*Proc. Natl. Acad. Sci. USA*
Vol. 86, pp. 5340–5344, July 1989
Biochemistry

# Characterization of the termini and transposition products of Tn*4399*, a conjugal mobilizing transposon of *Bacteroides fragilis*

(cloning/vectors/DNA sequencing/inverted repeats/target sites)

DAVID W. HECHT*, JEFFREY S. THOMPSON, AND MICHAEL H. MALAMY†

Department of Molecular Biology and Microbiology, Tufts University Health Sciences Campus, 136 Harrison Avenue, Boston, MA 02111

*Communicated by Stanley Falkow, May 1, 1989*

**ABSTRACT** **We have isolated a 9.6-kilobase conjugal transposon, Tn*4399* from *Bacteroides fragilis*, that is capable of mobilizing nonconjugal plasmids in cis. Here we characterize the ends of the transposon, its target-site requirements, and the products of transposition into the *B. fragilis* chromosome and two sets of *B. fragilis–Escherichia coli* shuttle vectors. With the exception of an additional cytosine residue in the left end, there are perfect 13-base-pair (bp) inverted repeats at the ends of Tn*4399*. Insertion of Tn*4399* resulted in a 3-bp target-site repeat in 8 out of 12 independent transpositions and showed a high insertion-site specificity. A remarkable feature of Tn*4399* insertions is the presence of an additional 5 bp located between the right inverted repeat and the target-site repeat. Four sequence variations of the 5 bp were found, with absolute conservation at positions 1, 2, and 5. Only two of the variations were present in junction fragments of all three copies of Tn*4399* contained in the chromosome of the original donor strain, *B. fragilis* TM4.2321. Tn*4399* appears to represent a new type of conjugal transposon. In contrast to Tn*916* and Tn*1545*, described in streptococci, Tn*4399* creates a target-site repeat and contains an additional 5 bp at the right end only, between the transposon and the target sequence. In addition, Tn*4399* can mobilize plasmids in cis.**

Transposon Tn*4399* is a 9.6-kilobase (kb) conjugal mobilizing transposon from *Bacteroides fragilis*. It was originally identified because in *B. fragilis–Escherichia coli* matings, it conferred transferability on Tra$^-$ derivatives of the autonomous transfer factor pBFTM10 contained in the chimeric plasmid pGAT400 (1). The ability of Tn*4399* to mobilize plasmids in cis contrasts with the other known conjugal transposons, Tn*916* and Tn*1545* (2, 3), isolated from *Streptococcus faecalis* and *Streptococcus pneumoniae*, respectively. While Tn*916* can transfer itself, it has not been observed to transfer plasmid pAD2 under similar conditions. To determine the structure of Tn*4399* and the nature of its transposition products, we sequenced several Tn*4399* insertion sites. We used a new series of *B. fragilis–E. coli* shuttle vectors (ref. 4 and unpublished data), containing the *Eco*RI endonuclease gene under the control of the $P_R$ promoter of bacteriophage λ, to isolate Tn*4399* insertions into this gene. The *Eco*RI gene is nonfunctional in *B. fragilis*, but its inactivation by insertion or mutation is required for survival in an *E. coli* λ nonlysogen. Since the complete sequence of the *Eco*RI gene is known (5), it was possible to utilize oligonucleotide primers homologous to the *Eco*RI gene to obtain the sequence of the ends of Tn*4399*. With this information, oligonucleotides specific for the ends of the transposon were used to sequence additional Tn*4399* insertion sites. In addition, we sequenced three independent insertions of Tn*4399* that conferred Tra$^+$ status on pGAT400ΔBglII. We also sequenced the ends of the three potential donor sites from the chromosome of TM4.2321. We report here the sequence of the ends of Tn*4399*, its target-site requirements, and the unique nature of its insertion products.

## MATERIALS AND METHODS

**Bacterial Strains and Media.** The characteristics of the bacterial strains and plasmids used in this study are listed in Table 1. *Bacteroides* strains were grown in supplemented brain/heart infusion broth (solidified with 1.5% agar when required) under culture conditions and antibiotic concentrations as described (1).

