Proc. of Special Lectures on Molecular Biological Approaches
to Plant Disease pp.1-19(1995)

Molecular characterization of a repetitive element of *Xanthomonas oryzae* pv. oryzae

Choong Hyo Yun

National Agricultural Science and Technology Institute Suwon, Korea

Abstract

The plasmid pJEL 101 contains a highly repetitive element from the genome of *Xanthomonas oryzae* pv. *oryzae* that has properties of an insertional element. The insertional nature of the element, hereto referred to as IS*203*, was confirmed by molecular analyses of the element and three related elements that were isolated from X. oryzae. The related sequences were isolated on the basis of transposition to the transposon-trapping vector pL3SAC and hybridization with pJEL101. The trapped elements (IS*203a*, IS*203b*, and IS*203c*) were each composed of 1,055 base pairs with 25 base pair terminal inverted repeats. The elements caused a three base pair target site duplication at the site of insertion in the *sacRB* gene. The sequence of pJEL 101 has 96% base pair identity with IS*203a* and 99% identity with IS*203b* and IS*203c* but lacks three nucleotides of the consensus left terminal repeat. IS*203b* has the same DNA sequences as IS*203c* but is inserted into the *sacRB* gene in the opposite orientation. The longest open reading frame of IS*203a* could code for a protein of 318 amino acids and molecular weight of 37, 151. A search of the Genbank database revealed that IS*203* has 51% identity with 909 nucleotides of IS*4551* from *Escherichia coli*. The predicted protein of ORF1 has 40% and 30% amino acid identity to the ORF1 of Tn*4551* and the transposase of IS*30*, respectively.

INTRODUCTION

Insertion elements and transposons are prominent features of bacterial genomes and generally are considered to play important roles in the adaptation of bacteria and, in particular, phytopathogenic bacteria, to the environment (4, 7, 11, 12, 19, 23, 24, 40). In *Pseudomonas cepacia*, IS elements, which often posses promoter sequences, have been shown to insert upstream of a weakly expressed gene and increased expression more than 30-fold (33). Inactivation of a deleterious gene by insertion also has been demonstrated to be a potential mode of adaptation of a

plant pathogen. Inactivation of the *avrBs1* gene, a gene in *Xanthomonas campestris* pv. *vesicatoria* that elicits a defense response in certain cultivars of pepper plants upon infection by the bacteria, can occur by insertion of IS*476* (15). The strain, which is nonpathogenic, subsequently attains pathogenicity upon inactivation of *avrBs1*.

A highly repetitive DNA sequence, present in approximately 81 copies per genome, was cloned from *Xanthomonas oryze* pv. *oryzae*, a bacterial pathogen of rice. The size and polymorphic nature of the sequence are suggestive of an insertion element (17). X. *oryzae* pv. *oryzae* is a highly adaptive pathogen and is under selective pressure due to the continuous cultivation of new rice cultivars with genes for resistance to bacterial infection. We are interested in determining if the repetitive DNA plays a role in race evolution of *X. oryaze* pv. *oryzae*. Toward that goal, we have characterized the element and report here the sequence analysis of four related elements and the transpositional activity of those elements.

MATERIALS AND METHODS

Plasmids, strains, and media. The plasmids used in this study are listed in Table 1. Strains of *X. oryzae* pv. *oryzae* were maintained on peptone-sucrose agar (41) at 28℃ *Escherichia coli* was maintained on Luria-Bertani (LB) agar (22) medium with appropriate antibiotics at 37℃. Antibiotic concentrations were carbenicillin (Cb) at 100 μg/ml and tetracycline (Tc) at 10 μg/ml. All bacterial isolates were stored at −80℃ in 30% glycerol (1).

Table 1. Plasmids and strains used in this study

plasmids or strain	Relevant characteristics	source on reference
pLAFR3	IncP, Tc^r, Mob$^+$, cos	Staskawicz et al. 1987
pL3SAC	IncP, Tc^r, Mob$^+$, cos, sacRB gene	Staskawicz et al. 1990
pL3SAC (203)	derivative of pL3SAC, contains inserted DNA of Xoo into sacRB gene, hybridizes with repetitive sequence of pBS101	This work
pL3SAC (203a)	derivative of pL3SAC contains inserted DNA of Xoo into SacRB gene, hybridizes with repetitive sequence of pBS101	R. Nelson, IRRI

plasmids or strain	Relevant characteristics	source on reference
pL3SAC (203b)	derivative of pL3SAC contains inserted DNA of Xoo into SacRB gene, hybridizes with repetitive sequence of pBS101	R. Nelson IRRI
pBluescript KS+/−	ColEl replicon, Cb^r	Stratagene, LaJolla, CA
pBS101 (+)	2.4-kb EcoRI-HindIII fragment in pBluescript KS+, contains repetitive sequence from Xoo	Leach et al. 1990
pBS101 (−)	2.4-kb EcoRI-HindIII fragment in pBluescript KS−, contains repetitive sequence (ISXoo1)	This work
pBS203(+)	2.8-kb BamHI-HindIII fragment of SacRB gene in pBluescript KS(+), contains insertion sequence (IS203)	This work
pBS203 (−)	2.8-kb BamHi-HindIII fragment of SacRB gene in pBluescript KS(−), contains insertion sequence (IS203)	This work
pBS203a (+)	2.8-kb BamHI-HindIII fragment of sacRBgene in pBluescript KS(+), contains insertion sequence (IS203a)	This work
pBS203a (−)	2.8-kb BamHI-HindIII fragment of sacRBgene in pBluescript KS(−), contains insertion sequence (IS203a)	This work
pBS203b (−)	2.8-kb BamHI-HindIII fragment of sacRBgene in pBluescript KS(−) contains insertion sequence (IS203b)	This work
E. coli strains		
MV 1190	Tc^r, F′	BRL
DH5α (203)	contains pL3SAC(203)	This work
DH5α (203a)	contains pL3SAC(203a)	R. Nelson, IRRI
DH5α (203b)	contain pL3SAC (203b)	R. Nelson, IRRI
X. oryzae pv. oryzae strains		
PX086	race 2 of *X. oryzae pv. oryzae*	
PX0112	race 6 of *X. oryzae pv. oryzae*	

Plasmid isolation. Plasmid DNA was isolated by the technique of Birnboim and Doly (2). Large-scale preparations were further purified by CsCl-ethidium bromide gradient centrifugation (31).

