The IS*1111* Family Members IS*4321* and IS*5075* Have Subterminal Inverted Repeats and Target the Terminal Inverted Repeats of Tn*21* Family Transposons

Sally R. Partridge[1,2] and Ruth M. Hall[1]*

CSIRO Molecular Science, Riverside Life Sciences Centre, North Ryde, New South Wales 2113,[1] and Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109,[2] Australia

IS*5075* and IS*4321* are closely related (93.1% identical) members of the IS*1111* family that target a specific position in the 38-bp terminal inverted repeats of Tn*21* family transposons and that are inserted in only one orientation. They are 1,327 bp long and have identical ends consisting of short inverted repeats of 12 bp with an additional 7 bp (TAATGAG) or 6 bp (AATGAG) to the left of the left inverted repeats and 3 bp (AGA) or 4 bp (AGAT) to the right of the right inverted repeat. Circular forms of IS*5075* and IS*4321* in which the inverted repeats are separated by abutting terminal sequences (AGATAATGAG) were detected. A similar circular product was found for the related ISPa11. Transposition of IS*4321* into the 38-bp target site was detected, but a flanking duplication was not generated. The precisely reconstituted target site was also identified. Over 50 members of the IS*1111* family were identified. They encode related transposases, have related inverted repeats, and include related bases that lie outside these inverted repeats. In some, the flanking bases number 5 or 6 on the left and 4 or 3 on the right. Specific target sites were found for several of these insertion sequence (IS) elements. IS*1111* family members therefore differ from the majority of IS elements, which are characterized by terminal inverted repeats and a target site duplication, and from members of the related IS*110* family, which do not have obvious inverted repeats near their termini.

The majority of mobile elements classified as insertion sequences (IS) are bounded by short terminal inverted repeats (TIR) and encode a transposase that includes a DDE motif (3, 20). Transposition generates a short duplication of the target sequence; as a result, IS are frequently found flanked by short (2- to 12-bp) direct repeats whose length is characteristic of the particular IS. However, other types of small mobile elements that do not share these features are also classified as IS. Members of one group, represented by IS*91*, do not have TIR, encode a distinct "transposase" with similarities to a family of replication proteins, and do not create a duplication of the target site (7). Movement of IS*91* and, by implication, other members of this family is believed to occur by a mechanism akin to rolling-circle replication (21).

The IS*110*/IS*1111* family is another atypical group. The founding member, IS*110*, and other original members of this group are not bounded by TIR and do not create a duplication of the target site (2, 3, 13). Subsequently, a distinct group of IS that have transposases that are closely related to the IS*1111* transposase (14) but that are only more distantly related to the IS*110* transposase were identified; these IS have short inverted repeats (IR) (15). In one classification (3), the IS*110* family includes these two different groups of IS, on the basis of the rather limited similarities between their transposases. However, it has been suggested that a separate IS*1111* family is needed (15), because there are distinct groups of transposases typified by IS*110* and IS*1111* and because the related IR ob-

served at the presumptive termini of members of the IS*1111* family are not found in IS*110* family members.

Members of this family encode transposases that were not initially thought to include a DDE motif. However, the Piv protein, which catalyzes site-specific inversion of a segment of the *Moraxella lacunata* genome, was also shown to be related to the IS*110* family transposases, with a small number of amino acid residues completely conserved in all proteins (16). This finding and the fact that some IS in this group are known to have preferred targets and to not produce a target site duplication led to the notion that this family of proteins represented a novel type of site-specific recombinase, rather than a classical DDE transposase (see reference 3). However, mutation of groups of conserved residues in the Piv protein was recently shown to lead to a loss of inversion activity without a loss of DNA binding, and a potential DDD (or DED) motif that appears to correspond to the DDE motif of classical transposases was identified (33).

IS*4321* is a member of the IS*1111* family that was first found in the multidrug resistance plasmid R751 (32). In R751, two variants of IS*4321* that are 96.8% identical and designated IS*4321*L and IS*4321*R are found at the boundaries of a putative cryptic compound transposon named Tn*4321*. The sequence surrounding IS*4321*L is derived from the *mer* (mercury resistance) end of Tn*501*, and IS*4321*L lies within one of the two 38-bp TIR found in this fragment (Fig. 1A). Closely related 38-bp TIR are normally found at the outer ends of transposons that belong to the Tn*21*/Tn*501* transposon family (9, 10), and IS*4321*R is found at the same position as IS*4321*L within a truncated (30-bp) copy of this 38-bp TIR.

While examining the flanking regions of a cluster of antibiotic resistance genes in a multidrug resistance plasmid from a

* Corresponding author. Present address: School of Molecular and Microbial Biosciences, Biochemistry and Microbiology Building G08, University of Sydney, New South Wales 2006, Australia. Phone: 61-2-93516014. Fax: 61-2-93514571. E-mail: ruth.hall@mmb.usyd.edu.au.
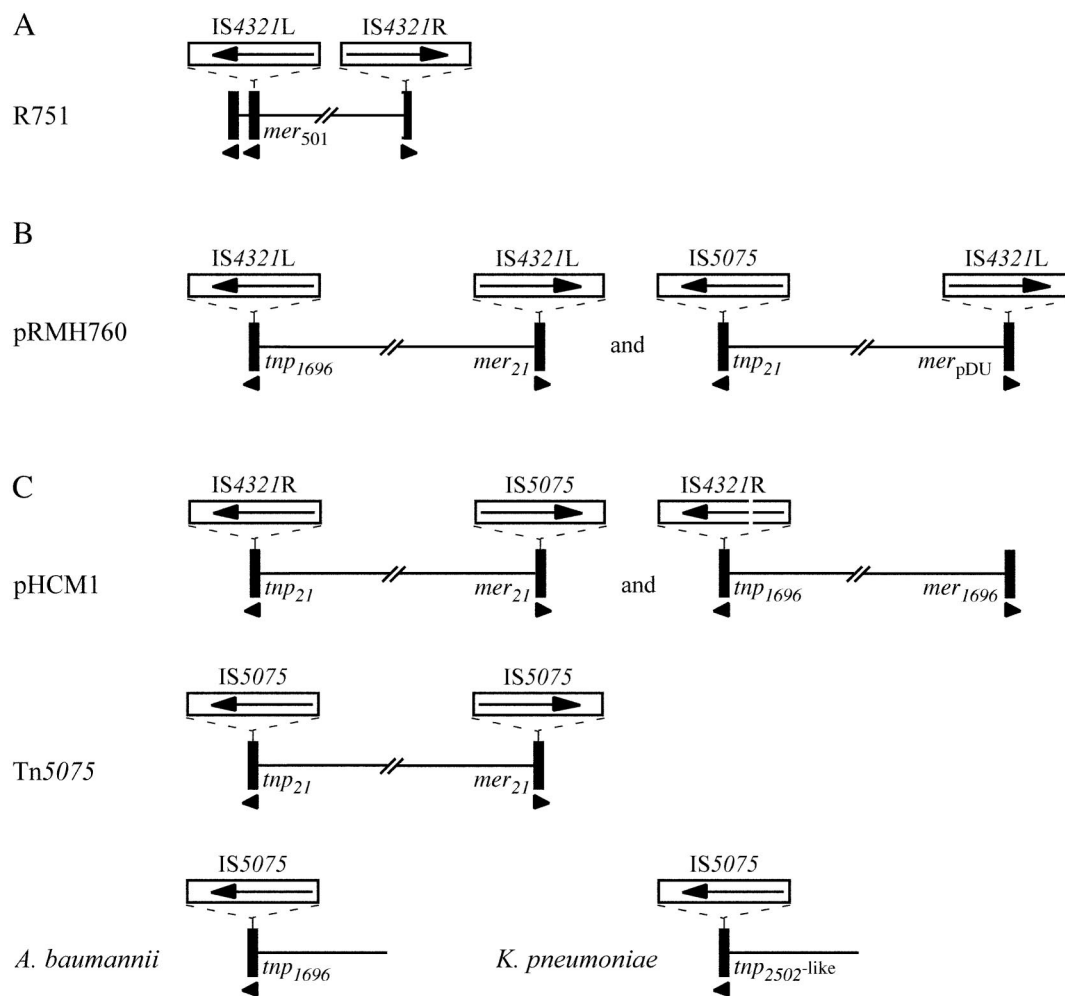
FIG. 1. Locations of known copies of IS*4321* and IS*5075*. The 38-bp TIR of Tn*21* family transposons are represented by black bars; arrowheads below the bars indicate their orientations. The identities of the adjacent sequences are indicated; e.g., tnp_{21} indicates the *tnp* end of Tn*21*, mer_{pDU} indicates one of the *mer* regions in plasmid pDU1358, and so forth. The IS*4321* and IS*5075* elements are represented by open boxes; arrows in the boxes indicate the positions and directions of the transposase genes. The plasmid, transposon, or organism where the IS were found are indicated on the left. Sequences are as follows (GenBank accession numbers): R751 from *Enterobacter aerogenes* (U67194), pRMH760 from *K. pneumoniae* (AY123253, AY242531, AY242532, and AY242533), pHCM1 from *S. enterica* serovar Typhi (AL513383), Tn*5075* from *E. coli* (AF457211), *A. baumannii* (AY196695), and *K. pneumoniae* (NC_002941).

clinical *Klebsiella pneumoniae* strain (27), we identified three copies of IS*4321*L and one copy of a closely related IS that was recently designated IS*5075* (6). Here, we describe the properties of these and related IS and use this information to better define the general features of the IS*1111* family.

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture media. *Escherichia coli* DH5α (*supE44* Δ*lacU169* [φ80 *lacZ*ΔM15] *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*) was used to propagate plasmids. pRMH760 is from a *K. pneumoniae* clinical strain isolated at Royal North Shore Hospital, Sydney, New South Wales, Australia, in 1997 (27). pRMH777 is a subclone of pRMH760 that includes IS*4321*L in the TIR at the *tnp* end of Tn*1696* (TIR$_{tnp1696}$) (27). pRMH901, which contains IS*5075* in the TIR at the *tnp* end of Tn*21* (TIR$_{tnp21}$), and other subclones of pRMH760 were recovered as described previously (27). The strain for transposition experiments was constructed by transforming pACYC184::Tn*21* (4) and pRMH777 (27) into UB1637 (F⁻ *his lys trp recA56 rpsL*) (4). *Pseudomonas aeruginosa* strain PAO222 is an auxotrophic derivative of PAO1 (11). Bacteria were routinely cultured at 37°C in Luria-Bertani medium or on Luria-Bertani agar, both containing ampicillin at 100 μg ml⁻¹.

Preparation of DNA. Plasmid DNA was isolated by using an alkaline lysis method (1) or purified for sequencing by using a Wizard maxiprep kit (Promega). Restriction enzymes were used in accordance with the manufacturers' instructions. Fragments were separated by electrophoresis on 1 to 2% (wt/vol) agarose gels and visualized by staining with ethidium bromide. *P. aeruginosa* PAO222 template DNA for PCR was prepared by suspending a single colony in 1 ml of Tris-buffered saline, pelleting the cells, resuspending them in 100 μl of Tris-EDTA, and boiling the mixture for 10 min. PCR products were purified for sequencing from Tris-acetate-EDTA–agarose gels by using an UltraClean DNA purification kit (Mo Bio Laboratories Inc.). *Eco*RI-digested bacteriophage SPP1 DNA (GeneWorks) and Hyperladder IV (Bioline) were used as size markers.

PCR amplification. The primers used are listed in Table 1. Amplification reactions were carried out with 50-μl volumes containing PCR Master Mix (Promega), additional MgCl₂ to 3 mM, 0.4 pmol of each primer, and approximately 20 ng of plasmid template or 1 μl of *P. aeruginosa* DNA. Reaction conditions were generally 94°C for 3 min; 35 to 40 cycles of 94°C for 1 min, 50 to 57°C (depending on the melting temperature of the primers) for 1 min, and 72°C for 30 s to 2 min (depending on the lengths of the expected products); and a final incubation at 72°C for 5 min. To obtain sufficient template for sequencing, some products were reamplified after purification of the relevant band.

TABLE 1. PCR primers

Primer	Sequence (5′-3′)	Location	GenBank accession no.	Positions
RH329	CAATCTGGTGACGGCGG	Tn*1696 tnp*	AY123253	1706–1690
RH337	GATAGATTCAAGCAAGC	IS*4321* and IS*5075*	AY123253	209–193
RH339	GTACGCCAGCAGTCGTG	IS*4321*	AY123253	701–685
RH385	GATTCGCCTCCTCTTCAACTATG	pRMH760		
RH407	GC**T**GCTTCACA**A**ATAGC[a]	IS*4321* and IS*5075*	AY123253	1120–1104
RH408	CCCTGGCACATTGAACC	IS*4321*	AY123253	562–578
RH429	CTCCTCCAAGCCAGTTACC	pACYC184	X06403	1269–1237
RH430	CGATGACTCGCTATTGCAG	Tn*21 tnp*	AF071413	169–151
RH432	GTGACTGGCGATGCTGTCG	pACYC184	X06403	1712–1693
RH435	CATTTCATCAGGGCTCG	ISPa11	AE004697	3428–3412
RH436	CAGACCAATGCGACTAAGG	ISPa11	AE004697	4581–4599

[a] T in bold type is A in IS*4321*; A in bold type is G in IS*5075*.

