Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus Veillonella: implications for phylogeny and taxonomy

Hélène Marchandin,¹ Corinne Teyssier,² Michèle Siméon de Buochberg,² Hélène Jean-Pierre,¹ Christian Carrière¹ and Estelle Jumas-Bilak²

¹Service de Bactériologie, Hôpital Arnaud de Villeneuve, 34295 Montpellier Cedex 5, France
²Laboratoire de Bactériologie, Faculté de Pharmacie, 15 Avenue Charles Flahault, 34093 Montpellier Cedex 5, France

Among the seven species characterized within the genus Veillonella, three (Veillonella dispar, Veillonella parvula and Veillonella atypica) have so far been isolated from human flora and during infectious processes. Sequencing and analysis of 16S rDNA (rrs) has been described as the best method for identification of Veillonella strains at the species level since phenotypic characteristics are unable to differentiate between species. rrs sequencing for the three species isolated from humans showed more than 98 % identity between them. Four rrs copies were found in the reference strains and in all the clinical isolates studied. The sequences of each rrs were determined for the clinical strain ADV 360.1, and they showed a relatively high level of heterogeneity (1-43%). In the majority of cases, polymorphic positions corresponded to nucleotides allowing differentiation between the three species isolated from humans. Moreover, variability observed between rrs copies was higher than that between 16S rDNA sequences of V. parvula and V. dispar. Phylogenetic analysis showed that polymorphism between rrs copies affected the position of strain ADV 360.1 in the tree. Variable positions occurred in stems and loops belonging to variable and hypervariable regions of the 16S rRNA secondary structure but did not change the overall structure of the 16S rRNA. PCR-RFLP experiments performed on 27 clinical isolates of Veillonella sp. suggested that inter-rrs heterogeneity occurs widely among the members of the genus Veillonella. These results, together with the lack of phenotypic criteria for species differentiation, give preliminary arguments for unification of V. dispar and V. parvula.

INTRODUCTION

The genus Veillonella consists of small, non-fermentative, strictly anaerobic, Gram-negative cocci that form part of the oral, genitourinary, respiratory and intestinal flora of man and animals. Three species, Veillonella dispar, Veillonella parvula and Veillonella atypica, have so far been isolated from human flora. Several reports have identified these bacteria as pathogens in severe infections such as bacteremia, endocarditis, osteomyelitis and arthritis (Houston et al., 1997; Liu et al., 1998; Marchandin et al., 2001; Singh & Yu, 1992). Moreover, Veillonella sp. strains are frequently isolated from clinical specimens in aero-anaerobic polymicrobial cultures. The genus Veillonella was first divided into two species, V. parvula and V. alcalescens, on the basis of biochemical characteristics. The species were further subdivided into subspecies on the basis of serological reactions (Rogosa, 1965, 1984). Using DNA–DNA homologies, strains representing all the species and subspecies were classified into eight homology groups (Mays et al., 1982): the species V. alcalescens was abolished, the subspecies were elevated to species rank and a new species, V. caviae, was created. Thus, today, the genus Veillonella is subdivided into seven species: V. parvula, V. atypica, V. dispar, V. criceti, V. ratti, V. rodentium and V. caviae. However, the identification of Veillonella at the species level remains uncertain and inconvenient owing to the lack of a conventional phenotypic and biochemical discriminating test (Kolenbrander & Moore, 1992). Moreover, serological tests (Rogosa, 1965) are no longer available. Recently, an identification method based on 16S rDNA amplification followed by RFLP analysis has been published. Using this approach, the authors were able to identify the seven Veillonella species (Sato et al., 1997a, b). These results suggested that the subdivision of the genus into seven species was valid.