Southern blot hybridization. The transfer of DNA from agarose gels onto nylon membranses and washes were done as described by the manufacturer of the GeneScreen Plus membrane (Du Pont Co., Wilmington, DE). The 2.4 kb *Eco* RI-*HinDIII* fragment of pBS101 was labeled with [^{32}P]-ATP using a nick translation kit (Bethesda Research Laboratories Life Technilogies, Inc. Gaitherberg, MD).

Blots were prehybridized at 65℃ for 1 hr in a solution composed of 25 mM K_2HPO_4(pH 7.4), 5X SSC (20X SSC contains 3 M NaCl and 0.3 M Na3 citrate), 5X Denhart's solution (2% Ficoll 400, 2% polyvinylpyrrolidone, 2% bovine serum albumin), and 50 μg/ml salmon sperm DNA. For hybridization, denatured labeled probe DNA (10^6 cpm/ml) was added directly to the prehybridization solution, and the blot was incubated at 65℃ for 6 hr. After hybridization, the blot was washed three times at 65℃ in 0.5X SSC containing 0.1% SDS and then two times in 0.1X SSC containing 0.1% SDS for 15 min each time. Autoradiographic exposure was at room temperature using Cronex film (Du Pont).

Trapping of transposable elements of *X. oryzae* pv. *oryzae*. Plasmid pL3SAC(15), which consists of the *sacRB* gene (39) cloned into pLAFR3 (38) was conjugated into *X. oryzae* pv. *oryzae* race 2 strain PXO86 and race 5 strain PXO112 by biparental mating from *E. coli* strain S17-1 (35). Five single colonies of each exconjugant were grown to mid-exponential phase in nutrient broth and plated on nutrient agar supplemented with tetracycline (10 mg/l) and 5% sucrose. Plasmid DNA was isolated from the sucrose-resistant *X. oryzae* pv. *oryzae* exconjugants and transformed into *E. coli* HB101. The transformants were subjected to Southern blot analysis using the [^{32}P]-ATP labeled 2.4-kb *EcoRI-HinD*III fragment of pBS 101 as a probe.

Recombinant DNA menthods. The DNA of *X. oryzae* pv. *oryzae* which had inserted into the *sacRB* fragment of pL3SAC (clones *203a, 203b,* and *203c*) was digested with both *BamHI* and *HinDIII* and subcloned into pBluescript KS+ and /or KS-(Stratagene) digested with *BamHI* and *HinDIII*. DNA was ligated with T_4 DNA ligase by the procedure of Sambrook *et al.* (31). Clones were transformed into competent *E. coli* MV1190 (25) and plated onto YT medium (per liter:8g tryptone, 5g yeast extract, 2.5g NaCl, 0.01g thiamine, 15g Bacto-agar, pH 7.0), containing

X-gal (40 mg/l), IPTG (20 mg/l), and carbenicillin (100 mg/l). Bacteria from white colonies were streaked to obtain single colonies, and the presence of the cloned fragment was confirmed by restriction enzyme analysis of plasmid preparations. The isolates containing the inserted DNA were subjected to deletion mutagenesis by digestion with exonuclease III as described by Ausubel et al. (1).

DNA sequencing. Single-stranded DNA from the deleted DNA were purified (20) and used as template in the dideoxy sequencing method of Sanger *et al.* (32). Some sequences were determined using double stranded sequencing; the template for double stranded sequencing was prepared by the Gene-Clean™(Bio 101) procedure. The sequencing reaction was done using a Sequenase version 2.0 kit as described by the manufacturer (USB, Cleveland, Ohio). Portions of the sequence were determined using synthetic oligonucleotide primers. Primer A (5′-TCGTGGCAAGTATTGGC-3′) and B (5′-CCCTCAGCAGGACCTCGATC-3′) were synthesized using a DNA synthesizer (Applied Biosystems). Primers [C(5′-CCCGACGGCACCGCCGA-3′), D (5′-CCTTTCCGTGCGAGGCC-3′), and E (5′- AGCCGCCTGGACCTGTC-3′)] were purchased (Operon, Alameda, CA).

Computer analysis of nucleotide and predicted amino acid sequences. DNA and protein sequence data were compared to the GenBank library using the FASTA software (28). Multiple alignments were obtained using GENALIGN software that was supplied by Intelligenetics (Intelligenetics, Mountain View, CA; 37). Open reading frame analysis was performed using SEQAID software (D. Rhoads and D. Roufa, Kansas State University, Manhattan, KS).

RESULTS

Trapping *IS203* related elements. To isolate transposable elements from *X. oryzae* pv. *oryzae*, we used a positive slection strategy similar to that used to isolate elements from *X. campestris* pv. *vesicatoria* (15). Plasmid pL3SAC (15), which contains *sacRB* (39), was introduced into *X. oryzae* pv. *oryzae*. The *sacRB* genes encodes levansucrase, which is secreted into the medium after induction of the genes by sucrose. The levan produced by levansucrase on medium containing high sucrose is lethal to most Gram negative bacteria, including *X. oryzae* pv. *oryzae*. Growth of *X. oryzae* pv. *oryzae* containing pL3SAC was arrested in the presence of 5% sucrose. Individual colonies of *X. oryzae* pv. *oryzae* were selected that were

capable of growth on 5% sucrose. Plasmid DNA from the sucrose-viable *X. oryzae* pv. *oryzae* mutants was purified and transformed into *E. coli* HB101. The plasmid DNA was isolated from *E. coli* transformants and examined in agarose gels after digestion with various restriction enzymes. Of twenty transformants screened, all had insertions in the 2.7 kb *PstI-BamHI sacRB* fragment (data not shown).

To determine if any of the DNA inserted into pL3SAC shared identity with the *EcorRI-HinDIII* fragment of pJEL101, 221 *E. coli* transformants were screened on colony blots. Three elements (pLSAC*203a*, pLSAC*203b* and pLSA*203c*) were trapped that hybridized with the pJEL101 fragment. Restriction enzyme analysis indicated that three elements represented three independent insertion events into the *sacRB* gene (Fig. 1).