**Plasmid Construction.** Plasmids pJST60, -61, and -62 are *B. fragilis–E. coli* chimeric plasmids based on the pEcoR251 plasmid used by Southern *et al.* (7) for cloning *B. fragilis* DNA. In addition to a clindamycin-resistance marker, which can be expressed in *B. fragilis*, and the pBFTM2006 replicon, which replicates in *B. fragilis*, these plasmids contain a pBR322-derived replicon expressing ampicillin resistance, an *oriT* sequence that allows mobilization in trans when plasmid RP4 or its derivatives are present in the same cell, and the *Eco*RI endonuclease structural gene under transcriptional control of the λ rightward promoter, $P_R$. Expression of the *Eco*RI protein from the high-copy-number plasmid in *E. coli* is lethal to the cell; however, the presence of λ in the chromosome suffices to repress *Eco*RI expression. A map of pJST61 is shown in Fig. 1.

**Plasmid Transfer Experiments.** Filter mating techniques were as described (1). Donor and recipient cultures were mixed and the cells were allowed to attach to the surface of a 0.45-μm Nalgene filter under aerobic conditions, then the filters were incubated overnight under anaerobic conditions. Transfer frequencies were calculated by dividing the number of antibiotic-resistant transconjugants by the number of input donor cells.

**Preparation of Plasmid DNA, Restriction Endonuclease Analysis, and DNA·DNA Hybridization Conditions.** These were as described (1).

**Cloning of *B. fragilis* Chromosomal DNA.** A complete *Hin*dIII digest of CsCl-purified TM4.2321 chromosomal DNA was electrophoresed in a 0.6% low-melting agarose gel (FMC). After ethidium bromide staining, regions of the gel containing specific size classes of DNA were excised and the DNA was purified by phenol/chloroform treatment and precipitated (6). Purified fragments were then ligated into the single *Hin*dIII site of pEcoR251 and used to transform HB101 to ampicillin resistance.

**DNA Sequencing.** Nucleotide sequences were determined by the Sequenase method (United States Biochemical) according to the supplier's instructions.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Loyola University Medical Center, Maywood, IL 60153.
†To whom reprint requests should be addressed.

Table 1. Bacterial strains and plasmids

| Strain or plasmid | Description | Reference |
|---|---|---|
| *B. fragilis* | | |
| TMP230 | Original source of transposon Tn*4399* | 1 |
| TM4000 | Standard recipient strain | 1 |
| TM4.2321 | Transconjugant of TMP230–TM4000; contains three copies of Tn*4399* | 1 |
| *E. coli* HB101 | — | 6 |
| Plasmids | | |
| pJST60 to -62 | *B. fragilis–E. coli* shuttle vectors | This paper |
| pJST100 to -109 | Tn*4399* insertion into pJST60, -61, or -62 | This paper |
| pGAT400ΔBglII | $Tra^-$ pGAT400 shuttle vector | 1 |
| pDWH20 to -22 | Tn*4399* insertion into pGAT400ΔBglII | This paper |
| pEcoR251 | — | 7 |

## RESULTS

**Analysis of the Products of Transposition of Tn*4399* to Plasmids pJST60, -61, and -62 in *B. fragilis*.** Plasmids pJST60, -61, and -62 were introduced into *B. fragilis* TM4000 and TM4.2321 (containing three independent copies of Tn*4399*) from *E. coli* HB101(λ) cells by RK231 mobilization, as described (1); selection was for clindamycin resistance. *B. fragilis* cells are not killed by the presence of the wild-type *Eco*RI gene. TM4.2321 transconjugants containing the pJST plasmids were mated with nonlysogenic HB101; by selecting for the transfer of ampicillin resistance, transconjugants were isolated at $10^{-8}$ per input donor. Transfer of the same plasmids from *B. fragilis* TM4000 to HB101 was not detected. Analysis of the products of transfer by restriction enzymes demonstrated that each plasmid had acquired a single copy of Tn*4399* at various positions within the *Eco*RI gene. The DNA sequence of a portion of the *Eco*RI gene showing the locations of nine independent Tn*4399* insertions is shown in Fig. 1. Each arrowhead represents the insertion of a single Tn*4399* from a separate mating.

