

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

Yung-An Lee<sup>1</sup> 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<sub>6</sub>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  $\alpha$ -helix-turn- $\alpha$ -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. *apii*, *P. syringae* pv. *tabaci*, *P. syringae* pv. *glycinea*, *P. syringae* pv. *coronafaciens*, *Erwinia amylovora*, *E. carotovora*, and *E. cypripedii*, 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 induc-

tion 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. *diffracta*, shared significant homology with *E. coli* IS5, and could be used as a probe to characterize strains from the pathovar *diffracta* (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

<sup>1</sup>Corresponding author: Fax: 886-2-2902-1124; E-mail: bio1007@fujens.fju.edu.tw

pISJ12 contains IS1403 on a 2.0-kb *Pst*I-*Cla*I fragment ligated into pBluescript SK(+) (Stratagene Co., La Jolla, CA), and plasmid pISC11 has IS1404 on a 2.0-kb *Sac*I-*Sma*I fragment. Whenever an *E. coli* host was necessary, strain DH5 $\alpha$  (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  $\mu$ g/ml) was added as necessary to maintain selection of the resistance marker in pBluescript SK(+).

### 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  $\mu$ g of DNA were blotted onto Nytran membranes (0.45- $\mu$ m 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 *Hind*III 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 *Hind*III-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 *Hind*III-digested genomic DNA. The Southern hybridization using 2.0-kb *Pst*I-*Cla*I fragment as a probe showed the same repetitive pattern as 3.8-kb *Hind*III 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 *Pst*I-*Cla*I 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 *Ava*II internal fragment of IS1403 was used as a probe to hybridize with *Hind*III-digested total genomic DNA of *X. campestris* pv. *campestris*. Multiple copies were also found, and one of them, a 3.0-kb *Hind*III fragment, was purified from a gel, ligated to *Hind*III-digested pBluescript SK(+) vector and transformed into *E. coli* DH5 $\alpha$ . The location of IS1403 homolog in the fragment was within 2.0-kb *Sac*I-*Sma*I fragment, determined by restriction mapping and hybridization with the 0.5 kb *Ava*II 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-

```

1 CGACGTCCCCCAATTTTGAGTAGCCCCCTCAGTTTGGAGTCCAATTCCTACCCCAAGGA
                                     LIR
61 GATTGGACGTGAAGAAGCGTTTACCGAAGAGCAGATCATCGGCTTCCTGCGCGAAGCCG
                                     RBS
   V K K R F T E E Q I I G F L R E A E
121 AAGGCGCGGTGGCGATCAAGGATCTGTGCCGGCGCCATGGCTTCAGCGAGGCCCTCGTACT
   G G V A I K D L C R R H G F S E A S Y Y
                                     AvaII
181 ATCTGTGGCGTAGCAAGTTCGGCGGCATGAGCGTGCCCGATGCCAAGCGGCTCAAGGACC
   L W R S K F G G M S V P D A K R L K D L
241 TTGAGGCCGAGAACCGCGGCTGAAGAAGTTGCTGGCCGAGCAGTTGTTTCGAGAACGACC
   E A E N A R L K K L L A E Q L F E N D L
301 TGATCAAGGATGCGTTGCGGAAAAAGTGGTGAGCGCACCGGCGCGTCGTATGCTGCTGCGG
   I K D A L R K K W *
361 CGAGTGGATCGGGCGTGGTGCCAGCGAGCGTCTGCGCTGGCAGTGATCGGCATGAGCGC
   E W I G R G A S E R R A L A V I G M S A
421 CAGCCGCTGCCCTATTGCCACGCCAAGACCGCAACGGCGAACTCGCGAGCGCATTCTTG
   S A L R Y C P R Q D R N G E L R E R I L
481 TGGCTGGCGCATCGCCATCGTCTGCTATGGCGTGGGGATGATCTATCTCAAACGCGACA
   A L A H R H R R Y G V G M I Y L K L R Q
541 GGAAGGGCGCATCGTGAAC TACAACCGCGTGAGCGGTTGTACC GCGAACAGCAGCTACA
   E G R I V N Y K R V E R L Y R E Q Q L Q
   AvaII
601 GGTCCGACGGCGCAAGCGCAAAAAAGTGCCAAATAGGCGAGCGTCAGCCGCTGCTGCGGCC
   V R R R K R K K V P I G E R Q P L L R P
661 AGCGCAGGCCAACCAGGTGTGGTCGATGGACTTCGTGTCGACCCTCCGCGAAGGCCG
   A Q A N Q V W S M D F V F D R S A E G R
721 AGTGATCAAGTGTCTGGTGATCGTGGACGACGCAACGCACGAAGCGGTCGCCATCGAGGT
   V I K C L V I V D D A T H E A V A I E V
781 GGAGCGCGCATCTCTGGGCACGGGTTGCGCGCTCCTGGATCGGTTGGCACACAGTCG
   E R A I S G H G V A R V L D R L A H S R
841 CGGCCTGCCGAGGTGATCCGCACCGACAACGGCAAGGAGTTTTCGGGTAAGGCAATGGT
   G L P Q V I R T D N G K E F C G K A M V
901 CGCCTGGGCGCATGCCCGTGGCGTGCAGTTGCGGCTGATCCAACAGGCAAACCGAACCA
   A W A H A R G V Q L R L I Q P G K P N Q
961 GAACGCCTACGTCGAATCCTTCAACGGCCGGCTACGGCATGAATGCCTCAACGAGCAGT
   N A Y V E S F N G R L R D E C L N E H W
1021 GTTCCCAGCGTTGCTGCACGCGCACCGAGATCGAACGCTGGCGACGCGAATACAACGA
   F P T L L H A R T E I E R W R R E Y N E
   AvaII
1081 GGACCGACCCAAGAAAGCAATTGGCGGCATGACGCCGGCTGCGTATGCCCAACATCTGGC
   D R P K K A I G G M T P A A Y A Q H L A
                                     EcoRV
                                     SmaI
1141 AAACACCGATATCATTAAACCGGACTCTAAACCAAGCCGCTACTCAGGGCGGGGGATG
   N T D I I N P G L *
                                     RIR
1201 TCG

