IS1403 and IS1404: analysis and distribution of two new insertion sequences in Xanthomonas campestris

Yung-An Lee1 and Shu-Ping Chiu

Department of Biology, Fu Jen Catholic University, Hsin Chuang 24205, Taipei, Taiwan, Republic of China

(Received October 9, 1997; Accepted April 28, 1998)

Abstract. A repetitive sequence in Xanthomonas campestris pv. juglandis was identified as a new insertion sequence, IS1403, by nucleotide sequence analysis. A homologous IS, IS1404, was also cloned from X. campestris pv. campestris. Both shared at least 91% identity on a nucleotide level. The sequences had 42-bp terminal inverted repeat sequences with eleven mismatches, and were flanked by 4-bp target site duplications. Comparison of target insertion sites flanking IS1403 and IS1404 revealed no similarities, and the G+C content of the flanking sequences of IS1403 was 45%, but that of IS1404 was 74%, indicating the elements transpose quite randomly and do not prefer A-T-rich regions. IS1403 and IS1404 contained two GTG-initiated open reading frames (orfA and orfB), and orfB was in phase -1 relative to orfA. There was an A,G motif followed by a potential stem-loop structure between two reading frames, which may have promoted a -1 translational frameshift to produce a transframe protein. The deduced amino acid sequences of orfA and orfB contained a potential α-helix-turn-α-helix DNA-binding motif and a D,D(35)E domain of transposases, respectively. Based on the features discussed, IS1403 and IS1404 were placed among the IS3 family with IS407 from Pseudomonas cepacia, IS476 from X. campestris pv. vesicatoria, IS1222 from Enterobacter agglomerans, and ISR1 from Rhizobium lupini. The insertion sequences were widely distributed and existed in multiple copies in most pathovars of X. campestris, but other plant pathogenic bacteria, such as P. syringae pv. api, P. syringae pv. tabaci, P. syringae pv. glycinea, P. syringae pv. coronafaciens, Erwinia amylovora, E. carotovora, and E. cyripedii, did not contain this insertion sequence based on Southern hybridization tests.

Keywords: Erwinia; Insertion sequence; IS1403; IS1404; Pseudomonas; Xanthomonas campestris.

Introduction

Insertion sequences (IS) are short segments of DNA, typically 1–2 kilobase pairs (kb) in length with inverted repeats (IRs) at each end. They can transpose to various sites within the chromosomes and plasmids, and duplicate 3- to 20-bp target sequences upon insertion (Galas and Chandler, 1989; Iida et al., 1983). ISs may mediate genomic rearrangement (Chow and Broker, 1978; Saedler et al., 1980), and regulate gene expression by insertional inactivation and/or polar effects on flanking genes (Iida et al., 1983; Schwartz et al., 1988). Thus overall, ISs have a strong impact on the genetic variability of microbial populations.

The genus Xanthomonas Dowson 1939 contains gram-negative, usually yellow-pigmented bacteria that occur worldwide and cause plant diseases. Several insertion sequences and transposons have been discovered in Xanthomonas spp. ISXc4 and ISXc5, were isolated from indigenous plasmids of X. campestris pv. citri and are capable of transposition in Escherichia coli (Tu et al., 1989). Subsequently, IS476 was discovered in copper-tolerant strains of X. campestris pv. vesicatoria. Transposition of IS476 on avirulence gene avrBs1 influences both induction of hypersensitivity and bacterial growth in planta. The IS476 shows similarity to E. coli IS3 and is a member of IS3 family (Kearney and Staskawicz, 1990; Steibl and Lewecke, 1995). IS1051 was found in the genome of X. campestris pv. diffenbachiae, shared significant homology with E. coli IS5, and could be used as a probe to characterize strains from the pathovar diffenbachiae (Berthier et al., 1994). Recently, another insertion sequence, ISXC6, was isolated from X. campestris pv. campestris and had no significant homology to other sequences in database (Weng et al., 1997). In this paper, we report that a repetitive sequence from X. campestris pv. juglandis is also an insertion sequence, given the name IS1403, by the Plasmid Reference Center (Stanford, CA, USA). The sequences contain general characteristics of insertion sequences in IS3 family. Furthermore, we compare IS1403 with a homologous IS, IS1404, isolated from X. campestris pv. campestris, and describe their distribution in plant pathogenic bacteria.

