Distribution of IS91 family insertion sequences in bacterial genomes: evolutionary implications

M. Pilar Garcillán-Barcia, Fernando de la Cruz *

Departamento de Biología Molecular (Unidad Asociada al C.I.B., C.S.I.C.), Universidad de Cantabria, C/Herrera Oria s/n, 39011 Santander, Spain

Received 1 February 2002; received in revised form 6 May 2002; accepted 17 May 2002

Abstract

IS91 is the prototype element of a family of bacterial insertion sequences that transpose by a rolling-circle mechanism. Although previously considered a rarity among IS elements, many new examples have been identified by sequence analysis of bacterial genomes. In this work we provide a summary of occurrences of IS91-like sequences in the GenBank database, characterise the genetic organisation of adjacent sequences, and analyse IS91 ecological significance under the light of current transposition mechanisms. Interestingly, IS91 family elements were usually found adjacent to pathogenicity- and virulence-related genes. Thus, this might constitute the niche for IS91 and IS91 family elements to play an important role in the dissemination and evolution of virulence and pathogenicity types of genes.

Keywords: Insertion sequence; Rolling-circle transposition; Virulence gene dissemination; IS91 ecology

1. Introduction

Twenty years ago, IS91 was discovered in α-haemolytic plasmids of Escherichia coli [1]. It was shown to be active in transposition and resulted in the formation of cointegrates between the α-haemolytic plasmid and plasmid pACYC184. IS91-like sequences were found in a number of Hly plasmids belonging to different incompatibility groups. In the α-haemolysin plasmid pHly152, the hly genes were flanked by a direct repeat of IS91-like elements [2]. Furthermore, an IS91 element was located close to an uropathogenic E. coli chromosomal hly determinant [3,4]. These findings suggested IS91-dependent dissemination of hly genes [2]. IS91 exhibits recombinational properties of an IS element [5]. It was characterised as a 1829-bp insertion sequence containing a transposase gene (tnpA) and lacking terminal inverted repeats [6,7]. It inserted specifically at the 3' end of the sequences 5'-CTTG or 5'-GTTC of the target DNA [8] without duplication of the target sequence upon insertion [9]. TnpA_IS91 is related to initiator proteins of rolling-circle replication plasmids and single strand DNA phages [10], having two tyrosines that align with the catalytic tyrosyl residues of the initiator protein gpA of phage φX174. It has been demonstrated that residues Y249 and Y253 of TnpA_IS91 are essential for transposition in vivo [11]. Despite the fact that most transposases act efficiently only in cis [12], TnpA_IS91 acted with similar efficiency in cis as in trans [13]. The IS91 termini play different roles in transposition, and were thus called ori (where transposition starts), and ter (where transposition stops). The ori region comprises the last 82 bp of the right end that contains two internal inverted repeats (IRCR and IRT). Furthermore, ori is absolutely indispensable for IS91 transposition, while deletion of ter results in one-ended transposition [14]. The tetranucleotide 5'-CTTG or 5'-GTTC located 3' adjacent to ori is also required for IS91 transposition [14]. IS91 one-ended transposition results in the insertion of fragments of variable length, flanked by ori at one end and by any available GTTC or CTTG sequence from the donor at the other end, providing the possibility of mobilising sequences adjacent to a single IS91 copy. A rolling-circle mechanism proposed for IS91 transposition is congruent with characteristics described above [14].
IS801 and IS1294 are functional elements related to IS91 [15–17]. Furthermore, several IS91-like elements have been found flanking virulence and toxin genes [18]. Recently, rolling-circle transposons were also detected in a variety of higher eukaryotes, including the higher plants *Oryza sativa*, *Arabidopsis thaliana* and the nematode *Cae- norhabditis elegans* [19]. In the present work, we pursue this genomic analysis by in silico search for the occurrence of new IS91-like elements. We analyze the distribution and characteristics of these IS91 isoforms for evolutionary scope. Finally, we propose that an IS91 idiosyncratic transposition mechanism is important in the ecological niche where IS91 family elements are found.

