Nucleotide Sequence and Characterization of a Repetitive DNA Element from the Genome of *Bordetella pertussis* with Characteristics of an Insertion Sequence

By MARTHA A. McLAFFERTY,* DAVID R. HARCUS AND ERIK L. HEWLETT

Departments of Medicine and Pharmacology, University of Virginia School of Medicine, Charlottesville, VA 22908, USA

(Received 24 February 1988)

A repeating element of DNA has been isolated and sequenced from the genome of *Bordetella pertussis*. Restriction map analysis of this element shows single internal *Cla*I, *Sph*I, *Bst*EII and *Sal*I sites. Over 40 DNA fragments are seen in *Cla*I digests of *B. pertussis* genomic DNA to which the repetitive DNA sequence hybridizes. Sequence analysis of the repeat reveals that it has properties consistent with bacterial insertion sequence (IS) elements. These properties include its length of 1053 bp, multiple copy number and presence of 28 bp of near-perfect inverted repeats at its termini. Unlike most IS elements, the presence of this element in the *B. pertussis* genome is not associated with a short duplication in the target DNA sequence. This repeating element is not found in the genomes of *B. parapertussis* or *B. bronchiseptica*. Analysis of a DNA fragment adjacent to one copy of the repetitive DNA sequence has identified a different repeating element which is found in nine copies in *B. parapertussis* and four copies in *B. pertussis*, suggesting that there may be other repeating DNA elements in the different *Bordetella* species. Computer analysis of the *B. pertussis* repetitive DNA element has revealed no significant nucleotide homology between it and any other bacterial transposable elements, suggesting that this repetitive sequence is specific for *B. pertussis*.

INTRODUCTION

Bacterial insertion sequence (IS) elements are discrete units of DNA which are capable of inserting themselves into multiple sites in a bacterial genome. The distinguishing properties of IS elements include their size (800–1800 bp), multiple copy number, presence of inverted repeats at their termini, and short duplication of target DNA at their insertion site. They also appear to encode only proteins that are involved in their own transposition (Calos & Miller, 1980; Grindley & Reed, 1985; Iida et al., 1983; Kleckner, 1981). This lack of encoded genetic markers, along with their small size, differentiates IS elements from larger bacterial transposons which encode phenotypically identifiable characteristics such as drug and antibiotic resistance. The exact function of IS elements is unknown, though they appear to be important for evolutionary processes by mediating chromosomal rearrangements such as deletions, inversions and duplications, and by altering expression of adjacent genes (Arber, 1983; Calos & Miller, 1980; Rappuoli et al., 1987; Saesler et al., 1974; Scordilis et al., 1987). These elements were first identified in *Escherichia coli* but now have been found in numerous other prokaryotes.

Recently we have been analysing the genome of *Bordetella pertussis* in search of DNA fragments for use as pertussis-specific DNA probes. In the course of these studies we have found that this organism contains a sequence of DNA which is repeated many times throughout its genome, but is not present in *B. parapertussis* or *B. bronchiseptica* (McLafferty et al., 1986). This

Abbreviations: IPTG, isopropyl β-D-thiogalactoside; X-gal, 5-bromo-4-chloro-3-indolyl β-D-galactoside.

0001-4726 © 1988 SGM
observation has been substantiated by McPheat & McNally (1987). With subsequent isolation and sequencing of this repeating DNA element from *B. pertussis*, we find that it has characteristics of bacterial IS elements.

**METHODS**

**Bacterial strains, plasmids and growth conditions.** *B. pertussis* Tohama has been described (Imaizumi *et al*., 1983). *B. parapertussis* CN8234 was obtained from P. Novotny (Wellcome Research Laboratories, Beckenham, UK). *B. parapertussis* 501 and *B. bronchiseptica* strains 213, 361 and 469 are clinical isolates maintained in our laboratory. The bacteria were cultivated on Bordet-Gengou plates (Difco) or modified Stainer-Scholte broth (Hewlett & Wolff, 1976). Cosmid p11-11 was obtained from A. A. Weiss (Medical College of Virginia, Richmond, VA, USA) and contains *B. pertussis* DNA cloned into cosmid pH79. Specific restriction fragments of the cosmid p11-11 insert were cloned into plasmid pBR325 by standard procedures.

