

Chlorobenzoate catabolic transposon Tn5271 is a composite class I element with flanking class II insertion sequences

(transposase/Tn3 family/evolution)

CINDY NAKATSU*, JAMES NG*, RAMA SINGH†‡, NEIL STRAUS†, AND CAMPBELL WYNDHAM*§

*Ottawa-Carleton Institute of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada; and †Department of Botany, University of Toronto, Toronto, ON M5S 2B3, Canada

Communicated by Allan M. Campbell, March 18, 1991 (received for review July 17, 1990)

ABSTRACT The structure of a transposon specifying the biodegradation of chlorobenzoate contaminants is described. Tn5271 is a 17-kilobase (kb) transposon that resides in the plasmid or chromosome of *Alcaligenes* sp. strain BR60 and allows this organism to grow on 3- and 4-chlorobenzoate. The transposon is flanked by a directly repeated sequence of 3201 base pairs (bp), which in turn is flanked by 110-bp inverted repeats. The 3.2-kb repeated sequence, designated IS1071, exists in multiple copies in the genome of *Alcaligenes* sp. strain BR60 and is involved in recombination of the catabolic genes into the chromosome of this strain. Sequence analysis revealed that the inverted repeat of IS1071 and the derived amino acid sequence of the single open reading frame within IS1071 are related to the inverted repeats and transposase (TnpA) proteins of the class II (Tn3 family) transposable elements. The absence of a resolvase gene within IS1071 suggests that this element is capable of determining the first step in class II transposition only. This was confirmed by observations on the IS1071-dependent formation of stable cointegrates in a recombination-deficient *Escherichia coli*. These results support an evolutionary scheme in which the class II transposable elements descended from simple insertion sequences.

Two of the most effective pressures driving natural selection in bacterial populations are the widespread use of antibiotics and the provision of unusual organic carbon sources in the form of environmental contaminants. In response to these pressures, bacteria have evolved varied mechanisms of resistance to antibiotics and different means of catabolizing unusual organic compounds. It is not surprising that the molecular mechanisms of adaptation to antibiotics and xenobiotics are similar. Plasmids have been implicated in the catabolism of complex organic compounds for two decades (1). The rearrangement of catabolic genes along with changes in their expression have been implicated as mechanisms in the evolution of bacteria exposed to recalcitrant carbon sources for almost as long (2). The discovery of transposable elements specifying resistance to antibiotics (3) was followed within a very few years by the first descriptions of transposition of the determinants for toluene degradation carried on plasmid pWWO (TOL) in *Pseudomonas putida* mt-2 (4–6).

Bacterial transposable elements fall into two well-defined structural classes (7). Class I elements include insertion sequences (IS), containing the genetic determinants for transposition only and composite transposons formed when flanking IS elements mobilize an intervening sequence (8). Class II elements, or the Tn3 family transposons, are related by inverted repeat (IR) similarities, transposase amino acid sequence homologies, and transposition mechanism (9). The class II catabolic transposons that have been characterized encode lactose (Tn951; ref. 10) and toluene metabolism

(Tn4653; ref. 11). Both are complex, nested class II elements. The naphthalene catabolic genes on plasmid NAH7 have recently been localized to a defective, class II transposon, Tn4655, which lacks a transposase function but has an intact resolution system (12).

Catabolic genes specifying the biodegradation of chlorinated aromatic compounds have been observed to undergo rearrangements; however, none of these has been shown to transpose as a defined element. Recently Tomasek *et al.* (13) mapped a 1477-base-pair (bp) IS element flanking 2,4,5-T catabolic genes of a *Pseudomonas cepacia* strain, suggesting a possible class I composite transposon structure.

In 1988 we described an *Alcaligenes* sp. strain BR60, isolated from a chlorobenzoate contaminated landfill that undergoes deletion of chlorobenzoate catabolic genes from the indigenous plasmid pBRC60 (formerly pBR60) and exhibits recombination of plasmid sequences into the chromosome (14). We have used this host–plasmid association to study gene transfer in the environment (15). In the following report, we describe the structure of the chlorobenzoate catabolic transposon in *Alcaligenes* sp. strain BR60, the nucleotide sequence of the flanking direct repeats in this element, and their relationship to both class I and class II transposable elements.¶

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions. *Alcaligenes* sp. strains BR60 (pBRC60), BR40 (pBRC40), BR6053, and BR6020 have been described (14). For plasmid or total genomic DNA isolations, these strains were grown on 1% tryptone/0.5% yeast extract/0.5% NaCl, at 25°C. *Escherichia coli* strains JM109, HB101, and DH5 α , with their plasmids pUC18, pUC19, pDPL13, pRK2013, or phage M13mp18 and M13mp19, have been described (16–19). A spontaneous rifampicin-resistant mutant of HB101 was also used. They were grown in Luria broth (19) at 37°C, with filter-sterilized ampicillin, kanamycin sulfate (Sigma), or rifampicin (Boehringer Mannheim) at 50, 50, or 200 μ g/ml, respectively.