### 2. Materials and methods

#### 2.1. Search for IS91 family isoforms in databases

IS91 family isoforms were searched in GenBank using BLAST on IS91, IS801 and IS1294 as queries [20]. In order to find more remote homologues, protein similarity searches were carried out on protein databases (SwissProt, translated GenBank/EMBL and various genome sequencing projects) using the iterative PSI-BLAST search algorithm [20] with the IS91 transposase sequence as query. Protein sequence alignments were done with CLUSTALW.

### 3. Results and discussion

#### 3.1. Occurrence and distribution of IS91-like transposases in protein databases

Relatives of IS91 transposase (TnpA_IS91) were found in a variety of bacterial genera, such as *Bergeyella*, *Fusobacterium*, *Rhizobium*, *Shewanella*, *Pseudomonas*, *Klebsiella*, *Shigella* and *Escherichia* as previously reported [18]. Interestingly, a recent study reported the existence of highly repeated IS91-like elements in several eukaryote genomes such as *A. thaliana*, *O. sativa* and *C. elegans* [19]. A PSI-BLAST search for TnpA_IS91 relatives converged in the fifth iteration, where all hits corresponded to TnpA proteins. Significant new findings were proteins from *Mesorhizobium loti* (26% identity), *Pseudomonas hut- tiniensis* (17%), *Vibrio cholerae* (16%) and *Salmonella typhi-

![Fig. 1. Alignment of IS91-like transposases. The alignment was constructed with CLUSTALW; the alignment of the helitron sequences was improved by eye. Five distinct most conserved motifs are represented. Residues invariant in at least half of the sequences are shown on a grey background. List of aligned proteins and their corresponding accession numbers by descending order: TnpA from *Pseudomonas pseudoalcaligenes* (AAB94124); TnpA from *S. typhimurium* (CAC81323); TnpA from *V. cholerae* (AAK64580); TnpA from *Pseudomonas syringae* (CAA40540); TnpA from *E. coli* (CAA34970); TnpA from *Ber- geyella zoohelcum* (AAA50501); Helitron 1-TnpA from *O. sativa* (BAA99532); Helitron 1-TnpA from *A. thaliana* (AAD15468); Helitron 1-TnpA from *C. elegans* (AAF60343); gpA protein from bacteriophage 6X-174 (P03631).](image-url)
and IS1294 cluster in a branch of the tree. This implies that the diversity of rolling-circle transposons is broader than that analysed in this work. A detailed analysis of the other tree branches has to wait until more data are available in the databases.

3.2. General properties of IS91 family occurrences in DNA databases

In order to analyse the genetic organisation of sequences containing IS91-like sequences, IS91, IS1294 and IS801 DNA sequences were used as queries in a BLAST search. Only initial hits with high scores that contained adjacent orfs were selected for further analysis. Twelve hits were IS91 isoforms with sequences exhibiting 87–96% nucleotide sequence identity to IS91. Fourteen sequences related to IS1294 and nine related to IS801 also appeared in the database search. Figs. 3–5 represent the genetic organisation of the sequences flanking these hits. Several generalisations could be made upon inspection of these figures.

First, most IS91 family isoforms do not appear as direct repeats flanking adjacent orfs. This result is somehow unexpected for classical IS elements, since well known DDE class insertion sequences disseminate genes through the formation of composite transposons [22]. Formation of such a compound transposon was proposed as a mechanism for IS91-mediated dissemination of the hly operon, since two IS91-like elements were found flanking the hly operon of the plasmid pHly152 [2]. Interestingly, we never detected bona fide transposition of this compound transposon. Rather, the left copy of IS91 was able to promote replicon fusions when complemented by the wild element. As we will see below, formation of a compound transposon is not necessary for IS91-mediated transposition of adjacent sequences since, contrary to other IS elements, one-ended IS91 transposition directly results in transloca-

