Integration and Excision of a Newly Discovered *Bacteroides* Conjugative Transposon, CTnBST[†]

Neil A. Wesslund, Gui-Rong Wang, Bo Song, Nadja B. Shoemaker,* and Abigail A. Salyers

Department of Microbiology, University of Illinois, Urbana, Illinois 61801

Received 19 August 2006/Accepted 10 November 2006

Conjugative transposons (CTns) are major contributors to the spread of antibiotic resistance genes among *Bacteroides* species. CTnBST, a newly discovered *Bacteroides* conjugative transposon, carries an erythromycin resistance gene, *ermB*, and previously has been estimated to be about 100 kbp in size. We report here the locations and sequencing of both of its ends. We have also located and sequenced the gene that catalyzes the integration of CTnBST, *intBST*. The integrase gene encodes a 377-amino-acid protein that has the C-terminal R-K-H-R-H-Y motif that is characteristic of members of the tyrosine recombinase family of integrases. DNA sequence comparisons of the ends of CTnBST, the joined ends of the circular intermediate, and the preferred site into which the circular form of CTnBST had integrated revealed that the preferred integration site (*attB1*) contained an 18-bp sequence of identity to the crossover region, *attBST*, on CTnBST. Although this site was used in about one-half of the integration events, sequence analysis of these integration events revealed that both CTnBST and a miniature form of CTnBST (miniBST) integrated into a variety of other sites in the chromosome. All of the sites had two conserved regions, AATCTG and AAAT. These two regions flanked a 2-bp sequence, bp 10 and bp 11 of the 18-bp sequence, that varied in some of the different sites and sometimes in the *attBST* sequences. Our results suggest that CTnBST integrates site selectively and that the crossover appears to occur within a 12-bp region that contains the two regions of conserved sequences.

Conjugative transposons (CTns) are widespread in *Bacteroides* species, but virtually all of the *Bacteroides* CTns so far identified, including the recently described CTn341, have proven to be closely related to a CTn designated CTnDOT (2, 22, 30). CTnDOT is a 65-kbp CTn that carries the tetracycline resistance gene *tetQ* and the erythromycin resistance gene *ermF* (17, 23, 30). Recently, a new CTn, CTnBST, was discovered in a *Bacteroides* isolate. This new CTn carried an *ermB* gene but no tetracycline resistance gene. On the basis of pulsed-field electrophoresis data, CTnBST was estimated to be about 100 kbp in size (9). An interior 13-kbp fragment that contained the *ermB* gene was cloned and sequenced. The sequence of this fragment had no similarity to any part of the sequence of CTnDOT, raising the possibility that CTnBST was not a member of the CTnDOT family. Also, CTnBST differed from CTnDOT in that whereas the transfer frequency of CTnDOT is stimulated 100- to 1,000-fold by tetracycline, CTnBST exhibited constitutive transfer (9).

Although CTnBST was assumed to be a conjugative transposon because it was integrated in the chromosome and was transferred by conjugation, a circular transfer intermediate of the type associated with previously characterized CTns could not be demonstrated because the sequences of its ends were not known. In this report, we describe the cloning and sequencing of the ends of CTnBST. Having the end sequences of CTnBST has allowed us to demonstrate that CTnBST does have a circular transfer intermediate. We also report the iden-

tification of the integrase of CTnBST, which is encoded by *intBST*, and analyze some of the CTnBST integration sites.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. *Bacteroides* sp. strains were grown anaerobically at 37°C in prereduced trypticase-yeast extract-glucose (TYG) broth (10) or agar plates incubated in BBL GasPak jars. *Escherichia coli* strains were grown aerobically at 37°C in Luria-Bertani (LB) medium or on brain heart infusion agar plates. The following antibiotic concentrations were used in this study: for ampicillin, 100 µg/ml; for cefoxitin, 15 µg/ml; for chloramphenicol, 20 µg/ml; for erythromycin, 3 µg/ml; and for rifampin, 10 µg/ml.

DNA manipulation and conjugation procedures. Cloning was performed by methods that have been previously described (18). Total DNA was prepared by a modification of a method described by Saito and Miura (16). Plasmids and fosmids were isolated by the use of a QIAGEN (Valencia, CA) miniprep kit. Restriction enzymes, T4 DNA ligase, and T4 DNA polymerase were obtained from Invitrogen (Carlsbad, CA), New England Biolabs (Beverly, MA), and Fisher Scientific. Calf intestinal alkaline phosphatase was purchased from Promega (Madison, WI), and oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Restriction digests and ligations were performed in accordance with the manufacturers' instructions. The procedures for filter matings between *E. coli* strains and between *E. coli* and *Bacteroides* sp. strains have been previously described (20, 27).

Construction of a CTnBST fosmid library. Approximately 4.5 µg of chromosomal DNA isolated from BT4020, a *Bacteroides thetaiotaomicron* 5482 strain containing CTnBST, was sheared by passing it through a 200-µl small-bore pipette tip 100 to 120 times. The extent of DNA shearing was first visualized on a 20-cm-long 1% agarose gel run at 30 V overnight. The DNA fragments were then end repaired using the end repair enzyme mix provided in the CopyControl fosmid library production kit (Epicenter, Madison, WI). After electrophoresis in a 1% low-melting-point agarose gel, DNA fragments in the range of 35 to 45 kbp were excised, the agarose fragment was melted at 65°C, and the DNA was phenol extracted from the gel. The DNA was ethanol precipitated, washed with 70% alcohol, and suspended in TE8 (0.01 M Tris, 0.001 M EDTA, pH 8). Ligation of the isolated DNA to the pCC1FOS fosmid vector was performed according to the manufacturer's instructions (Epicenter). Fosmid clones were packaged using MaxPlax λ packaging extracts (Epicenter) and transduced into the *E. coli* EPI300-T1^R plating strain. *E. coli* transformants were selected on LB agar containing chloramphenicol. Nine hundred sixty fosmid-containing colonies were

* Corresponding author. Mailing address: Department of Microbiology, 601 S. Goodwin Ave., Rm. B103, University of Illinois, Urbana, IL 61801. Phone: (217) 333-2061. Fax: (217) 244-6697. E-mail: nadjasho@life.uiuc.edu.

[†] Published ahead of print on 22 November 2006.