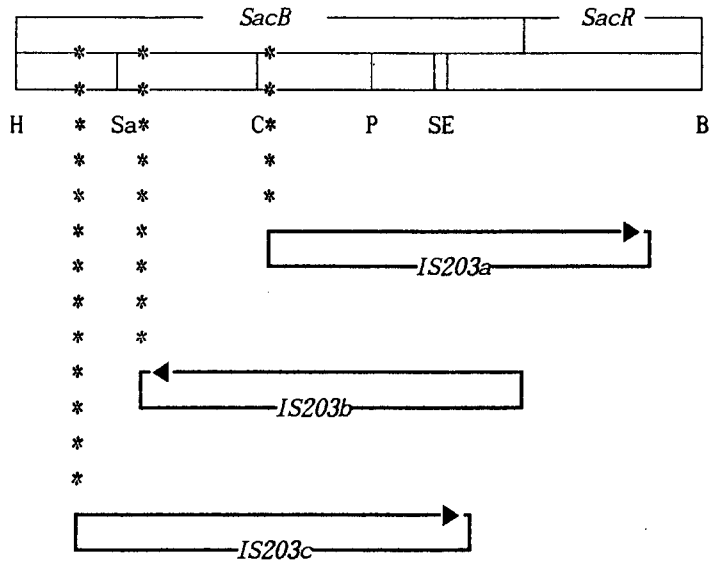

Figure 1. IS*203* elements inserted into *sac*RB gene of pL3SAC vector.
H, *Hind* Ⅲ; Sa, *Sac* Ⅱ; C, *Cla* Ⅰ; P, *Pvu* Ⅱ; S, *Stu* Ⅰ; E, *Eco* Ⅰ; B, *BamH* Ⅰ

DNA sequence analysis of pJEL101 and related elements. The IS elements inserted into the *sacRB* gene were subcloned into pBluescript (+ and/or −) as *BamHI-HinDIII* fragments. Nucleotide sequence was determined for the IS element portion of pBS101 by a combination of deletion mutagenesis and the use of synthetic primers (Fig. 2). IS*203* elements (*203a*, *203b*, and *203c*) were sequenced using a set of primers derived from pBS101 (Fig. 2). A sequence of 1049 bp within pBS101 that was bordered by 22bp imperfect inverted repeats was found. This sequence was labeled IS*Xoo1*. The three elements that werre trapped in *sacRB*

were 1055-bp in length and longer than IS*Xoo1* by six nucleotides.

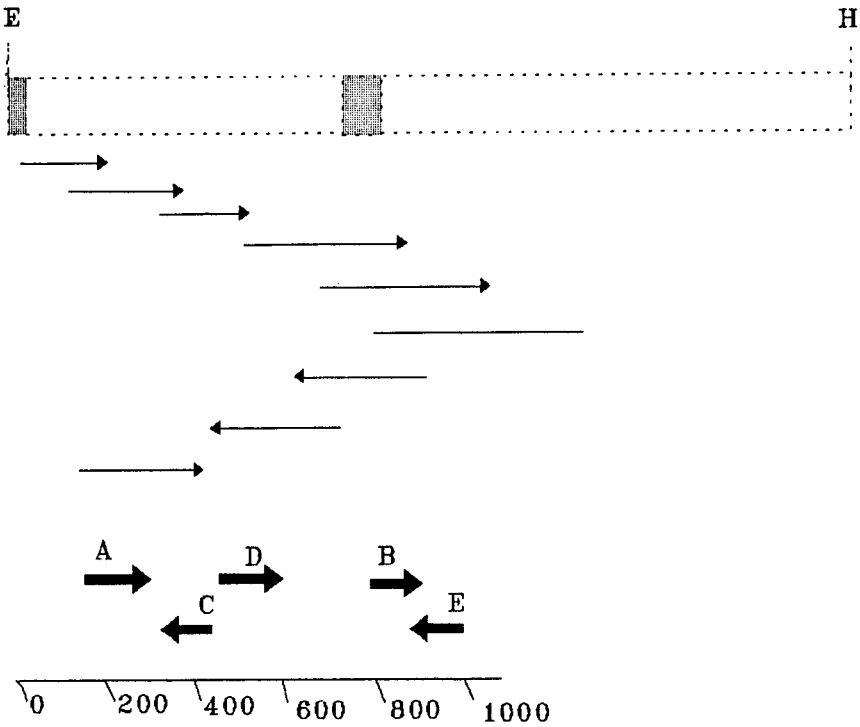

Figure 2. DNA sequence strategy. Arrows represent direction and extent sequence from a deletion clone. Wide arrows indicates the approximate position of sequencing primers A, B, C, D, and E. Lower scale is in base pairs. E, EcoRI; H, HindⅢ.

Alignment of all sequences indicates the ISXoo1 is actually 3-bp shorter at the left terminus of IS*203*, and that the inverted repeat is 25-bp in length. Alignment by computer analysis revealed that IS*203a* and IS*203b* have 96% and 99% identity, respectively, with the repetitive sequence (IS*Xoo1*) of pJEL101(Fig. 3). IS*203c* has the same DNA sequence as IS*203b* but the two elements had inserted into the *sacRB* gene in opposlite orientations. IS*Xoo1* probably does not contain the full-length IS element because three bases were removed by digestion with *EcoRI* prior to the cloning of the *EcoRI-HinDlll* fragment into plasmid pUC18 to construct pJEL101 (17).

