1954–1958 Nucleic Acids Research, 1994, Vol. 22, No. 11

An intervening sequence (IVS) in the 16S rRNA gene of the eubacterium *Helicobacter canis*

Dennis Linton, Jonathan P.Clewley[1], André Burnens[2], Robert J.Owen and John Stanley*
National Collection of Type Cultures, [1]Virus Reference Division, Central Public Health Laboratory, 61 Colindale Avenue, London NW9 5HT, UK and [2]Swiss National Reference Laboratory for Foodborne Diseases, Institute for Veterinary Bacteriology, University of Berne, Langass-Strasse 122, CH 3012, Berne, Switzerland

Received April 28, 1994; Accepted May 4, 1994 EMBL accession nos U04344 – U04348 (incl.)

ABSTRACT

PCR amplicons enlarged by ~250bp were generated from the 16S rRNA (*rrs*) genes of certain strains of the recently described *Helicobacter* species, *H.canis*. The DNA sequence of the *rrs* gene of one such strain was determined, and it was shown that an intervening sequence (IVS) of 235bp followed nucleotide 199 in the *rrs* sequence. In four other *H.canis* strains, identical or similar IVSs were found, always at the same location in the *rrs* gene. The secondary structures of the RNA transcripts of the IVSs were predicted. They were characterised by the presence of a conserved stem-loop structure, a potential recognition site for RNA processing enzymes. Ribosomal RNA was compared from a strain of *H.canis* with and without the IVS-containing *rrs* gene. In the former 16S rRNA appeared as two fragments, whose sizes were consistent with cleavage at either side of the IVS, and which were not subsequently religated. The IVS sequence was not represented elsewhere in the *H.canis* genome. Its evolutionary significance is discussed.

INTRODUCTION

The genes of eukaryotes, archaebacteria and eubacteria are sometimes interrupted by intervening sequences (IVSs), which may be either spliced (1) or cleaved from transcribed RNA without splicing (2). The primary transcript from ribosomal RNAs of eubacteria is processed into mature 16S, 23S and 5S rRNAs, whose respective conserved sizes are approximately 1550, 2900 and 120 nucleotides. However, atypical fragmented 23S rRNAs occur in certain eubacteria. For example when the 23S rRNA (*rrl*) gene of *S.typhimurium* (2) and *Y.enterocolitica* (3) was sequenced, an IVS of about 90bp (*S.typhimurium*) or 100bp (*Y.enterocolitica*) was shown to interrupt the *rrl* gene. All these IVSs are excised during 23S rRNA maturation, giving rise to fragmented rRNAs: in *S.typhimurium* IVS excision was shown to be catalysed by the processing enzyme RNAse III (2). Secondary structures have been described for these eubacterial *rrl* gene IVSs. A common feature in them is the presence of lengthy terminal inverted repeats which can form stable helices; potential sites for RNAseIII processing.

The genus *Helicobacter is* a distinct rRNA homology group within rRNA Superfamily VI (4). Its most studied member is the human gastric pathogen *H.pylori* (5, 6). One of several other *Helicobacter* species subsequently isolated from mammalian hosts is the canine species *H.canis* (7). In the course of a study of rRNA genes in this species, we detected atypically large PCR amplicons from the 16S rRNA (*rrs)* genes of certain strains. We show in this report that this is due to the presence of an *rrs* gene IVS with conserved sequence and location. We have carried out a secondary structure prediction for this *rrs* gene IVS and shown that it is removed to give rise to split 16S rRNA.