Direct sequencing of 16S rDNA (rrs) has proven to be a stable and specific marker for bacterial identification. While the copy number of 16S rRNA genes may vary
from 1 to 15 among eubacterial genomes, it is generally believed that all the copies in an organism are identical or nearly identical in nucleotide sequence. However, scattered nucleotide differences between 16S rDNA copies (so-called micro-heterogeneity) have been described in a few cases. Micro-heterogeneity has been identified in eubacteria, *Escherichia coli* (Cilia et al., 1996; Martinez-Murcia et al., 1999), *Mycobacterium terrae* (Ninet et al., 1996), *Mycobacterium celatum* (Reischl et al., 1998), *Paenibacillus polymyxa* (Nubel et al., 1996), members of the Mollicutes (Bascunana et al., 1994; Heldtander et al., 1998; Liefting et al., 1996; Pettersson et al., 1996) and some actinomycetes (Wang et al., 1997; Yap et al., 1999). These cases may represent a fairly common phenomenon, particularly if we consider the limited number of organisms for which the nucleotide sequences are available for each RNA gene copy. The direct sequencing of the 16S rRNA gene of clinical isolates belonging to the genus *Veillonella* showed ambiguities suggesting micro-heterogeneity. We investigated the copy number of 16S rRNA genes in the *Veillonella* genome and each copy was sequenced for a clinical isolate. A relatively high level of heterogeneity in sequences between copies was found. The intra-strain sequence heterogeneity was then tested by PCR-RFLP on 27 clinical strains. The results are discussed in terms of phylogenetic and taxonomic implications.

**METHODS**

**Bacterial strains and culture conditions.** *V. atypica* ATCC 17744<sup>2</sup> and *V. dispar* ATCC 17748<sup>3</sup> were purchased from the American Type Culture Collection (ATCC), Manassas, VA, USA. *V. atypica* DSM 1399 (corresponding to ATCC 14894) was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany (DMSZ), and *V. parvula* CIP 60.1 from the Institut Pasteur Collection (CIP), Paris, France. Twenty-seven clinical isolates were obtained from the Bacteriology Laboratory of the University Hospital of Montpellier, France. These 27 strains were collected over a 1 year period (from February 1999 to January 2000) from clinical samples of 26 adult patients in various units of the hospital. Strains were isolated in pure culture (one strain), in mixed aero-anaerobic cultures (25 isolates from 24 different patients) or in mixed anaerobic cultures (one strain). They were recovered from bronchoalveolar fluid (*n* = 12), joint aspiration (*n* = 2), wound (*n* = 6), periodontal flora (*n* = 3), semen (*n* = 1), abscess (*n* = 2) and ear (*n* = 1). Cultures were performed on Columbia sheep blood agar incubated for 2 days in anaerobic jars with the AnaeroGen System (Oxoid Unipath). Each isolate was purified by two subcultures of a single colony with particular care. The second subculture was frozen at −80°C and used as bacterial stock for all subsequent experiments. The clinical isolates were identified as *Veillonella* sp. on the basis of morphology, Gram stain, culture and biochemical profile (Gram-negative cocci, strictly anaerobic, catalase production and nitrate reduction).

**DNA amplification, sequencing and RFLP analysis.** DNAs were rapidly extracted by a boiling/freezing method (Carlier et al., 2002). Two-microlitre samples of extracts were used as the template in the PCR experiments. 16S rDNA was selectively amplified by PCR using 5′-GTGCTCGAGAAGGTGTGATCTCGGCTACAGC-3′ (27<sup>l</sup>), position 8–36 (*E. coli* numbering), as the forward primer and 5′-CACCAGATCTACGGGTACCTGTGTTAAAGGCTTT-3′ (1492r), position 1478–1508 (*E. coli* numbering), or 5′-CCCCAACACTTCTCCACAGACAGA-3′ (1090r), position 1083–1102 (*E. coli* numbering), as reverse primers in the conditions previously described (Carlier et al., 2002). Selective amplification of one 16S rDNA copy was performed by cutting *Smal* macrorestriction fragments obtained after PFGE migration in low-melting agarose gel (see below). Then, 2 μl melted agarose was used as the PCR template. A 0.7 kb fragment of the heat-shock protein 70 (*dnaK*) gene was amplified using the following degenerate primers deduced from the alignment of Hsp70 sequences available for the more phylogenetically related species: B1 5′-ATTGAYTTAGGWACAAACA-3′ and B2 5′-GCTTTTTCACGG-HGCDTGYTT-3′. The amplification profile was as follows: 35 cycles of 1 min at 94°C, 1.5 min at 55°C, and 1 min at 72°C. The 16S rDNA and *dnaK* PCR products were directly sequenced on an Applied Biosystems Automatic Sequencer (Genome Express) in both directions by using forward and reverse primers. The RFLP analysis of 16S rDNA PCR products consisted in the digestion of 10 μl amplified samples with 10 U of either *Sac3AI* or *MseI* (Biolabs). The digested samples were analysed by electrophoresis on a 2% agarose gel.

