Microbiology (2005), **151**, 2141–2150 **DOI** 10.1099/mic.0.27840-0

Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes

Sabri M. Naser,[1,2] Fabiano L. Thompson,[2] Bart Hoste,[2] Dirk Gevers,[1,3] Peter Dawyndt,[1] Marc Vancanneyt[2] and Jean Swings[1,2]

[1,2]Laboratory of Microbiology[1] and BCCM/LMG Bacteria Collection[2], Ghent University, K. L. Ledeganckstraat 35, Ghent 9000, Belgium

[3]Bioinformatics and Evolutionary Genomics, Ghent University/VIB, Technologiepark 927, Ghent 9052, Belgium

Correspondence
Sabri M. Naser
Sabri.Naser@Ugent.be

Received 19 December 2004
Revised 20 April 2005
Accepted 20 April 2005

The aim of this study was to evaluate the use of RNA polymerase α subunit (*rpoA*) and phenylalanyl-tRNA synthase (*pheS*) gene sequences as species identification tools for enterococci. Ninety-six representative strains comprising all currently recognized *Enterococcus* species were examined. *rpoA* gene sequences generated a robust classification into species groups similar to the one based on 16S rRNA gene sequence analysis. On the other hand, the *pheS* gene is a fast-evolving clock even better suited for species delineation than the *rpoA* gene, but not for recognition of species groups within *Enterococcus* as determined by both *rpoA* and 16S rRNA genes. All enterococcal species were clearly differentiated on the basis of their *rpoA* and *pheS* sequences. Evaluation of intraspecies variation showed that both *rpoA* and *pheS* genes have a high degree of homogeneity among strains of the same species. Strains of the same enterococcal species have at least 99 % *rpoA* and 97 % *pheS* gene sequence similarity, whereas, different enterococcal species have at maximum 97 % *rpoA* and 86 % *pheS* gene sequence similarity. It was concluded that both genes can be used as reliable tools for identification of clinical and environmental species of *Enterococcus* and are efficient screening methods for the detection of novel species. The sequence data obtained in this study were compared to the available *atpA* and 16S rRNA gene sequences. The MLSA approach to *Enterococcus* taxonomy provides portable, highly reproducible data with lower costs for rapid identification of all enterococcal species.

INTRODUCTION

The taxonomy of the genus *Enterococcus* has been exposed to considerable changes in recent years as a consequence of a progressive increase in the number of novel species. Members of the genus *Enterococcus* belong to the lactic acid bacteria (LAB). Enterococci are natural inhabitants of the gastrointestinal tract of man and animals (Franz *et al.*, 1999). *E. faecalis*, *E. faecium*, *E. hirae* and *E. durans* have been found in association with human faeces (Franz *et al.*, 1999; Gelsomino *et al.*, 2002), *E. faecalis* often being the dominant species (Godfree *et al.*, 1997; Murray 1990). *Enterococcus* strains are also found in the oral cavity and the urogenital tracts of man (Morrison *et al.*, 1997; Sedgley *et al.*, 2004). However, enterococci are ubiquitous, occurring in traditional fermented food and dairy products, water surfaces, plants and birds (Klein, 2003; Niemi *et al.*, 1993; Svec *et al.*, 2001; Vancanneyt *et al.*, 2002).

The increased association of enterococci with human disease has raised concern regarding their use as probiotics (Franz *et al.*, 2003). *E. faecalis* and *E. faecium* are among the leading causes of nosocomial infections and may cause endocarditis, urinary tract infections and bacteraemia (Ratanasuwan *et al.*, 1999; Saxena *et al.*, 2003; Fernandez-Guerrero *et al.*, 2002). *E. faecalis* predominates among enterococci isolated from the environment and from human infections (more than 80 %), while *E. faecium* is associated with the majority of the remaining infections (Jett *et al.*, 1994). The recent increase in vancomycin-resistant *E. faecium* (VREF) strains among clinical isolates is a cause of serious concern and has gained

Abbreviation: MLSA, multilocus sequence analysis.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AJ843476–AJ843515, AJ843517–AJ843523, AJ843525–AJ843534, AJ843537–AJ843553, AJ843555–AJ843556 and AJ843558–AJ843560 (*rpoA* partial gene sequences), and AJ843373–AJ843387, AJ843389–AJ843402, AJ843404–AJ843410, AJ843412–AJ843435, AJ843438–AJ843444 and AJ843446–AJ843474 (*pheS* partial gene sequences).