MATERIALS AND METHODS

Bacterial growth, DNA preparation and blot hybridization analysis

H.canis strains were cultured on 5% (v/v) horse blood agar plates in a Variable Atmosphere Incubator (Don Whitley Scientific) which maintained a microaerobic atmosphere of 5% O_2, 5% CO_2, 2% N_2 and 88% H_2 and a temperature of 37°C. Five of the *H.canis* isolates have been previously described (7) : they are the type strain NCTC 12739^T, NCTC 12741, NCTC 12743 and CCUG 19561 all isolated from dogs, and NCTC 12740 isolated from a human (8). Five further isolates from dogs were analyzed, they were A800-92, A804-92, A806-92, NCTC 12220 and NCTC 12410. Genomic DNA preparation and Southern blot analysis was carried out as previously described (9).

PCR amplification of rrs genes and DNA sequencing

An almost full length *rrs* gene amplicon was generated from various strains of *H.canis* using the 'ϵ' forward primer 5'AAGAGTTTGATCCTGGCTCAG 3' (corresponding to nts. 7–27 in the *E.coli* numbering scheme) and the '1510' reverse primer 5'GGTTACCTTGTTACGACTT 3' (corresponding to nts. 1492–1510 in the *E.coli* numbering system). Standard PCR

*To whom correspondence should be addressed

conditions were meployed and amplicons were analyzed by standard gel electrophoresis in 1% agarose. For DNA sequencing, the amplicon was purified by precipitation with polyethylene glycol 6000 (10) and subjected to dideoxy chain termination sequencing with a single primer and *Taq* polymerase according to the method of Embley (11).

Preparation of rRNA and Northern blot analysis

Total RNA was purified according to the method of Chomczynski and Sacchi (12), gloxylated, subjected to agarose electrophoresis under standard conditions (13) and 'Northern' blotted onto Hybond N nylon membrane. Blots were hybridized with 16S rRNA gene probes generated by PCR using the primer pair ϵ/1510.

Sequence analysis and secondary structure prediction

The RNA secondary structure was predicted by free energy minimisation using the algorithm of Zuker (14–16) as implemented in the programs 'Mulfold' and 'loopDloop' (17) and 'MacDnasis' (Hitachi). The sequences were examined for open reading frames (ORFs) and potential protein coding characteristics using Fickett's algorithm (18) as implemented in the program 'MacVector' (IBI). GenBank and EMBL nucleotide and combined translated databases were searched using 'MacVector' (IBI), 'LaserGene' (DNAStar) and 'Entrez' (National Center for Biotechnology Information, USA).

RESULTS

Enlarged *rrs* gene PCR amplicons are found in some strains of *H.canis*

Ten strains of *H.canis* were identified by concurrent phenotypic tests and homology with a species-specific probe (7). The type strain of *H.canis*, NCTC 12739^T contained an *rrs* gene with a 'conventional' physical–genetic map and sequence (7). When screened by PCR with the primers described in Materials and Methods 'conventional' amplicons of ~1.5 kbp were generated from the type strain and four other isolates of *H.canis*. From the remaining five strains an amplicon enlarged by ~250 bp was generated (see Fig 1). One of these strains, NCTC 12220, was chosen for detailed analysis.

Nucleotide sequence of the *rrs* gene of *H.canis* NCTC 12220

The nucleotide sequence of the *rrs* gene of NCTC 12220 was obtained. When compared with that of the type strain NCTC 12739^T, marked divergence was observed in the 5′ region following position 199 of the *E.coli* numbering system (Fig. 2). The full sequence was aligned with that of NCTC 12739^T (data not shown) indicating that 235bp of DNA were inserted following nt. 199 (*E.coli* numbering scheme). Three Ts which follow nt. 199 in *H.canis* NCTC 12739^T (and *H.pylori*) are also deleted as a result of IVS insertion. Other than the exceptional feature of the large inserted sequence there were 8 nucleotide substitutions and a single deletion in the *rrs* gene of NCTC 12220 compared to that of NCTC 12739^T (*viz.* nt. 189 in Fig. 2, nts. 79, 592, 797, 1006, 1257, 1278, 1331 and a deletion at nt. 888).

