

## Genomic Loci of the *Porphyromonas gingivalis* Insertion Element IS1126

HONG DONG,<sup>1</sup> TSUTE CHEN,<sup>1</sup> FLOYD E. DEWHIRST,<sup>1</sup> ROBERT D. FLEISCHMANN,<sup>2</sup>  
CLAIRE M. FRASER,<sup>2</sup> AND MARGARET J. DUNCAN<sup>†1\*</sup>

*Department of Molecular Genetics, The Forsyth Institute, Boston, Massachusetts 02115,<sup>1</sup> and  
The Institute for Genomic Research, Rockville, Maryland 20850<sup>2</sup>*

Received 9 December 1998/Returned for modification 23 February 1999/Accepted 16 April 1999

The *Porphyromonas gingivalis* genome contains multiple copies of insertion element IS1126. When chromosomal DNA digests of different strains were probed with IS1126, between 25 and 35 hybridizing fragments per genome were detected, depending on the strain. Unrelated strains had very different restriction fragment length polymorphism (RFLP) patterns. When different laboratory copies of a specific strain were examined, the IS1126 RFLP patterns were very similar but small differences were observed, indicating that element-associated changes had occurred during laboratory passage. Within the next year, genome sequencing, assembly, and annotation for *P. gingivalis* W83 will be completed. Because repetitive elements complicate the assembly of randomly sequenced DNA fragments, we isolated and sequenced the flanking regions of IS1126 copies in strain W83. We also isolated and sequenced the flanking regions of IS1126 copies in strain ATCC 33277 in order to compare insertion sites in phylogenetically divergent strains. We identified 37 new sequences flanking IS1126 from strain ATCC 33277 and 30 from strain W83. The insertion element was found between genes except where it transposed into another insertion element. Examination of identifiable flanking genes or open reading frames indicated that the insertion sites were different in the two strains, except that both strains possess an insertion adjacent to the Lys-gingipain gene (J. P. Lewis and F. L. Macrina, *Infect. Immun.* 66:3035–3042, 1998). Most of the genes or sequences flanking IS1126 in ATCC 33277 were present in W83 but were contiguous and not insertion element associated. Thus, where genes were identified in both strains, their order was maintained, indicating that the two genomes are organized similarly, but the loci of IS1126 are different. In both strains, insertion element-associated duplicated target sites were lost from several copies of IS1126, providing evidence of homologous recombination between elements. Larger organizational differences between the genomes, such as deletions and inversions, may result from insertion element-mediated recombination events.

*Porphyromonas gingivalis* is a gram-negative oral anaerobe associated with adult periodontitis. The bacterium possesses many potential virulence factors: adhesins, potent proteases, hemolysins, and an iron acquisition system. Considerable effort is expended on the identification and inactivation of putative virulence genes to determine their role in pathogenesis. These studies have revealed the existence of at least three insertion sequences (IS) in the *P. gingivalis* genome: IS1126 (14), PGIS2 (23), and most recently IS195 (12).

Plasmid vectors developed for colonic *Bacteroides* spp. have proved invaluable for genetic studies with *P. gingivalis*, and IS1126 was discovered after an *Escherichia coli*-*Bacteroides* shuttle vector was recovered from *P. gingivalis* and found to contain extra DNA identified as an IS element (15). Sequencing revealed that the element encoded a putative transposase, was named IS1126, and assigned to the IS5 subgroup of the IS4 family of elements (15). The number of copies of IS1126 in several strains has been estimated to be 6 to 12 (15, 22), and the element has been encountered with increasing frequency as more *P. gingivalis* genes are cloned. It was present in *E. coli* clones of a *P. gingivalis* library identified after screening for epithelial cell attachment (7), and a truncated copy was located 3' to the Lys-gingipain (porphypain) gene in several strains (12).

There were approximately 25 to 30 IS1126 elements in the laboratory strains we surveyed. While restriction fragment length polymorphism (RFLP) patterns of IS1126 were very similar in copies of strains we received from different laboratories, there were one or two band differences indicative of IS1126-mediated changes in the genomes. Within the next year, the genomic sequence of *P. gingivalis* W83, a virulent strain, will be completed. The presence of numerous repetitive regions complicates the assembly of randomly sequenced fragments. To facilitate this process, we have sequenced flanking regions of IS1126 from strains ATCC 33277 and W83 after direct cloning or inverse PCR, allowing a comparison of the organization of the two genomes with respect to IS1126. DNA sequences flanking the element were different in the two strains, and although most ATCC 33277 IS1126 flanking sequences were found in the W83 contig database, they were not directly associated with IS. Our results indicate that in some regions, the two genomes are organized similarly but the loci of IS1126 are different. In both ATCC 33277 and W83, there was evidence of IS1126-mediated recombination and the homology provided by IS and other repeat regions may be responsible for larger differences in genome organization, such as deletion and inversion of intervening sequences.

\* Corresponding author. Mailing address: Department of Molecular Genetics, The Forsyth Institute, 140 Fenway, Boston, MA 02115. Phone: (617) 262-5200, ext. 344. Fax: (617) 262-4021. E-mail: mduncan@forsyth.org.