A presentation of the polymorphic sites present in the RpoA and PheS dataset is available as supplementary data with the online version of this paper.

clinical significance in the last decade (Michel *et al.*, 1997; Dzidic & Bedekovic, 2003; Homan *et al.*, 2002; Rybak & Coyle, 1999). Although most human enterococcal infections are caused by *E. faecalis* and *E. faecium*, various studies have revealed an increase in infections caused by *E. durans*, *E. hirae*, *E. gallinarum* and *E. casseliflavus* (Baele *et al.*, 2000; Kirschner *et al.*, 2001; Knijff *et al.*, 2001; Willey *et al.*, 1999). There is, therefore, a need for rapid and accurate identification of enterococci at species level, as a means of effective infection control.

16S rRNA gene sequencing, DNA–DNA hybridization and SDS-PAGE of whole-cell proteins are among the most common techniques currently used for *Enterococcus* species identification (Angeletti *et al.*, 2001; Domig *et al.*, 2003; Vancanneyt *et al.*, 2001). However, 16S rRNA gene sequences have limited discriminating power for several closely related enterococcal species, e.g. the *E. faecium* species group (Devriese *et al.*, 2002; Poyart *et al.*, 2000; Vancanneyt *et al.*, 2001). SDS-PAGE may present problems concerning reproducibility and data portability, and DNA–DNA hybridization presents several inconveniences, i.e. few laboratories can execute this technique, the method is the slowest and most problematic step in species description and DNA–DNA data are not cumulative (Stackebrandt, 2003). *rpoB* gene sequence analysis (Drancourt *et al.*, 2004) and multiplex *sodA* PCR (Jackson *et al.*, 2004) have been used for identification of several *Enterococcus* species. So far, these molecular techniques are not yet used for routine identification. The use of protein-coding gene sequence data for the determination of genomic relatedness is emerging as an alternative to overcome these problems (Stackebrandt *et al.*, 2002; Zeigler, 2003).

In the present paper, a new approach is applied to discriminate between different species of *Enterococcus*, multilocus sequence analysis (MLSA). MLSA compares the primary DNA sequences from multiple conserved protein-coding loci for assessing the diversity and relationship of different isolates across related taxa, thereby using an appropriate phylogenetic or cladistic approach. Two studies of complete genomes provided the groundwork for establishing sets of genes useful for MLSA in large numbers of bacterial lineages (Zeigler, 2003; Santos & Ochman, 2004). Recently, *atpA*, the gene that encodes the ATP synthase α-subunit has been used as an identification tool for all enterococcal species (Naser *et al.*, 2005). In the present study, we have investigated the usefulness of the genes that encode the α-subunit of bacterial RNA polymerase (*rpoA*) and phenylalanyl-tRNA synthase α-subunit (*pheS*) as alternative identification tools for all enterococci species. We also compared the sequence data of *rpoA* and *pheS* genes with the available *atpA* and 16S rRNA gene sequences.

METHODS

Enterococcal strains and growth conditions. Ninety-six strains were analysed in this study (Table 1). To investigate the intraspecies *pheS* and *rpoA* gene sequence similarity, several representative strains of each *Enterococcus* species were included. These strains were selected on the basis of our own AFLP and SDS-PAGE data from whole-cell protein databases and, if available, other polyphasic data, and represent the known heterogeneity of *Enterococcus* species. Strains were grown on blood agar medium (Columbia Agar base) under microaerophilic conditions using CO_2-Gen (Oxoid) at 37 °C for 48 h. All strains included in this study are deposited in the BCCM/LMG Bacteria Collection at the Ghent University, Belgium. Bacterial genomic DNA was extracted following the methodology described by Gevers *et al.* (2001).

Primers. The sequences of the primers used for amplification and sequencing of *pheS* and *rpoA* genes are listed in Table 2. The primer combinations rpoA-21-F/rpoA-23-R and pheS-21-F/pheS-22-R amplified the target genes of all strains, except for *E. casseliflavus* and *E. flavescens*, in which an alternative primer combination, pheS-21-F/pheS-23-R, was used. *E. casseliflavus* and *E. flavescens* might have nucleotide sequences that are highly divergent from the degenerated primer pheS-22-R. These primers were designed using 12 *rpoA* and *pheS* gene sequences of lactic acid bacteria, i.e. *E. faecalis* (V583), *Lactobacillus plantarum* (WCFS1), *Lactococcus lactis* subsp. *lactis* (IL1403), *Streptococcus pneumoniae* (TIGR4 and R6), *Streptococcus agalactiae* (NEM316 and 2603 V/R), *Streptococcus pyogenes* (MGAS8232, SSI-1, MGAS315 and SF370) and *Streptococcus mutans* (UA159), all of which were originated from publicly available data of whole-genome sequence projects. The sequences used have the following accession numbers: AE016830, AL935263, AE005176, AE005672, AE007317, AL732656, AE009948, AE009949, BA0 00034, AE014074, AE004092 and AE014133.

