JOURNAL OF BACTERIOLOGY, Feb. 1999, p. 1021–1024 Vol. 181, No. 3
0021-9193/99/$04.00+0


# NOTES

## Characterization of IS*1547*, a New Member of the IS*900* Family in the *Mycobacterium tuberculosis* Complex, and Its Association with IS*6110*

Z. FANG,[1†] C. DOIG,[2] N. MORRISON,[1] B. WATT,[2] AND K. J. FORBES[1*]

*Medical Microbiology, Aberdeen University, Foresterhill, Aberdeen AB25 2ZD,[1] and Scottish Mycobacteria Reference Laboratory, The City Hospital, Edinburgh EH10 5SB,[2] United Kingdom*

Received 23 February 1998/Accepted 23 November 1998

**Unlike classically defined insertion sequence (IS) elements, which are delimited by their inverted terminal repeats, some IS elements do not have inverted terminal repeats. Among this group of atypical IS elements, IS*116*, IS*900*, IS*901*, and IS*1110* have been proposed as members of the IS*900* family of elements, not only because they do not have inverted terminal repeats but also because they share other features such as homologous transposases and particular insertion sites. In this study, we report a newly identified IS sequence, IS*1547*, which was first identified in a clinical isolate of *Mycobacterium tuberculosis*. Its structure, insertion site, and putative transposase all conform with the conventions of the IS*900* family, suggesting that it is a new member of this family. IS*1547* was detected only in isolates of the *M. tuberculosis* complex, where it had highly polymorphic restriction fragment length polymorphism patterns, suggesting that it may be a useful genetic marker for identifying isolates of the *M. tuberculosis* complex and for distinguishing different strains of *M. tuberculosis*. *ipl* is a preferential locus for IS*6110* insertion where there are eight known different insertion sites for IS*6110*. Surprisingly, the DNA sequence of *ipl* is now known to be a part of IS*1547*, meaning that IS*1547* is a preferential site for IS*6110* insertion.**

Bacterial insertion sequences (ISs) are genetic entities which are able to translocate to new genetic locations either within a replicon or between different replicons in the host cell. Typically, IS elements are 0.7 to 2.5 kb in length and end in perfect or nearly perfect inverted terminal repeats, which are proposed to play a role in transposition and in the selection of insertion targets. ISs encode only proteins related to their transposition activity, such as transposases. With few exceptions, IS elements generate a duplication of the DNA sequence at the insertion site on transposition; thus, after insertion, the duplication borders the IS element as a direct repeat (for a review, see reference 7).

Unlike classically defined IS elements, which are delimited by their inverted terminal repeats, some IS elements, for example, IS*900* identified in *Mycobacterium paratuberculosis* (8) and IS*901* and IS*1110* from *Mycobacterium avium* (10, 12), do not have these repeats. Other IS elements without inverted terminal repeats include IS*1000* from *Thermus thermophilus* (2), IS*116* from *Streptomyces clavuligerus* (14), IS*117* from *Streptomyces coelicolor* (9), and HBSI from *Bradyrhizobium japonicum* (11). Among these atypical IS elements, IS*116*, IS*900*, IS*901*, and IS*1110* have been proposed as members of a group of closely related IS elements, designated the IS*900* family, not only because they do not have inverted terminal repeats but because they share other features as well, such as homologous transposases and particular insertion sites (10). Since inverted terminal repeats are believed to be important in the selection of target sites and in the mechanics of transposition of IS elements (7), greater understanding of this group of atypical IS elements is important for the elucidation of transposition and the evolution of insertion sequences.