**Isolation and manipulation of DNA.** Plasmid DNA was isolated by the alkaline lysis procedure of Birnboim & Doly (1979), and then purified by density centrifugation through a CsCl/ethidium bromide gradient (Maniatis *et al*., 1982). Genomic DNA was isolated by the method of Yee & Inouye (1981). Restriction endonucleases were used according to the specifications of the manufacturer (New England BioLabs). DNA restriction fragments were resolved by electrophoresis through 0.9% agarose gels in Tris/acetate buffer (Maniatis *et al*., 1982). Specific DNA fragments were purified from the agarose gels using NA-45 DEAE paper (Schleicher & Schuell) as described by the manufacturer.

**DNA hybridization.** The transfer of DNA from agarose gels to nylon membranes was done as recommended by the manufacturer (AMF Cuno). DNA probes were labelled with $^{32}$P by nick-translation (Maniatis *et al*., 1982) and hybridized to the nylon membranes using the method of Church & Gilbert (1984).

**Preparation of DNA for sequencing.** Fragments to be sequenced were cloned directly into M13mp19 or M13mp18 phage vectors, transfected into *E. coli* JM101 and plated onto YT plates (per litre: 8 g tryptophane, 5 g yeast extract, 2.5 g NaCl, pH 7-0), containing X-gal and IPTG as recommended by the supplier (New England BioLabs). Each of the resulting colourless plaques was inoculated into 3 ml YT broth containing 30 p1 of an overnight culture of *E. coli* JM101. These cultures were incubated at 37°C with shaking at 400 r.p.m. for approximately 5 h. The bacteria were then pelleted by centrifugation in a microfuge at 12000 r.p.m. for 15 min. From the resulting supernatants, 2 p1 was removed and spotted onto a nylon membrane and the remainder was purified for DNA sequencing. The nylon membranes were dried at room temperature and used in a hybridization reaction with the $^{32}$P-labelled DNA probes (Church & Gilbert, 1984) to test for the presence of specific *B. pertussis* DNA inserts.

**DNA sequencing.** This was done by the dideoxy chain-termination method (Sanger *et al*., 1977) using reagents obtained from New England BioLabs. For sequencing inserts longer than approximately 300 bp, the Cyclone Biosystem (IBI) was used. Briefly, this system allows the construction of overlapping subclones from the DNA fragment cloned into the M13 vector through controlled deletions using T4 DNA polymerase. The final consensus sequence of the *B. pertussis* IS element was the result of sequencing each nucleotide at least twice in both directions.

**Computer analysis.** DNA sequence data were analysed using the programs of Conrad & Mount (1984). The DNA sequences were compared using the FASTN program (Lipman & Pearson, 1985), which searches the GenBank library for DNA homologies.

**RESULTS**

**Identification of the repeating DNA element in the genome of B. pertussis**

We have recently been studying the genome of *B. pertussis* to identify a fragment of DNA suitable for use as a pertussis-specific DNA probe. For these studies, we have used cosmid p11-11, which is one of a series of cosmid clones containing *B. pertussis* DNA generated by Dr Alison Weiss. Cosmid p11-11 was digested with *Clai* and an 1100 bp fragment was isolated and subcloned into plasmid pBR325. This 1100 bp fragment, designated Cla-4, was labelled with $^{32}$P by nick-translation and hybridized to a Southern blot of *Clai* digests of genomic DNA from *B. pertussis*, *B. parapertussis* and *B. bronchiseptica* (Fig. 1). It is evident from Fig. 1 that the 1100 bp Cla-4 probe contains a sequence of DNA which is repeated many times throughout the *B. pertussis* genome. There are at least 40 fragments which hybridize to the Cla-4 probe in the *B. pertussis* Tohama strain. This 1100 bp fragment also hybridizes to DNA fragments from *B. parapertussis* and *B. bronchiseptica*, but there are fewer copies. There are nine fragments in *B. parapertussis* strains 501 and CN8234, ranging in size from 1 to 20 kb, which hybridize to the Cla-4 probe, and these fragments show the same chromosomal distribution in both strains.
Repetitive DNA element in *B. pertussis*