Sequencing. Automated sequencing was performed at the sequencing facility at the Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia, with an ABI-PRISM 377 sequencer and a Big Dye system. The DNA sequences of both strands of each IS element were determined from pRMH760 subclones and some PCR products by using appropriate primers. DNA sequences were assembled by using MacVector6.5 and AssemblyLIGN (Oxford Molecular). Sequences of PCR products were determined for both strands by using the amplification primers.

Sequence analysis. GenBank searches were performed by using the BLASTN and FastA programs available through WebANGIS (Australian National Genomic Information Service) or through the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/). Programs in the Genetics Computer Group Wisconsin Package, version 8.1.0, were used via WAG (WebANGIS GCG) to align and analyze DNA sequences. Unpublished sequence data for bacterial genomes was obtained from the Welcome Trust Sanger Institute (http://www.sanger.ac.uk/Projects/Microbes/), The Institute for Genomic Research (http://www.tigr.org), the U.S. Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/JGI_microbial/html/), and the Department of Microbiology at the University of Illinois (http://www.salmonella.org). All of the IS described here have been assigned names, and their sequences have been submitted to the IS Finder database (http://www-is.biotoul.fr/is.html).

Nucleotide sequence accession numbers. The sequence data for the three copies of IS*4321* and the one copy of IS*5075* have been submitted to the GenBank database under accession numbers AY123253, AY242531, AY242532, and AY242533.

RESULTS

IS*5075* and IS*4321* target the 38-bp TIR of Tn*501*/Tn*21* family members. The sequence of the regions surrounding the multidrug resistance cluster of pRMH760 revealed the presence of two mercury resistance transposons. One has a transposition (*tnp*) module identical to that of Tn*1696* (26, 27) and a *mer* module derived from Tn*21* (17), and the other has a Tn*21 tnp* module and a *mer* module nearly identical to that found in pDU1358 (8, 23). Each of the four 38-bp TIR from these transposons is interrupted by a 1,327-bp insertion sequence that is related to IS*4321* (Fig. 1B). Three are variants of IS*4321*L (99.7 to 99.9% identity), while the fourth is a related (93.1% identical) but distinct element. Searches of sequences deposited in the GenBank database revealed that the novel IS differs by no more than 7 nucleotides (<1% difference) from unnamed IS in multidrug resistance plasmid pHCM1 found in *Salmonella enterica* serovar Typhi strain CT18 (GenBank accession no. AL513383) (25) and in sequences derived from an *Acinetobacter baumannii* strain (AY196695) and a *K. pneumoniae* strain (NC_002941). The DNA sequence also differs at only five and seven positions from two additional copies, named IS*5075*, found in a mercury resistance transposon, Tn*5075*, in a plasmid from a preantibiotic-era *E. coli* strain (GenBank accession no. AF457211) (6); the IS in pRMH760 therefore has been named IS*5075*. The putative transposase encoded by IS*5075* is 98.5% identical to the transposase encoded by IS*4321*L and 65.9 and 69.8% identical, respectively, to those encoded by IS*1328* (GenBank accession no. Z48244) (29) and an IS*1328*-like element (AF326777) (18), which are the next closest relatives (Table 2).

In pRMH760, all copies of IS*4321*L and IS*5075* are found in locations equivalent to those of IS*4321*L and IS*4321*R in R751. All of these IS are at the same position and in the same orientation in the 38-bp TIR of the transposons (Fig. 1 and 2A). In addition to the single copy of IS*5075*, the pHCM1 sequence contains a complete copy of IS*4321*R and a second copy with a short internal deletion (Fig. 1C). These IS, as well as the additional examples shown in Fig. 1, are also in the same position and orientation in the TIR of transposons that appear to be derived from Tn*21* and Tn*1696* or a Tn*2052*-like transposon with related TIR. Additional partially sequenced copies of IS*4321*-like elements found in pDU1358 (GenBank accession no. M24940) (23), pCFF04 (X64523) (19), and R831b (U77087) (24) are flanked by the appropriate part of the 38-bp TIR. The IS-interrupted 38-bp TIR do not appear simply to have become distributed directly by vertical or horizontal transmission or to have become disseminated by homologous recombination, as the IS are found within three distinct variants of the 38 bp TIR and different versions of these IS are found in the same individual TIR variant (Fig. 1). For example, TIR*tnp21* contains IS*5075* in pRMH760 and Tn*5075* but IS*4321*R in pHCM1, and TIR*mer21* contains IS*4321*L in pRMH760, IS*4321*R in pHCM1, and IS*5075* in Tn*5075*. Therefore, it appears that these IS target a specific position in the 38-bp TIR of transposons that belong to the Tn*21*/Tn*501* family.

Boundaries of IS*4321* and IS*5075*. The fact that IS*5075* and IS*4321* are found within a known sequence permits their boundaries to be deduced. The 12-bp IR identified previously (32) are not located at the IS boundaries; they are separated from the termini by 6 or 7 bp on the left and 4 or 3 bp on the right (Fig. 2A). The assignment of a T residue to the right or left is not possible because the IS are located within a tract of four T residues in the 38-bp TIR. The left-hand terminal sequence (5′-AATGAG-3′ or TAATGAG) is not obviously related to the right-hand terminal sequence (5′-AGAT-3′ or

TABLE 2. IS related to IS*1111*

Name[a]	GenBank accession no.[b]	Species	Location[c]	No. of copies[d]	Length (bp)[e]	IRl sequence[f]	Protein[g] Length (amino acids)	% Identity to: IS*4321*L	% Identity to: IS*1111*A
IS*4321*	U67194	Various[h]	Plasmids[h]	1–3	1,327	ATGGTCACTCCC	334	100.0	39.8
IS*5075*	AF457211	Various[h]	Plasmids[h]	1 or 2	1,327	ATGGTCACTCCC	334	98.5	39.8
ISSfl8	AF326777	*Shigella flexneri*	GI	1	1,357	ATGGTCACCCCC	334	69.8	39.8
IS*1328*	Z48244	*Yersinia enterocolitica*	Chrom	4 or 5	1,343	ATGGTCACCCCC	334	65.9	36.5
ISPpu5	AJ233397	*Pseudomonas putida*	Chrom	1	>1,342	atggtcACCCCC	~340	~60.7	~38.2
ISPsy7	AE016867	*Pseudomonas syringae*	Chrom	10	1,348	ATGGTCAGCCC	340	50.0	38.8
ISVpa1	AP005088	*Vibrio parahaemolyticus*	Chrom	1	(1,318)	ATGGTACGCCC	338	44.6	35.5
ISPa11	AE004697	*Pseudomonas aeruginosa*	Chrom	6	1,372	ATGGACTCCTCCC	338	42.6	46.9
ISRsp4	AE000090	*Rhizobium* sp. strain NGR234	pNGR234a	2	1,398	GTGGACGGCTCCC	344	41.4	37.2
ISBj4	AP005952	*Bradyrhizobium japonicum*	Chrom	6	(1,576)	A<u>T</u>GGTCCGGCCGTGC	354	41.0	37.0
IS*1111*A	M80806	*Coxiella burnetii*	Chrom	20	1,374	ATGGACCCACCC	339	39.8	100.0
ISShsp1	AB091693	*Sphingomonas* sp.	—	1	1,404	GTGGACGGCTCCC	347	39.8	37.8
ISSfr1	AF229441	*Sinorhizobium fredii*	Plasmid	1	1,359	ATGTGGACGCCCCC	~343	~39.9	~39.8
ISCc4	AE005940	*Caulobacter crescentus*	Chrom	1	>1,174	gtggacggcccc	340	39.6	37.3
ISMlo1	AP003008	*Mesorhizobium loti*	GI	1	1,405	GTGGACGGCCTCC	347	38.0	37.8
IS*1383*	AF052750	*P. putida*	pPGH1	4	1,419	ATGGACGCCTCCC	341	37.8	37.5
ISMlo2	AP003008	*M. loti*	GI	2	1,402	GTGGACGGC<u>C</u>CCC	346	37.7	37.8
ISMlo3	AP003008	*M. loti*	GI	1	(1,405)	GTGGATGGCTCCC	342	37.5	37.3
ISAtu1	AF242881	*Agrobacterium tumefaciens*	Ti plasmid	1	>1,092	atggacgcctccc	~343	~36.7	~40.1
ISKpn2	1919464A	*Klebsiella pneumoniae*	Plasmid	1	1,308	ATGGACTACCT	~338	~35.7	~42.1
ISSo14	AE015826	*Shewanella oneidensis*	Chrom	2	(1,439)	ATGGTCACCTCGG	344	34.2	37.9
ISPpu11	AE016775	*P. putida*	Chrom	2	1,438	ATGGACTCGCCC	352	32.9	35.7
ISUnCu1	AY139602	Uncultured	pSp7	1	1,381	ATGGACTC**C**CCC	~344	~30.5	~36.0
IS*1618*	AF074611	*Yersinia pestis*	pMT1	1	1,365	AAGGG**A**CTTCCCC	340	38.9	38.6
ISPsy16	AF141883	*P. syringae*	pAV511	2	1,461	AAGGG**A**CTTCCCC	344	37.9	38.3
IS*5708*	AJ010745	*K. pneumoniae*	—	1	1,364	AAGGG**A**CTTCCCC	340	37.8	39.2
ISSm1	AF389912	*Serratia marcescens*	—	1	1,369	AAGGG**A**CTTCCCC	340	37.7	38.3
IS*1492*	AJ288908	*Pseudomonas fluorescens*	pL6.5	2	(1,459)	AAGGG**A**CTTCCCC	344	36.6	38.6

[a] Unnamed IS were assigned a name in consultation with the IS database (http://www-is.biotoul.fr/is.html). ISRsp4 is equivalent to y4pF in earlier compilations.

[b] Only one GenBank accession number is listed for IS when multiple sequences are available. For ISKpn2, a protein database number is given, and the nucleotide sequence is given in reference 37.

[c] Plasmid names are given when possible; chrom, chromosomal location; GI, chromosomal genomic island; —, unknown location.

[d] Information on copy number was obtained from references named in the relevant database entry or from an analysis of complete genomes.

[e] Lengths in parentheses were calculated assuming that the terminal sequences total 10 bp or 9 bp for IS*1492*; a "greater-than" symbol indicates that the IS is incomplete.

[f] IRr is identical to IRl except for the following: IRr was not found for sequences shown in lowercase type; in IRr of IS*4321*L and IS*4321*R of R751, the underlined C is T; in ISPpu5, the first 6 nucleotides of IRl are missing due to insertion of ISPpu4 (36), and the equivalent sequence of IRr is shown in lowercase type; in IRr of ISBj4_1 and ISBj4_5, the underlined T is C; in IRr of ISMlo2 and ISUnCu1, the underlined C is T; A residues in boldtype are missing from "lower case type;" the corresponding IRr.

[g] Approximate values are due to the presence of frameshifts.

[h] See the legend to Fig. 1.

AGA). We conclude that IS*5075* and IS*4321* are atypical IS with subterminal IR.

To confirm these boundaries, we sought evidence for circular intermediates of IS*5075* and IS*4321*. pRMH777 (27) and pRMH901, subclones of pRMH760 containing a single IS, were used as templates with outward-facing PCR primers, which should detect only circular intermediates or tandem copies of the IS (Fig. 3A). A product of the correct size (433 bp) was detected, and digestion with restriction enzymes that should either cut both products or distinguish between the IS*5075* and IS*4321* products yielded fragments of the predicted sizes (data not shown). The sequences of the PCR products (Fig. 2B) revealed that the right and left 12-bp IR (IRr and IRl, respectively) were separated by a 10-bp sequence, IRr-AGATAATGAG-IRl, that comprises the abutted terminal sequences of the IS predicted above. This configuration indicates that the IS can be excised precisely and circularized and is consistent with the conclusion that the outlying bases are an intrinsic part of the IS. A circular form of IS*1383*, which also belongs to the IS*1111* family, was identified previously (22), and it also includes an additional 10-bp sequence between the IR of the IS

(Fig. 2B). Although these authors concluded that the intervening 10 bp consisted of 5 bp from each side of the IS, the same 10-bp sequence could also consist of 7 or 6 bp from the left end and 3 or 4 bp from the right end, as described here for IS*4321* and IS*5075*, as two bases (bold type in Fig. 2B) could be derived from either the left or the right end of IS*1383*.

