Partial recN gene sequencing: a new tool for identification and phylogeny within the genus *Streptococcus*

Olga O. Glazunova, Didier Raoult and Véronique Roux

Laboratoire de Bactériologie – Virologie, Hôpital de la Timone, CNRS UMR 6236, IFR48, 264 rue Saint-Pierre, 13385 Marseille, Cedex 05, France

Partial sequences of the recN gene (1249 bp), which encodes a recombination and repair protein, were analysed to determine the phylogenetic relationship and identification of streptococci. The partial sequences presented interspecies nucleotide similarity of 56.4–98.2 % and intersubspecies similarity of 89.8–98 %. The mean DNA sequence similarity of recN gene sequences (66.6 %) was found to be lower than those of the 16S rRNA gene (94.1 %), rpoB (84.6 %), sodA (74.8 %), groEL (78.1 %) and gyrB (73.2 %). Phylogenetically derived trees revealed six statistically supported groups: *Streptococcus salivarius*, *S. equinus*, *S. hyovaginalis*/*S. pluranimalium*/*S. thailandensis*, *S. pyogenes*, *S. mutans* and *S. suis*. The ‘mitis’ group was not supported by a significant bootstrap value, but three statistically supported subgroups were noted: *Streptococcus sanguinis*/*S. cristatus*/*S. inisens*, *S. anginosus*/*S. intermedius*/*S. constellatus* (the ‘anginosus’ subgroup) and *S. mitis*/*S. infantis*/*S. peroris*/*S. oralis*/*S. oligofermentans*/*S. pneumoniae*/*S. pseudopneumoniae*. The partial recN gene sequence comparison highlighted a high percentage of divergence between *Streptococcus dysgalactiae* subsp. *dysgalactiae* and *S. dysgalactiae* subsp. *equisimilis*. This observation is confirmed by other gene sequence comparisons (groEL, gyrB, rpoB and sodA). A high percentage of similarity was found between *S. intermedius* and *S. constellatus* after sequence comparison of the recN gene.

To study the genetic diversity among the ‘anginosus’ subgroup, recN, groEL, sodA, gyrB and rpoB sequences were determined for 36 clinical isolates. The results that were obtained confirmed the high genetic diversity within this group of streptococci.

Representatives of the genus *Streptococcus* are Gram-positive, catalase-negative, non-motile cocci. They are commensals of the tegument or mucous membranes in both humans and animals. The normal presence of streptococci on the mucous membranes and the skin explains how they can frequently contaminate biological samples. Although their pathogenicity is variable, they can be responsible for meningitis, pneumonia, endocarditis and fasciitis. At the time of writing, 65 species and 12 subspecies with validly published names are recognized as members of the genus (Euzeby, 2010). Novel species are continually being isolated from human, animal and environmental samples and have been characterized by 16S rRNA gene sequence comparisons in association with phenotypic criteria (Glazunova et al., 2006; Milinovich et al., 2008; Shewmaker et al., 2007; Takada & Hirasawa, 2008; Vela et al., 2009). The precise identification of these bacteria at the species level is quite laborious, particularly in the ‘viridans’ group. Streptococci were first identified based on the haemolytic reaction, group carbohydrate antigens (Lancefield serotyping) and phenotypic tests (Facklam, 2002). Commercial systems based on comparisons of phenotypic characteristics do not permit unambiguous identification at the species level, however, because of the limited number of biochemical traits that can be analysed, the variability of traits within a species, the poor reproducibility of some tests and the fact that not all species are included in the database. Therefore, it is impossible to obtain a real idea of the host–pathogen relationships of streptococci. According to Hoshino et al. (2005), the rate of correct identification of non-haemolytic streptococcal strains by commercial identification kits was below 50 % and varied significantly with regard to the studied species.

**Abbreviation:** ILD, incongruence length difference.

The GenBank/EMBL/DDBJ accession numbers for the partial recN gene sequences determined in this study are EU917226–EU917315, as detailed in Supplementary Table S1, and those of the other sequences determined in this study are FJ712038–FJ712187 (Supplementary Table S4).

