Journal of Medical Microbiology (2007), **56**, 504–513 DOI 10.1099/jmm.0.47027-0

Identification of species of *Abiotrophia*, *Enterococcus*, *Granulicatella* and *Streptococcus* by sequence analysis of the ribosomal 16S–23S intergenic spacer region

Sheng Kai Tung,[1] Lee Jene Teng,[2] Mario Vaneechoutte,[3] Hung Mo Chen[4] and Tsung Chain Chang[1]

Correspondence
Tsung Chain Chang
tsungcha@mail.ncku.edu.tw

[1]Department of Medical Laboratory Science and Biotechnology, School of Medicine, National Cheng Kung University, 1 University Road, Tainan 701, Taiwan, ROC

[2]School of Medical Technology, National Taiwan University College of Medicine, Taipei, Taiwan, ROC

[3]Department of Clinical Chemistry, Microbiology and Immunology, Ghent University Hospital, Ghent, Belgium

[4]Division of Clinical Microbiology, Department of Pathology, National Cheng Kung University Hospital, Tainan, Taiwan, ROC

The feasibility of sequence analysis of the ribosomal 16S–23S intergenic spacer region (ITS) was evaluated for identification of 24 species of *Streptococcus*, one species of *Abiotrophia*, 18 species of *Enterococcus* and three species of *Granulicatella*. As GenBank currently lacks ITS sequence entries for many species of these four genera, the ITS sequences of 38 type strains were first sequenced and submitted to GenBank to facilitate species identification of these genera. Subsequently, the ITS sequences of 217 strains (84 reference strains and 133 clinical isolates) were determined and species identification was made by BLAST search for homologous sequences in public databases. Species other than *Streptococcus* contained multiple ITS fragments and only the shortest fragment was analysed. A total of 25 isolates (11.5 %) produced discrepant identification by ITS sequencing. The 25 discordant strains were analysed further by sequencing of the 16S rRNA gene for species clarification, and 21 were found to be identified correctly by ITS sequence analysis. The correct identification rate by ITS sequencing was 98.2 % (213/217). Several closely related enterococcal and streptococcal species/subspecies contained specific ITS signature sequences that were useful for differentiating these bacteria. In conclusion, ITS sequencing provides a useful approach towards identifying this group of pathogens on a molecular platform alongside 16S rRNA gene sequencing.

Received 25 October 2006
Accepted 15 December 2006

INTRODUCTION

Enterococci, nutritionally variant streptococci (NVS) – of which most have been allocated to the genera *Abiotrophia* and *Granulicatella* – and streptococci are Gram-positive, catalase-negative bacteria causing a wide variety of opportunistic and invasive infections (Facklam, 2002; Teixeira & Facklam, 2003). No single system of classification suffices for the differentiation of this heterogeneous group of organisms. Instead, classification depends on a combination of features including patterns of haemolysis on blood agar plates, antigenic composition, growth characteristics and biochemical reactions (Facklam, 2002). The routine procedures based on phenotypic tests do not allow unequivocal identification of many species of these bacteria (Hoshino *et al.*, 2005; Ruoff, 2003). In addition, some recently described species such as *Streptococcus cristatus*, *Streptococcus infantarius* and *Streptococcus gallolyticus* (Arbique *et al.*, 2004; Poyart *et al.*, 2002) have complicated the problems with identification of these micro-organisms.

A variety of molecular methods have been developed for the identification of enterococci, NVS and streptococci to species level. The targets used for molecular diagnoses include genes encoding rRNA (Clarridge *et al.*, 2001), RNA polymerase (Drancourt *et al.*, 2004), the

Abbreviations: ITS, intergenic spacer region; NVS, nutritionally variant streptococci.

The phylogenetic trees and table of ITS sequences submitted to GenBank are available as supplementary material in JMM Online.

Table 1. Bacterial strains used in this study and results of identification by ITS sequencing

Species	Reference strain*†	Clinical isolate†	No. of discrepant strains/total no. of strains
E. avium	CCUG 34661	2510, 2955, 5691, 7334, 7830	1/6
E. casseliflavus	ATCC 12755, ATCC 12817	363-90, 394-2, 2525, 3669-1, 3931, 4799, 5259-2, 6140, 8553, 7539, 9539	3/13
E. cecorum	ATCC BAA-150, CCUG 38939	2158, 3125, 160902	0/5
E. columbae	CCUG 27893		0/1
E. dispar	CCUG 37857		0/1
E. durans	CCUG 46232, CCUG 37858, CCUG 44816, LMG 16886	790	1/5
E. faecalis	ATCC 27332	351-21, 357-39, 622-2, 1003, 3783, 3938, 6728-1, 6777-2, 8092-2, 9147, 9470, d333, d714, b1083, g808	1/16
E. faecium	CCUG 34851, CCUG 46070	350-11, 350-71, 520-2, 1082, 1199, 2069, 2796, 6143, 7535n, 9672	0/12
E. flavescens	CCUG 30568, CCUG 30569		0/2
E. gallinarum	CCUG 29831, CCUG 34517	10, 3669-1, 3801, 6140, 6777-1, 9016, sh528	1/9
E. gilvus	LMG 13600		0/1
E. hirae	BCRC 11547, BCRC 12496	5502, 9227, 6545-a	0/5
E. mundtii	LMG 12308		0/1
E. pallens	CCUG 45554		0/1
E. pseudoavium	CCUG 44888		1/1
E. raffinosus	CCUG 37864, CCUG 37865		0/2
E. saccharolyticus	NCIMB 702609		0/1
E. villorum	LMG 19177		0/1
S. agalactiae		402, 496, 3229, 355-15, c133, c240, c273, d50, d766, 355-80, 356-73, d3, d36	0/13
S. alactolyticus	CCUG 41502, CCUG 41503		0/2
S. anginosus		1845	1/1
S. australis	CCUG 45974, CCUG 45975		0/2
S. bovis	CCUG 4214	7017, 8903-2, 9624-2	3/4
S. bovis biotype I‡		5833, 8727, sv238, sv279	0/4
S. bovis biotype II.1‡		195, 731, 1757, 1832, 3129, 5106, 6621, 7452, 7618	1/9
S. bovis biotype II.2‡		9, 42, 256, 0079, 1429, 1443, 2255-2, 3873, 6107, 6250, 6854, 7421, 7499, 9901, b439, sv268	0/16
S. canis	CCUG 27660		0/1
S. cristatus	CCUG 30424, CCUG 35233		1/2
S. dysgalactiae subsp. *equisimilis*	CCUG 27483		1/1
S. equi subsp. *equi*	BCRC 14191		0/1
S. equi subsp. *zooepidemicus*	BCRC 15414, CCUG 43890		0/2
S. equinus	CCUG 4213		0/1
S. gallolyticus subsp. *gallolyticus*	CCUG 35224, CCUG 46101, CCUG 46667		0/3
S. gallolyticus subsp. *macedonicus*	CCUG 39969, CCUG 43003		0/2
S. gallolyticus subsp. *pasteurianus*	CCUG 19454, CCUG 35885, CCUG 46034, CCUG 46150		0/4
S. lutetiensis	CCUG 38926, CCUG 43823		0/2
S. infantarius subsp. *infantarius*	CCUG 43820, CCUG 43821, CCUG 44960, CCUG 47548		0/4
S. infantis	CCUG 39817, CCUG 39818		0/2
S. iniae	CCUG 27623		0/1
S. mitis	ATCC 15910, ATCC 15914		2/2
S. mutans	BCRC 15255, BCRC 15256		2/2
S. oralis		9429	1/1