Insertion sites of the IS*203* elements into sacRB were determined by comparison of the IS*203* sequence with that of the intact *sacRB* (Fig. 4). The sequence of DNA flanking the IS*203* elements was unique in each case. The number of dupli-

```
IS.Yoo1   GAATTCAACTCTGGATCGCAACACCAACGGTTGTGAAGTGATCCAGGCGACCTGACCTGAGCGACTTCACTACCAAATGGAGTTGC
IS203b GGCGCC----------------------------------------------------------------------------------
IS203a GGCGCC----------at--------------c----------g----------------------------t-g----g---------

CCATGTCATCCAGCCGcCTGGACCTGTCAGAACGATACCGCCTACATGCGTTATATGAAACCGGGATGTCGATGCGCGCCATCGCCGAT
---------------g--------------------------------------------------------------------------
-t-----c--------c-----------g--------------------c--c-----------------------------------

GCAGTGGCGCGTGCGCCCAGCACGATCAGTCGTGAGCTGCGCCGCAACCGGCACGCGGCGAAGTATCGGCCCGATCACGCGCAGCGCAT
--------------------------------------------------------------------------------------------
---t---a--------------------c--c--a--------t--t-a------t---cg---c-----------------------

CAGCGAGCATCGGCGCgCACAGGCCAGCCGGCGaCCACGCATCGACGCTGAGCGTATCcGtCAGATCGAGGtCCTGCTGAGGGAGGACT
---------------A---------------T----------------------G-C----------A-----------------
---------------A---------------T----------------------G-C----------A-----------------

TCAGTCCCGAACAGATTGCCGGTCGCACCGGCTTGGCCAGTCACGcATGGATCTATCGGCACATCGACGCCGATCAGAAGCGCGGTGGT
-------------------------------------------A--------------------------------------------
-------------------------------------------A------------------t---------------------

CAGTTGTTCATGCATCTACGCAAACGCCGCCGCAAGCGCCGTCGGCGTGGCGTGCGCGATGGCCGCGGGCAGCTGACGCATCGGCGCAG
--------------------------------------------------------------------------------------------
--a-----------------------------------------a-------------------------------------------

CTGGACACAGCGCCCCAGcGTGGTTGAGCAGCGAAGCCGTATCGGCGACTGGGAGCTGGAgACCATCAGGGCCTCGCACGGAAAGGGCG
-----------------T-------------------------------------------T----------------------------
-----------------T--------------c-----c-------------------------------------------------t-

TGGTGGTCAGCATGACCGAACGCCGCAGTCGCCTGCATCTGCTGGCTTACTCGCCCGACGGCACCGCCGAGAACGTGCGCAACGCCATT
--------------------------------------------------------------------------------------------
-------------------------------t--------------------c-----------------------------------

GTCCAGCGACTGGGCGGCCTGCGCCATGCAGTTCACACCCTCACCGCCGACAACGGCAAGGAGTTCGCCGATCATCGGCTCATTGCCGC
--------------------------------------------------------------------------------------------
---------------------------a----------g-------------------------------------------------

CTGCCTGCAGAGCGATTTCTATTTCGCAGATCCGTACTGCCCATGGCAGCGCGGCAGCAACGAGAATGCCAACGGATTGACACGCCAAT
--------------------------------------------------------------------------------------------
----t-----------------------------------g----------------------------------g-------------

ACTTGCCACGACAGACCGATTTCAGCACCATCACCGATGCGCACCTGCGATGGATCGAGCAGCGGCTCTACAATCGTCCGCGCAAGATA
--------------------------------------------------------------------------------------------
-----------------------------------a--------------------------------------------------------

CTTGGATTCAAAACGCCCCTCGAAGTCTTCTCCGAGGAGGTCCTCAAAAGCGTTGCGAATCAGAGTTGAATTc
-------------------------------------------------------------------------tGCC
-------------------------------------------c----------------------cGCC
```

Figure 3. Aligment of nucleotide sequence of IS*203* elements.

cated nucleotides was the same (3), but the nucleotide base composition of each target site was different. IS*203a*, IS*203b*, and IS*203c* were CGG, CAC, and GGC, respectively(Fig. 4). Each IS*203* element has partially matched terminal inverted repeats: IS*203a* matched 23 of 25 base pairs and IS*203b* and IS*203c* matched 20 of 25 base pairs (Fig. 4). Among all of the elements, 20 base pairs of the terminal inverted repeat were conserved (Fig. 5).

Based on the right termini, the predicted target site and three missing bases of IS*Xoo1* are GCC and GGC, respectively.

IS*Xoo1* has 50.8% identity with IS*4551* of *E. coli* (36) over a 909 nucleotide stretch based on a search of Genbank database (Fig. 6). Twelve base pairs of the

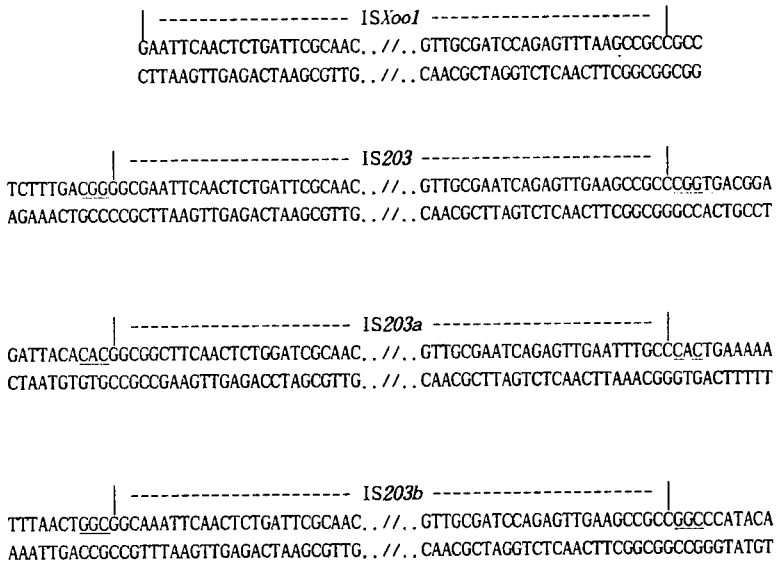

Figure 4. Target sites and terminal inverted repeats of each IS*203* element (IS*203a*, IS*203b*, and IS*203c*) at insertion into the *sacRB* gene. The sequence shown outside IS*203* element were part of *sacB* gene. Underlined three bases are duplicated target site. IS*Xool* has not contain whole sequence of IS element.

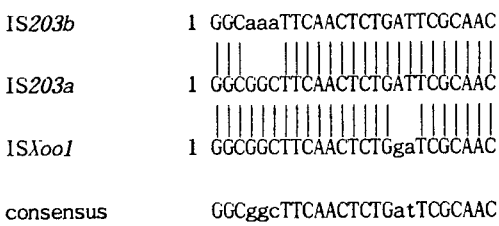

Figure 5. Consensus sequence of left terminal inverted repeat of IS*203* elements.