Strain and primer details, including details of accession numbers, and details of inter- and intraspecies divergence and phylogenetic trees for the ‘anginosus’ subgroup are available as supplementary material with the online version of this paper.
Various molecular DNA-based methods have been proposed for the identification of *Streptococcus* species. Genes encoding d-alanine:d-alanine ligase (*ddl*) (Garnier et al., 1997), autolysin (*lytA*) (Kawamura et al., 1999), dextranase (*dex*) (Igarashi et al., 2001), the 10 kDa and 60 kDa heat-shock proteins (*groESL*) (Hung et al., 2005; Teng et al., 2002), the RNA subunit of endoribonuclease P (*rnpB*) (Tapp et al., 2003), elongation factor Tu (*tuf*) (Picard et al., 2004), quinolone-resistance-determining regions of DNA gyrase (*gyrA*) and topoisomerase subunit C (*parC*) (Kawamura et al., 2005) and the 16S–23S rRNA intergenic spacer region (Chen et al., 2004; Nielsen et al., 2009) have been suggested. A good candidate for identification at the species level should present enough variability between different species but little variability within a species.

Zeigler (2003) studied 32 protein-encoding genes and determined a small set of protein-encoding genes in which the percentage of similarity between isolates could assign novel strains reliably to bacterial species and replace DNA–DNA hybridization. This set of genes included *recN* (recombination/repair protein), *thdF* (GTP-binding, thiophene oxidation) and *rpoA* (RNA polymerase, alpha subunit). For members of the family *Pasteurellaceae*, sequence comparisons of these three genes were proposed to replace DNA–DNA hybridization, to estimate the genome G+C content of a species and to infer phylogenetic association in association with other markers (Kuhnert & Korczak, 2006). Sequence comparison of the *recN* gene was applied to phylogenetic analysis and the assignment of isolates to species within the genus *Geobacillus* and presented a resolving power that was nearly an order of magnitude greater than that of the 16S rRNA gene (Zeigler, 2005). In this study, we explored the use of partial *recN* gene sequence comparisons in order to investigate taxonomy and phylogeny in the genus *Streptococcus*. In addition, our results were compared with those obtained previously for four other genes (*rpoB*, *sodA*, *groEL* and *gyrB*; Glazunova et al., 2009).

The 89 streptococcal strains studied (representing 54 species and nine subspecies) were obtained from the Collection de l’Institut Pasteur (CIP, Paris, France) and the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany) and are listed in Supplementary Table S1 (available in IJSEM Online). *Enterococcus faecalis* CIP 103015\(^3\) was chosen as an outgroup for phylogenetic tree reconstructions. Clinical isolates obtained by our laboratory are characterized in Table 1.

Bacterial DNA was extracted using the MagNA Pure LC DNA isolation kit III (Roche) with the MagNA Pure LC instrument, as described by the manufacturer. Primers for amplification of the *recN* gene fragment (recNd/recNr) were selected from conserved regions after alignment of sequences from complete streptococcal genomes (*Streptococcus pneumoniae* R6 (GenBank accession no. AE007317), *S. pneumoniae* TIGR4 (AE005672), *S. agalactiae* 2603V/R (AE009948), *S. pyogenes* SF370 (AE004092) and *S. mutans* UA159 (AE014133)) and phylogenetically related bacterial genomes (*Enterococcus faecalis* V583 (AE016830), *Lactococcus lactis* IL1403 (AE005176), *Listeria innocua* Clip11262 (AL592022), *Listeria monocytogenes* EGD (AL591824) and *Staphylococcus aureus* N315 (BA000018)) which were available when the study was started. The PCR annealing temperature was 48°C. Other primers were designed for the sequencing reactions. Primers are described in Supplementary Table S2. Fragments of the *rpoB*, *sodA*, *gyrB* and *groEL* genes were amplified by PCR as described previously (Glazunova et al., 2009).

PCR products were purified using MultiScreen PCR plates (Millipore) and sequenced using a DNA sequencing kit (BigDye Terminator version 1.1 cycle sequencing kit; PE Biosystems) according to the manufacturers’ instructions. Sequence products were purified and electrophoresis was performed with a 3100 Genetic Analyzer (Applied Biosystems).