Second, most occurrences contained ori plus, an adjacent consensus tetranucleotide. For instance, six of the IS91 isoforms (Fig. 3) contained ori91 (isoforms A1, A2, A3, A4, A6 and A10). Four of the ori91 were linked to the 3’ end of the consensus sequences 5’-CTTG (isoforms A3, A6 and A10) or 5’-GTTC (isoform A4). Likewise, 10 IS1294 isoforms (Figs. 3 and 4) contained ori1294, many of the versions truncated (isoforms B5, B9 and B10 started within ori, while isoforms A1, B4 and B6 contained only 38, 58 and 56 bp of the terminal part of ori, respectively). In six instances, ori1294 was adjacent to the tetranucleotide 5’-GTTC (isoforms B3, B4, B6, B7, B11 and B12). Finally, seven of the IS801 isoforms contained ori801 (Fig. 5). Of these, two were placed at the 3’ end of 5’-CTTG (isoforms C6 and C7) while two more laid adjacent to 5’-GTTC (isoforms C1 (right one) and C3). In the instances where the canonical tetranucleotides were not found, the sequences found immediately adjacent to ori do not display obvious similarity to the consensus tetranucleotides. This overall result can be interpreted if we take into account that IS91 is able to perform one-ended transposition, in which only the ori91 region is indispensable [14]. Interestingly, the tetranucleotides 5’-CTTG or 5’-GTTC adjacent to ori91, which is known to be also essential for IS91 functionality [14], was found in many occurrences. Also taking into account the fact that TnpA_IS91 transposes efficiently even if the transposase is located in trans [13], partial IS91 isoforms containing ori91 and the adjacent tetranucleotide could transpose by the action of a transposase supplied by a functional element located anywhere else. For instance, a sequence of 392 bp, 96% identical to the segment comprising the end of tsp4 and ori91 (isoform A1), precedes the eltAB region of the Ent plasmid in the enterotoxigenic E. coli strain S6 [23]. The eltAB region codes for both subunits of the heat-labile enterotoxin, one of the main virulence determinants of the ETEC strains. This IS91 segment contains ori91 but lacks the rest of the element as well as the consensus tetranucleotide, being thus TnpA+. To render it transposable, the following experiment was carried out [23]. The essential 5’-CTTG tetranucleotide was located adjacent to ori91 via homologous recombination with a wild-type IS91 provided in trans. Thus, a functional ori91 was generated close to eltAB. Its functionality was proven by a subsequent experiment in which ori91–eltAB transposed to a new location, integrating precisely so that ori91 was adjacent to the 3’ end of a GTTC in the target DNA. This experiment strongly suggests that IS91 is involved in the eltAB rearrangements that led to its dissemination among a variety of plasmids and bacteria. Therefore, it is not surprising to find that IS91 elements, even active elements, are widely distributed among ETEC strains [23]. Besides, IS1294-like elements (isoforms B1 and B2) also appeared flanking genes eltAB in various Ent plasmids [23]. Curiously, 38 nucleotides belonging to the end of ori1294 lay close together to ori91 in isoform A1, perhaps suggesting
that IS/294 may have kidnapped \textit{eltAB} from an original IS91-based transposon, or vice versa.

Third, most of the isoforms corresponded to partial IS91 elements. Only in one of the 12 occurrences did an intact IS91 element appear (isoform A4). It would be expected that the transposase included in this isoform is functional, because it was 90\% identical to TnpA\_IS91 and showed all the invariant amino acid residues of the family (Fig. 1). In the same way, truncated IS/294 and IS/801 isoforms were more abundant than entire elements (three out of 14, and one out of nine, respectively). There is no obvious reason to explain this fact, except, perhaps,