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Relevant phenotype ^a	Source and/or description (reference)
<i>E. coli</i> strains		
EPI300-T1 ^R		Epicenter
S17-1	RecA Tra _{RP4} Tp ^r Str ^r	Contains the transfer functions of RP4 integrated in the chromosome (25)
EM24NR	RecA Nal ^r Rif ^r Pir ⁻	<i>recA</i> derivative of LE392 (18)
BW19851	RecA Tra _{RP4} ⁺ Tp ^r Str ^r Pir ⁺	<i>E. coli</i> S17-1 with RK6 <i>pir</i> in <i>uidA</i> ; donor strain for Pir-dependent vectors such as pEPE (14)
<i>Bacteroides</i> sp. strains		
BT4001	Rif ^r	Spontaneous rifampin-resistant mutant of <i>Bacteroides thetaiotaomicron</i> 5482
BT4100	Thy ⁻ Tp ^r	Spontaneous thymidine requiring mutant of <i>Bacteroides thetaiotaomicron</i> 5482
WH207	Em ^r	Community <i>Bacteroides uniformis</i> isolate that contains CTnBST (9)
BT4020	Rif ^r Em ^r	BT4001 with a copy of CTnBST from WH207 Thy ⁻ donor in the <i>attB1</i> site (9)
BT4021	Thy ⁻ Tp ^r Em ^r	BT4100 with a copy of CTnBST in the <i>attB1</i> site (9); used as donor of CTnBST to BT4001 recipient
Plasmids		
pEPE	Cm	A Pir-dependent suicide vector for <i>E. coli</i> that is mobilized from BW19851 to Pir ⁻ recipients to test for integration (29)
pGEM-T Easy	Ap ^r	<i>E. coli</i> PCR cloning vector (Promega)
pGERM	(Em ^r Rep ⁻) Ap ^r	<i>Bacteroides</i> shuttle suicide vector that contains <i>ermG</i> and RK2 <i>oriT</i> cloned on pUC19; mobilized by RP4 transfer functions in <i>E. coli</i> S17-1 chromosome to <i>Bacteroides</i> recipients to test for integration (21)
pLYL7 _{oriTRK2}	(Cef ^r Rep ⁺) Ap ^r	<i>Bacteroides-E. coli</i> shuttle vector that replicates in both hosts; contains RK2 <i>oriT</i> , which is also on pGERM; mobilized by RP4 transfer functions in <i>E. coli</i> S17-1 chromosome to <i>E. coli</i> or <i>Bacteroides</i> recipients (12)
Fosmid 4-E8	Cm ^r	Fosmid clone containing an ~40-kbp internal fragment of CTnBST that includes the 13-kbp <i>ermB</i> region previously sequenced (this study)
Fosmid 8-D12	Cm ^r	Fosmid clone containing an ~41-kbp internal fragment of CTnBST that overlaps the right end of fosmid 4-E8 (this study)
Fosmid 10-C7	Cm ^r	Fosmid clone containing an ~37-kbp fragment of CTnBST that overlaps the left end of fosmid 4-E8 and contains the left junction of CTnBST with 582 bp from BT4001 (this study)
Fosmid 6-E4	Cm ^r	Fosmid clone containing an ~33-kbp fragment of CTnBST that overlaps the right end of fosmid 8-D12 and contains <i>intBST</i> and right junction with 861 bp from BT4001 (this study)
pBJE1.7	(Em ^r Int ⁻) Ap ^r	1.7-kbp PCR product containing the CTnBST joined ends and the 5' end of <i>intBST</i> subcloned from pGEM-T Easy into SstI-SphI sites of pGERM (this study)
pBJE2.1	(Em ^r Int ⁺) Ap ^r	2.1-kbp PCR product amplified with primers that contain SphI and SstI cloned into pGERM; contains the CTnBST joined ends and all of <i>intBST</i> ; miniBST that integrates into the chromosome of <i>Bacteroides</i> hosts (this study)
pBJE2.1mut1	(Em ^r Int ⁺) Ap ^r	Same as pBJE2.1 except for a 1-bp change (bp 10) in the 18-bp region of the joined ends of CTnBST (this study)
pBJE2.1mut2	(Em ^r Int ⁺) Ap ^r	Same as pBJE2.1 except for a 2-bp change (bp 10 and bp 11) in the 18-bp region of the joined ends of CTnBST (this study)
pNAW2.1	Cm ^r Int ⁻	2.1-kbp SphI-SstI fragment of pBJE2.1 subcloned from pGEM-T Easy into ApaI-SstI sites of pEPE; contains CTnBST joined ends and <i>intBST</i> but cannot integrate into <i>E. coli</i> hosts (this study)

^a Phenotypes in parentheses are expressed only in *Bacteroides*, and phenotypes not in parentheses are expressed in *E. coli*. Abbreviations used for antibiotic resistance: Ap, ampicillin; Cm, chloramphenicol; Cef, cefoxitin; Em, erythromycin; Nal, nalidixic acid; Rif, rifampin; Str, streptomycin; Tp, trimethoprim. Other phenotype abbreviations: Int⁺ or Int⁻, ability or inability to integrate; Rep⁻, unable to replicate; Thy⁻, requires thymidine.

picked and inoculated in Luria broth supplemented with chloramphenicol in 96-well microtiter plates. This library of 960 fosmid clones represents an estimated 1.4-times genome coverage, based on an average insert size of 40 kbp and a *B. thetaiotaomicron* 5482 genome size of 6.2 Mbp (18).

Screening of the fosmid library for the ends of CTnBST. In order to find the ends of CTnBST, a screening of the fosmid library by colony hybridization was performed. The entire library was inoculated onto 10 GeneScreen (PerkinElmer Life Sciences, Boston, MA) hybridization transfer membranes overlaying brain heart infusion agar plates. After incubation at 37°C overnight, the membranes were placed on a pad of absorbent filter paper soaked with denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30 min. This was followed by a 30-min neutralization step (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.4). The DNA was then cross-

linked to the membranes by use of a Stratelinker 2400 instrument (Stratagene, La Jolla, CA). A 13-kbp segment of internal CTnBST DNA (accession number AY345595) had previously been sequenced and had been estimated to be near the middle of the CTn (9). This segment was subcloned to obtain a 2-kbp PstI-SstII probe and a 3-kbp HindIII probe, which marked the ends of the known sequence (Fig. 1). The probes were labeled with fluorescein-dUTP by use of random primers according to the protocol in the Renaissance kit from PerkinElmer Life Sciences. Hybridization was done at 42°C overnight. Two washes of increasing stringency were used to remove nonspecifically hybridized probes, and the blocking agent from the Renaissance kit was used to prevent nonspecific binding of proteins. Membranes were developed using a chemiluminescent substrate and exposure to film.

FIG. 1. Map of CTnBST derived from overlapping fosmid clones. The CTnBST map at the top was constructed from sequences on the four fosmids (4-E8, 8-D12, 10-C7, and 6-E4) shown below the map. Chromosomal sequences from *Bacteroides thetaiotaomicron* 5482 are represented by dashed lines for both the composite CTnBST map and from the respective fosmids, 10-C7 and 6-E4. The *HindIII*-*HindIII* and *PstI*-*SstI* fragments obtained from a 13-kbp region of CTnBST that had previously been sequenced (9), indicated as filled-in boxes, were the probes used for the original screen of the fosmid library. Both probes hybridized to 4-E8. The sequencing of the ends of the fosmids, the PCR primers (Table 2) used to make probes from the sequences shown as open boxes at the ends of 4-E8 and 8-D12, and the identification of overlapping fosmids from the library by dot blotting are all described in Materials and Methods. The approximate distance between the location of the probe and the end of the respective fosmid or CTnBST is indicated, and the locations of the probes used are also indicated on the composite map of CTnBST. 10-C7 contained the left-end junction between CTnBST and the chromosomal site in *B. thetaiotaomicron* 5482, designated as *attL*, and 6-E4 contained the right junction, *attR*. CTnBST is approximately 100 kbp with the 13-kbp *ermB* region located near the middle of the element.

Fosmid clones that hybridized with the initial probes were induced to high copy number according to the protocol from the CopyControl fosmid library production kit (Epicenter). After DNA isolation, restriction analysis was used to eliminate fosmid clones that appeared to be similar. The ends of the clones were then sequenced using pCC1/pEpiFOS forward sequencing primer and pCC1/pEpiFOS reverse sequencing primer (Epicenter). The sequence was analyzed, and new probes were made with PCR products from the ends of these clones. The screening process was repeated until the fosmid clones that contained the ends of CTnBST were found (Fig. 1).

Since the genome of *B. thetaiotaomicron* strain BT5482 (also called BT4000) has been sequenced (31), fosmid clones that contained an end of CTnBST could be identified by analyzing the sequence obtained with the pCC1/pEpiFOS for-

ward and reverse sequencing primers. A clone that contained an end of CTnBST should have DNA sequences found in the BT5482 genome sequence as well as DNA sequences not found there. The DNA segments carrying the ends of cloned CTnBST segments were sequenced until the CTn ends adjacent to chromosomal sequences were located. The PCR primers used to make the probes for the fosmid library screening are listed sequentially as follows: BST4E8-FR and BST4E8-RR for the right end of fosmid 4-E8, 4E8R; BST4E8FL and BST4E8RL for the left end of fosmid 4-E8, 4E8L; and BST8D12-F and BST8D12-R for the right end of 8-D12, 8D12R (Table 2 and Fig. 1).