Gene sequences were aligned using the multiple alignment program CLUSTAL_X version 1.8 (Thompson et al., 1997). Phylogenetic relationships were determined using MEGA version 4 (Tamura et al., 2007). Distance matrices were determined following the assumptions described by Kimura (1980) and were used to reconstruct dendrograms using the neighbour-joining method (Saitou & Nei, 1987). The maximum-parsimony algorithm was also used for phylogenetic analysis (Farris, 1970). Phylogenetic trees were also calculated by the maximum-likelihood method using the program DNAML in the PHYLIP software package (Felsenstein, 1989). The percentage of similarity between nucleotide sequences was calculated using BioEdit software. Bootstrap values were obtained for a consensus tree based on 100 randomly generated trees. DNA sequences of clinical isolates were compared using the BLAST2 program (Altschul et al., 1990) (http://www.ncbi.nlm.nih.gov/blast/b12seq/wblast2.cgi) with our local database.

The TaxonGap software tool (Slabbinck et al., 2008) was applied to represent the resolution of the gene within and between taxonomic units. TaxonGap represents the *s*-heterogeneity and *s*-separability values as light-grey and dark-grey horizontal bars, respectively, for individual biomarkers. The same scale is used to plot both *s*-heterogeneity and *s*-separability in order to support optimal comparability of the values across the biomarkers. The name of the closest neighbour is identified to the right of the dark-grey bar. Distances used for the calculation of *s*-heterogeneity and *s*-separability values were determined using pairwise nucleotide sequence alignments with MEGA version 4.

To investigate incongruence between the *recN* data and the dataset for each gene sequence, the incongruence length difference (ILD) test of Farris et al. (1994) was performed using the PAUP 4.0 beta version (Swofford, 1998). The maximum-parsimony trees exploited for the ILD test were obtained from parsimony-informative sites determined using MEGA version 4.
Phylogeny of streptococci derived from partial recN sequences

The sequenced fragment represented 75.3% of the full-length recN gene sequence of S. mutans UA159 (NP_721018). A 1249 bp fragment was amplified from the majority of the 89 streptococcal strains studied using the recN primers designed in this study, with the exception of the fragment from Streptococcus gallinaceus CIP 107087T, which was 1246 bp. A fragment of only 1212 bp was amplified from Streptococcus parauberis CIP 103956T as a result of unsuccessful sequencing of the 3’ end of the gene. Accession numbers for all submitted sequences are given in Supplementary Table S1.

Phylogenetic analysis was inferred from partial recN sequences using the neighbour-joining, maximum-parsimony and maximum-likelihood methods. Identical organization was obtained for the three analyses with slight variations. The highest bootstrap values were obtained for the tree inferred from the neighbour-joining method and therefore we have concentrated on this analysis in the discussion. The neighbour-joining phylogenetic tree for type strains of species and subspecies is presented in Fig. 1.

Compared with the phylogenetic trees inferred from previously described sequence comparisons of various genes, the tree inferred from sequence comparison of the

<table>
<thead>
<tr>
<th>Strain</th>
<th>Isolate no.</th>
<th>Biological sample</th>
<th>Identification based on:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>recN</td>
<td>groEL</td>
</tr>
</tbody>
</table>

**Table 1. Identification of clinical isolates and reference strains of S. anginosus, S. constellatus and S. intermedius using phylogenetic analysis and BLAST comparison**

Analysis resulted in identification as: SA, S. anginosus; SC, S. constellatus; SI, S. intermedius. Strains that could not be identified by phylogenetic analysis were identified by BLAST comparison, with identifications given in parentheses. NK, Not known.
Fig. 1. Phylogenetic tree of members of the genus *Streptococcus* inferred from comparison of partial recN gene sequences using the neighbour-joining method. Bar, 0.05 changes per nucleotide position. The sequence of *Enterococcus faecalis* CIP 103015T was used as the outgroup. Bootstrap values $\geq 50$ (from 100 replicates) are indicated.
recN gene resulted in the most stable nodes. The numbers of nodes presenting bootstrap values higher than 50, 80 and 95% are detailed in Table 2. In the recN tree, 29 (49.2%) nodes presented bootstrap values ≥95%, which was higher than the value found for groEL (24, 40.7%), which we had considered the best tool for inference of phylogenetic relationships of streptococci in a previous study (Glazunova et al., 2009).