DNA Manipulations and Restriction Mapping. Plasmids pBRC60, pBRC40, and *E. coli* plasmids were isolated as described (14, 19). Single-stranded M13mp18 and M13mp19 were isolated by the method of Dale *et al.* (20). Total genomic DNA was prepared from *Alcaligenes* strains by the procedure of Wheatcroft and Watson (21). Genomic DNA, pBRC40, and pBRC60 were single or double digested with HindIII, EcoRI, Pst I, Sal I, Nar I, BamHI, Nru I, Bgl II, Sac

Abbreviations: IS, insertion sequence(s); IR, inverted repeat; IHF, integration host factor.

‡Present address: Institute for Marine Bioscience, National Research Council of Canada, Halifax, NS B3H 3Z1, Canada.

§To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M65135).

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.

I, and *Sma* I (Bethesda Research Laboratories) and resolved in 0.7% agarose in Tris borate/EDTA buffer (19). The catabolic region of pBRC60 was mapped using the 10 restriction enzymes listed above by hybridizing [α - 32 P]dCTP-labeled (Amersham) fragments to pBRC60 digests immobilized on nylon (Hybond; Amersham) (14).

Insertion sequence IS1071 on a 4.1-kilobase (kb) *Nru* I fragment of pBRC40 was cloned into the *Nru* I site of pDPL13 (17) to give plasmid pBRN4029. The outside flanking ends of the two IS1071 sequences in Tn5271 were cloned as 1.2- and 1.9-kb *Eco*RI/*Nru* I fragments into *Eco*RI/*Sma* I-cut pUC18 to give pBREN12 and pBREN19, respectively. The inside flanking ends were obtained from existing clones (pBRE11 and pBRH3; ref. 14). Transformations into *E. coli* JM109 or DH5 α were carried out either by the method of Hanahan (22) or by electroporation with a Gene Pulser apparatus (Bio-Rad).

DNA Sequencing. *Hind*III fragment 4 cloned in pUC18 (pBRH4; ref. 14) was subcloned into the replicative forms of M13mp18 and M13mp19 (19) and then sequenced in both directions by using the commercial Sequenase kit (United States Biochemical) and deoxyadenosine 5'-[α - 35 S]thio-triphosphate (Amersham). Purified plasmid pBRN4029 was then used for plasmid sequencing in both directions away from the H4 internal sequence. Plasmid sequencing was also used to characterize the termini of the two IS1071 elements cloned from pBRC60. Oligonucleotide primers were prepared by using a Cyclone DNA synthesizer (Biosearch). Sequence analyses were performed by using Microgenie version 6.0 (Beckman). Tn5271, IS1071, pBRC60, and pBRC40 are Plasmid Reference Center listings (*E. Lederberg, Stanford*).

RESULTS

The catabolic region of the 85-kb plasmid pBRC60 is shown in Fig. 1. Tn5271 (17 kb) is flanked by 3.2-kb direct repeat sequences designated IS1071. Within the internal 10.6-kb unique region are sequences required for the growth of *Alcaligenes* sp. strain BR60 on chlorobenzoates. Deletion of the catabolic region from pBRC60 occurs at a frequency of 1.6×10^{-3} per cell per generation, in the absence of 3-chlorobenzoate, to give strain BR40 (14). The deletion plasmid pBRC40 (71 kb) retains one copy of IS1071 entirely within a 4.1-kb *Nru* I fragment generated by the deletion event. The DNA sequence of this copy of IS1071 is shown in Fig. 2. The sequence shown represents the product of an intramolecular recombination at an unknown point between the two direct repeats of IS1071 present on pBRC60. This was confirmed by comparison of the sequence in Fig. 2 with sequences at the four IS1071 termini of Fig. 1:

left outside,
5'-GCCGGTCTCGGGTC GGGGTCTCCTCGTTT-3';
left inside,
5'-CGTCGAGCAGGGCAT GGGGTCTCCTCGTTT-3';

right inside,
5'-CGGGCAGCTTGTCGC GGGGTCTCCTCGTTT-3';
right outside,
5'-GGTTTTGCTCGTAAA GGGGTCTCCTCGTTT-3'.

The termini of IS1071 in Fig. 2 correspond to the left outside and right outside sequences of Tn5271. There are no duplicated sequences flanking the IS1071 termini.

The distribution of IS1071 in the genomes of *Alcaligenes* sp. strain BR60, strain BR6020 (mitomycin C cured), and strain BR6053 (chromosomal Tn5271 recombinant) (14) was determined by digestion of genomic DNAs with *Bgl* II and probing with *Hind*III fragment 4 (IS1071 probe; data not shown). The BR60 genomic digest had seven fragments homologous to the IS1071 probe, ranging in size from 5 to 28 kb. Two of these (11 and 24 kb) corresponded to pBRC60 fragments containing the IS element (fragments *Bgl* II-2 and -3; Fig. 1). Strain BR6053 genomic digests also contained seven copies of IS1071, but the pBRC60 fragments were missing and two new chromosomal fragments were observed. Strain BR6020 genomic digests contained only three copies of IS1071 on chromosomal *Bgl* II fragments of 9, 19, and 22 kb.

DNA homology searches revealed that the 110-bp IR of IS1071 was related to the class II (Tn3 family) IRs (Fig. 3). The regions of homology in the outer 38 bp have been recognized as DNA-binding and cutting sites for the class II transposases (25). There was no significant homology of the IS1071 IR sequence extending from 39 to 110 bp with sequences flanking the 38-bp IRs of the Tn3 family transposons shown. However, an integration host factor (IHF) consensus binding sequence occurs at nucleotides 71–83 within the left and right IR of IS1071 (Fig. 3), at positions analogous to IHF binding sites near the Tn γ δ IR (23).