Detection of the joined ends of the excised circular form of CTnBST. The primers BST6E4FJ and BST10C7RJ2 (1.43 kb) (Table 2) were designed with the sequence obtained from the ends of CTnBST integrated in BT4020. The primers

TABLE 2. PCR primers used in this study and their functions

Primer	Function(s) ^a	Sequence (5'-3') ^b
BST4E8-FR	Right-end probe from 4-E8, 4E8R	CAC AAG GTT GTC ACG CAT TTT CA
BST4E8-RR	Right-end probe from 4-E8, 4E8R	CAA CAG GAG ACT ATC GTG CGT ACA
BST4E8FL	Left-end probe from 4-E8, 4E8L	CAG CTG TCG GTA CAG CTT CAT AC
BST4E8RL	Left-end probe from 4-E8, 4E8L	GAC TAC GGC ATG TGG ATG ACA GGA
BST8D12F	Right-end probe from 8-D12, 8D12R	GTC ACG ATG ATG CTG TAT GTC CT
BST8D12R	Right-end probe from 8D12, 8D12R	CCG GAT AAC GTG TTT CCT TGC T
BST6E4FJ	F primer for 1.4-kb BJE and for right chromosomal junctions	GTT GGT CAG CTC GAT AAT ACG
BST10C7RJ2	R primer for 1.4-kb BJE and for left chromosomal junctions	TGC TTC CAG GTG AGT CAA CC
BST6E4FJ2	F primer for 1.7-kb DJE	CAA TCC GTT GAC CAG GTT GA
BST10C7RJ1	R primer for 1.7-kb DJE	GAG TCC ACT TTC TGC TGT CT
BJESPHI	F primer for 2.1-kb DJE with <i>SphI</i> site	GCT CTT TAT <u>GCA TGC</u> ₁ TTT CAC GAC CA
BJESSTI	R primer for 2.1-kb DJE with <i>SstI</i> site	GAA GGC CTT TGT <u>GAG CTC</u> ₂ TTG TCA TG
BST12-1-2RJF	R primer used with BST6E4FJ to detect insertions in <i>attB1</i>	GTC CAA AGC TAG GCC AAT TG
BST12-4-1RJF	R primer used with BST6E4FJ to detect insertions in <i>attB3</i>	CGA TCG TCC AAC CTG GAT TT
BST11-1-2 RJF	R primer used with BST6E4FJ to detect insertions in <i>attB_{mut1}</i>	GCT CAA GTA GCC CTT TGT GT
BST2-1-1 RJF2	R primer used with BST6E4FJ to detect insertions in <i>attB_{mut2a}</i>	GGT TGA CTT GTG CAA CCA TG
BST2-3-3 RJF2	R primer used with BST6E4FJ to detect insertions in <i>attB_{mut2b}</i>	CCG TCA GAT TGA GAG AAG GA
BST2-4-3 RJF2	R primer used with BST6E4FJ to detect insertions in <i>attB_{mut2c}</i>	GGT ATG AAC GGC AGG TCA AA
F743	F primer for BJE, 410 bp	TGC GAA TTC CTG GAT ATA TGA
R1153	R primer for BJE, 410 bp	GGT CGA TAT TGA GAA ACA AAC GA
F1418	F primer for 423-bp integrase probe of right-end junctions	TAT TAT CGA GCT GAC CAA CGG CGA
R1841	R primer for 423-bp integrase probe of right-end junctions	CAA TGC ACG CAC CTC TTC AAT GGT

^a F, forward; R, reverse.

^b Underlined sequences are restriction sites inserted into the primers. 1, *SphI*; 2, *SstI*.

			*		*
IntBST	224-	VKGAYLFSCFCGLRISDIIKLQWK--DVFVDRGQYRLSVSMQ	KTKEP		
IntTn5520	245-	VRDLFVFSVFTGLAYS DVKNLTDR LQTFFD-GNLWII TRRK	KTNTTE		
IntDOT	238-	YRDLYLFCFAFTGLSFSDMRNLT DENIRTYFD-EHEWININRQ	KTGVV		
IntCTn3Bf	225-	VKSAYLFSCFCGLRISDIIKLQWK--DVFIDNGQYRLAVAMQ	KTKEP		
IntSGI1	226-	LKRAFLFSCCTTGLRWSDIQKLTKWEIEEFQD-GHYRIIFKQAK	LLNA		
Int-lambda	199-	LRLAMELAVVTGQRVGDLCMKWS-----DIVDGYLYVEQS	KTGVK		
			*	*	*
IntBST	321-	FHTARHTFAT-MMLTLGADLYTTSKLLGHADV KMTQVYAKIINQKK-	<12> -377		
IntTn5520	343-	YHVARHTNATTVLLSHGVPIETVSRLLGHNTNIKTQIYAKITAKIS	<21> -410		
IntDOT	344-	WHQSRHTAATTVFLSNGVPIETVSSMLGHKSIKTTQIYAKITKEKLN-	<21> -411		
IntCTn3Bf	319-	FHTARHTFAT-MMLTLGADLYTVSKLLGHTSVKMTQVYAKIVNKKK-	<12> -376		
IntSGI1	326-	FHVGRHTFAV-AQLNRGVDIYSLSRLLGHSELRTTEIYADILESRR-	<34> -405		
Int-lambda	307-	FHELRLSLSAR--LYEKQISDKFAQHLLGHKSDTMASQYR-----	<13> -356		

FIG. 2. ClustalW alignment of the C-terminal amino acid sequence of IntBST and with those of some known tyrosine recombinases. The closest homolog (75% identity over the entire protein) in the BLASTP search to IntBST was the putative integrase from CTn3Bf designated IntCTn3 (BF2788, BAD49538) of *Bacteroides fragilis* YCH46 (11). The integrase of the *Salmonella* genomic island (AAG03003), designated IntSGI1 (8), was the closest characterized integrase, at 33% identity over the entire protein. The percent identities to the other integrases in the alignment are confined to smaller areas mostly in the C-terminal region of the proteins shown in the figure. IntTn5520 (31% C identity) is the transposase encoded by *bipH* (AAB53787) carried on the *Bacteroides* MTn Tn5520 (28), and IntDOT (27% identity) is the integrase (CAC47921) of the *Bacteroides* conjugative transposon CTnDOT (6), which is nearly identical (>97%) to the integrases of CTnERL and CTn341 (AAS83518) (2). Lambda integrase (P03700) (15) is the well-characterized tyrosine recombinase from bacteriophage λ in *E. coli*, and it has only 23% identity over a portion of the protein. The amino acid number for each protein is indicated at the beginning of each line of the two C-terminal alignment regions, and the total sizes of the integrases are indicated at the ends of the second set of C-terminal alignments. The conserved amino acids found in most of the tyrosine recombinases (RKHRHY) are indicated in bold with asterisks. The locations of the canonical amino acids were found at positions R237, K264, H322, R325, H348, and Y357 of IntBST. Note that IntDOT and IntTn5520 have A and S, respectively, instead of the first conserved arginine (R), R237, of IntBST.

were directed out of the ends of the CTnBST element so that a PCR product could not form unless the integrated element had excised from the host chromosome and formed a circular intermediate. Template DNA was isolated from WH207 and BT4020 by use of either chromosomal or plasmid preparation techniques. About 10 to 100 ng of template DNA was combined with 40 pmol of each primer in 100 μ l of reaction buffer (Invitrogen PCR buffer, 1.5 mM MgCl₂, 0.2 mM deoxynucleoside triphosphate mixture) and amplified with *Taq* polymerase. The cycling conditions were as follows: 95°C for 5 min and 25 cycles of 94°C for 1 min, 51°C for 1 min, and 72°C for 2 min, followed by a final extension for 5 min at 72°C. The PCR products were gel purified and cloned into pGEM-T Easy (Promega) for sequencing.

Construction of the CTnBST minielement. To determine whether the gene that encoded a putative tyrosine recombinase-type integrase, *intBST* (Fig. 2), was necessary and sufficient for integration, we used PCR to amplify two different segments carrying the CTnBST joined ends (BJE) by use of DNA from BT4020 cells. The smaller BJE product (primers BST6E4FJ2 and BST10C7RJ1) was 1.7 kbp and lacked the C-terminus-encoding region of *intBST*, whereas the larger 2.1-kbp product (primers BJESPH and BJESST1) contained the entirety of *intBST* (Fig. 3). The smaller product was first cloned into the pGEM-T Easy PCR product cloning vector. The BJE region was then isolated on a 1.7-kbp SphI-SstI fragment and cloned into pGERM, a *Bacteroides* suicide vector based on pUC19 (21). pGERM contains the *oriT* of RK2 and the *Bacteroides* selectable marker *ermG*.