Conservation of the IVS sequence and its site of insertion

The nucleotide sequence from position 150 across the junction with the divergent sequence at nt 199, and that of any potential IVS, was obtained for four more strains from which enlarged *rrs* gene amplicons had been generated. In all these strains, a

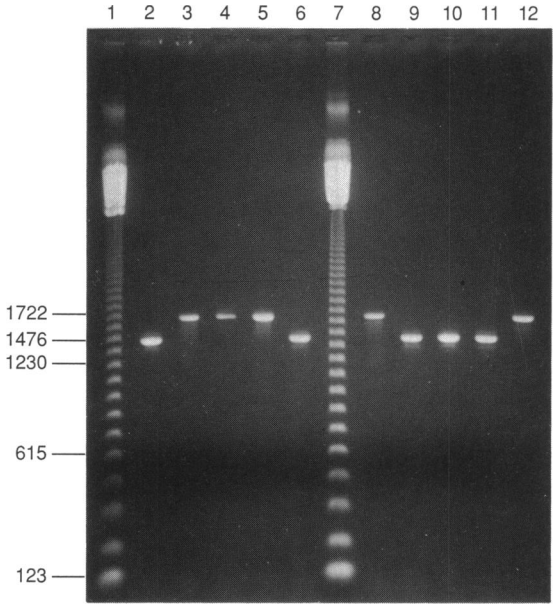

Figure 1. PCR amplicons from the *rrs* gene of ten *H.canis* strains. Lanes 1 and 7 contain molecular size markers, the 123bp ladder (Gibco BRL). Lanes 2,6,9,10, and 11 contained the type strain NCTC 12739^T, NCTC 12741, NCTC 12743, CCUG 19561 and NCTC 12740 respectively. Lanes 3,4,5,8, and 12 contained strains NCTC 12220, A800-92, A804-92, A806-92 and NCTC 12410 respectively, all of which produced amplicons enlarged by ~250bp.

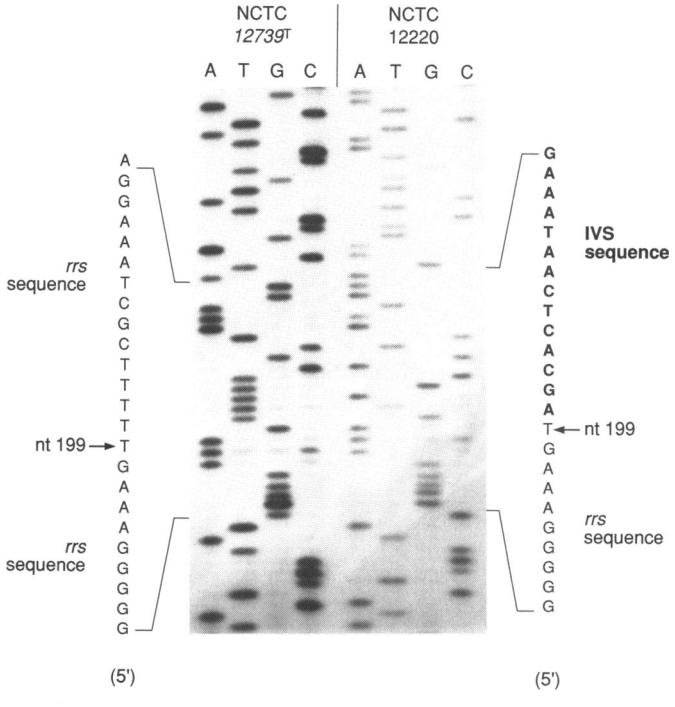

Figure 2. Autoradiograph showing comparative *rrs* gene sequences at the beginning of the IVS. The *rrs* gene amplicons of NCTC 12739^T and NCTC 12220 (fig. 1, lanes 2 and 3) were sequenced directly (see methods) and aligned in the 5′ region. Marked sequence divergence is seen in NCTC 12220 (shown at right) corresponding to the beginning of an IVS.

conserved sequence identical to that in NCTC 12220 was shown to follow nt. 199. The inserted DNA was invariably 235 bp in length, and terminated in the same 3′ junction (see Fig. 3). In two strains the IVS was identical to that of NCTC 12220, in a third strain there was a single base pair substitution and in a fourth strain, two base pairs were substituted.