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Bacterial source</th>
<th>% identity to IS91</th>
<th>Diagram representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>pEnt (ETEC S6)</td>
<td>96%</td>
<td>orf1294 eltAB</td>
</tr>
<tr>
<td>A2</td>
<td>pWR100 pWR501 (\textit{S. flexneri})</td>
<td>90%</td>
<td>IS629 ISJ IS629 \textit{virA} \textit{virG}</td>
</tr>
<tr>
<td>A3</td>
<td>Idem</td>
<td>89%</td>
<td>IS/1294 \textit{ipaH1.4}</td>
</tr>
<tr>
<td>A4</td>
<td>Chromosome O cluster (\textit{S. sonnei} 53G)</td>
<td>92%</td>
<td>wbgY' wbgW wbgX wbgY wbgZ \textit{aapZ} IS629 IS911</td>
</tr>
<tr>
<td>A5</td>
<td>Idem</td>
<td>87%</td>
<td>IS630 wzz wbgT</td>
</tr>
<tr>
<td>A6</td>
<td>pO157 (EHEC O157:H7)</td>
<td>95%</td>
<td>IS609 \textit{katP}</td>
</tr>
<tr>
<td>A7</td>
<td>pB171 (EPEC B171)</td>
<td>91%</td>
<td>ISJ toxB toxB</td>
</tr>
<tr>
<td>A8</td>
<td>pEnt (ETEC)</td>
<td>87%</td>
<td>orf1 orf2 orf3 orf4 orf5 orf6 pilin CS3</td>
</tr>
<tr>
<td>A9</td>
<td>pEnt (ETEC)</td>
<td>87%</td>
<td>IS102 cssA cssB cssC cssD</td>
</tr>
<tr>
<td>A10</td>
<td>pEnt (ETEC)</td>
<td>88%</td>
<td>\textit{cfd}</td>
</tr>
<tr>
<td>A11</td>
<td>pRI8801 (ETEC G7)</td>
<td>89%</td>
<td>faeA faeB</td>
</tr>
</tbody>
</table>

Fig. 3. Diagram representation of IS91 family occurrences. IS91 family elements are represented by grey boxes, in which \textit{ori} sequences are represented by black triangles, and \textit{ter91} by a spotted triangle. \textit{inpt} genes are represented by a grey arrow if complete or partially deleted at the N-terminal region, or by a grey rectangle if deleted at the C-terminal region. White arrows represent \textit{orfs} near IS91-like elements. White boxes with diagonal black lines represent sequences interrupting IS91-like elements. The diagram representations are not drawn to scale. For each isoform, the bacterial source and percentage homology to the prototype family member are indicated. An identifier has been assigned to each isoform for easy reference. ETEC, enterotoxigenic \textit{E. coli}; EHEC, enteroinvasive \textit{E. coli}; EAEC, enteropathogenic \textit{E. coli}; APEC, avian pathogenic \textit{E. coli}; Accession Nos.: A1 (AF190927); A2, A3 (AL391753 and AF348706); A4 (AF294823); A5 (AF455358); A6 (AF074613); A7 (AB024946); A8 (X16944); A9 (U04844); A10 (M55609); A11 (X77671). The location of each isoform is stated either as 'chromosome', as the specific plasmid name if it is known, or as 'plasmid' in case the name of the plasmid is unknown (and the host strain in brackets). Idem: same as above.
that IS91 insertion events are old, and present day sequences are the result of genetic wear and tear due to the accruement of successive genetic events.