Materials and Methods

Bacterial Strains, Plasmids and Culture Conditions

Xanthomonas campestris pv. juglandis C5 was isolated from infested walnut buds as described in Lee et al. (1994), and X. campestris pv. campestris XCC1-1 was isolated from diseased cabbage in this study. Plasmid

---

1Corresponding author: Fax: 886-2-2902-1124; E-mail: bio1007@fujens.fju.edu.tw
pISJ12 contains IS1403 on a 2.0-kb PstI-ClaI fragment ligated into pBluescript SK(+) (Stratagene Co., La Jolla, CA), and plasmid pISC11 has IS1404 on a 2.0-kb SacI-SmaI fragment. Whenever an E. coli host was necessary, strain DH5α (BRL Life Technologies, Inc., Gaithersburg, MD) was used. Escherichia coli and X. campestris were cultured in Luria-Bertani medium (Miller, 1972) supplemented with 1.5% Bacto agar (Difco) for plates at 37°C and 28°C, respectively. Ampicillin (50 µg/ml) was added as necessary to maintain selection of the resistance gene.

**General DNA Manipulations**

Mini-scale preparations of E. coli plasmid DNA were made by the alkaline lysis method (Sambrook et al., 1989). Total genomic DNAs of xanthomonads were isolated by a CTAB method (Wilson, 1987). Restriction endonuclease treatments, DNA ligation, transformation, Southern and dot blot assays were made as described by Sambrook et al. (1989). In blot assays, 0.5 µg of DNA were blotted onto Nytran membranes (0.45-µm pore size; Schleicher & Schull, Keene, NH). Fragments used in blot assays as probes were separated on agarose gel, and extracted with a QIAEX II Gel Extraction Kit (QIAGEN Inc. Chatsworth, CA). The fragments were then labeled with digoxigenin-11-dUTP (DIG) using the Genius nonradioactive DNA labeling and detection kit (Boehringer Mannheim Biochemicals, Indianapolis, IN). Prehybridization, hybridization, and washing for DIG-labeled probes were performed at 68°C according to the manufacturer’s protocol.

**DNA Sequencing**

Subclones in the pBluescript SK(+) vector or in the pGEM-3Zf(+) (Promega Corp., Madison, WI) were produced by digestion with appropriate restriction endonucleases. Double-stranded sequencing was performed by Sanger dideoxy chain termination method (1977) adapted for the AutoRead Sequencing Kit (Pharmacia Biotech, Uppsala, Sweden) with standard universal and reverse M13 primers. The alignment and analyses of sequences were accomplished by using the programs of BIONET (Intelligenetic, Palo Alto, CA) and the Pileup program of GCG (Devereux et al., 1984) and adjusted to a better match by visual examination. For amino acid sequence comparison, conservative amino acid substitutions were defined as residues belonging to one of the following six groups: Cys; Ser, Thr, Pro, Ala, Gly; Asn, Asp, Gln, Glu; His, Arg, Lys; Met, Ile, Leu, Val; Phe, Tyr, Trp.

**Results and Discussion**

**Identification of a Repetitive Sequence in X. campestris pv. juglandis and Isolation of a Homologous IS from X. campestris pv. campestris**

In the course of the study of a copper-resistance locus in the genome of X. campestris pv. juglandis C5 (Lee et al., 1994), it was observed that a 3.8-kb HindIII fragment was present in multiple copies in X. campestris pv. juglandis genome when the fragment was used as a probe in the Southern hybridization with HindIII-digested genomic DNA of X. campestris pv. juglandis. To further localize the repetitive sequence, several subclones from the fragment were used as probes to hybridize with HindIII-digested genomic DNA. The Southern hybridization using 2.0-kb PstI-ClaI fragment as a probe showed the same repetitive pattern as 3.8-kb HindIII fragment; in contrast, the probes outside the 2.0-kb fragment hybridized only to a single genomic restriction fragment, indicating that the repetitive sequence was within the 2.0-kb PstI-ClaI fragment. The 2.0-kb fragment was ligated into pBluescript SK(+), resulting in plasmid pISJ12. The insertion sequence-like element in the fragment was named IS1403 afterward.