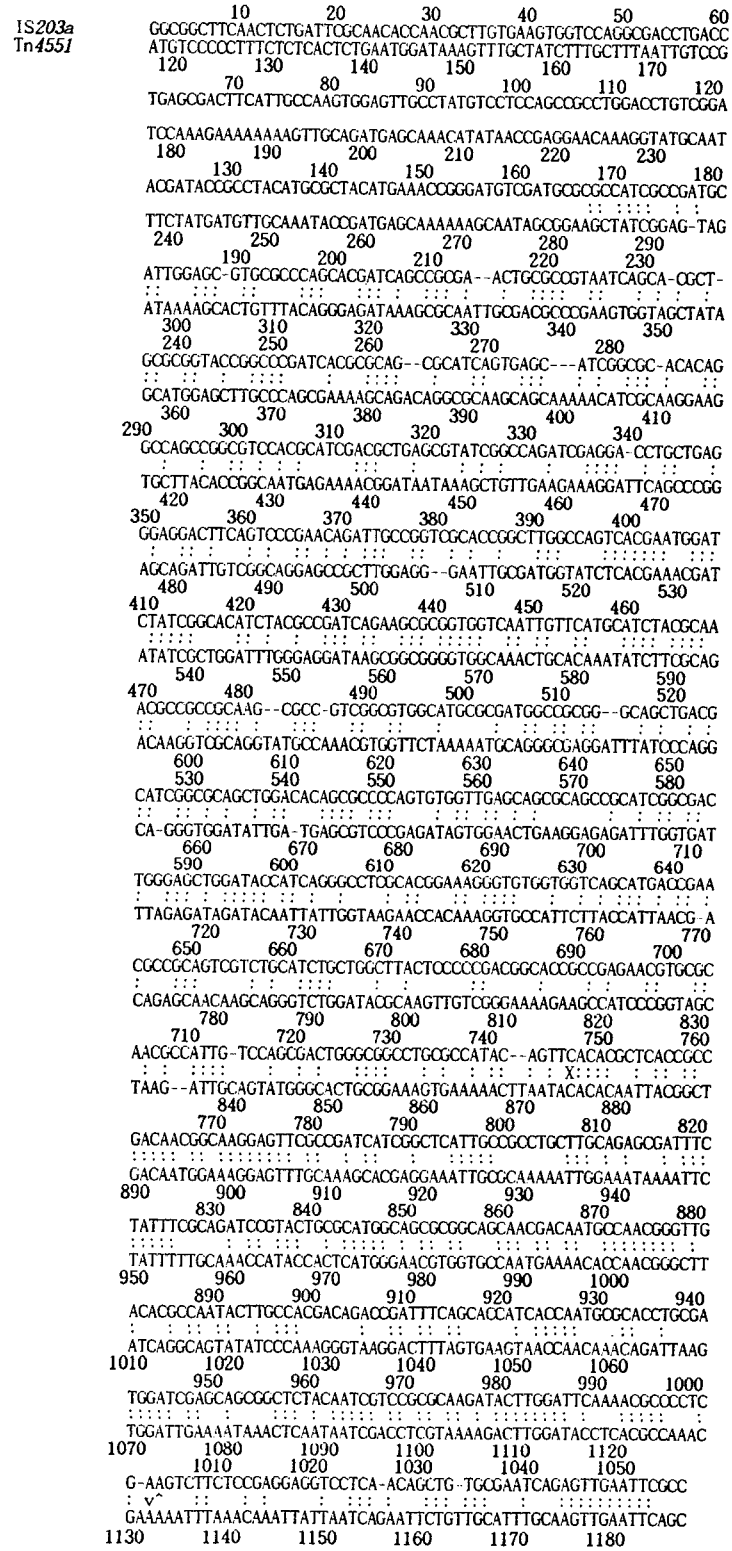

Figure 6. Comparison of IS*203a* with transposon Tn*4551*.

terminal inverted sequence were conserved between the IS*203* elements and IS*4351* of *Bacteroides fragilis* (29) (Fig. 7). Each IS*203* element sequence has six possible open reading frames (ORFs) and each has 50 or more codons between an AUG and stop codon (Fig. 8).

```
IS4351      1 CTTGAGTTCAACTTATAAATGCAAC
                ¦ ¦¦¦¦¦¦  ¦  ¦¦¦¦¦
IS203a      1 GGCGGCTTCAACTCTGAtTCGCAAC
              ¦¦¦¦¦¦¦¦¦¦¦¦¦¦¦  ¦¦¦¦¦¦¦
ISXoo1      1 GGCGGCTTCAACTCTGgaTCGCAAC
              ¦¦¦  ¦¦¦¦¦¦¦¦¦¦  ¦¦¦¦¦¦¦
IS203b      1 GGCaaaTTCAACTCTGatTCGCAAC

consensus     ggcga-TTCAACTctgaatcGCAAC
```

Figure 7. Consensus sequence of left terminal inverted repeat of IS*203* elements and IS*4351*.