Movement of IS*4321* and reconstitution of the target sequence. To search for RecA-independent movement of IS*4321*, pACYC184::Tn*21* (4), which has two 38-bp TIR that could act as targets for IS*4321*, was used. Movement of IS*4321* from pRMH777, which contains IS*4321* in TIR$_{tnp1696}$ to pACYC18::Tn*21* was detected by PCR with pairs of primers (Fig. 3 and Table 1) that can detect IS*4321* in either TIR of Tn*21* but not in its original location. PCR fragments of the expected sizes were obtained, and their sequences (Fig. 3B) revealed that the TIR of Tn*21* now contained a copy of IS*4321*. The IS was inserted precisely, with no duplication of the target sequence.

To detect a reconstituted 38-bp target sequence, which would arise if the flanking sequence could be rejoined after the IS was excised, plasmid DNA from a strain containing
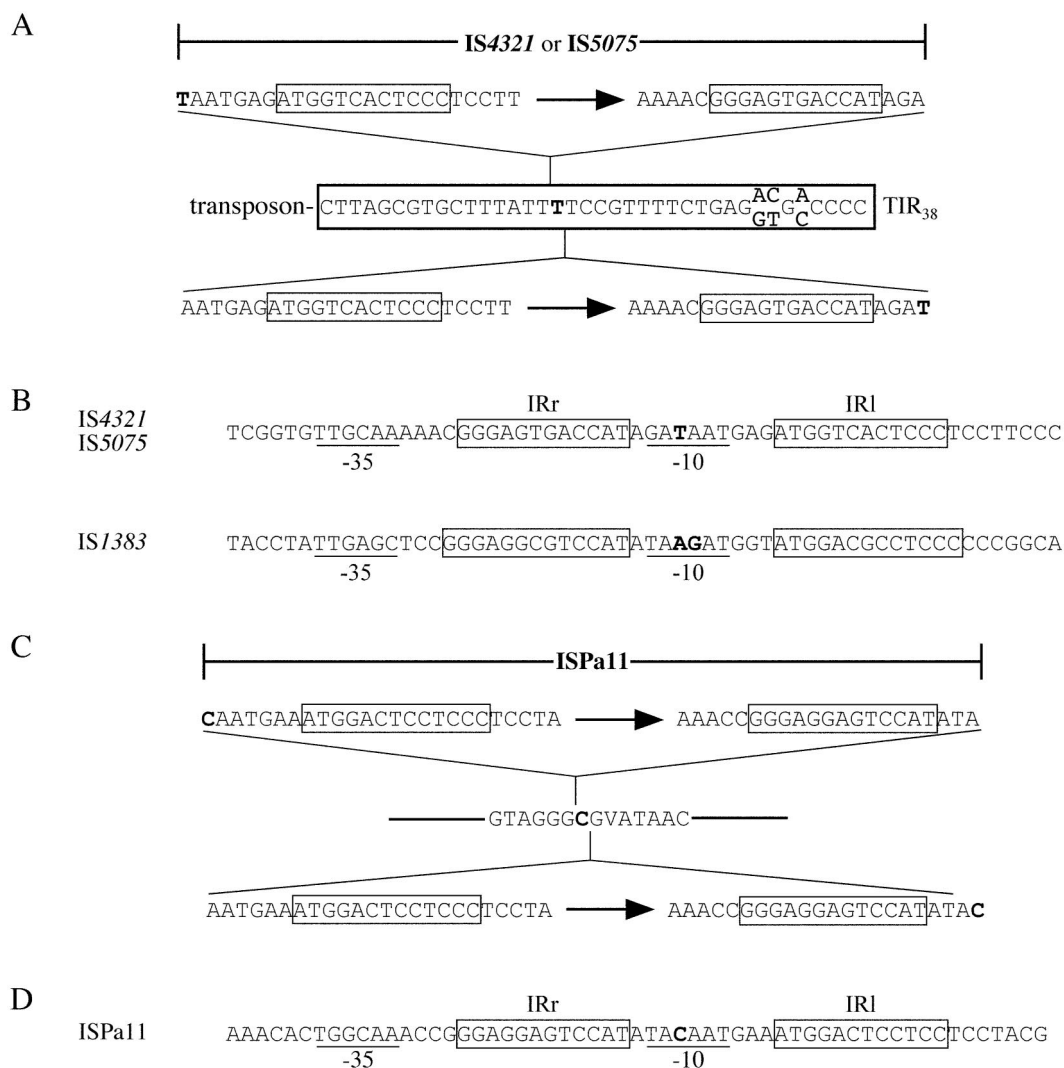
A

IS*4321* or IS*5075*

TAATGAG ATGGTCACTCCC TCCTT ➤ AAAAC GGGAGTGACCAT AGA

transposon- CTTAGCGTGCTTTATT**T**TCCGTTTTCTGAG $^{AC\ A}_{GT\ C}$ G CCCC TIR$_{38}$

AATGAG ATGGTCACTCCC TCCTT ➤ AAAAC GGGAGTGACCAT AGA**T**

B

IS*4321*
IS*5075*

IRr   IRl

TCGGTGTTGCAAAAAC GGGAGTGACCAT AGA**T**AATGAG ATGGTCACTCCC TCCTTCCC
            -35   -10

IS*1383*

TACCTATTGAGCTCC GGGAGGCGTCCAT ATA**AG**ATGGT ATGGACGCCTCCC CCCGGCA
         -35   -10

C

ISPa11

CAATGAA ATGGACTCCTCCC TCCTA ➤ AAACC GGGAGGAGTCCAT ATA

GTAGGG**C**GVATAAC

AATGAA ATGGACTCCTCCC TCCTA ➤ AAACC GGGAGGAGTCCAT ATA**C**

D

ISPa11

IRr   IRl

AAACACTGGCAAACCG GGAGGAGTCCAT ATA**C**AATGAA ATGGACTCCTCC TCCTACG
            -35   -10

FIG. 2. Boundaries of IS*4321*, IS*5075*, and ISPa11. (A and C) Extents of IS indicated by bars. The IR are boxed (thin lines); arrows indicate the directions of the transposase genes. The base that may originate from either the left or the right end of the IS and the base in the target adjacent to which the IS is inserted are shown in bold type. The 38-bp transposon TIR is boxed (thick lines), and residues that differ in the two common alternate types are shown. (B and D) Sequences of circular intermediates. IRr and IRl are boxed, and potential −35 and −10 regions are underlined. The sequence of the IS*1383* circular intermediate (22) is shown for comparison.

pRMH777 was also analyzed with a pair of PCR primers flanking the TIR$_{tmp1696}$ in pRMH777 (Fig. 3A). In addition to a fragment of 1.74 kb, corresponding to the TIR containing IS*4321*, a faint band of the expected size (415 bp) following the loss of IS*4321* was seen. The sequence of this band (Fig. 3B) revealed that IS*4321*, including the extra bases outside the IR, had been excised precisely to recreate the 38-bp TIR.

***P. aeruginosa* ISPa11.** The *P. aeruginosa* PAO1 chromosome (31) contains six complete copies of a gene encoding a transposase (open reading frames PA0445 in GenBank accession no. AE004481-2, PA2319 in AE004658, PA2690 in AE004697, PA3434 in AE004764, PA3993 in AE004817, and PA4797 in AE004892-3) that is 42.6% identical to that of IS*4321*, but the boundaries of the IS have not been reported. Analysis of the DNA sequences of these six regions revealed that they are identical over a stretch of about 1,380 bp that includes 13-bp IR related to those of IS*4321* (8 of 12 bp identical) and IS*1383* (12 of 13 bp identical) but extends for several base pairs to the

left and right (Fig. 2C); these findings suggest that the IS also may target a specific sequence. Searches with the flanking sequence, reconstituted by assuming that the left- and right-hand terminal sequences have lengths equivalent to those of IS*4321*, IS*5075*, and IS*1383*, revealed over 50 uninterrupted copies of the target sequence in the PAO1 chromosome, permitting the ends to be deduced. The circular intermediate of ISPa11 was detected by PCR with DNA from strain PAO222, an auxotrophic derivative of PAO1 (11), and appropriate primers (Table 1). The sequence of the junction (Fig. 2D) confirmed that the IS does indeed include 6 or 7 bases beyond IRl and 4 or 3 bases beyond IRr. It appears that the presence of at least one residue of ambiguous origin (left or right) is a feature of this family of IS, although the identity of the residue(s) is not conserved.

Promoter in the circular intermediate. A promoter consisting of a −35 region located just inside the right-hand end of the IS and a −10 region created by the fusion of the right- and
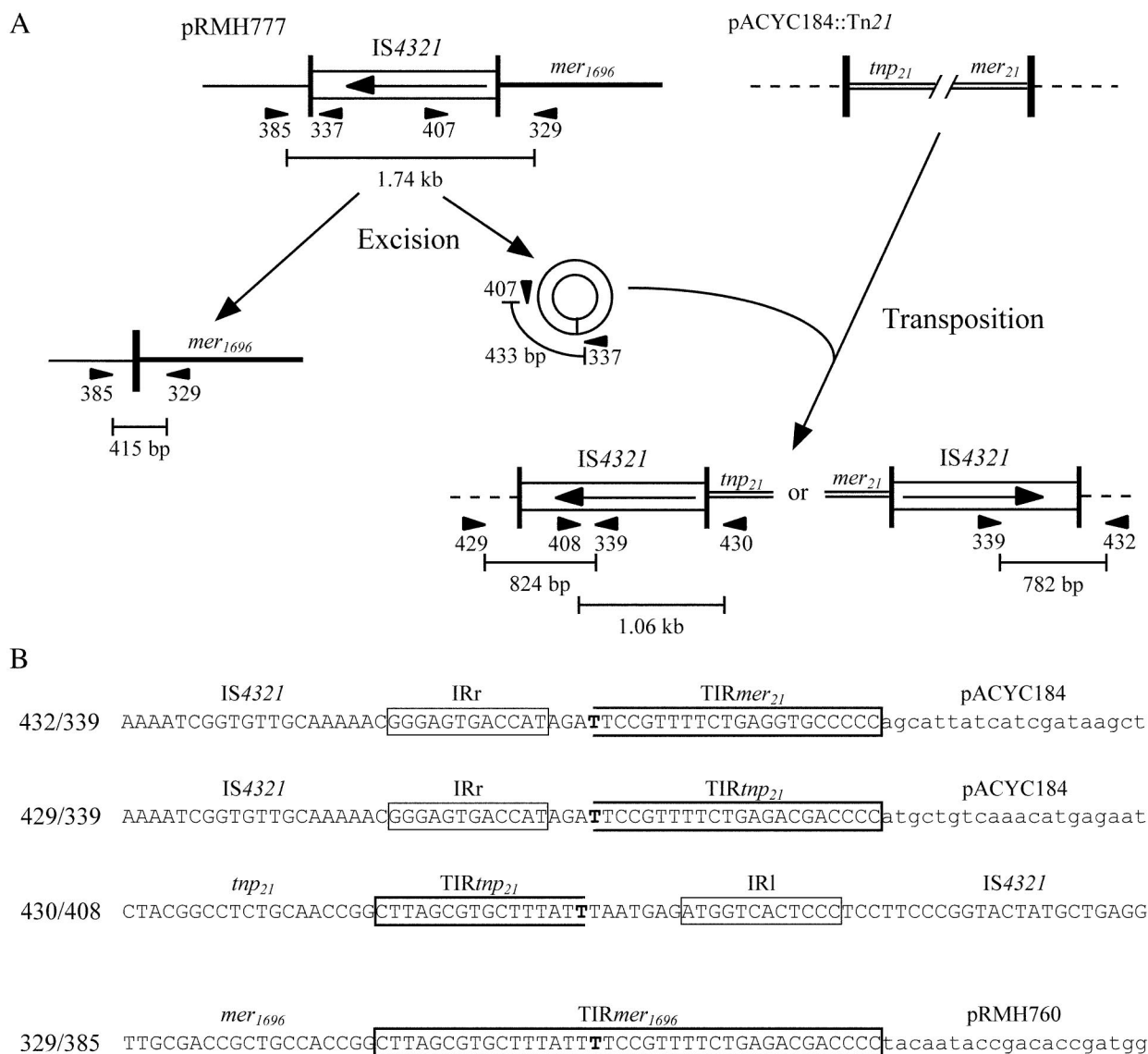
A

B

FIG. 3. Movement of IS4321. (A) Substrates used to detect the movement of IS4321. Features of the IS, TIR, and adjacent regions are as described in the legend to Fig. 1. pACYC184 and pRMH777 backbone sequences are indicated by broken and thin lines, respectively. The positions and orientations of primers used are indicated by arrowheads with numbers to identify the RH primer (Table 1). The extents of PCR products are indicated by bars; their lengths are shown below the bars. (B) Sequences of PCR products. PCR products are named by the primer pairs giving rise to them. TIR, IR, and ambiguous bases are indicated as described in the legend to Fig. 2; the name of each region is shown above the sequence. The plasmid backbone sequence is shown in lowercase type.

left-hand terminal sequences was detected in the circular form of IS1383 (22). Equivalent −35 and −10 regions are present in the circular intermediates of IS5075 and IS4321 and the *P. aeruginosa* IS detected here (Fig. 2). Equivalent promoters are also generated by fusion of the deduced ends of several additional IS1111 and IS4321 relatives (see below), indicating that this promoter in the circular intermediate is also a characteristic of this family. It is likely to play a role in the movement of IS1111 family members, perhaps transiently increasing the expression of the transposase. As the expression of the transposase presumably is required first to form the circular intermediate, we searched for expression signals upstream of the transposase gene but within the left end of the IS. A potential

promoter was found in IS4321, IS5075, and the related IS1328 and ISSfl8 but not in IS1383 and ISPa11, which may rely on promoters in the adjacent sequence for the initial expression of the transposase.