**Phylogenetic analysis.** The 16S rDNA sequences were compared and aligned with sequences deposited in the GenBank database using the programs BLAST (Altschul et al., 1997), LAALIGN and DIALIGN (http://www.expasy.ch). The computed alignments were then manually checked and corrected. Pairwise evolutionary distances were computed using the Jukes and Cantor equation implemented in the DNADIST program and a phylogenetic tree was constructed by the neighbour-joining method (PHYLIP programs package available online at http://www.pasteur.fr). A total of 100 bootstrapped trees were sampled to determine a measure of the support for each node on the consensus tree, using the SEQBOOT and CONSENSE programs (PHYLIP package). Prediction of RNA secondary structure by energy minimization was performed by the MFOLD program (Walter et al., 1994).

**PFGE and DNA electrophoresis.** Intact genomic DNA was extracted in agarose plugs as described previously (Marchandin et al., 2001) and submitted to PFGE after restriction by either 40 U of the endonuclease *Smal* or 1 U of the intronic endonuclease I-CeuI (Jumas-Bilak et al., 1998). The macrorestriction fragments were separated with a contour-clamped homogeneous field electrophoresis apparatus, CHEF-DRII (Bio-Rad) in 0.5× TBE at 10°C. The PFGE parameters for resolution of *Smal* fragments were 42 h at 4.5 V cm<sup>−1</sup>, with switch times ramped from 1 to 15 s in a 1% agarose gel. I-CeuI fragments were separated using pulse times from 90 to 150 s during 24 h at 5.1 V cm<sup>−1</sup> in a 0.8% agarose gel.

Genomic DNAs digested with 10 U of *HindIII*, *NheI* or *Spel* were submitted to electrophoresis for 3 h at 80 V in a 0.8% agarose gel in 0.5× TBE using a SubCell apparatus (Bio-Rad). After electrophoresis, DNA fragments were stained in 0.5× TBE containing 0.5 μg ethidium bromide ml<sup>−1</sup> and visualized under UV illumination.

**Southern blotting, probes and hybridization.** Electrophoresis gels were transferred onto Nitran N (Scheicher and Schuell) nylon membrane by vacuum blotting in 20× SSC. 16S rDNA digoxigenin-labelled probe was obtained by PCR using primers 27<sup>l</sup>F/1090r as described above with a dNTP mixture containing 0.1 mM digoxigenin-dUTP (Roche). Hybridization of the probe was detected by using chromogenic substrate NBT/BCIP (Roche).

**Nucleotide sequence accession numbers.** GenBank accession numbers for 16S rDNA sequences determined in this work are as follows: *V. atypica* ATCC 17744<sup>2</sup>, AF439641; *V. dispers* ATCC 17748<sup>2</sup>, AF439640; *V. parvula* CIP 60.1, AF439639; *Veillonella* sp. strain ADV 360.1, AF439641; *Mycobacterium terrae* CIP 60.1, AF439639; *Mycobacterium celatum* ATCC 17748T, AF439640; *Paenibacillus polymyxa* strain HGCD 17744<sup>T</sup>, AF439641; *Mycobacterium terrae* CIP 60.1, AF439639; *Mycobacterium terrae* CIP 60.1, AF439639; and *Mycobacterium terrae* CIP 60.1, AF439645, respectively. GenBank accession numbers

---

*H. Marchandin and others*
for 70 kDa heat-shock protein (dnaK) gene sequences determined in this work are as follows: V. atypica ATCC 17744\textsuperscript{T}, AF440436; V. dispar ATCC 17748\textsuperscript{T}, AF440435; V. parvula CIP 60.1, AF440437; Veillonella sp. strain ADV 360.1, AY220521.

GenBank accession numbers for previously deposited 16S rDNA sequences used in this work are as follows: V. atypica sequences used in this work are as follows: GenBank accession numbers for previously deposited 16S rDNA sp. strain ADV 360.1, AY220521.