*ipl* is a hot spot for IS*6110* insertion in the genome of *Mycobacterium tuberculosis* (EMBL, GenBank, and DDBJ database accession no. X95799 [5]). At this locus we have extended the DNA sequence of *ipl* on each side to 2732 nucleotides (nt) in clinical isolate *M. tuberculosis* 151, an isolate which does not harbor an IS*6110* copy in *ipl*. Analysis of this DNA sequence revealed several open reading frames (ORFs), and translation products of these were used in searches of protein databases with BLAST (1). One of the translated amino acid sequences, ORF1, showed significant homology to several peptide sequences, most of which were transposases, such as those from IS*116* (BLAST score of more than $1.1 \times 10^{-16}$) from *S. clavuligerus*, IS*1110* ($4.2 \times 10^{-10}$) from *M. avium*, IS*901* ($2.3 \times 10^{-10}$) from *M. avium*, IS*900* ($1.6 \times 10^{-10}$) from *M. paratuberculosis*, and IS*110* ($5.6 \times 10^{-9}$) from *S. coelicolor*. Consequently, along with having other features (see below), the translation product of ORF1 is proposed to be the transposase of a new IS element, designated IS*1547*.

Two different insertion sites of IS*1547* from two clinical isolates were identified and sequenced. Sequence comparison of the DNAs from these two insertion sites, together with ORF1 peptide sequence common to both of these two sites (see below), revealed that IS*1547* is 1351 nt in length without terminal inverted repeats but with target direct repeats (CCTT) in Y13470 and imperfect direct repeats (CCTT/CCTC) in Y16254. These two insertion sites were also found in

* Corresponding author. Mailing address: Medical Microbiology, Aberdeen University, Foresterhill, Aberdeen AB25 2ZD, United Kingdom. Phone: 44 1224 663123, ext. 54953. Fax: 44 1224 685604. E-mail: mmb001@abdn.ac.uk.

Present address: Department of Biomedical Sciences, University of Bradford, Bradford, West Yorkshire, BD7, 1DP, United Kingdom.