**DNA Sequences of the Ends of Tn*4399*.** To determine the sequences of the ends of Tn*4399*, we first mapped the position of Tn*4399* in derivative pJST102 by using a series of restriction enzymes. Twenty-base-pair primers containing the sequence of the *Eco*RI gene flanking this insertion were prepared and used to determine the precise site of Tn*4399* insertion and the sequences at the ends of the transposon. Fig. 2 shows the DNA sequences of the right and left ends of Tn4399 within the *Eco*RI gene in pJST102. With the exception of an additional cytosine residue in the left end, the ends consist of 13-base-pair (bp) inverted repeats. An additional 8 bp between the right terminus and the *Eco*RI structural gene were also detected; the significance of these bases will be discussed below.

**Sequences of Nine Sites of Tn*4399* Insertion into the *Eco*RI Gene.** We used 20-bp primers corresponding to unique sequences at the left and right ends of Tn*4399* (Fig. 2) to determine the sequences of eight additional Tn*4399* insertion sites (Fig. 3 *Upper*). In each case the left inverted repeat begins precisely after the target site. However, in every case, there are an additional 8 bp between the right inverted repeat and the target site. In five of the nine sequences, the last 3 bp of the 8 bp added match the 3 bp immediately adjacent to the left inverted repeat; this may represent a 3-bp target-site

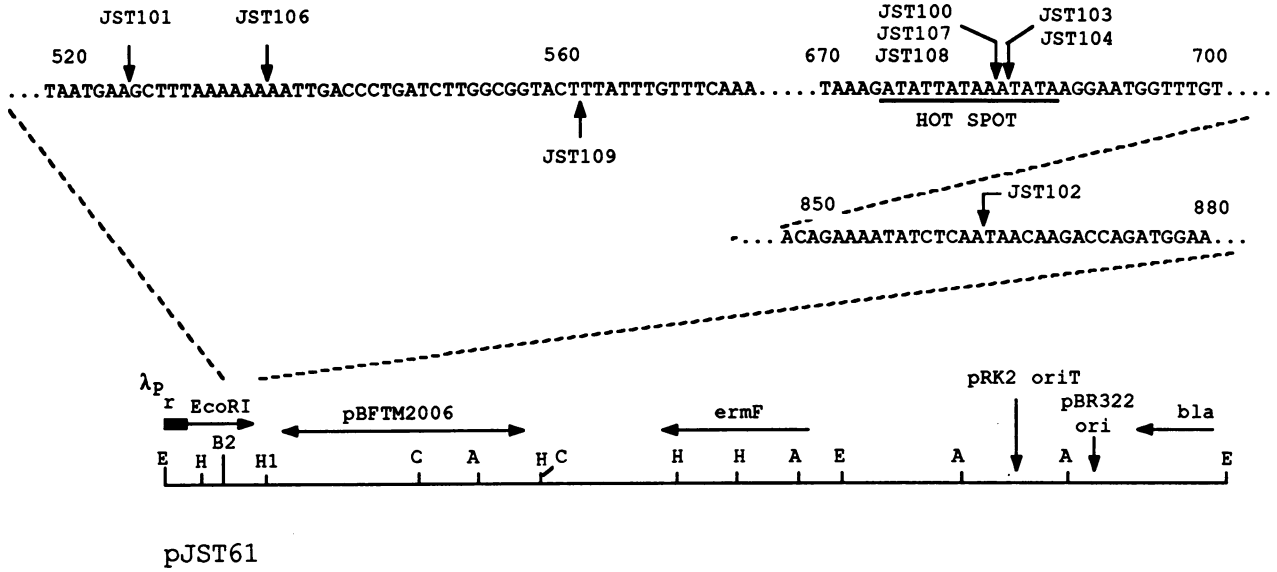


FIG. 1. (*Upper*) Sites of Tn*4399* insertion into the *Eco*RI gene of pJST60, pJST61, and pJST62. Base numbering for the *Eco*RI sequence is according to ref. 5. Vertical arrows indicate the sites of Tn*4399* insertions and their isolation numbers; the arrows above the line denote insertions with orientation LTR . . . RTR (LTR, left terminal repeat; RTR, right terminal repeat), while the arrow below the line (pJST109) denotes the opposite orientation. The underlined "hot spot" corresponds to the site at which five of the insertions occurred. (*Lower*) A partial restriction map of pJST61 indicating the components of the shuttle vector and the position of the *Eco*RI gene. pJST62 differs from pJST61 only in the orientation of the *B. fragilis* replicon; the *B. fragilis* replicon of pJST60 is in the same orientation as in pJST61. pJST61 and pJST62 both differ from pJST60 in that a 1.0-kilobase *Cla* I fragment containing the aerobic tetracycline-resistance gene and insertion sequence IS*4400*R has been deleted (4). Resistance to clindamycin and ampicillin are provided by the *ermF* and *bla* genes, respectively. Restriction sites: A, *Ava* I; E, *Eco*RI; H, *Hin*dIII; C, *Cla* I; H1, *Bam*HI; B2, *Bgl* II.