The larger product was PCR amplified with SphI and SstI sites in the primers, which allowed us to clone the fragment directly into pGERM (Fig. 3). The pGERM:BJE miniBST elements (pBJE1.7, pBJE2.1) were transformed into the *E. coli* strain S17-1 and mobilized into BT4001 to test for their ability to integrate into the *Bacteroides* chromosome. A third miniBST element, pNAW2.1, was constructed to see whether the joined ends and *intBST* were sufficient for integration into *E. coli* recipients. The 2.1-kbp SphI-SstI fragment containing *intBST* and the joined ends of CTnBST was cloned into pGEM-T Easy before being subcloned as an ApaI-SstI fragment into pEPE (29). This vector was transformed into the Pir⁺ BW19851 strain and mobilized into EM24NR, a Pir⁻ recipient, to select for possible integration as described by Cheng et al. (7).

During amplification and sequencing of several different excision events in

BT4020 involving BJE, we found two miniBST sequences that differed in sequence from the most commonly obtained BJE sequence. These two BJE miniBST variants, called mut1 and mut2, were identical to each other and the main or "wild-type" miniBST except for a 1- or 2-bp sequence that lay in the middle (at bp 10 and 11) of the 18-bp *attBST* sequence region. Accordingly, we used minielements containing these two variants, *attBST*_{mut1} and *attBST*_{mut2}, to determine if these mutants could integrate at the same frequencies as the wild type and if they could provide some insight as to where the cut sites for the integration event might occur.

Southern blot analysis of the integration specificity of CTnBST and the minielement. To determine whether CTnBST integrated into a single site or multiple sites in the *B. thetaiotaomicron* chromosome, we isolated chromosomal DNA from independent integration events as described previously (16). DNA from each isolate was digested with PstI and analyzed by Southern blotting as described in the work of Sambrook et al. (18). Similarly, DNA isolated from strains which contained insertion(s) of the miniBST constructs was digested with HindIII, selected because of sites contained on the shuttle vector. The probe in both cases was a 423-bp PCR fragment (primers F1418 and R1841 in Table 2) that lay within the *intBST* open reading frame (ORF), so that each insertion event would give only one cross-hybridizing restriction fragment. Multiple bands indicate multiple insertions and/or the excised joined ends (BJE) of the element.

Site specificity was assessed in two ways. First, Southern blot analysis was done with a 423-bp probe that contained part of *intBST* to determine whether CTnBST and the miniBST both integrated into different sites as was done previously with CTnDOT and the miniDOT construct (6). The insertion into *attB1* should result in a 5.3-kbp PstI right junction band for the CTnBST insertions and a 3.5-kbp HindIII fragment for the miniBST insertions indicated as *attB1* in Fig. 4A and B. Apparent insertions of both elements into *attB1* were confirmed by PCR using primers to produce the right-end junction fragment (primers BST6E4FJ and BST12-1-2RJF). Lack of a PCR product indicated integration into another site (Fig. 4A and B).

Plasmid rescue technique used to isolate the end sequences of integrated minielements. A plasmid rescue technique was utilized to determine the DNA sequences of the locations of some of the target sites into which the miniBST had integrated. In the plasmid rescue experiments, genomic DNAs from independent

FIG. 3. Construction of the CTnBST miniBST elements. CTnBST integrated into its preferred site, *attB1*, is flanked by an 18-bp sequence, ATAAATCTGGTAAATTTA. When the element excises, this 18-bp sequence is found between the two ends of the element within the *attBST*. PCR primers were used to produce two products from the joined ends of the element containing the region involved in integration, *attBST*, and part or all of the gene believed to encode the CTnBST integrase, *intBST*. The minielements pBJE1.7 and pBJE2.1 were constructed in the *Bacteroides* insertional vector pGERM (21). The smaller miniBST, pBJE1.7, was constructed using the 1.7-kbp PCR product from the excised circular form of CTnBST that included *attBST* and a truncated version of *intBST*. The larger miniBST, pDJE2.1, was constructed using a 2.1-kbp PCR fragment that included the *attBST* region and the entire *intBST* gene. The same 2.1-kbp PCR fragment, including the entirety of *intBST*, was cloned into pEPE, a Pir-dependent vector that contains *cat* (29). The resulting miniBST, pNAW2.1, was used to test for integration in *E. coli* hosts that lack the *pir* gene required for replication of the vector.

transconjugants from *E. coli*-to-*Bacteroides* matings were digested with either BspEI or SstI, which left the majority of pBJE2.1 intact and included the left junction with a portion of adjacent chromosomal DNA (Fig. 5). Southern blot analysis was performed using an internal portion of *ermG*, the marker on the plasmid, as a probe to determine the approximate size of the rescued fragment for each digest. The genomic DNA from the transconjugants was digested with the enzyme that appeared likely to have a site near an end, and ligation conditions that favored monomeric circularization were used. The ligation mixture was transformed into *E. coli* DH5 α MCR, and selecting for ampicillin resistance identified transformants that received the intact vector. The vectors were isolated, and the chromosomal DNA adjacent to the minielement was sequenced using BST10CRJ2 as a primer (Table 2).

This procedure gave the DNA sequence on the left side, i.e., the nonintegrase end, of the integrated element. The sequence of the integration site (*attB*) was deduced by comparing this sequence with the sequence of the joined ends of the circular form and the genome sequence of BT4001, accession number NC 004663 (31). The DNA sequence of the right side of the integrated element was obtained by using BST6E4FJ, a PCR primer seated inside the right integrase end of CTnBST and a PCR primer seated in the adjacent chromosomal sequence to amplify the region. The primers that hybridized with chromosomal DNA were designed using the genome sequence information for the region that contained the other junction. The primers used in this study are as follows: BST12-1-2RJF for *attB1*, BST12-4-1RJF for *attB3*, BST11-1-2RJF for *attB_{mut1}*, BST2-1-1RJF2 for *attB_{mut2a}*, BST2-3-3RJF2 for *attB_{mut2b}*, and BST2-4-3 RJF2 for *attB_{mut2c}*. The PCR amplicons were then sequenced to provide the right junctions of the integrated miniBST elements.

DNA sequencing. Sequencing reactions were performed at the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois, Urbana, IL. DNA and amino acid sequences were analyzed using the NCBI BLAST server (1).

Nucleotide sequence accession number. The nucleotide sequence of the joined ends with *intBST* contained on the 2.1-kbp miniBST element has been submitted to the GenBank nucleotide sequence database; the GenBank accession number is EF067916.

RESULTS

Cloning and sequencing of the ends of CTnBST. Figure 1 provides a schematic map of CTnBST derived from an analysis of four overlapping fosmid clones that covered the entire 100-kbp element and its junctions (*attL* and *attR*) with the chromosomal site in *B. thetaiotaomicron* 5482. The fosmid library was initially screened with two probes made from the ends of the *ermB*-containing 13-kbp region that had been cloned and sequenced previously (9). Based on Southern blot analysis, this region was estimated to be located somewhere near the middle of CTnBST, and this estimate proved to be correct. A fosmid clone that hybridized to both of these probes was 4-E8. Approximately 500- to 700-bp regions of both ends of the 4-E8