IS1111 subgroup of the IS1111 family. A small number of IS with transposases related to that of IS1111 and with IR at or near their ends were identified previously (3, 15, 29, 39). Searches of DNA sequences in GenBank revealed many additional predicted proteins related to the IS1111 and IS4321 transposases. Those that are complete and exhibit more than 30% identity to the IS4321 and IS1111 transposases, together with a few examples that include frameshifts in the transposase gene (either errors or mutations) that can be tentatively cor-

rected by using the sequence of their next closest relative as a guide, were selected for further analysis (Table 2). The proteins range in length from 334 to 354 amino acids, and alignment of the sequences (Fig. 4) revealed a small number of completely conserved amino acids and several more that are conserved in all but a few of these proteins. At several additional positions, only one of two alternative amino acids (e.g., D or E, R or K, and S or T) is present. Searches of available incomplete bacterial genome sequences revealed many more IS*1111* and IS*4321* relatives (Table 3), and an alignment (data not shown) of all 46 complete transposase sequences that did not require frameshift correction revealed the same conserved residues (residues conserved in 45 or 46 of these 46 sequences are indicated by asterisks in Fig. 4). In the N-terminal domain, the conserved amino acids include five acidic residues, D residues at positions 15, 100, and 103 and E residues at positions 60 and 156, and four basic residues, K18, R41, R70, and K98. The D15, E60, and D100 or D103 residues correspond to the DDD residues in the Piv protein that are thought to correspond to a transposase DDE motif (16). Additional blocks of conserved residues are found in the C-terminal portions at positions equivalent to stretches of conserved residues identified in previous alignments of Piv with IS*110*/IS*1111* family transposases (16).

As IR were identified previously for only a small number of these IS (15, 29, 32, 36), the DNA sequences surrounding the transposase genes were examined for the presence of IR, which were found in all but five that appear to lack one end (Tables 2 and 3). Most of the IR were 11 to 13 bp long, and the left- and right-hand copies were, with a few exceptions, identical. The sequences were related to the IR of IS*4321*, ISPa11, and IS*1383*; variants of 5'-ATGGACGCCCCC(C)-3', with occasional T residues in the run of C residues, were most common. A number of subgroups could be identified based on the variation in the first eight residues of the IR. For example, the IS*1328* group, to which IS*4321* belongs, begins with ATGGT CAC/G, and all of the IS with transposases closely related to the IS*4321* transposase (>50% identity) belong to this group (Tables 2 and 3). However, the IR are not completely predictive, as the IR of ISSo14 also belong in the IS*1328* group, whereas the transposase does not. The groups that have IR beginning with GTGGACGG or ATGAACGC also include mostly, but not exclusively, IS whose transposases are most closely related to one another. An additional group, typified by IS*1618*, had distinctive IR with AAGGGA**C**TTCCCC on the left and a shorter sequence lacking the A residue in bold type on the right (Tables 2 and 3). Six IS fell into this group, and their transposases are also closely related (>70% identity) and include characteristic differences, an insertion of 3 amino acids at positions 91 to 93 and a deletion of 5 amino acids at positions 168 to 172 (Fig. 4). One IS, ISSfr1, had unusual IR that included the sequence GTGGACGG preceded by AT.

It was previously assumed that, by analogy with the majority of IS, the IR represent the termini, and similarities noted between the sequences flanking the IR for IS*4321* and IS*1328* (32) and for several IS belonging to the IS*1111* group (15, 38, 39) were believed to indicate an element of target site specificity. As the terminal sequences could account for part of these regions, we sought evidence for their presence in additional IS. Where both IR could be identified, the flanking

sequences (up to 100 bp from each side) were joined, assuming that 10 extra bases are part of the IS and that there is no duplication of the target sequence. The resulting sequence was used to search for copies of the uninterrupted target sequence that would allow deduction of the ends of the IS. Sufficiently closely related sequences were found in many cases, and in all of them it appeared that terminal sequences were present. The sequence of the junction in each of the predicted circular intermediates is shown in Fig. 5, with the base(s) that could be derived from either the right- or the left-hand side indicated in bold type. The terminal sequences are related and, when joined, most include a potential −10 region at a suitable distance (16 to 18 bp) from a −35 region in the right-hand end. An additional feature of the left-hand terminal sequence is that it includes a tetranucleotide similar to the outer 4 bp of the IR. This tetranucleotide is ATGG or ATGA in most examples but TTGG for the group with IR beginning with GTGG. All but a few of the examples found include at least 1 bp that could have originated from either the right- or the left-hand end (bold type in Fig. 5), confirming that this feature is a general property of IS*1111* family members.

The exceptions included two IS, ISYen1 and ISEch3, where multiple copies of the IS and of the uninterrupted target sequence were found in the same organism. For these IS, the seventh base to the left of IRl was not always identical to the fourth base to the right of IRr. ISYen1 has T on the left and T, C, or G on the right (T shown in Fig. 5), and the three copies of ISEch3 have T with T, C with C (shown in Fig. 5), and T with C. In both examples, the residue found at the equivalent position in the target sequence is also variable. In the IS of GenBank accession no. AY139602, the T shown is found to the left and the C found to the right is identical to a residue in the target sequence. For a second group, ISAzvi4 and ISPsy7, multiple copies of the target sequence or more than one target sequence was found, and individual comparisons predicted a total of either 10 bp (shown in Fig. 5) or 9 bp (the T shown in bold type in Fig. 5 was missing) of terminal sequences. A −10 region was not found in the latter.