Fig. 1. Southern blot of ClaI-digested genomic DNA from *Bordetella* species hybridized to the 1100 bp Cla-4 probe, showing the presence of the repeating sequence. Lanes: A, ClaI digest of cosmid p11-11, with arrows marking the 1100 bp Cla-4 probe and the 5 kb Cla-3 fragment which also hybridizes to the Cla-4 probe; B, *B. pertussis* Tohama; C, *B. parapertussis* CN8234; D, *B. parapertussis* 501; E, *B. bronchiseptica* 213; F, *B. bronchiseptica* 361; G, *B. bronchiseptica* 469.

*B. bronchiseptica* strains 361 and 213 each contain one band of approximately 6 kb which hybridizes to the Cla-4 probe. There is no hybridization of the Cla-4 probe to DNA from *B. bronchiseptica* strain 469.

Isolation of the repeating element

To identify the complete repeating element, it was necessary to have at least two different copies of it for sequence comparison. This enabled the boundaries of the element to be accurately determined. Since the 1100 bp Cla-4 fragment hybridizes to a 5 kb fragment also obtained from the ClaI digest of cosmid p11-11, this indicates that the 5 kb fragment is likely to contain at least one copy of the repeating element (Fig. 1). A comparison of the 5 kb fragment (designated Cla-3) and the Cla-4 sequences should thus identify the repeating element. To test this hypothesis, the entire Cla-4 fragment was cloned into M13mp19 and sequenced. Restriction
site analysis of the Cla-4 sequence shows two internal SalI sites and one internal SphI site. Because the SalI and SphI sites were believed to be within the repeating element, these two enzymes were used to digest the 5 kb Cla-3 fragment and thus increase the likelihood of obtaining matching sequences. The fragments of Cla-3 obtained from these digests were shotgun cloned into the SalI and SphI sites of M13mp19 and the resulting colourless plaques which hybridized to the Cla-4 probe were picked and sequenced. From a comparison of the sequence data, the 5' terminus of the repeating element was determined and it showed that there are approximately 465 nucleotides of the Cla-4 fragment which precede the start of the repeating element.

**Determination of the 3' terminus of the repeating element**

Sequence comparison of the Cla-3 and Cla-4 restriction fragments was not adequate to identify the entire repeating sequence, due to the presence of a ClaI site in the middle of the repeating element (Fig. 2). To obtain the sequence of the repeat to its 3' terminus, the cosmid p11-11 was digested with SalI. The SalI fragments were shotgun cloned into M13mp19 and the resulting colourless plaques were screened by plaque hybridization to a 460 bp SalI–ClaI fragment isolated from the 1100 bp Cla-4 probe (Fig. 2, fragment D). This 460 bp fragment was chosen for a probe since it contains DNA only within the repeating element. Using this method, three clones were identified and sequenced and the 3' terminus of the repeating element was determined from the termination of sequence homology among the three fragments.

**Sequence analysis**

The entire DNA sequence of the repeating element identified in *B. pertussis* is shown in Fig. 3. The size of the repeating element is 1053 bp and its termini (underlined in Fig. 3) consist of 28 bp of near-perfect inverted repeats. These termini may hybridize to form the double-stranded stem-loop structure shown in Fig. 4. This repetitive DNA element has a G + C base content of 63.7% (19.1% A, 17.2% T, 28.1% G, 35.6% C), which is consistent with the high G + C content reported for *B. pertussis* DNA (Locht & Keith, 1986). The terminal inverted repeats of the *B. pertussis* repetitive DNA element have a much higher A + T base content (64.3% A + T) than the rest of the element. This is similar for other IS elements, where the ends of the element contain a 3–13% higher percentage of A–T base pairs than the elements themselves (Kleckner, 1981).