FIG. 4. Southern blot analysis of miniBST and CTnBST insertions in BT4001. (A) Southern blot of miniBST insertions in *Bacteroides thetaiotaomicron* BT4001. The miniBST (pBJE2.1) was transferred from *E. coli* S17-1 to BT4001 by conjugation. DNA was extracted from 13 independent isolates (1 each from separate mating filters to avoid siblings), digested with HindIII, and run on an agarose gel as described in Materials and Methods. The Southern blot of the gel was probed with a labeled 423-bp region of the *intBST* which detected right-end junctions. Multiple bands indicate multiple insertions in lanes 2, 5, 6, 7, 11, and 13. The 3.8-kb HindIII fragment, expected for a hybridizing band due to insertion into the preferred target site (labeled *attB1*) for the miniBST, is indicated by an arrow at the right. Insertions into *attB1* were confirmed by PCR amplification of the right junction using the primers for the right end of the element and the right side of the chromosomal *attB1* sequence (Table 2 and Materials and Methods). The HindIII fragments of λ are in the λ lane, and the sizes of the bands in kbp are indicated on the left. (B) Southern blot of CTnBST insertions in BT4001. CTnBST was transferred by conjugation from BT4021 to BT4001. In the experiment shown, DNA was extracted from 11 independent isolates (from separate mating filters to avoid siblings) and was digested with PstI. The location and sizes in kbp of the λ HindIII standards are shown to the left of the panel. The Southern blot of the gel was probed with the same probe used in panel A, which detected right-end junctions and also the excised CTnBST joined-end (BJE) fragments. The expected size of a PstI fragment containing the BJE derived from the end sequences of CTnBST is 2.3 kbp and is indicated on the right. The expected size of the PstI fragment containing the right junction of CTnBST integrated into the preferred site of *B. thetaiotaomicron* 5482, *attB1*, is 5 kb and is also indicated. PCR amplification of the right junction using *attB1* and right-end CTnBST primers (Table 2) was used to verify insertion into *attB1* in the isolates which had the 5-kbp band. Strains that lacked the 5-kbp band did not produce a PCR amplicon (data not shown). Lanes 2 and 8 have two right-junction bands indicating two elements integrated in the strain.

clone were sequenced. BLAST searches revealed that the sequences of the ends of the 4-E8 clone did not match any sequences in the genome sequence of the host, *Bacteroides thetaiotaomicron* 5482. This finding indicated that the cloned region was an internal region of CTnBST.

New probes, 4E8R and 4E8L, were then constructed from

FIG. 5. Plasmid rescue based on the miniBST used to obtain the left junctions of the integrated elements. Chromosomal DNA containing the integrated miniBST, pBJE2.1, was digested with enzymes that cut near the integrase of the integrated element within the vector and adjacent to the left end of the integrated plasmid (*SstI* or *BspEI*). The digested DNA was ligated and used to transform *E. coli* with selection for the marker on the miniBST (Ap^R). The vector isolated from the transformants contained the left junction of the integrated element and its chromosomal site.

the ends of the 4-E8 clone and used to rescreen the fosmid library. The 4E8R hybridized with fosmid 8-D12, whereas 4E8L hybridized with fosmid 10-C7. Sequence analysis of 10-C7 showed that it contained an end of the CTn left junction, called *attL*, plus about 1 kbp of *B. thetaiotaomicron* chromosomal DNA. Fosmid clone 8-D12, however, consisted entirely of CTnBST DNA. A new probe was made from the right end of 8-D12, 8D12R, and the library was rescreened. 8D12R hybridized to fosmid clone 6-E4. Sequence analysis revealed that 6-E4 contained chromosomal sequences. The right junction of CTnBST, called *attR*, was located within 1 kbp of the end of 6-E4.

Once the left and right junctions of CTnBST, *attL* and *attR*, respectively, were identified, primer walking was used to look for ORFs near the ends of the element. Because the integrase genes of bacteriophages and CTns are generally located near an end, we expected to find an ORF that was similar to known integrases at one of the ends. The left end of CTnBST did not contain any ORFs of significant length within the first 1 kbp from the end. The sequence obtained from the right end, however, contained a 1,131-bp ORF, tentatively designated *intBST*, whose first possible start codon was 213 bp from the right junction (Fig. 1).

According to BLASTP search results, the deduced amino acid sequence of this ORF exhibited a low identity to known integrases of the tyrosine recombinase family. The most closely related known integrase, with 33% identity, was from the *Salmonella* genomic island SGI1 (8). However, the highest-identity matches (75% and 42% identity) were to predicted amino acid sequences from two ORFs, BF2788 (BAD49538) and BF1677 (BAD48425), found in the newly released *Bacteroides fragilis* YCH46 genome sequence (11). Although these ORFs had been annotated as integrase genes based on sequence

comparisons, no characterization has been reported for them. A comparison of IntBST with C-terminal sequences of known tyrosine recombinases revealed that IntBST contains the signature C-terminal amino acids R-K-H-R-H-Y, which are conserved in members of the lambda family of site-specific integrases (15) and integrases of most of the CTNs characterized to date (Fig. 2). The locations of the canonical amino acids were found at positions R237, K264, H322, R325, H348, and Y357 of IntBST.

By searching the chromosomal sequences flanking CTnBST in BT4020 against the *B. thetaiotaomicron* 5482 genome sequence, we identified a possible 18-bp target site. This 18-bp sequence was found at each end of the integrated form of the element, indicating that these direct repeats were a duplication of the target site. To determine whether CTnBST had a circular intermediate, primers BST6E4FJ and BST10C7RJ2, which were directed outward from the integrated ends of CTnBST, were used to amplify a 1.4-kbp product that could be made only if the element had excised and formed a closed circular intermediate. The PCR product was cloned and sequenced, revealing that the excised joined-end sequence of CTnBST contained an 18-bp sequence that was identical to an 18-bp sequence found in the site in the genome sequence where CTnBST had integrated in BT4020.

Minimal region required for integration into the *Bacteroides* chromosome includes the integrase gene (*intBST*) and the joined-end sequence (*attBST*). To determine the minimal region required for the integration of CTnBST and to obtain a smaller version of CTnBST that would be more manageable for analyzing sequences of integration events, we constructed a form of CTnBST (miniBST) that is designated pBJE2.1 in Fig. 3. pBJE2.1 contained the entire *intBST* gene and the joined ends of CTnBST (*attBST*), as well as a replication origin that allowed it to replicate in *E. coli* and an RK2 *oriT* that allowed it to be mobilized from an *E. coli* donor into *B. thetaiotaomicron* strain BT4001. pBJE2.1 was mobilized by the RP4 transfer functions in the *E. coli* donor S17-1 (Table 1) into BT4001 recipients with the selection for erythromycin resistance. The miniBST transferred to and integrated into the chromosome at a frequency of about 10^{-2} to 10^{-3} integrants per recipient. Considering that the frequency of transfer of a replicating vector with the same transfer origin, pLYL7_{oriTRK2}, from S17-1 to BT4001 is 10^{-3} to 10^{-4} per recipient (12), this result indicates that nearly every minielement that is transferred into the *Bacteroides* BT4001 recipient integrated into the host chromosome.

To confirm that the gene identified as *intBST* was essential for integration, we also constructed pBJE1.7 (Fig. 3), which contained the joined ends of CTnBST plus a truncated integrase gene. This truncated miniBST element was also mobilized into BT4001. No transconjugants ($<10^{-9}$ integrants per recipient) were obtained. The plasmid, however, transferred from *E. coli* to *E. coli* at a high frequency (10^{-1}), indicating that no changes in the mobilization region of the vector had occurred during cloning. This result supports our contention that IntBST, together with the joined ends which contain the *attBST* sequences, is necessary and sufficient for integration in BT4001.

In the process of making the miniBST, a process during which we sequenced several PCR clones obtained from differ-

ent BT4020 colonies, two of the cloned regions proved to have either a change in bp 10 (G to A) or a 2-bp change in bp 10 and bp 11 (GT to CC) within the 18-bp region that spanned the joined ends. When the *attBST* regions of these variant minielements, labeled *attBST*_{mut1} and *attBST*_{mut2}, respectively, were used in integration assays, they had an integration frequency of 10^{-3} to 10^{-2} , the same as that of the minielement with the 18-bp region that was identical to the 18-bp sequence in the chromosomal *attB1* site in BT4020. Thus, complete identity throughout the 18-bp sequence was not required for integration.