The large open reading frame within IS1071 (Fig. 2) spans 2910 bp, coding for a putative 970-amino acid polypeptide of molecular weight 108,413. This putative transposase exhibited significant homology to the class II transposases of Tn3 and Tn21 (Fig. 4). These alignments place greater weight on the known alignment of the Tn3 and Tn21 TnpA sequences. There was 21% identity of amino acids in comparing either IS1071 to Tn3 or IS1071 to Tn21. Only 12% of amino acids were conserved for all three TnpA sequences. Including conservative substitutions raised similarities to 37% (with Tn3) or 39% (with Tn21).

The absence of the class II transposon signature of a 5-bp direct repeat of target DNA flanking IS1071 in pBRC60 and pBRC40 suggests that these copies of IS1071 are not the immediate products of transposition events. Indeed, this element may represent only the remnant of a once-functional class II transposon. The latter possibility was eliminated by characterizing transposition products of IS1071 in recombination-deficient mutants of *E. coli*. Plasmid pBRN4029 carrying a single copy of IS1071 formed cointegrates with the conjugative plasmid pRK2013 at a frequency of 5×10^{-4} per transconjugant in mating-out assays to a rifampicin-resistant *E. coli* HB101 recipient. Restriction mapping of these cointe-

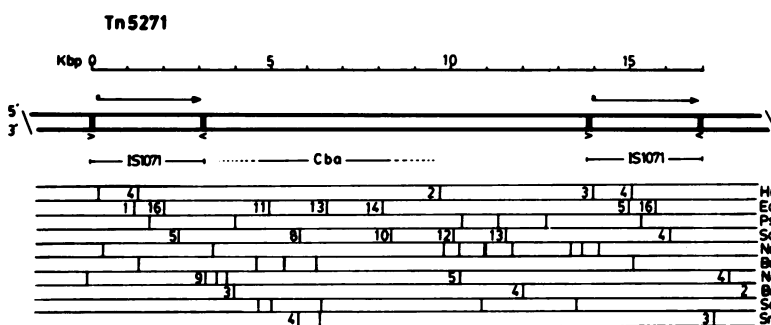


FIG. 1. Restriction enzyme digest map of the catabolic transposon Tn5271 from *Alcaligenes* sp. strain BR60. Open reading frames within IS1071 are indicated by arrows. The positions of inverted repeat sequences are indicated by vertical bars and arrowheads. The location of the catabolic genes (—Cba—) has been determined by subcloning (unpublished data). Only cloned restriction fragments or fragments used as probes are numbered.

-35 NruI
 TCGCGAAAGCCACAGCCGGGGCGGTCTCGCGGTC 1 15 25
GGGGT CTCCTCGTTT TCAGTGCAT

35 45 55 65 75 85
AAGTGACGGT ACGCAAAAGT AGCACTGGCG CCGGGGTGGT CTGGGTAGAC CGTTGATTC

95 105 115 125 135 145
ATTGACTTTC CTGTTCGGCT TGTAAACGGG TATGGTGGCC TCCCACTTTT GAGGTTCCAG

160 190
 ATG CAG GGT TGG CAC ACA ACG TTT TTG GGG ATG CGT GGG CTC CCC GAT ATC
 Met Gln Gly Trp His Thr Thr Phe Leu Gly Met Arg Gly Leu Pro Arg Asp Ile

220 250
 AGC GAC TTC GAG ATG AAG GCA TTT TTC ACC TTC GAT GGT GCC GAG CGC GAC GCA
 Ser Asp Phe Glu Met Lys Ala Phe Phe Thr Phe Asp Gly Ala Glu Arg Asp Ala

280 HindIII
 ATC AAT GCA CGC CGA GGT GAT TCC CAC AAG CTT GGT CTG GCG CTC CAT ATT GGT
 Ile Asn Ala Arg Arg Gly Asp Ser His Lys Leu Gly Leu Ala Leu His Ile Gly

310 340
 TTC CTG CGC ATG AGT GGG CGT TTG CTC GGT GCC TTT CGG GTA ATT CCA GTA GCC
 Phe Leu Arg Met Ser Gly Arg Leu Leu Gly Ala Phe Arg Val Ile Pro Val Ala

370 400
 TTG TGG CGC CAC CTT GGC AAC GAG CTT GGC ATT GCA GCA CCA GAA GTC GCC TCG
 Leu Trp Arg His Leu Gly Asn Glu Leu Gly Ile Ala Ala Pro Glu Val Ala Ser

430 460
 CTG AGA GCC ATG TAT GAA CGC GGG CGC ACG CTA TTC GAT CAC CAA CAA GTA GCC
 Leu Arg Ala Met Tyr Glu Arg Gly Arg Thr Leu Phe Asp His Gln Gln Val Ala

490 520
 TGC ACG GTC CTT GGA TTC CAG TGG ATG AGC GAG CAC CAG CGC CGC TCA CTG GTA
 Cys Thr Val Leu Gly Phe Gln Trp Met Ser Glu His Gln Arg Arg Ser Leu Val