Veillonella ATCC 33048\textsuperscript{T}, X82500; sintes http://mic.sgmjournals.org 1495

Percentage identity between 16S rDNAs of V. atypica The 16S rDNA sequences of V. atypica 16S rDNA sequences and copy number in (dnaK

Veillonella ATCC 17748\textsuperscript{T}, AF440437; V. parvula CIP 60.1, AF440438; Veillonella sp. strain ADV 360.1, AY220521.

Veillonella ATCC 17748\textsuperscript{T}, AF440437; V. parvula CIP 60.1, AF440438; Veillonella sp. strain ADV 360.1, AY220521.

The copy number of 16S rRNA genes was determined by Southern blotting experiments using a 16S rDNA probe on bacterial DNA digested by Hin\textsuperscript{III} (restriction site absent from the 16S rDNA of Veillonella spp.). Four hybridizing fragments were found in V. parvula CIP 60.1, V. dispar ATCC 17748\textsuperscript{T} and all the 27 clinical isolates tested (data not shown). But, oddly enough, only three fragments of V. atypica ATCC 17744\textsuperscript{T} hybridized with 16S rDNA probe (data not shown). These results were confirmed by Southern blotting on SmaI, NheI and SpeI restriction patterns (data not shown). In all eubacterial genomes examined so far, the intronic endonuclease I-Ceu cleaved specifically a 26 bp site in 23S rDNA (Liu & Sanderson, 1998). Thus, the number of I-Ceu restriction fragments is taken as a measure of rrn copy number in a genome. Genomic digestion with I-Ceu gave four restriction fragments for V. parvula CIP 60.1, V. dispar ATCC 17748\textsuperscript{T} and the clinical isolates, whereas only three fragments were obtained for V. atypica ATCC 17744\textsuperscript{T} (Fig. 1). Southern blotting experiments and I-Ceu digestion gave matching results showing the presence of four ribosomal operons in all Veillonella sp. strains tested except V. atypica ATCC 17744\textsuperscript{T}. When we tested V. atypica ATCC 14894 by 16S rDNA hybridization and I-Ceu digestion, four ribosomal operons were found (Fig. 1). Since the rrn skeleton appears to be conserved in a

**Table 1.** Percentage identity between 16S rDNAs of V. atypica ATCC 17744\textsuperscript{T}, V. dispar ATCC 17748\textsuperscript{T} and V. parvula CIP 60.1 and (in italic) between each 16S rRNA gene copy of Veillonella sp. strain ADV 360.1

<table>
<thead>
<tr>
<th></th>
<th>V. atypica ATCC 17744</th>
<th>V. dispar ATCC 17748</th>
<th>V. parvula CIP 60.1</th>
<th>Strain ADV 360.1 rrsA</th>
<th>Strain ADV 360.1 rrsB</th>
<th>Strain ADV 360.1 rrsC</th>
<th>Strain ADV 360.1 rrsD</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. atypica ATCC 17744</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. dispar ATCC 17748</td>
<td>98.1</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. parvula CIP 60.1</td>
<td>98.2</td>
<td>99.0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain ADV 360.1 rrsA</td>
<td>97.6</td>
<td>99.5</td>
<td>98.9</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain ADV 360.1 rrsB</td>
<td>98.2</td>
<td>98.7</td>
<td>99.7</td>
<td>98.6</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain ADV 360.1 rrsC</td>
<td>98.1</td>
<td>98.6</td>
<td>99.5</td>
<td>98.5</td>
<td>99.8</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Strain ADV 360.1 rrsD</td>
<td>97.6</td>
<td>99.5</td>
<td>98.9</td>
<td>100</td>
<td>98.6</td>
<td>98.5</td>
<td>100</td>
</tr>
</tbody>
</table>
large population of *Veillonella* spp., the unexpected lack of one copy in *V. atypica* ATCC 17744T was probably a peculiarity of this collection strain.