To isolate a homologous IS from X. campestris pv. campestris XCC1-1, a 0.5-kb AvalI internal fragment of IS1403 was used as a probe to hybridize with HindIII-digested total genomic DNA of X. campestris pv. campestris. Multiple copies were also found, and one of them, a 3.0-kb HindIII fragment, was purified from a gel, ligated to HindIII-digested pBluescript SK(+) vector and transformed into E. coli DH5α. The location of IS1403 homolog in the fragment was within 2.0-kb SacI-SmaI fragment, determined by restriction mapping and hybridization with the 0.5 kb AvalI fragment. This 2.0-kb fragment was ligated into pBluescript SK(+), resulting in plasmid pISC11. The insertion sequence-like element in the fragment was named IS1404. Double-stranded pISJ12 and pISC11 DNAs and their subclones were sequenced.

**Sequence Analysis and Feature of the Insertion Sequences**

The presence of imperfect inverted repeats at the ends of the repetitive sequences indicated the presence of a potential insertion sequence. IS1403 and IS1404 shared 91% identity on a nucleotide level. Both were 1,203 nucleotides long (Figure 1) and had putative 42-bp terminal inverted repeats (IR) with 11-bp mismatches. The overall G+C contents of IS1403 and IS1404 were 61.9% and 60.8%, respectively, similar to that of the whole X. campestris chromosome (Bradbury, 1984; Lin and Tseng, 1997).

IS elements characteristically generate small, directly repeated duplications of the target DNA at the point of insertion, and the length of the duplication, but not the sequence, is usually specific to a particular IS element. Thus, the ISs are grouped according to the size of these duplications (Calos and Miller, 1980; Galas and Chandler, 1989). IS1403 and IS1404 are flanked by 4-bp target site duplications (Figure 2) and should be assigned to the IS3 family. ISs in IS3 family generate a 3 or 4-bp repeated sequence. For example, IS3 and IS911 generated 3-bp duplications, and IS407, IS476, and ISR1 generated 4-bp target duplications (Wood et al., 1991). In addition, comparison of the nucleotide sequence of IS1403 with inser-
Figure 1. Nucleotide sequence of IS1403. The deduced aa sequence of ORF is given below the nt sequence. The left and right terminal IR sequences of IS1403, LIR and RIR, are underlined. A RBS is also underlined. The A-G frameshift motif is boxed. The motif is followed by internal IR indicated by solid arrow, which may form a potential pseudoknot. Translation of orfA is shown, starting at position 69 and extending for 87 amino acids, and that of orfB starting at position 356 and extending for 271 amino acids. A homologous IS, IS1404, was cloned from X. campestris pv. campestris. Both elements are 1,203 nucleotides long and share 91% identity on a nucleotide level. IS1403 and IS1404 are listed in the GenBank library under accession Nos. U45993 and U45994, respectively.

With the exception of IS476, IS1403 had no significant homology to other IS elements from X. campestris in database.

The target insertion sites of IS1403 and IS1404 were analyzed. Comparison of a 30-bp sequence flanking IS1403 and IS1404 elements revealed no similarities (Figure 2). The G+C content of the flanking sequences of IS1403 is 45%, but that of IS1404 is 74%, indicating that IS1404 may not prefer A-T rich regions as mentioned in some other ISs (Vaughan and de Vos, 1995).

Moreover, the IRs of IS1403 and IS1404 are similar to those of the IS3 family. Within the IRs, three sections can be recognized (Figure 3). The length of first two sections is 20 bp, but that of the last section is variable.
Figure 2. Sequences of the 5' and 3' flanking regions of IS1403 and IS1404. Duplicated target sequences are underlined. The flanking sequences of IS1403 are not similar to those of IS1404.

