JOURNAL OF BACTERIOLOGY, Aug. 1996, p. 4429–4437 Vol. 178, No. 15
0021-9193/96/$04.00+0


# The Integrons In0, In2, and In5 Are Defective Transposon Derivatives

HEIDI J. BROWN,[1,2] H. W. STOKES,[1] AND RUTH M. HALL[2]*

*School of Biological Sciences, Macquarie University, New South Wales 2109,[1] and Sydney Laboratory, CSIRO Division of Biomolecular Engineering, North Ryde, New South Wales 2113,[2] Australia*

Received 5 February 1996/Accepted 24 May 1996

**The class 1 integrons In0, In2, and In5, found in different locations in pVS1, Tn*21*, and pSCH884, have closely related structures. All three integrons contain an insertion sequence, IS*1326*, that is a new member of the IS*21* family. IS*1326* has caused deletions of adjacent 3′-conserved segment and transposition module sequences, and all three integrons retain a complete copy of only one of four genes required for transposition of related transposons and are thus defective transposon derivatives. In2 contains an additional insertion sequence, IS*1353*, located within IS*1326*. IS*1353* is a member of the IS*3* family and appears to have been acquired after the integron was inserted into an ancestral mercury resistance transposon to create the ancestor of Tn*21* and several other transposons that are close relatives of Tn*21*.**

Many of the antibiotic resistance genes found in clinical isolates of gram-negative bacteria are known to be located in integrons (reviewed in references 12, 14, 15, and 31). The resistance genes are contained in discrete mobile elements known as gene cassettes and are integrated at a specific site, *attI*, in the integron. The integron encodes a site-specific recombinase that belongs to the lambda integrase family, is responsible for the insertion of cassettes at *attI*, and also provides the promoter responsible for expression of the cassette-encoded antibiotic resistance genes.

Three classes of integrons have been identified (see reference 31). However, the integrons most commonly isolated from antibiotic-resistant clinical isolates of members of the family *Enterobacteriaceae* and pseudomonads belong to class 1 and most of them contain the *sulI* sulfonamide resistance gene. These integrons appear to be mobile elements, as they are found in many distinct locations, in plasmids such as R46 and R388 and within transposons such as Tn*21* and Tn*1696* (16, 43). However, early studies aimed at identifying the sequences common to all integrons failed to identify features consistent with the classification of integrons as transposons (43). Two conserved regions located on either side of the integrated gene cassettes were identified. The 5′-conserved segment (5′-CS) includes a gene, *intI1*, encoding the integrase; a gene *attI*, encoding the cassette integration site; and the promoter, $P_{ant}$, that is responsible for expression of the cassette genes (13, 16, 32, 43). The 3′-CS defined by Stokes and Hall (43) included the *sulI* sulfonamide resistance determinant (43, 44) and a further two open reading frames (ORFs), one of which has recently been identified as a defective version of the quaternary ammonium compound resistance determinant *qacE* and designated *qacEΔ1* (26).

Subsequently, a more extensive study of the sequences that were common to six integrons from independent locations revealed that all contained the same 5′-CS but the 3′-CS extended different distances beyond the *sulI* gene (13). It was concluded that a variety of events, such as deletions and insertions, that removed part of or interrupted the 3′-CS had occurred to give rise to the integrons studied. The previous 3′-CS endpoint (43) was found to represent one endpoint of a deletion in one of the two integrons (In1 and In2) used initially to define the 3′-CS, and the sequences assigned to the 3′-CS were extended to include a further 359 bp (13). In three cases, In2 from Tn*21*, In0 from pVS1, and In5 from pSCH884, two further common regions were found. The first directly abuts the 3′-CS endpoint and extends at least 0.7 kb. The second is located 4 to 6 kb away, and an outer end that terminates with a 25-bp sequence that is an inverted repeat of the sequence found at the outer end of the 5′-CS was identified (13). This 25-bp inverted repeat had previously been found at the boundaries of the 11.2-kb insertion in Tn*21*, identified by comparison of the sequences of Tn*501* and Tn*21* (2), and this insertion corresponds to In2. This finding is consistent with the identification of integrons as transposons; however, previous attempts to demonstrate movement of In2 were unsuccessful (see reference 11), suggesting that In2 is a defective transposon derivative.

Recently, the sequence of Tn*402* (Tn*5090*), which is a class 1 integron that does not include the *sulI* gene, has been reported (30). Tn*402* was first identified as a transposon present in IncP plasmid R751 (39). Tn*402* is bounded by the same 25-bp inverted repeats and includes a 5′-CS identical to that found in the *sulI*-containing integrons and two integrated cassettes, *dfrB3* (*dhfrIIc*) and orfD. Tn*402* also includes the complete *qacE* gene (26) and a 4-kb region that contains four ORFs, three of which were identified as encoding potential transposition functions, on the basis of similarities to known transposition proteins (30). Close relatives of these four genes are also found in the transposition module of mercury resistance transposon Tn*5053*, and it has been shown that three of the genes *tniA*, *tniB*, and *tniQ*, are essential for transposition and the fourth, *tniR*, is required for cointegrate resolution (18). Partial sequencing of In2 revealed that a 2.7-kb segment of the Tn*402* transposition module is also present in In2 from Tn*21*, but only one complete gene (*tniA*) and a truncated form of a second gene (*tniB*) are present in this fragment (30). Thus, it appears that In2 may be a defective transposon that does not contain a complete set of transposition functions.

In In2, the 3′-CS sequences (13, 43) and the transposition module segment (30) are separated by a region of about 4 kb, and short inverted repeats (27 bp) have been identified at or close to the boundaries of this segment in In2 (45). The se-

* Corresponding author. Mailing address: CSIRO Division of Biomolecular Engineering, Sydney Laboratory, P.O. box 184, North Ryde NSW 2113, Australia. Phone: 61-2-8864961. Fax: 61-2-8864805. Electronic mail address: hall@pelican.dbe.csiro.au.

TABLE 1. Plasmids used in this study

| Plasmid | Description | Relevant phenotype(s)[a] | Reference |
|---|---|---|---|
| pVS1 | Contains In0 and Tn*501* | $Su^r$ $Hg^r$ | 42 |
| pUB2401 | pACYC184::Tn*21*, contains In2 | $Cm^r$ $Sm^r$ $Sp^r$ $Su^r$ $Hg^r$ | 8 |
| pSCH884 | Contains In5 | $Gm^r$ $Tb^r$ $Su^r$ | 40 |
| pRMH467 | 2.06-kp *Eco*RI fragment of pVS1 in pUC19 | $Ap^r$ | This work |
| pRMH473 | 6.8-kb *Hin*dIII fragment of pVS1 in pUC19 | $Ap^r$ | This work |
| pRMH451 | 1,643-bp *Eco*RI fragment of Tn*21* in pUC19 | $Ap^r$ | This work |
| pRMH456 | 1,847-bp *Eco*RI fragment of Tn*21* in pUC19 | $Ap^r$ | This work |
| pRMH458 | 661-bp *Sph*I fragment of Tn*21* in pUC19 | $Ap^r$ | This work |
| pRMH474 | 4.2-kb *Eco*RI fragment of Tn*21* in pUC19 | $Ap^r$ | This work |
| pRMH475 | 1,573-bp *Sph*I fragment of Tn*21* in pUC19 | $Ap^r$ | This work |
| pRMH477 | 4.6-kb *Eco*RI fragment of pSCH884 in pUC19 | $Ap^r$ | This work |

[a] Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Hg, mercury; Sm, streptomycin; Sp, spectinomycin; Su, sulfonamide; Tb, tobramycin.

quence of the left-hand end of this region has been determined (13), and identical sequences were found in In0 and In5, suggesting that In2, In0, and In5 share a common lineage. However, in In0 and In2, a deletion of 359 bases of the 3′-CS that are present in In5 has occurred (13). Together, these findings suggest that a further transposon or insertion sequence may be present in In2, In0, and In5.