Table 1. cont.

Species	Reference strain*†	Clinical isolate†	No. of discrepant strains/total no. of strains
S. parasanguinis		2000	1/1
S. pneumoniae	ATCC 6301, ATCC 27336, ATCC 49619	104c, 2519, 9863, d163, e659-1, e659-2, e757, z289,	0/11
S. porcinus	CCUG 7982, CCUG 41363		0/2
S. pyogenes	ATCC 12344	481, 665-a, c207-2, 350-88, 352-9, 354-65, 355-99, 350-82	0/9
S. ratti	CCUG 27641		0/1
S. salivarius		9054, 9624-2	0/1
S. sobrinus	CCUG 21019, CCUG 35254		1/2
S. suis	CCUG 42755, CCUG 42756		0/2
S. thermophilus	CCUG 35458		0/1
S. vestibularis	CCUG 32749 B		1/1
S. uberis	ATCC 13386		1/1
S. urinalis	CCUG 41825		0/1
A. defectiva	CCUG 27805	1716, 4400, 8559, 9021, MJ	1/6
G. adiacens	CCUG 27811, CCUG 44406, CCUG 44407, CCUG 44408	828-1, 4972, 5684, 6313, 6943, 9807, YM	0/11
G. elegans	CCUG 13462, CCUG 26024, CCUG 27554		0/3
Total no. of strains	85	133	25/218

*ATCC: American Type Culture Collection, USA; BCRC, Bioresources Collection and Research Center, Taiwan; CCUG, Culture Collection of the University of Göteborg, Sweden; LMG, Laboratorium voor Microbiologie, Belgium; NCIMB, The National Collection of Industrial, Marine and Food Bacteria, UK.

†Underlined numbers indicate discrepant identification by ITS sequencing.

‡Three biotypes (I, II.1 and II.2) are recognized by the Rapid ID 32 STREP system. *S. bovis* biotype I and II.2 have been renamed as *S. gallolyticus* subsp. *gallolyticus* and *S. gallolyticus* subsp. *pasteurianus*, respectively (Schlegel *et al.*, 2000, 2003).

D-alanine-D-alanine ligase (Garnier *et al.*, 1997), the *β*-subunit of the elongation factor (Picard *et al.*, 2004), the manganese-dependent superoxide dismutase ($sodA_{int}$) (Poyart *et al.*, 1998, 2000), the heat-shock proteins (*groESL*) (Teng *et al.*, 2002) and the tRNA gene intergenic spacer region (ITS) (Baele *et al.*, 2000, 2001). Recently, optimal identification of non-haemolytic streptococci was performed by phylogenetic sequence analysis of four housekeeping genes (*ddl*, *gdh*, *rpoB* and *soda*) (Hoshino *et al.*, 2005).

The ribosomal 16S–23S ITS sequence has been suggested as a good candidate for bacterial identification and strain typing (Gürtler & Stanisich, 1996; Hassan *et al.*, 2003; Roth *et al.*, 1998). Sequences of the ITS region have been found to have low intraspecies variation and high interspecies divergence (Hassan *et al.*, 2003; Whiley *et al.*, 1995). In our previous study, the feasibility of using the ITS sequence to identify 11 species of viridans streptococci was established (Chen *et al.*, 2004). This study aimed to expand the results of this technique to cover a total of 57 species of *Abiotrophia*, *Enterococcus*, *Granulicatella* and *Streptococcus*.

METHODS

Bacterial strains. As GenBank currently lacks ITS sequence entries for many species of enterococci, NVS and streptococci, the ITS sequences of 38 type strains were determined and submitted to GenBank (see Supplementary Table S1 available in JMM Online) to facilitate ITS sequence comparison. Following this, a total of 217 strains comprising 84 reference and 133 clinical isolates (Table 1) was analysed by ITS sequencing. Reference strains were obtained from the American Type Culture Collection (ATCC; USA), the Bioresources Collection and Research Center (BCRC; Taiwan) and the Culture Collection of the University of Göteborg (CCUG; Sweden). Clinical isolates were obtained from the National Taiwan University Hospital (Taiwan), the National Cheng Kung University Medical Center (Taiwan) and the Ghent University Hospital (Belgium). Most of the clinical isolates were phenotypically identified using the Rapid ID 32 STREP system (bioMérieux) or by tRNA gene PCR (Baele *et al.*, 2000, 2001). A limited number of clinical isolates of NVS were tested in this study, as the feasibility of ITS sequence analysis for identification of these bacteria has been reported previously (Chen *et al.*, 2004). All strains were cultured on sheep blood agar, except for strains of *Abiotrophia* and *Granulicatella*, which were cultured on chocolate agar.

Amplification and sequencing of the ITS. A boiling method was used to extract DNA from bacteria (Millar *et al.*, 2000). The bacteria-specific universal primers 13 BF (5′-GTGAATACGTTCCCGGGCCT-

3′) and 6R (5′-GGGTTYCCCCRTTCRGAAAT-3′) (Y=C or T, R=A or G) (Relman, 1993) were used to amplify a DNA fragment encompassing a small portion of the 16S rRNA gene, the ITS and a small portion of the 23S rRNA gene, as described previously (Chen *et al.*, 2004). Sequences of the PCR products were edited to remove portions of the 16S and 23S rRNA genes to obtain the exact ITS sequences. Species of *Abiotrophia*, *Enterococcus* and *Granulicatella* produced multiple ITS amplicons by PCR, as revealed by agarose gel electrophoresis (Fig. 1). Under this condition, only the shortest ITS fragment of a strain, which was usually the dominant band (Fig. 1), was eluted from the agarose gel and sequenced. Moreover, the amplicons of *Enterococcus raffinosus* and *Granulicatella adiacens* eluted from gels could not be sequenced successfully due to the presence of multiple amplicons in the shortest fragments. For this reason, the eluted PCR products were cloned with a Topo TA cloning kit (Invitrogen). The ITS regions of positive clones were reamplified and sequenced in accordance with the manufacturer's instructions. All reference sequences submitted to GenBank were sequenced on both strands using primers 13 BF and 6R.