### MATERIALS AND METHODS

**Bacterial strains, growth conditions, and isolation of genomic DNA.** The *P. gingivalis* strains used in this study are listed in Table 1 and were grown on blood agar plates as described previously (8). Genomic DNA was extracted from 72-h

TABLE 1. *P. gingivalis* strains used in this study

Strain	Laboratory stock collection	Origin of oral infection <sup>a</sup>	Site of isolation <sup>a</sup>
W50 FDC	The Forsyth Institute, Boston, Mass.	Unknown	Unknown
W50 BEI <sup>b</sup>	Same as above		
381 FDC	Same as above	Periodontitis	Subgingival
ATCC 33277 FDC	Same as above	Periodontitis	Subgingival
W50 OMGS	P. Papapanou, Department of Oral Microbiology, Faculty of Odontology, Goteborg University, Goteborg, Sweden	Unknown	Unknown
381 K	H. Kuramitsu, Department of Oral Biology, State University of New York, Buffalo	Periodontitis	Subgingival
HG66M	Christian Mouton, Groupe de Recherche en Ecologie Buccale, Faculté de Médecine Dentaire, Université Laval, Quebec, Canada	Unknown	Unknown
W83M	Same as above	Unknown	Unknown
W83G	Genevieve Gallagher, Department of Oral Biology, Indiana University School of Dentistry, Indianapolis	Unknown	Unknown
HG66vW	A. van Winkelhoff, Department of Oral Microbiology, ACTA, Vrije Universiteit, Amsterdam, The Netherlands	Unknown	Unknown

<sup>a</sup> As listed in reference 17.

<sup>b</sup> A spontaneous nonpigmented mutant derived from W50 (16).

cultures (11) extracted three times with phenol-chloroform-isoamyl alcohol, ethanol precipitated, and finally resuspended in Tris-EDTA buffer.

In the present study, all DNA sequence comparisons were made between the *P. gingivalis* type strain, ATCC 33277, and strain W83M, obtained from C. Mouton, Laval University, Quebec, Canada. The genome of W83M is being sequenced at The Institute for Genomic Research.

**Southern hybridization.** To determine the copy number of IS1126 in each strain, genomic DNA was digested with *Pst*I (approximately 5 U of enzyme/ $\mu$ g of DNA) overnight at 37°C and fragments were fractionated by agarose gel electrophoresis. After transfer to nylon membranes, genomic blots were probed with the 0.86-kb *Ban*I-*Sma*I fragment from an intact copy of IS1126 isolated from strain ATCC 33277. Hybridization conditions and signal development were as recommended in the enhanced chemiluminescence (ECL) kit from Amersham-Pharmacia.

**Pulsed-field gel electrophoresis.** Genomic DNA was extracted from agarose-embedded log-phase bacteria by standard procedures. *Xba*I digestion was done with 40 U of enzyme/plug of DNA at 37°C overnight, and *Spe*I digestion was done with 4.5 U of enzyme/plug of DNA at 37°C for 2 h. Pulsed-field gels were run on a Bio-Rad CHEF DRII for 25 h at 14°C and 200 V with an initial switch time of 6.75 s to a final switch time of 26.29 s with linear ramping. Fragment sizes were estimated by comparison with the mobility of lambda concatemers.

**Library construction.** Most of the IS1126 flanking sequences from *P. gingivalis* ATCC 33277 and all those from strain W83M were isolated after screening of genomic libraries. Genomic DNAs were partially digested with *Sau*3A1, and 2- to 5-kb fragments were isolated after size fractionation on agarose gels. Genomic fragments were ligated with *Bam*HI and bacterial alkaline phosphatase-treated pUC18 (Amersham-Pharmacia), and ligation mixtures were transformed into *E. coli* DH5 $\alpha$  (Gibco BRL) with selection for ampicillin resistance (50  $\mu$ g/ml) on Luria-Bertani plates. Colonies were lysed for hybridization (4) and probed with the 0.86-kb *Ban*I-*Sma*I fragment from IS1126. Plasmid DNA was isolated from positively reacting clones, and IS flanking regions were sequenced by using vector or inverse PCR primers.

**Inverse PCR.** Several sequences flanking IS1126 from strain ATCC 33277 sequences were derived by inverse PCR. Genomic DNA was digested with enzymes which cut outside IS1126, and fragments were fractionated in agarose gels to estimate size ranges. *Pst*I and *Hae*III yielded suitable-size fragments. DNA digests were diluted to less than 0.5  $\mu$ g/ml, and fragments were self-ligated with T4 DNA ligase. The inverse PCR primers used were E81 (5'ATGCCATGGAGGAATAT3'), E83 (5'CGTTTTGTGGTTTGCGATA3'), E86 (5'ATCTTGAAAGCATCGCCT3'), and E88 (5'GATTTACAACACTTTCCTC3') (see Fig. 1). PCRs were carried out in a total volume of 100  $\mu$ l containing 10  $\mu$ l of 10 $\times$  PCR buffer (Perkin-Elmer, Foster City, Calif.), 25 ng of self-ligated template DNA, each primer at 0.4  $\mu$ M, each deoxynucleoside triphosphate at 200  $\mu$ M, 2.5 mM MgCl<sub>2</sub>, and 2.5 U of AmpliTaq Gold. Amplifications were carried out in a PE 480 Thermocycler for a total of 30 cycles.

**DNA sequencing.** Sequencing reactions were carried out with either dRhodamine or Big Dye Terminator cycle sequencing kits (Perkin Elmer) by using a PE 9700 Thermocycler. Reactions were run on an ABI 377 Sequencer.

## RESULTS

**Structure of IS1126.** Figure 1 shows a schematic diagram of IS1126, which is 1,338 bp in length and is bounded by 12-bp perfect inverted repeats. The original IS1126 sequence from strain W83 (15) contained several frameshifts within the open

reading frame (ORF) of the putative transposase. The ORF encodes a protein of 346 amino acids and a molecular mass of 41 kDa. Analysis of IS1126 sequences obtained from ATCC 33277 and W83M by either cloning or inverse PCR showed that approximately 60% of the inverted repeats from ATCC 33277 and 41% of those from W83M were imperfect, with mismatches always occurring in the 3' sequence, and in most cases, the change was from G to T at the second base of the inner domain of the inverted repeat, as depicted in Fig. 1.