Sites involved in integration events are located in different parts of the chromosome. To assess whether insertion of the minielement was site specific and to determine if multiple insertions could occur within the same strain, we isolated a number of independent transconjugants into which the wild-type minielement had integrated. The donor was *E. coli* S17-1 containing the miniBST pDJE2.1, and the recipient was BT4001. DNA preparations from isolates taken from separate mating filters were probed with a probe that hybridized to only one end of the element (see Materials and Methods). Thus, a single insertion event should have yielded only one band. As is evident from Fig. 4A, 6 of the 13 transconjugants contained more than one minielement insertion. In one strain into which the miniBST was transferred, there were four bands (Fig. 4A, lane 5). Also, not all of the insertions occurred in the initially identified insertion site in BT4020, *attB1*, although that site seemed to be preferred. An insertion into *attB1* would produce a 3.5-kbp HindIII band indicated on the right as *attB1*. This same integration pattern into *attB1* was observed in the case of intact CTnBST (Fig. 4B). CTnBST was transferred from BT4021 to BT4001 at a frequency of 10^{-6} transconjugants per recipient. For CTnBST, integration into *attB1* should produce the 5.3-kbp PstI fragment indicated on the right blot. Two or more insertions were seen in 3 of the 11 recipients for CTnBST. The DNAs from transconjugants containing the miniBST or the CTnBST were also tested by PCR to confirm that elements that produced the expected right-junction fragments, indicating that insertion into *attB1* produced a PCR right-junction product and that those that lacked the junction bands did not produce a PCR product (see Materials and Methods). All the transconjugants shown in Fig. 4A and B that produced the hybridizing junction band for *attB1* insertion also produced a right-junction PCR amplicon for *attB1*. The ones lacking the band labeled *attB1* did not produce a PCR junction amplicon, but they did produce the *attB1* PCR product (data not shown). Thus, although *attB1* is a preferred site, it is not the only insertion site in the recipient BT4001. This observation indicated that the *attB* sequence required for integration was smaller than 18 bp or that complete identity to the incoming *attBST* within this sequence was not required.

Excision of CTnBST could be detected, as shown in Fig. 4B. The joined-end product (BJE) should produce a 2.4-kbp PstI fragment (location is indicated on the right of the blot) that would hybridize to the right-end probe used for the Southern blot. A band of this size was observed in all lanes, although it is weak in lane 8. All of these samples produced a PCR product with primers used to detect the joined ends (data not shown). Thus, excision of CTnBST occurred at a level high enough for the joined ends of the excised circular form (BJE) to be de-

FIG. 6. Sequences of *attB* sites used by miniBST constructs of CTnBST and the consensus sequence of the *attB* sites relative to *attBST*. These sequences were derived from a combination of left-end sequences obtained with the plasmid rescue technique and right-end sequences obtained by PCR amplification based on primers designed from the genome sequence of BT4001 and the right-end sequences of CTnBST (see Materials and Methods). (A) Six panels showing the 18-bp sequences of an integrating miniBST and the corresponding 18-bp sequences of the chromosomal *attB* site. At the top of each of the six panels is the sequence of the 18-bp region that spans the joined ends of the miniBST (*attBST*) involved. Panels 1 and 2 are integration events with the miniBST which had the wild-type 18-bp sequence in *attBST*. Panel 3 is the miniBST derivative *attBST_{mut1}*, which has one change at bp 10, indicated by a lowercase "a," and panels 4 to 6 are examples for the miniBST *attBST_{mut2}*, which has base pair changes at bp 10 and bp 11, indicated as "cc." Immediately below each *attBST* sequence is the sequence of the site on the chromosome (*attB*) into which that miniBST integrated. One preferred site (*attB1* in panel 1) and a secondary site (*attB3* in panel 2) for the wild-type minielement are shown. One site for *attB_{mut1}* into *attBST_{mut1}* is in panel 3, and three target sites (*attB_{mut2a}*, *attB_{mut2b}*, and *attB_{mut2c}*) are shown for *attBST_{mut2}* in panels 4 to 6. For each example, the regions of identity to the wild-type 18-bp sequence for all of the *attBST*s and the *attBs* are indicated in capital letters, and the differences in the sequences are in lowercase letters. The regions where staggered cuts possibly could occur for each integration event are shown as bold and underlined bases and are indicated with X's between the *attBST* and the *attB* sites. Below each pair of *attBST* and *attB* sequences are given the left and right junctions of the integrated miniBST. The left junction of each insertion was cloned by plasmid rescue (Fig. 5) and sequenced. The right junction was obtained by PCR using a primer designed from the identified chromosomal site and the right end of the miniBST (see Materials and Methods). (B) Sequences of the three *attBST*s and the six *attB* sites are aligned. The wild-type sequence of the CTnBST for the 18 bp in the *attBST* region is shown at the top and the joined-end sequences of the three minielements that were either identical for the wild-type miniCTnBST or had 1-bp (bp 10) or 2-bp (bp 10 and bp 11) changes for the *attBST_{mut1}* and *attBST_{mut2}*, respectively, are shown below. The six *attB* sequences in panel A are shown with spaces between bp 9 and bp 10 and between bp 11 and bp 12 to emphasize the base pairs that are sometimes changed in the *attBST*s or are different in the *attBs*. The sequences where staggered cuts could occur, which would explain the junctions observed for each of the insertions shown in panel A, are underlined boldface capital letters for each *attB* site. The consensus sequence at the bottom shows identical sequences in the *attBST* and *attB* sites as boldface capital letters. The boxed and shaded regions indicate where staggered cuts flanking bp 10 and bp 11 could be made in the *attBST*s of the miniBSTs and the *attBs* to give the resulting junctions for all six examples shown in panel A. The arrows above the wild-type *attBST* sequence in panel B indicate a set of 5-bp inverted repeats that flank the shaded regions.

tectable by Southern hybridization. No circular form was detected on the Southern blot in the case of the minielement in Fig. 4A, which would have been seen as a 2.4-kbp HindIII band. However, if high concentrations (>500 ng) of template DNA were used in the PCRs instead of the 10 to 100 ng usually used (see Materials and Methods), very faint PCR products

could sometimes be detected (data not shown). This result demonstrates that the minielement may be able to excise using IntBST alone but at a frequency considerably lower than that of the full-length CTnBST and that the excision product is too weak to be detected by Southern blotting.

From the results shown in Fig. 4, it appeared that both

miniBST and intact CTnBST used the *attB1* site with about equal frequencies. This is the result expected if *intBST* and *attBST* contain all the information necessary to determine integration specificity. To test this hypothesis further, we used PCR analysis to compare the frequencies with which miniBST and intact CTnBST integrated into the preferred site in separate experiments. This was done by determining whether the *attB1* site was still intact or whether a right-junction PCR product could be obtained. In three independent matings in addition to the experiment shown in Fig. 4B, CTnBST integrated into the preferred site with frequencies of 15/28, 16/26, and 22/30. These frequencies were similar to those observed for the miniBST which integrated into *attB1* with frequencies of 8/13 and 7/15 from two independent matings in addition to the experiment shown in Fig. 4A. Thus, the intact CTnBST element appears to have an amount of secondary integration sites similar to that seen with the miniBST, as determined both by Southern blotting verified by PCR and by PCR analysis alone.

Sequence characteristics of CTnBST integration events. Having the miniBST vector allowed us to obtain sequences from the left ends of independent insertions, using the plasmid rescue technique shown in Fig. 5. The left-end sequence was obtained by sequencing the cloned left junction by use of the plasmid rescue technique. The right-end junction sequence was obtained by designing a primer from the expected right-side chromosomal sequence, which was deduced from the genome sequence and a primer seated in the right end (integrase end) of the integrated element to PCR amplify the right junction (see Materials and Methods). Results of this analysis for six integration events are shown in Fig. 6A. Four of the examples shown were obtained using the miniBST constructs containing *attBST* with bp 10 or bp 10-and-bp 11 differences in the middle of the 18-bp region. The integration frequencies for these variant or "mutant" miniBST vectors and their site specificities for *attB1* were the same as those observed for the wild-type miniBST (data not shown). The mutant miniBST insertions were analyzed for integration into sites other than *attB1* in an attempt to define where within the 18-bp region integration occurred. We also hoped that these mutant miniBST elements would help define the conserved sequences that were required for efficient integration by varying the internal bp 10/11 residues. Comparison of the sequences of the input molecules (miniBST and the chromosomal *attB* target involved in the insertion) and the end or junction sequences of the resulting integrated elements revealed a 12-bp subset of the 18 bp that contained the region(s) in which the crossovers appeared to have occurred (Fig. 6A, panels 2 to 5).