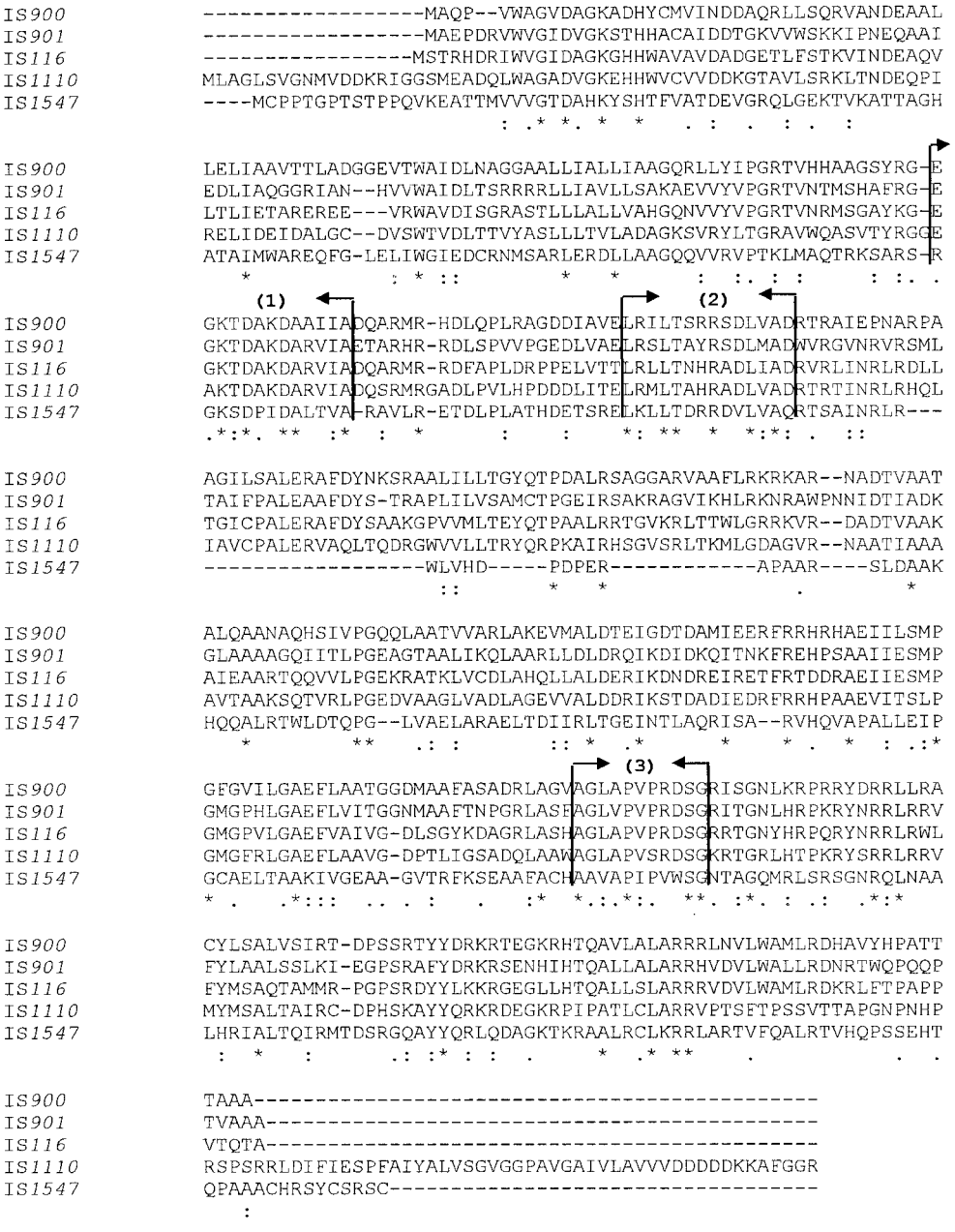


FIG. 1. Amino acid sequence alignment of the translation product of IS*1547* ORF1 and the transposases of the IS elements of the IS*900* family, derived via program Clustal W (version 1.7). The * symbols in the conservation line indicate the positions where there are identical residues across all the amino acid sequences; colons indicate strongly conserved residues, and period indicate weakly conserved residues (16). The three regions indicated by numbers and delimited with bent arrows are conserved regions in the IS*900* family, one of which (region 1) was found to contain the motif in reverse transcriptase and the other of which (region 3) has the motif involving DNA inversion (Kunze et al. [12] and Hernandez Perez et al. [10]). See the text for more details.

the genome of *M. tuberculosis* H37Rv, which is being sequenced (14a).

Two large ORFs, ORF1 and ORF2, were identified in the DNA sequence of IS*1547*. ORF1 has no stop codon within the DNA sequence of IS*1547*. The predicted translation product of ORF1 of IS*1547* itself was considered the peptide sequence which is common to the ORF1s of IS*1547* at the two insertion sites. This gave a 380-amino-acid peptide of about 41 kDa in mass. In accession no. Y13470, ORF1 was 1182 nt long with a stop codon outside the IS*1547* sequence, giving three extra amino acid residues at the C terminus of the peptide; ORF1 of IS*1547* in Y16254 was even longer. The use of flanking DNA sequences to provide the stop codon for an ORF has also been observed in IS*1110* of *M. avium* (10) and in IS*870* of *Agrobacterium vitis*, which uses its specific insertion site (CTAG) to generate the stop codon (6). The use of external stop codons to terminate ORFs of IS elements is thus not uncommon and is not limited to the IS*900* family of elements; however, the reason for this strategy is not clear.

The transposases of the IS*900* family of elements have been reported to have two conserved peptide regions, one of which is a motif found in reverse transcriptases (region 1 in Fig. 1)