**RFLP analysis of IS1126 in *P. gingivalis* strains.** To assess the relative stability of IS1126, copies of frequently used strains of *P. gingivalis* were collected from several laboratories (Table 1). The assumption was that differences in the handling of laboratory strains, i.e., the number of passages, the age of stock cultures, and the frequency of strain revival, might induce IS-mediated genomic changes, resulting in different IS1126 RFLP patterns. Of several enzymes which did not cut within IS1126, *Pst*I was chosen for genomic digests because it gave the widest range of fragment sizes. *Pst*I-digested chromosomal DNA was fractionated by electrophoresis through long agarose gels to optimize fragment separation. DNA was transferred to nylon membranes and probed with the internal 0.86-kb *Sma*I-*Ban*I fragment from IS1126. Hybridization patterns are shown in Fig. 2. Between 25 and 35 bands hybridized with the probe, depending on the strain. Because IS1126 does not contain a *Pst*I site, it could be assumed that each hybridizing band con-

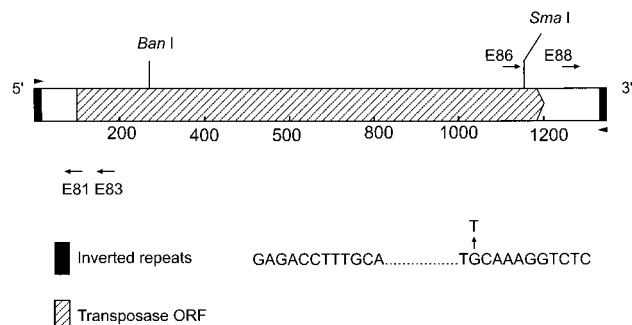


FIG. 1. Schematic of IS1126. Shown are the locations of the restriction enzyme sites and the PCR primers (E81, E83, E86, and E88) used in this work. Also shown are the 12-bp inverted repeats and the G-to-T transversion found in several copies of the element.

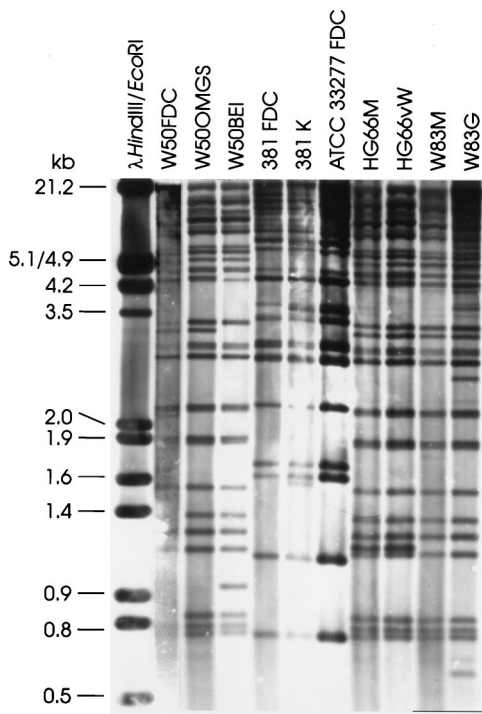


FIG. 2. RFLP analysis of *IS1126* in laboratory strains of *P. gingivalis*.

tained a single copy of the element. However, some bands may be doublets and high-molecular-weight bands may contain more than one copy of the IS; thus, the number of hybridizing bands might be underestimated. Conversely, although every effort was made to ensure that genomic DNAs were digested to completion, partial digestion would lead to an increased number of hybridizing bands. In addition, from the alignment of several *IS1126* sequences, e.g., 3' to the *prtP* gene from strains W12 and W83 (2, 12), it appears that not every copy of the element is intact. Thus, residual fragments smaller than *IS1126* would still hybridize with the probe.

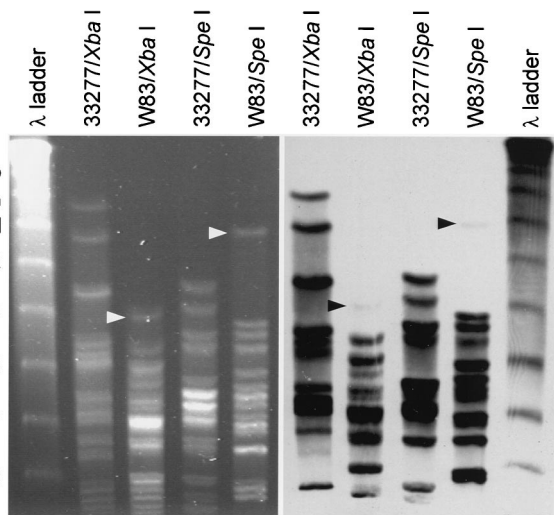


FIG. 3. Pulsed-field gel electrophoresis of large genomic fragments from strains ATCC 33277 and W83M. Left, ethidium bromide-stained gel; right, Southern blot probed with the *SmaI-BanI* fragment from *IS1126*.