Fourth, in most cases other IS elements appear close to the IS91 family elements. Some of the most conspicuous are IS629 (near isoforms A2, A4, B3 and B7), IS100 (near isoforms B3, B7, B10 and C1), IS3 (near isoforms A2 and A7), IS600 (near isoforms A6 and B3), IS630 (near isoforms A5 and B11) and IS911 (near isoforms A4 and B10). Even IS91 and IS1294 coincide in isoforms A1, A3 and B10. This striking accumulation of IS elements was the rule rather than the exception, as can be seen in Figs. 3–5. A rationale for this could be that activities of other transposable elements lead to partial deletion of IS91 elements, rendering one-ended versions more efficient in hitchhiking adjacent genes. Interestingly, IS91 (four cop-
ies) and IS1294 (six copies) cohabited in the large virulence plasmids pWR100 and pWR501 of *Shigella flexneri*, both prolific in IS elements and virulence-associated genes [24,25]. These gene arrangements have markedly lower G+C composition than the average content of the plasmid and the *Shigella* chromosome. At least in one case, IS91 and IS1294 were located in close proximity. It is tempting to speculate that IS91 family elements are hot spots for insertion of other transposable elements, the accumulation of IS elements being the remnants of the evolutionary history of IS-mediated gene flux.

Fifth, IS91 family elements occur in the neighbourhood of genes related to virulence and pathogenicity (all hits except isoforms B12 and C9). Many IS91 and IS1294 isoforms occurred in plasmids and/or chromosomes of pathogenic strains of *E. coli* and *Shigella* that cause several infections of the gastrointestinal–urinary tract. Ten IS91 isoforms and thirteen IS1294 isoforms occurred in plasm-
mids, while two and one, respectively, occurred in chromosomes (Figs. 3 and 4). IS801 isoforms sat in plasmids and/or chromosomes of various pathovars of *Pseudomonas syringae*, responsible for different plant diseases: six isoforms occurred in plasmids and three in chromosomes. The association between IS91-like elements and virulence is perhaps the most intriguing feature that sticks out from Figs. 3–5. The fact that most determinants are hosted in mobile genetic elements has obvious implications in the context of horizontal gene transfer.

### 3.3. A specific link between IS91 family elements and pathogenicity determinants of animal cells

The successful colonisation of a new ecological niche by bacteria is the result of a conglomerate of factors: invasion and dissemination, obtaining nutrients, protection against host defenses, adherence, regulation of virulence genes, etc. A compilation of virulence genes occurring in the vicinity of IS elements is presented in Table 1. This possibility is underscored by the presence of numerous mobile elements in the vicinity of virulence genes, like transposable elements (see Figs. 3 and 4).

Horizontal gene transfer appears to be responsible for acquisition by *E. coli* and *Shigella* strains of some of the genes described in Table 1. This possibility is underscored by the presence of numerous mobile elements in the vicinity of virulence genes, like transposable elements (see Figs. 3 and 4 and the fourth point of Section 3.2), and of remnant plasmids (for example, a region containing *oriV* of plasmid R100 and the replication protein RepA1 is located immediately upstream from the *sfp* gene cluster [39]). Furthermore, many DNA sequences comprising virulence determinants have a G+C content lower than surrounding sequences (for example, *toxB, css* operon, *cfa d, perABC* operon). In summary, acquisition of IS91-containing virulence plasmids by horizontal gene transfer seems to have been an essential speciation event in the emergence of pathogenic *E. coli* strains.

### 3.4. IS801 is more specifically related to plant pathogens

In contrast to IS91, IS801 copies were found mostly in virulence plasmids and chromosomes of plant-pathogenic variants of *P. syringae*. This is in spite of the similarity in genetic organisation, and transpositional properties detailed in Section 3.2. For instance, TnpA_IS801 is also able to act in *trans* and is capable of one-ended transposition albeit at low frequencies [48]. Furthermore, TnpA_IS91 can recognise IS801 as a terminally deleted derivative of itself [48]. In spite of their similarity, both elements clearly differ in the bacterial pathogens in which they are found: IS91 and IS1294 seem to be animal pathogen-specific, while IS801 is mostly found in plant pathogens.