The length of the left-hand terminal sequence was not absolutely conserved. Some IS included 7 or 6 bp on the left and 3 or 4 bp on the right, as was found experimentally for IS*4321*, ISPa11, and IS*1383*; others included 6 or 5 bp on the left and 3 or 4 bp on the right. The latter group would have 9 bp separating the IR in the circular intermediate and would include four members of the IS*1618* group described above (Fig. 5). Alignment of the circular junctions reveals that, for the IS*1618* group, the first residue of IRl (which is 1 bp longer than IRr) may be equivalent to the 10th base of the 10-bp spacer sequence, raising the possibility that the two IR may be recognized in slightly different manners. Additional IS with predicted terminal sequences totaling 9 bp fell into two groups. For ISPpu11 and ISSfr1, a −10 promoter region was formed, and there was at least 1 base of ambiguous origin. For IS*1328* and ISAzvi2, these features were not apparent, and indeed close relatives have terminal sequences totaling 10 bp. It is possible that in these IS, there is an error in one of the sequences compared. This notion is supported by the fact that the terminal sequences of the closest relative are very similar but include an additional base at the position normally occupied by the one that could be derived from either end. Fur-

```
              1        *    *                                 *                  *           *        *                90
IS1111B   .....MKDIK ILGVDIAKDV FQLCGIDEWG K.VIYTRRVK RAQYVSTVAS LKVGC.VVME ACGGANHWYR TFMGMGIPTQ LISPQHVKPY
ISSo14    ......MNYK LISIDLAKTV FQVAAFNPDN T.IAFNRKIN RPALLDMLRQ CEPTL.VVME ACYSANYWGR TIEKLGHTVK LIPARMVKAM
IS1328    .....MPNVT LIGIDLGKNS FHIHCQEKHG N.TLLRKRFS RIQLTQFLAT CPPCI.VAME SCAGAHFMAR HISQLGHQVK LISPQFVRPF
ISSf18    .....MQNIT LIGIDLGKHS FHIHCQDKSG K.ALLRKKFT RTKLMQFLAT CPPSV.VVME ACAGAHFMAR CISDFGHETK LISPQFVRPF
IS4321L   ....MENIA LIGIDLGKHS FHIHCQDKSG K.AVYRKKFT RPKLIEFLAT CPATT.IAME ACGGSHFMAR KLEELGFFPK LISPQFVRPF
ISPsy7    ...MKNGLK LLGIDIGKNS FHLIGHDAGG HQVL.RKCFN RNRLLQFAAQ IEPCT.IAME SCGGSHWLAR KLASLGRHVK LIAPQHVKPY
ISVpa1    .....MSAIK VIGIDLGKST FHLVGHDYSG REQY.RHKLS RSKLLQFISV HEPTL.IAME ACCGAHWLAR KCQEYGHQVK LIPPQYVKPY
ISBj4_1   MSQTPSTAIA VIGIDLGKNS FHVVGHDTRG T.VVLRQKWS RGQVEARLAN MPPCL.IGME ACVGAHHLSR KLASLGHDAR LMPAKYVRPY
IS5708    ......MTIT TVGIDLAKNV FAVHCIDQNG KAVLVKPKVS RAALSELIAG LPPCV.IGME ACSGAHYWAR LFRQYGHDVR LMAPKFVSPY
IS1618    ......MTIT TVGIDLAKNV FAVHCVDQNG KTVLVKPKVS RAALPELIAG LPPCV.IGME ACSGAHYWAR LFREYGHEPR LMAAKFVSPY
IS1492    ......MSVI TVGIDLAKTV FAVHGVNEAG RDQLVPLIAQ LPPCL.IGME ACSGAHHWAR VFVQHGHTVR LMAPKLVAPY
ISMlo1    ....VEQII RIGMDTSKHV FQLHGVNAAE E.VVLRKKMR RKEMVAFFER AAPTV.VALE ACGGSHHWAR LLEALGHHVK LIPPQYVKPY
ISMlo2    ....VEQII RIGIDTSKHV FQLHGVNAAE E.PVLRKKLR RKEVVALFEK LSPTV.VAIE ACGGSHHWAR LLQSFGHEVK LIPPQYVKPY
ISShsp1   ....VEQIT RIGMDTSKHV FQLHGVNAAE KAVL.RKKLT RREMVRFFEK LPPTV.VAIE ACGGSHHWAR LLGSLGHDVR LIAPQLAKPY
ISRsp4    ....VEQI RIGMDTSKSV FQLHGVNAVE Q.PILRKKLS RREMVKFFEK TPPTI.IALE ACGGSHHWAR LLSSFGHEVK LIAPQLAKPY
ISCc4     ......MMTT TIGLDLAKSV FQVHGVSSEG Q.VTVRRKLR RGELVSFFRD LPPCL.IGME ACATAHYWAR ELSALGHEVR LMPPSYVKAY
ISMlo3    ......MQIT TIGLDLAKSV FQFYAVDAEG H.VVVRKALR RAQILSFFAK LPCCL.VGME ACGTSHHWAR ELTKLGHEVR LMPPAYVKPY
ISPal1    ......MNLS RIGIDLAKQV FQVHGVDRHE H.VVCRRQLK RAQVRDFFRQ LPPCL.VAME ACGSAHYWAR ELRELGHIVR LIAPQFVKPY
IS1383    ......MKHS VIGMDIAKKV FQLHTVDQCT GRIE.RIKLR RDEVLAFFAN YPTSL.VAIE ACGSAHWWAR QLQQQGHEVR LLAPRSVRPF
ISPpu11   .....VELCT TICVDLAKQV FQLAGEDATG RVIYEDRIKS RQAFHDFLLR LPTTVAVLME CGPGAQAWAR LLQAKCNPVR ILPAQRVAEH
Consensus --------- --G-D--K-- F--------- ---------- R--------- ---------E AC------R -----GH--r L-p--V---
                                                                                       k            a

             91      *  *      ** *      *                        *                             * *              180
IS1111B   ...VKSNKND RNDAQAIAPA ASRASMRFVR GKTVEQQDVQ ALLKVRDRLV KSRTALINEI RGLLQEYGLT MARGAKRFYE E.LPLILASE
ISSo14    ...LVGNKND ANDAVAIGET AGRPKVSLLR PKTVEQQDIQ SLLRIRERLV ENRTANCNQL RGFLAEYGLI VAKTRRQLML A.IPLILEDA
IS1328    ...VKSNKND FIDAAAICEA TSRPSMHFVT PRTEDQQAMS ALHRVRDALV RERVKATNQM HAFLLEFGIS MPRGIAVIK. R.LATVPE..
ISSf18    ...VKSNKND FVDAAAICEA ASRPSMRFVQ PRTETQQAMR ALHRVRESLI RDKVKTTNQI HGFLLEFGIS LPIGDAVIK. R.LSLVLA..
IS4321L   ...VKSNKND FVDAAAICEA ASRPSMRFVQ PRTESQQAMR ALHRVRESLV QDKVRTTNQL HAFLLEFGIS VPRGAAVIS. R.LSTLLE..
ISPsy7    ...VTGNKND FIDAAAICEA ASRPRTHYVA VKTVEQQVLS VQHRLRESLV CHRTEAINRV HGFLLEFGIA LPATKAAIA. KVPE.LIDDS
ISVpa1    ...VKAHKND FIDAAAIABA STRENMRFVS PKSEQAQVGV VIRRVRTGYI RERTATMNRI GSMLIEFGIS FPRGHANMK. KLFQ.WLADN
ISBj4_1   ...SKGPKND FNDABAIABA VQRPTMKFVA TKTAEQLDLQ ALHRVERLV SQRTGIINQI RAFMLERGIA VRQGIGFLRT E.LPTILATR
IS5708    RMAGKTGKND AADAIAICEA VQRPHMRFVP VKDESQQAMQ CLHRTRQGFI QERTATYNRL RGLISEFGVI APQSTDA... ..LRHVVSDQ
IS1618    HMAGKSGKND AADAQAICEA VRRPHMRFVP VKDESQQAMQ CLHRTRQGFI EEKTATYNRL RGLISEFGVI APQSTDA... ..LRRMVSEQ
IS1492    RMSGKRGKND AADAAAICEA VTRPSMRFVP VKDEHQQATL CLHRTRQGFI EERTATYNRL RGLLSEFGVV LPQSPER... ..LRREIGPY
ISMlo1    ...VKRGKND AADABALCEA VSRPTMRFVR VKTTDNQAAL MLVGMRDRLV RNRTQLANAI RGYAAEFGLI AATGTSKVR. SLLERIATDE
ISMlo2    ...VKRGKND TADABALCEA MSRPTMRFVP VKTAKQQAAL MLVGLRDRLI RNRTQLANAI RGYAAEYGLI AARGMCKIE. PLLERIAADK
ISShsp1   ...VKRGKND AADABALCEA VSRPTMRFVP VKTAEQQAAL MLVGVRERLL QKRTQLVNAI RGYGAEFGIT VATGACRVV. PLLERIAADE
ISRsp4    ...VKRGKND AADABALCEA MSRPTMRFVP MKTAAQQAAL MLVGMRERLI RNRTQLANAI RGFAMEFGIV AAKGMCRIE. ALLERIAADP
ISCc4     ...LRRQKND AADAQAICEA VTRPTMRFVP VKSAERQAIL TLHRTRELLV RQKTTLIRAI RGHCAEFGLV VPQGARRV.F DLLNLMREAS
ISMlo3    ...VKRGKTD AGDABAICEA VTRPTMRFVP VKSADQQAAL ALHRTRDLLV RQRTQLVNML RGLLAEFGIE IARGLHHA.L RLAGRLSNGE
ISPal1    ...VKGDKND AHDARAICEA ASRESMRYVP VKSAEQQAVQ SMHRVRSRLV RARTALQNEV RGLLGEFGLI ATRRGRAATM ALLETVMATE
IS1383    ...VLRNKTD AADAQAIWTA VQQPGACLVA IKQADQQAIL SLHRIRSQLL KFRIMQGSNGL RGLLYEFGIV LPEGYTQLSK A.VPEAFANA
ISPpu11   ...RSGAKND RNDCYAILRA GRDTSIASIP IKSTTALAIQ ALHRIRRGNI KRHTALGNQI RGLLLEHGVS MPQGDAAINK H.VPQALEDA
Consensus -------KnD --DA-AI--A ---P----V- -k---QQ--- -----R---- -------N-- rg---E-G-- ---------- ----------
                 t             r                                   h                 ha

             181                                                       *                  *         ** *     270
IS1111B   AVGLTPRMKR VLNCLYTELL NRDEAIGDYE EELKAVAKAN EDCQRVQSIP GVGYLTALSV YASVGDIHQF HRSEQLSAFI GLVPEQHSSG
ISSo14    DNELSTVARG FIQRLYEFQL YLNEQIDQVA QDILALAQAM PQYPLLLTIP GVCPIVASTL LASINEVNDF KNGRQLAAWL GLTPSQYSSG
IS1328    EHELPPYLVR LLTRLHQHYG YLCEQIEEIE RELKNHLADD ETAQRLLTIP GIGTITASLL ATKLGDGKNY LSSRDFGAST GLVPRQYSTG
ISSf18    EHEIPEYLSR LLVRLHTHYF YLVEQIAELE SELSQSINAD DTAQRIMTIP GVCPITASTL SSQLGDGKQF SCSRDFAAST GLVPRQYSTG
IS4321L   DNSLPLYLSQ LLLKLQQHYH YLVEQIKDLE SQLRKKLDED EIGQRLLSIP CVGTLTASTI STEIGDGKQV ASSRDFAAAT GLVPRQYSTG
ISPsy7    RSDLPLQFKR VMEHLLEDIH RLDTEIKACD AQIKQQLAQD DAGTRLMTIP GIGPITASAF VADLGDASNF RASRDLSAFL GLVPRQFSTG
ISVpa1    KEPIPPLLVR ELQNQLDYYN QLNERIKEQD RKIEKLSSED ELYTLLQTIP GVCPMTASCC LSCVGKPSDF SNGRNFAAWI GLVPRQYSTG
ISBj4_1   TDALSPRMLC VIEELAGDWR RLDQRIDRLS GEIEALARQD QACSRLMTVP GIGPIISSAM VAAIGTGDVF SKGRDFGAWL GLVEKQISTG
IS5708    KETLPVQVRL CIDDLLTHVT NIEEKIAEYD RVLSRVAKTD HRSQQLMELK GVGPTTASAL VASIGNAHDF KNGRQLAAWL GLTPSQYSSG
IS1618    KNSLPFQVQQ CIDDLLEHVD RIEANIADYD RILSRMAKTD HRSQRLMELK GVGPTTACAL VASIGNAHDF KNGRQLAAWL GLTPSQYSSG
IS1492    LDSLPGWAKR CMSELLEHTS HIEDRLVELD RAISQIARED DRSRRLLQLR GVGSTTASAL VASIGDDARQF RNGRQVAAWL GLTPGQYSSG
ISMlo1    T..LPELARE LFGLHGREYD RLEDEIEKVE ARLMAWHRTN ECSRRLTRIP GVCPIGALMM VMKTPAPETF RSGRHFAAWL GLTPKDHSTA
ISMlo2    M..LPDLARE LFALHAKEYA QLQTQLKDVD AKLMAWHRAE ACSQRLAQIP GVCPIGASLL VMKTPAPETF RSARHFAAWL GLTPKDHSTA
ISShsp1   T..IPELARE LFALQGKEYA QLQCEIERLD AKLMEWHRND ECSQRLAKIP GVCPIGATLL MMKTPDPHLF KSGRDFAAWM GLTPKDHSTA
ISRsp4    S..LPELAQD LFALHGQEYR ELACQIKTLD EKLMKLHRAD ECSKRLAEIP GVCPIGASLL LMKTPDPRMF KSGRDFAAWI GLTPKDHSTA
ISCc4     SSSLPETAMV TMEMLGAQME ALALQIRTLE RRLLAWHREN QTSQRLATIP GVCIITSTAL AASISDPRLF RSGREFAAFL GLVPRQNSSG
ISMlo3    APEVPPLARL VVKVLADQIG TIQLQLARLE KELLAWHRSN ELSQRLATIP GICIVSATAL AASVIEPERF RSGRQFAASL GLTPIQNCSG
ISPal1    PAPLPAPMGE LLRELKDELQ TLEARIARLE RQIQAHVRGD ARIQRLLAVE GICPISASAV AASAGDARQF RTGRQFAAWL GLVPRQHSTG
IS1383    EHRVPSMLLD SLRDQWTRII QLDEEIRKIE LRLKQCLRES SDCQKIAEIP GIGLLTATAA VASLGDATTF HSGRQFAAWL GLVPRQTGTG
ISPpu11   SLPLPDLLRE LLDELLMDWL SLGERITALS RRLEATAQED KVAQRLITIR GVCPIIATAI VAKQTDPSRF ASGRMYSAFF GIVPDQHSSG
Consensus ---------- ---------- ---------- ---------- ---------- ---------- GvG---A--- --------- ---R---A-- GLvP-Q--tG
                                                                              i            y              t
                                                                                                           s

             271       *        *                       *                     *        *                  361
IS1111B   NKEVLLGISK RGNVMLRTLL IHGARALLRH V.KNKTDKKS LLLKALIERR GMNRACVALA NKNARIIWAL LTRQETYRCG A.........
ISSo14    TTNRLGKITK RGNQTLRRLL IHGARTVVNW C.DKKDDSLS LWLQRLKQRQ HVCKVTVALA NKMARMVFVV MSRGQAFDMN KACKLAT...
IS1328    GKSTLMGISK RGDKNLRRFL VQCARVYMQR L.EYQSGRLA EWWNGQLTRH HSNVVACALA NKLARIAWVV TTQGTVFSK. ..........
ISSf18    GKSTLLGISK RGDKNLRRLL VQCARSFIMR L.ERKHGKLA EWWREQLNKK HSNVVTCALA NKLARIAWAI TTQKKEYQA. ..........
IS4321L   GRTTLLGISK RGNKKIRTLL VQCARVFIQK L.EHQSGKLA DWWRDLLCRK SNFVVTCALA NKLARIAWAL TARQQTYVA. ..........
ISPsy7    GRPVLLGISK RGDRHLRTLL VQGSRALMIR IERRDN.ALG QWVRDLLARK HPNKVACALA NKLARIVWAV LRRGSTYNAS MI........
ISVpa1    CKQRMLGISK RGNKKLRELF VHAARAVMWR EETAER.YFG SWLLDLKRRK PFNVALVALA NKLARIAWSV MTSHKPFELR V.........
ISBj4_1   DRTILGKISK RGNRYLRVLF VQAAWVLVR IKNWERYGLR SWIEAAKKRL HHNVLALALA NKLARIAAW LAKGRAFELT RTNDAGVRPA
IS5708    GKSKLGRITK AGDAYLRTLL VQGARSVMVG A.ENRTDPFS RWVCSLIGRR GYWRAVVAIA AKNARLCWAS LHYGDDFRLY STT.......
IS1618    GKSKLGRITK AGDSYLRTLL VQGARSVLIG A.DKRTDSFS RWVCTLVERR GYWRAVVAIA AKNARLCWAS LHYGDDFRLY SAS.......
IS1492    GKQRLGSITK AGDAYLRTLL VLGARSVMAN L.GDKQDTFS RWIRNLIERR GYWRAAVAIA AKNARMAAV LHYGDTFKLE QVEPTGA...
ISMlo1    CRVRLGVITR AGDENLRSVL VAGATAVLRH VREDRGKNAS PWLLDMLNRK PPKLVAVALA NKLARVAWKL MVTGERYNPN AALSAYGQAA
ISMlo2    CRVRLGVITR AGDEALRSVL VAGATAVIRH VRRGRG.AVS PWLVDLLRKK PPKLAAVALA NKLARIAWKL MVSGEIYGAK PMPPASARAA
ISShsp1   CKTRHLGVITR AGDELLRSTL VVGATAVVQH ARNGRGRNLT PWLIGLIGRK SPKLVAVAVA NKAARVAWKL MVSGERYQCV APPALTAVAA
ISRsp4    CKVRLGVITR AGDEMLRSIL VVGATSLLQQ VRTGRSRHAS AWLMGLLQRK RPKLVAVALA NKLARIAWKL MTSGESYRQA EGQAQTS...
ISCc4     GKDRLGRITK KGDGYLRRLL VVGATSVIRR ARTGVT.PSA GWWQSLLERR PSRVVTVAMA NKLARIAWAV MTRGESYRAM ATV.......
ISMlo3    GRERMGRISR MGDRYLRMLL VVGMTSLVRR ARIQPA.SVD PRLSAMLARK PARLVTVAAA NRTARVAAWV MTRGGTFRAP ISAAA.....
ISPal1    GQQRLGNISK RGDTYLRTLL IHGARAVVRC CANKTD.ARS RLLQGLLQRR PANVVAVALA NKNARIILWAL LSRETCYRPG ..........
IS1383    GRVRQLGLSK RGDTYLRSLL MHGARSILNR ......GHRS EWVERLLTRR PYNVAVAAVA NKLARIIWAV LAKGSTYRAE LFTACPTGH.
ISPpu11   NKIRLGKMSK RGDGYIRSLM IQGAHSVLSQ LRPDSDQPDD RRLLRWLSRL GRKEAAIRLA NRNLRIIWAL LQSDQVYQRK PNEKPEAAMS L
Consensus G---L--I-k -Gd-LR--- ---A------ ---------- -W------R- -------A-A -K-AR--W-- ---------- ----------
               r       n
```