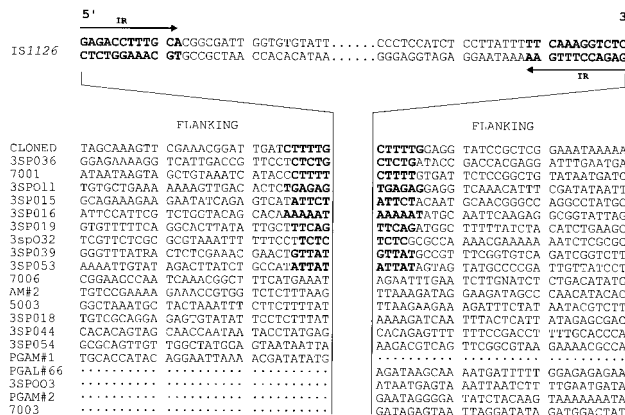


FIG. 4. Sequences flanking *IS1126* from *P. gingivalis* ATCC 33277. From each clone, 30-bp flanking sequences are shown 5' and/or 3' of *IS1126*, which is depicted centrally in abbreviated form. IR, inverted repeat.

Copies of the same strain from different laboratories showed very similar RFLP patterns; however, one or two band differences were observed. This result indicates that although *IS1126* was relatively stable under standard culture conditions, either transposition or recombination between IS copies had occurred during laboratory passage. Strain W50BEI is a spontaneous nonpigmented, avirulent mutant which was isolated from W50 after 49 to 73 generations of growth in a chemostat (16). Compared with its parent, W50BEI lost an *IS1126* hybridizing band of approximately 3.3 kb and gained a band at 0.96 kb. However, RFLP patterns differed between strains, indicating that the loci of *IS1126* were different.

**Distribution of *IS1126* throughout the *P. gingivalis* genome.** *XbaI*- and *SpeI*-digested large genomic fragments of ATCC 33277 and W83M were separated by pulsed-field gel electrophoresis, transferred to membranes, and probed with the *SmaI-BanI* fragment from *IS1126* to determine distribution around the genomes. The results are shown in Fig. 3A and B. In both strains, the probe hybridized to most of the ethidium bromide-staining bands, indicating that *IS1126* copies were not clustered in specific regions of the genome. However, extremely weak hybridization was observed with the largest *XbaI* and *SpeI* fragments of W83M (estimated at approximately 169 and 302 kb, respectively).

**Flanking sequences of *IS1126* in *P. gingivalis* ATCC 33277 and W83M.** Thirty base pairs of flanking sequence adjacent to

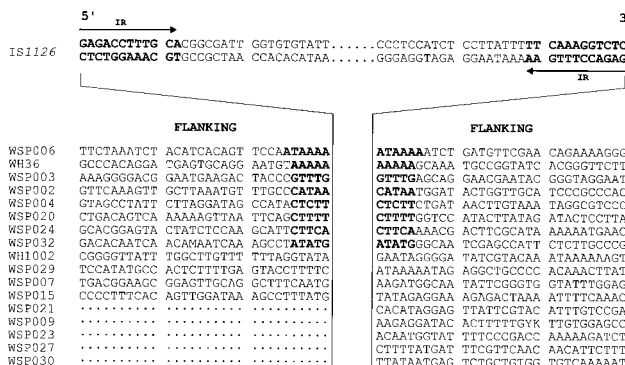


FIG. 5. Sequences flanking *IS1126* from *P. gingivalis* W83M. For details, see the legend to Fig. 4.