ITS sequence similarity and multiple sequence alignment. The PILEUP algorithm of the Wisconsin Genetics Computer Group (GCG) package (version 10.3; Accelrys) was used to calculate the interspecies ITS sequence similarity. To create an alignment of multiple sequences, the PRETTYBOX program of the GCG package was used. To simplify comparison, only the smallest amplicons of enterococci and NVS that possessed multiple ITS fragments (Fig. 1d–f) were used for sequence comparison and alignment.

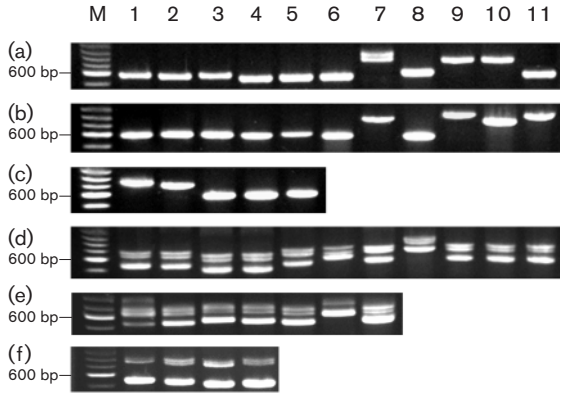

Fig. 1. Amplification of the ITS regions and separation of PCR products by 2 % agarose gel electrophoresis. Lane M, 100 bp ladder. (a) Lanes: 1, *S. agalactiae*; 2, *S. alactolyticus*; 3, *Streptococcus australis*; 4, *S. bovis*; 5, *Streptococcus canis*; 6, *S. cristatus*; 7, *S. downei*; 8, *S. dysgalactiae* subsp. *equisimilis*; 9, *S. equi* subsp. *equi*; 10, *S. equi* subsp. *zooepidemicus*; 11, *S. equinus*. (b) Lanes: 1, *Streptococcus gallolyticus* subsp. *gallolyticus*; 2, *S. gallolyticus* subsp. *macedonicus*; 3, *S. gallolyticus* subsp. *pasteurianus*; 4, *S. lutetiensis*; 5, *S. infantarius* subsp. *infantarius*; 6, *S. infantis*; 7, *S. iniae*; 8, *S. pneumoniae*; 9, *Streptococcus porcinus*; 10, *S. pyogenes*; 11, *Streptococcus ratti*. (c) Lanes: 1, *Streptococcus sobrinus*; 2, *Streptococcus suis*; 3, *S. thermophilus*; 4, *S. vestibularis*; 5, *Streptococcus urinalis*. (d) Lanes: 1, *E. avium*; 2, *E. casseliflavus*; 3, *Enterococcus cecorum*; 4, *E. columbae*; 5, *Enterococcus dispar*; 6, *E. durans*; 7, *E. faecalis*; 8, *E. faecium*; 9, *E. flavescens*; 10, *Enterococcus gallinarum*; 11, *Enterococcus gilvus*. (e) Lanes: 1, *E. hirae*; 2, *Enterococcus mundtii*; 3, *Enterococcus pallens*; 4, *Enterococcus pseudoavium*; 5, *E. raffinosus*; 6, *Enterococcus saccharolyticus*; 7, *Enterococcus villorum*. (f) Lanes: 1, *A. defectiva*; 2, *G. adiacens*; 3, *Granulicatella balaenopterae*; 4, *G. elegans*.

Species identification by ITS sequencing. A total of 217 strains including 84 reference strains and 133 clinical isolates was analysed (Table 1). The ITS region of each strain was amplified by PCR. The amplicons were sequenced with forward primer 13 BF and, in case the forward sequence was not readable, reverse primer 6R was also used for generating the sequences. PCR products of all species of streptococci (Fig. 1a–c) were sequenced directly, except for *Streptococcus downei*, which apparently contained two ITS fragments (Fig. 1a, lane 7). For *S. downei*, enterococci and NVS that possessed multiple ITS fragments (Fig. 1d–f), only the smallest amplicon was eluted from the agarose gel and sequenced for species identification. Species identification was done by searching databases using the BLAST (http://www.ncbi.nlm.nih.gov/blast/) sequence analysis tool. Species identification was determined from the best-scoring reference sequence of the BLAST output and whether the best-scoring reference sequence in the databases had a sequence identity of $\geqslant$98 % with the query sequence. Strains producing discrepant species names by ITS sequence analysis were analysed further by sequencing of the near-complete 16S rRNA gene (Relman, 1993).

Phylogenetic analysis. Unrooted phylogenetic trees were constructed by the neighbour-joining method listed in the MEGA (version 3.0) analytical package (Kumar *et al.*, 2004). For neighbour-joining analysis, the distance between two ITS sequences was calculated using Kimura's two-parameter model (Li & Tsoi, 2002). The robustness of the neighbour-joining method was evaluated statistically by bootstrap analysis with 1000 bootstrap samples.

RESULTS AND DISCUSSION

Amplification and sequencing of ITS fragments

The ITS fragments of *Abiotrophia*, *Enterococcus*, *Granulicatella* and *Streptococcus* strains were amplified by PCR. A single amplicon was observed for each species of *Streptococcus* except for *S. downei*, which possessed two ITSs (Fig. 1a, lane 7). Multiple PCR products were observed for strains of *Abiotrophia*, *Enterococcus* and *Granulicatella* (Fig. 1d–f). For a strain that possessed multiple ITSs with different lengths and sequences, only the shortest fragment (usually the dominant band on the agarose gel) was eluted and sequenced. However, the amplicons eluted from *E. raffinosus* (Fig. 1e, lane 5) still contained two fragments with lengths of 224 and 246 bp, respectively, as revealed by cloning of the PCR products (see Supplementary Table S1 available in JMM Online). In addition, the PCR product eluted from *G. adiacens* (Fig. 1f, lane 2) contained two amplicons of identical size (222 bp) but which differed in sequence, as revealed by cloning and sequencing of the eluted amplicons. As specific primers were used for ITS amplification, multiple bands on agarose gels implied the presence of different rRNA operons in species of the three genera. The presence of multiple ITS operons has also been found in many species of *Acinetobacter* (Chang *et al.*, 2005). To simplify sequence comparison among species, only the shortest ITS fragments of the three genera were used for sequence determination.