**Analysis of the target DNA flanking the insertion site**

Another characteristic of bacterial insertion sequence elements is that they form short duplications in the target DNA at the insertion site (Calos & Johnsrud, 1978; Calos & Miller, 1980). To determine whether this occurred with the *B. pertussis* repetitive DNA element, we isolated one entire copy so that both the terminal inverted repeats as well as the flanking DNA
Repetitive DNA element in *B. pertussis*

1

CTAGGTGTGAGATTTCAAATTGGTTGATAGCATGTTCTACGCAACCAGATTGTGAGAAACT

61

GGGAATCGCACCACACCCGAGTCTAGCAGGACCACGGGGGATGAACACCTATAGAT

Sall

121

GCCCCATGACCTCTATTGAGTCTGCAATCGCAATAGTGATACCACCACTAGAATTT

BstEII

181

TGTTGCTGCAAGCGCGGCTAGAAGTTGCTCTGATGAGCTTACGGCTGCGGCAAGG

241

GGCGGTCTGCTGCTAGGCGGCGGCTGCTGCTGGTCAGGGCCAGGCGCTC

301

TGGCCGTCGCGGATTTCTGACTGAGTCTGGGATGGCTCTGCGCGGAGGG

361

CTGACCACAGGCGCAGTCCCGCCAAGGGGCTGCTGCTGATAGGTGACACCCGACGG

421

CTGGGCCGCGCGCGCCCTCTGCTGACCTGCGCAAGCCGCGGCGGCGGCGGCTG

481

TACGACATCGGCGCCGCTGACTGACATCCAGACAGTGAGCTGAGCGTAT

541

CAGCGCGGGTGGCCGCGGCTGATACCCTGAGGAGGAGGAGGGCGGCTG

ClaI

601

GACTTCGCTTTTGGCAAGCCCGCGATGCGCCGCGGCGGCGGCTG

661

GACGAGCCTTTCGCTGCTGCTGATTCAAGGGGCTGCCTGATCTCGAGCAGG

721

CTGGCCGGTCAGCCCTGGGATGGGCTGCTGCTGCTGCTGCTGCTG

781

TTCGGCGGGTGGGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG

841

CAGCGACATGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG

901

CACACCTACAGAAAACTCCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG

961

AATGCGAGAGGAGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG

1021

GACGAAATACAGCCTATGGAGATCTCCAGCAGACTGAGACTG

Fig. 3. Complete nucleotide sequence of the putative *B. pertussis* IS element. The sequence is written 5'→3', with the relevant restriction sites underlined and labelled. The 28 bp terminal inverted repeats are underlined and mark the boundaries of the element.

could be examined. Since restriction site analysis shows that there are no internal BamHI or EcoRI sites, cosmid p11-11 was digested with BamHI and a 3.5 kb fragment containing one copy of the repeating element was cloned into M13mp19 and sequenced. As shown in Fig. 4, the target DNA flanking the *B. pertussis* repeating element was not duplicated during the integration process.

Absence of the *B. pertussis* repetitive DNA element in *B. parapertussis* and *B. bronchiseptica*

Since the Cla-4 probe used in the hybridization studies with genomic DNA of *Bordetella* species contains DNA from the repetitive sequence as well as approximately 465 bp of adjacent DNA, it was unclear whether the repetitive element itself is present in all of the *Bordetella* species or whether the flanking *B. pertussis* DNA is responsible for the hybridization to *B. parapertussis* and *B. bronchiseptica* DNA. To distinguish between these possibilities, a 365 bp BstEII–ClaI fragment internal to the *B. pertussis* repetitive sequence (Fig. 2, fragment C), and a 440 bp HpaI–ClaI fragment outside the repetitive sequence (Fig. 2, fragment B), were isolated
from the 1100 bp Cla-4 fragment (Fig. 2, fragment A). These two different fragments were hybridized to identical Southern blots containing ClaI digests of genomic DNA from B. pertussis, B. parapertussis and B. bronchiseptica. The results, shown in Fig. 5, indicate that fragment C, internal to the B. pertussis repetitive element, is specific for B. pertussis and is not found in B. parapertussis. A 6 kb fragment of B. bronchiseptica DNA hybridizes to fragment C (Fig. 5a, lanes D and E), but this fragment does not appear to be a copy of the repetitive DNA element due to its absence when different internal fragments of the element are hybridized to B. bronchiseptica DNA (data not shown). The HpaI–ClaI fragment B, external but adjacent to the B. pertussis repetitive DNA sequence, hybridizes to nine fragments in B. parapertussis, four fragments in B. pertussis and none in B. bronchiseptica (Fig. 5b). This fragment accounts for the hybridization to B. parapertussis DNA seen with the Cla-4 probe (Fig. 1) and suggests that there is another, different repeating sequence in B. parapertussis with a few copies in B. pertussis. Further characterization of this fragment will determine whether it is also a bacterial IS element.