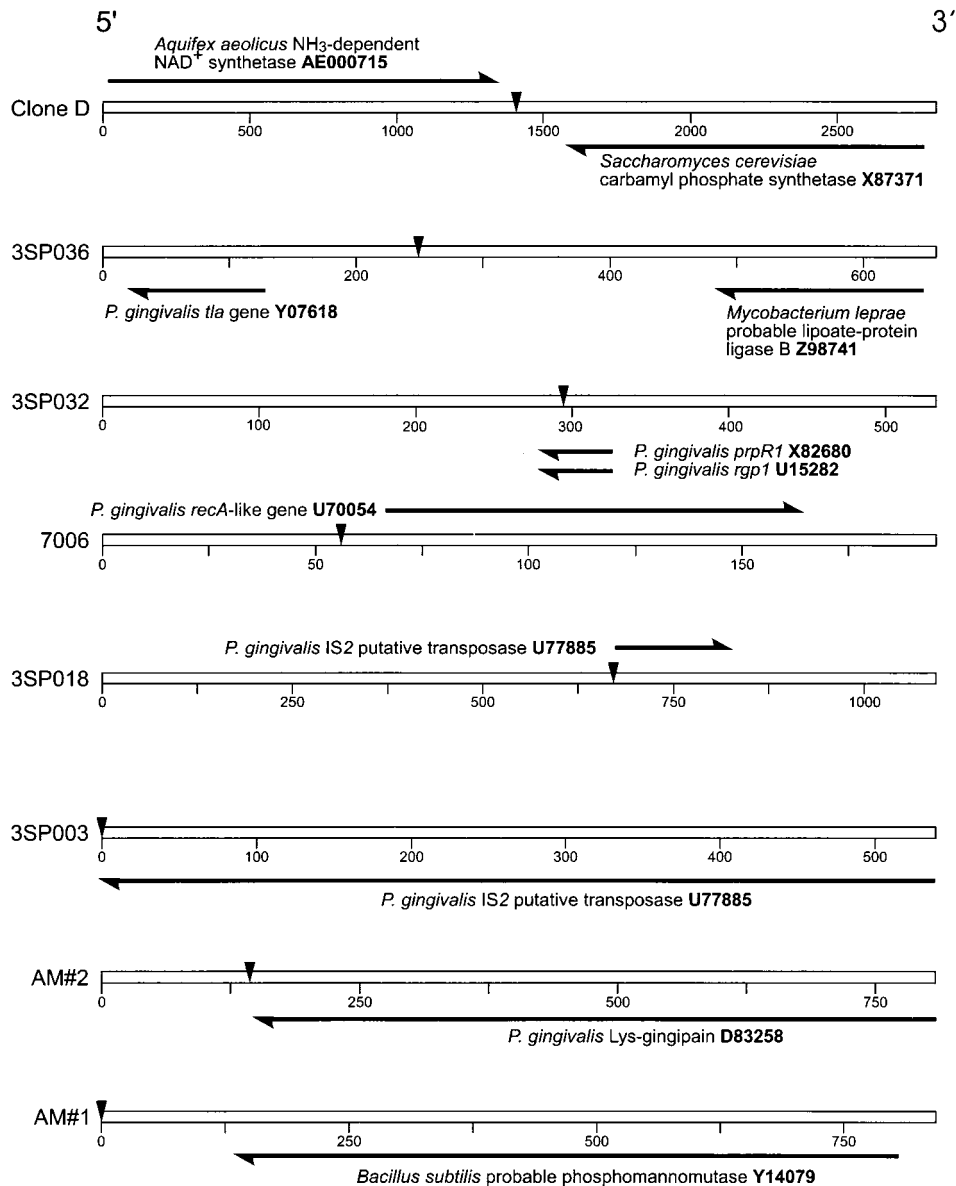


FIG. 6. Nearest neighbor ORF of IS1126 from *P. gingivalis* ATCC 33277. The open bar represents the total flanking sequence, and the markers are in base pairs. IS1126 is depicted as the inverted arrowhead, and its sequence (1,338 bp) has been deleted from the schematic. Homologous genes and ORFs are labeled and represented as solid arrows, which are drawn to scale. Also shown is the accession number of each homologous sequence.

the 5' and/or 3' ends of IS1126 from ATCC 33277 and W83M were obtained by either cloning or inverse PCR and are shown in Fig. 4 and 5. Like many insertion elements, IS1126 integrated at relatively AT-rich sites (51 to 75% AT). In more than half of the cases (18 of 28), the pairing of flanking sequences with a specific copy of the IS could be confirmed because of the presence of a duplicated target site (4 to 6 bases in length and shown in boldface). However, at least six copies from ATCC 33277 and four copies from W83M did not have target site duplications, indicative of past recombination events between copies of the IS which had led to genomic rearrangements, e.g., deletions and inversions; thus, the original pairing of the IS flanking sequences was lost. The 5' flanking regions for several clones, and one 3' sequence, have not been identified in our libraries.

When the 30-bp sequences flanking IS1126 in ATCC 33277

were compared with those obtained from W83M, no matches were found. We reasoned that possible sequence variation between strains might compromise the comparison of short sequences; therefore, similarity searches were extended by using the complete flanking sequences (100 to 500 bp) of the IS1126 copies we isolated. Using the BLAST algorithm (3), database homology searches were carried out with the extended sequences to identify the nearest neighbor genes and ORFs flanking IS1126 in both strains. Those with the highest homology scores are shown in Fig. 6 and 7. Again, no matches were observed between strains; however, the ORF encoded by the gene for Lys-gingipain (*kgp*; 19) was found 5' to IS1126 in ATCC 33277, as reported for W83 and several other strains (12). In ATCC 33277 and W83M, neighboring the IS were fragments of several genes: *rgp-1* (20), *prpR1* (1), *tla* (2), and *recA* (9). ATCC 33277 and W83M flanking sequences not