Although the representatives of the ‘mitis’ group formed the major cluster, it was not supported by a significant bootstrap value. Three subgroups were observed and were supported by high bootstrap values. The first included Streptococcus cristatus, S. sanguinis and S. sinensis, the second included Streptococcus anginosus, S. intermedius and S. constellatus and the third included Streptococcus infantis, S. peroris, S. oralis, S. oligofermentans, S. pneumoniae, S. pseudopneumoniae and S. mitis. In addition, Streptococcus parasanguinis and S. entericus grouped together, and Streptococcus gordonii and S. massiliensis were isolated as singletons.

Six other clusters were noted: (i) Streptococcus thermophilus, S. vestibularis and S. salivarius; (ii) the representatives of the ‘equinus’ group, which included Streptococcus alactolyticus, S. galloyticus subsp. galloyticus, S. galloyticus subsp. pasteurianus, S. galloyticus subsp. macedonicus, S. infantarius subsp. infantarius, S. infantarius subsp. coli and S. equinus; (iii) Streptococcus hyovaginalis, S. phuranimalium and S. thoraliensis; (iv) Streptococcus pyogenes, S. dysgalactiae subsp. dysgalactiae, S. dysgalactiae subsp. equisimilis, S. equi subsp. equi, S. equi subsp. zooepidemicus and S. castoreus; (v) the representatives of the ‘mutans’ group, which included Streptococcus mutans, S. macacae, S. devriesi and S. ratti; and (vi) Streptococcus suis, S. acidominimus, S. gallinaceus, S. minor and S. ovis. All of the other species clustered with non-significant bootstrap values at the nodes.

In the tree constructed by the maximum-parsimony method, the same clusters were noted. By this method, S. entericus was not included in the ‘mitis’ group but rather was closely related to the ‘mutans’ group, even though the bootstrap value at the node was not significant (20%). S. equi and S. castoreus were closely related to the ‘pyogenes’ group, but the bootstrap values at the nodes were low (56 and 54%, respectively). In the tree constructed by the maximum-likelihood method, the major groups were congruent with those obtained using the neighbour-joining method and were supported by statistically significant bootstrap values that ranged from 85 to 100%. Slight modifications were noted in the ‘mitis’ group. S. massiliensis was closely related to the ‘anginosus’ subgroup, but the bootstrap value at the node was low (60%). S. gordonii grouped with the ‘cristatus’ subgroup, also with a low bootstrap value (59%).

Compared with the tree obtained from sequence comparison of the groEL gene fragment (Glazunova et al., 2009), some differences were noted. S. acidominimus and S. suis were not included in the ‘mitis’ group and clustered with S. gallinaceus, S. minor and S. ovis. Streptococcus feris was closely related to the ‘mutans’ group with a bootstrap value of 89%. S. equi clustered with S. pyogenes, S. dysgalactiae and S. castoreus. In the ‘mitis’ group, S. gordonii did not cluster with S. sanguinis, S. cristatus and S. sinensis. The other groups or subgroups that were described for phylogenetic analysis inferred from recN sequence comparisons were identical to those inferred from groEL sequence comparisons (Glazunova et al., 2009).

To improve some of these inconsistencies, we considered concatenating the sequences that were obtained from different gene fragments. Most systematists consider data partitions to be combinable if they are not strongly incongruent with one another and use tests, such as the ILD test, to evaluate incongruence (Planet, 2006). Our results from the ILD test demonstrated no incongruence between the recN data and the data for the previously studied genes (not shown). Therefore, we did not concatenate sequences that were obtained from different genes.