Eight out of nine IS801 isoforms were found near to virulence and avirulence genes in our database search (Fig. 5). A summary of the characteristics of IS801 neighbour genes is presented in Table 2. The term avirulent is used to describe a potentially virulent pathogen which is

### Table 1

<table>
<thead>
<tr>
<th>Isoform</th>
<th>Neighbour genes</th>
<th>Proteins encoded</th>
<th>Function of proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td><em>virA</em> [24–26], <em>virG</em> [24,25,27]</td>
<td>VirA: member of the invasion regulon; VirG: autotransporter</td>
<td>Invasiveness of <em>Shigella</em>; VirG: actin tail assembly to propel bacteria through the cell cytoplasm and into adjacent cells</td>
</tr>
<tr>
<td>A3</td>
<td><em>ipaH1.4</em> [24,25]</td>
<td>Antigen</td>
<td>Invasion of <em>Shigella</em></td>
</tr>
<tr>
<td>A4 and A5</td>
<td>O-Antigen cluster [28,29]</td>
<td>Antigen</td>
<td>Evasion of host defense</td>
</tr>
<tr>
<td>A6</td>
<td><em>katP</em> [30,31]</td>
<td>Catalase–peroxidase</td>
<td>Protection of <em>Shigella</em> against oxidative stress</td>
</tr>
<tr>
<td>A7 and A8</td>
<td><em>pilin CS3</em> [32] and <em>css</em> operon [33], respectively</td>
<td>Pilin</td>
<td>Colonisation</td>
</tr>
<tr>
<td>A9</td>
<td><em>cfa d</em> [34]</td>
<td>DNA-binding protein, homologous to VirF</td>
<td>Positive regulation of fimbrial genes</td>
</tr>
<tr>
<td>A10</td>
<td><em>faeA</em> [35]</td>
<td>Regulator protein</td>
<td>Negative regulation of fimbrial genes</td>
</tr>
<tr>
<td>A11</td>
<td><em>toxB</em> [36]</td>
<td>Toxin</td>
<td>Toxicity</td>
</tr>
<tr>
<td>B3 and B4</td>
<td><em>ShET2-I</em> [37] and <em>senB</em> [38], respectively</td>
<td>Fimbrial antigens and haemagglutination proteins</td>
<td>Osmotic leak of mucosal epithelium</td>
</tr>
<tr>
<td>B5</td>
<td><em>sfp</em> operon [39]</td>
<td>Fimbrial adhesin</td>
<td>Adhesion of ETEC to the small intestine mucosa</td>
</tr>
<tr>
<td>B6</td>
<td><em>enfA</em> [40]</td>
<td>Fimbrial adhesin</td>
<td>Colonisation</td>
</tr>
<tr>
<td>B7</td>
<td><em>virF</em> [41]</td>
<td>DNA-binding protein</td>
<td>Positive regulation of proteins involved in invasiveness of <em>S. flexneri</em> and <em>Shigella sonnei</em></td>
</tr>
<tr>
<td>B8</td>
<td><em>perABC</em> operon [42]</td>
<td>DNA-binding protein</td>
<td>Transcriptional activation of EPEC virulence genes (bundle-forming pilus operon and <em>eae</em> gene for intimate adherence)</td>
</tr>
<tr>
<td>B9</td>
<td><em>hbp</em> [43]</td>
<td>IgA1 protease-like autotransporter protein</td>
<td>Provides an iron source to EPEC</td>
</tr>
<tr>
<td>B10</td>
<td><em>tsh</em> [44–46]</td>
<td>IgA1 protease-like autotransporter protein</td>
<td>Adhesion and proteolytic activity</td>
</tr>
<tr>
<td>B11</td>
<td><em>sepA</em> [47]</td>
<td>IgA1 protease-like autotransporter protein</td>
<td>Tissue invasion</td>
</tr>
</tbody>
</table>

*IS element notation follows the nomenclature of Figs. 3 and 4.*
Table 2
Association between IS801 elements and virulence determinants