The six insertion sites are aligned in Fig. 6B relative to the *attBST* sequences of CTnBST excised from *attB1* in BT4020 and the three miniBSTs to further clarify the region within which the crossover occurred. The regions of identity flank the two base pairs that appear to be variable both in the target sites and in the two variant *attBST* sequences. The two conserved regions are highlighted in Fig. 6B. The consensus sequence for the *attBST* and the *attB* sequences is AATCTG nn AAAT ("n" indicates any base). Staggered cleavage sites located at the ends or within these two regions of identity flanking the variable bp 10 and bp 11 indicated in Fig. 6A could explain the sequences seen for the junctions in Fig. 6A. There is a short inverted repeat sequence in the *attBST* sequences, indicated at

the top of Fig. 6B, that end within the highlighted regions, but the importance of these sequences, if any, is not yet known.

The minielement does not integrate in *E. coli*. A *pir*-dependent variant of the miniBST, pNAW2.1, was constructed to test the ability of a miniBST to integrate into the *E. coli* chromosome. Like pBJE2.1, this pEPE-based minielement contained the 2.1-kbp segment of DNA from the circular form of CTnBST that included the joined ends and the entire *intBST*. pNAW2.1 was mobilized from BW19851 (Table 1) into EM24NR (*pir* minus) with selection for Cm^r. No transconjugants were obtained ($<10^{-9}$ transconjugants per recipient). This vector transferred into a *pir*⁺ recipient at a frequency of 10^{-1} . In this miniBST, *intBST* was under the control of its *Bacteroides* promoter, so the failure of the minielement to integrate in *E. coli* could well have been due to the failure of the *Bacteroides* promoter to function in *E. coli*, or it could be due to the lack of an adequate integration site. There is no 18-bp preferred site in the sequence of the *E. coli* chromosome; however, as seen in Fig. 6, complete identity is not necessarily required.

DISCUSSION

An important part of the currently accepted definition of a CTn is that the integrated element excises from the chromosome to form a circular intermediate. We have now shown that there is in fact a circular intermediate that is formed when CTnBST excises. Another feature shared by many CTns is that their integrases are members of the tyrosine recombinase family, although they are only distantly related (usually $<23\%$ identity) to such well-studied tyrosine recombinases as the integrase of bacteriophage λ . IntBST shares this feature and has the carboxy-terminal amino acid signature (R-K-H-R-H-Y) that typifies most members of the tyrosine recombinase family (Fig. 2). In contrast to CTnDOT and the *Bacteroides* mobilizable transposon (MTn) Tn5520, which have a Ser(S) and an Ala(A), respectively, instead of the first Arg(R) of the signature amino acids (6, 28), the CTnBST integrase has the more commonly seen Arg(R) residue.

IntBST proved to be related to a variety of proteins that have been called integrases on the basis of sequence and motif comparisons, but only 1 of the 10 closest relatives of IntBST has actually been shown to be part of a mobile element, the *Salmonella* genomic island 1 (SGI1) (8). SGI1 was recently shown to be mobilizable when transfer functions are provided in *trans* by the conjugative IncC plasmid R55 (8). This feature puts SGI1 into the same category as the *Bacteroides* MTns, including the *Bacteroides* NBUs, Tn5520, and other mobilizable transposons. MTns are elements smaller than CTns that require a self-transmissible element such as a plasmid or CTn to provide transfer functions in *trans* (30). The closest known *Bacteroides* integrase homologs are on the mobilizable transposons, including Tn5520 (28), Tn4555 (26), and cLV25 (3), followed by the integrases of the CTnDOT group. However, the closest homologs of IntBST found in the database searches were putative uncharacterized integrases or transposases that were found in the genome sequences of *Bacteroides*, *Porphyromonas*, and *Prevotella* spp. It is notable that so far every one of the genome sequences available from this related group of genera has multiple homologs of IntBST attributed to bacteriophage or integrated transmissible elements. These putative

integrases are not always associated with antibiotic resistance genes in the strains as they are in the cases of CTnDOT and CTnBST. BF YCH46, for example, has *tetQ* on a CTnERL-type element (CTn1Bf), but this putative CTn is not located near either of the close IntBST homologs, BF2788 (assigned to a suspected CTn [CTn3Bf]) or BF1677 (annotated as a phage integrase), in this strain (11).

CTnBST appeared at first to resemble SGI1 in that it had integrated into an 18-bp site, but we have shown that CTnBST differs from SGI1 in that it can integrate into many chromosomal locations, both between *orfs* and within *orfs*. *attB1*, for example, is between two genes encoding small hypothetical proteins (BT4040 and BT4041). An analysis of the sequences of several integration sites for CTnBST revealed that the integration of this CTn is more site specific than at first appeared to be the case from the Southern analysis. The six examples share a region that has 10/12 bases of identity. Apparently CTnBST, like phage λ , has secondary sites that are used less frequently than the primary site (4), but in the case of CTnBST, the secondary sites are used much more frequently. The heterology at bp 10 and bp 11 may also be similar to that for coupling sequences where mismatches are seen for CTnDOT and Tn916 integration and excision (6, 19).

If the *attB* sequence used for searching the *B. thetaiotaomicron* chromosomal sequence is just the internal 18-bp region, the following possibilities are found: one site with 18/18 bp of identity, which is *attB1*, and four sites with 16/18 bp of identity. One of these 16/18 sites is the *attB_{mut1}* site shown in Fig. 6A. The sites other than *attB1* do not appear to cause a hot spot for integration. Thus, what we interpreted at first as a lack of integration specificity by CTnBST by Southern blot analysis proved to be due to the fact that there are hundreds of copies of the 12-bp consensus sequence in the *B. thetaiotaomicron* chromosome shown in Fig. 6B. In other words, there are four copies of the 18-bp sequence in the *B. thetaiotaomicron* chromosome with only 2 bp of sequence difference, but there are many copies of the 12-bp consensus with the 2-bp region of variability at bp 10 and bp 11 in the *B. thetaiotaomicron* chromosome.

The closest characterized element, SGI1, integrates into the 3' end of a *thdF* gene just as the *Bacteroides* mobilizable transposons NBU1 and NBU2 integrate into the 3' end of tRNA genes (24, 29). However, CTnBST does not integrate into the 3' end of any known gene. As previously mentioned, the preferred site, *attB1*, is between two genes. The secondary sites are sometimes within a gene but none duplicate either the 5' or the 3' end of the wild-type gene. The discoverers of SGI1 have not yet determined whether the integration of this element is site specific or, like CTnBST, has a variety of integration sites. We do not know if CTnBST is site specific in some strains or species of *Bacteroides*, for example WH207, but not in other strains, such as *B. thetaiotaomicron* 5482.

The six sites used by CTnBST and the miniBST derivatives in *Bacteroides thetaiotaomicron* shown in Fig. 6B all share a core consensus sequence, AATCTG nn AAAT. An odd feature of these different sites is that integration appears to occur, probably by staggered cuts, in two regions flanking a 2-bp sequence that can vary. Variations in this 2-bp sequence in the *attBST* region cloned in the variant mini-BSTs, *mut1* and *mut2*, had no significant effect on integration frequency or on target

site specificity as shown by Southern blot or PCR analysis (data not shown), a finding that is explained if the 12-bp consensus sequence contains the true *attBST* core site. The fact that the minielement had the same pattern of integration of ~50% integration in the *attB1* as the intact CTnBST suggests that the IntBST and the *attBST* are sufficient for site selection for integration frequency. By contrast, in at least one other case, that of the *Bacteroides* MTn Tn4555, there is another protein needed to ensure site specificity (26).