550 610
 CGT GAA CTG CGC GAC GAA VTG GCC GGC TGC GAC CGC GAT CAG CTA CTC GTG CGG
 Arg Glu Leu Arg Asp Leu Ala Gly Cys Asp Arg Asp Gln Leu Leu Val Arg

580 640
 GCG CGT CAA TGG CTG TAC AAG AAC AAG CTG GTG ATC GTC CAC GAG CGG GCA ATT
 Ala Arg Gln Trp Leu Tyr Lys Asn Lys Leu Val Ile Val His Glu Arg Ala Ile

670 730
 CGG ACA CTG ATT GCG CGC GCA CTT GCC CAG CTT GAA GTT GAA ACA GGC ACC GCC
 Arg Thr Leu Ile Ala Ala Ala Leu Ala Gln Leu Glu Val Glu Thr Gly Thr Ala

700 790
 ATC GCC GCC AGC GTT GAT CCA GCA ACA CTT GAT CGC TGG CGA GCC TCA GTT TCA
 Ile Ala Ala Ser Val Asp Pro Ala Thr Leu Asp Arg Trp Arg Ala Ser Val Ser

760 790
 GAG CTG CGC CCA GAT GGA CAA ACC CAG CAG AGT TGG CTA TGG GCT GCA CGC GCG
 Glu Leu Arg Pro Asp Gly Gln Thr Gln Gln Ser Trp Leu Trp Ala Ala Pro Ala

820 880
 AAA CAC TCA ACC CGC CAA ATC AGC GAG GTA CTG GAG CGC ATC GAC CTG CTT TAC
 Lys His Ser Thr Arg Gln Ile Ser Glu Val Leu Glu Arg Ile Asp Leu Leu Tyr

850 910
 ACG CTG GAC GTT CAT AAG CAC CTG GCA GAC ATC CCC GAT CTC ATC TTG CGC CGC
 Thr Leu Asp Val His Lys His Leu Ala Asp Ile Pro Asp Leu Ile Leu Arg Arg

940 1000
 TAC CGC CGC CGA CTT GTC TCC AGG CCG CCC TCA GCC GGA GCC AAG ATC AAA GAG
 Tyr Ala Arg Arg Leu Val Ser Arg Pro Pro Ser Ala Gly Ala Lys Ile Lys Glu

970 1060
 CCA GCG CGC ACC GTG GAG GTC GCA TGC TTT CTT CGG TAT TGC CTG TTC ACC ACC
 Pro Ala Arg Thr Val Glu Val Ala Cys Phe Leu Arg Tyr Cys Leu Phe Thr Thr

1030 1060
 ACA GAC CAG TTG ATC CTT ATG GTG CAG CGC CGG ATC GCC GAT CTG TGG CGT CAG
 Thr Asp Gln Leu Ile Leu Met Val Gln Arg Arg Ile Ala Asp Leu Trp Arg Gln

1090 1150
 GCT GCC GCC GAT GTC CCC GCT ACC GTC AAT TGG GCC GCA ATG TAC AAA ACG CTG
 Ala Ala Ala Asp Val Pro Ala Thr Val Asn Trp Ala Ala Met Tyr Lys Thr Leu

1120 1210
 CTC GGC GAA CTT GTT GCC TTG AGC GCG CAA GGT GCG GTG CCA GAT GCT GAG TTG
 Leu Gly Glu Leu Val Ala Leu Ser Ala Gln Gly Ala Val Pro Asp Ala Glu Leu

1180 1270
 CGT GCC CGT CTT GAA GCC TTG ATC ACC GAA ACC CAG AAA CGC AAA CCA CCG AGC
 Arg Ala Arg Leu Glu Ala Leu Ile Thr Glu Thr Gln Lys Arg Lys Pro Pro Ser

1240 EcoRI 1270
 AGG GCC TCC CTG GTC CGC GAG GGA TTG ATT GAT GGA ATT CGC CCC CTG GCG TCG
 Arg Ala Ser Leu Val Arg Glu Gly Leu Ile Asp Gly Ile Arg Pro Val Arg Ser

1300 1330
 TTG CTC GTC GCC ATT GCA AAG CTG CCC TGG CAG GCC ACC GGC GAG CAT CCB GCC
 Leu Leu Val Ala Ile Ala Lys Leu Pro Trp Gln Ala Thr Gly Glu His Pro Ala

HindIII
 ATC GAG TAC CTT GCC AAG CTG CAA GCT TTA TAT CTC AAA GGA TCC AGA AAG CTG
 Ile Glu Tyr Leu Ala Lys Leu Gln Ala Leu Tyr Leu Lys Gly Ser Arg Lys Leu

1390 1480
 CCA GTT GAA GTG GTG GCA CCA AGT CTG GGA ATG ATC TGG CAG GTT TCG ATC TCC
 Pro Val Glu Val Val Ala Pro Ser Leu Gly Met Ile Trp Gln Val Ser Ile Ser

1450 1540
 AGC CCA GAC CGG GAA CGG CGG TTT CAG GCG TTG GAG GTG GCC ACC CTG TTT GCC
 Ser Pro Asp Arg Glu Arg Ala Phe Gln Ala Leu Glu Val Ala Thr Leu Phe Ala