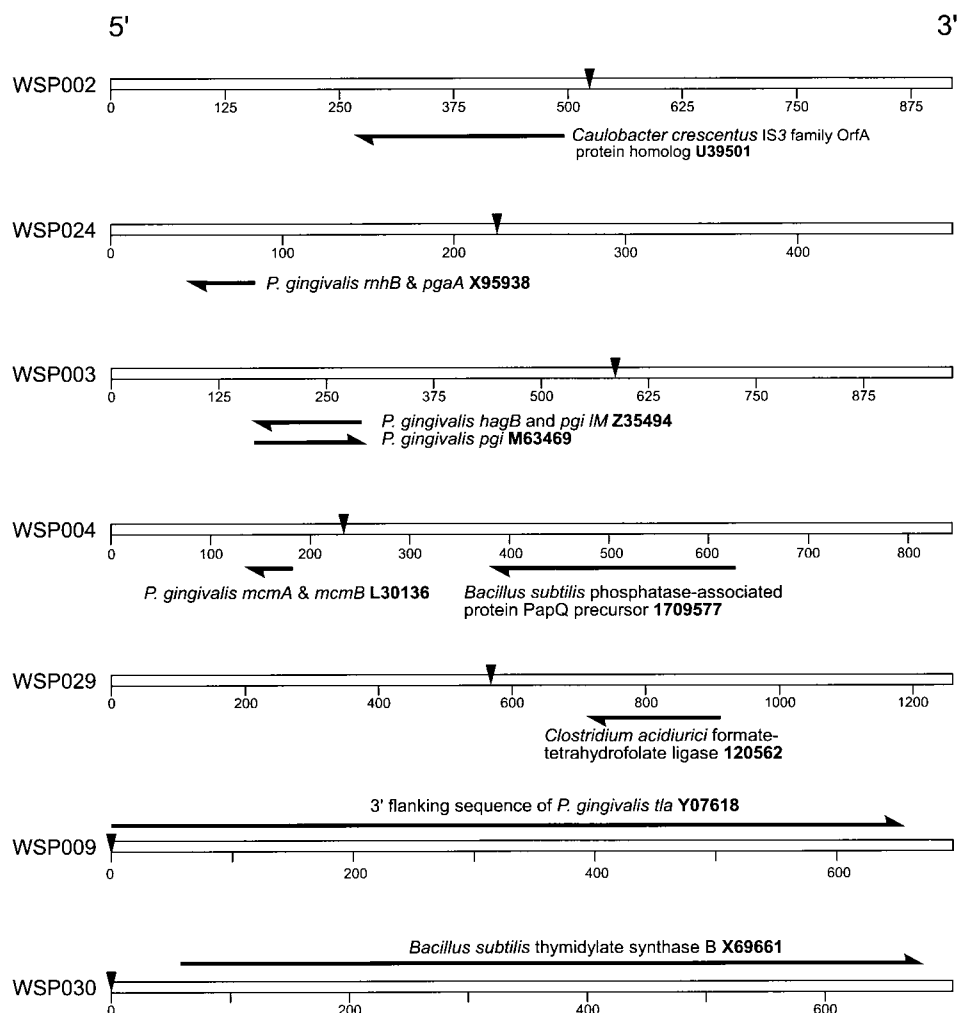


FIG. 7. Nearest neighbor ORF of *IS1126* from *P. gingivalis* W83M. For details, see the legend to Fig. 6.

shown in Fig. 6 and 7 did not contain homology to known genes or proteins.

**Homology searches of the W83M contig database with *P. gingivalis* ATCC 33277 *IS1126* flanking sequences.** The total *IS1126* flanking sequences obtained from ATCC 33277 were mapped to the W83M contig database (as deposited in July 1998 with the National Center for Biotechnology Information). In 8 of 10 cases in which the duplicated target site sequence was present, both the 5' and 3' flanking sequences from ATCC 33277 were found in W83; however, these sequences were contiguous and not disrupted by *IS1126* (e.g., classes I and II in Table 2). In some cases, e.g., clone 7001, there were multiple copies of the flanking sequences in the W83M database. Of these, two sequences were associated with *IS1126*, but it was located either before the 5' and/or after the 3' sequence and not between them, as in ATCC 33277. Two ATCC 33277 sequences with duplicated sites (3sp016 and 3sp053) could not be mapped to W83M. While it is possible that these are located within a physical or sequencing gap in the W83M genome, they may also be from a region which is unique to that of ATCC 33277. Several of the ATCC 33277 flanking sequences did not contain duplicated target sites, presumably because of prior recombinations. For two of these, 7006 and 3sp044 (class V; Table 2), the 5' and 3' sequences were not together in W83M



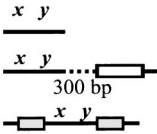



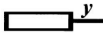
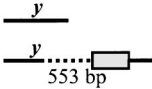

and were not associated with *IS1126*. Clone 3sp003 of ATCC 33277 contained the 3' flanking sequence of *IS1126* within PGIS2; i.e., *IS1126* had hopped into another IS. Thus, because the flanking sequence was PGIS2, it was also present in multiple copies in the W83M database.

## DISCUSSION

Originally, *IS1126* was assigned to the *IS4* family of IS. Two conserved sequence motifs within the transposase define the *IS4* family: a D-(1)-(G/A)-(Y/F) consensus core within the N-terminal region (N3) and a Y-(2)-R-(3)-E-(6)-K core in the C-terminal region (C1). A perfect N3 and a possible C1 consensus core were described (15). An alternate C1 core sequence, RALTEEEKQGNK, including amino acids 289 to 300 is also possible. The close proximity of the N3 and C1 sequences, together with the GAG external trinucleotides of the inverted repeats defined *IS1126* as a member of the *IS5* subgroup of the *IS4* family of insertion sequences (15, 21). Recently, this subgroup was reassigned as a family in its own right (13); thus, *IS1126* belongs to the *IS5* family of IS.

There have been previous estimates of the number of copies of *IS1126* in the *P. gingivalis* genome (15, 22). When W83 genomic DNA was digested with *Bam*HI, at least 12 poorly

TABLE 2. Results of homology searches of the W83 genome database with *P. gingivalis* ATCC 33277 IS1126 flanking sequences

ATCC 33277 clone	Class	Configuration <sup>b</sup> in:		Comment
		ATCC 33277	W83	
Clone D <sup>a</sup> 3SP036 <sup>a</sup> 3SP011 <sup>a</sup> 3SP015 <sup>a</sup> 3SP019 <sup>a</sup> 3SP032 <sup>a</sup> 3SP039 <sup>a</sup>	I			
7001 <sup>a</sup>	II			3 copies associated with other IS
3SP054	III			x not found
AM#2 PGAM#1	IV			y not found
7006 3SP044 3SP044	V			
3SP003	VI			4 copies associated with PGIS2
PGAL#66	VII			5 copies

<sup>a</sup> Clones with duplicated target sequences.

<sup>b</sup> The letter *x* corresponds to the 5' flanking sequences, and *y* corresponds to the 3' flanking sequences. Open boxes represent complete sequences of IS1126, and shaded boxes are incomplete sequences of IS1126. Dots represent extra W83 sequences found between the IS and the corresponding ATCC 33277 flanking region.

resolved hybridizing bands were observed (15). Digestion with this enzyme yields a smaller number of fragments than does digestion with *Pst*I; therefore, a single *Bam*HI fragment may contain more than one copy of IS1126. We were surprised at the large number of hybridizing bands we first observed in strain ATCC 33277 after *Pst*I digestion. When RFLP analyses were repeated with other strains, they too contained more than 25 *Pst*I fragments which hybridized with the IS1126 probe. Because the element was discovered after transposition into a shuttle vector, an immediate question was: how active is the element as a transposon? IS1126 RFLP patterns of copies of the same strain obtained from different laboratories were very similar, but different strains showed different patterns. Strain patterns are the result of millions of years of evolution, and IS1126 RFLP analysis may be a useful tool for strain typing and epidemiological studies. Strains 381 and ATCC 33277 had very similar RFLP patterns, despite being originally isolated at different geographic locations. Randomly amplified polymorphic DNA fingerprinting also showed that these strains are very closely related (17). In addition, strains W83 and HG66, showed very similar RFLP patterns; apparently, the same strain was assigned a different name in different laboratories (17). Single-band differences were observed between some copies of the same strain, indicating recent IS1126-mediated changes, i.e., transposition by either a replicative or "cut-and-

paste" mechanism or recombination between IS1126 sequences. For example, 1 of the 381 strains gained a band of approximately 1.6 kb, which could be the result of replicative transposition of IS1126. Interestingly, strain W50 BEI, a spontaneous, nonpigmented, avirulent mutant of W50, lost a hybridizing band of 3.3 kb and gained a band of 0.96 kb. This could be explained by either cut-and-paste transposition of IS1126 or recombination between copies of the element. Whether such events are responsible for the phenotypic changes observed with W50BEI is unknown.

In many of the IS copies we identified, the inverted repeats were imperfect and carried a mutation in the inner domain of the 3' repeat sequence. In IS903, also a member of the IS5 family, the transposase specifically recognizes and binds to this region in vitro and binding is sensitive to sequence changes (5). While we do not know if this base change similarly affects the transposase of IS1126 and reduces transposition frequency, a compound transposon containing the imperfect inverted repeat did transpose in *P. gingivalis* (6). Also, certain *P. gingivalis* DNA sequences deposited in GenBank contained adjacent IS1126 sequences from which part of the transposase ORF was deleted, precluding transposition (2, 12). Lastly, when IS1126 flanking sequences were used for database homology searches to identify the closest neighboring genes or ORFs, the IS was found to be inserted between these sequences, except when it

transposed into the ORF of another IS element. The mechanisms of *IS1126* transposition and regulation are unknown; however, it is possible that the element has long been associated with *P. gingivalis*, a level of stasis has been attained with the element confined to noncoding regions of the genome, and transposition of individual copies has been compromised by mutation and deletion.

As visualized by pulsed-field gel electrophoresis, *IS1126* appeared to be uniformly distributed around the genome of ATCC 33277. However, the large 169-kb *Xba*I and 302-kb *Spe*I fragments from W83M showed only background or low levels of hybridization with the IS probe, indicating that they did not contain the element. Given a genome size of about 2,500 kb and almost 30 copies of *IS1126*, there should be 1 copy of the element approximately every 83 kb; however, a region of up to 300 kb without an insertion could be expected by chance.

While chromosomal insertion of an IS element does not require host recombination functions, the host recognizes IS as substrates for homologous recombination. Loss of the duplicated target sequences flanking some *IS1126* copies was the first indication that they had been involved in previous recombination events (Fig. 3 and 4). This was confirmed by searching the W83M contig database with ATCC 33277 sequences which had lost the duplicated target sites. For example, the flanking sequences of ATCC 33277 clones 7006 and 3SP044 (Table 2) were present in W83M, but they were in separate contigs. These flanking sequences were probably once separate in ATCC 33277 also and associated with two different copies of *IS1126*. Recombination between these copies brought the sequences together as the flanking regions of the single recombinant IS. Whether such recombinations were accompanied by deletions or inversions of the intervening genetic material will become known when the genome sequences of ATCC 33277, W83M, and other *P. gingivalis* strains are compared.

None of the flanking regions of *IS1126* obtained from ATCC 33277 were identical to those flanking the element from W83M, and more extensive sequence searches for neighboring genes or ORFs associated with the IS identified only Lys-gingipain as a flanking sequence common to both strains. However, most of the ATCC 33277 flanking sequences were found in W83M contigs and in the same sequence order but without the IS. Among the possible explanations for this result is that either genome organization differs between the two strains or the two genomes are organized similarly but the locations of *IS1126* (and potentially other IS elements) are different; hence the different *IS1126* RFLP patterns observed with the two strains. Comparison of the sequences identified in this study indicates that in many regions the two strains are similar in genome organization. Differences in gene order could result from homologous recombination between IS elements and other repeat sequences.

IS elements play a role in the inter- and intraspecies spread of virulence factors between microorganisms. Recently, a copy of *IS1126* was identified flanking the *ragAB* locus of strain W50 (10). The lower G+C content of *ragAB* (42%) compared to *IS1126* and the *P. gingivalis* genome (46 to 48%) indicates that these genes were acquired relatively recently. Copies of *IS1126* flank the *kgp* and *tla* genes, putative virulence factors which also contain hemagglutinin domains with homology to that of *rgp-I*. The G+C content of these genes is similar to that of the *P. gingivalis* genome, suggesting they are older residents of the genome. Thus, if they were acquired by horizontal transfer, it was not a recent event. Duplication of the hemagglutinin domains may have resulted from intrachromosomal rearrangements involving gene conversion, as proposed by Okamoto et

al. (19), rather than from IS transposition. This type of recombination has previously been demonstrated in *P. gingivalis* (18).

It can be speculated that an ancestral strain of *P. gingivalis* contained a small number of copies of *IS1126* and that the original copy, together with the *kgp* gene, was acquired by horizontal transfer. Replicative transposition of *IS1126* would result in a copy remaining next to Lys-gingipain and an independent insertion at another locus. Eventually, independent copies would accumulate throughout the genome. Thus, although the genomic backbone is unchanged, the IS loci are different. Superimposed on IS transposition are homologous recombinations between these sequences and the resulting genome rearrangements. These two events could generate the separate lineages from which strains ATCC 33277 and W83 are derived.