**Table 2. Nodes that presented bootstrap values greater than the indicated threshold values**

Bootstrap values were calculated from 100 trees. Corresponding percentages of all nodes in the trees are indicated in parentheses. Phylogenetic trees were inferred from different genetic targets using the neighbour-joining method. Accession numbers of the sequences used are listed in Supplementary Table S1 and by Glazunova et al. (2009).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Bootstrap values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≥50</td>
</tr>
<tr>
<td>recN</td>
<td>42 (71.2)</td>
</tr>
<tr>
<td>groEL</td>
<td>38 (64.4)</td>
</tr>
<tr>
<td>gyrB</td>
<td>40 (67.8)</td>
</tr>
<tr>
<td>rpoB</td>
<td>33 (55.9)</td>
</tr>
<tr>
<td>sodA</td>
<td>34 (57.6)</td>
</tr>
</tbody>
</table>

**Sequence similarity for streptococcal species**

The TaxonGap output for the Streptococcus species that were studied by recN gene sequence comparisons is shown in Fig. 2. The s-separability is a measure of the divergence between the different species, which is subsequently referred to as interspecies divergence. The s-heterogeneity is a measure of the heterogeneity between different isolates of the same species, which is subsequently referred to as intraspecies divergence. The left panel shows a neighbour-joining tree inferred from sequence comparison of partial recN genes of type strains.

The similarity of the recN sequences between members of different species ranged from 56.4% between Streptococcus halichoeri and S. entericus to 98.2% between S. constellatus and S. intermedius. Compared to previously studied genes, recN gene sequence comparisons presented lower mean similarity values between different species (Table 3).
Sequence comparison of the partial recN gene was also applied to streptococci at the subspecies level. In the genus *Streptococcus*, most of the subspecies were highly related (95.6–98%). The most closely related subspecies were *S. galolyticus* subsp. *galolyticus* and *S. galolyticus* subsp. *macedonicus* (98%) and the least closely related subspecies were *S. dysgalactiae* subsp. *dysgalactiae* and *S. dysgalactiae* subsp. *equisimilis* (89.8%). Higher mean divergence values between subspecies were obtained with the recN gene (Tables 3 and 4). Fewer than 1% of species and subspecies pairs exhibited levels of sequence similarity ≥90%. The majority of species and subspecies pairs (85.6%) were situated in the range of 60–70% similarity. Several isolates from some species or subspecies were studied: *S. mitis* (3), *S. oralis* (2), *S. constellatus* (3), *S. intermedius* (3), *S. anginosus* (2), *S. gordonii* (2), *S. infantarius* subsp. *coli* (2), *S. infantarius* subsp. *infantarius* (3), *S. gordonii* subsp. *gordonii* (2), *S. gordonii* subsp. *galolyticus* (2), *S. gordonii* subsp. *macedonicus* (2), *S. gordonii* subsp. *galolyticus* (2), *S. gordonii* subsp. *macedonicus* (2), *S. gordonii* subsp. *macedonicus* (2), *S. gordonii* subsp. *macedonicus* (2), *S. gordonii* subsp. *macedonicus* (2)}.<ref>

**Fig. 2.** TaxonGap output for *Streptococcus* species studied by partial recN gene sequence comparison. The left panel shows a neighbour-joining tree inferred from partial recN gene sequence comparison. For each of the species in the phylogenetic tree, the right panel depicts the interspecies and intraspecies variability. Separability values between different species are represented as dark-grey bars with the names of the closest relatives. Heterogeneity values among isolates of the same species are represented as pale-grey bars.
Table 3. Nucleotide sequence similarities between species and divergence between subspecies of the genus *Streptococcus*

Values are means with minimum–maximum ranges in parentheses. Accession numbers of the sequences used are listed in Supplementary Table S1 and by Glazunova *et al.* (2009).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Interspecies similarity (%)</th>
<th>Intersubspecies divergence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA</td>
<td>94.1 (88.8–99.7)</td>
<td>1.11 (0.03–2.1)</td>
</tr>
<tr>
<td>rpoB</td>
<td>84.6 (76.1–97.6)</td>
<td>1.31 (0.57–2.1)</td>
</tr>
<tr>
<td>sodA</td>
<td>74.8 (63.6–100)</td>
<td>5.92 (1.2–12.9)</td>
</tr>
<tr>
<td>groEL</td>
<td>78.1 (72.6–96.6)</td>
<td>2.96 (1.2–6.1)</td>
</tr>
<tr>
<td>gyrB</td>
<td>73.2 (64.6–97.5)</td>
<td>3.29 (1.6–6.5)</td>
</tr>
<tr>
<td>recN</td>
<td>66.6 (56.4–98.2)</td>
<td>4.32 (1.9–10.2)</td>
</tr>
</tbody>
</table>

subsp. *gallolyticus* (4), *S. galloxyticus* subsp. *pasteurianus* (3), *S. equinus* (3) and *S.agalactiae* (8). Although only relatively few isolates were studied, the intraspecies variability seemed low for most species (<4 %), with the exception of *S. mitis* (6.7 %). These observations will need to be confirmed by studying more isolates.