1510 1600
 CTG CGC CGC GCG GTG CGC AAT GGC TCG GTC TGG ATT GAG CAC AGC CTG AGC TTT
 Leu Arg Arg Ala Val Arg Asn Gly Ser Val Trp Ile Glu His Ser Leu Ser Phe

1570 1600
 CGG GGT CGT GCG CGC TTG TTC TCC ACG GAG GAG CGT TGG CAG GCA GAG TCC AAG
 Arg Gly Arg Ala Arg Leu Phe Thr Asp Glu Arg Trp Gln Ala Glu Ser Lys

1630
 AAA CAC TAT GCC CGT CTA TCG TTA CCC AGC AAG GCT GCC ACT TTC TTG AAG CCT
 Lys His Tyr Ala Arg Leu Ser Leu Pro Ser Lys Ala Ala Thr Phe Leu Lys Pro

1660 1690
 TTG CTG GCC AGA GTA ACT GCC GGT GTC GAT GCG GTG GCC GCT GCA GCC CGC AGT
 Leu Leu Ala Arg Val Thr Ala Gly Val Asp Ala Val Ala Ala Ala Ala Arg Ser

1720 1750
 GGC GTA CTG CGC GTG GAT GAT GAA CTC CAT TTG TCG CCA TTG CCC GCA GAG GAC
 Gly Val Leu Arg Val Asp Asp Glu Leu His Leu Ser Pro Leu Pro Ala Glu Asp

1780 1810
 GAA GAC CCA GAA GTG ACC AAG CTG CGC GCG GCT TTG GAT CAC CGC ATC GGT GAG
 Glu Asp Pro Glu Val Thr Lys Leu Arg Ala Ala Leu Asp His Arg Ile Gly Glu

1840 1870
 GTT CAA TTG CCG GAA GTG ATT CTG GCC GTT GAC GCC CAG GTG CGC TTT AGC TGG
 Val Gln Leu Pro Glu Val Ile Leu Ala Val Asp Ala Gln Val Arg Phe Ser Trp

1900
 ATC ATG CTC GGA CGT GAG CCG CGC TCT ACC GAC GAG CTG CTG ATG GTC TAT GCC
 Ile Met Leu Gly Arg Glu Pro Arg Ser Thr Asp Glu Leu Leu Met Val Tyr Ala

1930 1960
 GGC ATC ATG GCC CAC GGC ACC AGT CTG ACT GCG GTC GAA TGC GCG CGC ATG ATT
 Gly Ile Met Ala His Gly Thr Ser Leu Thr Ala Val Glu Cys Ala Arg Met Ile

1990 2020
 CCG CAA TTG TCT GCC ACC AGC ATT CGC CAG GCC ATG CGC TGG GCG CGG GAC GAA
 Pro Gln Leu Ser Ala Thr Ser Ile Arg Gln Ala Met Arg Trp Ala Arg Asp Glu

2050 EcoRI 2080
 CGC CGT CTG AGC CAG GCC TGC CAG GCT GTG CTG GAA TTC ATG CAG CGA CAC CGC
 Arg Arg Leu Ser Gln Ala Cys Gln Ala Val Leu Glu Phe Met Gln Arg His Pro

2110 2140
 ATT GCC GCC ACC TGG GGG CGG TCC GAT TTG GCA TCT TCT GAC ATG ATG AGC ATG
 Ile Ala Ala Thr Trp Gly Arg Ser Asp Leu Ala Ser Ser Asp Met Met Ser Met

2170
 GAG ACC ACC AAA CGG GTG TGG CAA GCC CGG CTT GAT CCT CGC CGC AAC ACA CCT
 Glu Thr Thr Lys Arg Val Trp Gln Ala Arg Leu Asp Pro Arg Arg Asn Thr Pro

2200 2230
 TCC ATT GGA ATC TAC TCC CAT GTA AAA GAC CGG TGG GGC ATC TTC CAT GCG CAG
 Ser Ile Gly Ile Tyr Ser His Val Lys Asp Arg Trp Gly Ile Phe His Ala Gln

2260 2290
 CCC TTT GTG CTC AAT GAG CGC CAG GCG GGC GTG GCC ATT GAA GGT GTC ATC CGC
 Pro Phe Val Leu Asn Glu Arg Gln Ala Gly Val Ala Ile Glu Gly Val Ile Arg

2320 2350
 CAA GAA AAG CTG GAG ACC AGC CAG CTT GCT GTG GAT ACC CAT GGC TAC ACC GAC
 Gln Glu Lys Leu Glu Thr Ser Gln Leu Ala Val Asp Thr His Gly Val Thr Asp

2380 2410
 TTT GCC ATG TCA CAT GCC CGT TTG CTT GGT TTT GAT CTT TGC CCG CGG TTG AAG
 Phe Ala Met Ser His Ala Arg Leu Leu Gly Phe Asp Leu Cys Pro Arg Leu Lys

2440
 GAA CTC AAA CAG CGC CAC CTC TTT GTG CCA CGC GGC ACC AAA GTG CCC GCA GAA
 Glu Leu Lys Gln Arg His Leu Phe Val Pro Arg Gly Thr Lys Val Pro Ala Glu