```

**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<sub>6</sub>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.

tion elements listed in the NIH GenBank Data Bases indicated that the element was related to members of the IS3 family of insertion sequences. For comparative analysis of IS1403 the following well studied IS elements were selected from IS3 family: IS3 (Timmerman and Tu, 1985), IS150 (Schwartz et al., 1988), IS407 (Wood et al., 1991), IS476 (Kearney and Staskawicz, 1990), IS861 (Rubens et al., 1989), IS911 (Prère et al., 1990), IS1222 (Steibl and Lewecke, 1995), IS3411 (Ishiguro and Sato, 1988), and ISR1 (Priefer et al., 1989). The extent of overall nucleotide sequence identity of IS1403 with elements in IS3 family ranged between 46 and 61%. The highest identity scores were ISR1 from *Rhizobium*, IS407 from *P. cepacia*, and IS476 from *X. campestris* (Table 1).

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

5' -TCACGCATTA~~AACTGAAAGTGGGCCTTTTG~~-IS1403-~~TTTGCTGCGAACCAGCCCAATTTAGTTGAG~~-3'

5' -TTCCGCGGCTATCCCCGCGCGGTGCGC~~CACG~~-IS1404-~~CACGGACAACGGCCCGGAGTTCACCAGCCG~~-3'

**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.

IS <sup>a</sup>	IS1403		
	DNA sequence identity (%) <sup>b</sup>	ORFA amino acid identity (%) <sup>b</sup>	ORFB amino acid identity (%) <sup>b</sup>
IS3	51	31	25
ISR1	61	58	35
IS150	47	24	26
IS407	61	20	53
IS476	61	54	47
IS861	46	21	24
IS911	49	20	19
IS1222	58	57	48
IS1404	91	95	97

<sup>a</sup>See Results and Discussion for references for IS elements.