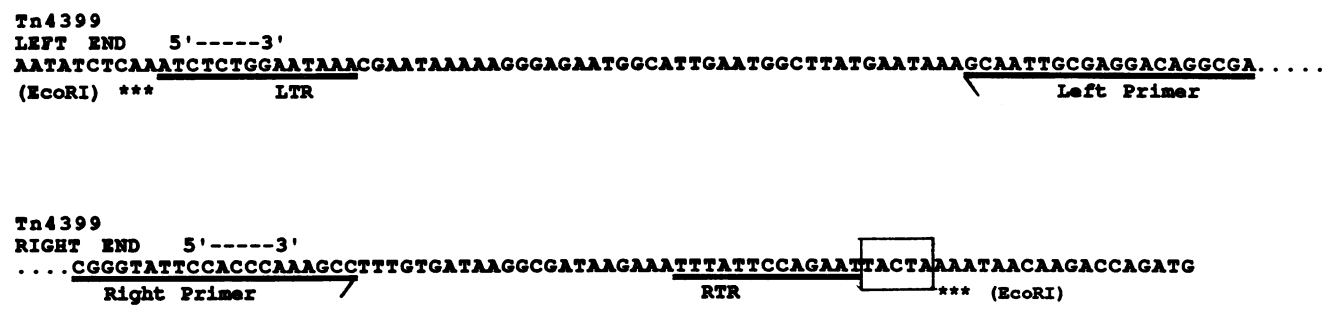


FIG. 2. DNA sequences of the right and left ends of Tn*4399* inserted into the *Eco*RI gene of pJST102. LTR, left terminal repeat; RTR, right terminal repeat. Left and right primers indicate the unique DNA sequences used to sequence Tn*4399* insertions. (EcoRI), DNA sequence of the *Eco*RI endonuclease gene (5). The box encloses the 5 additional bases, and asterisks indicate the target-site repeats within the 8 additional bases between the right terminal repeat and the *Eco*RI sequence.

repeat. The target-site sequences were A+T-rich and the target-site repeats always contained the sequence AA at positions 2 and 3. The remaining 5 bp between the right inverted repeat and the 3-bp repeat represented four different sequences (Table 2). The first, second, and fifth positions were conserved in all cases, but variability was observed at positions 3 and 4. Five independent Tn*4399* insertions occurred at a hot spot within the *Eco*RI gene (see Fig. 1). Three 5-bp sequences were found among insertions into this hot spot: TAAGA, TACTA, and TACGA.

**Sequences of Three Tn*4399* Insertions into pGAT400ΔBglII.** The same oligonucleotide primers were used to sequence three Tn*4399* insertions into pGAT400ΔBglII, which conferred transferability on the previously Tra$^-$ plasmid (1). The results are presented in Fig. 3 *Lower*. In these cases the sequence of the DNA target site is not known. Nevertheless, the sequences at the junction of Tn*4399* and the target site show the same basic pattern as the Tn*4399* insertions into the *Eco*RI gene: 3-bp repeats flanking the transposon, an additional 5 bp between the right inverted repeat and the 3-bp target-site repeat, and an A+T-rich target site. The 5 added bases observed in these three cases are the same as those observed for Tn*4399* insertions into the *Eco*RI gene.