**16S rRNA gene copies of a Veillonella sp. clinical isolate are heterogeneous in sequence**

The direct sequencing of total 16S rRNA PCR products obtained from several clinical isolates of *Veillonella* sp. gave ambiguous results consisting of undetermined positions. Moreover, visual inspection of the chromatograms showed positions with a double sequencing signal (data not shown). Owing to the presence of multiple 16S rDNA copies in the genome of *Veillonella*, this phenomenon could be explained by some heterogeneity in sequences between copies. Each 16S rDNA copy of the clinical isolate ADV 360.1 was directly sequenced using the following method. The four *SmaI*-restricted fragments hybridizing with the 16S rRNA probe were cut from the PFGE gel. The endonuclease *SmaI* gave hybridization patterns with clearly separated bands, avoiding cross-contamination when the gel slices were cut (data not shown). PCR amplification and direct sequencing were performed on each of the four gel slices. The four *rrs* copies were named *rrs* A, *rrs* B, *rrs* C and *rrs* D following the sizes of the *SmaI* fragments. Due to the presence of a *SmaI* site in the 3' end of the *Veillonella* 16S rDNA, shorter amplification products of about 1100 bp were obtained for each *rrs* copy using primers 27f and 1090r. Alignment of partial sequences of 1048 bp (from 82 to 1109 following *E. coli* numbering) with the sequences we have deposited for *V. atypica* and *V. parvula* showed that *rrs A* and *rrs D* had maximum identity with the sequence of *V. dispar* whereas *rrs B* and *rrs C* had maximum identity with that of *V. parvula* (Table 1). The sequences of *rrs A* and *rrs D* were identical whereas *rrs B* and *rrs C* varied in three nucleotide positions. We found that 15 of the 1048 nucleotide positions (1.43%) were variable between the four 16S rDNA copies. The *V. dispar* and *V. parvula* 16S rDNA sequences displayed a higher percentage of identity than the two *rrs* copies of the *Veillonella* sp. isolate ADV 360.1 (Table 1).

To confirm that strain ADV 360.1 was clonal and thus rule out the hypothesis that the heterogeneity observed could be due to a mix of different strains, we performed the following experiment. The *dnaK* gene was amplified and sequenced for *V. atypica* ATCC 17744T, *V. dispar* ATCC 17748T, *V. parvula* CIP 60.1 and *Veillonella* sp. strain ADV 360.1. The sequences showed inter-species variability higher than that observed after comparison of 16S rDNA sequences. For example, *V. dispar* and *V. parvula* differed by 7.2% of nucleotide positions. We also observed the absence of a double sequencing signal corresponding to heterogeneous positions in the *dnaK* sequence of ADV 360.1. This result was a strong argument in favour of the clonality of this isolate since a mix of strains would have led to a *dnaK* sequence with several heterogeneous positions.

The differences in nucleotides between the three *Veillonella* species, and between the four 16SrRNA gene copies of strain ADV 360.1, are listed in Table 2. Nine positions have been determined to be variable between *V. dispar* and *V. parvula*. Among them, eight were proved to be polymorphic between *rrs* copies of the strain ADV 360.1. In fine, only one of these variable nucleotides (position 193) could be considered as valuable for differentiation between *V. parvula* and *V. dispar*. However, comparison of the two currently available 16S rDNA sequences for *V. dispar* ATCC 17748T (accession numbers X84006 and AF439640) showed that position 193 is variable. Consequently, none of the variable positions between *V. dispar* and *V. parvula* could be retained as species-specific. In contrast, 11 nucleotides differentiated *V. atypica* from *V. dispar* and *V. parvula*. Among them, eight were not polymorphic between operons and can be retained for species specification (Table 2). Thus, *V. atypica* could be separated from the two other human species on the basis of 16S rDNA sequences.

The reading of the chromatogram corresponding to the bulk sequence obtained by amplification of total DNA of strain ADV 360.1 clearly exhibited double sequencing signals in polymorphic positions. For *V. dispar* ATCC 17748T and *V. parvula* CIP 60.1, the chromatograms of bulk sequences showed some double peaks in the same positions but the secondary peaks were small and could be interpreted as sequencing artefacts in a usual or automatic reading. The absence of clear double signals on the chromatograms could result either from the absence of inter-*rrs* polymorphism or from a preferential amplification of one copy.