2470 2500
 ATC GCT GCG GTG TGC GAA GCC AAT GTC GAC GTC GCT TTG ATC GAA AAG CAT TGG
 Ile Ala Ala Val Cys Glu Ala Asn Val Asp Val Ala Leu Ile Glu Lys His Trp

2530 2560
 GAT AGT CTG GTG CAC CTG GCA GCC TCG GTC ATG AGC GGA CAT GCC AGT GCG GTG
 Asp Ser Leu Val His Leu Ala Ala Ser Val Met Ser Gly His Ala Ser Ala Val

2590 2620
 GCA GCT CTT CCG CGG TTC GGT TCT GCC GCC CAG GGC GAT CCA ATC TAT GAG GCT
 Ala Ala Leu Ala Arg Phe Gly Ser Ala Ala Gln Gly Asp Pro Ile Tyr Glu Ala

2650 2680
 GGC GTG CAA TTG GGG CGG TTG CTG CGT ACG GCG TTT TTG GCT GAC TAC TTT GTC
 Gly Val Gln Leu Gly Arg Leu Leu Arg Thr Ala Phe Leu Ala Asp Tyr Phe Val

2710
 AAG GAC GCT TTC AGG AAC GAG TTG CGC CGG GTG CTC AAT CCG GCC GAG GCT GTT
 Lys Asp Ala Phe Arg Asn Glu Leu Arg Arg Val Leu Asn Arg Gly Glu Ala Val

2740 2770
 AAC GCC CTC AAG CGC GCC ATT TAT ACC GGC CGG ATC AGC CCG GCG CAG GCC AAA
 Asn Ala Leu Lys Arg Ala Ile Tyr Thr Gly Arg Ile Ser Pro Ala Gln Ala Lys

2800 2830
 CGT GTC GAT GAA ATG CAG GCT GTG GCC GAT GCG TTG AGC CTG ATG GCC AAC ATC
 Arg Val Asp Glu Met Gln Ala Val Ala Asp Ala Leu Ser Leu Met Ala Asn Ile

2860 2890
 GTG ATG GCG TGG AAT ACC TCA CAG ATG CAG GCG GTC CTG GAT CCG TGG TCG AAC
 Val Met Ala Trp Asn Thr Ser Gln Met Gln Ala Val Leu Asp Arg Trp Ser Asn

2920 2950
 CGC CGC CAG CTC ATT CCA CCG GAA CTG ATC GGG AAG ATT GCG CCC ACC AGG CTG
 Arg Arg Gln Val Ile Pro Pro Glu Leu Ile Gly Lys Ile Ala Pro Thr Arg Leu

2980
 GAG AGC ATC AAC TTG CCG GGT GTG TTT CGC TTC CCG GTT GAC CGC TAT GCT GAC
 Glu Ser Ile Asn Leu Arg Gly Val Phe Arg Phe Pro Val Asp Arg Tyr Ala Asp

3010 3040 3060
 CAA ATC CTG CCT TCG CGG CCA AAT GCA TCG ATA ACT GGC ACC AAT GGA TGA
 Gln Ile Leu Pro Ser Arg Pro Asn Ala Ser Ile Thr Gly Thr Asn Gly *
CGACCCACGG TTTGACGCCA CGAATCGCAG ATTTGAAAGT GAACGGGAAA GTCAAATGAAA

3070 3080 3090 3100 3110 3120
CGACCCACGG TTTGACGCCA CGAATCGCAG ATTTGAAAGT GAACGGGAAA GTCAAATGAAA

3130 3140 3150 3160 3170 3180
TCAACGATCT ACCAACACCA CCTCCGCGCC AGTCTAGCT TTTGTAACCG TCACTTAATG

3190 3200
CACTGAAAAC GAGGAGACCC TTTAGCAGAAAACCTACTGTCCGACACCAT

FIG. 2. Nucleotide sequence of IS1071. Numbering is relative to the first G in the IR sequence (underlined). Short sequences on either side of the insertion element are shown. The translation product of the open reading frame beginning at position 146 is shown. There is a potential ribosome binding site (GAGG) at position 135 and a stem-loop structure underlined beginning at the A in the stop codon at position 3058.