The size of the ITS ranged from 189 bp (*Enterococcus columbae*) to 601 bp (*Streptococcus equi* subsp. *equi*) (see Supplementary Table S1 available in JMM Online). At the start of this study, GenBank contained ITS sequences of the following 22 species: *Abiotrophia defectiva*, *Enterococcus durans*, *Enterococcus faecalis*, *Enterococcus faecium*, *Enterococcus hirae*, *G. adiacens*, *Streptococcus agalactiae*, *Streptococcus alactolyticus*, *Streptococcus anginosus*, *Streptococcus bovis*, *Streptococcus constellatus* subsp. *constellatus*, *Streptococcus intermedius*, *Streptococcus mitis*, *Streptococcus mutans*, *Streptococcus oralis*, *Streptococcus parasanguinis*, *Streptococcus parauberis*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus salivarius*, *Streptococcus sanguinis* and *Streptococcus uberis*. After deposition of the 38 ITS sequences determined in this study (see Supplementary Table S1 available in JMM Online), most clinically relevant species of the four genera *Abiotrophia*, *Enterococcus*, *Granulicatella* and *Streptococcus* could be identified by ITS sequencing and BLAST search analysis.

Interspecies sequence similarity of the ITS region

Pairwise comparisons of the ITS sequences between the type strains of any two given enterococcal species showed that sequence similarity ranged from 0.42 (*Enterococcus avium* vs *E. faecium*) to 0.996 (*Enterococcus casseliflavus* vs *Enterococcus flavescens*) (data not shown). In general, interspecies ITS sequence similarities were less than 0.90. The interspecies ITS sequence similarities of streptococci ranged from 0.36 (*S. anginosus* vs *Streptococcus iniae*) to as high as 0.95–1.0 (*S. pneumoniae* vs *S. mitis*) (Chen *et al.*, 2004) and 1.0 (*S. bovis* vs *Streptococcus equinus*). However, based on the 16S rRNA gene sequence and total DNA–DNA hybridization, *S. equinus* and *S. bovis* were proposed

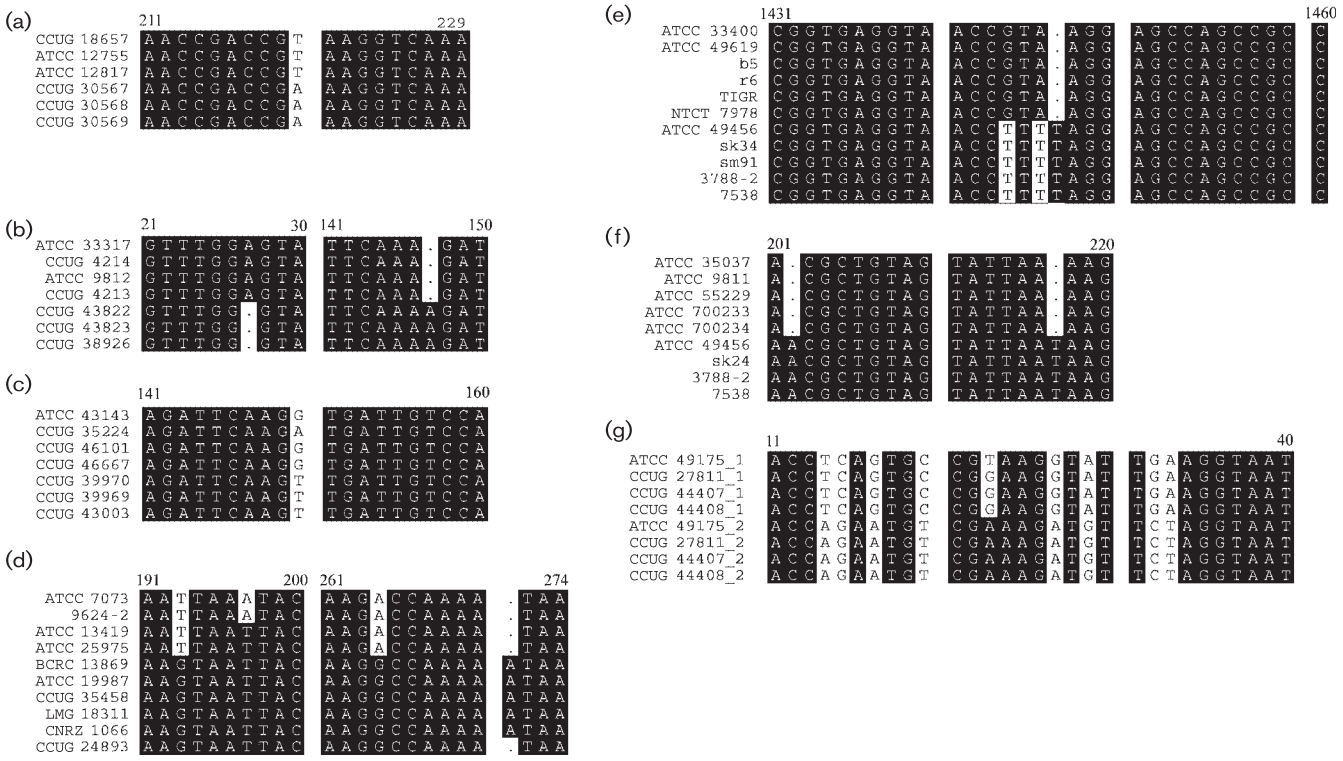

Fig. 2. Sequence alignment of the partial ITS regions or the 3′ end regions of the 16S rRNA gene to reveal species- or subspecies-specific signature sequences: (a) *E. casseliflavus* (CCUG 18657^T, ATCC 12755, ATCC 12817) and *E. flavescens* (CCUG 30567^T, CCUG 30568, CCUG 30569); (b) *S. bovis* (ATCC 33317^T, CCUG 4214), *S. equinus* (ATCC 9812^T, CCUG 4213) and *S. lutetiensis* (CCUG 43822^T, CCUG 43823, CCUG 38926); (c) *S. gallolyticus* subsp. *gallolyticus* (ATCC 43143^T, CCUG 35224, CCUG 46101, CCUG 46667) and *S. gallolyticus* subsp. *macedonicus* (CCUG 39970^T, CCUG 39969, CCUG 43003); (d) *S. salivarius* (ATCC 7073^T, 9624-2, ATCC 13419, ATCC 25975), *S. thermophilus* (BCRC 13869, ATCC 19987, CCUG 35458, LMG 18311, CNRZ 1066) and *S. vestibularis* (CCUG 24893^T); (e) The 3′ end sequences of 16S rRNA genes of *S. pneumoniae* (ATCC 33400^T, ATCC 49619, b5, r6, TIGR, NCTC 7978) and *S. mitis* (ATCC 49456^T, sk34, sm91, 3788-2, 7538); (f) *S. oralis* (ATCC 35037^T, ATCC 9811, ATCC 55229, ATCC 700233, ATCC 700234) and *S. mitis* (ATCC 49456^T, sk24, 3788-2, 7538); (g) the two short ITS fragments of strains of *G. adiacens* (ATCC 49175^T, CCUG 27811, CCUG 44407 and CCUG 44408). Type strains are indicated with a superscript T.