6378

TABLE 3. IS related to IS*1111* in genome sequencing projects

Name[a]	Species	No. of copies[b]	Length (bp)[c]	IRI sequence[d]	Protein Length (amino acids)	Protein % Identity to: IS*4321*L	Protein % Identity to: IS*1111*A	Source[e]
ISAfe1	*Acidithiobacillus ferrooxidans*	2	(1,424)	ATGGATGCCTCC	342	41.5	41.6	TIGR
ISAzvi4	*Azotobacter vinelandii*	1	1,351	ATGGTCACCCCC	341	62.3	38.4	JGI
ISAzvi2		1	1,396	ATGGACTCCCCCC	338	42.3	47.3	
ISAzvi3		1	(1,424)	ATGGACGCCTCCC	341	37.8	41.6	
ISBcen1	*Burkholderia cenocepacia*	1	(1,357)	ATGGTCACCCCC	341	65.3	41.1	SI
ISBcen2		1	(1,363)	ATGGTCAGCCCGAT	341	57.8	41.1	
ISBcen3		2	1,413	ATGGACGCCCCCG	344	44.0	43.0	
ISBcen4		1	>1,085	atggacgcctcc	341	39.1	39.2	
ISBcen5		3	(1,426)	ATGAACGCGTCCC	339	37.7	39.6	
ISBfun1	*Burkholderia fungorum*	1	(1,466)	AAGGG**A**CTTCCCC	341	37.8	38.6	JGI
ISBfun2		1	>1,125	atgaacgcgtccc	332	36.6	39.2	
ISCps1	*Colwellia psychroerythraea*	1	(1,347)	ATGGACTCCTCC	342	33.2	36.6	TIGR
ISEch4	*Erwinia chrysanthemi*	4	(1,354)	ATGGTCACCCCC	334	65.6	37.4	TIGR
ISEch3		3	1,363	ATGGACGCTCC	339	40.4	41.2	
ISMmg1	*Magnetospirillum magnetotacticum*	1	1,359	ATGGCCCGCCCC	342	42.3	41.1	JGI
ISMmg2		1	(1,381)	GTGGATGGCTCCCGC	342	37.7	41.7	
ISPfl1	*Pseudomonas fluorescens*	1	1,342	ATGGTCAGCCCA	340	48.5	39.9	JGI
ISRle1	*Rhizobium leguminosarum*	1	1,409	ATGGCGCCC	349	38.9	39.2	SI
ISSaen1	*Salmonella enterica* serovar Enteritidis	1	1,294	ATGGACTAC<u>T</u>TCC	340	35.4	42.3	UIUC
ISSdy1	*Shigella dysenteriae*	1	>1,120	aagggacatc	338	35.7	39.5	SI
ISSso1	*Shigella sonnei*	2	1,443	ATGAACGCATCCC	339	37.8	38.7	SI
ISSpo1	*Silicibacter pomeroyi*	1	(1,363)	ATGTACCGGCTGCT	352	39.8	36.9	TIGR
ISYen1	*Yersinia enterocolitica*	7	1,376	ATGGACGCTCC	339	40.4	40.1	SI

[a] IS were assigned names as in Table 2.

[b] When contigs have not yet been assembled, the number of copies is not definite.

[c] Lengths in parentheses were calculated assuming that the terminal sequences total 10 bp or 9 bp for ISBfun1; a "greater-than" symbol indicates that the IS is incomplete.

[d] IRr is identical to IRl except for the following: IRr was not found for sequences shown in lowercase type; in IRr of ISBfun1, the A in boldtype is missing; in IRr of ISSaen1, the underlined T is C.

[e] Preliminary sequence data were produced by the following: TIGR, The Institute for Genomic Research (can be obtained at http://www.tigr.org); JGI, U.S. Department of Energy Joint Genome Institute (can be obtained at http://www.jgi.doe.gov/JGI_microbial/html/); SI, Welcome Trust Sanger Institute (can be obtained at http://www.sanger.ac.uk/Projects/Microbes/); and UIUC, Department of Microbiology at the University of Illinois (can be obtained at http://www.salmonella.org).

thermore, if a 10-bp spacer is assumed, a −10 region is expected. For the remaining IS listed in Tables 2 and 3, an uninterrupted target sequence was not found, and the length of the element was calculated assuming that the terminal sequences total 10 bp or, for the IS*1618* group, 9 bp. Generally, the sequences of the termini of these IS were conserved, and a base of ambiguous origin was found in most of them.

Target site specificity. A few IS*1111* family members are known to have a preferred target site into which they insert in a preferred orientation. IS*1111* is found at one end of related stem-loop structures that were originally thought to form part of the IS (14), and all of the copies in the *Coxiella burnetii* genome (30) were found to be similarly located (Fig. 6A). Over 50 additional copies of this target sequence were identified in the *C. burnetii* genome. IS*1383* targets the 12-bp TIR of IS*1384* (22), and preferred targets for IS*4321* or IS*5075* and ISPa11 were also identified in this study (Fig. 2). However, there is no obvious relationship among the sequences of the targets for these IS, indicating that each group of closely related IS selects

a different target. For all of the remaining IS for which multiple copies have been sequenced, when the sequences adjacent to their ends were aligned, a potential target site sequence was found; for some, however, sequence identity extended for only a few base pairs on either side of the IS (Fig. 6C). For ISPpu11, ISPsy7, ISAzvi2, ISEch3, and ISYen1, the targets were longer, and multiple copies of them were found in the genome of the organism where the IS was found; these results indicate that genomic repeat sequences were used as targets. For four members of the IS*1618* group, the targets were a set of three 38-bp TIR that are closely related to one another and to the TIR of Tn*6901* (GenBank accession no. AP004237) and that are related to but distinct from the TIR of Tn*21* (Fig. 6B). The IS inserts at a position different from that recognized by IS*4321* and IS*5075*. Short regions of similarity to the target site region of the TIR were found adjacent to two additional members of the IS*1618* group (Fig. 6B). These findings suggest that preferred targets may be a general feature of the IS*1111* family but

FIG. 4. Alignment of IS*1111* family transposases. Twenty transposases from the IS shown in Table 2 are aligned. Only one example was included for pairs of proteins that were >90% identical (IS*4321*L and IS*5075*, IS*1618* and ISSm1, and IS*1492* and ISPsy16). An alignment of all 46 transposases without frameshifts from Tables 2 and 3 was used to define conserved amino acids. Residues conserved in all or all but 1 of the 46 sequences are indicated by asterisks above the sequence. Residues differing in 6 or fewer of the 46 sequences are shown as white on black and are indicated by uppercase letters below the sequence. Where only two alternative amino acids in 45 or 46 sequences are found, the two alternatives are indicated by lowercase letters below the sequence. The sequences used are from translations of nucleotide sequences identified by the GenBank accession numbers listed in Table 2.

```
                                    ┌────10 bp────┐
                          IRr       ├──────────┤        IRl
IS4321
IS5075      CGGTGTTGCAAAAAC[GGGAGTGACCAT]AGATAATGAG[ATGGTCACTCCC]TCCTTCCC-  40-ATG
                         -35                   -10

ISSfl8      CAGTGTTGCAAAAAC[GGGGGTGACCAT]AGATAATGAT[ATGGTCACCCCC]ACCTTAAC-  44-ATG

ISAzvi4     CACTATTGCAAAAAC[GGGGGTGACCAT]AGATAATGAG[ATGGTCACCCCC]ACACCCTG-  46-ATG

ISPfl1      CGTGATTGCAGGAGA[TGGGCTGACCAT]ACATAATGAG[ATGGTCAGCCCA]GTACGGGA-  53-ATG

ISPsy7      TGCGATTGCAGGAGGT[GGGCTGACCAT]ACATAATGAG[ATGGTCAGCCC]GATCCGAGA-  54-ATG


IS1383      ACCTATTGAGCTCC[GGGAGGCGTCCAT]ATAAGATGGT[ATGGACGCCTCCC]CCCGGCA-109-ATG

ISPa11      AACACTGGCAAACC[GGGAGGAGTCCAT]ATACAATGAA[ATGGACTCCTCCC]TCCTACG-  53-ATG

ISBcen3     CGAACCTGGCAAAA[CGGGGGCGTCCAT]GTATAATGGA[ATGGACGCCCCCG]CCCGGCA-  60-ATG

ISEch3      ACAGATTGCAAATGTG[GGAGCGTCCAT]ATACGATGAA[ATGGACGCTCC]AACTTGACG-  56-ATG

ISYen1      AACACTTGGTTTATTG[GGAGCGTCCAT]ATATAATGAA[ATGGACGCTCC]AACTTTTAC-  53-ATG

ISKpn2      AGTACTTGCAAAACGG[AGGTAGTCCAT]AGATAATGGG[ATGGACTACCT]ACTCCCGCT-  40-ATG

ISMmg1      GAGCCACTTGCGGAC[GGGGCGGGCCAT]ACGTTTATGG[ATGGCCCGCCCC]TCACCGGG-  51-ATG

IS1111      GATTGTTGACAACAA[GGGTGGGTCCAT]ATACAATGAA[ATGGACCCACCC]CTTAAAGA-  55-ATG
                                                 ↓
ISUnCu1     TGCTCACTACTTGTG[GGGAGAGTCCAT]ATATGCTGTA[ATGGACTCCCCC]GCACCACC-  44-ATG


ISRsp4      AAACACTTGCAAGC[GGGAGCCGTCCAC]ATATGTTGAG[GTGGACGGCTCCC]AACGGCA-  52-GTG

ISMlo1      ACATCTTGCAGGGA[GGAGGCCGTCCAC]ATATGTTGGT[GTGGACGGCCTCC]TGCACGG-  54-GTG

ISMlo2      CAATCTTGCGAGAA[GGGTGCCGTCCAC]ATATGTTGGG[GTGGACGGCCCCC]GCACGGC-  52-GTG

ISShsp1     CAGCCCTTGCAAGA[GGGAGCCGTCCAC]ATATGTTGGG[GTGGACGGCTCCC]AACGGCA-  52-GTG


                                    ┌────9 bp────┐
                                    ├──────────┤
IS5708      CGGCGTTGACTGTTC[GGGGAAGCCCTT]ATAGAATGG[AAGGGACTTCCCC]TAGATAT-  56-ATG

IS1618      CGACGTTGACATGCC[GGGGAAGCCCTT]ATAGAATGG[AAGGGACTTCCCC]TGATTAT-  57-ATG

ISPsy16     GCGCTTGCGCTCAAT[GGGGAAGCCCTT]GTAGAATGA[AAGGGACTTCCCC]GTCCAAG-  79-ATG

ISSm1       CGGCGTTGACATACC[GGGGAAGCCCTT]ATAGAATGG[AAGGGACTTCCCC]TGATTAT-  57-ATG


ISPpu11     GCAGCTAGACAGT[GTGGGCGAGTCCAT]ACATAATCG[ATGGACTCGCCCCC]GCCGAG-100-GTG

ISSfr1      AGCCCTTGCCATA[GGGGGCGTCCACAT]ATGATGATG[ATGTGGACGCCCCC]GCGCCG-  54-ATG

                                              ↓
IS1328      CGGTGTTGCAAAAAC[GGGGGTGACCAT]AGAAATGAG[ATGGTCACCCCC]ACCCTGTG-  40-ATG
                                              ↓
ISAzvi2     ATCACTGGCAAACC[GGGGGGAGTCCAT]ATAAATGAA[ATGGACTCCCCCC]TCCTACG-  59-ATG
```
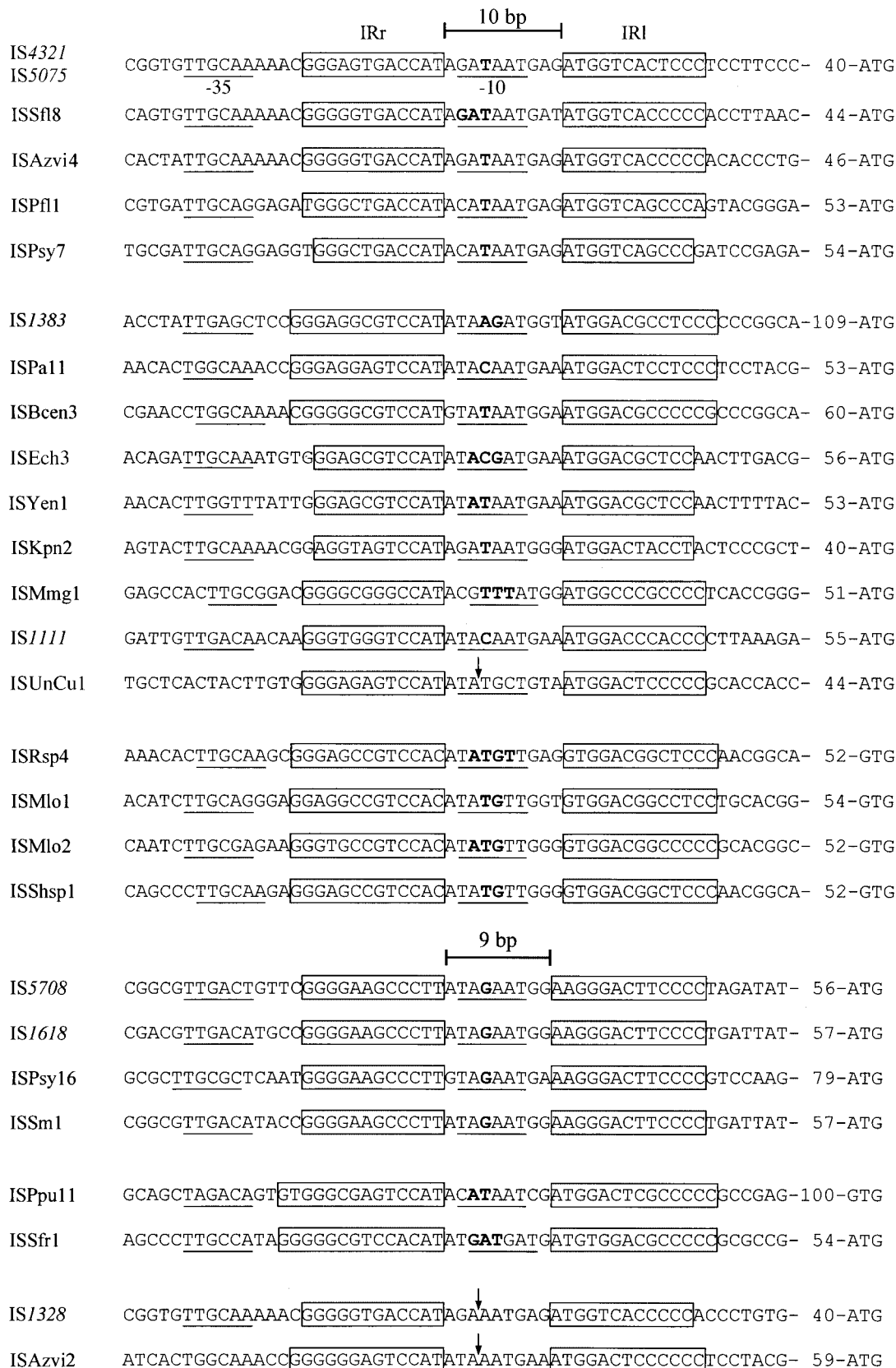
FIG. 5. Sequences of predicted circular forms of IS1111 family elements. The sequences shown are deduced from comparisons of sequences flanking the IS and uninterrupted target sequences. Features are indicated as described in the legend to Fig. 2B. The distances to the initiation codons of the transposase genes are also shown. Where the comparison predicts the origins of all central bases, a vertical arrow indicates the junction of the left and right ends.
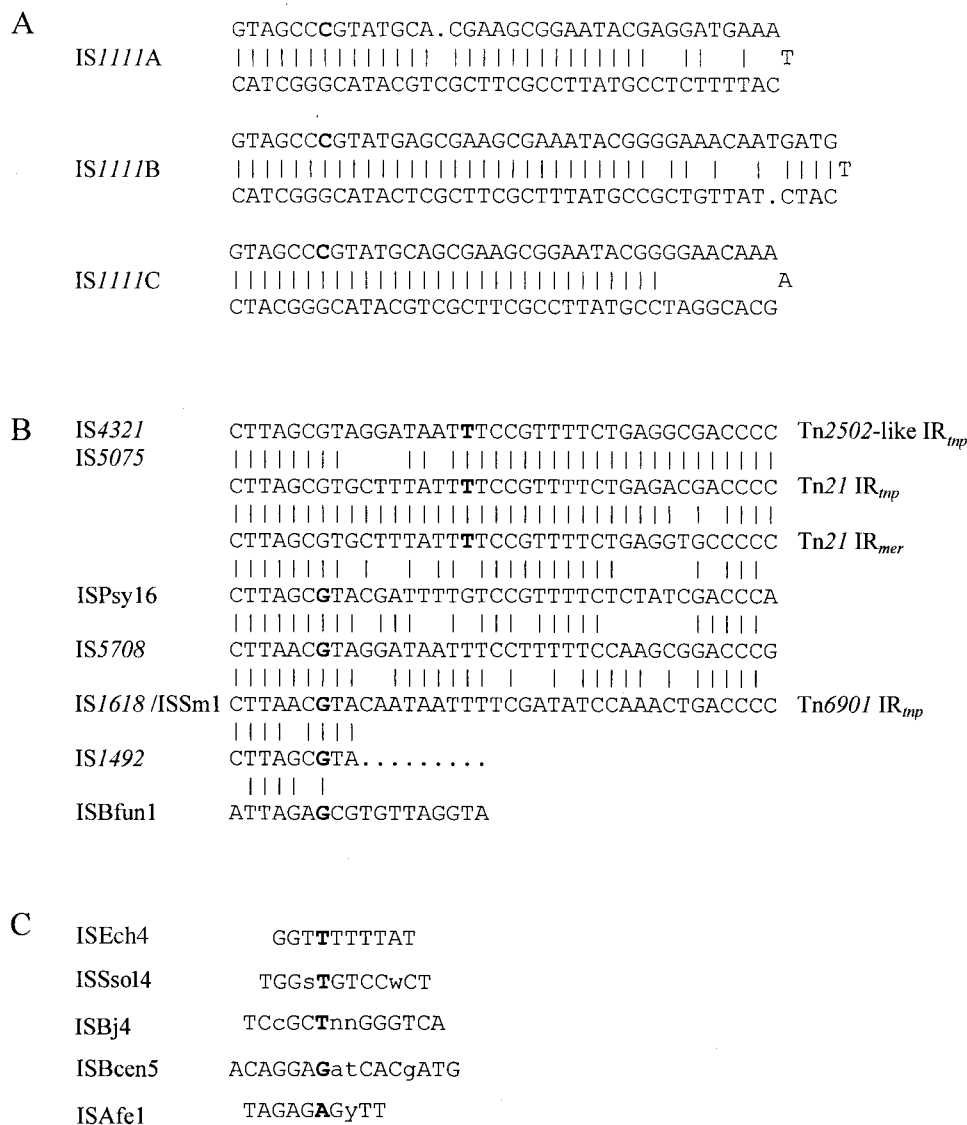
A