<sup>b</sup>Calculated 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 TGANCTGCCCCAN<sub>3-7</sub>-TGN<sub>3-6</sub>AGN<sub>2-10</sub>-TN<sub>2</sub>-AGN. 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* extends 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 (Prère 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  $\alpha$ -helix-turn- $\alpha$ -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

		----- 20 nt -----   ----- 20 nt -----   ----- 17 nt -----
IS 3	R	TGATCCTACCC-ACGTAA--TATGGAC-ACAGGCC-----TAAGCG--AG
	L	TGATCTTACCC-AGCAA---TAGTGGACACGCGGC-----TAAGTG--AG
IS 150	R	TGTACTGACCCCAAAAAGT-TGGAC---AGTTAAA-----CACGA
	L	TGTACTGCACCCATTTTGT-TGGAC---AGTTAAA-----TGGAA
IS 407	R	TGACCTGCCCCCATCAA---TAGGGCCAATGGGCTC-----TAGCAA--AGTCC
	L	TGACCTGCCCCCTGCAAAC-AGGGCC--AGCCGGAGTC--TAGCAA--AGTTC
IS 476	R	TGACCTGCCCCCATCG----TCCGTACCAGCTGAAGC---TATAAA
	L	TGACCTGCCCCCAC-----TGAGCCGTACCAGTGATTACTGATAA
IS 861	R	TGTACTGACCCCAAAAAGT-TGGAC---AGTTTATTT---TAAG
	L	TGAACTGCACCCCAAAAAGT-TAGAC---AGAAAATT----TAACG
IS 911	R	TGAAGTGGCAC-ACTGAATTTGGCC---ACC-----TGAAC-AGAGGT
	L	TGAAGTGGTCA-ACAAAACTGGCC---ACCGAGT-----TAG-----AGTT
IS 1222	R	TGACCTGCCCCCAGGAT---TAGAT---ACAACGC-----TCACTT--AGTAATGTC
	L	TGACCTGCTCCCAGT-----TGATT---AGTACACCCCGATGTT----AGTAATGTC
IS 3411	R	TGAACCGCCCCGGGTTTCC-TGGAG---AGTGTTTTATC-TGTGA
	L	TGAACCGCCCCGGGAATCC-TGGAG---ACTAAACTTCC-TG-----AGA
IS RI	R	TGACGTGACCCCCCTTTC--TGATCC--AGCCATAAC---GAG-----AGTCC
	L	TGACGTGACCCCCCTGAAAC-TCCTCC--AGGAATAGC---TAG-----AGTCC
IS 1403	R	CGACGTCCCCCAATTT---TGAGT---AGCCCCTCAGTTTGG-----AGTCC
	L	CGACATCCCCCGCCC-----TGAGT---AGCGGCT-----TGGTTTAGAGTCC
IS 1404	R	CGACGTCCCCCAATTT---TGAGT---AGCGCCTCAGTTTGG-----AGTCC
	L	CGACGCCCCCGCCC-----TGAGT---AGCGGCC-----TGGTTTAGAGTCC
Consensus		TGA CTGCCCCCA TG AG T AG

Figure 3. Alignment of the terminal IR of IS1403 and IS1404 with other IRs from IS3 family. R and L define the right and left IR, respectively. Dashes represent gaps introduced for optimal alignment. Three sections can be identified in these IRs. The length of each section is indicated at top. Consensus nt are listed in the last line indicating bases conserved >60%. References of nt sequences for the listed ISs are in the paper of Prère et al. (1990).