PCR and sequencing. PCR reactions were composed of 33·5 µl sterile MilliQ water, 5·0 µl 10× PCR buffer, 5·0 µl dNTPs (2 mM each), 0·5 µl forward primer (50 µM), 0·5 µl reverse primer (50 µM), 0·5 µl AmpliTaq DNA Polymerase (1 U μl^{-1}) and 5·0 µl template DNA (0·01 µg μl^{-1}). PCR was performed using a GeneAmp PCR System 9600 thermocycler (Applied Biosystems). The thermal programme consisted of (1) 5 min at 95 °C, (2) 3 cycles of 1 min at 95 °C + 2 min 15 s at 46 °C + 1 min 15 s at 72 °C, (3) 30 cycles of 35 s at 95 °C + 1 min 15 s at 46 °C + 1 min 15 s at 72 °C and (4) a final 7 min at 72 °C. In a few cases, an annealing temperature of 42 °C was used for the amplification of *rpoA*. PCR products were checked by RESult 1 % LE Agarose (Biozym) gel electrophoresis. The products of positive PCRs, giving a product with the expected size, were purified using the Nucleofast 96 PCR clean-up membrane system (Macherey–Nagel). Subsequently, 3·0 µl purified and concentration-normalized PCR product was mixed with 1·0 µl ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Mix version 3.1 (Applied Biosystems), 3·0 µl sequencing primer (4 µM), 1·5 µl 5× dilution buffer and 1·5 µl MilliQ water. The primers listed in Table 2 were used for sequencing, with the exception of primer pheS-22-R. The thermal program consisted of 30 cycles of 15 s at 96 °C + 1 s at 35 °C + 4 min at 60 °C. Sequencing products were purified using a Montage SEQ_{96} sequencing reaction clean-up kit (Millipore). Purified sequencing reactions were recovered in 20 µl injection solution and mixed with 20 µl deionized formamide. Sample preparation was assisted by a Tecan Genesis Workstation 200. Subsequently, separation of the DNA fragments was obtained in an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). The time and voltage of sample injection were 20 s at 1·25 kV. Each run was performed at 50 °C for 6500 s at 0·1 mA and 12·2 kV.

Sequence analysis. Raw sequence data were transferred to Factura 1.2.Or6 and AutoAssembler software 1.4.0 (Applied Biosystems) or GeneBuilder (Applied Maths) where consensus sequences were determined using two reads for each *rpoA* and *pheS* gene, respectively. Consensus sequences were imported into BioNumerics 3.5 software (Applied Maths), where a similarity matrix and phylogenetic trees

Table 1. Strain list

Species name	Strain no.	Place of isolation	Year of isolation	Source
E. asini	LMG 18727^T			Donkey (*Equus asinus*), caecum
E. avium	LMG 16309	Sweden	1995	Human, sacral wound
	LMG 10744^T			Human, faeces
	LMG 12900			Human
	LMG 15118			
E. canis	LMG 12316^T	Belgium	1991	Dog, anus
	LMG 21545	Belgium	2000	Dog, anal swabs
	LMG 21555	Belgium	2000	Dog, anal swabs
E. casseliflavus	LMG 10745^T			Plant material
	LMG 16286	Belgium	1990–1991	Horse
	LMG 14406	Belgium		Horse manure
E. cecorum	LMG 12902^T	Belgium		Chicken, caecum
	LMG 17333	Switzerland	1995	44-year-old woman
E. columbae	LMG 11740^T	Belgium		Pigeon
	LMG 12296	Belgium		Pigeon, intestine
	LMG 11747	Belgium		Pigeon
E. dispar	LMG 13521^T	USA		Human, synovial fluid
E. durans	LMG 10746^T			Dried milk
	LMG 16193			
	LMG 20231	Ireland	1999	Farm, human faeces
	LMG 16888			Human, blood
	LMG 16892			Human with septic infection, synovia
E. faecalis	LMG 7937^T			
	LMG 17122	Belgium		Pig
	LMG 16301	Belgium	1990–1991	Horse
	LMG 20925	Italy	1971	Raw milk
	LMG 8222			Urine
	LMG 11207			Young chicken, intestine
	LMG 11396			Abscess
	LMG 13566			
E. faecium	LMG 11423^T			
	LMG 8147	Sweden	1968	
	LMG 16198	Sweden	1995	Human newborn, umbilical catheter
	LMG 16270	Belgium	1995	Dog
	LMG 20730	The Netherlands	1995	Human, wound
	LMG 20721	Belgium	1997	Human, faeces
	LMG 20789	Ireland	1994	Cheddar cheese
	LMG 20769	Greece	1995	Feta cheese
	LMG 20760	Greece	1995	Feta cheese
	LMG 20943	Ireland	1997	Human, faeces
	LMG 20772	Greece	1995	Feta cheese
	LMG 20635	Germany	1990	Swiss cheese salad
	LMG 20909	Italy	1997	Scamorza cheese
	LMG 20630	Burkina Faso	1995	Dawadawa
	LMG 20908	Italy	1997	Scamorza cheese
	LMG 20768	Greece	1995	Feta cheese
E. flavescens	LMG 13518^T	Italy	1985	Human suffering from septicaemia
	LMG 16313	Italy		Patient suffering from lung cancer
	LMG 16314	Italy		Patient suffering from osteomyelitis
E. gallinarum	LMG 13129^T			Chicken, intestine
	LMG 12904			
	LMG 16204	Sweden	1995	Human, blood
	LMG 12290	Belgium		Chicken, intestine
	LMG 16197	Belgium		Pheasant