**Cloning of the Tn*4399* Junction Sites from the Chromosome of *B. fragilis* TM4.2321.** The variability of the 5 bp located between the right inverted repeat and the target-site repeat

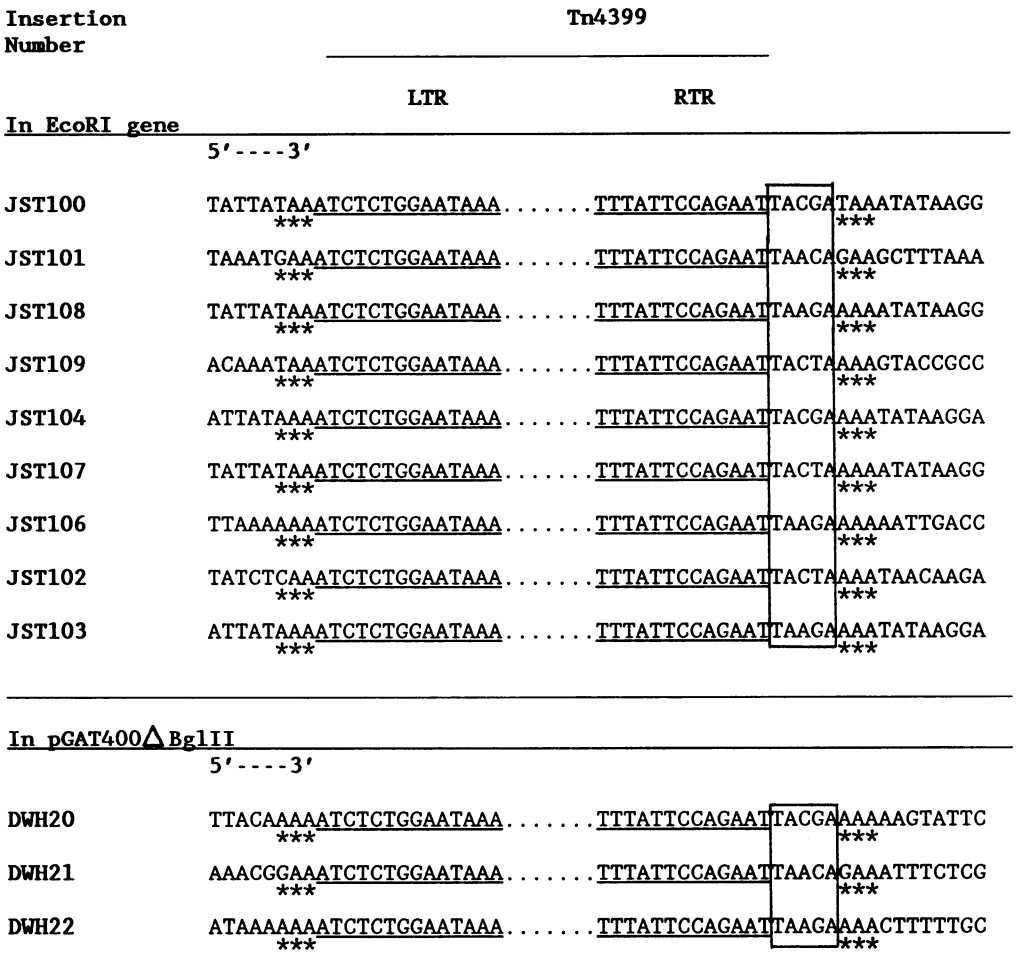


FIG. 3. DNA sequences of the junction regions of 12 Tn*4399* insertions. (*Upper*) The sites of insertion of JST100–109 are given in Fig. 1. (*Lower*) DHW20–22 are insertions into pGAT400ΔBglII (1). Abbreviations are as in Fig. 2. The target-site repeats are indicated by asterisks; the 5-bp regions are boxed.

Table 2. Sequence and occurrence of the 5-bp regions

| Sequence (5′ → 3′) | No. of copies in TM4.2321 | No. of products of transposition |
|---|---|---|
| TAAGA | 2 | 4 |
| TAACA | 1 | 2 |
| TACTA | 0 | 3 |
| TACGA | 0 | 3 |

could result if several Tn*4399* elements with different 5-bp regions were involved in the transposition events. We previously demonstrated that the chromosome of the donor cell used as the source of Tn*4399* contains three unlinked copies of the transposon (1). To determine the DNA sequence at the sites of insertion in TM4.2321, we cloned all of the Tn*4399* junction fragments from this strain. Purified fragments from *Hin*dIII-digested chromosomal DNA of TM4.2321 were cloned into the pEcoR251 plasmid. Transformants were screened by colony blot hybridization using cloned $^{32}$P-labeled left and right ends (pDWH11L and pDWH12R) of Tn*4399* as probes (1). Colonies showing homology were purified, and their hybridization was confirmed by using partially purified plasmid DNA. Restriction analysis of plasmid DNA confirmed that three different chromosomal fragments corresponding to the right junctions and three corresponding to the expected left junctions were obtained. In addition, a fourth clone showing weak homology with the left probe was also obtained.