As described for members of a multigene family, *rrn* operons are subject to homogenization processes. Thus, *rrn* sequences generally show low variability between copies in a genome and within species or subspecies (Liao, 2000). However, a total of 15 positions variable between copies have been identified in the same strain in this work (Table 2). Despite a few descriptions of micro-heterogeneity, the general extent of 16S rRNA heterogeneity within a bacterial species has received little attention, except in few works (Cilia et al., 1996; Clayton et al., 1995; Martinez-Murcia et al., 1999; Ueda et al., 1999). Clayton et al. (1995) observed that slightly different 16S rRNA sequences were deposited in the databases for strains belonging to the same species. They thought most of the differences were not caused by sequencing errors but by real intra-specific variations. Part of these variations could be due to inter-operon heterogeneity. Thus, the authors showed that 18% of *rrs* sequence pairs deposited for the same strain exhibited 1–5% heterogeneity. We compared sequences of *rrs* copies obtained from 32 completely sequenced genomes of Gram-positive bacteria and *Proteobacteria* using the BLAST program. With respect to the 16S rDNA region analysed in this work (from 82 to 1109 following *E. coli* numbering), the percentage of variation in sequence between copies of the same organism ranges between 0 and 1.2%. The distribution of the strains according to the percentage of
Table 2. Nucleotides of the 16S rRNA gene polymorphic between *V. atypica* ATCC 17744<sup>T</sup>, *V. dispar* ATCC 17748<sup>T</sup> and *V. parvula* CIP 60.1, or between *rrs* copies of strain ADV 360.1

Helix numbering is according to the system of the Gutell Lab (http://www.rna.icmb.utexas.edu/). Positions follow *E. coli* numbering. The underlined positions correspond to an imprecise *E. coli* numbering due to the presence of additional nucleotides in the *Veillonella* sequences. Dark grey: positions allowing species differentiation but polymorphic between copies of strain ADV 360.1. Light grey: positions allowing species differentiation and conserved between copies of strain ADV 360.1. The row labelled ADV 360.1 corresponds to the bulk sequence obtained by sequencing a mix of the four *rrs* copies. The nucleotides in parentheses indicate the presence of double sequencing signals on the chromatogram. Nucleotides in bold type belong to *Sau*3AI or *Mnl*I restriction sites.

<table>
<thead>
<tr>
<th>Helix:</th>
<th>H184</th>
<th>H198</th>
<th>H441</th>
<th>H588</th>
<th>H1006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positions:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV 360.1 rrsA</td>
<td>T</td>
<td>C</td>
<td>GA</td>
<td>TA</td>
<td>CATG</td>
</tr>
<tr>
<td>ADV 360.1 rrsB</td>
<td>T</td>
<td>T</td>
<td>GA</td>
<td>AA</td>
<td>TGTA</td>
</tr>
<tr>
<td>ADV 360.1 rrsC</td>
<td>T</td>
<td>C</td>
<td>GA</td>
<td>AA</td>
<td>TGTA</td>
</tr>
<tr>
<td>ADV 360.1 rrsD</td>
<td>T</td>
<td>C</td>
<td>GA</td>
<td>TA</td>
<td>CATG</td>
</tr>
<tr>
<td>ADV 360.1</td>
<td>T</td>
<td>(C/T)</td>
<td>GA</td>
<td>(A/T)A</td>
<td>(C/T)/(A/G)/(T/A/G)</td>
</tr>
<tr>
<td><em>V. parvula</em></td>
<td>T</td>
<td>C</td>
<td>GA</td>
<td>AA</td>
<td>TTTA</td>
</tr>
<tr>
<td><em>V. dispar</em></td>
<td>T</td>
<td>C</td>
<td>GA</td>
<td>TG</td>
<td>TTTA</td>
</tr>
<tr>
<td><em>V. atypica</em></td>
<td>C</td>
<td>T</td>
<td>AC</td>
<td>TG</td>
<td>TTTA</td>
</tr>
</tbody>
</table>

Restriction site | *Mnl* | *Sau*3AI | *Sau*3AI |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CCTCN&lt;sub&gt;7&lt;/sub&gt;</td>
<td>GATC</td>
<td>GATC</td>
<td></td>
</tr>
</tbody>
</table>
variation was: <0·1 %, 19 strains; 0·1–1 %, 12 strains; and >1 %, one strain. The same distribution was obtained when complete sequences were compared. Thus, the level of variations in the sequences of rrs copies of Veillonella sp. strain ADV 360.1 (1·43 %) could be considered high and somewhat atypical.