IS 3	1	VSTSKKPRKQHSPEFRSEALKLAER. <b>IGV</b> TAAARELSLYESQLYNWRSK	
IS 476	1	.....MKKSRFSTEQ <b>I</b> GF <b>I</b> KQADAGMA <b>V</b> AELCRRHGFSPASFYQWRAK	
IS 1222	1	.....MKKRFSDQ <b>Q</b> I <b>S</b> ILREAEAGV <b>S</b> ARELCRKHAISDATFYTWRRK	
IS RI	1	.....MKRSRFTEE <b>Q</b> I <b>G</b> ILREQEAG <b>V</b> A <b>T</b> AECRRHGVSSATFYKWKAK	
IS 1403	1	.....VKKRFTEE <b>Q</b> I <b>G</b> FLREAE <b>G</b> GV <b>A</b> IKDLCRRHGFSEASYLWRSK	
IS 1404	1	.....VKKRFTDE <b>Q</b> V <b>I</b> GFLREAE <b>S</b> GV <b>A</b> IKDLCRRHGFSEASYLWRSK	
IS 3		QQNQQTSSEREL.EMSTEIARLKRQLAERDEELAILQKAATYFAKRLK	99
IS 476		<b>Y</b> GGMEADEAKRLKELEVQ <b>N</b> TRLK <b>K</b> LLAE <b>A</b> HLDIEALK <b>V</b> GF <b>G</b> VKR	88
IS 1222		<b>Y</b> GGMEVPEVKRLKSLE <b>E</b> NARL <b>K</b> LLAE <b>A</b> MLDKEAL <b>Q</b> VALGRKY	87
IS RI		<b>F</b> GGLDVSEARRLKALEDENARL <b>K</b> RMLAD <b>A</b> MLDNVAL <b>K</b> DLLG <b>K</b> KW	88
IS 1403		<b>F</b> GGMSVPDAKRLKDLE <b>A</b> ENARL <b>K</b> LLAE <b>Q</b> LFENDLI <b>K</b> DALR <b>K</b> KW	87
IS 1404		<b>F</b> GGMSVPDAKRLKDLE <b>S</b> ENARL <b>K</b> LLAE <b>Q</b> LFENDLI <b>K</b> DALR <b>K</b> KW	87

Figure 4. Amino acid sequences comparison between several ORFs 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 Prère et al. (1990). 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).

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<sub>6</sub>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; Prère 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 *Ava*II fragment, as a probe, dot-blotting was performed with genomic DNA from different pathovars of *X. campestris*. Thirty-two pathovars were hybridized to the *Ava*II 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*.

IS 3	132	<b>WAGD</b>	I	T	Y	L	R	T	D	E	G	W	L	Y	L	A	.	<b>VVIDLWSRAVIGWSMSPRMTAQLACDALQMALWRR</b>																																		
IS 150	225	<b>WVTD</b>	V	T	E	F	A	V	N	G	R	K	L	Y	L	S	.	<b>PVIDLFNNEVISYLSERPVMNMVENMLDQAFKKL</b>																																		
IS 861	118	<b>CYTD</b>	V	T	E	F	A	L	P	E	G	K	L	Y	L	S	.	<b>PVL</b> DGYNSEIIDFTLSRSPALKQVQTMLERAFPAA																																		
IS 911	142	<b>WCGD</b>	V	T	Y	I	W	T	G	K	R	W	A	Y	L	A	.	<b>VVLDL</b> FARKPVGWAMSFSPDSRLTMKALEMAWETR																																		
ISR1	117	<b>LVAD</b>	F	V	A	D	Q	L	T	D	G	R	R	F	R	I	L	T	V	I	D	N	C	T	R	E	C	L	A	L	V	A	D	T	S	L	S	G	A	R	V	V	R	E	L	D	A	V	I	R	Q	R
IS 407	17	<b>WSID</b>	F	V	M	D	A	L	S	N	G	R	R	V	K	C	L	T	V	V	D	D	F	T	K	E	A	V	D	I	V	V	D	H	G	I	S	G	L	Y	V	A	R	A	L	D	R	A	A	R	F	R
IS 476	108	<b>WSMD</b>	F	V	F	D	A	L	A	N	A	R	R	I	K	C	L	T	V	V	D	D	F	T	R	E	S	V	D	I	A	V	D	H	G	I	S	G	A	Y	V	V	R	L	L	D	Q	A	A	C	F	R
IS 1222	110	<b>WSDD</b>	F	V	M	D	A	L	A	T	G	R	R	I	K	C	L	T	C	V	D	D	F	T	K	E	C	L	T	V	T	V	A	F	G	I	S	G	V	Q	V	T	R	I	L	D	S	I	A	L	F	R
IS 1403	111	<b>WSMD</b>	F	V	F	D	R	S	A	E	G	R	V	I	K	C	L	V	I	V	D	D	A	T	H	E	A	V	A	I	E	V	E	R	A	I	S	G	H	G	V	A	R	V	L	D	R	L	A	H	S	R
IS 1404	91	<b>WSMD</b>	F	V	F	D	R	T	A	E	G	R	V	I	K	C	L	V	I	V	D	D	A	T	H	E	A	V	A	I	E	V	E	R	A	I	S	G	H	G	V	T	R	V	L	D	R	L	A	H	S	R