Table 1. Percent identities of IS1403 and IS elements of the same family.

<table>
<thead>
<tr>
<th>IS</th>
<th>DNA sequence identity (%)</th>
<th>ORFA amino acid identity (%)</th>
<th>ORFB amino acid identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS3</td>
<td>51</td>
<td>31</td>
<td>25</td>
</tr>
<tr>
<td>ISR1</td>
<td>61</td>
<td>58</td>
<td>35</td>
</tr>
<tr>
<td>IS150</td>
<td>47</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>IS407</td>
<td>61</td>
<td>20</td>
<td>53</td>
</tr>
<tr>
<td>IS476</td>
<td>61</td>
<td>54</td>
<td>47</td>
</tr>
<tr>
<td>IS861</td>
<td>46</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>IS911</td>
<td>49</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>IS1222</td>
<td>58</td>
<td>57</td>
<td>48</td>
</tr>
<tr>
<td>IS1404</td>
<td>91</td>
<td>95</td>
<td>97</td>
</tr>
</tbody>
</table>

aSee Results and Discussion for references for IS elements.
bCalculated by using the FASTDB program of IntelliGenetics.

and its maximal length is 17 bp. There is an appreciable divergence between sections. The consensus sequences of IR are TGANCTGCCCCCAN3-7-TGN3-6AGN2-10-TN2-7AGN. Most IRs terminate in 5'-TG-3' at 5' end in the first subunit. This terminal dinucleotide was shown to be the substrate for transposases, carrying the D,D(35)E motif (Khan et al., 1991; Steibl and Lewecke, 1995). IS1403 and IS1404, however, terminate in 5'-CG-3' at both ends. To our knowledge, this is the first report indicating members in IS3 family can terminate in different dinucleotides. The importance of other consensus nucleotides, such as TG and AG in the second section and T and AG in the third, is not known. Since these three sections begin with T and are almost the same length, one possibility to generate these sequences is that three sections might be homologous and current IRs might have been generated in the past by duplicating a 20-bp or less repeating unit, after which mutations accumulated, including point mutation, deletions, and insertions (e.g. gaining an additional C) to create divergence.

Characterization of orfA and orfB

Sequence analysis revealed that IS1403 and IS1404 contained two open reading frames (orf) indicated in Figure 1. No consensus E. coli -10 and -35 promoter sequences (Hawley and McClure, 1983) were found in the region upstream from these coding regions. The actual promoter sequences have not yet been defined. The orfA begins with an GTG at position 69 and ends at position 332. The GTG is preceded at a distance of 6 bp by a sequence, AGGAGA, resembling the consensus Shine-Dalgarno ribosome-binding site sequence (Shine and Dalgarno, 1974; Lin and Tseng, 1997). The orfA could encode a protein of 87 amino acids. The orfA is followed by orfB in phase -1 relative to orfA. For orfB, two potential start codons are in IS1403, ATG at position 350 and GTG at 356, but the ATG at position 350 is absent in IS1404. Therefore, the GTG at position 356 is assumed to be the correct start codon. The orfB extents to position 1171, hypothetically encoding a gene product of 271 amino acids. Such genetic organization with two consecutive open reading frames flanked by imperfect terminal repeats is common to members of the IS3 family (Prére et al., 1990).

ORFA (gene product of orfA), with a size of 87 aa, displays several characteristics of DNA-binding proteins. Comparison of this ORFA of IS1403 with the ORFs of other ISs of IS3 family gave values of 20–58% identity. When conservative amino acid changes were considered, the similarity became 29–79%. The highest identity scores were shown by ISR1 from Rhizobium class IV (Priefer et al., 1989) and IS476 from X. campestris pv. vesicatoria (Kearney and Staskawicz, 1990). Further sequence analysis revealed that the ORFA polypeptide had a potential α-helix-turn-α-helix DNA-binding motif between amino acid 23 and 42. This motif and its flanking regions display significant similarity to those of some characterized upstream ORF of the IS3 family (Figure 4). The analysis predicted that a glycine in residue 9 of the motif breaks the alpha helix and begins the turn, similar to several other bacterial DNA-binding proteins (Galas and Chandler, 1989; Brennan and Matthews, 1989). This conserved domain seems the most likely candidate for DNA-binding capacity of transposases, specifically interacting with the ends of the element.