To clarify the relationships between the structures found in different integrons, we examined the sequence of the region between the 3′-CS and transposition module regions in In2, In0, and In5. In all three integrons, an insertion sequence, IS*1326*, was found to be present, and in In2, a further insertion sequence, IS*1353*, inserted just inside one of the terminal inverted repeats of IS*1326* was found. The regions adjacent to IS*1326* were identical in In2 and In0, indicating that these two integrons have a recent common origin. In5 includes extra 3′-CS sequences to the left of IS*1326* and has 439 bp less of the transposition module to the right, indicating that IS*1326* has caused deletions of adjacent sequences. Possible routes for the evolution of the In2, In0, and In5 group from an integron with a backbone structure similar to that of Tn*402* are discussed, and a coherent tree for the evolution of the Tn*21* transposon family is presented.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** *Escherichia coli* JM109 [(Δ*lac-proAB*) *supE thi* F′ (*traD36 proAB*$^+$ *lacI*$^q$ *lacZ* ΔM15)] was used to propagate plasmid DNA. The plasmids used are listed in Table 1. The pRMH series of plasmids listed were constructed by randomly cloning fragments from *Eco*RI, *Hin*dIII, and *Sph*I digests of the relevant parental plasmid (pVS1, Tn*21*, and pSCH884) into pUC19 (52) by using standard procedures (37). Plasmids containing the appropriate fragments were identified by restriction mapping and by sequencing of the fragment ends with a universal primer. Cells were grown in L broth (37) at 37°C if plasmid free and in L broth supplemented with ampicillin (Sigma Chemical Co., St. Louis, Mo.) at 100 μg/ml if plasmids were present.

**DNA isolation and restriction mapping.** Plasmid DNA was recovered from 2-ml samples of overnight cultures by the alkaline lysis method (37). Plasmid DNA for sequencing was purified by using the Magic Minipreps DNA Purification System (Promega Corp., Madison, Wis.). Restriction enzymes were used in accordance with the manufacturers' instructions. Fragments were separated by electrophoresis on 1% (wt/vol) agarose gels and visualized by staining with ethidium bromide. The size standard used was an *Eco*RI digest of bacteriophage SPP1 (Bresatec, Adelaide, South Australia, Australia).

**DNA sequencing.** Nucleotide sequencing was performed with a Sequenase 2.0 system (46) as recommended by the manufacturer (U.S. Biochemicals, Cleveland, Ohio) by using dITP reaction mixtures followed by a 30-min incubation with a 1 mM deoxynucleoside triphosphate mixture and terminal deoxynucleotidyl transferase (Boehringer GmbH, Mannheim, Germany). The specific primers used were 17-mers synthesized with an Applied Biosystems 380A DNA synthesizer.

**Sequence analysis.** Sequence data were collated, assembled, and interpreted by using Staden programs BAP and NIP (7). Database searches (GenBank, February 1996 version) for sequence similarity were performed by using the Match and Match-Translate programs (50). The sequences reported were compiled with previously published contiguous sequences.

**Nucleotide sequence accession numbers.** The GenBank accession numbers of the sequences determined in this study are as follows: In0, U49101; In2, U42226; In5, U38230; IS*1326*, U38187; IS*1353*, U40482.

## RESULTS

**Maps of the region between the 3′-CS and the integron boundary differ.** In a previous study, examination of the sizes of PCR products revealed significant differences in the length of the region between the *Eco*RI site that marks the end of available sequences (Fig. 1) and the right-hand outer boundary of three integrons, In0, In2, and In5 (13). To resolve the structures of these three integrons, restriction fragments from this region, generated by digestion with *Hin*dIII, *Eco*RI, or *Sph*I, were cloned and mapped. By using these data, together with available maps of In0 (17) and In2 (25, 45), detailed maps were generated (Fig. 1). All three integrons include similarly located *Hin*dIII and *Sph*I sites immediately to the right of the *Eco*RI site at the end of the sequenced region, and thereafter the maps diverge. Although in In0 and In2, further *Eco*RI sites are present (one in In0 and three in In2), a single 4.6-kb *Eco*RI fragment from pSCH884 includes the remainder of In5.

**Insertion sequence IS*1326* has caused adjacent deletions in In0 and In5.** The regions shown in Fig. 1 were sequenced, and the sequence of In0 extending from the last *Bam*HI site in the 3′-CS is shown in Fig. 2. The sequence of In5 (accession no. U38230) is identical to this sequence, except that a 439-bp region (boxed in Fig. 2) is not present in In5 and a further region of 359 bp is present in In5 (13) at the position indicated in Fig. 2. A simple explanation for these findings is that In0 and In5 have arisen from a common ancestor by deletion of sequences to the left or right of bases 152 and 2625, respectively (Fig. 2). Careful examination of the region between these bases revealed features consistent with its identification as an insertion sequence, which was designated IS*1326*. IS*1326* is 2,470 bp long, is bounded by inverted repeats of 26 bp with 22 of the 26 bp identical, and includes two ORFs (see below for details). The apparent deletion in In0 relative to In5 originates precisely at the left boundary of IS*1326*, and the apparent deletion in In5 relative to In0 originates precisely at the right boundary of IS*1326* (Fig. 3). This suggests that In0 and In5 have arisen from a common ancestor by IS*1326*-mediated deletion of sequences adjacent to the IS*1326* insertion in that ancestor.

The sequence to the right of IS*1326* in In0 (bases 2626 to 3110 in Fig. 2) is identical to bases 2891 to 2406 in Tn*402* (30), which is also a class 1 integron, and the sequence identity extends to and includes the 25-bp repeat at the right-hand terminus of In0 (13) and Tn*402*. This 2.7-kb region lies within the transposition module of Tn*402*, which contains the ORFs encoding transposition functions (18, 30). Only 2.26 kb of the transposition module is present in In5. As these regions in-