In the phylogenetic analysis inferred from the recN gene sequence comparison, the two subspecies of *S. dysgalactiae* did not cluster together. The species *S. dysgalactiae* was proposed by Garvie *et al.* (1983), and its description was amended by Farrow & Collins (1984). The latter authors showed, based on DNA–DNA hybridization, that the species 'Streptococcus equisimilis' should be included in *S. dysgalactiae*. Two subpopulations of this species were subsequently characterized by Vandamme *et al.* (1996), from which two subspecies, *S. dysgalactiae* subsp. *dysgalactiae* (animal origin) and *S. dysgalactiae* subsp. *equisimilis* (human origin), were proposed. A multicilose enzyme electrophoresis study confirmed the existence of the two subgroups in *S. dysgalactiae*, but the isolates included in each subspecies differed (Vieira *et al.*, 1998). In our study, the divergence between the two subspecies of *S. dysgalactiae* was high (10.2 %). This particular divergence was the highest for each of the gene sequence comparisons: *groEL* (6.1 %), *rpoB* (2.1 %) and *sodA* (12.9 %). The values found for other genes are reported in Table 4. DNA–DNA hybridization is considered the standard reference technique in bacterial identification. Values ≥70 % (commonly held as the threshold value) were found for strains of the two subspecies of *S. dysgalactiae* by Farrow & Collins (1984) (92–99 %) and Vieira *et al.* (1998) (75–77 %), although the reported values are very different. Therefore, *S. dysgalactiae* subsp. *dysgalactiae* and *S. dysgalactiae* subsp. *equisimilis* are representatives of the same species, although genetic analysis of more isolates could improve their clustering into subspecies.

**The ‘anginosus’ subgroup**

Another interesting result of this study was the high similarity between *S. constellatus* and *S. intermedius*. *S. anginosus*, *S. constellatus* and *S. intermedius* were described in 1906, 1924 and 1925, respectively. Previous studies have shown considerable genetic similarity within the group, even though it is biochemically heterogeneous (Farrow & Collins, 1984; Coykendall *et al.*, 1987), and proposed that the three species should be included in a single species. Based on DNA–DNA hybridization, Whiley & Beighton (1991) again proposed the recognition of three distinct species. However, a lack of reciprocity was observed in the levels of DNA relatedness between the type strains of *S. intermedius* and *S. constellatus* (36 and 65 %). Taken together, these observations demonstrated the complexity in defining species within the 'anginosus' subgroup. To understand better the phylogenetic evolution of this group, we determined the sequences of the recN, *groEL*, *gyrB*, *sodA* and *rpoB* genes for 30 clinical strains isolated in our laboratory as well as isolates of *S. anginosus* (1), *S. intermedius* (2) and *S. constellatus* (3) obtained from the CIP collection (Supplementary Table S3).

Two approaches were used to identify the isolates at the species level: phylogenetic analysis (Supplementary Fig. S1) and BLAST comparison when identification by phylogenetic analysis was impossible. In the phylogenetic analysis, sequences that grouped into one cluster with a significant bootstrap value at the corresponding node (>85 %) were considered to belong to the same species. Sequences that were isolated or which grouped on separated branches with non-significant bootstrap values at the corresponding nodes were subjected to analysis by the BLAST program. Strain identifications are detailed in Table 1. Two strains identified as *S. anginosus* in the CIP, strains CIP 103053 and CIP