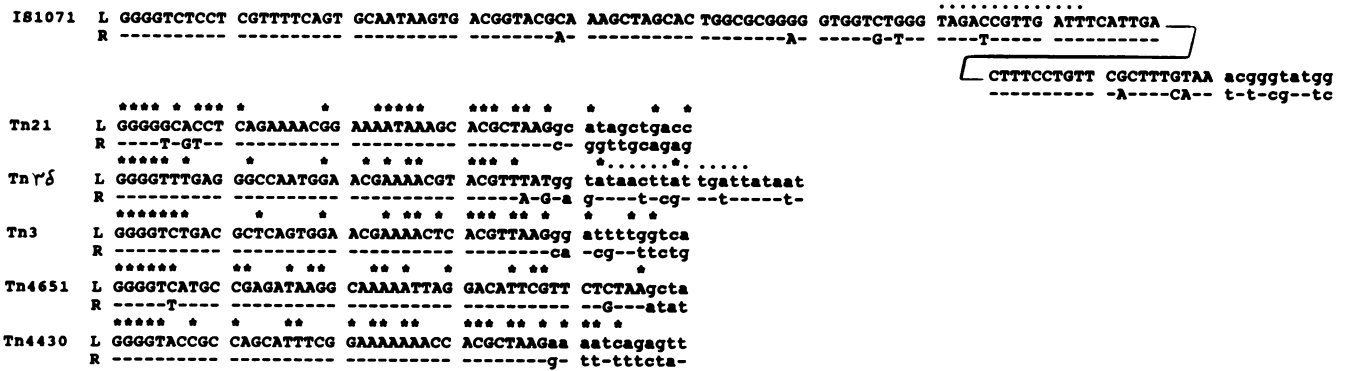


FIG. 3. Comparison of IR sequences defining the ends of class II transposable elements. L and R, left and right IRs, defined as distal and proximal to the *tnpA* gene (11). Nucleotides that differ in the right IR compared to the left are indicated. The ends of the IRs are indicated by a shift to lowercase letters. An asterisk above the sequence indicates identity at that position to the left IR of IS1071. Dotted lines mark sequences in IS1071 and Tn $\gamma\delta$ homologous to the IHF binding site consensus sequence. Sources: Tn21 and Tn3, GenBank; Tn4651, ref. 11; Tn $\gamma\delta$, ref. 23; Tn4430, ref. 24.

grates showed that they consist of the entire pBRN4029 replicon, flanked by direct repeats of IS1071, integrated into various sites on pRK2013 (data not shown). These cointegrates were quite stable in a *recA* host, although resolution was observed with subculturing in some cases.

DISCUSSION

The chlorobenzoate catabolic genes on plasmid pBRC60 are subject to deletion at a high frequency due to recombination

between IS1071 copies on the plasmid. Recombination of the catabolic genes with the chromosome of BR60 also occurs at a high frequency (14) and is probably a result of recombination between plasmid and chromosomal IS1071 copies. This background of host-encoded recombination activity, combined with the fact that BR60 was recently isolated from an environment in which natural genetic exchange is likely to be important (15), makes it difficult to attribute the structure of Tn5271 or the rearrangements of IS1071 in *Alcaligenes* sp.



FIG. 4. Alignment of TnpA sequences for Tn3, IS1071, and Tn21. Identical amino acids are indicated by asterisks; similar amino acids are indicated by colons. The numbers above each set of three sequence lines refer to the amino acid numbers for the IS1071 sequence. The Tn3 sequence is 1004 and the Tn21 sequence is 988 amino acids long. Data for the Tn3 and Tn21 transposase sequences were taken from GenBank.

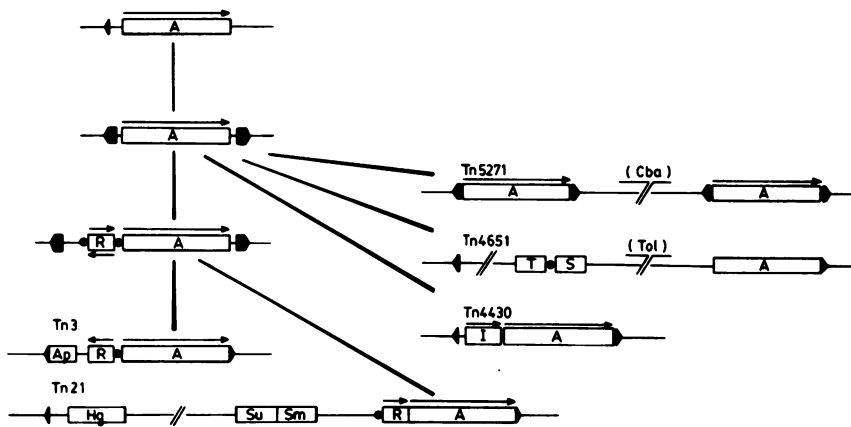


FIG. 5. Evolution of class II transposable elements. An ancestral element is indicated by a transposase gene (A) and its recognition sequence (A). IR sequences including the 38-bp recognition sequence are indicated by thick arrowheads (A); the direction of transcription of *tnpA* (A), *tnpR* (R), and integrase (I) genes, where known, is indicated by arrows; ●, *res* site; Ap, Su, Sm, and Hg, ampicillin-, sulfonamide-, streptomycin-, and mercury-resistance loci, respectively; Cba and Tol, chlorobenzoate and toluene catabolic loci, respectively; T and S, trans-acting resolution genes for the toluene transposon.

strain BR60 to transposition events. However, we have observed transposition of the cloned IS1071 element to generate cointegrates in a recombination-deficient *E. coli*. Therefore, it would seem likely that transposition has played a part in the mobilization of catabolic genes in Tn5271. We are currently characterizing transposition products of both Tn5271 and IS1071 in recombination-deficient hosts in order to define target sites.

Class II transposable elements share structural features that clearly delineate them from other mobile genetic elements (9). IS1071 has inverted repeat sequences and a transposase sequence that place it among the class II elements; however, it has no sequence homologous to class II resolvases. A search of the internal 10.6-kb sequence of the transposon (unpublished data) has failed to reveal a candidate for a resolvase-like gene. Given the complementary nature of the activities of transposase and resolvase, and the evidence that resolvase catalyzes site-specific recombination between cis-oriented *res* sites (26), it is unlikely that a resolvase gene linked to a specific transposon would exist outside of the IR defining the element. We therefore conclude that IS1071 is a true IS carrying the coding information for replicative transposition only. Our observation of the IS1071-dependent formation of stable cointegrates in a *recA* host supports this conclusion.