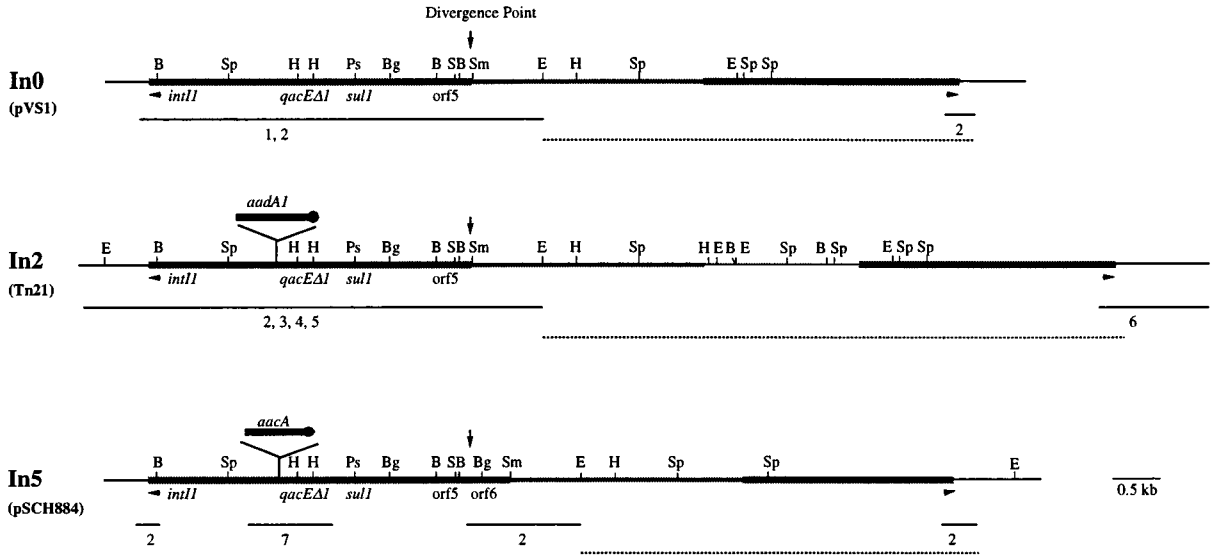


FIG. 1. Restriction maps of *sulI*-type integrons In0, In2, and In5. Arrowheads indicate the 25-bp inverted repeats at the outer boundaries, and the vertical arrows indicate the point at which the In0 and In2 sequences diverge from that of In5. Solid lines below the maps indicate the extent of published sequences collated from the following sources: 1, Bissonnette and Roy (1); 2, Hall et al. (13); 3, Diver et al. (9); 4, Stokes and Hall (43); 5, Sundström et al. (44); 6, Brown et al. (2); 7, Shaw et al. (40). Dashed lines show the extent of sequences determined in this study. Restriction enzyme sites: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Ps, *Pst*I; S, *Sal*I; Sm, *Sma*I; Sp, *Sph*I.

clude only the complete *tniA* gene and part of the *tniB* gene, it seems reasonable to conclude that the ancestor of In0 and In5 contained the complete transposition module, part of which was subsequently lost as the result of one or more IS*1326*-mediated deletion events.

**Insertion sequence IS*1326* is a member of the IS*21* family.** IS*1326* includes two long ORFs that overlap by 13 bp (Fig. 2). The polypeptides predicted by these ORFs were found to be related to the two polypeptides encoded by the *istA* and *istB* genes of IS*21* (34) and to the corresponding polypeptides in three further known relatives of IS*21*, namely, IS*232* (23), IS*5376* (51), and IS*100* (27). On the basis of the similarity of the IS*1326* ORFs to IstA and IstB from IS*21*, which have been shown to be required for transposition (33), the IS*1326* genes are also designated *istA* and *istB*. A detailed description of the features common to members of the IS*21* family, including IS*1326*, will be presented elsewhere (2a). Briefly, IstA includes an unusual D, D(45), E transposase motif and a helix-turn-helix motif and IstB has a distinctive set of NTP-binding domains. Inverted repeats of 27 bp are present at the boundaries of IS*1326* in In0, In2, and In5, as has been noted previously for In2 (45). However, as many of the mobile elements that encode a transposase with a D, D(35), E motif are bounded by 5′-TG and CA-3′ (10, 28), we have assumed that IS*1326* is bounded by the same dinucleotides and that the inverted repeats are most likely to be 26 bp long.

**In2 is closely related to In0 but contains IS*1353* inserted within IS*1326*.** The restriction map of In2 includes an extra segment not found in In0 and In5 (see Fig. 1). As this region has previously been shown to be absent from transposons such as Tn*2411*, which are otherwise very close relatives of Tn*21* (20), the possibility that this extra segment corresponds to an insertion sequence was examined. The sequence of In2 from the *Eco*RI site in IS*1326* to the outer, right-hand end of In2 was determined (GenBank accession number U42226), and this completes the sequences of In2 and Tn*21*. In2 contains a complete copy of IS*1326*, but a further new insertion sequence, designated IS*1353*, has been inserted just within the one end (right-hand end as shown in Fig. 3) of IS*1326*. The part of the In2 sequence containing IS*1353* is shown in Fig. 4. IS*1353* is 1,613 bp long and is bounded by inverted repeats of 12 and 13 bp (boxed in Fig. 4). Comparison of the In2 sequence with that of In0 (Fig. 2) revealed that IS*1353* is flanked by a 2-bp direct duplication of the target sequence (indicated in Fig. 2). IS*1353* is a member of the IS3 family (see below). Beyond the end of IS*1326*, the In2 sequence is identical to that of In0, and identity to the sequence of part of the transposition module of Tn*402* extends up to the right-hand end of In2, as suggested by Rådström et al. (30), who sequenced several parts of this region. Thus, In2 is identical to In0, except that it includes an integrated *aadA1* gene cassette, that determines resistance to streptomycin and spectinomycin, and IS*1353*.

**IS*1353* is a member of the IS*3* family.** Analysis of the IS*1353* ORFs longer than 50 amino acids revealed that on one strand there were two ORFs, orfA of 222 codons and orfB of 304 codons, spanning almost the entire insertion sequence (Fig. 4). Several ORFs were also identified on the other strand. Only ORFB was found to be related to known sequences. ORFB shares 20 to 37% identity with the ORFB products of IS*3*, IS*911*, and the other 32 known members of the IS*3* family (10, 21, 28, 49), and several conserved regions found in the transposases of IS*3* family members are shared by ORFB. In particular, the acidic residue motif D, D(35), E, which was originally identified in the IS*3* family (10), is present in ORFB. This motif is also present in other bacterial transposases, including those of Tn*552* (36) and the IS*4* family (35), as well as in the integrases of many retrotransposons and retroviruses (10, 22), and is believed to be part of the catalytic domain of these transposases and integrases (see reference 38).

IS*1353* also exhibits an organization similar to that of members of the IS*3* family. Most members of the IS*3* family are bounded by 5′-TG and CA-3′ (10, 28), and IS*1353* shares this feature. Members of this group generally code for two ORFs, one short ORF (orfA) that overlaps a larger downstream ORF (orfB), which is in phase −1 with respect to ORFA. Expression studies with IS*911* (29) and IS*150* (48) have shown that these