Predicted secondary structure of the IVS and sequence similarity search

The secondary structures of the IVSs were predicted, and the two different conformations are displayed in figure 3. These conformations, observed from among five separate IVSs, had similar free energies (−67 and −68 kcal/mol at 37°C). Four suboptimal structures could be found within 10% of the computed free energy. nine suboptimal structures could be found up to −58.2 kcal/mol computed free energy. However, the stem (nucleotides 1−22 or 23) was conserved in all suboptimal structures within this range. In an attempt to determine the relationship of these IVSs to any other sequences, the appropriate databases were searched with both the nucleotide sequence of the IVSs and those of two small regions of predicted amino acid coding (reading frames 3 and 6). No homologies were detected in the databases (GenBank/EMBL release 79). There was no apparent sequence similarity with the 16S or 23S rRNA IVS of *Campylobacter sputorum* recently described by van Camp and colleagues (19). The nucleotide 187−212 stem of the *H.canis* IVSs (Fig. 3) had 37% and 43% similarity respectively with regions of the *rrl* gene IVSs of *L.noguchii* and *L.weilii* as

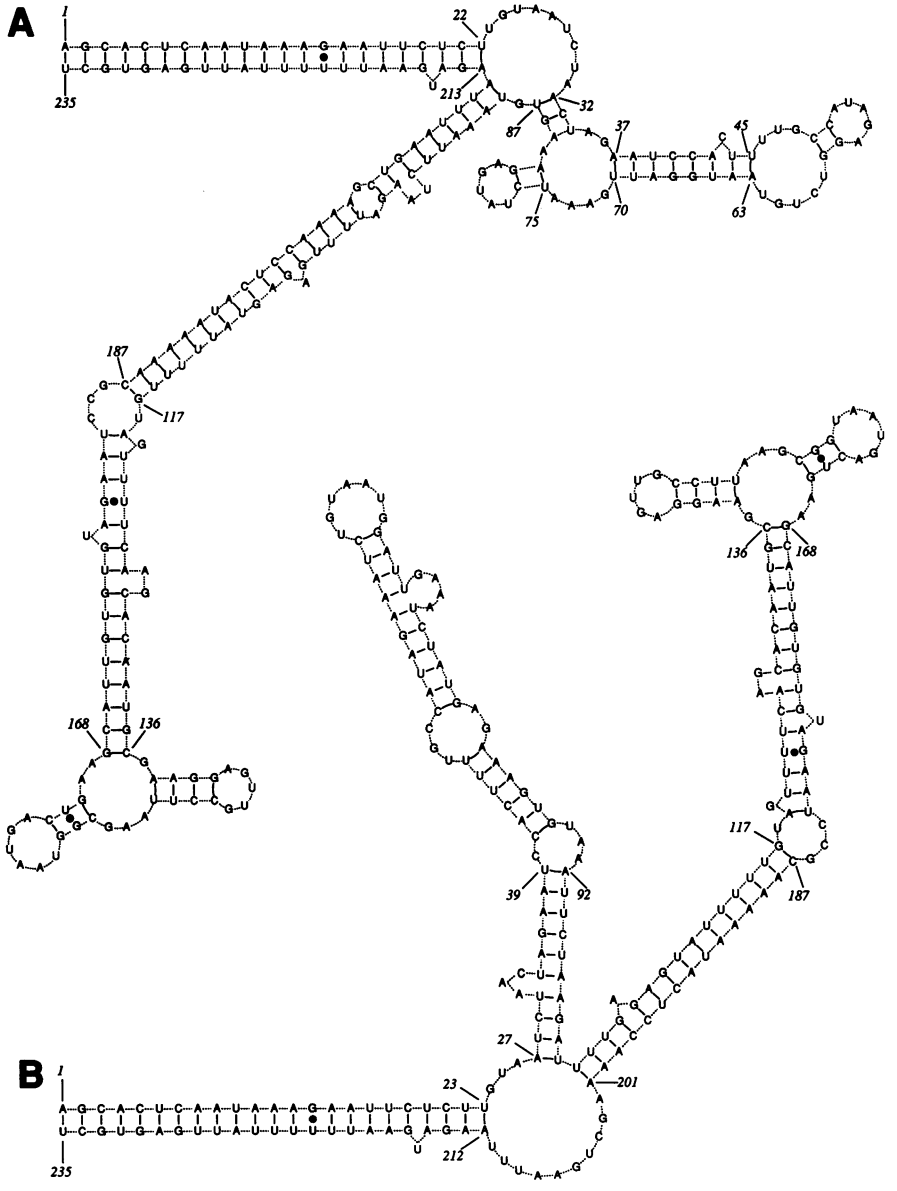

Figure 3. Predicted RNA secondary structures of the IVSs. **A.** NCTC 12220, A800-92, A806-92 and NCTC 12410. Strain NCTC 12410 had a single base pair change (G to U) at position 175 with only a minor local effect on the predicted stem , nucleotides 119−183. **B.** A804-92. This strain had two base pair substitutions (G to A), at positions 56 and 57. Phosphodiestser bonds are represented by broken lines, hydrogen bonds by unbroken lines. GU base pairs are represented by ●.