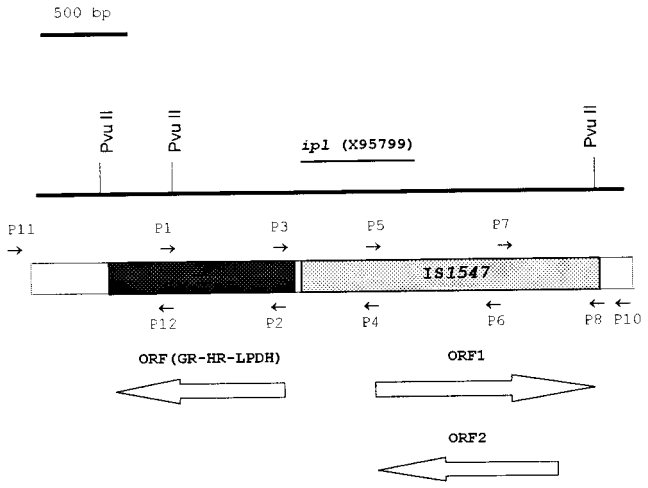


FIG. 2. Schematic illustration of the IS*1547* insertion site in accession no. Y13470, restriction map of the DNA sequence (EMBL, GenBank, and DDBJ accession no. Y13470), and location of *ipl* (EMBL, GenBank, and DDBJ accession no. X95799). The smaller arrows represent primers used in this study, while the larger, open arrows indicate locations of the ORFs. There are three restriction sites for *Pvu*II within this DNA fragment, as presented in the line, and there are 17 sites for *Alu*I, but there are no restriction sites for *Asn*I, *Dra*I, and *Hin*dIII. GR, glutathione reductase; HR, mercuric reductase; LPDH, dihydrolipoamide dehydrogenase.

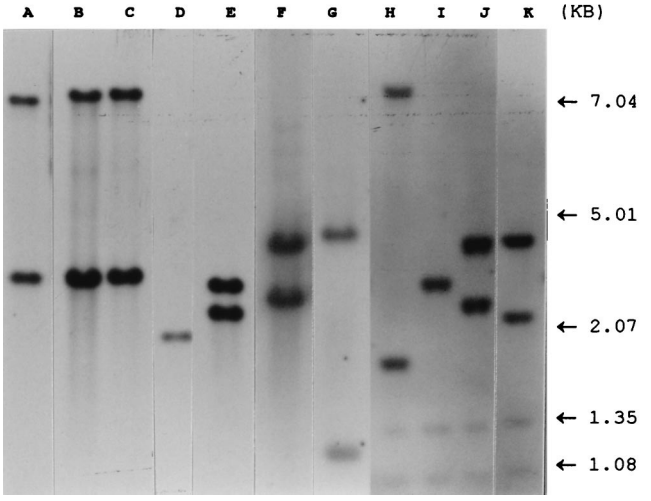


FIG. 3. Autoradiograph of a Southern blot of *Pvu*II-digested chromosomal DNA of mycobacterial isolates probed with a digoxigenin-labelled IS*1547* probe. Lane A, *M. bovis* BCG (Glaxo); lane B, clinical isolate 8189/96 of *M. bovis* BCG; lane C, clinical isolate B1 of *M. bovis*; lane D, clinical isolate L2523 of *M. africanum*; lane E, clinical isolate 11804/93 of *M. africanum*; lane F, type strain H37Ra of *M. tuberculosis*; lane G, IS*6110* RFLP reference strain mtb14323 of *M. tuberculosis*; lanes H to K, clinical isolates 9407, 9212, 9308, and 9101 of *M. tuberculosis*. Faint bands in lanes H to K are internal size standards.

(12), the other of which shows homology to a motif involved in inverting DNA (region 3 in Fig. 1) (10). Comparison of the sequences of the IS*1547* ORF1 peptide and the transposases of the other members of the IS*900* family revealed the presence of both of these conserved peptide regions in the IS*1547* ORF1 peptide. In addition, the comparison disclosed a third conserved region (region 2 in Fig. 1) with the consensus sequence L--LT--R--L-A. This consensus sequence did not significantly match sequences in the motif database PROSITE (released in November 1995; Amos Bairoch, Medical Biochemistry Department, University of Geneva, Geneva, Switzerland). These three motifs further support the function of ORF1 as the transposase of IS*1547*.

A second large ORF, ORF2, of IS*1547* is on the DNA strand complementary to and overlapped by ORF1, and these two ORFs share their third codon positions (Fig. 2). It is predicted that ORF2 encodes a peptide of 296 amino acids (Y13470:e339203) with a molecular mass of about 32 kDa. A second ORF is also found in IS*900* and IS*116*, both of which are again on the strand complementary to ORF1, while this ORF is not found in IS*1110* (8, 10). The IS*1547* ORF2 translation product showed only 45% similarity and 23% identity to IS*900* ORF2, suggesting little similarity between these peptides. In addition, a protein database search of the IS*1547* ORF2 translation product did not recover any significantly similar peptides.