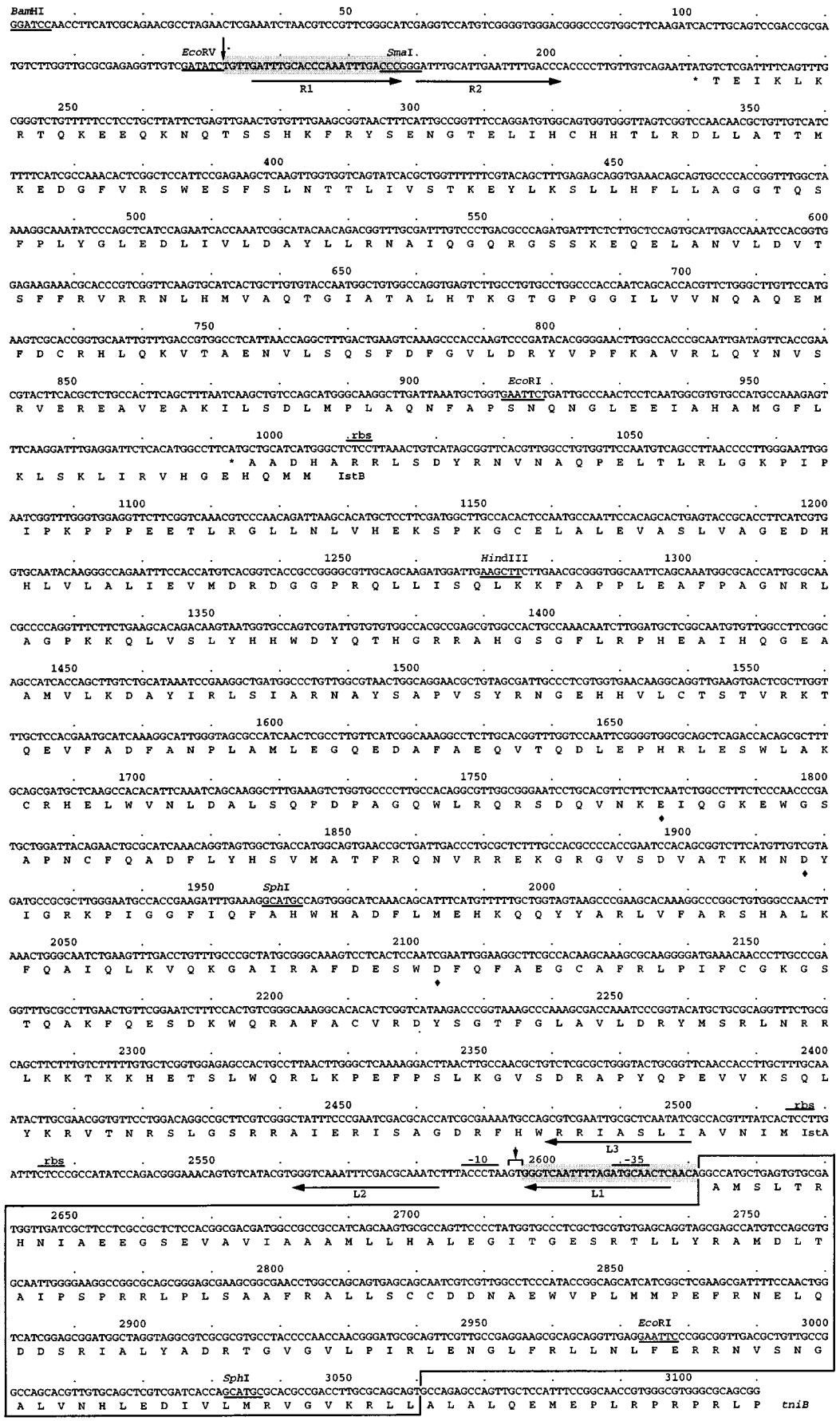
BamHI
EcoRV
SmaI
R1
R2
rbs
IstB
HindIII
EcoRI
SphI
L3
L2
L1
-10
-35
tniB

FIG. 2. Nucleotide sequence of In0, including the sequence of IS*1326*. The sequence shown extends from the *Bam*HI site commencing at base 1876 of the 3′-CS (13). The sequence of the corresponding fragment of pSCH884 is identical, except that pSCH884 contains an extra 359 bp located at the position indicated by the arrow between bases 151 and 152 (13) and a deletion of 439 bp (boxed). The terminal 26-bp inverted repeats of IS*1326* are shaded, and repeats of 22 bp found near the ends of the insertion sequence are indicated by arrows and designated R1, R2, L1, L2, and L3. Potential ribosome-binding sites (rbs) and promoter −10 and −35 regions are indicated. Diamonds indicate the conserved acidic residues of the D, D(45), E motif. The vertical arrow at bases 2594 and 2595 marks the position of insertion of IS*1353* in In2 (see Fig. 3 and 4).

elements produce three polypeptides, ORFA, ORFB, and fusion protein ORFAB. ORFAB, which appears to be the active transposase for IS*911* (29), is a transframe product, resulting from a translational −1 frameshift that occurs in the region where orfA and orfB overlap. Specific signals induce the −1 frameshift; a heptanucleotide motif, AAAAAAG, in the overlap region; and a downstream sequence capable of forming a strong secondary structure, such as a stem-loop structure (29, 48). These signals are also present in many other members of the IS*3* family (3, 21, 49), and in the region of overlap between ORFA and ORFB in IS*1353*, the potential frameshift motif $A_6G$ is present and a pair of inverted repeats that have the potential to form a stem-loop structure are found downstream (Fig. 4). This suggests that IS*1353*, in common with other members of the IS*3* family, also utilizes translational frameshifting. It is highly likely that in IS*911*, translation of ORFB is initiated at an ATT codon that lies immediately upstream of the $A_6G$ motif (29). In IS*1353*, there is also an ATT codon 3 bp upstream of the $A_6G$ motif; however, an ATG codon is also present six codons downstream, and which codon is in fact used for initiation of ORFB needs to be determined experimentally.

## DISCUSSION

The sequence data presented here show that integrons In0, In2, and In5 all contain IS*1326* and are also otherwise closely related (Fig. 3). The two differences between In0 and In2 are readily explained by known mechanisms for DNA insertion.