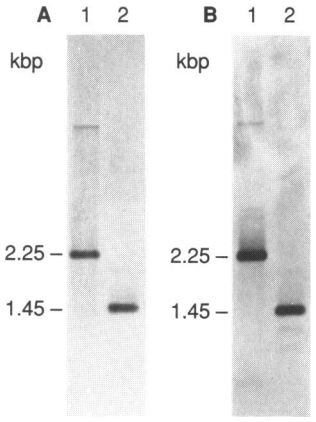

Figure 4. The IVS is present only in the *rrs* gene. **A.** NCTC 12220 (lane 1) and NCTC 12739^T (lane 2) hybridized with an almost full length fragment of the conventional *rrs* gene amplified from NCTC 12739^T. **B.** The same strains hybridized with a 1.3 kbp fragment of the enlarged *rrs* gene containing the IVS, amplified from NCTC 1220.

described by Ralph and McClelland (1993): the significance of this similarity is unclear, but it might represent a conserved structure in IVSs. There are small regions of predicted amino acid coding but these showed no obvious similarity to any protein sequence in the database. However, prediction of the coding potential of the *H.canis* IVS by Fickett's method (18) was not inconsistent with this element being derived from an ancestral coding sequence.

16S rRNA is cleaved during maturation

In order to establish whether or not mature 16S rRNA retained the IVS, total RNA was prepared from NCTC 12739^T and 12220, glyoxylated, electrophoresed and Northern-blotted as described in Methods. In strain NCTC 12739^T two rRNA bands appeared in the denaturing gel. Upon Northern blot analysis, the upper band did not hybridize but the lower (1500 nt.) did hybridize with the 16S (*rrs*) gene probe. In strain NCTC 12220 on the other hand, this lower band was absent and was replaced by two new bands (1300 and 200 nts.). Both new bands hybridized with the *rrs* gene probe (data not shown). Mature 16S rRNA therefore exists as two molecules in NCTC 12220, rather than one conventionally-sized species in NCTC 12739^T. This is consistent with excision of the IVS from the primary transcript of the *rrs* gene in NCTC 12220, without subsequent religation of the products.