Comparison of the ORFB with those of other IS3 family members gave values of 19–53% identity and 26–72% similarity (Table 1). The highest identity scores were

\[
5' - TCACGCATTTAAGCTGGACCTTTTG - IS1403 - TTTGCTGCGAACCACGCTTATTAGTGGAG - 3'
\]

\[
5' - TCCGCGGTCTATCGCGCCTGGTGCTGGACUG - IS1404 - CACGGAACACGGCCCGAGTTCACGACGCCG - 3'
\]
Figure 4. Amino acid sequences comparison between several ORFA of ISs in the IS3 family. Dots represent gaps introduced for optimal alignment. The numbers on the left and right indicate the number of residues from the N terminus. Identical amino acids and conserved substitutions are in boldface type. Potential \( \alpha \)-helix-turn-\( \alpha \)-helix regions are boxed. The boxed regions are aligned according to the study of Pröre et al. (1990). Specific references for the sequence of each IS elements are given in the text. The sequences were aligned by using the Pileup program of GCG (Devereux et al., 1984).
IS407 from *P. cepacia* (Wood et al., 1991) and IS1222 from *Enterobacter agglomerans* (Steibl and Lewecke, 1995). All of these ORFBs contained a D,D(35)E region (two consecutive aspartic acid residues and a glutamic acid residue separated by 35 residues). This region is highly conserved for most IS3 elements (Figure 5). Except for this region, comparison of other portions shows a lower degree of similarity. The D,D(35)E motif is thought to be at the active site of transposase (DNA-binding/cutting domain), taking part in binding, cutting, and transfer of DNA strands during transposition (Fayet et al., 1990; Khan et al., 1991; Kulkosky et al., 1992; Sekine et al., 1994). Closer inspection of the sequences revealed that six other amino acids also consistently existed in the transposases, and the regions of ISR1, IS407, IS476, IS1222, IS1403, and IS1404 are in the same length and highly homologous (Figure 5).

Moreover, there is an A,G motif followed by a potential pseudoknot structure with a spacing of 6 nucleotides between two reading frames, as marked in Figure 1, displaying a high similarity to the -1 ribosomal frameshift site in the IS3 family insertion sequences and in certain retroviruses (Brierly and Boursnell, 1987; Brierly et al., 1989; Prêtre et al., 1990). This structure was shown to be necessary for ribosomal to slip one base to 5'-end to make reading frame in the same frame of *orfB* (Sekine et al., 1994). Three elements in the IS3 family, IS3, IS150, and IS911, have been shown to produce transposase by a programmed -1 translational frameshifting between the two reading frames, leading to the combination of a potential DNA-binding domain in ORFA and the D,D(35)E motif in ORFB into a single transframe protein (Chandler and Fayet, 1993; Polard et al., 1991; Vögele et al., 1991).

**Genomic Distribution of IS1403**

Thirty nine pathovars of *X. campestris* were examined for the presence of IS1403. Using an internal fragment of IS1403, 0.5-kb AvaII fragment, as a probe, dot-blotting was performed with genomic DNA from different pathovars of *X. campestris*. Thirty-two pathovars were hybridized to the AvaII 0.5-kb fragment (Table 2). These results suggest that IS1403 is widely distributed within the pathovars of *X. campestris*. The same approach was used to test other plant pathogenic bacteria. Southern blot analysis of genomic DNAs from pseudomonads *P. syringae* pv. *apii*, *P. syringae* pv. *tabaci*, *P. syringae* pv. *glycinea*, *P. syringae* pv. *coronafaciens*, and *P. syringae* pv. *atrofaciens*, and erwinias *E. amylovora*, *E. carotovora*, *E. cypripedii*, and *E. rhapontici*, showed no hybridization signal. These results indicated that distribution of IS1403 is limited to *X. campestris*.