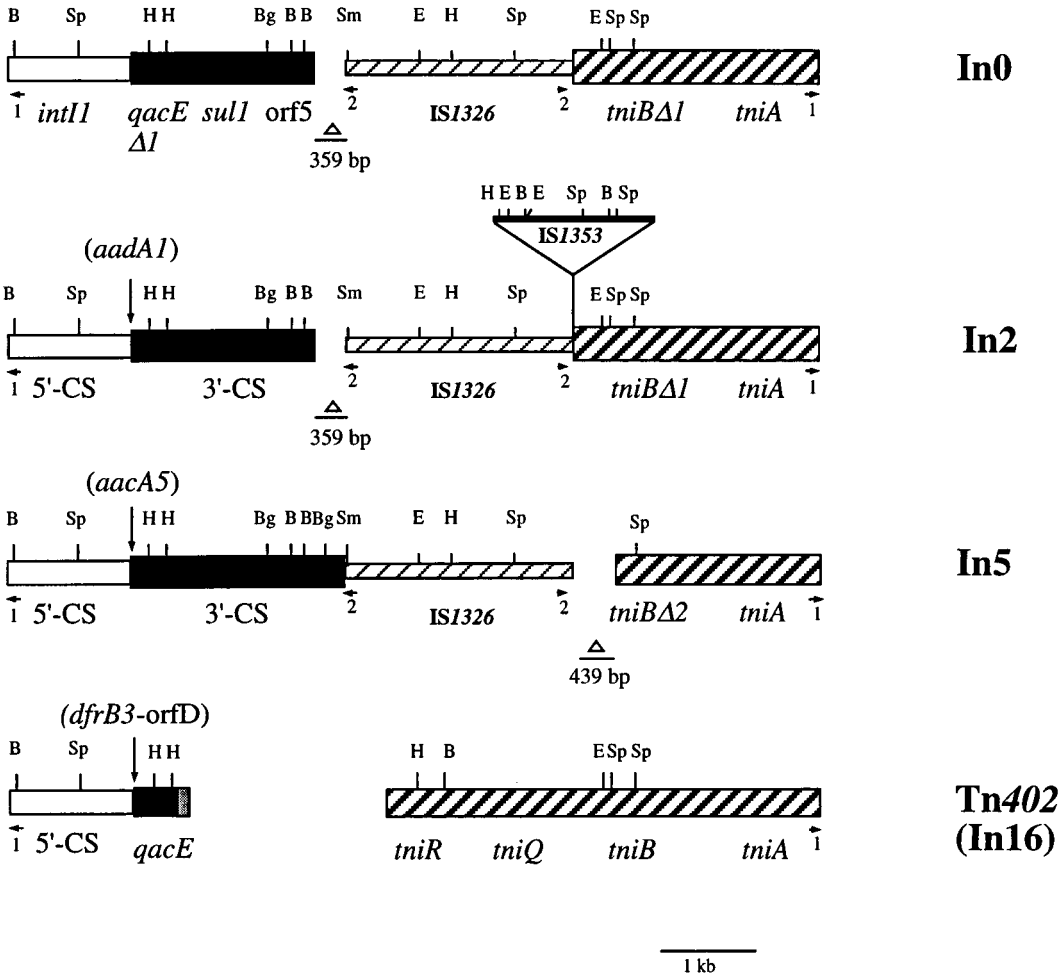


FIG. 3. Structures of In0, In2, In5, and Tn*402*. The 5′-CS is shown as an open box, the 3′-CS is a solid box, and the *tni* module is a bold hatched box. The region of the Tn*402 qacE* gene that is not present in the 3′-CS is stippled. Vertical arrows indicate the positions of the integrated cassettes (in parentheses). IS*1326* is represented by a narrow hatched box, and IS*1353* is a solid line. The regions shown are contiguous, and the gaps were introduced only to align identical sequences in the four integrons shown. The extents of IS*1326*-mediated deletions in In0, In2, and In5 are indicated below the maps. Arrowheads indicate the terminal inverted repeats of the integrons (designated 1) and IS*1326* (designated 2). Abbreviations for restriction sites are the same as in Fig. 1.

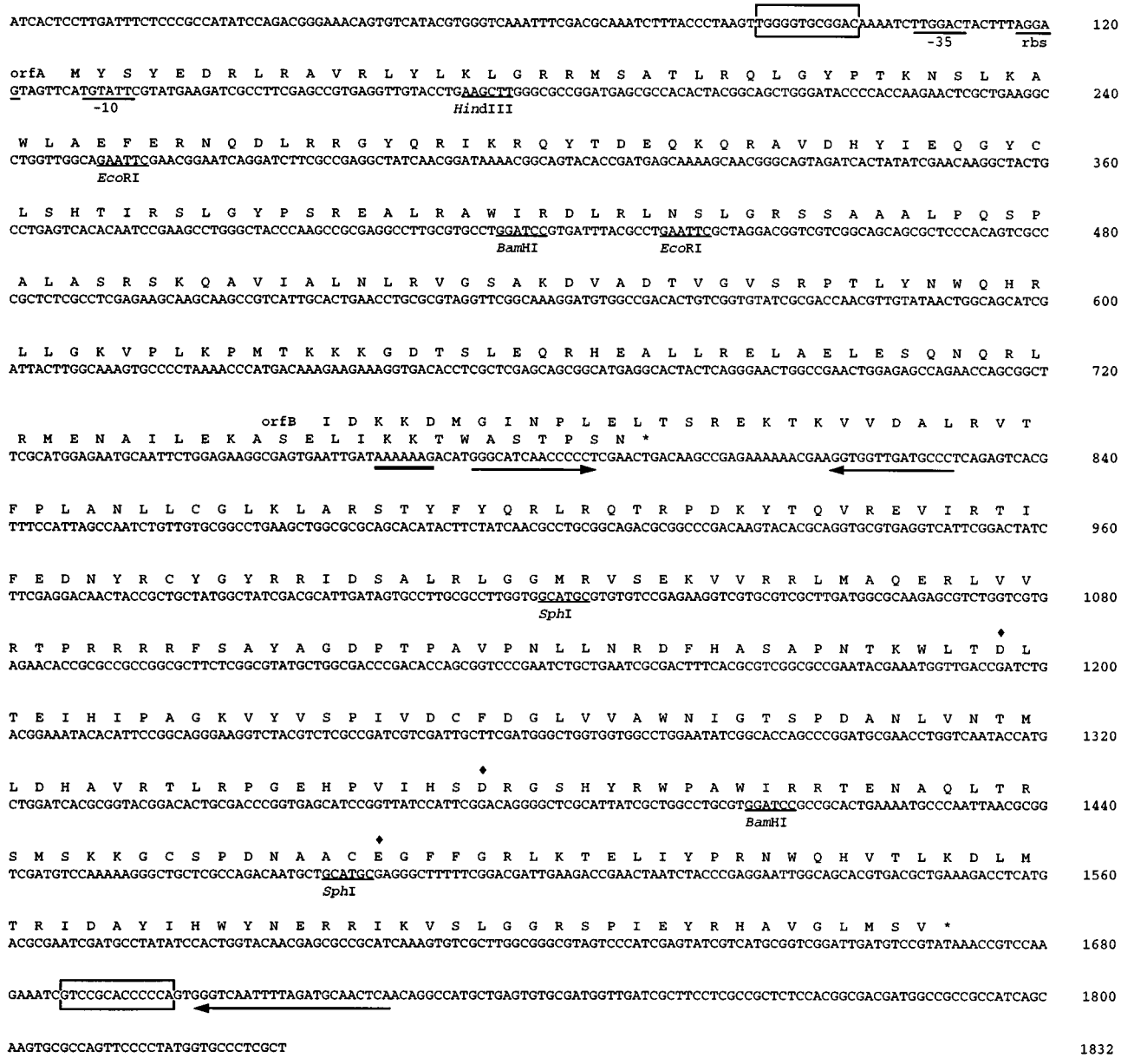


FIG. 4. Nucleotide sequence of the In2 region containing IS*1353*. The IS*1353* inverted repeats are boxed. The potential heptanucleotide frameshift signal $A_6G$ is underlined, and the inverted repeats of the downstream potential stem-loop structure are indicated by arrows. Diamonds indicate the conserved acidic residues of the D, D(35), E motif. The terminal inverted repeat from IS*1326* is underlined with an arrow.

Whereas In0 contains no integrated cassettes (1), In2 contains the *aadA1* cassette (44) that has been integrated at *attI* by the action of the integron integrase IntI1 (4). The further extra segment in In2 corresponds to an acquired insertion sequence, IS*1353*. There are only three differences between In0 and In5, one of which is the presence of an *aacA* cassette integrated at *attI* in In5 (40). The remaining differences are both deletions with one endpoint at an IS*1326* boundary and can be explained by invoking deletion events originating at the boundaries of IS*1326*. These events are presumably mediated by the action of the insertion sequence-encoded transposition functions in a manner analogous to the formation of adjacent deletions by other insertion sequences and transposons (19).