IS1111A
```
       GTAGCCCGTATGCA.CGAAGCGGAATACGAGGATGAAA
       ||||||||||||||| |||||||||||||||  |   |   T
       CATCGGGCATACGTCGCTTCGCCTTATGCCTCTTTTAC
```

IS1111B
```
       GTAGCCCGTATGAGCGAAGCGAAATACGGGGAAACAATGATG
       |||||||||||||||||||||||||||||||||| ||  |  | ||||T
       CATCGGGCATACTCGCTTCGCTTTATGCCGCTGTTAT.CTAC
```

IS1111C
```
       GTAGCCCGTATGCAGCGAAGCGGAATACGGGGAACAAA
       ||||||||||||||||||||||||||||||||||||||          A
       CTACGGGCATACGTCGCTTCGCCTTATGCCTAGGCACG
```

B

IS4321
IS5075
```
       CTTAGCGTAGGATAATTTCCGTTTTCTGAGGCGACCCC   Tn2502-like IRtmp
       ||||||||        ||  ||||||||||||||||||||||||
       CTTAGCGTGCTTTATTTTCCGTTTTCTGAGACGACCCC   Tn21 IRtmp
       |||||||||||||||||||||||||||||||| |  ||||
       CTTAGCGTGCTTTATTTTCCGTTTTCTGAGGTGCCCCC   Tn21 IRmer
       |||||||| |   |  || ||||||||||   |  |||
```

ISPsy16
```
       CTTAGCGTACGATTTTGTCCGTTTTCTCTATCGACCCA
       ||||||||| |||   |  |||  |||||      |||||
```

IS5708
```
       CTTAACGTAGGATAATTTCCTTTTTCCAAGCGGACCCG
       |||||||||  |||||||  |  |  |||||  |  |||||
```

IS1618 /ISSm1 `CTTAACGTACAATAATTTTCGATATCCAAACTGACCCC   Tn6901 IRtmp`
```
               ||||  ||||
```

IS1492
```
       CTTAGCGTA.........
         ||||  |
```

ISBfun1
```
       ATTAGAGCGTGTTAGGTA
```

C

ISEch4
```
       GGTTTTTTAT
```

ISSsol4
```
       TGGsTGTCCwCT
```

ISBj4
```
       TCcGCTnnGGGTCA
```

ISBcen5
```
       ACAGGAGatCACgATG
```

ISAfe1
```
       TAGAGAGyTT
```

FIG. 6. Targets for additional IS. The IS is inserted to the right or left of the base in bold type. (A) Stem-loop structures targeted by IS1111. (B) The 38-bp TIR and related sequences recognized by members of the IS1618 group are compared to the 38-bp TIR recognized by Tn4321 and Tn5075. Dots indicate unknown sequences. (C) Sequences targeted by additional family members. Bases in lowercase letters are not completely conserved.

that a variety of different sequences are used as targets for individual IS.

DISCUSSION

IS4321 and IS5075 were found to differ from members of other IS families in that they include short extensions beyond the IR and do not create a direct duplication of the target site. The terminal sequences, an additional 7 or 6 bp on the left end and 3 or 4 bp on the right end, clearly are an intrinsic part of the IS and move with it. The ambiguity in assigning 1 bp to the left- or right-hand end of the IS arises from the fact that this nucleotide is the same as the one targeted by the IS in the sequence that it enters. When the ends are brought together to form the circular intermediate (Fig. 2D), a −10 region is

formed and, as it is separated by 17 bp from a −35 region found near the right-hand end of the IS, a promoter that may be important for the expression of the transposase is formed. Sequenced copies of IS4321 and IS5075 are always at the same position in the TIR of Tn21 family transposons, and transposition into this location was detected. We conclude that the IS targets part or all of the TIR sequence. A possible consequence of disruption of the TIR caused by IS4321 or IS5075 is that the transposon is inactivated; consistent with this notion, the movement of Tn5705 could not be detected (6).

Most of the properties of IS4321 and IS5075 appear to be general features of IS1111 family members, as many additional family members were found to include short IR, usually 11 to 13 bp, and terminal sequences (Tables 2 and 3 and Fig. 5). The IR sequences fall into a number of distinguishable groups but

clearly are related. In most of them, the sequence at the outer ends of the IR is related to the consensus sequence 5′-ATGG ACGC-3′ and is followed by a short tract of predominantly C residues. Evidence for terminal sequences comes from the sequence of the circular intermediate of ISPa11 and from comparisons to uninterrupted target sequences, which were identified in many examples. The terminal sequences generally total 10 bp, although in a few examples they appear to be 1 bp shorter. For members of the IS1618 group, where 9 bp separate the IR in the predicted circular intermediate, other distinguishing features are shared: a characteristic 3-amino-acid insertion and a 5-amino-acid deletion are present in the transposase, the outer ends of the IR are distinctive (5′-AAGGG and CCCTT-3′), and IRl includes 1 bp more than IRr. The latter feature may be related to the apparent 9-bp spacer if IRl is recognized in a manner different from that of IRr, i.e., as AGGG or as AAGG. A comparison of the 9-bp sequence to the 10-bp consensus sequence seen in other circular intermediates reveals that the 10th base is missing and that the 1st base of IRl may be equivalent to it. For some of the remaining examples where a 9-bp spacer was predicted, the 4th base to the right (or the 7th base to the left) is missing, and close relatives have 10-bp spacers. Here, it is possible that the prediction has missed 1 bp. An examination of the sequences of the appropriate circular intermediates should clarify this matter.

An additional general feature of the IS1111 family members examined here is that a −10 promoter region is formed when the ends are brought together in the circular intermediate (Fig. 2 and 5). This −10 region is generally separated by an appropriate distance from a −35 region found at the right-hand end of the IS, and this promoter may be important for expression of the transposase, as has been shown for IS911 (34, 35). The −35 region is also likely to influence the expression of adjacent genes if a −10 region at an appropriate distance, i.e., at the boundary of the IS, is formed. A potential promoter was also found completely within the left-hand end of the IS in a fraction of examples. The overall organization of these IS is also conserved, with 50 to 100 bp separating IRl from the beginning of the transposase gene and 200 to 300 bp between the termination codon and IRr. It seems unlikely that such a long downstream region would be retained if it had no function, but a short open reading frame was found in only some examples and the features it contains thus remain to be established.

A number of IS1111 family members clearly have a preferred target site. These targets include short sequences of up to 16 bp as well as the TIR of other IS or transposons and various sequences found in multiple copies in the genome of a particular organism. In some of the examples evaluated here, the genomic repeats are comprised of IR that can fold to form a stem-loop structure. How these targets are recognized and recruited remains to be established, but it is possible that some structural feature of the DNA rather than the sequence itself is important, as has been proposed for IS231, which also targets a 38-bp TIR sequence (12).