  

IS 3		K	R	P	R	N	V	I	V	H	T	D	R	G	G	Q	Y	C	S	A	D	Y	Q	A	Q	L	K	R	H	N	L	R	G	S	M	S	A	K	G	C	C	Y	D	N	A	C	V	E	S	F	F	H	S	L	K	238
IS 150		N	P	H	E	H	P	V	L	H	S	D	Q	G	W	Q	Y	R	M	R	R	Y	Q	N	I	L	K	E	H	G	I	K	Q	S	M	S	A	K	G	N	C	L	D	N	A	V	V	E	C	F	F	G	T	L	K	331
IS 861		S	Y	S	E	T	.	I	L	H	S	D	Q	G	W	Q	Y	Q	H	K	S	Y	H	Q	F	L	E	D	K	G	I	R	P	S	M	S	R	K	G	N	S	P	D	N	G	M	M	E	S	F	F	G	I	L	K	233
IS 911		G	K	P	G	G	V	M	F	H	S	D	Q	G	S	H	Y	T	S	R	Q	F	R	Q	L	L	W	R	Y	Q	I	R	Q	S	M	S	R	R	G	N	C	W	D	N	S	P	M	E	R	F	F	R	S	L	K	248
ISR1		G	R	P	D	.	.	T	I	V	S	D	N	G	T	E	Y	T	S	N	A	V	L	A	W	A	T	T	P	A	F	G	W	H	Y	I	A	P	G	K	P	Q	N	G	F	N	E	S	F	N	G	R	L	R	222	
IS 407		G	Y	P	K	.	.	A	V	R	T	D	Q	G	P	E	F	T	S	R	A	L	D	Q	W	A	N	G	V	T	L	K	L	I	Q	A	G	K	P	T	Q	N	A	I	E	S	F	N	G	K	F	R	122			
IS 476		G	Y	P	R	.	.	A	V	R	T	D	Q	G	P	E	F	T	S	R	A	F	I	A	W	T	Q	H	G	I	E	H	I	L	I	E	P	G	A	P	T	Q	N	A	I	E	S	F	N	G	K	F	R	213		
IS 1222		G	Y	P	A	.	.	T	I	R	T	D	Q	G	P	E	F	T	C	R	A	L	D	Q	W	A	F	E	H	G	V	E	L	R	L	I	Q	P	G	K	P	T	Q	N	G	F	I	E	S	F	N	G	R	F	R	215
IS 1403		G	L	P	Q	.	.	V	I	R	T	D	Q	G	K	E	F	C	G	K	A	M	V	A	W	A	H	A	R	G	V	Q	L	R	L	I	Q	P	G	K	P	N	Q	N	A	Y	V	E	S	F	N	G	R	L	R	216
IS 1404		G	L	P	K	.	.	V	I	R	T	D	Q	G	K	E	F	C	G	K	A	M	V	A	W	A	H	A	R	N	V	Q	L	R	L	I	Q	P	G	K	P	N	Q	N	A	Y	V	E	S	F	N	G	R	L	R	196

**Figure 5.** Amino acid sequences comparison between several ORFB 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 bacteria<sup>a</sup>.**Positive species and strain<sup>b</sup>***Xanthomonas campestris* pathovar

Armoraciae XC132; betlicola XB104; campestris XCC1-1; carotae XC139; cassiae XC112; corylina ATCC1931; eucalypti XE104; gardneri XG101; holcicola XH3; incanae XI4; juglandis XJ123; khayae XK101; lespedezae XL1; manihotis XM125; martyniicola XM107; melhusii XM123; nakataecorchori XN104; papavericola XP5; patelii XP186; phlei ATCC33805; plantaginis XP195; poae ATCC33804; poinsettiicola XP137; pruni ATCC19316; ricini ATCC19317; secalis XT129; spermacoces ATCC17998; tamarindi XT105; taraxaci XT11; vasculorum ATCC35938.