Recent studies on Tn*5053*, which has a transposition module closely related to that of Tn*402*, have demonstrated that three genes, *tniA*, *tniB*, and *tniQ*, are essential for transposition of Tn*5053* and that a further gene, *tniR*, that encodes a resolvase is required for efficient resolution of cointegrate intermediates (18). Because the complete *tni* module found in Tn*402* is not present in any of the three integrons examined here, it seems reasonable to conclude that these integrons are all defective derivatives of an ancestral transposon and are unable to direct their own movement. However, the fact that In2 and In0, which contain identical 3′-CS and *tni* module segments, are found in independent locations, in Tn*21* and in pVS1, respectively (see reference 13), suggests that movement can occur. As movement of Tn*5053* derivatives with defects in any or all of the *tniA*, *tniB*, and *tniQ* genes could be restored by supplying

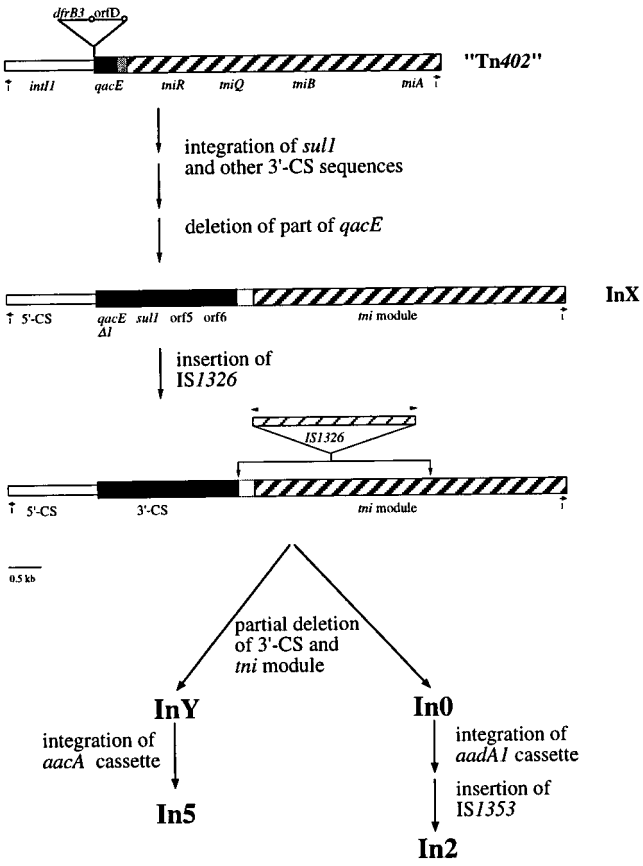


FIG. 5. Evolution of In0, In2, and In5 from an ancestral, transposition-proficient integron similar to Tn*402*. The 5′-CS, 3′-CS, transposition modules, and IS*1326* are indicated as in Fig. 3. Only Tn*402* is shown as containing integrated gene cassettes, as any cassette configuration at the position indicated by a vertical arrow can be readily achieved by IntI-mediated recombination. InX and InY are hypothetical intermediates.

the defective function(s) in *trans* (18), it is likely that integrons such as In0, In2, and In5, which retain intact ends, could be mobilized when a transposon with a complete *tni* module, such as Tn*402* or Tn*5053*, is present in the same cell. Indeed, movement of In0 and In2 in the presence of mercury resistance transposon Tn*502* (41) has recently been demonstrated (40a).

A model for the evolution of In0, In2, and In5 from an ancestral transposition-proficient class 1 integron similar to Tn*402* is shown in Fig. 5. Acquisition and loss of gene cassettes are well documented (4–6, 14, 31) and can occur at any time. For simplicity, cassette exchange events have been assumed to occur last. The ancestor, represented by Tn*402*, consists of four distinct domains: the 5′-CS encoding the integrase responsible for the movement of the gene cassettes, the variable region which contains the integrated gene cassettes (*dfrB3* and orfD in Tn*402*), the *qacE* gene region, and the transposition module which includes the *tniA*, *tniB*, *tniQ* (orf6), and *tniR* (*tniC*) genes (18, 30).

The proposed transposition-proficient ancestor of the *sulI* family of class 1 integrons contains the complete *tni* module and is derived from the Tn*402* ancestral configuration by integration of the *sulI* gene and other 3′-CS sequences downstream of *qacE* and deletion of the end of the *qacE* gene to generate *qacEΔ1*, which is found in the 3′-CS. The exact mechanism of the evolution of the 3′-CS module remains a matter for speculation, but the 3′-CS may represent a particular series of gene cassettes that have become fixed. We have previously suggested that the *sulI* gene may have originally been part of a cassette which was integrated at *attI* and subsequently fixed by loss of the functional 59-base element recombination site (43). The same scenario has recently been suggested for the origin of the *qacEΔ1* gene (30), and the fact that the *qacE* gene of Tn*402* has since been shown to be part of a functional mobile gene cassette (31a) supports this conclusion.

Insertion of IS*1326* into the ancestral *sulI*-containing integron is assumed to have occurred either within a further, unidentified, region of the 3′-CS (dotted region in Fig. 5) or within the *tni* module but between the positions shown by the horizontal bar in Fig. 5. Thereafter, deletion events originating at the ends of IS*1326* and leading to loss of known 3′-CS sequences from In0 and In2 and to loss of part of the *tni* module in In0, In2, and In5 have occurred. To explain the different extents of 3′-CS and *tni* modules retained in In5 and In0 (In2), it is necessary to invoke at least two insertion sequence-mediated events in the evolution of one of them. The acquisition of IS*1353* by In2 is presumed to be a recent event (see below).

**Evolution of the Tn*21* transposon family.** By using information from this study, together with data from previously published restriction maps of Tn*21* and related transposons and heteroduplex analyses involving pairs of transposons from this family (11, 20, 24, 25, 38, 47), it is possible to construct a coherent model for the evolution of members of this family (Fig. 6). Only close relatives of Tn*21* identified by the presence of the characteristic motif of *Eco*RI and *Hin*dIII sites found in the Tn*21* transposition gene (*tnpA*, *tnpR*, and *res*) region are considered here as members of the Tn*21* family. This group includes a little over half of the transposons listed as belonging to the Tn*21* subgroup of the Tn*3* transposon family in a recent review (see Table 1 in reference 11). Tn*1401*, Tn*1406*, and Tn*1409* were not considered because the most recently published maps of these transposons (24) are not sufficiently accurate to permit analysis. One further transposon, Tn*5086*, that has subsequently been identified as a member of this group (45) was considered.

In the simplest evolutionary tree consistent with all of the published data, the first step in the evolution of this family is the insertion of an integron identical to In0 (or equivalent to In0 but containing one or more integrated gene cassettes) into an ancestral mercury resistance transposon, possibly Tn*2613* (47). As most members of this group include the *aadA1* cassette, for simplicity it has been assumed that this cassette was present in the incoming integron. Tn*2411*, rather than Tn*21*, is assumed to be the true ancestor of the integron-containing members of this family because the acquisition of IS*1353* by In2 (in Tn*21*) appears to be a recent event that occurred after the movement of the integron to its present location. The region corresponding to IS*1353* is not present in transposons that are otherwise very closely related to Tn*21*, such as Tn*4* and Tn*2411* (20, 25, 47), and it has been suggested that this region was acquired in Japan, as it is generally present in Tn*21*-like transposons isolated in Japan (11).