IS*1547* shares several features with members of the IS*900* family of elements; one of them is that they tend to insert into the promoter regions of genes (10). In the sequences flanking the two IS*1547* copies, ORFs were identified on the complementary strand at the 5′ end of the IS*1547* copies, with their direction of expression opposite to that of the putative transposase of IS*1547* (Fig. 2). The ORF in accession no. Y16254 encoded a peptide of 172 amino acids (Y16254:e1240541) which had no significant matches in protein database searches. However, the ORF in Y13470 encoded a 499-amino-acid peptide (Y13470:e3215020) which showed strong homology (BLAST scores of more than $10^{-28}$) to enzymes of the pyridine nucleotide-disulfide oxidoreductase class I family, including seven mercuric reductases, four glutathione reductases, and four dihydrolipoamide dehydrogenase, all from different species.

To examine the distribution of IS*1547* in mycobacteria, a digoxigenin-labelled IS*1547* probe was applied to Southern blots of *Pvu*II-digested genomic DNAs of the following isolates: 61 isolates of *M. tuberculosis*, including strain H37Ra and IS*6110* restriction fragment length polymorphism (RFLP) reference strain Mt14323; 3 isolates of *M. bovis*; 2 vaccine strains of *M. bovis* BCG (Glaxo and Copenhagen) and 3 clinical isolates of *M. bovis* BCG; 2 isolates of *Mycobacterium africanum*; 3 isolates of *M. avium*; and 1 isolate each of *M. paratuberculosis*, *Mycobacterium malmoense*, *Mycobacterium fortuitum*, *Mycobacterium marinum*, and *Mycobacterium kansasii*. The results suggest the following. (i) Hybridizing DNA fragments were found in all of the isolates of *M. tuberculosis*, *M. bovis*, *M. bovis* BCG, and *M. africanum* (Fig. 3) but not in any of the *M. avium*, *M. paratuberculosis*, *M. malmoense*, *M. fortuitum*, *M. marinum*, or *M. kansasii* isolates. IS*1547* may therefore be useful as a genetic marker in distinguishing the *M. tuberculosis* complex from other mycobacterial species. (ii) Within the isolates of the *M. tuberculosis* complex, many different IS*1547* RFLP patterns were observed; for example, three clinical isolates of *M. bovis* BCG had two IS*1547* copies with the same banding pattern (Fig. 3, lane B), while the vaccine strains of *M. bovis* BCG (Glaxo and Copenhagen) had a slightly different banding pattern (Fig. 3, lane A versus lane B). Unlike in *M. bovis*, the IS*1547* banding patterns in *M. africanum* and *M. tuberculosis* exhibited a totally different picture.

*ipl* was identified as a preferential locus in the genome of *M. tuberculosis* for IS*6110* insertions (5). In addition to the six different insertion sites (*ipl*-1::IS*6110* to *ipl-6*::IS*6110*) described previously, two more have since been found: *ipl-7*::IS*6110* (Y14613) and *ipl-8*::IS*6110* (Y14614). It is now apparent that the original *ipl* locus is in fact in the DNA sequence of IS*1547* and is located at nt 1718 to 2370 of sequence Y13470 (Fig. 3) and that there are two such sites in the genomes of

many *M. tuberculosis* isolates. That is to say, IS*1547* is a preferential site for IS*6110* insertion. Interactions between IS elements have also been observed in other bacteria, although they have been poorly studied. For instance, IS*53* from a plasmid of *Pseudomonas syringae* subsp. *savastanoi* was found to insert into IS*51* (15) and the target of IS*Rm3* transposition in *Rhizobium meliloti* is the insertion sequence IS*Rm5* (13). Recently, the genome sequence of *Escherichia coli* K-12 (3) revealed that two IS*911*-related sequences (IS*911*A and IS*911*B) had been interrupted by IS*30* and IS*600*. Further investigations are being carried out to clarify the basis of the interaction between IS*1547* and IS*6110* and its implications for transposition and strain similarity assessments based on these elements.

**Nucleotide and peptide sequence accession numbers.** DNA fragments sequenced in this study have been deposited in the EMBL, GenBank, and DDBJ data banks under accession no. Y13470, Y14613, Y14614, and Y16254. Predicted peptide sequences have been deposited in TREMBL data bank under accession no. Y13470:e3215020 and Y13470:e339203.