CTnBST appears to require some identity between *attBST* and *attB* in the 12-bp region for integration. If so, this feature contrasts with the conjugative transposon Tn916, which seems not to have such a requirement for regions of identity between the cleavage sites used during integration sites. It is worth noting that the actual cleavage sites used by the integrase of Tn916 have only been inferred, not demonstrated directly. Initially, we thought that CTnDOT also had no requirement for sequence identity between the *attDOT* and the *attB* sites used for cleavage. Nonetheless, we have recently identified the cleavage sites used by the CTnDOT integrase. As with lambda Int, the cleavage sites are 7-bp staggered cuts (13) but the sequence between these cleavage sites consists of two homologous base pairs that are 5' to the five base pairs, previously called coupling sequences, that appear never to be identical. The integration mechanism mediated by IntBST appears to represent an intermediate between the recombination mediated by IntDOT and that mediated by lambda Int. Although CTns are now known to be widespread in bacteria, relatively little information is available about their integration requirements and mechanisms. This study provides new information about the likely diversity of the integration mechanisms of these elements.

In four of the integration events shown in Fig. 6A, the bases at positions 10 and 11 found at the junctions were the ones expected from the input sequences if there were staggered cuts within the two conserved regions flanking bp 10 and bp 11. That is, the sequence of the *attBST* was found at one junction and the sequence of *attB* was at the other end, as seen for the coupling sequences of CTnDOT and Tn916 (6, 19). In two of the integration events, however (examples 5 and 6 for *attBST_{mut2}*), one of the 2-bp sequences (the one in *attBST_{mut2}*) was replaced by or converted to the one in *attB_{mut2b}* or *attB_{mut2c}*, perhaps by a mismatch repair mechanism. We do not have experimental data to explain this phenomenon. It is noteworthy, however, that this type of event has also been seen in the case of phage λ integration into secondary sites and that sequences observed at the junctions varied with the locations and extents of the mismatches (5).

ACKNOWLEDGMENTS

This work was supported by a grant from the U.S. National Institutes of Health (AI/GM 22383) and by a grant from the Ellison Foundation (IDSS042703).

REFERENCES

1. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
2. Bacic, M., A. C. Parker, J. Stagg, H. P. Whitley, W. G. Wells, L. A. Jacob, and C. J. Smith. 2005. Genetic and structural analysis of the *Bacteroides* conjugative transposon CTn341. *J. Bacteriol.* **187**:2858–2869.
3. Bass, K. A., and D. W. Hecht. 2002. Isolation and characterization of cLV25,

- a *Bacteroides fragilis* chromosomal transfer factor resembling multiple *Bacteroides* sp. mobilizable transposons. *J. Bacteriol.* **184**:1895–1904.
4. Bauer, C. E., J. F. Gardner, and R. I. Gumpert. 1985. Extent of sequence homology required for bacteriophage lambda site-specific recombination. *J. Mol. Biol.* **181**:187–197.
5. Bauer, C. E., J. F. Gardner, R. I. Gumpert, and R. A. Weisberg. 1989. The effect of attachment site mutations on strand exchange in bacteriophage lambda site-specific recombination. *Genetics* **122**:727–736.
6. Cheng, Q., B. J. Paszkiet, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers. 2000. Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. *J. Bacteriol.* **182**:4035–4043.
7. Cheng, Q., N. Wesslund, N. B. Shoemaker, A. A. Salyers, and J. F. Gardner. 2002. Development of an in vitro integration assay for the *Bacteroides* conjugative transposon CTnDOT. *J. Bacteriol.* **184**:4829–4837.
8. Doublet, B., D. Boyd, M. R. Mulvey, and A. Cloeckert. 2005. The *Salmonella* genomic island 1 is an integrative mobilizable element. *Mol. Microbiol.* **55**:1911–1924.
9. Gupta, A., H. Vlamakis, N. Shoemaker, and A. A. Salyers. 2003. A new *Bacteroides* conjugative transposon that carries an *ermB* gene. *Appl. Environ. Microbiol.* **69**:6455–6463.
10. Holdman, L. V., E. P. Cato, and W. E. C. Moore. 1997. Anaerobe laboratory manual, 4th ed. Virginia Polytechnic Institute and State University, Blacksburg, VA.
11. Kuwahara, T., A. Yamashita, H. Hirakawa, H. Nakayama, H. Toh, N. Okada, S. Kuhara, M. Hattori, T. Hayashi, and Y. Ohnishi. 2004. Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc. Natl. Acad. Sci. USA* **101**:14919–14924.
12. Li, L. Y., N. B. Shoemaker, and A. A. Salyers. 1995. Location and characteristics of the transfer region of a *Bacteroides* conjugative transposon and regulation of transfer genes. *J. Bacteriol.* **177**:4992–4999.
13. Malanowska, K., A. A. Salyers, and J. F. Gardner. 2006. Characterization of a conjugative transposon integrase, IntDOT. *Mol. Microbiol.* **60**:1228–1240.
14. Metcalf, W. W., W. Jiang, and B. L. Wanner. 1994. Use of the rep technique for allele replacement to construct new *Escherichia coli* hosts for maintenance of R6K gamma origin plasmids at different copy numbers. *Gene* **138**:1–7.
15. Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26**:391–407.
16. Saito, H., and K. I. Miura. 1963. Preparation of transforming deoxy-ribonucleic acid by phenol treatment. *Biochim. Biophys. Acta* **72**:619–629.
17. Salyers, A. A., N. B. Shoemaker, A. M. Stevens, and L. Y. Li. 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev.* **59**:579–590.
18. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Scott, J. R., F. Bringel, D. Marra, G. Van Alstine, and C. K. Rudy. 1994. Conjugative transposition of Tn916: preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. *Mol. Microbiol.* **11**:1099–1108.
20. Shoemaker, N. B., C. E. Getty, E. P. Guthrie, and A. A. Salyers. 1986. Regions of *Bacteroides* plasmids pBFTM10 and pB8-51 that allow *Escherichia coli*-*Bacteroides* shuttle vectors to be mobilized by IncP plasmids and by a conjugative *Bacteroides* tetracycline resistance element. *J. Bacteriol.* **166**:959–965.
21. Shoemaker, N. B., G. R. Wang, and A. A. Salyers. 2000. Multiple gene products and sequences required for excision of the mobilizable integrated *Bacteroides* element NBU1. *J. Bacteriol.* **182**:928–936.
22. Shoemaker, N. B., H. Vlamakis, K. Hayes, and A. A. Salyers. 2001. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Appl. Environ. Microbiol.* **67**:561–568.
23. Shoemaker, N. B., R. D. Barber, and A. A. Salyers. 1989. Cloning and characterization of a *Bacteroides* conjugal tetracycline-erythromycin resistance element using a shuttle cosmid vector. *J. Bacteriol.* **171**:1294–1302.
24. Shoemaker, N. B., G. R. Wang, and A. A. Salyers. 1996. The *Bacteroides* mobilizable insertion element, NBU1, integrates into the 3' end of a Leu-tRNA gene and has an integrase that is a member of the lambda integrase family. *J. Bacteriol.* **178**:3594–3600.
25. 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. *Bio/Technology* **1**:784–791.
26. Tribble, G. D., A. C. Parker, and C. J. Smith. 1999. Transposition genes of the *Bacteroides* mobilizable transposon Tn4555: role of a novel targeting gene. *Mol. Microbiol.* **34**:385–394.
27. Valentine, P. J., N. B. Shoemaker, and A. A. Salyers. 1988. Mobilization of *Bacteroides* plasmids by *Bacteroides* conjugal elements. *J. Bacteriol.* **170**:1319–1324.
28. Vedantam, G., T. J. Novicki, and D. W. Hecht. 1999. *Bacteroides fragilis* transfer factor Tn5520: the smallest bacterial mobilizable transposon containing single integrase and mobilization genes that function in *Escherichia coli*. *J. Bacteriol.* **181**:2564–2571.
29. Wang, J., N. B. Shoemaker, G. R. Wang, and A. A. Salyers. 2000. Characterization of a *Bacteroides* mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene. *J. Bacteriol.* **182**:3559–3571.
30. Whittle, G., N. B. Shoemaker, and A. A. Salyers. 2002. The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cell. Mol. Life Sci.* **59**:2044–2054.
31. Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper, and J. L. Gordon. 2003. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**:2074–2076.