The IVS is not represented elsewhere in the genome of *H.canis*

In order to establish whether the IVS was present in any other genetic locus in *H.canis* i.e. whether it had originated or transposed outside the *rrs* gene, genomic Southern blots were made from *Hin*dIII-digested DNA of five strains containing enlarged *rrs* genes, and five strains containing conventional *rrs* genes. They were hybridized with probes for the 1.5kbp (conventional) gene of NCTC 12739^T and a 1.3kbp fragment of the NCTC 12220 gene containing the IVS. There was no difference in the number of bands detected with these probes for any strain, as would be expected if the IVS were located uniquely in the enlarged *rrs* genes of strains which contain it. Results for NCTC 12739^T, which typifies the strains with conventional genes, and NCTC 12220, which typifies those containing enlarged genes, are shown in Figure 4.

CONCLUSIONS

Half the strains of *H.canis* so far identified contain enlarged *rrs* genes. In the 5′ region of these genes is found a 235bp IVS with conserved nucleotide sequence and intragenic location. The secondary structure predicted for the primary RNA transcript of these IVSs are characterized by several domains, but the principal and conserved feature is a 22–23bp stem, which may be a recognition site for RNAse III or related processing enzymes, as was experimentally demonstrated in the case of the IVS int he 23S rRNA gene of *Salmonella typhimurium* (2). The IVS is not retained in mature 16S rRNA, which appears as two separate molecular species, the sizes of which are consistent with cleavages at the insertion site.

With respect to the origin of the IVS in *H.canis*, the existence of strains with and without the element, i.e. their sporadic occurrence, and the apparent nonfunctionality of the IVS transcript, could be taken as evidence of recent evolutionary acquisition. By similar reasoning, the finding of terminal inverted repeats in the IVS might indicate that it is a remnant of an ancestral transposon, which suffered excision from the 16S rRNA gene. However we detected no evidence of the presence of a homologous insertion element elsewhere in the genome of *H.canis*. The fact that some *H.canis* strains have an IVS in the *rrs* gene and other strains do not, may point to the existence of evolutionarily distinct lineages. The low level (<1%) of substitution outside the IVS would be consistent with sub-specific divergence. This should be further investigated from a taxonomic pespective.

The presence of intervening sequences in rRNA genes has been observed for the 23S *(rrl)* gene of certain eubacteria; *Salmonella typhimurium* (20–22), *Leptospira interrogans, L.santasori, L.noguchi, L.weillii* and *L.borgpetersenii* (23, 24) and *Yersinia enterocolitica* (3). Whilst the *S.typhimurium* and *Y.enterocolitica* 23S rRNA gene IVs are small and noncoding, the larger IVSs found in *Leptospira* vary from 485 to 759 nts and contain a small ORF (121–133 amino acids) oriented so that the sense strand appears in the transcript rRNA (24). It was earlier observed that the 23S rRNA appears as three smaller molecular species in *Wolinella recta (Campylobacter rectus), W.curva (C.curvus), Bacteroides gracilis* and *B.ureolyticus* (25), four species belonging to RNA Superfamily VI (4), which also contains the genus *Helicobacter*.

Unlike 23S rRNA, there was no evidence of enlarged 16S (*rrs*) genes or fragmented 16S rRNAs in eubacteria, until two very recent reports of this phenomenon in the *Campylobacter* species *C.helveticus* (26) and *C.sputorum* (19). The present study confirms that the linear continuity of bacterial 16S rRNA genes can (like that of certain bacterial 23S rRNA genes) be interrupted by transcribed IVSs. It remains to be seen whether the existence of IVSs in *rrs* genes is unique to rRNA Superfamily VI, or will also be found in other eubacterial taxa. In our laboratory, sequencing of the 5′ end of an enlarged *rrs* gene amplicon from the related species *Helicobacter fennelliae* also detected a distinct IVS in the same location, i.e. following nt. 199 of the *rrs* gene as in *H.canis*. The conservation of intragenic location of distinct IVSs can be seen from the studies of *C.helveticus* (26), *C.sputorum* (19), *H.canis* (this report) and *H.fennelliae* (our

unpublished observations): IVS insertion sites are conserved between strains and species of *Helicobacter* and *Campylobacter*. A factor in the selection of these conserved insertion sites must be that the upstream and downstream transcripts can remain functional in the ribosome without retaining their covalent contiguity. The finding of four different IVSs in the same location within the 16S rRNA gene of four distinct species, indicates that this site may be a potential hot spot for recombination.