**Sequences of the Tn*4399* Insertion Sites in *B. fragilis* TM4.2321.** CsCl-purified DNA of the cloned junction fragments from TM4.2321 was used for DNA sequencing. Fig. 4 presents the sequences obtained. The underlined bases represent the inverted repeats at the left and right ends. The region to the right of the right inverted repeat in the three junction fragments contained two of the four previously identified 5-bp regions: TAAGA (isolated in two junction fragments) and TAACA. The three left inverted repeat junctions did not show any bases corresponding to the 5-bp regions identified in the right junction fragments of the chromosomal or plasmid insertion sites. They presumably represent flush junctions of the left inverted repeat with the chromosomal target site. A fourth clone, L-W, showing weak homology with the left-hand probe differed from the previous three but exhibited a similar organization. The insertion sites in the TM4.2321 chromosome are A+T-rich regions, and each junction contains adenine residues at positions 2 and 3 of the proposed 3-bp target site repeat. (However, since the chromosomal sites have not been sequenced, we do not know which left-hand sequence goes with which right-hand fragment.)

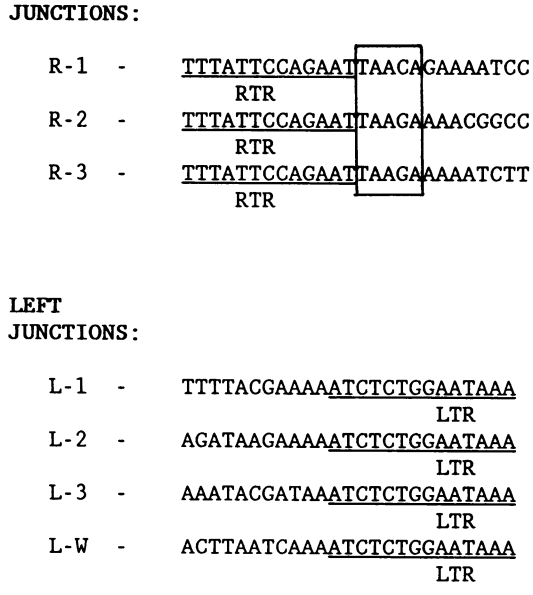


FIG. 4. DNA sequences of the right and left junction regions of the three copies of Tn*4399* in TM4.2321. R1–R3 are the right junction fragments; L1–L3 are the left junction fragments. L-W indicates the fourth fragment, showing weak homology to the left-hand probe. Abbreviations are as in Fig. 2.

## DISCUSSION

Tn*4399* is a conjugal transposon isolated from *B. fragilis* with the unusual property of causing the transfer of nonconjugal plasmids in cis. We have now determined the structure of the ends of this transposon and analyzed its target-site requirements and the products of its insertion into the chromosome of *B. fragilis* TM4000 and the plasmids pGAT400ΔBglII and pJST60, -61, and -62.

The ends of Tn*4399* are perfect 13-bp inverted repeats with an extra residue within the left inverted repeat. This property is shared with most simple transposons, which have short or long imperfect inverted repeats at their termini (8).

The insertion of Tn*4399* results in a 3-bp target-site repeat in 8 out of the 12 insertions reported here. The only variability occurs at position 1 of the 3-bp repeat, in which we find the substitution of an adenine in the right-hand copy for a thymine, cytosine, or guanine residue in the target sequence present at the left end. The target site is A+T-rich and consists of at least two adenine or two thymine residues. This results in a consistent AA or TT sequence at positions 2 and 3 of the target-site repeat. Five out of nine independent insertions in the *Eco*RI gene occurred within the same target site (ATATTATAAATATA) but at two different positions. This nonrandom distribution of events into a certain highly A+T-rich region suggests that this site is an insertional hot spot. There is a remarkable orientation preference for insertions within the *Eco*RI gene carried by pJST61; eight of nine have occurred with the orientation Tn*4399*LTR . . . Tn*4399*RTR (where LTR and RTR are the left and right terminal repeats) with respect to the direction of transcription of this gene. A·T target-site preference is common to many other transposons; 3-bp target-site repeats are rare but occur after transposition of the *Tam* transposons of maize (9) and have been reported for a single sequenced insertion site of insertion sequence IS*4351* from pBF4 of *B. fragilis* (10).