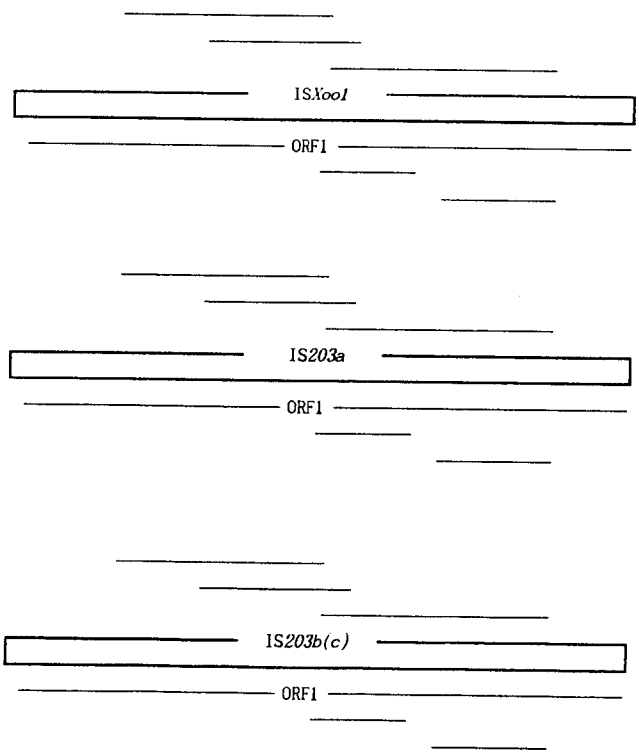

Figure 8. Open reading frames(ORFs) of IS*203* elements. All ORFs were shown for frames with 50 or more codons between an AUG and a stop codon. The ORFs reading from left to right are shown on top of each element and those from right to left on the bottom. The longest ORF were designated as ORF1.

```
        10         20         30         40         50         60
GGCGGCTTCAACTCTGATTCGCAACACCAACGCTTGTGAAGTGGTCCAGGCGACCTGACC
                                                      -35

        70         80         90        100        110        120
TGAGCGACTTCATTGCCAAGTGGAGTTGCCTATGTCCTCCAGCCGCCTGGACCTGTCGGA
                     RBS      M  S  S  S  R  L  D  L  S  E

       130        140        150        160        170        180
ACGATACCGCCTACATGCGCTACATGAAACCGGGATGTCGATGCGCGCCATCGCCGATGC
  R  Y  R  L  H  A  L  H  E  T  G  M  S  M  R  A  I  A  D  A

       190        200        210        220        230        240
ATTGGAGCGTGCGCCCAGCACGATCAGCCGCGAACTGCGCCGTAATCAGCACGCTGCGCG
  L  E  R  A  P  S  T  I  S  R  E  L  R  R  N  Q  H  A  A  R

       250        260        270        280        290        300
GTACCGGCCCGATCACGCGCAGCGCATCAGCGAGCATCGGCGCACACAGGCCAGCCGGCG
  Y  R  P  D  H  A  Q  R  I  S  E  H  R  R  T  Q  A  S  R  R

       310        320        330        340        350        360
TCCACGCATCGACGCTGAGCGTATCGGCCAGATCGAGGACCTGCTGAGGGAGGACTTCAG
  P  R  I  D  A  E  R  I  G  Q  I  E  D  L  L  R  E  D  F  S

       370        380        390        400        410        420
TCCCGAACAGATTGCCGGTCGCACCGGCTTGGCCAGTCACGAATGGATCTATCGGCACAT
  P  E  Q  I  A  G  R  T  G  L  A  S  H  E  W  I  Y  R  H  I

       430        440        450        460        470        480
CTACGCCGATCAGAAGCGCGGTGGTCAATTGTTCATGCATCTACGCAAACGCCGCCGCAA
  Y  A  D  Q  K  R  G  G  Q  L  F  M  H  L  R  K  R  R  R  K

       490        500        510        520        530        540
GCGCCGTCGGCGTGGCATGCGCGATGGCCGCGGGCAGCTGACGCATCGGCGCAGCTGGAC
  R  R  R  R  G  M  R  D  G  R  G  Q  L  T  H  R  R  S  W  T

       550        560        570        580        590        600
ACAGCGCCCCAGTGTGGTTGAGCAGCGCAGCCGCATCGGCGACTGGGAGCTGGATACCAT
  Q  R  P  S  V  V  E  Q  R  S  R  I  G  D  W  E  L  D  T  I

       610        620        630        640        650        660
CAGGGCCTCGCACGGAAAGGGTGTGGTGGTCAGCATGACCGAACGCCGCAGTCGTCTGCA
  R  A  S  H  G  K  G  V  V  V  S  M  T  E  R  R  S  R  L  H

       670        680        690        700        710        720
TCTGCTGGCTTACTCCCCCGACGGCACCGCCGAGAACGTGCGCAACGCCATTGTCCAGCG
  L  L  A  Y  S  P  D  G  T  A  E  N  V  R  N  A  I  V  Q  R

       730        740        750        760        770        780
ACTGGGCGGCCTGCGCCATACAGTTCACACGCTCACCGCCGACAACGGCAAGGAGTTCGC
  L  G  G  L  R  H  T  V  H  T  L  T  A  D  N  G  K  E  F  A

       790        800        810        820        830        840
CGATCATCGGCTCATTGCCGCCTGCTTGCAGAGCGATTTCTATTTCGCAGATCCGTACTG

  D  H  R  L  I  A  A  C  L  Q  S  D  F  Y  F  A  D  P  Y  C

       850        860        870        880        890        900
CGCATGGCAGCGCGGCAGCAACGAGAATGCCAACGGGTTGACACGCCAATACTTGCCACG
  A  W  Q  R  G  S  N  E  N  A  N  G  L  T  R  Q  Y  L  P  R

       910        920        930        940        950        960
ACAGACCGATTTCAGCACCATCACCAATGCGCACCTGCGATGGATCGAGCAGCGGCTCTA
  Q  T  D  F  S  T  I  T  N  A  H  L  R  W  I  E  Q  R  L  Y

       970        980        990       1000       1010       1020
CAATCGTCCGCGCAAGATACTTGGATTCAAAACGCCCCTCGAAGTCTTCTCCGAGGAGGT
  N  R  P  R  K  I  L  G  F  K  T  P  L  E  V  F  S  E  E  V

      1030       1040       1050
CCTCAACAGCGTTGCGAATCAGAGTTGAATTCGCC
  L  N  S  V  A  N  Q  S
```

Figure 9. Open reading frame of IS*203a*. Ribosome binding site(RBS) and pomoter region(−35) were underlined.

Figure 10. Comparison of each open reading frame of IS*203* elements.
1) M1 = IS*Xoo1*, 2) M2 = IS*203a*, 3) M3 = IS*203b(c)*.

Of these frames, only the longest open reading frame (hereafter ORF1) of each element was preceded by a putative ribosome binding site ('5-GGAGTT-3') (34)at nucleotide 82 and a promoter-like sequence ('5-CTGACC-3') at nucleotide 55. The ORF1 of each IS*203* element [IS*Xoo1*, IS*203a*, and *203b(c)*] extended from nucleotide 92 to nucleotide 1048 (Fig. 9) and could code for a 318 amino acid polypeptide of molecular weight 37, 045, 37, 151, and 37, 036, respectively. The amino acid sequences of each IS*203* element were more than 95% conserved (Fig. 10). The putative polypeptide of IS*203* ORF1 has 39.2% identity in 316 amino acids of ORF1 of Tn*4551* (36) and 30.6% identity over a 216-amino acid stretch of IS*30* transposase(5) (Fig. 11).