**Figure 5.** Amino acid sequences comparison between several ORFBs of ISs in the IS3 family. Dots represent gap introduced for optimal alignment. The numbers on the left and right indicate the number of residues from the N-terminus. Identical amino acids and conserved substitutions are in boldface type. Invariant aspartate (D) and glutamate (E) in D,D(35)E region proposed by Kulkosky et al. (1992) are boxed. Other invariant amino acids are indicated by a star at top. Specific references for the sequence of each IS element are given in the text. The sequences were aligned by using the Pileup program of GCG (Devereux et al., 1984).
Table 2. Hybridization of an internal fragment of IS1403 to plant pathogenic bacteriaa.

<table>
<thead>
<tr>
<th>Positive species and strainb</th>
<th>Xanthomonas campestris pathovar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armoraciae XC132; betlicola XB104; campestris XCC1-1; carotae XC139; cassiae XC112; corylina ATCC1931; eucalypti XE104; gardneri XG101; holricola XH13; incanae XI4; juglandis XJ123; khayae XK101; lespedaeae XL1; manihotis XM125; martynicola XM107; melhusi XM123; nakataecorchori XN104; papavericola XP5; patelii XP186; phlei ATCC33805; plantaginis XP195; poae ATCC33804; poinssetticola XP137; pruni ATCC19316; ricini ATCC19317; secalis XT129; spermacoeces ATCC17998; tamarindii XT105; taraxaci XT11; vasculorum ATCC35938.</td>
<td></td>
</tr>
</tbody>
</table>

X. oryzae pathovar
Oryzae ATCC35933; oryzicola XO111

Negative species and strain
X. campestris pathovar
Badri XB103; biophyi XB117; celebensis XC145; citri XCI1-1; phyasdicola XP172; physalisid ATCC17994; vignicola XV18

Pseudomonas syringae pathovar
Apii PA 102; atrofaciens UCBPP 260; coronafaciens UCBPP 470; glycinea UCBPP 527; tabaci PT 124

Erwinia spp.
Amylovora EA169; carotovora subsp. carotovora EC250; cypripedii EC155; rhapontici ER102

aDot blots of total DNA from the strains listed above were hybridized with the DIG-labeled 0.5 kb AvaII fragment of IS1403. The hybridization method is described in Materials and Methods.
bThe strains were obtained from the International Collection of Phytopathogenic Bacteria, University of California, Berkeley, and ATCC (American Type Culture Collection), Rockville, MD, USA.

The number and distribution of IS within the genome changes more rapidly than conventional genetic markers, which may be of important practical value in tracing the genetic ancestry of bacterial strains for classification (Green et al., 1984; Sawyer et al., 1987). In light of the increasing amount of genetic data, the species now classified as X. campestris appears to be heterogeneous. More analyses will be needed to permit the further reinstatement of species from this heterogeneous taxon. Since IS1403 is widely distributed in pathovars of X. campestris, the analyses of presence, copy-number, and similarity on nucleotide level among homologs of IS1403 should be useful for understanding phylogenetic relationship between different pathovars of X. campestris.

Furthermore, in X. campestris pv. vesicatoria pepper race 2 strains, the insertional inactivation of avirulence gene avrBsI caused by IS476 could extend the host range of the pathogen to include plants previously resistant (Kearney and Staskawicz, 1990). It is possible, then, that wide distribution of IS1403 in different pathovars of X. campestris will increase the ability of pathogens to spontaneously mutate to overcome genetically defined disease resistance. Our laboratory has found at least 13 different hybridization banding patterns in X. campestris pv. campestris using an internal fragment of IS1404 as a probe, indicating that there are locational variations of IS1404 in different strains, and that IS1404 might transpose frequently in the field. It should be possible, then, to use IS1404 as a natural tagging transposon and a genetic marker to study the evolution of pathogen populations in the field when plants with specific resistance genes are released.

Acknowledgements. This research was supported by a grant from the National Science Council Project, NSC 85-2311-B-030-002.

Literature Cited


Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic