JOURNAL OF BACTERIOLOGY, Apr. 1992, p. 2724–2728
0021-9193/92/082724-05$02.00/0

Vol. 174, No. 8

Homology of a Plasmid from the Spirochete *Treponema denticola* with the Single-Stranded DNA Plasmids

JANE MACDOUGALL, DANIELLE MARGARITA, AND ISABELLE SAINT GIRONS*
Unité de Bactériologie Moléculaire et Médicale, Institut Pasteur, 75724 Paris Cedex 15, France

Received 6 November 1991/Accepted 5 February 1992

The 2,647-bp nucleotide sequence of cryptic plasmid pTD1, isolated from the oral spirochete *Treponema denticola*, was determined. The sequence revealed two open reading frames, A and B, which encode polypeptides of 335 and 235 amino acids, respectively. Open reading frame A shows sequence similarity to genes that encode replication proteins from a group of plasmids common in gram-positive bacteria, which replicate via a single-stranded intermediate.

The order *Spirochaetales* comprises two families of morphologically similar but otherwise diverse bacteria (9, 16); these include species that are pathogenic for both humans and animals. Despite their medical significance, genetic manipulation of these organisms is still at a preliminary level, in part because of the lack of any known system of genetic transfer. The identification of a small, circular 2.6-kb plasmid, pTD1, in the cultivable oral spirochete *Treponema denticola* raised the possibility of a plasmid-based genetic transfer system within this group of bacteria (8). As the first step in the development of pTD1 as a shuttle vector, we determined its nucleotide sequence to distinguish between essential regions required for replication and nonessential areas that would be suitable as cloning sites. This is the first reported sequence of an extrachromosomal element from within this genus.

Identification of proteins encoded by pTD1. Gene expression in pTD1 was demonstrated in a cell-free prokaryotic DNA-directed in vitro transcription-translation assay, based on a 30S ribosomal *Escherichia coli* extract (Amersham, France SA). The gene products were identified by autoradiography of incorporated L-[^{35}S]methionine (37 TBq/mmol; Amersham, France SA), after conventional sodium dodecyl sulfate-polyacrylamide gel electrophoresis in a 10% discontinuous system (11), and this revealed the presence of four major protein bands. In comparison to molecular size markers, these corresponded to 32, 28, 25, and 21 kDa (Fig. 1). These size estimations suggest total coding regions in the plasmid of approximately 3,300 bp. Since the plasmid is only 2,647 bp long, the amino acid sequence of each of the four protein products cannot be unique. This could be a consequence, for example, of the initiation of translation at alternate sites within a coding region.

pTD1 structure and DNA sequence analysis. Plasmid pTD1 was obtained from *T. denticola* ATCC 33520 which was maintained anaerobically as previously described (8). The entire nucleotide sequence, comprising 2,647 bp (Fig. 2), was determined for both strands. Figure 2 indicates the restriction endonuclease sites utilized to subclone fragments of pTD1 into pTZ18R (Pharmacia) for sequencing by the dideoxy method of Sanger et al. (19). The methods used for ligation, transformation, preparation of single-stranded DNA (ssDNA), running of agarose gels, and restriction endonuclease reactions were as described by Sambrook et al. (18). The relative ease with which clones could be made, and also their stability, varied considerably; ssDNA could not be obtained from clones carrying some regions of the plasmid. These latter regions were sequenced from double-stranded templates obtained by transformation of the recombinant plasmids into *E. coli* XL-1 Blue (*endA1 hsdR17 supE44 thi-1* λ$^-$ *recA1 gyrA96 relA1*[F'*proAB lacI*^{q}ZΔM15 Tn*10*]) and Qiagen preparation of the DNA (Hybaid Ltd.).

The nucleotide sequence was searched for regions containing inverted repeats of 10 bp or more, and three such regions were found. Between bp 152 and 344 are one inverted repeat of 10 bp and two 10-bp direct and two 10-bp indirect repeats with greater than 80% homology (Fig. 2). A further three inverted repeats, the longest of which spans 20 bp, occur at the end of the sequence. Such regions have the potential to form hairpin loops and secondary structures in RNA transcripts and may also be important for regulation of plasmid replication (3).

Analysis of the sequence revealed the presence of three open reading frames (ORFs) of more than 50 codons. The largest ORF, A, extends from bp 167 to 1220 and contains 336 codons (Fig. 2). The first ATG codon in ORF A is at bp 212 to 215. A protein expressed from ORF A, initiated by this codon, would have a molecular weight of 38,999; however, this does not correspond to the data from the in vitro assay, according to which the largest protein product has an apparent mobility of 32 kDa (Fig. 1). The sequence of the 3' end of the 16S rRNA of *T. denticola* is 5'-CCGAAG GUACGUUU-3' (15). The sequence 5'-ACCT-3' is thus a possible ribosome-binding site (20), 7 bp from the ATG at bp 212 to 215 (Fig. 2). A sequence reading 5'-GAAG-3', which bears a resemblance to the consensus sequences described for *E. coli* ribosome-binding sites, occurs at bp 371 to 374, 17 bp from the third ATG codon at bp 392 to 394. If this served as a site for translation initiation in *E. coli*, it would account for the accumulation of a protein smaller (32 kDa) than that predicted (39 kDa) in the in vitro transcription-translation assay.

ORF A homology with replicative proteins and conservation of the active site. Comparison of ORF A both as a nucleotide sequence and as a peptide sequence with the EMBL (release 26), GenBank (release 66), and NBRF (release 27) data banks showed similarity between this ORF and the sequences for the replicative (Rep) proteins from *Bacillus subtilis* plasmid pBAA1 (4), pFTB14 from *B. amyloliquefaciens* (14), pLAB1000 from *Lactobacillus hilgardii* (10), pTB913 from *Bacillus* sp., and pMV158 from *Streptococcus*

* Corresponding author.

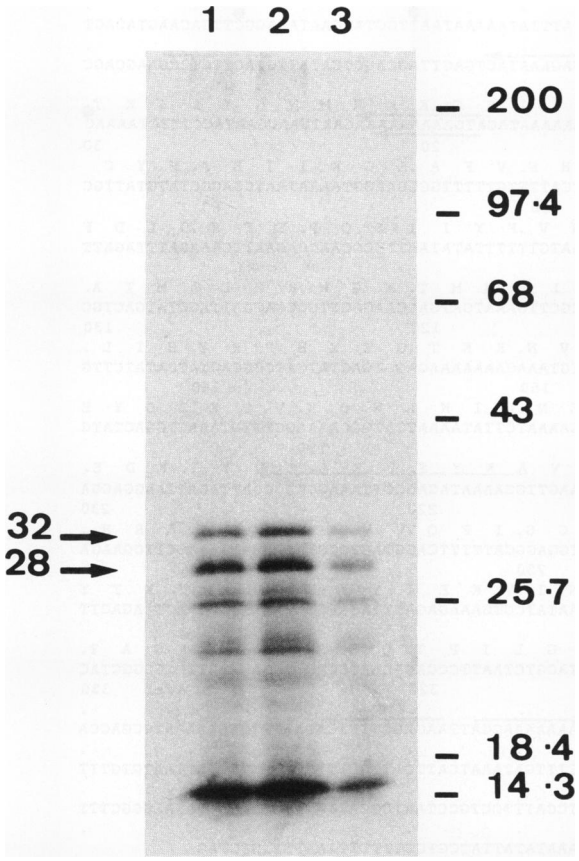


FIG. 1. In vitro transcription-translation products of pTD1 in a sodium dodecyl sulfate–10% polyacrylamide gel, with 6, 4, and 2 μl of the reaction product loaded in lanes 1 to 3, respectively. Protein molecular sizes are given in kilodaltons. The arrows mark the positions of the proteins thought to be the products of ORFs A (32 kDa) and B (28 kDa).

agalactiae (21). If two peptide sequences were aligned by class of amino acid (e.g., hydrophilic) rather than by strict conservation of the amino acid, the level of homology between the protein encoded by ORF A of pTD1 and the Rep protein of pBAA1 could be increased to 45%, and that with the other three Rep proteins could be increased to approximately 40% (Fig. 3A).

An interesting feature of this homology search was that these replicative proteins are all products of plasmids from the group known as ssDNA plasmids, which are highly interrelated and widespread in gram-positive organisms (6). The Rep proteins of those plasmids which share the pBAA1-type consensus sequence for the origin of replication have a conserved amino acid motif which is also present at the active site of the bacteriophage ϕX174 Rep protein (6). Figure 3B shows that this motif is also highly conserved within ORF A of pTD1, at amino acids 217 to 225.

ORF B, which is encoded by the strand opposite to that which encodes ORF A, commences at bp 2297 and terminates at the TAA at bp 1590. The first ATG codon is at bp 2297 to 2295. A protein initiated at this codon would have a predicted molecular weight of 27,000 and could correspond to the 28-kDa protein band observed in the in vitro transcription-translation assay (Fig. 1). A hypothetical *T. denticola* ribosome-binding site, 5′-GACT-3′, is within 3 bp of this ATG (Fig. 2), immediately adjacent to an *E. coli* ribosome-binding site. Computer searches revealed weak homology between ORF B and the plasmid-encoded recombinase proteins in pG12 from *B. thuringiensis* (13) and pLB4 from *L. plantarum* (2), the "hypothetical protein" from *Staphylococcus aureus* plasmid pC403 (7), and regions of plasmids pTB913 and pMV158 (21).

The third ORF, C, runs from bp 2567 to 2316. Although this ORF is large enough to code for a peptide of 84 amino acids, in the absence of a product of the predicted molecular weight from the in vitro transcription-translation assay it was concluded to be a noncoding region.

Features of ssDNA plasmids. ssDNA plasmids are alike in that they all replicate via an ssDNA intermediate, by rolling-circle replication. Three plasmid-encoded elements are essential for rolling-circle replication: the plus origin, a replication protein (Rep), and a minus origin. The sequence homology between ORF A and known Rep proteins suggested that it too could function as a Rep protein and that pTD1 is an ssDNA plasmid. Further examination of the sequence revealed that there are also regions in pTD1 which could function as the plus and minus origins. In ssDNA plasmids, the minus origin serves as an efficient initiation site, recognized by host factors for conversion of circular plus-strand ssDNA to double-stranded DNA, and marks the completion of one productive cycle of replication. One major feature of the minus origin is the presence of palindromic sequences forming secondary structures. One such region of the pTD1 sequence is located between bp 2306 and 2467 (Fig. 2).

Plus origin of replication. Consensus sequences for the plus origin of replication in ssDNA plasmids fall into three groups. There are two regions within the sequence of pTD1 that have a degree of homology with known plus origin sequences from plasmids which have bacteriophage ϕX174-type Rep proteins (6). One is within ORF A, from bp 915 to 923, and has the sequence 5′-CTCTTGATA-3′. For comparison, the sequence of the origin of plasmid pBAA1 is 5′-CTTAT**CTTGATA**-3′ and that of bacteriophage ϕX174 is 5′-CCCGCAA**CTTGATA**-3′. The other is upstream of ORF A, between bp 2338 and 2351, and has the sequence 5′-TTTCTT**CTTAGATA**-3′. This latter region shows better homology over a longer-length sequence and is rich in palindromic motives (Fig. 2) which would allow formation of hairpin loops and secondary structures, which are common features of this type of origin. The plus origin is typically situated either upstream from or within the Rep protein (6), so each of the above-described regions satisfies this criterion.

Conclusion. According to the criteria specified by Gruss and Ehrlich (6), plasmid pTD1 has sufficient features in common with ssDNA plasmids to be classified as a novel member of this group. Identification of pTD1 as an ssDNA plasmid has been based on homologies found between both nucleotide and peptide sequences; however, the high incidence of unstable clones and deletions may be another indication of its ssDNA nature (6). Replication of pTD1 through an ssDNA intermediate could be confirmed, using *B. subtilis* as the host, by the observation in vivo of pTD1 ssDNA. Alternatively, hybrid derivatives of pTD1 containing inserts of foreign (e.g., from *E. coli*) DNA should yield readily observable high-molecular-weight tandem multimers (5). The presence of a plasmid with ssDNA features suggests that expression of spirochete genes, previously reported to be poor in *E. coli* (1, 12, 17, 22), may be higher in a gram-positive host and that use of shuttle vectors designed

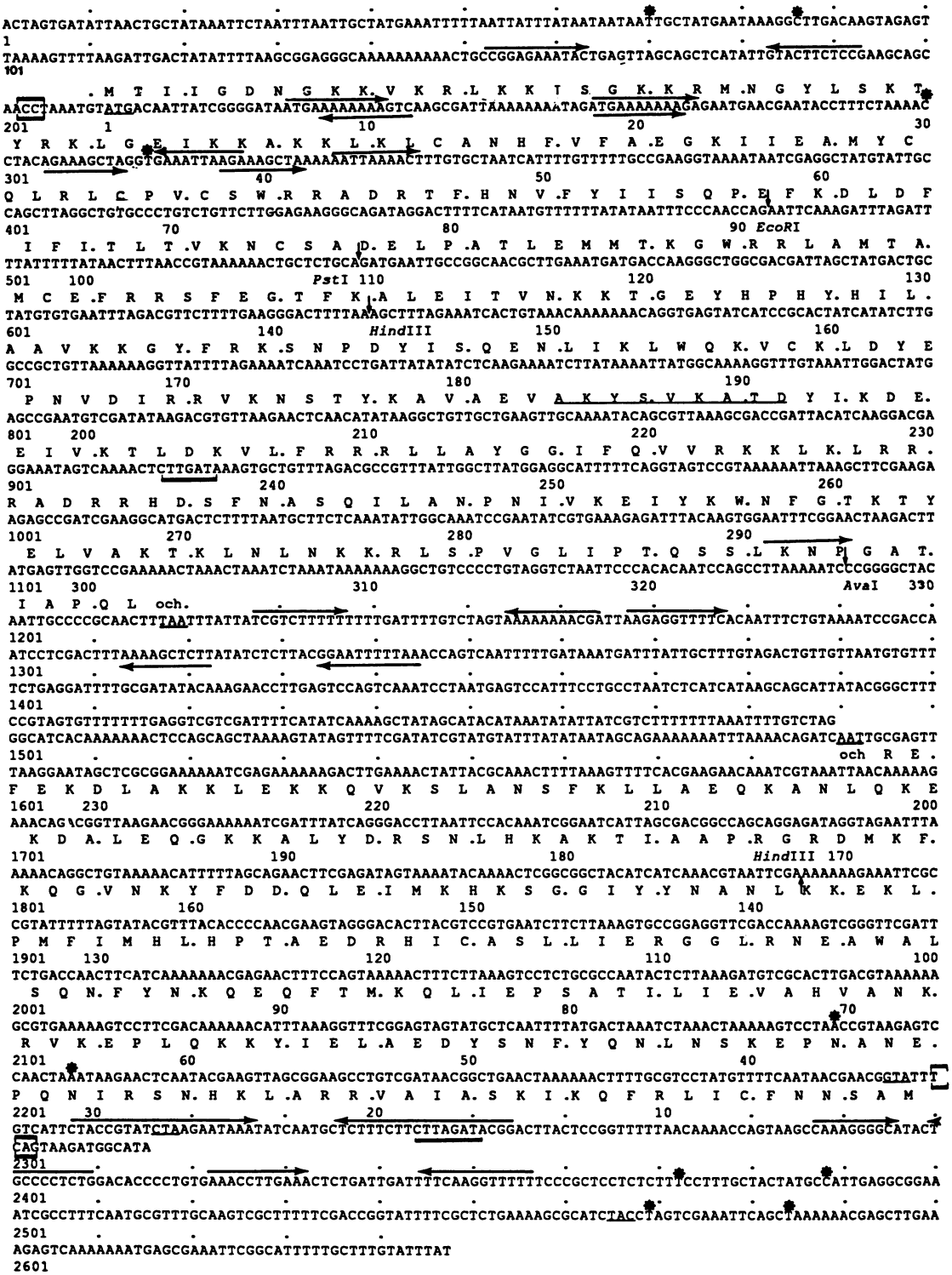
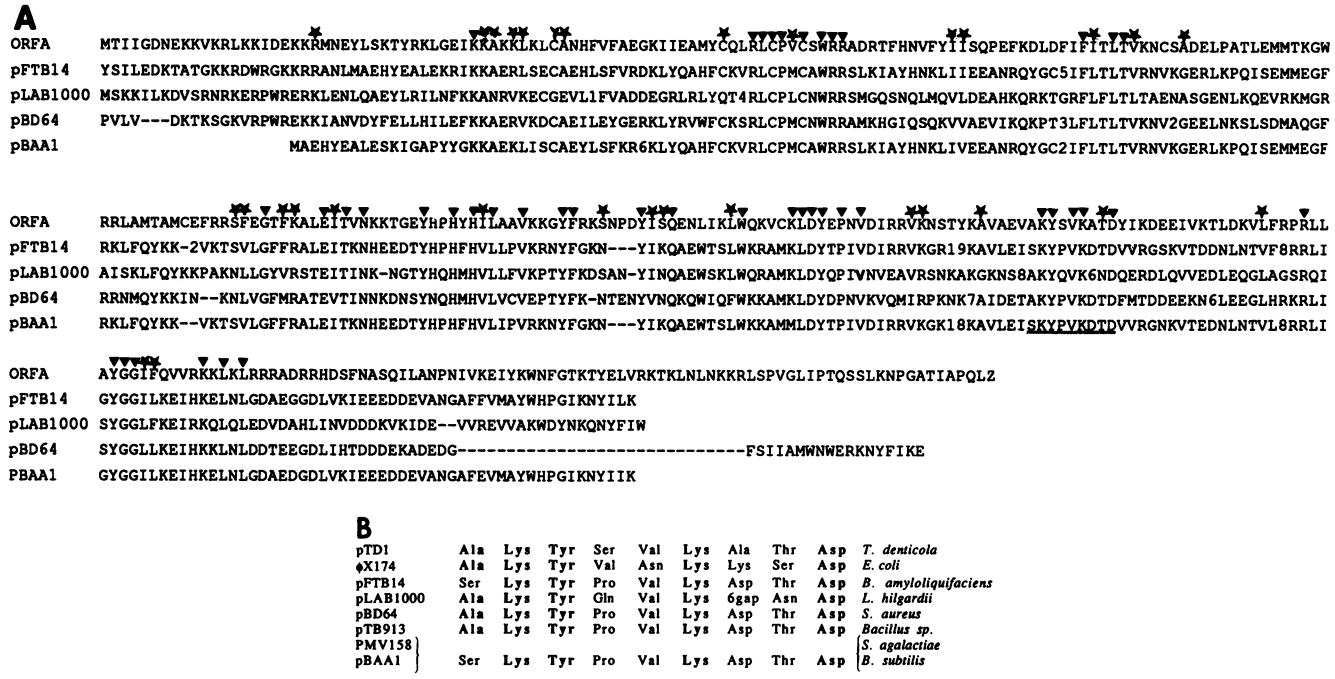