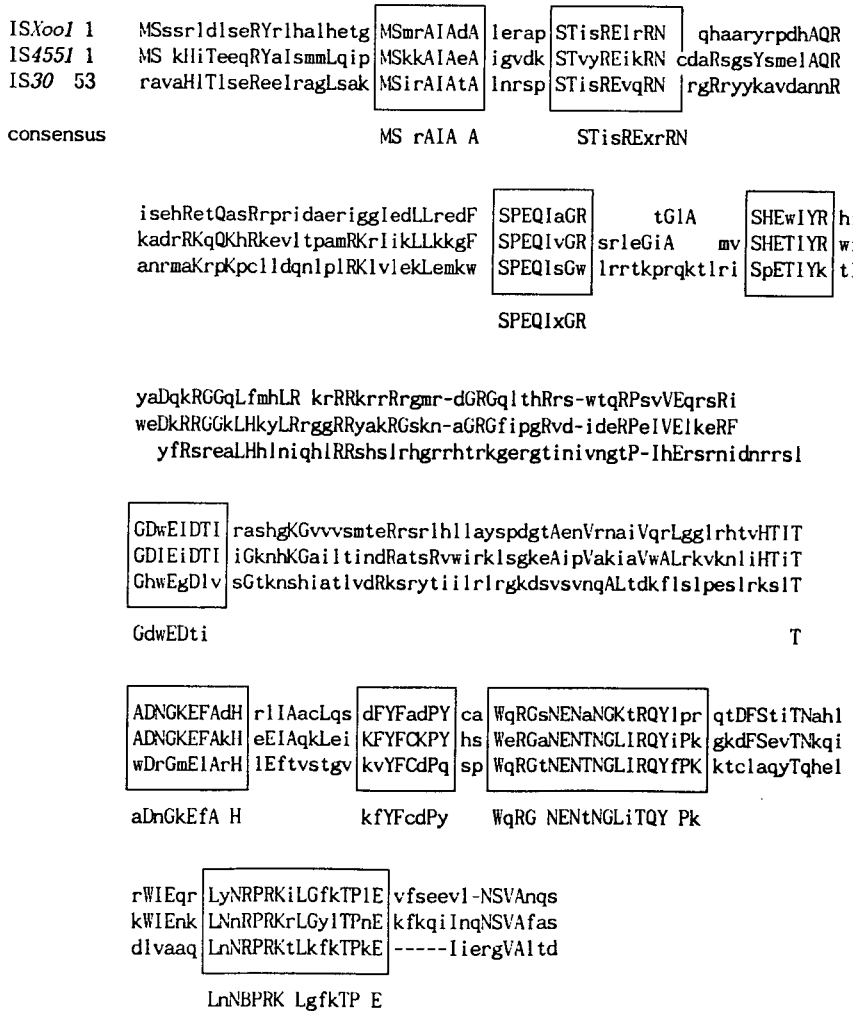

Figure 11. Comparison of the open reading frame of IS*Xool* with ORF-1 of Tn *4551* and transposase of IS*30*. Conserved regions are blocked.

DISCUSSION

Earlier studies of the genetic diversity among strains of *X. oryzae* pv. *oryzae* revealed that the species appears to contain a high number of repetive elements (17). The repetitive element that was represented on pJEL101 is the first of the elements from *X. oryzae* pv. *oryzae* to be characterized and was found to be a member of an insertion element family which we collectively refer to as IS*203*. All of the elements characterized in this work were 1055 base pairs in length with the exception of the element that was contained on pJEL101. This element was probably shortened due to the removal of three terminal nucleotides upon cleaving the element with the restriction endonuclease *EcoRI* during the cloning process.

IS*203* has all the hallmarks of a procaryotic insertion sequence. The intactelements are bounded by 25-bp imperfect inverted repeats and create a 3-bp duplication at the site of insertion. The elements of 1055bp were competent for transposition as evidenced by their insertion into the *sacRB* gene suggesting that the sequence is the functional unit for transposition. All three of the intact elements contain an open reading frame of 318 amino acid residues that has 40% and 30% sequence identity with putative transposases of Tn*4551* (36), Tn*4351* (29), and IS*30* (44), respectively. The absence of structural defects in the coding frame provides support to the hypothesis that the intact elements are autonomous elements. It remains unknown whether any of the three are capable of autonomous transposition, and experiments are in progress to determine the frequency of transposition in *E. coli*.

The intact elements possess inverted imperfect repeats of 25 base pairs. Similar repeats were identified in the direct repeats of Tn*4351* and Tn*4551* (36, 29). Twelve of the 25 bases appear to be conserved between the *B. fragilis* and *X. oryzae* pv. *oryzae* elements, an observation which leads to the suggestion that the conserved sequence may be an important core sequence required by the transposition mechanism. The elements from both species cause a 3-bp duplication at the insertion site.

Regardless of the implications for the transposition mechanism, the conserved nature of the inverted repeats and the entire IS*203* and IS*4351* element was proposed to have arisen in *Bacteroides* (29). The data here bring to light additional possibilities. Original experiments with the genomic fragment (on pJEL101) revealed hybridization to DNA fragments in a wide variety of *Xanthomonas* species (17), and the element is likely to be a member of a large family of elements present

in many gram-negative bacteria species. IS*30* of *E. coli* appears to be a even more distant member of the family (44). It may be impossible to determine in which species the element arose. In this regard, the percentage of the G+C content is interesting. The percentage of the IS*4351* element is similar to the G+C content of *Bacteroides* (41−44%, 13), while the G+C content of IS*203* is similar to the G+C content of *X. oryzae* pv. *oryzae* (63−71%; 3). The G+C content would appear not to be a good indicator for sequence origin. However, the content may suggest a mechanism that leads to acquisition of a G+C content of the whole organism. Alternatively, the elements may have been present in the ancestral organism, and may simply refect the evolutionary distance between the species. Further sequence comparisons between these species would be required for more insight regarding the evolution of the elements.

IS*203* was previously estimated to have 81 copies in the *X. oryzae* pv. *oryzae* (17), and the prevalence of elements may reflect the selective pressure placed on the species by rice cultivation practices. It seems likely that IS*203* and other elements play an important role in the adaptation of the bacteria to new rice cultivars and growing conditions. In this regard it is interesting to note that the related IS*4351* and IS*4551* elements have promoter activities as well as transposition capabilities (36, 29). The promoter activity of IS*203* remains to be determined.

LITERATURE CITED

1. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, and J. A. Smith, 1988. Current Protocols in Molecular Biology. John Wiley & Sons, New York.
2. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7: 1513-1523.
3. Bradbury, J. F. 1984. Xanthomonas. pp. 199-210. In: N. R. Kreig (ed.), Bergey's manual of systematic bacteriology. Vol. 1. Williams and Wilkins co., Baltimore.
4. Comai, L., and T. Kosuge. 1983. Transposable element that causes mutations in a plant pathogenic *Pseudomonas*. J. Bacteriol. 154: 1162-1167.
5. Dalrymple, B., P. Caspers and W. Arber. 1984. Nucleotide sequence of the prokaryotic mobile genetic element IS*30*. EMBO J. 3: 2145-2149.
6. Dodson, K. W., and D. E. Berg. 1989. Factors affecting transposition activity of IS*50* and Tn5 ends. Gene 76: 207-213.

7. Flores, M., V. Gonazalez, S. Brom, E. Martinez, D. Pinero, D. Romero, G. Davila, and R. Palacios. 1987. Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. J. Bacteriol. 169: 5782-5788.
8. Galas, D. J., and M. Chandler. 1989. Bacterial insertion sequences. p 10-162. *In* D. E. Berg, and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, DC.
9. Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman, and C. I. Kado. 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. 164: 918-921.
10. Hartl, D. L., and S. A. Sayer. 1988. Why do unrelated insertion sequences occur together in the genome of *Escherichia coli?* Genetics 118: 537-541.
11. Haugland, R. A., U. M. X. Sangodkar, and A. M. Chakrabarty. 1990. Repeated sequences including RS1100 from *Pseudomonas cepacia* AC1100 function as IS elements. Mol. Gen. Genet. 220: 222-228.
12. Higgins, C. F., G. Ames, W. Barnes, J. Clement, and M. Hofnung. 1982. A novel intercistronic regulatory element of prokaryotic operons. Nature (London) 298: 760-762.
13. Holdeman, L. V., R. W. Kelly, and W. E. C. Moore. 1984. *Bacteroides.* spp. p. 604-631. *In*: N. R. Krieg (ed.), Bergey's Manual of Systematic Bacteriology, Vol. 1. Williams and Wilkins Co., Baltimore.
14. Lida, S., J. Meyer, and W. Arber. 1983. Prokaryotic IS elements. p. 159-221 In: J. A. Shapiro, (ed.), Mobile Genetic Elements. Academic Press, Inc., New York.
15. Kearney, B., and B. J. Staskawicz. 1990. Characterization of IS*476* and itsrole in bacterial spot disease of tomato and pepper. J. Bacteriaol. 172: 143-148.
16. Kleckner, N. 1981. Transposable elements in prokaryotes. Annu. Rev. Genet. 15: 341-404.
17. Leach, J. E., F. F. White, M. L. Rhoads, and H. Leung. 1990. A repetitive DNA sequence differentiates *Xanthomonas campestris* pv. *oryzae* from other pathovars of *X. campestris.* Mol. Plant-Microbe Interact. 3: 238-246.
18. McClure, W. R. 1985. Mechanism and control of transcription initiation in prokaryotes. ann. Rev. Biochem. 54: 171-204.
19. McLafferty, M. A., D. R. Harcus, and E. L. Hewlett. 1988. Nucleotide sequence and characterization of repetitive DNA element from the genome of *Bordetella pertussis* with characteristics of an insertion sequence. J. Gen. Microbiol. 134: 2297-2306.
20. Messing, J. 1983. New M13 vectors for cloning. Meth. Enzymol. 101: 20-78.

21. Mew, T. W., 1987. Current status and future propects of research on bacterial blight of rice. Annu. Rev. Phytopathol. 25: 359-382.
22. Miller, J. H. 1972. Experiments in Molecular Genetics. Cold spring Harbor Laboratory, Cold Spring Harbor, NY. 466pp.
23. Mogen, B. D., and A. E. Oleson. 1987. Homology of pCSI plasmid sequences with chromosomal DNA in *Clavibacter michiganese* subsp. *sepedonicum:* Evidence for the presence of repeated sequence and plasmid integration. Appl. Environ. Microbiol. 53: 2467-2481.
24. Mogen, B. D., H. R. Oleson, R. B. Sparks, N. C. Gudmestad, and A. E. Oleson. 1990. Genetic variation in strains of *Clavibacter michiganese* subsp. *sepedonicum*: Polymorphisms in restriction fragments containing a highly repeated sequence. Phytopathology 80: 90-96.
25. Morrison, D. A. 1979. Transformation and preservation of competent bacterial cells by freezing. Meth. Enzymol. 68: 326-331.
26. Nyman, K. K., K. Nakamura, H. Ohtsubo, and E. Ohtsubo. 1981. Distribution of the insertion sequence IS*1* in gram-negative bacteria. Nature (London) 289: 609-612.
27. Ohtsubo, E., H. Ohtsubo, W. Doroszkifwicz, K. Nyman, D. Allen, and D. Davison. 1984. An evolutionary analysis of ISO-IS*1* elements from *Escherichia coli* and *Shigella* strains. J. Gen. Appl. Microbiol. 30: 359-376.
28. Pearson, W. R. and D. J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natil. Acad. Sci., 85: 2444-2448.
29. Rasmussen, J. L., D. A. Odelson, and F. L. Macrina. 1987. Complete nucleotide sequence of insertion element IS*4351* from *Bacteroides fragilis*. J. Bacteriol. 169: 3573-3580.
30. Reznikoff, R. N. 1983. Some bacterial transposable elements:their organization, mechanisms of transposition, and roles in genome evolution. p. 229-252. In. J. Beckmith, J. Davies, and J. A. Gallant(ed.), Gene function in prokaryotes. Cold Spring Harbor Laboratory, cold Spring Harbor, N. Y.
31. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, N. Y.
32. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.
33. Scordilis, G. E., H. Ree, and T. G. Lessie. 1987. Identification of transposable elements which activate gene expression in *Pseudomonas cepacia*. J. Bacteriol. 169: 8-13.
34. 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. Nat. Acad. Sci. USA 71: 1342-1346.

35. Simon, R., U. Priefer, and A. Puhler. 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram-negative bacteria. Biotechnology 1: 784-791.
36. Smith, C. J. 1987. Nucleotide sequence analysis of Tn*4551*: use of *erm*FS operon fusions to detect promoter activity in *Bacteroides fragilis*. J. Bacteriol. 169: 4589-4596.
37. Sobel, E. and Martinez, H. M. 1985. A multiple sequence alignment program. Nucl. Acid Res. 14: 363-374.
38. Staskawicz, B., D. Dahlbeck, N. Keen, and C. Napoli. 1987. Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. J. Bacteriol. 169: 5789-5794.
39. Steinmetz, M., D. Le Coq, S. Aymerich, G. Gonzy-Treboul, and P. G. 1985. The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. Mol. Gen. Genet. 200: 220-228.
40. Szabo, L. J., and D. Mills. 1984. Integration and excision of pMC7105 in *Pseudomonas syringae* pv. *phaseolicola*: Involvement of repetitive sequences. J. Bacteriol. 157: 821-827.
41. Tsuchiya, K., T. W. Mew, and S. Wakinoto. 1982. Bacteriological and pathological characteristics of wild types and induced mutants of *Xanthomonas campestris* pv. *oryzae*. Phytopathology 72: 43-46.
42. Umeda, M., and E. Ohtsubo. 1989. Mapping of insertion elements IS*1*, IS*2* and IS*3* on the *Escherichia coli* K-12 chromosome role of the insertion elements in formation of Hfrs and F′ factors and in rearrangement of bacterial chromosomes. J. Mol. Biol. 208: 604-614.
43. Umeda, M., and E. Ohtsubo. 1990. Mapping of insertion element IS*30* in the *Escherichia coli* K12 chromosome. Mol. Gen. Genet. 222: 317-322.
44. Umeda, M. and E. Ohtsubo. 1991 four types of IS1 with differences in nucleotide sequence reside in the *E. coli* K12 chromosome. Gene 98: 1-5.