The transposase alignments of IS1071 with Tn3 and Tn21 revealed only 12% conservation of amino acids, comparable to the identity observed when Tn3, Tn21, Tn501, Tn917, Tn2501, and Tn4430 sequences were all aligned (24). These transposons are therefore representative of the known evolutionary limits of the class II elements. Schmitt *et al.* (27) have proposed a scheme for the evolution of the class II transposons based on the observation that a single copy of an IR sequence and a transposase are the minimal requirements for transposition. Inverted repeats of the recognition sequence flanking the transposase may have been acquired by a process such as strand exchange during hairpin loop replication. Such an element would be structurally identical to IS1071 (Fig. 5). One prediction of this scheme is that the ancestral IRs, because of the nonspecific events occurring at a hairpin loop replication fork, may have been longer than the minimum transposase recognition sequence of 38 bp. This is the case observed with IS1071 (110-bp IR), Tn2501 (48-bp IR), and the toluene transposon Tn4651 (46-bp IR). Acquisition of site-specific recombination systems (resolvase or integrase), antibiotic-resistance determinants, or catabolic genes by an element like IS1071 would give rise to the class II transposons depicted in Fig. 5 (24, 27, 28).

The class II transposons have been very successful in promoting the fitness of a range of prokaryotes under very different environmental conditions. With the discovery of

IS1071, there is now good evidence to suggest that the transposase gene alone confers a selective advantage on its host under some environmental conditions.

The assistance of Mary Barbara, Ellen Schwartzel, and Maysoon Salih in this work is greatly appreciated. Research was supported by grants to C.W. from the Natural Sciences and Engineering Research Council of Canada.

- Chakrabarty, A. M. & Gunsalus, I. C. (1971) *Genetics* **68**, S10.
- Bayley, S. A., Duggleby, C. J., Worsey, M. J., Williams, P. A., Hardy, K. G. & Broda, P. (1977) *Mol. Gen. Genet.* **154**, 203–204.
- Hedges, R. W. & Jacob, A. F. (1974) *Mol. Gen. Genet.* **132**, 31–40.
- Jacoby, G. A., Rogers, J. E., Jacob, A. F. & Hedges, R. W. (1978) *Nature (London)* **274**, 179–180.
- Nakazawa, T., Hayashi, E., Yokota, T., Ebina, Y. & Nakazawa, A. (1978) *J. Bacteriol.* **134**, 270–277.
- Chakrabarty, A. M., Friello, D. A. & Bopp, L. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3109–3112.
- Kleckner, N. (1981) *Annu. Rev. Genet.* **15**, 341–404.
- Sylvanen, M. (1988) in *Genetic Recombination*, eds. Kucherlapati, R. & Smith, G. R. (Am. Soc. Microbiol., Washington), pp. 331–356.
- Sherratt, D. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington), pp. 163–184.
- Michiels, T., Cornelis, G., Ellis, K. & Grinstead, J. (1987) *J. Bacteriol.* **169**, 624–631.
- Tsuda, M., Minegishi, K.-I. & Iino, T. (1989) *J. Bacteriol.* **171**, 1386–1393.
- Tsuda, M. & Iino, T. (1990) *Mol. Gen. Genet.* **223**, 33–39.
- Tomasek, P. H., Frantz, B., Sangodkar, U. M. X., Haugland, R. A. & Chakrabarty, A. M. (1989) *Gene* **76**, 227–238.
- Wyndham, R. C., Singh, R. K. & Straus, N. A. (1988) *Arch. Microbiol.* **150**, 237–243.
- Fulthorpe, R. R. & Wyndham, R. C. (1991) *Appl. Environ. Microbiol.* **57**, 1546–1553.
- Yannisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
- Gendel, S., Straus, N., Pulleyblank, D. & Williams, J. (1983) *J. Bacteriol.* **156**, 148–154.
- Figurski, D. H. & Helinski, D. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1648–1652.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
- Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–46.
- Wheatcroft, R. & Watson, R. J. (1988) *Appl. Environ. Microbiol.* **54**, 574–576.
- Hanahan, D. (1984) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Washington), p. 115.
- Wiater, L. A. & Grindley, N. D. F. (1988) *EMBO J.* **7**, 1907–1911.
- Mahillon, J. & Lereclus, D. (1988) *EMBO J.* **7**, 1515–1526.
- Ichikawa, H., Ikeda, K., Amemura, J. & Ohtsubo, E. (1990) *Gene* **86**, 11–17.
- Krasnow, M. A. & Cozzarelli, N. R. (1983) *Cell* **32**, 1313–1324.
- Schmitt, R., Rogowsky, P., Halford, S. E. & Grinstead, J. (1985) in *Evolution of Prokaryotes*, eds. Schleifer, K. H. & Stackebrandt, E. (Academic, London), pp. 91–104.
- Mercier, J., Lachapelle, J., Couture, F., Lafond, M., Vezina, G., Boissinot, M. & Levesque, R. C. (1990) *J. Bacteriol.* **172**, 3745–3757.