*X. oryzae* pathovar

Oryzae ATCC35933; oryzicola XO111

**Negative species and strain***X. campestris* pathovar

Badrii XB103; biophyti XB117; celebensis XC145; citri XC11-1; physalidicola XP172; physalidis 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

<sup>a</sup>Dot blots of total DNA from the strains listed above were hybridized with the DIG-labeled 0.5 kb *Ava*II fragment of IS1403. The hybridization method is described in Materials and Methods.

<sup>b</sup>The 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 *avrBs1* 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**

- Berthier, Y., D. Thierry, M. Lemattre, and J.-L. Guesdon. 1994. Isolation of an insertion sequence (IS1051) from *Xanthomonas campestris* pv. *dieffenbachiae* with potential use for strain identification and characterization. *Appl. Environ. Microbiol.* **60**: 377–384.
- Bradbury, J.F. 1984. Genus *Xanthomonas*. In N.R. Krieg and J.G. Holt (eds.), *Bergey's Manual of Systematic Bacteriology*, vol.1. The Williams & Wilkins Co., Baltimore, pp. 199–209.
- Brennan, R.G. and B.W. Matthews. 1989. The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264**: 1903–1906.
- Brierly, I. and M.E.G. Bournsnel. 1987. An efficient ribosomal frame-shifting signal in the polymerase encoding region of the coronavirus IBV. *EMBO J.* **12**: 3779–3785.
- Brierly, I., P. Digard, and S.C. Inglis. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement of an RNA pseudoknot. *Cell* **57**: 537–547.
- Calos, M.P. and J.H. Miller. 1980. Transposable elements. *Cell* **20**: 579–595.
- Chandler, M. and O. Fayet. 1993. Translational frameshifting in the control of transposition in bacteria. *Mol. Microbiol.* **7**: 497–503.
- Chow, L.T. and T.R. Broker. 1978. Adjacent insertion sequence IS2 and IS5 in bacteriophage Mu mutants and IS5 in lambda *darg* bacteriophage. *J. Bacteriol.* **133**: 1427–1436.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic*