Table 2. List of 25 strains that produced discrepant identification by ITS sequence analysis

Strain	Species received as:	Species identification by:		Result of discrepant analysis
		ITS sequence* (%)	16S rRNA gene sequence† (%)	
Reference strain				
CCUG 44888	*E. pseudoavium*	*E. avium* (99.6) *E. pseudoavium* (85)	*E. avium* (99.1) *E. pseudoavium* (98.8)	*E. avium*
CCUG 35233‡	*S. cristatus*	*S. oralis* (99.6) *S. cristatus* (84.1)	*S. oralis* (98) *S. mitis* (98)	*S. oralis*
CCUG 27483	*S. dysgalactiae* subsp. *equisimilis*	*S. dysgalactiae* subsp. *dysgalactiae* (100) *S. dysgalactiae* subsp. *equisimilis* (100)	*S. dysgalactiae* (100)	*S. dysgalactiae* subsp. *equisimilis*
ATCC 15910	*S. mitis*	*S. vestibularis* (100) *S. thermophilus* (99) *S. salivarius* (99)	*S. vestibularis* (99.3) *S. thermophilus* (99.3) *S. salivarius* (99.3)	*S. vestibularis*
ATCC 15914‡	*S. mitis*	*S. oralis* (99.1) *S. mitis* (94)	*S. oralis* (98) *S. mitis* (98) *S. pneumoniae* (98)	*S. oralis*
BCRC 15255	*S. mutans*	*S. ratti* (100) *S. mutans* (47)	*S. ratti* (99) *S. mutans* (92)	*S. ratti*
BCRC 15256	*S. mutans*	*S. ratti* (100) *S. mutans* (47)	*S. ratti* (99) *S. mutans* (92)	*S. ratti*
CCUG 35254	*S. sobrinus*	*S. mutans* (100) *S. sobrinus* (41)	*S. mutans* (99) *S. sobrinus* (90)	*S. mutans*
CCUG 32749B§	*S. vestibularis*	*S. salivarius* (99.6) *S. vestibularis* (98.5)	*S. salivarius* (99.7) *S. vestibularis* (99.1)	*S. salivarius*
ATCC 13386	*S. uberis*	*S. parauberis* (99.7) *S. uberis* (52)	*S. parauberis* (99) *S. uberis* (92)	*S. parauberis*
Clinical isolate‖				
4400	*A. defectiva* (52.1)¶/*S. suis* (D)	*A. defectiva* (99.4)	*A. defectiva* (99)	*A. defectiva*
2510	*E. avium* (UA)	*E. raffinosus* (99.1)	*E. raffinosus* (100)	*E. raffinosus*
3669-1	*E. casseliflavus* (UA)	*E. gallinarum* (99.5)	*E. gallinarum* (100)	*E. gallinarum*
3931	*E. casseliflavus* (55.4)/*E. gallinarum* (45.5)	*E. gallinarum* (99.5)	*E. gallinarum* (100)	*E. gallinarum*
9539	*E. casseliflavus* (UA)	*E. cecorum* (99)	*E. cecorum* (99)	*E. cecorum*
790	*E. durans* (76.7)/*E. hirae* (23.1)	*E. hirae* (100)	*E. hirae* (99)	*E. hirae*
622-2	*E. faecalis* (UA)	*E. raffinosus* (100)	*E. raffinosus* (100)	*E. raffinosus*
sh528	*E. gallinarum* (53.6)/*E. faecium* (46.3) (D)	*E. faecium* (100)	*E. faecium* (98.2)	*E. faecium*
1845	*S. anginosus* (D)	*S. gallolyticus* subsp. *pasteurianus* (100)	*S. gallolyticus* subsp. *pasteurianus* (99.4)	*S. gallolyticus* subsp. *pasteurianus*
7017	*S. bovis* (UA)	*S. parasanguinis* (98.5)	*S. parasanguinis* (98)	*S. parasanguinis*
9624-2	*S. bovis* (UA)	*S. salivarius* (99.2)	*S. salivarius* (99)	*S. salivarius*
5106#	*S. bovis* biotype II.1	*S. lutetiensis* (98.9)	*S. gallolyticus* subsp. *pasteurianus* (99.4)	Not resolved
9429	*S. oralis* (UA)	Not identified	*S. mitis* (99.6)	Not resolved
2000	*S. parasanguinis* (UA)	*S. infantis* (99.4)	*S. infantis* (99.2)	*S. infantis*

Table 2. cont.

Strain	Species received as:	Species identification by:		Result of discrepant analysis
		ITS sequence* (%)	16S rRNA gene sequence† (%)	
8903-2**	*S. thermophilus* (UA)	*S. pneumoniae* (99.1)	*S. oralis* (99.2) *S. pseudopneumoniae* (99.7) *S. pneumoniae* (99.6)	Not resolved

D, Doubtful identification by the Rapid ID 32 STREP system; UA, unacceptable profile by the Rapid ID 32 STREP system.
*The value in parentheses is the ITS sequence identity of the strain analysed compared with the type strain of the indicated species.
†Results of BLAST.
‡The ITS sequences of strains CCUG 35233 and ATCC 15914 had the ITS signature sequence of *S. oralis* (Fig. 2f).
§The ITS sequences of strain CCUG 32749B had the ITS signature sequence of *S. salivarius* (Fig. 2d).
||Most clinical isolates were identified by the Rapid ID 32 STREP system or tRNA gene PCR.
¶The value in parentheses is the identification percentage reported by the Rapid ID 32 STREP system.
#Strain 5106 was identified as *S. infantarius* subsp. *infantarius* by sequence analysis of the *sodA* gene (99 % identity).
**Strain 8903-2 was optochin resistant.

to form a single DNA group, i.e. the same species (Facklam, 2002; Schlegel *et al.*, 2003). In addition, a low level of interspecies ITS sequence homology was observed among the four species of NVS, with sequence similarities ranging from 0.37 (*A. defectiva* vs *Granulicatella elegans*) to 0.57 (*G. adiacens* vs *G. elegans*) (data not shown).