A unique feature of Tn*4399* is the presence of an additional 5 bases between the end of the right inverted repeat and the target site. The sequence of the 5 bases varies only at the third and fourth position, with limited substitution resulting in four different combinations. There must be at least one G·C pair within the 5 bp, at position 3 or 4 or both. Insertion of Tn*4399* into a target-site hot spot within the *Eco*RI gene leads to the formation of three different 5-bp sequences. In addition, the insertion of Tn*4399* into the chromosome of *B. fragilis* TM4000 or into plasmid pGAT400ΔBglII or plasmid pJST60, -61, or -62 gave rise to the same 5-bp sequences. These results imply either that the sequences of these 5-bp regions reflect the strict conservation of four different donor sites and are part of the transposon or that the variability is introduced during the transposition process. Analysis of the three right junction fragments obtained from the original Tn*4399* donor, strain TM4.2321, showed only two patterns of the 5 bases, TAAGA and TAACA. This suggests that there is conservation of the 5 bp at the right end of the transposon but that there can be occasional variation at positions 3 and 4. This explanation is supported by the fact that the one copy of Tn*4399* in TMP230 with a unique 5-bp sequence results in three copies with two 5-bp combinations of TM4.2321. An alternative explanation for the additional bases would be that the first 2 residues, TA, are actually part of the right end of the inverted repeat. This would leave only 3 additional bases

between the right inverted repeat and the target-site repeat. However, we favor the first hypothesis.

The structure at the right end of Tn*4399* is unique to this transposon. Conjugal transposons Tn*916* from *S. faecalis* and Tn*1545* from *S. pneumoniae* have been reported to contain an additional 4 random bases at either end of the transposon (11, 12).

The mechanism of Tn*4399* transposition is unknown, as is the process that introduces the variation in the 5-bp regions. We can postulate that the 5-bp region plays an important role in transposition, serving perhaps as the site of transposase action leading to a staggered cut of the transposon and a 3-bp staggered cut of the target sequence. The variability would be introduced as a result of mismatch fill-in of the resulting single-stranded regions. Why variability is limited only to the right junctions and only to positions 3 and 4 of the 5-bp repeat and position 1 of the target-site repeat on the right side is not clear.

This work was supported by National Institutes of Health Grant AI19497 (to M.H.M.). D.W.H. was the recipient of National Research Services Award AI07609-01.

1. Hecht, D. W. & Malamy, M. H. (1989) *J. Bacteriol.* **171,** in press.
2. Clewell, D. B. & Gawron-Burke, C. (1986) *Annu. Rev. Microbiol.* **40,** 635–659.
3. Courvalin, P. & Carlier, C. (1987) *Mol. Gen. Genet.* **206,** 259–264.
4. Robillard, N., Tally, F. P. & Malamy, M. H. (1985) *J. Bacteriol.* **164,** 1248–1255.
5. Newman, A. K., Rubin, R. A., Kim, S. H. & Modrich, P. (1981) *J. Biol. Chem.* **256,** 2131–2140.
6. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
7. Southern, J. A., Parker, J. R. & Woods, D. R. (1986) *J. Gen. Microbiol.* **132,** 2827–2835.
8. Kleckner, N. (1981) *Annu. Rev. Genet.* **15,** 341–404.
9. Wessler, S. R. (1988) *Science* **242,** 399–405.
10. Rasmussen, J. L., Odelson, D. A. & Macrina, F. L. (1987) *J. Bacteriol.* **169,** 3573–3580.
11. Clewell, D. B., Flannagan, S. E., Ike, Y., Jones, J. N. & Gawron-Burke, C. (1988) *J. Bacteriol.* **170,** 3046–3052.
12. Caillaud, F. & Courvalin, P. (1987) *Mol. Gen. Genet.* **209,** 110–115.