IS families are based on similarities in organization, transposases, and IR (3, 20). The analysis presented here supports the proposal of Lauf et al. (15) that the IS1111 group represents a family. IS that were previously designated as belonging to the IS1111 subgroup of the IS110 family (3), and the addi-

tional ones that were examined here have transposases that are generally more closely related to one another (>30% identity) than to the transposases of the IS110 family (generally less than 20% identity). Members of the IS1111 family also differ from members of the IS110 family in that they include short, 11- to 13-bp IR. This separation will greatly simplify the analysis of this burgeoning IS type.

It is possible that additional subgroups of the IS1111 family will be found in the future as further detailed analyses are undertaken. For example, IS1533, which is found in multiple copies in the chromosome of several *Leptospira* species, has been grouped with IS1111 (3, 15). IS1533 was one of the first IS110/IS1111 family members sequenced, and IR flanked on both sides by additional conserved residues were identified (38, 39). Although in one example the N-terminal portion of the predicted transposase sequence showed similarities to the sequences of a group of IS110-related transposases, including IS1111 (39), the other predicted product was not the same, and both sequences appear to include multiple frameshifts in the transposase gene. An additional IS1533 sequence (GenBank accession no. X77623) found in our searches predicts a transposase that is about 24% identical to the IS4321 and IS1111 transposases and aligns over its full length with those in Fig. 4, making it a member of the broader IS1111 family. Searches with this sequence revealed a small number of IS1533 relatives (>30% identity), and their alignment revealed clusters of conserved residues at positions similar to those in the IS1111 subgroup (Fig. 4), with several completely conserved residues common to both groups. However, there were significant differences. In particular, only three of the four possible acidic residues that form the DED motif in the IS1111 subgroup were also completely conserved in the IS1533 subgroup. The G-D--K and EA motifs were conserved, but the K-D--DA motif in the IS1111 subgroup was K-D/N--DA in the IS1533 subgroup, indicating that the critical D residue is likely to be the second one. A fifth conserved E residue lying to the right of the DED motif was present in IS1533 relatives, but it was closer to the K-D-DA motif than E156 in Fig. 4. For most members of this IS1533 subgroup, IR and possible terminal sequences could be found. However, the IR were shorter and, with the exception of IS1533, usually more closely related to one another than to those of the IS1111 subgroup. Although a more detailed analysis of this group is needed to resolve the precise relationships, the members of the IS1533 subgroup are most closely related to those of the IS1111 subgroup and belong to the IS1111 family.

Members of the IS110 family share features with the IS1111 family in addition to the similarity in their transposase sequences, which all include a small number of completely conserved residues. A comparison of the sequences of integrated copies with the target sequence allowed their boundaries to be identified and led to the conclusion that these IS were integrated conservatively, i.e., without the creation of a short duplication of the target site. Although for IS492 a 5-bp duplication was proposed (28), one copy of this sequence was found in a circular intermediate and the data would be equally well explained if one copy were part of the IS and the other copy were present in the target but fortuitously or necessarily identical. IS4321 and IS5075 also changed locations without duplicating any residues in the target. Circular forms of IS117 (13)

and IS*492* (28) have been reported and, for IS*492*, the formation of a strong promoter at the circle junction has been demonstrated. A number of other IS which belong to the IS*110* subgroup of the IS*110* family appear to have preferred targets (3, 5).

A number of questions about the IS*1111* family and its relationship to the IS*110* family remain to be answered. First, we have not extended our bioinformatic analysis exhaustively, and it is possible that additional subgroups of the IS*1111* family beyond the IS*1111* and IS*1533* subgroups will be found among the IS whose transposases were less than 30% identical to that of IS*1111*, the value which was used as the cutoff in our analysis. Examination of a small number of examples revealed the presence of IR and, although the IR that we found are shorter than those reported here for the IS*1111* family, they were related to the outer 8 bp. Second, whether other IS presently classified as IS*110* family members have IR has not always been examined. Imperfect IR have been noted near the boundaries of IS*492*, as deduced from the sequence of the circular form (28), and we found the configuration CCAT-10 bp-ATGG in that sequence. It is possible that very short IR are also present in other IS. Third, the way in which these IS recognize their preferred target and the features of the target that are recognized remain to be elucidated. In vitro analysis to identify the regions in the IS and in the target bound by individual transposases should assist in clarifying some of these issues. It also remains to be established whether the conserved DDD motif in the IS*110* family and the DED motif in the IS*1111* family are indeed equivalent to the DDE motifs of classical IS or whether the mechanism of movement of IS*110* and IS*1111* family members is indeed more akin to site-specific-recombination.

REFERENCES

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
2. **Bruton, C. J., and K. F. Chater.** 1987. Nucleotide sequence of IS*110*, an insertion sequence of *Streptomyces coelicolor* A3(2). Nucleic Acids Res. **15:**7053–7065.
3. **Chandler, M., and J. Mahillon.** 2002. Insertion sequences revisited, p. 305–366. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, D.C.
4. **de la Cruz, F., and J. Grinsted.** 1982. Genetic and molecular characterization of Tn*21*, a multiple-resistance transposon from R100.1. J. Bacteriol. **151:**222–228.
5. **Doran, T., M. Tizard, D. Millar, J. Ford, N. Sumar, M. Loughlin, and J. Hermon-Taylor.** 1997. IS*900* targets translation initiation signals in *Mycobacterium avium* subsp. *paratuberculosis* to facilitate expression of its *hed* gene. Microbiology **143:**547–552.
6. **Essa, A. M. M., D. J. Julian, S. P. Kidd, N. L. Brown, and J. L. Hobman.** 2003. Mercury resistance determinants related to Tn*21*, Tn*1696*, and Tn*5053* in enterobacteria from the preantibiotic era. Antimicrob. Agents Chemother. **47:**1115–1119.
7. **Garcillán-Barcia, M. P., I. Bernales, M. V. Mendiola, and F. de la Cruz.** 2002. IS*91* rolling circle transposition, p. 891–904. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, D.C.
8. **Griffin, H. G., T. J. Foster, S. Silver, and T. K. Misra.** 1987. Cloning and DNA sequence of the mercuric- and organomercurial-resistance determinants of plasmid pDU1358. Proc. Natl. Acad. Sci. USA **84:**3112–3116.
9. **Grindley, N. D. F.** 2002. The movement of Tn*3*-like elements: transposition and cointegrate resolution, p. 272–302. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, D.C.
10. **Grinsted, J., F. de la Cruz, and R. Schmitt.** 1990. The Tn*21* subgroup of bacterial transposable elements. Plasmid **2:**163–189.
11. **Haas, D., and B. W. Holloway.** 1976. R factor variants with enhanced sex factor activity in *Pseudomonas aeruginosa*. Mol. Gen. Genet. **144:**243–251.
12. **Hallet, B., R. Rezsöhazy, J. Mahillon, and J. Delcour.** 1994. IS*231*A insertion specificity: consensus sequence and DNA bending at the target site. Mol. Microbiol. **14:**131–139.
13. **Henderson, D. J., D. J. Lydiate, and D. A. Hopwood.** 1989. Structural and functional analysis of the mini-circle, a transposable element of *Streptomyces coelicolor* A3(2). Mol. Microbiol. **3:**1307–1318.
14. **Hoover, T. A., M. H. Vodkin, and J. C. Williams.** 1992. A *Coxiella burnetii* repeated DNA element resembling a bacterial insertion sequence. J. Bacteriol. **174:**5540–5548.
15. **Lauf, U., C. Müller, and H. Herrmann.** 1999. Identification and characterisation of IS*1383*, a new insertion sequence isolated from *Pseudomonas putida* strain H. FEMS Microbiol. Lett. **170:**407–412.
16. **Lenich, A. G., and A. C. Glasgow.** 1994. Amino acid sequence homology between Piv, an essential protein in site-specific DNA inversion in *Moraxella lacunata*, and transposases of an unusual family of insertion elements. J. Bacteriol. **176:**4160–4164.
17. **Liebert, C. A., R. M. Hall, and A. O. Summers.** 1999. Transposon Tn*21*, flagship of the floating genome. Microbiol. Mol. Biol. Rev. **63:**507–522.
18. **Luck, S. N., S. A. Turner, K. Rajakumar, H. Sakellaris, and B. Adler.** 2001. Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes. Infect. Immun. **69:**6012–6021.
19. **Mabilat, C., J. Lourencao-Vital, S. Goussard, and P. Courvalin.** 1992. A new example of physical linkage between Tn*1* and Tn*21*: the antibiotic multiple-resistance region of plasmid pCFF04 encoding extended-spectrum β-lactamase TEM-3. Mol. Gen. Genet. **235:**113–121.
20. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
21. **Mendiola, M. V., I. Bernales, and F. de la Cruz.** 1994. Differential roles of the transposon termini in IS*91* transposition. Proc. Natl. Acad. Sci. USA **91:**1922–1926.
22. **Müller, C., U. Lauf, and H. Hermann.** 2001. The inverted repeats of IS*1384*, a newly described insertion sequence from *Pseudomonas putida* strain H, represent the specific target for integration of IS*1383*. Mol. Genet. Genomics **265:**1004–1010.
23. **Nucifora, G., L. Chu, S. Silver, and T. K. Misra.** 1989. Mercury operon regulation by the *merR* gene of the organomercurial resistance system of the plasmid pDU1358. J. Bacteriol. **171:**4241–4247.
24. **Ogawa, H. I., C. L. Tolle, and A. O. Summers.** 1984. Physical and genetic map of the organomercury resistance (Omr) and inorganic mercury resistance (Hgr) loci of the IncM plasmid R831b. Gene **32:**311–320.
25. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848–852.
26. **Partridge, S. R., H. J. Brown, H. W. Stokes, and R. M. Hall.** 2001. Transposons Tn*1696* and Tn*21* and their integrons In4 and In2 have independent origins. Antimicrob. Agents Chemother. **45:**1263–1270.
27. **Partridge, S. R., and R. M. Hall.** 2003. In34, a complex In5 family class 1 integron containing orf513 and *dfrA10*. Antimicrob. Agents Chemother. **47:**342–349.
28. **Perkins-Balding, D., G. Duval-Valentin, and A. C. Glasgow.** 1999. Excision of IS*492* requires flanking target sequences and results in circle formation in *Pseudoalteromonas atlantica*. J. Bacteriol. **181:**4937–4948.
29. **Rakin, A., and J. Heesemann.** 1995. Virulence-associated *fyuA/irp2* gene cluster of *Yersinia enterocolitica* biotype 1B carries a novel insertion sequence IS*1328*. FEMS Microbiol. Lett. **129:**287–292.
30. **Seshadri, R., I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser, and J. F. Heidelberg.** 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc. Natl. Acad. Sci. USA **100:**5455–5460.
31. **Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, and I. T. Paulsen.** 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature **406:**959–964.
32. **Thorsted, P. B., D. P. Macartney, P. Akhtar, A. S. Haines, N. Ali, P. David-**

son, T. Stafford, M. J. Pocklington, W. Pansegrau, B. M. Wilkins, E. Lanka, and C. M. Thomas. 1998. Complete sequence of the IncPβ plasmid R751: implications for evolution and organisation of the IncP backbone. J. Mol. Biol. 282:969–990.

33. Tobiason, D. M., J. M. Buchner, W. H. Thiel, K. M. Gernert, and A. C. Glasgow Karls. 2001. Conserved amino acid motifs from the novel Piv/MooV family of transposases and site-specific recombinases are required for catalysis of DNA inversion by Piv. Mol. Microbiol. 39:641–651.

34. Ton-Hoang, B., M. Betermier, P. Polard, and M. Chandler. 1997. Assembly of a strong promoter following IS911 circularization and the role of circles in transposition. EMBO J. 16:3357–3371.

35. Ton-Hoang, B., P. Polard, and M. Chandler. 1998. Efficient transposition of IS911 circles in vitro. EMBO J. 17:1169–1181.

36. van Beilen, J. B., S. Panke, S. Lucchini, A. G. Franchini, M. Rothlisberger, and B. Witholt. 2001. Analysis of Pseudomonas putida alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the alk genes. Microbiology 147:1621–1630.

37. Vasselon, T., P. J. Sansonetti, and X. Nassif. 1991. Nucleotide sequence of rmpB, a Klebsiella pneumoniae gene that positively controls, colanic biosynthesis in Escherichia coli. Res. Microbiol. 142:47–54.

38. Woodward, M. J., and G. J. Sullivan. 1991. Nucleotide sequence of a repetitive element isolated from Leptospira interrogans serovar hardjo type hardjobovis. J. Gen. Microbiol. 137:1101–1109.

39. Zuerner, R. L. 1994. Nucleotide sequence analysis of IS1533 from Leptospira borgpetersenii: identification and expression of two IS-encoded proteins. Plasmid 31:1–11.