Signature sequences of some enterococcal and streptococcal species

Only one nucleotide difference was observed in the ITS regions (229 bp) in strains of *E. casseliflavus* and *E. flavescens*. Multiple ITS sequence alignment revealed that nt 220 was T in strains of *E. casseliflavus*, but A in strains of *E. flavescens* (Fig. 2a). *E. casseliflavus* and *E. flavescens* are intrinsically vancomycin resistant (possessing the *van*C genotype), motile and capable of producing pigments. However, Baele *et al.* (2000) pointed out that *E. casseliflavus* and *E. flavescens* are most probably synonymous, as revealed by tRNA ITS PCR and other evidence (Descheemaeker *et al.*, 1997; Teixeira *et al.*, 1997).

For the three species of the *S. bovis* complex (*S. bovis*, *S. equinus* and *Streptococcus lutetiensis*), ITS sequence alignment demonstrated that there was a single nucleotide deletion at position 27 in strains of *S. lutetiensis*, whereas strains of *S. bovis* and *S. equinus* had a single nucleotide deletion at position 147 (Fig. 2b). Strains of *S. bovis* have been found to be associated with endocarditis and colonic cancer (Ruoff *et al.*, 1989; Tripodi *et al.* 2004), so accurate identification of the micro-organism is of clinical importance. For strains of the two subspecies of *S. gallolyticus* (subsp. *gallolyticus* and *macedonicus*), there was only one nucleotide difference in the ITS regions (Fig. 2c). For strains of subspecies *gallolyticus*, nt 150 was G or A, but in strains of subspecies *macedonicus* this was T (Fig. 2c).

S. salivarius, *Streptococcus thermophilus* and *Streptococcus vestibularis* are members of the *S. salivarius* group and are closely related micro-organisms. Multiple ITS sequence alignment revealed that nt 193 and 264 were T and A, respectively, in strains of *S. salivarius*, whilst both nucleotides were G in strains of *S. thermophilus* and *S. vestibularis* (Fig. 2d). In addition, there was a single nucleotide (A) insertion at position 271 in strains of *S. thermophilus* and this insertion could be used to differentiate this species from *S. vestibularis*.

S. mitis and *S. pneumoniae* are members of the *S. mitis* group and have a high level of ITS sequence homology (Chen *et al.*, 2004). However, sequence analysis of the co-amplified 3′ end of the 16S rRNA gene revealed that the nucleotides at positions 1444 and 1446 were G and A, respectively, in strains of *S. pneumoniae*, whilst both nucleotides were T in strains of *S. mitis* (Fig. 2e). Moreover, there was a single nucleotide deletion at position 1447 in strains of *S. pneumoniae*. The ITS sequences of *S. mitis* and *S. oralis* also had high similarity (Chen *et al.*, 2004). In comparison with *S. oralis*, two single nucleotide

insertions at positions 202 (A) and 217 (T) were found in strains of *S. mitis* (Fig. 2f). In addition, *S. pneumoniae*, *S. oralis* and *S. mitis* are difficult to differentiate by sequencing of the 16S rRNA gene (Bosshard *et al.*, 2004; Kawamura *et al.*, 1995). Phylogenetic trees constructed using the genes encoding 16S rRNA (Bentley *et al.*, 1991; Kawamura *et al.*, 1995), *sodA* (Poyart *et al.*, 2000, 2002), *groESL* (Teng *et al.*, 2002) and the ITS region (Chen *et al.*, 2004) grouped *S. mitis*, *S. oralis* and *S. pneumoniae* together. This is the first report indicating the presence of signature sequences in some species of enterococci and streptococci. These signature sequences are useful for differentiating these closely related enterococcal and streptococcal species/subspecies.

Identification of reference strains by ITS sequence analysis

To validate ITS sequencing for species identification, a total of 84 reference strains was analysed (Table 1). Nine strains produced discrepant identification and one strain (*Streptococcus dysgalactiae* subsp. *equisimilis* CCUG 27483) was not identified by ITS sequencing (Table 2). Strain CCUG 27483 also was not identified to subspecies level by 16S rRNA gene sequencing (Table 2). Identification of six of the nine discrepant strains, as obtained by ITS sequencing, was confirmed by 16S rRNA gene sequencing (Table 2). However, unambiguous identification of the remaining three strains (*S. cristatus* CCUG 35233, and *S. mitis* ATCC 15910 and ATCC 15914) was not obtained by 16S rRNA gene sequencing. The identification of *S. cristatus* CCUG 35233, and *S. mitis* ATCC 15910 and ATCC 15914 as *S. oralis*, *S. vestibularis* and *S. oralis*, respectively, by ITS sequencing was further confirmed by their signature sequences shown in Fig. 2(d, f). In brief, nine of the ten discrepant reference strains were considered to be given incorrect species names but were correctly identified by ITS sequence analysis, as confirmed by 16S rRNA gene sequencing (Table 2). Therefore, the identification rate of reference strains by ITS sequencing was 98.8 % (83/84). The mislabelling of up to 10.7 % (9/84) of the reference strains reflects the difficulties in identifying some streptococcal and enterococcal species by phenotypic methods.

Identification of clinical isolates by ITS sequence analysis

A total of 133 clinical isolates was identified by ITS sequencing (Table 1). Most clinical isolates were first identified to species level with the Rapid ID 32 STREP system or using tRNA gene PCR (Baele *et al.*, 2000, 2001). Overall, there was an agreement of 88.7 % (118/133) in species identification between the Rapid ID 32 STREP/tRNA gene PCR and ITS sequence analysis, with the remaining 15 isolates (11.3 %) producing discrepant identification by the two methods (Table 2). Clinical isolates producing an unacceptable profile, a low identification percentage or doubtful identification by the Rapid ID 32 STREP system were identified accurately to species level by ITS sequence analysis, as confirmed by 16S RNA gene sequencing (Table 2).