FIG. 2. Complete nucleotide sequence of pTD1 linearized through the unique *Spe*I site at bp 1. Restriction enzymes utilized for cloning are marked with vertical arrows, and the primers constructed to complete the sequence are marked with asterisks. The primer sequences read as follows: I, 5′-TGGCATTCTCAGGTTGGTTGATT-3′ (bp 2189 to 2207); II, 5′-GGCATAGTAGCAAAGGA-3′ (bp 2472 to 2488); III, 5′-TAGTCGAAATTCAGCT-3′ (bp 2569 to 2584); IV, 5′-ACCTAGCTTTCTGTAGG-3′ (bp 300 to 316); V, 5′-TTGCTATGAATAAAGGC-3′ (bp 70 to 86). The bases at the start and finish of each of the three ORFs are underlined, as are the ATG start codons and the conserved amino acid region. Potential ribosome-binding sites are indicated by brackets, and the two possible locations of the plus origin of replication are indicated by half brackets. The positions of repeats are indicated by horizontal arrows. The translations of the ORFs are in the single-letter amino acid code (the first base of each codon) at the top for ORF A and the bottom for ORF B. The numbering at the amino acid level is given below every 10th letter.



B

pTD1	Ala	Lys	Tyr	Ser	Val	Lys	Ala	Thr	Asp	*T. denticola*
ϕX174	Ala	Lys	Tyr	Val	Asn	Lys	Lys	Ser	Asp	*E. coli*
pFTB14	Ser	Lys	Tyr	Pro	Val	Lys	Asp	Thr	Asp	*B. amyloliquifaciens*
pLAB1000	Ala	Lys	Tyr	Gln	Val	Lys	6gap	Asn	Asp	*L. hilgardii*
pBD64	Ala	Lys	Tyr	Pro	Val	Lys	Asp	Thr	Asp	*S. aureus*
pTB913	Ala	Lys	Tyr	Pro	Val	Lys	Asp	Thr	Asp	*Bacillus sp.*
PMV158	Ser	Lys	Tyr	Pro	Val	Lys	Asp	Thr	Asp	*S. agalactiae*
pBAA1	Ser	Lys	Tyr	Pro	Val	Lys	Asp	Thr	Asp	*B. subtilis*

FIG. 3. (A) Alignment of the deduced amino acid sequence of the ORF A gene product and the Rep proteins from four different ssDNA plasmids. The amino acid sequences, in the single-letter code, were aligned to find homology by class of amino acid. Numbers indicate amino acids omitted to maximize homology; dashes show where insertions are required. Triangles indicate amino acids common to all five sequences, and asterisks mark conservative substitutions of amino acids. The conserved amino acid motif of the pBAA1 Rep protein is underlined. (B) Conserved amino acid motif of the putative Rep protein from ORF A of pTD1. Comparison is with the active sites of replicative proteins from bacteriophage ϕX174 and other ssDNA plasmids that share the pBAA1-type origin of replication (6). Conserved amino acids are in boldface.

for gram-positive organisms could enable genetic transformation within spirochetes.

Nucleotide sequence accession number. The sequence shown in Fig. 2 has been submitted to GenBank and assigned accession no. M87856.

We thank Barrie Davidson for critical review of the text and many helpful suggestions with the sequence analysis and Iain Pemberton for oligonucleotide III. We are also grateful to Charles Penn and Roy Russell for continued interest in and encouragement of this project.

This work was supported by CNRS (UA1129) and by an ADIP fellowship (Institut Pasteur) awarded to Jane MacDougall.