#### ACKNOWLEDGMENTS

Thanks are due to M. Malamy, Tufts University Medical School for discussions on insertion sequences, and R. Goldstein, Boston University Medical School for use of PFGE equipment.

This research was supported by NIH grants DE 10510 (M.J.D.) and DE 12082 (R.D.F.).

#### REFERENCES

1. Aduse-Opoku, J., J. Muir, J. M. Slaney, M. Rangarajan, and M. A. Curtis. 1995. Characterization, genetic analysis, and expression of a protease antigen (PrpRI) of *Porphyromonas gingivalis* W50. *Infect. Immun.* **63**:4744-4754.
2. Aduse-Opoku, J., J. M. Slaney, M. Rangarajan, J. Muir, K. A. Young, and M. A. Curtis. 1997. The Tla protein of *Porphyromonas gingivalis* W50: a homolog of the RI protease precursor (PrpRI) is an outer membrane receptor required for growth on low levels of hemin. *J. Bacteriol.* **179**:4778-4788.
3. 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.
4. Becker, J. M., G. A. Caldwell, and E. A. Zachgo. 1996. Biotechnology: a laboratory course, 2 ed. Academic Press, Inc., New York, N.Y.
5. Derbyshire, K. M., and N. D. Grindley. 1992. Binding of the IS903 transposase to its inverted repeat in vitro. *EMBO J.* **11**:3449-3455.
6. Dong, H., and M. J. Duncan. Unpublished data.
7. Duncan, M. J., S. A. Emory, and E. C. Almira. 1996. *Porphyromonas gingivalis* genes isolated by screening for epithelial cell attachment. *Infect. Immun.* **64**:3624-3631.
8. Duncan, M. J., S. Nakao, Z. Skobe, and H. Xie. 1993. Interactions of *Porphyromonas gingivalis* with epithelial cells. *Infect. Immun.* **61**:2260-2265.
9. Fletcher, H. M., R. M. Morgan, and F. L. Macrina. 1997. Nucleotide sequence of the *Porphyromonas gingivalis* W83 *recA* homolog and construction of a *recA*-deficient mutant. *Infect. Immun.* **65**:4592-4597.
10. Hanley, S. A., J. Aduse-Opoku, and M. A. Curtis. 1999. A 55-kilodalton immunodominant antigen of *Porphyromonas gingivalis* W50 has arisen via horizontal gene transfer. *Infect. Immun.* **67**:1157-1171.
11. Kado, C. I., and S. T. Liu. 1981. Rapid procedure for detection and isolation of large and small plasmids. *J. Bacteriol.* **145**:1365-1373.
12. Lewis, J. P., and F. L. Macrina. 1998. IS195, an insertion sequence-like element associated with protease genes in *Porphyromonas gingivalis*. *Infect. Immun.* **66**:3035-3042.
13. Mahillon, J., and M. Chandler. 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**:725-774.
14. Maley, J., and I. S. Roberts. 1994. Characterisation of *IS1126* from *Porphyromonas gingivalis* W83: a new member of the IS4 family of insertion sequence elements. *FEMS Microbiol. Lett.* **123**:219-224.
15. Maley, J., N. B. Shoemaker, and I. S. Roberts. 1992. The introduction of colonic-*Bacteroides* shuttle plasmids into *Porphyromonas gingivalis*: identification of a putative *P. gingivalis* insertion-sequence element. *FEMS Microbiol. Lett.* **72**:75-81.
16. McKee, A. S., A. S. McDermid, R. Wait, A. Baskerville, and P. D. Marsh. 1988. Isolation of colonial variants of *Bacteroides gingivalis* W50 with a reduced virulence. *J. Med. Microbiol.* **27**:59-64.
17. Menard, C., and C. Mouton. 1995. Clonal diversity of the taxon *Porphyromonas gingivalis* assessed by random amplified polymorphic DNA fingerprinting. *Infect. Immun.* **63**:2522-2531.
18. Nakayama, K. 1994. Rapid viability loss on exposure to air in a superoxide dismutase-deficient mutant of *Porphyromonas gingivalis*. *J. Bacteriol.* **176**:1939-1943.
19. Okamoto, K., T. Kadowaki, K. Nakayama, and K. Yamamoto. 1996. Cloning and sequencing of the gene encoding a novel lysine-specific cysteine protein-

- ase (Lys-gingipain) in *Porphyromonas gingivalis*: structural relationship with the arginine-specific cysteine proteinase (Arg-gingipain). *J. Biochem.* **120**: 398–406.
20. **Pavloff, N., J. Potempa, R. N. Pike, V. Prochazka, M. C. Kiefer, J. Travis, and P. J. Barr.** 1995. Molecular cloning and structural characterization of the Arg-gingipain proteinase of *Porphyromonas gingivalis*. Biosynthesis as a proteinase-adhesin polyprotein. *J. Biol. Chem.* **270**:1007–1010.
  21. **Rezsohazy, R., B. Hallet, J. Delcour, and J. Mahillon.** 1993. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* **9**:1283–1295.
  22. **Suzuki, H., T. Ikeda, T. Noguchi, and F. Yoshimura.** 1997. Detection and prevalence of IS1126, an insertion sequence, in *Porphyromonas gingivalis* by Southern blot analysis. *Dent. Jpn.* **33**:111–115.
  23. **Wang, C. Y., V. C. Bond, and C. A. Genco.** 1997. Identification of a second endogenous *Porphyromonas gingivalis* insertion element. *J. Bacteriol.* **179**: 3808–3812.

---

*Editor:* V. A. Fischetti