Three clinical isolates (5106, 9429 and 8903-2) were not identified by ITS sequencing (Table 2). Isolate 5106 was identified as *S. lutetiensis* by ITS sequence analysis, but was identified as *S. gallolyticus* subsp. *pasteurianus* by 16S rRNA gene sequencing and as *S. infantarius* subsp. *infantarius* (99 % identity) by sequence analysis of the *sodA* gene (Poyart *et al.*, 1998). Isolate 5106 did not display the signature sequence of *S. lutetiensis* (Fig. 2b). Therefore, the identity of isolate 5106 was not resolved. Isolate 9429 was not identified by ITS sequence analysis (<95 % identity to all known sequences), but was identified as *S. mitis* by 16S rRNA gene sequencing. Isolate 8903-2 (optochin resistant) was identified as *S. pneumoniae* by ITS sequencing and was identified as *Streptococcus pseudopneumoniae* or as *S. pneumoniae* by 16S RNA gene sequencing. Isolate 8903-2 did not display the ITS signature sequence typical of *S. pneumoniae* (Fig. 2h) and therefore the identity of the strain was not resolved. Although isolate 2000 (*S. parasanguinis*) was identified as *Streptococcus infantis* by ITS sequencing (99.4 % identity), a BLAST search of its 16S rRNA gene sequence revealed that this strain might be either *S. infantis* or *S. oralis* (Table 2). In brief, 97.7 % (130/133) of the clinical isolates were identified correctly by ITS sequencing. If reference strains and clinical isolates were taken together, the identification rate was 98.2 % (213/217) by ITS sequencing.

Sequence analysis of the 16S rRNA gene (approx. 1.5 kb) has been widely used for bacterial identification (Clarridge, 2004; Relman, 1993). As the ITS region is relatively short (from 189 to 601 bp) (see Supplementary Table S1 available in JMM Online), sequencing of the ITS region would be more efficient and accurate than sequencing of the 16S rRNA gene. This study shows an interesting approach towards identifying this group of pathogens on a molecular platform alongside 16S rRNA gene sequencing. It should be noted that species other than *Streptococcus* contained multiple ITS fragments and only the shortest fragment was analysed in this study (Fig. 1). A limitation of using the ITS sequence for bacterial identification is the relatively limited database compared with that of 16S rRNA genes. However, ITS sequences in the public database have increased substantially in recent years (D'Auria *et al.*, 2006).

Phylogenetic analysis

The phylogenetic tree derived from the ITS sequences of *Streptococcus* (35 species) is presented in Supplementary Fig. S1 available in JMM Online. The tree was generally in agreement with the tree constructed from the 16S rRNA gene sequences (see Supplementary Fig. S2 available in JMM Online). Species of each of the following groups (*S. bovis* group, *S. salivarius* group, *S. mitis* group, *S. sanguinis* group, *S. anginosus* group and *S. mutans* group) were clustered together in the ITS sequence-based tree, as was the case in the tree derived from the 16S rRNA genes.

Compared with the tree based on 16S rRNA gene sequences, two streptococcal species (*S. parasanguinis* and *S. anginosus*) branched differently or formed different clusters in the ITS sequence-based tree. It was interesting to find that the phylogenetic tree derived from ITS sequences of streptococci generally displayed a better resolution for basal branches compared with that constructed from the sequences of the 16S rRNA genes. For several major deep branches, the bootstrap values were too low in the 16S rRNA gene sequence-based tree to permit reliable interpretations.

ACKNOWLEDGEMENTS

This project was supported by grants (NSC 94-2323-B006-007, NSC 94-2320-B006-078) from the National Science Council.

REFERENCES

Arbique, J. C., Poyart, C., Trieu-Cuot, P., Quesne, G., Carvalho Mda, G., Steigerwalt, A. G., Morey, R. E., Jackson, D., Davidson, R. J. & Facklam, R. R. (2004). Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. *J Clin Microbiol* **42**, 4686–4696.

Baele, M., Baele, P., Vaneechoutte, M., Storms, V., Butaye, P., Devriese, L. A., Gillis, M. & Haesebrouck, F. (2000). Application of tRNA intergenic spacer PCR for identification of *Enterococcus* species. *J Clin Microbiol* **38**, 4201–4207.

Baele, M., Storms, V., Haesebrouck, F., Devriese, L. A., Gillis, M., Verschraegen, G., de Baere, T. & Vaneechoutte, M. (2001). Application and evaluation of the interlaboratory reproducibility of tRNA intergenic length polymorphism analysis (tDNA-PCR) for identification of *Streptococcus* species. *J Clin Microbiol* **39**, 1436–1442.

Bentley, R. W., Leigh, J. A. & Collins, M. D. (1991). Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences. *Int J Syst Bacteriol* **41**, 487–494.

Bosshard, P. P., Abels, S., Altwegg, M., Bottger, E. C. & Zbinden, R. (2004). Comparison of conventional and molecular methods for identification of aerobic catalase-negative Gram-positive cocci in the clinical laboratory. *J Clin Microbiol* **42**, 2065–2073.

Chang, H. C., Wei, Y. F., Dijkshoorn, L., Vaneechoutte, M., Tang, C. T. & Chang, T. C. (2005). Species-level identification of isolates of the *A. calcoaceticus–A. baumannii* complex by sequence analysis of the 16S–23S rRNA gene spacer region. *J Clin Microbiol* **42**, 1632–1639.

Chen, C. C., Teng, L. J. & Chang, T. C. (2004). Identification of clinically relevant viridans group streptococci by sequence analysis of the 16S–23S ribosomal DNA spacer region. *J Clin Microbiol* **42**, 2651–2657.

Clarridge, J. E., III (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* **17**, 840–862.

Clarridge, J. E., III, Attorri, S. M., Zhang, Q. & Bartell, J. (2001). 16S ribosomal DNA sequence analysis distinguishes biotypes of *Streptococcus bovis*: *Streptococcus bovis* biotype II/2 is a separate genospecies and the predominant clinical isolate in adult males. *J Clin Microbiol* **39**, 1549–1552.

D'Auria, G., Pushker, R. & Rodriguez-Valera, F. (2006). IWoCS: analyzing ribosomal intergenic transcribed spacers configuration and taxonomic relationships. *Bioinformatics* **22**, 527–531.

Descheemaeker, P., Lammens, C., Pot, B., Vandamme, P. & Goossens, H. (1997). Evaluation of arbitrarily primed PCR analysis and pulsed-field gel electrophoresis of large genomic DNA fragments for identification of enterococci important in human medicine. *Int J Syst Bacteriol* **47**, 555–561.

Drancourt, M., Roux, V., Fournier, P. E. & Raoult, D. (2004). *rpoB* gene sequence-based identification of aerobic Gram-positive cocci of the genera *Streptococcus*, *Enterococcus*, *Gemella*, *Abiotrophia*, and *Granulicatella*. *J Clin Microbiol* **42**, 497–504.