<table>
<thead>
<tr>
<th>Isoform*</th>
<th>Neighbour genes</th>
<th>Encoded protein</th>
<th>Function of the proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>virPphA [49]</td>
<td>Avirulence protein</td>
<td>Control of rapid hypersensitive reaction in soybean</td>
</tr>
<tr>
<td>C2</td>
<td>psvA [50]</td>
<td>Virulence protein</td>
<td>Required for pathogenesis</td>
</tr>
<tr>
<td>C3</td>
<td>ptc [51]</td>
<td>Homologous to citoquin biosynthetic protein</td>
<td>Tumour formation. Olive-knot disease</td>
</tr>
<tr>
<td>C4, C5, C6 and C7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>avrB [52], avrPphC [53], avrRpml [54] and avrA [55], respectively</td>
<td>Avirulence proteins</td>
<td>Activation of plant defense. Pathogenicity in the absence of the interacting resistance gene in the host plant</td>
</tr>
<tr>
<td>C8</td>
<td>taa operon [56]</td>
<td>Auxin indole-3-acetic acid biosynthetic proteins</td>
<td>Tumour induction. Oleander-knot disease</td>
</tr>
</tbody>
</table>

*IS element notation follows the nomenclature of Fig. 5.

unable to induce disease symptoms in specific cultivars. The phenotype of avirulence is a conditional expression since a pathogen which is avirulent on one cultivar can be virulent on a different cultivar. So, certain avirulence genes (avr) of P. syringae, although recognised by their ability to activate plant defenses (the host resistance genes), may also have a role in pathogenicity, constituting virulence genes (vir) in the absence of the interacting resistance gene in the host plant.

There are other examples of IS801 elements in close proximity to genes related to plant pathogenesis, although they did not appear in the sequence database. For instance, an IS801-like element is located nearby the avirulence gene avrD [57], and another interrupts an orf potentially encoding an avirulence gene of P. syringae pv. phaseolicola race 1 [58]. IS801 transposition into avirulence genes would generate new pathogen races as avr genes are inactivated. This mechanism allows evolution of plant pathogens to avoid recognition and response by hosts overcoming their resistance genes. Besides, multiple copies of IS801-like elements are present in some COR plasmids, flanking the COR biosynthetic gene cluster responsible for production of the phytotoxin coronatine, that substantially contributes to the virulence of several pathovars of P. syringae [59].

3.5. IS91 family elements in the vicinity of antibiotic resistance genes: an exception to the rule

Compound transposons usually disseminate antibiotic resistance genes. Nevertheless, as can be inferred from the discussion above, IS91-like elements rarely appear close to these genes. Only one of the hits of our database search, isoform B12, occurred in the proximity of an antibiotic resistance gene. In this hit, the prototype IS1294 was a neighbour of gene aph in plasmid pUB2380 [16,17]. Gene aph encodes an aminoglycoside-3'-phosphotransferase type 1 that confers resistance to kanamycin.

Another exceptional case is the multiple-antibiotic resistance deletable element (MRDE) on the chromosome of S. flexneri 2a strain YSH6000, encoding resistance determinants to streptomycin, ampicillin, chloramphenicol and tetracycline. This MRDE was found to be flanked by two identical intact IS91 elements. Deletion of MRDE took place by recA-mediated homologous recombination between two flanking IS91 elements, resulting in a single IS91 copy spanning the two original IS91 loci [60]. Maybe the same mechanism of homologous recombination is responsible for acquisition of MRDE antibiotic resistance genes.

The last exception in the IS91 family is provided by an IS801 element in E. coli BM2506 plasmid pTZ3721. This plasmid contains five orfs whose deduced products are similar respectively to penicillin-binding protein 4 of Streptomyces lactamharus, to repressor protein AcrR of acrilavine resistance operon akrAB, to the enzyme RdmC involved in biosynthesis of the antibiotic aklavinone, to macrolide 2'-phosphotransferase II which inactivates macrolide antibiotics, and to an IS801 transposase-like protein [61]. DNA sequence analysis suggests that the gene mphB (encoding macrolide 2'-phosphotransferase II) is not endogenous to E. coli. The proximity of an IS801 copy would be the key to interpret the origin of mphB.