**Computer analysis of the B. pertussis repetitive DNA sequence**

The B. pertussis repetitive DNA sequence was examined for nucleotide homology with other known prokaryotic transposable elements (including IS elements) and no significant homology was found. The DNA sequence was also translated in all six reading frames to identify potential
Repetitive DNA element in *B. pertussis*

**DISCUSSION**

We have identified and sequenced a segment of DNA from *B. pertussis* that is 1053 bp long and is found in multiple copies in the *B. pertussis* genome. This sequence has properties consistent with bacterial IS elements, including its size, multiple copy number and presence of 28 bp terminal inverted repeats. It does not appear that the integration of this element into the *B. pertussis* genome results in a short duplication of the target DNA at the insertion site.

Duplication of the target DNA at the IS insertion site is a common feature of IS elements and the only other IS element known to lack this property is IS91 (Diaz-Aroco *et al.*, 1987). A
comparison of IS91 and the *B. pertussis* repetitive DNA sequence shows that the ends of both elements are defined by restriction sites. There are *Taq*I sites (TCGA) at the ends of IS91 and *Mae*I sites (CTAG) at the ends of the *B. pertussis* element. It appears that having restriction sites at the ends of an IS element is novel, and that it may influence the site of insertion as well as the mechanism by which the element inserts. It is possible, however, that these restriction sites are actually part of the target DNA rather than the IS element. Since these restriction sites are palindromes they could either hybridize to each other to be part of the terminal inverted repeats of the IS element, or could be the direct repeats (target DNA duplications) normally formed upon insertion of the IS element. If these restriction sites were part of the target DNA rather than the IS element, it would suggest that IS91 and this *B. pertussis* IS element insert specifically at *Taq*I or *Mae*I sites in the target DNA. Diaz-Aroco et al. (1987) have shown, though, that the *Taq*I sites are actually part of the IS91 sequence and the IS91 does not insert specifically at *Taq*I sites in the target DNA. They suggest that the lack of target DNA duplications with insertion of IS91 may be explained by the method of insertion. Currently it is hypothesized that IS elements insert by creating a staggered cleavage in the target DNA at the insertion site. The short target DNA duplications would form by repair of the staggered ends for ligation with the IS element. Diaz-Aroco et al. (1987) suggest that transposition of IS91 may produce a blunt rather than a staggered cleavage in the target DNA, and thus would not produce the duplications usually seen.

The similarity between the ends of IS91 and the putative *B. pertussis* insertion element suggests that their mechanisms of integration into the target DNA may be the same. Further research into the actual transposition of this *B. pertussis* DNA element should clarify the role, if any, of these terminal restriction sites in the transposition of this insertion sequence.