Facklam, R. (2002). What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin Microbiol Rev* **15**, 613–630.

Garnier, F., Gerbaud, G., Courvalin, P. & Galimand, M. (1997). Identification of clinically relevant viridans group streptococci to the species level by PCR. *J Clin Microbiol* **35**, 2337–2341.

Gürtler, V. & Stanisich, V. A. (1996). New approaches to typing and identification of bacteria using the 16S–23S rDNA spacer region. *Microbiology* **142**, 3–16.

Hassan, A. A., Khan, I. U., Abdulmawjood, A. & Lammler, C. (2003). Inter- and intraspecies variations of the 16S–23S rDNA intergenic spacer region of various streptococcal species. *Syst Appl Microbiol* **26**, 97–103.

Hoshino, T., Fujiwara, T. & Kilian, M. (2005). Use of phylogenetic and phenotypic analyses to identify nonhemolytic streptococci isolated from bacteremic patients. *J Clin Microbiol* **43**, 6073–6085.

Kawamura, Y., Hou, X. G., Sultana, F., Miura, H. & Ezaki, T. (1995). Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int J Syst Bacteriol* **45**, 406–408.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.

Li, Y. J. & Tsoi, S. C. (2002). Phylogenetic analysis of vertebrate lactate dehydrogenase (LDH) multigene families. *J Mol Evol* **54**, 614–624.

Millar, B. C., Jiru, X., Moore, J. E. & Earle, J. A. (2000). A simple and sensitive method to extract bacterial, yeast and fungal DNA from blood culture material. *J Microbiol Methods* **42**, 139–147.

Picard, F. J., Ke, D., Boudreau, D. K., Boissinot, M., Huletsky, A., Richard, D., Ouellette, M., Roy, P. H. & Bergeron, M. G. (2004). Use of *tuf* sequences for genus-specific PCR detection and phylogenetic analysis of 28 streptococcal species. *J Clin Microbiol* **42**, 3686–3695.

Poyart, C., Quesne, G., Coulon, S., Berche, P. & Trieu-Cuot, P. (1998). Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J Clin Microbiol* **36**, 41–47.

Poyart, C., Quesnes, G. & Trieu-Cuot, P. (2000). Sequencing the gene encoding manganese-dependent superoxide dismutase for rapid species identification of enterococci. *J Clin Microbiol* **38**, 415–418.

Poyart, C., Quesne, G. & Trieu-Cuot, P. (2002). Taxonomic dissection of the *Streptococcus bovis* group by analysis of manganese-dependent superoxide dismutase gene (*sodA*) sequences: reclassification of '*Streptococcus infantarius* subsp. *coli*' as *Streptococcus lutetiensis* sp. nov. and of *Streptococcus bovis* biotype 11.2 as *Streptococcus pasteurianus* sp. nov. *Int J Syst Evol Microbiol* **52**, 1247–1255.

Relman, D. A. (1993). Universal bacterial 16S rDNA amplification and sequencing. In *Diagnostic Molecular Microbiology*, pp. 489–495. Edited by D. H. Persing, T. F. Smith, F. C. Tenover & T. J. White. Washington, DC: American Society for Microbiology.

Roth, A., Fischer, M., Hamid, M. E., Michalke, S., Ludwig, W. & Mauch, H. (1998). Differentiation of phylogenetically related slowly growing mycobacteria based on 16S–23S rRNA gene internal transcribed spacer sequences. *J Clin Microbiol* **36**, 139–147.

Ruoff, K. L. (2003). *Aerococcus, Abiotrophia*, and other infrequently isolated aerobic catalase-negative, Gram-positive cocci. In *Manual of Clinical Microbiology*, 8th edn, pp. 434–444. Edited by P. R. Murray, E. J. Baron, J. H. Jorgensen, M. A. Pfaller & R. H. Yolken. Washington, DC: American Society for Microbiology.

Ruoff, K. L., Miller, S. I., Garner, C. V., Ferraro, M. J. & Calderwood, S. B. (1989). Bacteremia with *Streptococcus bovis* and *Streptococcus salivarius*: clinical correlates of more accurate identification of isolates. *J Clin Microbiol* **27**, 305–308.

Schlegel, L., Grimont, F., Collins, M. D., Regnault, B., Grimont, P. A. & Bouvet, A. (2000). *Streptococcus infantarius* sp. nov., *Streptococcus infantarius* subsp. *infantarius* subsp. nov. and *Streptococcus infantarius* subsp. *coli* subsp. nov., isolated from humans and food. *Int J Syst Evol Microbiol* **50**, 1425–1434.

Schlegel, L., Grimont, F., Ageron, E., Grimont, P. A. & Bouvet, A. (2003). Reappraisal of the taxonomy of the *Streptococcus bovis*/*Streptococcus equinus* complex and related species: description of *Streptococcus gallolyticus* subsp. *gallolyticus* subsp. nov., *S. gallolyticus* subsp. *macedonicus* subsp. nov. and *S. gallolyticus* subsp. *pasteurianus* subsp. nov. *Int J Syst Evol Microbiol* **53**, 631–645.

Teixeira, L. M. & Facklam, R. R. (2003). *Enterococcus*. In *Manual of Clinical Microbiology*, 8th edn, pp. 422–433. Edited by P. R. Murray, E. J. Baron, J. H. Jorgensen, M. A. Pfaller & R. H. Yolken. Washington, DC: American Society for Microbiology.

Teixeira, L. M., Carvalho, M. G., Merquior, V. L., Steigerwalt, A. G., Teixeira, M. G., Brenner, D. J. & Facklam, R. R. (1997). Recent approaches on the taxonomy of the enterococci and some related microorganisms. *Adv Exp Med Biol* **418**, 397–400.

Teng, L. J., Hsueh, P. R., Tsai, J. C., Chen, P. W., Hsu, J. C., Lai, H. C., Lee, N. & Ho, S. W. (2002). *groESL* sequence determination, phylogenetic analysis, and species differentiation for viridans group streptococci. *J Clin Microbiol* **40**, 3172–3178.

Tripodi, M. F., Adinolfi, L. E., Ragone, E., Durante Mangoni, E., Fortunato, R., Iarussi, D., Ruggiero, G. & Utili, R. (2004). *Streptococcus bovis* endocarditis and its association with chronic liver disease: an underestimated risk factor. *Clin Infect Dis* **38**, 1394–1400.

Whiley, R. A., Duke, B., Hardie, J. M. & Hall, L. M. (1995). Heterogeneity among 16S–23S rRNA intergenic spacers of species within the '*Streptococcus milleri* group'. *Microbiology* **141**, 1461–1467.