We thank A. Rayner and G. Harris at the Scottish Mycobacteria Reference Laboratory for bacteriological assistance, P. Carter, and K. Reay for DNA sequencing and synthesis of the oligonucleotide primers. DNA sequence analysis benefited from SEQNET, the SERC facility (Daresbury, United Kingdom).

This study was financially supported by the Department of Health, the Scottish Office; Chest, Heart and Stroke Scotland; and a Milner Scholarship from the University of Aberdeen.

## REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Ashby, M. K., and P. L. Bergquist.** 1990. Cloning and sequence of IS1000, a putative insertion sequence from *Thermus thermophilus* HB8. Plasmid **24:**1–11.
3. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of Escherichia coli K-12. Science **277:**1453–1474.
4. **Butler, W. R., W. H. Haas, and J. T. Crawford.** 1996. Automated DNA fingerprinting analysis of *Mycobacterium tuberculosis* using fluorescent detection of PCR products. J. Clin. Microbiol. **34:**1801–1803.
5. **Fang, Z., and K. J. Forbes.** 1997. A *Mycobacterium tuberculosis* IS*6110* preferential locus (*ipl*) for insertion into the genome. J. Clin. Microbiol. **35:**479–481.
6. **Fournier, P., F. Paulus, and L. Otten.** 1993. IS*870* requires a 5′-CTAG-3′ target sequence to generate the stop codon for its large ORF1. J. Bacteriol. **175:**3151–3160.
7. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 939–958. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
8. **Green, E. P., M. L. Tizard, M. T. Moss, J. Thompson, D. J. Winterbourne, J. J. McFadden, and J. Hermon-Taylor.** 1989. Sequence and characteristics of IS900, an insertion element identified in a human Crohn's disease isolate of Mycobacterium paratuberculosis. Nucleic Acids Res. **17:**9063–9073.
9. **Henderson, D. J., D. J. Lydiate, and D. A. Hopwood.** 1989. Structural and functional analysis of the mini-circle, a transposable element of Streptomyces coelicolor A3(2). Mol. Microbiol. **3:**1307–1318.
10. **Hernandez Perez, M., N. G. Fomukong, T. Hellyer, I. N. Brown, and J. W. Dale.** 1994. Characterization of IS1110, a highly mobile genetic element from Mycobacterium avium. Mol. Microbiol. **12:**717–724.
11. **Judd, A. K., and M. J. Sadowsky.** 1993. The *Bradyrhizobium japonicum* serocluster 123 hyperreiterated DNA region, HRS1, has DNA and amino acid sequence homology to IS*1380*, an insertion sequence from *Acetobacter pasteurianus*. Appl. Environ. Microbiol. **59:**1656–1661.
12. **Kunze, Z. M., S. Wall, R. Appelberg, M. T. Silva, F. Portaels, and J. J. McFadden.** 1991. IS901, a new member of a widespread class of atypical insertion sequences, is associated with pathogenicity in Mycobacterium avium. Mol. Microbiol. **5:**2265–2272.
13. **Laberge, S., A. T. Middleton, and R. Wheatcroft.** 1995. Characterization, nucleotide sequence, and conserved genomic locations of insertion sequence ISRm5 in Rhizobium meliloti. Comp. Appl. Biosci. **3:**239–241.
14. **Leskiw, B. K., M. Mevarech, L. S. Barritt, S. E. Jensen, D. J. Henderson, D. A. Hopwood, C. J. Bruton, and K. F. Chater.** 1990. Discovery of an insertion sequence, IS116, from Streptomyces clavuligerus and its relatedness to other transposable elements from actinomycetes. J. Gen. Microbiol. **136:**1251–1258.

14a. Sanger Centre Website. 1997, copyright date. [Online.] Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom. http://www.sanger.ac.uk. [December 1997, last date accessed.]

15. **Soby, S., B. Kirkpatrick, and T. Kosuge.** 1993. Characterization of an insertion sequence (IS53) located within IS51 on the iaa-containing plasmid of Pseudomonas syringae pv. savastanoi. Plasmid **29:**135–141.
16. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence-weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.