- Acids Res. **12**: 387–395.
- Fayet, O., P. Ramond, P. Polard, M.F. Prère, and M. Chandler. 1990. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol. Microbiol.* **4**: 1771–1777.
- Galas, D.J. and M. Chandler. 1989. Bacterial insertion sequences. In D.E. Berg and M.M. Howe (eds.), *Mobile DNA*. American Society for Microbiology, Washington, D.C., pp. 109–162.
- Green, L., R.D. Miller, D.E. Dykhuizen, and D.L. Hartl. 1984. Distribution of DNA insertion element IS5 in natural isolates of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **81**: 4500–4504.
- Hawley, D.K. and W.R. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter sequences. *Nucleic Acids Res.* **8**: 2237–2255.
- Iida, S., J. Meyer, and W. Arber. 1983. Procaryotic IS element. In J.A. Shapiro (ed.), *Mobile Genetic Elements*. Academic Press, Inc. New York, pp. 159–221.
- Ishiguro, N. and G. Sato. 1988. Nucleotide sequence of insertion sequence IS3411, which flanks the citrate utilization determinant of transposon Tn3411. *J. Bacteriol.* **170**: 1902–1906.
- Kearney, B. and B.J. Staskawicz. 1990. Characterization of IS476 and its role in bacterial spot disease of tomato and pepper. *J. Bacteriol.* **172**: 143–148.
- Khan, E., J.P.G. Mack, R.A. Katz, J. Kulkowsky, and A.M. Skalka. 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* **19**: 851–860.
- 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.
- Lee, Y.-A., M. Henderson, N.J. Panopoulos, and M.N. Schroth. 1994. Molecular cloning, chromosomal mapping and sequence analysis of copper-resistance genes from *Xanthomonas campestris* pv. *juglandis*: homology with small blue copper proteins and multicopper oxidases. *J. Bacteriol.* **176**: 173–188.
- Lin, N.T. and Y.H. Tseng. 1997. Sequence and copy number of the *Xanthomonas campestris* pv. *campestris* gene encoding 16S rRNA. *Biochem. Biophys. Res. Commun.* **235**: 276–280.
- Miller, J.H. 1972. *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- 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 IS911. *J. Mol. Biol.* **222**: 465–477.
- Prère, M.F., M. Chandler, and O. Fayet. 1990. Transposition in *Shigella dysenteriae*: isolation and analysis of IS911, a new member of the IS3 group of insertion sequences. *J. Bacteriol.* **172**: 4090–4099.
- Priefer, U.B., J. Kalinowski, B. Rüger, W. Heumann, and A. Pühler. 1989. ISRI, a transposable DNA sequence resident in *Rhizobium* Class IV strains, shows structural characteristics of classical insertion elements. *Plasmid* **21**: 120–128.
- Rubens, C.E., L.M. Heggen, and J.M. Kuypers. 1989. IS861, a group B Streptococcal insertion sequence related to IS150 and IS3 of *Escherichia coli*. *J. Bacteriol.* **171**: 5531–5535.
- Saedler, H., G. Cornelis, J. Cullum, B. Schumacher, and H. Sommer. 1980. IS1 mediated DNA rearrangements. *Cold Spring Harbor Symp. Quant. Biol.* **45**: 93–98.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: a Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.
- Sawyer, S., D. Dykhuizen, R. DuBose, L. Green, T. Mutangadura-Mhlanga, D. Wolczyk, and D. Hartl. 1987. Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. *Genetics* **115**: 51–63.
- Schwartz, E., C. Herbeberger, and B. Rak. 1988. Second-element turn-on of gene expression in an IS1 insertion mutant. *Mol. Gen. Genet.* **211**: 282–289.
- Sekine, Y., N. Eisaki, and E. Ohtsubo. 1994. Translational control in production of transposase and in transposition of insertion sequence IS3. *J. Mol. Biol.* **235**: 1406–1420.
- Shine, J. and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**: 1341–1346.
- Steibl, H.-D. and F.-M. Lewecke. 1995. IS1222: analysis and distribution of a new insertion sequence in *Enterobacter agglomerans* 339. *Gene* **156**: 37–42.
- Timmerman, K. and C.-P.D. Tu. 1985. Complete sequence of IS3. *Nucleic Acids Res.* **13**: 2127–2139.
- Tu, J., H.-R. Wang, S.-F. Chang, Y.-C. Charng, R. Lurz, B. Dobrinski, and W.-C. Wu. 1989. Transposable elements of *Xanthomonas campestris* pv. *citri* originating from indigenous plasmids. *Mol. Gen. Genet.* **217**: 505–510.
- Vaughan, E.-E. and W.M. de Vos. 1995. Identification and characterization of the insertion element IS1070 from *Leuconostoc lactis* NZ6009. *Gene* **155**: 95–100.
- Vögele, K., E. Schwartz, C. Welz, E. Schiltz, and B. Rak. 1991. High-level ribosomal frameshifting directs the synthesis of IS150 gene products. *Nucleic Acids Res.* **19**: 4377–4385.
- Weng, S.F., S.M. Tang, Y.H. Tseng, and J.W. Lin. 1997. IS element ISXC6 of *Xanthomonas campestris* pv. *campestris*. *DNA Sequence* **7**: 221–223.
- Wilson, K. 1987. Preparation of genomic DNA from bacteria. In F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (eds.), *Current Protocols in Molecular Biology*, vol. 1. Wiley Interscience, Cambridge, Massachusetts, pp. 241–245.
- Wood, M.S., A. Byrne, and T.G. Lessie. 1991. IS406 and IS407, two gene-activating insertion sequences from *Pseudomonas cepacia*. *Gene* **105**: 101–105.