Another property of the *B. pertussis* repetitive DNA element is that it is found in many copies in the *B. pertussis* genome. This large copy number is uncommon for IS elements in general but has been reported for *Shigella* species, which have 30–40 copies of insertion element IS1 (Nyman et al., 1981). We also report the absence of this repetitive DNA element in *B. parapertussis* and *B. bronchiseptica* genomic DNA. These results are similar to those of McPheat & McNally (1987), who reported a single band of 1.45 kb in genomic digests of *B. parapertussis* strain BPAH1 and two bands in genomic digests of *B. bronchiseptica* strain BBRH1 which hybridized to a *Bam*HI fragment probe that contained a repeating sequence of *B. pertussis* DNA. This hybridization to *B. parapertussis* and *B. bronchiseptica* DNA is most likely due to the presence of DNA external to the repeating sequence and therefore suggests that we have isolated the same repetitive DNA sequence. Also, the size of approximately 1.0 kb and presence of one internal *Cla*I site reported by McPheat & McNally (1987) for the repeating sequence they identified is consistent with our results. However, the fact that we have identified another repeating sequence present in nine copies in *B. parapertussis* and four copies in *B. pertussis* suggests that more than one kind of repetitive DNA sequence may be present in *Bordetella* species. Other bacteria, such as *E. coli*, possess several different types of IS elements (Calos & Miller, 1980). It is also possible that different strains of *B. pertussis* have different copy numbers of this repeating sequence. We have tested eight different clinical isolates of *B. pertussis*, and have found them all to contain approximately the same number of copies of the repeating sequence (data not shown).

None of the three strains of *B. parapertussis* or four strains of *B. bronchiseptica* that we have examined contain the *B. pertussis* insertion sequence. Study of additional strains of *Bordetella* species will determine whether there is consistency in the copy number of the repeating element as well as identify other possible repetitive DNA sequences.

The exact function of IS elements is unknown, but they appear to play an important evolutionary role by mediating chromosomal rearrangements and altering expression of adjacent genes, which may help organisms adapt to new environmental conditions (Arber, 1983; Syvanen, 1984). Examination of *Bordetella* species by multilocus enzyme electrophoresis has shown *B. bronchiseptica* to be the ancestral strain from which *B. parapertussis* and *B. pertussis* evolved and also that *B. parapertussis* is more closely related to *B. bronchiseptica* than to *B. pertussis* (Musser et al., 1986). If the putative IS element is indeed involved in evolutionary processes, then it is interesting to note the large copy number of this element in the most evolutionarily advanced species. This suggests that the element plays an important functional
Repetitive DNA element in B. pertussis

role for B. pertussis and may be aiding in its adaptation to new environmental conditions by controlling such things as phase variation and thus production of virulence determinants. Transposable elements have been found to activate gene expression in Pseudomonas cepacia (Scordilis et al., 1987) and E. coli, and to inactivate diphtheria toxin production in Corynebacterium diphtheriae (Rappuoli et al., 1987). Therefore, it is conceivable that this element is involved in gene expression in B. pertussis.

Even though the exact function of this IS element is unknown, its properties suggest that it might be useful for studying the epidemiology of pertussis infection. The large copy number of this element indicates that it would be a sensitive diagnostic probe for detecting B. pertussis organisms in clinical specimens. In fact, we find that the 32P-labelled Cla-4 probe, which contains a portion of the B. pertussis IS element as well as the repeating sequence found in B. parapertussis, is able to detect 103 B. pertussis organisms and 105 B. parapertussis organisms filtered onto nylon membranes (McLafferty et al., 1986). The fact that this repetitive DNA sequence may be species specific, as evidenced by its lack of significant homology with other known bacterial DNA sequences, suggests that it would be a specific as well as a sensitive tool for use in diagnosis of B. pertussis infection. We propose to designate this newly characterized B. pertussis IS element IS481 (Lederberg, 1987).

The authors gratefully acknowledge the assistance of the following individuals in this work: Miss Cynthia Kogut, and Mr Jonathan Skinner for preparation of the manuscript; Dr Pavel Novotny for supplying Bordetella strains; Dr Alison Weiss for providing cosmid pl1-11; and Dr Kevin Lynch and Dr Robert Kadner for important discussion and suggestions. This work was conducted with the support of a contract from the Centers for Disease Control (DHHS 200-84-0784), a grant from the National Institutes of Health (AI18000) and the facilities of the University of Virginia Diabetes Center Molecular Biology Core Laboratory (supported by NIH grant AM22125).

REFERENCES