Table 4. Nucleotide sequence divergence between subspecies

<table>
<thead>
<tr>
<th>Subspecies comparison</th>
<th>Mean divergence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>recN</td>
<td>10.2</td>
</tr>
<tr>
<td>groEL</td>
<td>6.1</td>
</tr>
<tr>
<td>gyrB</td>
<td>3.3</td>
</tr>
<tr>
<td>rpoB</td>
<td>2.1</td>
</tr>
<tr>
<td>sodA</td>
<td>12.9</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>1.3</td>
</tr>
<tr>
<td><em>S. dysgalactiae</em> subsp. <em>dysgalactiae</em> vs subsp. <em>equisimilis</em></td>
<td></td>
</tr>
<tr>
<td><em>S. equi</em> subsp. <em>equi</em> vs subsp. <em>zoopneumonicus</em></td>
<td>3.7</td>
</tr>
<tr>
<td><em>S. infantarius</em> subsp. <em>infantarius</em> vs subsp. <em>coli</em></td>
<td>2.1</td>
</tr>
<tr>
<td><em>S. galloxyticus</em> subsp. <em>galloxyticus</em> vs subsp. <em>macedonicus</em></td>
<td>2.0</td>
</tr>
<tr>
<td><em>S. galloxyticus</em> subsp. <em>galloxyticus</em> vs subsp. <em>pasteurianus</em></td>
<td>3.5</td>
</tr>
<tr>
<td><em>S. galloxyticus</em> subsp. <em>macedonicus</em> vs subsp. <em>pasteurianus</em></td>
<td>4.4</td>
</tr>
</tbody>
</table>

Table S1 and by Glazunova *et al.* (2009).
recN sequence comparison of streptococci

103055, grouped with the species S. intermedius and S. constellatus, respectively. Identification at the species level (S. intermedius) was impossible using recN gene sequence comparison for isolate S5. Isolates of S. constellatus and S. intermedius clustered in the phylogenetic tree inferred from sequence comparisons of the recN, groEL, sodA and gyrB genes (Supplementary Fig. S1). Using rpoB gene sequence comparison, the type strain of S. intermedius, CIP 103248T, clustered with the ‘anginosus’ species isolates, and another cluster included the other isolates that represented S. intermedius. However, this cluster was not related to the cluster that included the S. constellatus isolates. The identification of S. anginosus and S. constellatus isolates was more difficult using rpoB gene sequence comparison, as it was impossible to identify five isolates (S11, S13, S14, S17 and CIP 105057) at the species level. Identification of two isolates (S15 and S26) was impossible at the species level; they were identified as either S. anginosus or S. constellatus, depending on which gene sequence was analyzed. To try to explain the difficulty in differentiating isolates of S. anginosus and S. constellatus, inter- and intraspecies percentages of divergence were calculated for each species and for each gene (Supplementary Table S4). The interspecies divergence was lower for S. intermedius/S. constellatus and the intraspecies divergence was higher for S. anginosus. The most likely explanation is that genetic recombination has occurred between some isolates of S. anginosus and S. constellatus.

Conclusions

Partial sequence comparison of the recN gene differentiates members of the genus Streptococcus clearly and allows phylogenetic relationships between the different species to be inferred with a high level of statistical support. In comparison with previously studied genes (16S rRNA, groEL, gyrB, rpoB and sodA), recN represents the best tool to identify streptococci and to study evolution in this group of bacteria. Six clusters are well defined: the ‘pyogenes’ group, the ‘equinus’ group, the ‘salivarivus’ group, the ‘mutans’ group, the ‘hyovaginalis’ group and the ‘suis’ group. In general, however, the streptococci constitute a heterogeneous group, as the phylogenetic position of many representatives differs with the studied gene. DNA–DNA hybridization has been the ‘gold-standard’ method for identification at the species level, but partial gene sequencing analyses highlight some peculiarities, such as the high divergence between S. dysgalactiae subsp. dysgalactiae and S. dysgalactiae subsp. equisimilis or the high similarity between S. intermedius and S. constellatus. In the past, contradictory results have been published concerning these species and subspecies and, in future, the use of new approaches could result in re-evaluation of their taxonomic status.

References


Milinovich, G. J., Burrell, P. C., Pollitt, C. C., Bouvet, A. & Trott, D. J. (2008). Streptococcus henryi sp. nov. and Streptococcus caballi sp. nov.,