3.6. A role for IS801 in the evolution of degradative pathways

TnpA_IS801-like proteins are coded in Pseudomonas sp. strains in the proximity of genes encoding organic compound degradation-related activities. For example, three homologues of TnpA_IS801 are located near genes degrading atrazine on the Pseudomonas sp. ADP atrazine catabolic plasmid pADP-1 [62]. Atrazine is an aromatic compound used as herbicide. Three genes are involved in atrazine degradation: atzA, atzB and atzC. One IS801 homologue is in the vicinity of each gene. For the last 40 years, more than one billion pounds of atrazine have been applied to soils. This application has provided selection pressure for the evolution of new pathways of microbial atrazine metabolism. The G+C percentages of atzA and atzB are within the range of the total Pseudomonas sp. DNA. Nevertheless, the atzC G+C content is significantly
less [62]. *Pseudomonas pseudoalcaligenes* JS45 grows on nitrobenzene by using a partially reductive pathway in which the intermediate hydroxyaminobenzene is enzymatically rearranged to 2-aminophenol by hydroxyaminobenzene mutase (HAB mutase). Two isozymes are encoded by chromosomal orfs habA and habB and divergently transcribed. There is an orf 48% similar to TnpA_IS081 between habA and habB [63]. The significant differences among the G+C contents of habA (53%), habB (71%), and *P. pseudoalcaligenes* genome (62–64%) suggest that the genes encoding the two HAB mutase isozymes may have evolved separately and in hosts other than *P. pseudoalcaligenes*.

A last curiosity, an IS081 isoform C9 is close to the ethylene-forming enzyme gene efe [64]. Ethylene is a hormone involved in the ripening of fruits. The enzyme is produced by several pathovars of *P. syringae* in their secondary metabolism.

### 3.7. Concluding remarks

It seems clear that one-ended transposition is the main avenue for IS91-mediated gene propagation. The fact that IS91 is able to perform one-ended transposition with high efficiency raises the possibility of IS91-mediated gene dissemination without the need to assemble a compound transposon. TnpA_IS91 trans activity implies that ‘non-autonomous’ IS91 isoforms containing orf91 and the consens flanking tetrancleotide 5'-CTTG or 5'-GTTC would be transposed. A significant proportion of the IS91 isoforms found in our genomic analysis fulfill these requirements, as discussed above. As ter91 is known not to be 100% efficient as a transposition terminator, the displaced strand during IS91 transposition may contain genes adjacent to the element. Looking at the neighbourhood of IS91 family occurrences, the striking finding is that rolling-circle transposable elements seem to be preferentially involved in the dissemination of pathogenicity-related genes. IS91 and IS1294 have specialized in animal cell enterobacterial pathogens, while IS801 is a *Pseudomonas* element involved in the dissemination of plant pathogenicity determinants and operons for degradation of xenobiotic compounds. Curiously enough, IS91 family elements are rarely found flanking antibiotic resistance genes. Apparently, IS91-like elements were not the best suited for the rapid dissemination mechanisms that accompanied massive antibiotic treatments in hospital for the last 50 years [65]. It seems, on the other hand, that they have been evolutionarily successful in the dissemination of pathogenicity determinants, in evolutionary schemes that last millions of years, the time elapsed in the adaptation of bacteria to their specific hosts [65]. During these long periods of time, many IS elements have left their traces, so the historical reconstruction of the transposition events pertinent to present day structures (shown in Figs. 3–5) is not always feasible.

### Acknowledgements

This work was supported by a DGICYT grant (PB98-1106) from the Ministerio de Ciencia y Tecnología (Spain) to F.C. The authors want to thank the European Concerted Action MECBAD (EU-BIOTECH: BIO4-CT98-0099) for providing a forum for meetings and discussion.

### References


FEMSEC 1406 5-11-02