**16S rDNA PCR-RFLP of 27 clinical isolates of Veillonella sp.**

The nucleotides variable between rrs copies of the strain ADV 360.1 occurring in positions 591 and 648 involved polymorphism of two Sau3AI restriction sites. In the same way, the variable position 184 involved MnlI polymorphism (Table 2). These enzymes have been previously described as effective in the identification of Veillonella species by 16S rDNA PCR-RFLP (Sato et al., 1997a, b). Particularly, Sau3AI digestion clearly distinguished V. dispar from V. parvula. We performed a PCR-RFLP assay by Sau3AI digestion of the PCR products obtained with primers 27f and 1090r. V. parvula gave a profile formed by five bands of 59, 116, 169, 308 and 491 bp. A profile consisting of three bands of 116, 169 and 910 bp was obtained for V. dispar (data not shown). These results were in accordance with the number and size of Sau3AI restriction fragments calculated on the sequences. Sau3AI digestion of PCR products obtained from each rrs copy of strain ADV 360.1 is shown in Fig. 2. The profiles obtained for rrsA and rrsD were similar to the V. dispar profile, whereas the rrsB and rrsC profiles corresponded to the V. parvula one. The RFLP profiles were in accordance with the polymorphism in rrs sequences described above. The PCR product obtained from total genomic DNA of strain ADV 360.1 gave a complex RFLP profile consisting of a mix of the bands obtained in lanes 1–4 (Fig. 2). Using this method, we rapidly searched for the occurrence of sequence heterogeneity between rrs copies in a panel of 27 human isolates of Veillonella sp. Among these strains, four (14·8 %) gave the profile described for V. parvula CIP 60.1, three (11·1 %) gave the profile characteristic of V. dispar type strain, and 20 (74·1 %) gave mixed profiles. These 20 isolates were shown to be unrelated after comparison of Smal macro-restriction patterns (data not shown). The general aspect of the PFGE profiles suggested that each DNA was extracted from a single strain and not from a mixture of strains. Thus, the clonality of each isolate proved for the strain ADV 360.1 was highly probable for the other isolates. These results suggested that inter-operon heterogeneity was a common phenomenon that occurred widely in the genus Veillonella.

**Pattern of sequence variation on the rrs secondary structure**

The four rrs sequences of strain ADV 360.1 and reference strains were manually aligned with the secondary structure of Clostridium perfringens, the closest relative to the genus Veillonella for which a 16S rRNA secondary structure model is available in the Gutell Lab Comparative RNA database (http://www.rna.icmb.utexas.edu/). The 15 nucleotides variable between rrs copies occurred in four variable or hypervariable regions of the secondary structure: (i) helices H184 and H198; (ii) helix H441; (iii) helix H588; (iv) helix H1006 (Table 2). The nomenclature of the helices was given following the recommendation of the Gutell Lab (http://www.rna.icmb.utexas.edu/). The impact of these substitutions (from position 184 to 1019, E. coli numbering) on the secondary structure was analysed using the Mfold program for each of the four rrs sequences of the strain ADV 360.1. Eleven substitutions fell in stem regions of the 16S tRNA secondary structure model. Four of these substitutions resulted in a change from a canonical base pair to a common non-canonical base pair GU or GA types (or vice versa). One substitution resulted in a change from the rare non-canonicalUU to the common non-canonical GU. In three cases, variations in stem regions resulted in a substitution in the complementary position of the stem, leading to two changes from canonical GC to canonical AU and to one change from the rare non-canonicalUU to the canonical CG. Three other substitutions occurred in the loop which capped the stems of helices H198, H441 and H588. The remaining variable position occurred in the loop between helices H184 and H198. Finally, the 15 substitutions observed could be considered as compensatory since the overall secondary structure was conserved.
It is striking that differences between rrs sequences did not appear at random. The variable positions occurred in stems and loops belonging to hypervariable and variable regions of the 16S rRNA secondary structure model. None of them coincided with invariant positions or positions that were conserved in a large majority of bacteria. Moreover, genes A and D have identical sequences but differed from genes B and C by 13 and 14 nucleotides, respectively, whereas rrsB and rrsC differed by only three nucleotides. This fact suggested that the gene duplication followed an independent acquisition of the two parental copies, probably by recombination events after lateral transfer.