Most of the differences between Tn*2411* and the known transposons in this family can be explained by site-specific integration of further cassettes into the integron or insertion of further transposons (e.g., Tn*3* in Tn*4* and IS*1353* in Tn*21*) at various positions. Two further differences can be explained by IS*1326*-mediated events. The large deletion in Tn*1831* (25) originates close to one boundary of IS*1326*, extends beyond the right-hand end of the integron into the *mer* region, and is likely to be due to a further IS*1326*-mediated deletion event. Trans-

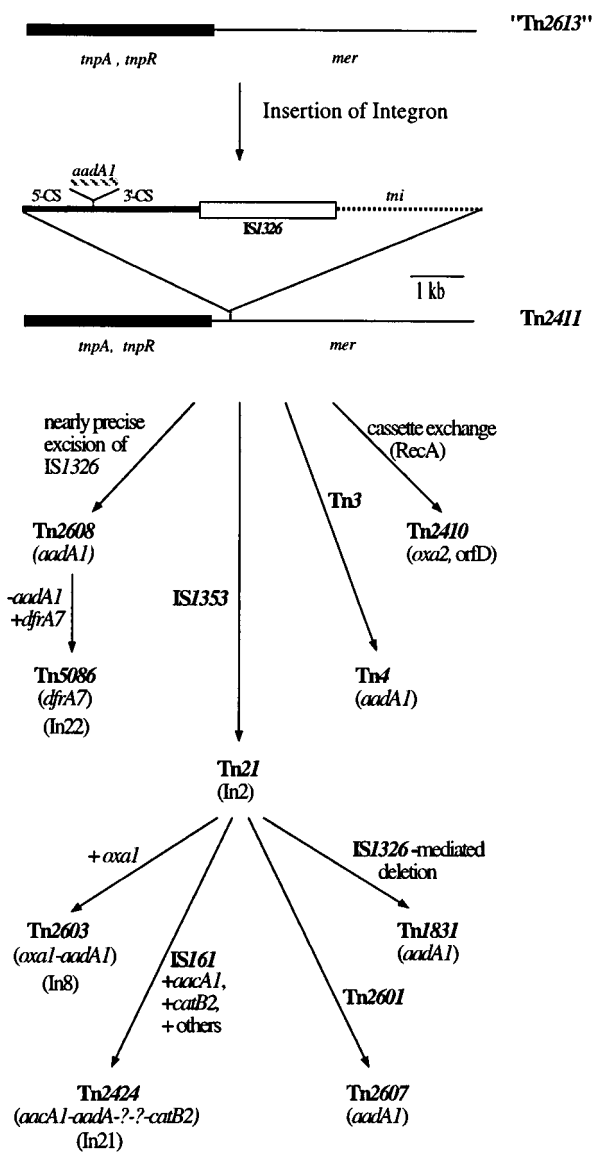


FIG. 6. Evolution of the Tn*21* transposon family. The relationships of all of the transposons in this group (see reference 11 for references) can be explained by invoking the insertion of an integron with the structure shown followed by subsequent events including integration, excision, or exchange of gene cassettes, insertion of transposons, or insertion sequences, insertion sequence-mediated deletions, or insertion sequence excision.

poson Tn*5086* contains all of the bases to the right of IS*1326* in In2 and all but the last four bases to the left (TATC), which are replaced by two bases (CA) (45). The simplest explanation for this configuration is excision of IS*1326* with imprecise rejoining and repair of the ends generated. A similar configuration is found in Tn*2608* (47), which may also have arisen by this route.

Other members of the Tn*21* subgroup are also known to include integrons, the best known being Tn*1696*, which contains the integron In4 (13, 43). However, Tn*1696* is not a close relative of Tn*2411* and the integron is located at a different position (38) and has a different structure, which includes IS*6100* (13). Thus, it appears that Tn*1696* has a different evolutionary history and that integrons have been inserted into ancestral transposons containing *tnp* and *mer* modules on more than one occasion. Studies to resolve the complete structure of In4 and Tn*1696* are in progress.

## ACKNOWLEDGMENTS

We thank Vilma Stanisich for communicating data prior to publication. Access to DNA databases and sequence analysis programs and assistance in their use were provided by the Australian National Genomic Information Service (A.N.G.I.S.).

This work was supported by a grant from the National Health and Medical Research Council.

## REFERENCES

1. **Bissonnette, L., and P. H. Roy.** 1992. Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multiresistance plasmids and transposons of gram-negative bacteria. J. Bacteriol. **174:**1248–1257.
2. **Brown, N. L., T. K. Misra, J. N. Winnie, A. Schmidt, M. Seiff, and S. Silver.** 1986. The nucleotide sequence of the mercuric resistance operons of plasmid R100 and transposon Tn*501*: further evidence for *mer* genes which enhance the activity of the mercuric ion detoxification system. Mol. Gen. Genet. **202:**143–151.

2a. **Brown, H. J., and R. M. Hall.** Unpublished data.

3. **Chandler, M., and O. Fayet.** 1993. Translational frameshifting in the control of transposition in bacteria. Mol. Microbiol. **7:**497–503.
4. **Collis, C. M., G. Grammaticopoulos, J. Briton, H. W. Stokes, and R. M. Hall.** 1993. Site-specific insertion of gene cassettes into integrons. Mol. Microbiol. **9:**41–52.
5. **Collis, C. M., and R. M. Hall.** 1992. Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. J. Bacteriol. **174:**1574–1585.
6. **Collis, C. M., and R. M. Hall.** 1992. Gene cassettes from the insert region of integrons are excised as covalently closed circles. Mol. Microbiol. **6:**2875–2885.
7. **Dear, S., and R. Staden.** 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. **19:**3907–3911.
8. **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.
9. **Diver, W. P., J. Grinsted, D. C. Fritzinger, N. L. Brown, J. Altenbuchner, P. Rogowsky, and R. Schmitt.** 1983. DNA sequences of and complementation by the *tnpR* genes of Tn*21*, Tn*501* and Tn*1721*. Mol. Gen. Genet. **191:**189–193.
10. **Fayet, O., P. Ramond, P. Polard, M. F. Prère, and M. Chandler.** 1990. Functional similarities between retroviruses and the IS*3* family of bacterial insertion sequences? Mol. Microbiol. **4:**1771–1777.
11. **Grinsted, J., F. de la Cruz, and R. Schmitt.** 1990. The Tn*21* subgroup of bacterial transposable elements. Plasmid **24:**163–189.
12. **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 crossover point. Mol. Microbiol. **5:**1941–1959.
13. **Hall, R. M., H. J. Brown, D. E. Brookes, and H. W. Stokes.** 1994. Integrons found in different locations have identical 5′ ends but variable 3′ ends. J. Bacteriol. **176:**6286–6294.
14. **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.
15. **Hall, R. M., and H. W. Stokes.** 1993. Integrons: novel DNA elements which capture genes by site-specific recombination. Genetica **90:**115–132.
16. **Hall, R. M., and C. Vockler.** 1987. The region of the IncN plasmid R46 coding for resistance to β-lactam antibiotics, streptomycin/spectinomycin and sulphonamides is closely related to antibiotic resistance segments found in IncW plasmids and in Tn*21*-like transposons. Nucleic Acids Res. **15:**7491–7501.
17. **Itoh, Y., J. M. Watson, D. Haas, and T. Leisinger.** 1984. Genetic and molecular characterization of the *Pseudomonas* plasmid pVS1. Plasmid **11:**206–220.
18. **Kholodii, G. Y., S. Z. Mindlin, I. A. Bass, O. V. Yurieva, S. V. Minakhina, and V. G. Nikiforov.** 1995. Four genes, two ends, and a *res* region are involved in transposition of Tn*5053*: a paradigm for a novel family of transposons carrying either a *mer* operon or an integron. Mol. Microbiol. **17:**1189–1200.
19. **Kleckner, N.** 1977. Translocatable elements in procaryotes. Cell **11:**11–23.
20. **Kratz, J., F. Schmidt, and B. Wiedemann.** 1983. Characterization of Tn*2411* and Tn*2410*, two transposons derived from R-plasmid R1767 and related to Tn*2603* and Tn*21*. J. Bacteriol. **155:**1333–1342.
21. **Kropinski, A. M., M. A. Farinha, and I. Jansons.** 1994. Nucleotide sequence of the *Pseudomonas aeruginosa* insertion sequence IS*222*: another member of the IS*3* family. Plasmid **31:**222–228.
22. **Kulkosky, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka.** 1992.

Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12:**2331–2338.
23. **Menou, G., J. Mahillon, M.-M. Lacadet, and D. Lereclus.** 1990. Structural and genetic organization of IS*232*, a new insertion sequence of *Bacillus thuringiensis*. J. Bacteriol. **172:**6689–6696.
24. **Mercier, J., J. Lachapelle, F. Couture, M. Lafond, G. Vézina, M. Boissinot, and R. C. Levesque.** 1990. Structural and functional characterization of *tnpI*, a recombinase locus in Tn*21* and related β-lactamase transposons. J. Bacteriol. **172:**3745–3757.
25. **Meyer, J. F., B. A. Nies, J. Kratz, and B. Wiedemann.** 1985. Evolution of Tn*21*-related transposons: isolation of Tn*2425*, which harbours IS*161*. J. Gen. Microbiol. **131:**1123–1130.
26. **Paulsen, I. T., T. G. Littlejohn, P. Rådström, L. Sundström, O. Sköld, G. Swedberg, and R. A. Skurray.** 1993. The 3′-conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. Antimicrob. Agents Chemother. **37:**761–768.
27. **Podladchikova, O. N., G. G. Dikhanov, A. V. Rakin, and J. Heesemann.** 1994. Nucleotide sequence and structural organization of *Yersinia pestis* insertion sequence IS*100*. FEMS Microbiol. Lett. **121:**269–274.
28. **Polard, P., and M. Chandler.** 1995. Bacterial transposases and retroviral integrases. Mol. Microbiol. **15:**13–23.
29. **Polard, P., M. F. Prère, M. Chandler, and O. Fayet.** 1991. Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS*911*. J. Mol. Biol. **222:**465–477.
30. **Rådström, P., O. Sköld, G. Swedberg, J. Flensburg, P. H. Roy, and L. Sundström.** 1994. Transposon Tn*5090* of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. J. Bacteriol. **176:** 3257–3268.
31. **Recchia, G. D., and R. M. Hall.** 1995. Gene cassettes: a new class of mobile element. Microbiology **141:**3015–3027.
31a. **Recchia, G. D., and R. M. Hall.** Unpublished data.
32. **Recchia, G. D., H. W. Stokes, and R. M. Hall.** 1994. Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. Nucleic Acids Res. **22:**2071–2078.
33. **Reimmann, C., and D. Haas.** 1990. The *istA* gene of insertion sequence IS*21* is essential for cleavage at the inner 3′ ends of tandemly repeated IS*21* elements *in vitro*. EMBO J. **9:**4055–4063.
34. **Reimmann, C., R. Moore, S. Little, A. Savioz, N. S. Willetts, and D. Haas.** 1989. Genetic structure, function and regulation of the transposable element IS*21*. Mol. Gen. Genet. **215:**416–424.
35. **Rezsöhazy, R., B. Hallet, J. Delcour, and J. Mahillon.** 1993. The IS*4* family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. **9:**1283–1295.
36. **Rowland, S.-J., and K. G. H. Dyke.** 1990. Tn*552*, a novel transposable element from *Staphylococcus aureus*. Mol. Microbiol. **4:**961–975.
37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
38. **Schmidt, F., and I. Klopfer-Kaul.** 1984. Evolutionary relationship between Tn*21*-like elements and pBP201, a plasmid from *Klebsiella pneumoniae* mediating resistance to gentamicin and eight other drugs. Mol. Gen. Genet. **197:**109–119.
39. **Shapiro, J. A., and P. Sporn.** 1977. Tn*402*: a new transposable element determining trimethoprim resistance that inserts in bacteriophage lambda. J. Bacteriol. **129:**1632–1635.
40. **Shaw, K. J., C. A. Cramer, M. Rizzo, R. Mierzwa, K. Gewain, G. H. Miller, and R. S. Hare.** 1989. Isolation, characterization, and DNA sequence analysis of an AAC(6′)-II gene from *Pseudomonas aeruginosa*. Antimicrob. Agents Chemother. **33:**2052–2062.
40a. **Stanisich, V. A.** Personal communication.
41. **Stanisich, V. A., R. Arwas, P. M. Bennett, and F. de la Cruz.** 1989. Characterization of *Pseudomonas* mercury-resistance transposon Tn*502*, which has a preferred insertion site in RP1. J. Gen. Microbiol. **135:**2909–2915.
42. **Stanisich, V. A., P. M. Bennett, and M. H. Richmond.** 1977. Characterization of a translocation unit encoding resistance to mercuric ions that occurs on a nonconjugative plasmid in *Pseudomonas aeruginosa*. J. Bacteriol. **129:**1227–1233.
43. **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.
44. **Sundström, L., P. Rådström, G. Swedberg, and O. Sköld.** 1988. Site-specific recombination promotes linkage between trimethoprim and sulfonamide resistance genes. Sequence characterization of *dhfrV* and *sulI* and a recombination active locus of Tn*21*. Mol. Gen. Genet. **213:**191–201.
45. **Sundström, L., G. Swedberg, and O. Sköld.** 1993. Characterization of transposon Tn*5086*, carrying the site-specifically inserted gene *dhfrVII* mediating trimethoprim resistance. J. Bacteriol. **175:**1796–1805.
46. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84:**4767–4771.
47. **Tanaka, M., T. Yamamoto, and T. Sawai.** 1983. Evolution of complex resistance transposons from an ancestral mercury transposon. J. Bacteriol. **153:** 1432–1438.
48. **Vögele, K., E. Schwartz, C. Welz, E. Schiltz, and B. Rak.** 1991. High-level ribosomal frameshifting directs the synthesis of IS*150* gene products. Nucleic Acids Res. **19:**4377–4385.
49. **Walker, D. C., and T. R. Klaenhammer.** 1994. Isolation of a novel IS*3* group insertion element and construction of an integration vector for *Lactobacillus* spp. J. Bacteriol. **176:**5330–5340.
50. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80:**726–730.
51. **Xu, K., Z. Q. He, Y. M. Mao, R. Q. Sheng, and Z. J. Sheng.** 1993. On two transposable elements from *Bacillus stearothermophilus*. Plasmid **29:**1–9.
52. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.