ACKNOWLEDGEMENT

We are grateful to Jon White (CPHL Medical Illustration Dept.) for assistance with drawing of figure 3.

REFERENCES

1. Cech, T. (1990) Annu. Rev. Biochem. 59:543–568.
2. Burgin, A.B., Parodos, K., Lane, D.J. and Pace, N.R. (1990) Cell. 60:405–414.
3. Skurnik, M. and Toivanen, P. (1991) Mol. Microbiol. 5:585–593.
4. Vandamme, P., Falsen, E., Rossau, R., Hoste, B., Segers, P., Tytgat, R. and De Ley, J. (1991) Int. J. Syst. Bacteriol. 41:88–103.
5. Blaser, J. (1993) Trends. Microbiol. 1:255–260.
6. Akopyanz, N., Bukanov, N., Westblom, T.U. and Berg, D.E. (1992) Nuc. Acids. Res. 20:6221–6225.
7. Stanley, J., Linton, D., Burnens, A.P., Dewhirst, F.E., Owen, R.J., Porter, A., On, S.L.W. and Costas, M. (1993) J. Gen. Microbiol. 139:2495–2504.
8. Burnens, A.P., Stanley, J., Schaad, U.B. and Nicolet, J. (1993) J. Clin. Micro. 31:1916–1917.
9. Stanley, J., Burnens, A.P., Linton, D., On, S.L.W., Costas, M. and Owen, R.J. (1992) J. Gen. Microbiol. 138:2293–2303.
10. Kusukawa, N., Uemori, T., Asada, K. and Kato, I. (1990) Biotechniques. 9:66–72.
11. Embley, T.M. (1991) Lett. Appl. Microbiol. 13:171–174.
12. Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162:156.
13. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory Press., New York.
14. Zuker, M. (1989) Science. 244:48–52.
15. Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) Proc. Nat. Acad. Sci. USA. 86:7706–7710.
16. Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) In Doolittle, R.F. (ed.), Molecular evolution: computer analysis of protein and nucleic acid sequences Vol. 183, pp. 281–306.
17. Gilbert, D.G., (1992) . Published electronically on the Internet, available via anonymous ftp to ftp.bio.indiana.edu, 1992,
18. Fickett, J.W. (1982) Nuc. Acids. Res. 10:5303–5318.
19. van Camp, G., van de Peer, Y., Nicolai, S., Neefs, J.-M., Vandamme, P. and de Wachter, R. (1993) Syst. Appl. Microbiol. 16:361–368.
20. Winkler, M.E. (1979) J. Bacteriol. 139:842–849.
21. Smith, N.H., Crichton, P.B., Old, D.C. and Higgins, C.F. (1988) J. Med. Microbiol. 25:223–228.
22. Hsu, D., Zee, Y.C., Ingraham, J. and Shih, L.-M. (1992) J. Gen. Microbiol. 138:199–203.
23. Hsu, D., Pan, M.-J., Zee, Y.C. and LeFebvre, R.B. (1990) J. Bacteriol. 172:3478–3480.
24. Ralph, D. and McClelland, M. (1993) Proc. Nat. Acad. Sci. USA. 90:6864–6868.
25. Sunday, G.J., Gillespie, M.J., Motley, S.T. and Zambon, J.J. (1991) J. Dent. Res. 70:961–5.
26. Linton, D., Dewhirst, F.E., Clewley, J.P., Owen, R.J., Burnens, A.P. and Stanley, J. (1994) Microbiology. 140:847–855.