**Phylogenetic analysis and taxonomy of the genus Veillonella**

The 1048 bp nucleotide sequences of rrsA, rrsB, rrsC and rrsD of strain ADV 360.1 were aligned together with seven deposited sequences obtained from the three human species of Veillonella, two human Veillonella strains unidentified at the species level and two species of Veillonella isolated from rodents (V. cricetii and V. ratti). Four related species belonging to the Sporomusa sub-branch, Anaeroglobus geminatus, Dialister pneumosintes, Megasphaera elsdenii and Acidaminococcus fermentans, were also included and aligned. The last one was added as outgroup organism. A phylogenetic tree obtained by the neighbour-joining method after computing a distance matrix using the DNADIST program implemented with the Jukes and Cantor algorithm is presented in Fig. 3. The tree topology confirmed that the rrsA and rrsD sequences were closely related to the sequence of V. dispar whereas rrsB and rrsC were related to that of V. parvula. These branches were associated with bootstrap scores of 96 % and 94 %, respectively. Thus, the sequence variation among rrs copies of strain ADV 360.1 affected the phylogenetic position of this strain. The level of heterogeneity observed in this work did not affect the phylogenetic results obtained with organisms exhibiting distant relationship but has to be considered for studies at the level of the species or the subspecies.

The origin of distinct types of rRNA genes in a single genome could be explained either by the divergent evolution of each copy or by lateral transfer between different species (Cilia et al., 1996; Wang et al., 1997). Our phylogenetic analysis did not solve the problem of the origin of the heterogeneity in the Veillonella genome since it reflects the actual diversity but not the mechanisms involved during evolution.

Previous studies have shown that comparison of 16S rRNA gene sequences gives results in accordance with the existence of the seven species previously described on the basis of serological data (Sato et al., 1997a, b; Tanner et al., 2000). The topology of the phylogenetic tree constructed in this work suggests that V. dispar type strain and V. parvula CIP 60.1 associated with strain ADV 360.1 and the two unidentified Veillonella sp. strains form a homogeneous group. No cluster representative of two species could be highlighted in this group, mainly for three reasons: (i) similarity in sequence of about 99 % between 16S rDNAs; (ii) heterogeneous distribution of the rrs copies in the same strain; (iii) branching of the two unidentified strains, which did not allow their affiliation to one of the two defined species. The Veillonella species have been described by serotyping (Rogosa, 1965, 1984) and then confirmed by DNA/DNA hybridization (Mays et al., 1982). However, the DNA/DNA hybridization data were hardly interpretable and they seem to be insufficient to support any taxonomic conclusions. Our results based on 16S rDNA analysis did not suggest the separation of V. parvula and V. dispar in two independent clusters. However, the value of this marker in splitting species is limited since an intra-operon variability higher than an inter-species variability was observed.
The sequencing of other housekeeping genes in a multi-locus sequencing typing (MLST) approach as recommended by Stackebrandt et al. (2002) could clarify the inter-species relations in the genus Veillonella. The sequencing of dnaK and rpoB genes is on-going in our laboratory in order to obtain additional arguments to discuss the unification of V. parvula and V. dispar. However, our study has already revealed that 16S rDNA-based methods are not suitable for the identification of V. dispar and V. parvula.

A polyphasic approach which includes phenotypic and sequencing data is recommended for bacterial systematics (Murray et al., 1990; Vandamme et al., 1996; Stackebrandt et al., 2002). As a result, and in relation to the development of molecular methods, a dramatic increase of the number of 16S rRNA sequences deposited in databases has been seen. However, analyses at the intra-species and/or at the intra-chromosomal level have been completely ignored in most sequencing analyses. Our results, together with previously published data (Martínez-Murcia et al., 1999; Ninet et al., 1996), suggest that it is necessary to determine rss copy number and to sequence each copy in a large panel of bacterial species. Indeed, evaluating the general extent of rss heterogeneity in the bacterial world could clarify its implications for 16S rDNA-based phylogeny, taxonomy and identification.

ACKNOWLEDGEMENTS

We wish to thank Josiane Campos, Jean-Luc Jeannot, Agnès Masnou and Isabelle Diego for technical assistance. We are grateful to Philippe Vande Perre, Georges Gal and Lynn Richardson for reading the manuscript. This work was supported by the association ADEREMPHA, Montpellier, France.

REFERENCES