REFERENCES

1. **Baril, C., C. Richaud, E. Fournié, G. Baranton, and I. Saint Girons.** 1992. Cloning of *dap*D, *aro*D and *asd* of *Leptospira interrogans* serovar *icterohaemorraghiae*, and nucleotide sequence of the *asd* gene. J. Gen. Microbiol. **136:**651–656.
2. **Bates, E. E. M., and H. J. Gilbert.** 1989. Characterisation of a cryptic plasmid from *Lactobacillus plantarum*. Gene **85:**253–258.
3. **Bouia, A., F. Bringel, L. Frey, B. Kammerer, A. Belarbi, A. Guyonvarch, and J.-C. Hubert.** 1989. Structural organization of pLP1, a cryptic plasmid from *Lactobacillus plantarum* CCM 1904. Plasmid **22:**185–192.
4. **Devine, K. M., S. T. Hogan, D. G. Higgins, and D. J. McConnell.** 1989. Replication and segregational stability of *Bacillus* plasmid pBAA1. J. Bacteriol. **171:**1166–1172.
5. **Gruss, A., and S. D. Ehrlich.** 1988. Insertion of foreign DNA into plasmids from gram-positive bacteria induces formation of high-molecular-weight plasmid multimers. J. Bacteriol. **170:**1183–1190.
6. **Gruss, A., and S. D. Ehrlich.** 1989. The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. Microbiol. Rev. **53:**231–241.
7. **Horinouchi, S., and B. Weisblum.** 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. J. Bacteriol. **150:**815–825.
8. **Ivic, A., J. MacDougall, R. R. B. Russell, and C. W. Penn.** 1991. Isolation and characterisation of a plasmid from *Treponema denticola*. FEMS Microbiol. Lett. **78:**189–194.
9. **Johnson, R. C.** 1977. The spirochetes. Annu. Rev. Microbiol. **31:**89–106.
10. **Josson, K., T. Scheirlinck, F. Michiels, C. Platteeuv, P. Stanssens, H. Joos, P. Dhaese, M. Zabeau, and J. Mahillon.** 1989. Characterisation of a gram-positive broad-host-range plasmid isolated from *Lactobacillus hilgardii*. Plasmid **21:**9–20.
11. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **277:**680–685.
12. **MacDougall, J. H., D. Beighton, and R. R. B. Russell.** 1991. Cloning and expression of protease genes from *Treponema denticola* in *Escherichia coli*. Oral Microbiol. Immunol. **6:**270–274.
13. **Mahillon, J., and J. Seurinck.** 1988. Complete nucleotide sequence of pG12, a *Bacillus thuringiensis* plasmid containing Tn*4430*. Nucleic Acids Res. **16:**11827–11828.
14. **Murai, M., H. Miyashita, H. Araki, T. Seki, and Y. Oshima.** 1987. Molecular structure of the replication origin of a *Bacillus amyloliquefaciens* plasmid pFTB14. Mol. Gen. Genet. **210:**92–100.
15. **Paster, B. J., F. E. Dewhurst, W. G. Weisburg, L. A. Tordoff, G. J. Fraser, R. B. Hespell, T. B. Stanton, L. Zablen, C. R. Woese, and L. Mandelco.** 1991. Phylogenetic analysis of the spirochetes. J. Bacteriol. **173:**6101–6109.
16. **Paster, B. J., E. Stackebrandt, R. B. Hespell, C. M. Hahn, and C. R. Woese.** 1984. The phylogeny of the spirochetes. Syst. Appl. Microbiol. **5:**337–351.

17. **Richaud, C., D. Margarita, G. Baranton, and I. Saint Girons.** 1990. Cloning of genes required for amino acid biosynthesis from *Leptospira interrogans* serovar *icterohaemorrhagiae*. J. Gen. Microbiol. **136:**651–656.
18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
19. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
20. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal DNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
21. **Van der Lelie, D., S. Bron, G. Venema, and L. Oskam.** 1989. Similarity of minus origins of replication and flanking open reading frames of plasmids pUB110, pTB913 and pMV158. Nucleic Acids Res. **17:**7283–7294.
22. **Yelton, D. B., and R. A. Cohen.** 1986. Analysis of cloned DNA from *Leptospira biflexa* serovar patoc which complements a deletion of the *Escherichia coli trpE* gene. J. Bacteriol. **165:**41–46.