Table 1. cont.

Species name	Strain no.	Place of isolation	Year of isolation	Source
E. gilvus	LMG 21841^T	Canada	1998	Patient with cholecystitis, bile
	LMG 13601	Belgium		Tonsil
	LMG 13600	Belgium		Tonsil
E. haemoperoxidus	LMG 19487^T	Czech Republic		Service water
E. hermanniensis	LMG 12317^T	Belgium	1991	Dog, tonsil
	LMG 21989	Finland		Modified packaged broiler legs
E. hirae	LMG 6399^T			
	LMG 20637	South Africa	1996	Ostrich colon
	LMG 11492			Chicken, crop
	LMG 14260	Belgium		Bovine, mastitis
	LMG 17110	Belgium		Ostrich
E. italicus	LMG 22039^T	Italy	2001	Toma cheese
E. malodoratus	LMG 10747^T			Gouda cheese
	LMG 15718	Belgium		Pig
	LMG 13601			
	LMG 12302	Belgium	1991	Cat, tonsil
E. moraviensis	LMG 19486^T	Czech Republic	1996	Spring Hájek
E. mundtii	LMG 10748^T			Soil
	LMG 20649	Italy	1997	Tapes decussatus (shellfish)
	LMG 20698	Italy	1995	Grass silage
E. pallens	LMG 21842^T	Canada	2000	Patient with peritonitis
E. phoeniculicola	LMG 22471^T			
E. pseudoavium	LMG 11426^T	France		Cow, udder
	LMG 11732			
	LMG 20707	The Netherlands	1996	Milk (cattle)
E. raffinosus	LMG 12888^T			Blood
	LMG 12172t2			
E. ratti	LMG 21828^T	USA	1992	Neonatal rat with diarrhoea
	LMG 21829	USA		Rat
E. saccharolyticus	LMG11427^T			Straw bedding
E. saccharominimus	LMG 21727^T	Belgium	2002	Contaminant pasteurized cow's milk
	LMG 22196	Romania	2002	Raw cow's milk
	LMG 22197	Morocco	1999	White soft cheese
E. solitarius (= *Tetragenococcus solitarius*)	LMG 12890^T			Ear exudate
E. sulfureus	LMG 13084^T			Plant material
E. villorum	LMG 12287^T	Canada		Piglet, intestine
	LMG 17496	USA		Pig, intestine
	LMG 19177	USA	1981	2-day-old piglet
	LMG 19179	Korea		Pig, intestine
Enterococcus CDC PNS E-1	LMG 22681^T	USA	1991	Human
Enterococcus CDC PNS E-2	LMG 22682^T	USA	1996	Human
Enterococcus CDC PNS E-3	LMG 22683^T	USA	1996	Human brain tissue

were created based on the maximum-parsimony and neighbour-joining methods (Saitou & Nei, 1987). The reliability of the groups was evaluated by bootstrap with 500 resamplings. 16S rRNA gene sequence data were obtained from EMBL. Splits decomposition tree analysis was done using software available on the web (http://bibiserv.techfak.uni-bielefeld.de/splits/) (Huson, 1998), while the GC content, the ratio of mean synonymous substitutions per synonymous site/mean non-synonymous substitutions per non-synonymous site (d_S/d_N) and Sawyer's test were calculated using the software package START obtained from http://pubmlst.org/software/analysis/start/ (Jolley *et al.*, 2001).

Table 2. Amplification and sequencing primers used in this study

Primer name	Sequence (5′→3′)	Position
pheS-21-F	CAYCCNGCHCGYGAYATGC	557
pheS-22-R	CCWARVCCRAARGCAAARCC	1031
pheS-23-R	GGRTGRACCATVCCNGCHCC	968
rpoA-21-F	ATGATYGARTTTGAAAAACC	1
rpoA-23-R	ACHGTRTTRATDCCDGCRCG	802

RESULTS AND DISCUSSION

The *rpoA* (533 nt) and *pheS* (455 nt) partial gene sequences had a GC content of $39{\cdot}3 \pm 1{\cdot}3$ mol% and $41{\cdot}5 \pm 2{\cdot}0$ mol%, respectively. The *rpoA* and *pheS* d_S/d_N ratios were 21 and 31, respectively, suggesting that these loci are under neutral selective pressure. Both Sawyer's test and Splits decomposition analysis did not reveal evidence of recombination for the whole set of 96 *Enterococcus* strains. A presentation of the polymorphic sites present in the RpoA and PheS dataset is available as supplementary data with the online version of this paper. In addition, *Enterococcus* species formed a monophyletic group within the lactic acid bacteria by both *rpoA* and *pheS* gene sequence trees. Overall, these analyses did not show evidence for horizontal gene transfer involving *rpoA* and *pheS* sequences of *Enterococcus* strains.

Interspecies and intraspecies heterogeneity of *rpoA* gene sequences

On the basis of *rpoA* gene sequences, all *Enterococcus* species were clearly differentiated, forming distinct branches (Fig. 1). At the interspecies level, the *rpoA* gene sequence similarity was at maximum 97 % for all species. Strains of the same species had at least 99 % *rpoA* gene sequence similarity. Members of the *E. avium*, *E. casseliflavus*, *E. faecalis*, *E. cecorum* and *E. faecium* (except for *E. canis*) species groups clustered together as in the 16S rRNA based phylogeny, although the closest neighbours were not the same as in the 16S rRNA analysis. Within the *E. faecium* species group, all species occupied distinct positions with at maximum 97 % *rpoA* gene sequence similarity. The closest neighbours of *E. faecium* were *E. villorum* (97 % *rpoA* gene sequence similarity), *E. durans* and *E. hirae* (96 %), *E. mundtii* (95 %) and *E. ratti* (94 %). The *rpoA* gene sequence tree revealed two subclusters within the *E. faecalis* species group, i.e. *E. faecalis* and *E. moraviensis*/*E. haemoperoxidus*. *E. faecalis* was more distantly related to *E. moraviensis* (90 %) and *E. haemoperoxidus* (89 %). Within the *E. casseliflavus* species group, *E. casseliflavus* LMG 10745^T and *E. flavescens* LMG 13518^T were highly related to each other having >99 % *rpoA* gene sequence similarity. Both species shared at maximum 94 % similarity with *E. gallinarum*, the other member of this species group. Similarly, *E. saccharominimus* LMG 21727^T (Vancanneyt *et al.*, 2004), *E. italicus* LMG 22039^T (Fortina *et al.*, 2004) and *Enterococcus* CDC PNS-E1 (= LMG 22681^T) (Carvalho *et al.*, 2004) were highly related, having about 100 % *rpoA* gene sequence similarity. All species within the *E. avium* species group occupy distinct positions. The closest neighbours of *E. avium* were *E. malodoratus*, *E. gilvus*, *E. raffinosus* (97 %), *E. pseudoavium* (96 %), *E. hermanniensis* (95 %) and *E. pallens* (90 %). The *E. cecorum* species group consists of *E. cecorum* and *E. columbae*. The type strains of both species had at maximum 87 % *rpoA* gene sequence similarity.

Interspecies and intraspecies heterogeneity of *pheS* gene sequences

At the interspecies level, *pheS* showed a high degree of resolution for differentiating the enterococcal species (Fig. 2). The *pheS* gene sequence similarity was at maximum 86 % for all species. Conspecific strains had at least 97 % *pheS* sequence similarity. The topology obtained by *pheS* gene sequences is not the same as the one based on the 16S rRNA gene sequences. The members of different species groups of *Enterococcus* clustered on the basis of 16S rRNA gene sequences were obviously split in the *pheS* tree (Fig. 2). With the exception of the highly related *E. casseliflavus* and *E. flavescens* (98·5 % *pheS* gene sequence similarity) as well as *E. saccharominimus* LMG 21727^T, *Enterococcus* CDC PNS-E1 (= LMG 22681^T) and *E. italicus* LMG 22039^T (99·5 %), the type strains of *E. faecium*, *E. durans*, *E. mundtii*, *E. gallinarum*, *E. haemoperoxidus*, *E. moraviensis*, *E. faecalis* and *E. phoeniculicola* were obviously differentiated as distinct branches. *E. hirae*, *E. villorum* and *E. ratti* clustered together (Fig. 2). *E. hirae* was more distantly related to *E. villorum* (85 % *pheS* gene sequence similarity), *E. ratti* and *Enterococcus* CDC PNS-E2 (= LMG 22682^T) (Carvalho *et al.*, 2004) (82 %). Similarly, other *Enterococcus* species occupied distinct positions. The closest neighbours of *E. avium* were *E. malodoratus* (83 %), *E. pseudoavium*, *Enterococcus* CDC PNS-E3 (= LMG 22683^T) (Carvalho *et al.*, 2004) (82 %), *E. hermanniensis* (81 %), *E. gilvus* (80 %), *E. raffinosus* (79 %) and *E. pallens* (78 %). Therefore, as novel species, *Enterococcus* CDC PNS-E2 and *Enterococcus* CDC PNS-E3, recently described by Carvalho *et al.* (2004), showed distinct branches based on *pheS* gene sequence analysis. This indicated that *pheS* gene is an efficient screening tool for detection of novel enterococcal species.

Comparisons of *rpoA*, *pheS*, *atpA* and 16S rRNA gene sequence data

Our data clearly show that *rpoA* and *pheS* gene sequences are much more discriminatory than 16S rRNA. 16S rRNA gene similarities above 97 % corresponded to *rpoA* and *pheS* pairwise similarities above 76 and 74 %, respectively (data not shown). Comparisons of *rpoA* and *pheS* 16S rRNA indicated that the *rpoA* gene correlated well with the 16S rRNA gene sequences. It is worth mentioning that the gene sequence trees of both *rpoA* and *pheS* genes (Figs 1 and 2) revealed a high degree of relatedness between *E. casseliflavus* (Collins *et al.*, 1984) and *E. flavescens* (Pompei *et al.*, 1992). The corresponding 16S rRNA (Patel *et al.*, 1998), *atpA* (Naser *et al.*, 2005), $sodA_{int}$ (Poyart *et al.*, 2000), *ddl* (Navarro *et al.*, 1994) and *vanC* gene sequences (Dutka-Malen *et al.*, 1995) of *E. casseliflavus* and *E. flavescens* type strains were also similar (98·8, 98·9, 98, 99·5 and 96 % sequence similarity, respectively). Therefore, MLSA of *rpoA*, *pheS*, *atpA* and 16S rRNA genes confirm that *E. flavescens* is not a separate species, but in fact, should be classified as *E. casseliflavus*, in complete agreement with previous studies (Descheemaeker *et al.*, 1997; Teixeira *et al.*,

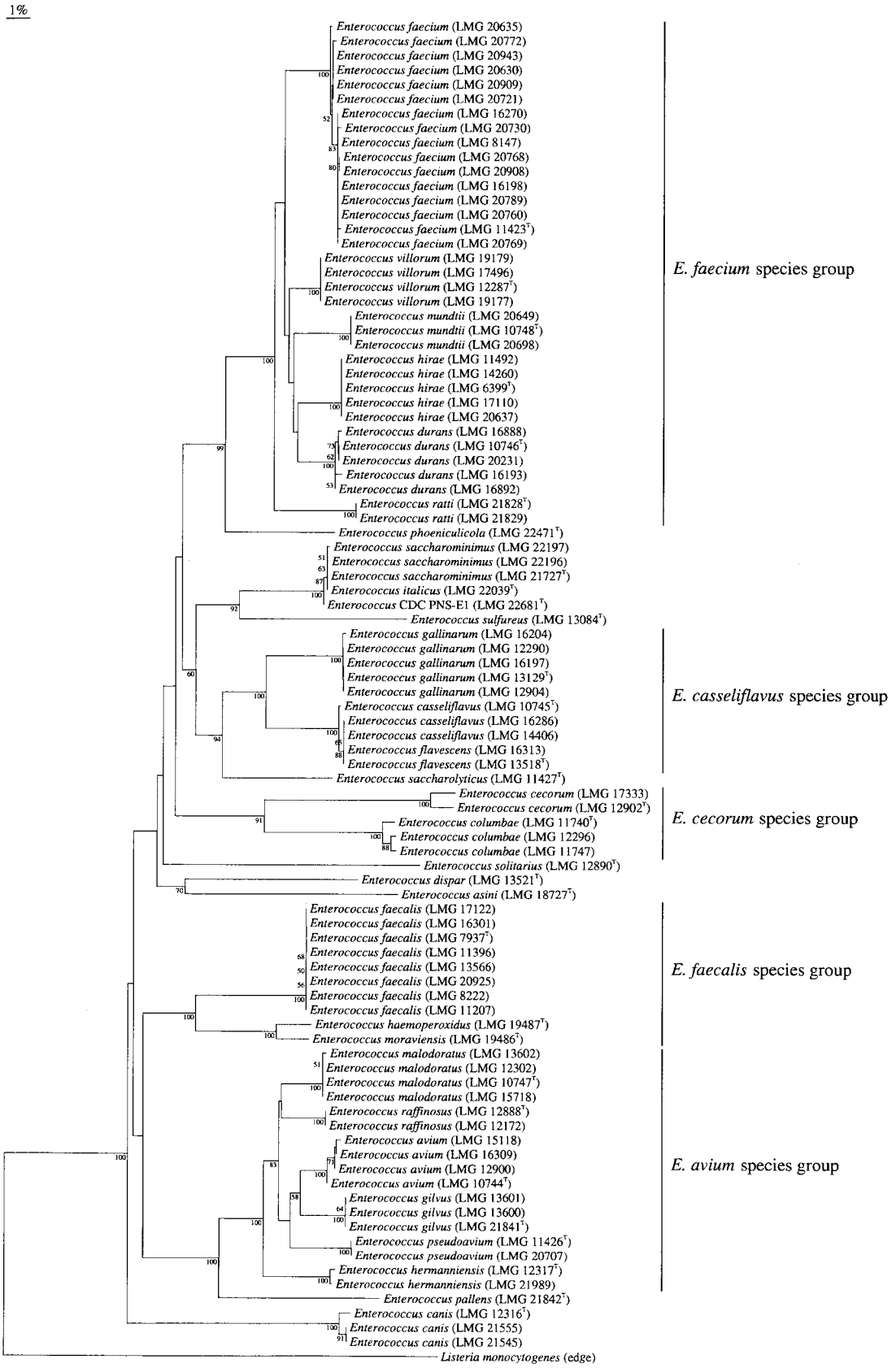

Fig. 1. Neighbour-joining tree based on the *rpoA* gene sequences of 92 enterococcal strains. Bootstrap percentages (⩾50) after 500 simulations are shown. *Listeria monocytogenes* was included as an outgroup.

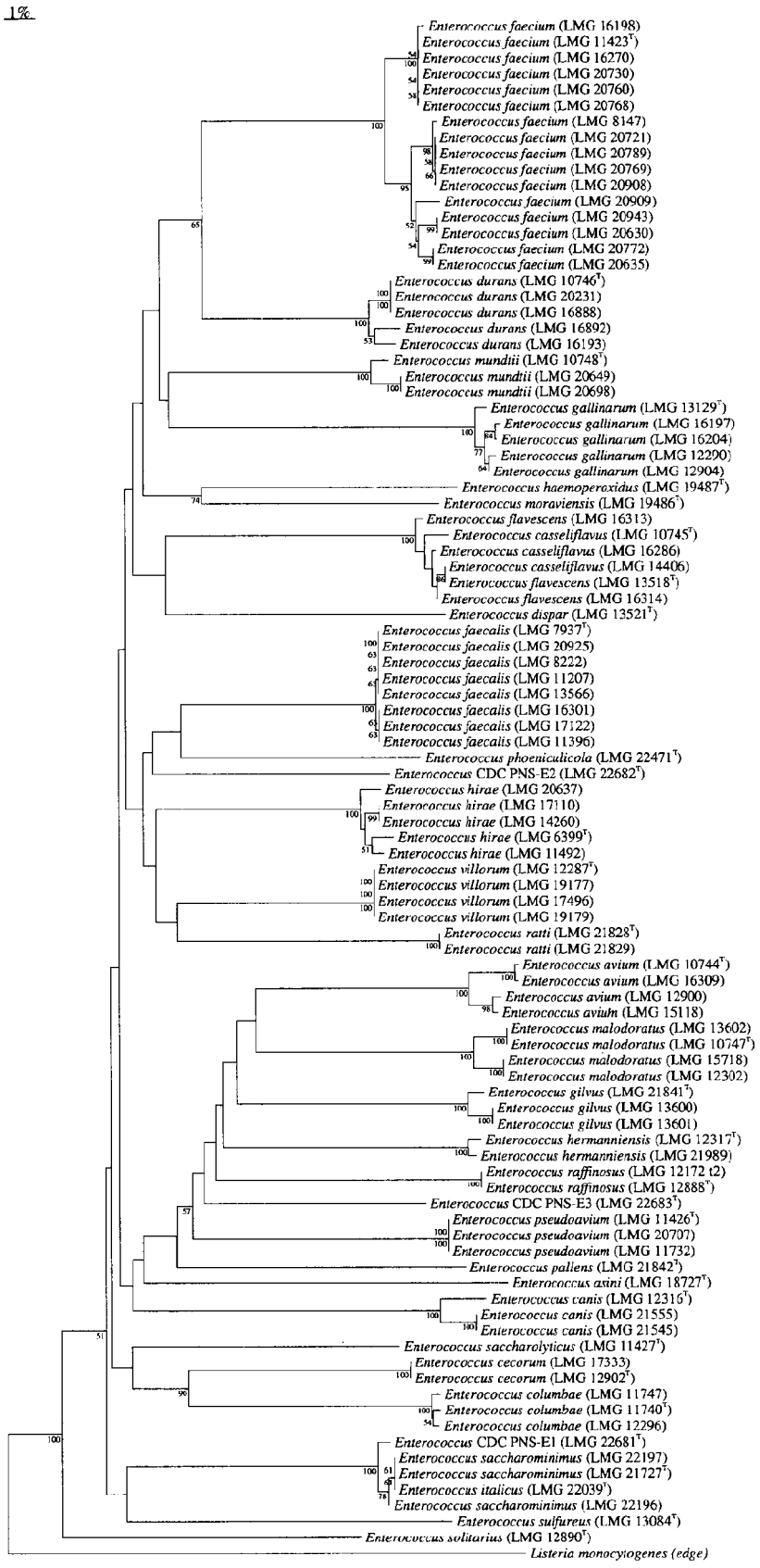

Fig. 2. Neighbour-joining tree based on the *pheS* gene sequences of 96 enterococcal strains. Bootstrap percentages (⩾50) after 500 simulations are shown. *Listeria monocytogenes* was included as an outgroup.

1997). *E. saccharominimus* LMG 21727^T (Vancanneyt *et al.*, 2004), *E. italicus* LMG 22039^T (Fortina *et al.*, 2004) and *Enterococcus* CDC PNS-E1 (=LMG 22681^T) (Carvalho *et al.*, 2004) were highly related, sharing about 100 % *rpoA*, *pheS* and *atpA* gene sequence similarity, confirming 16S rRNA data. Consequently, this indicates that *E. saccharominimus* LMG 21727^T (Vancanneyt *et al.*, 2004), *Enterococcus* CDC PNS-E1 (=LMG 22681^T) (Carvalho *et al.*, 2004) and *E. italicus* LMG 22039^T (Fortina *et al.*, 2004) are not separate species, but should be classified as *E. italicus*. Both *rpoA* and *atpA* gene sequences clustered *E. canis*, a member of the *E. faecium* species group, as a separate branch, in contrast to the phylogenetic tree of 16S rRNA. *pheS* gene sequence analysis showed a higher interspecies resolution to differentiate closely related species. For instance, *E. haemoperoxidus* LMG 19487^T and *E. moraviensis* LMG 19486^T have 99·4 % 16S rRNA gene sequence similarity, but only 82 % *pheS* gene sequence similarity. At the interspecies level, all enterococcal species were clearly differentiated on the basis of *rpoA*, *pheS* and *atpA* gene sequences, with at maximum 97, 86 and 92 % similarity, respectively.

The aim of this study was to focus particularly on tools for rapid, reliable and inexpensive identification for discrimination among different species of enterococci, and not for typing or phylogeny purposes, and to that end our data convincingly prove that the partial sequences of *pheS* and *rpoA* gene sequences perfectly fulfil this aim and show high resolution for differentiating all enterococcal species, even better than 16S rRNA gene sequences. At species level, the bootstrap values for *rpoA* and *pheS* gene sequences are always 100 %, proving that both *rpoA* and *pheS* genes are reliable genomic markers for species differentiation within the genus *Enterococcus* (Figs 1 and 2). Our choice for using partial sequences neither gave enough resolution for typing at the intraspecies level nor provided sufficient evidence at deeper phylogenetic branches. To fulfil these approaches, the inclusion of more loci and full sequences would be essential, but this is beyond the scope of this paper. The full gene sequences can be obtained by gene cloning, but the amount of work and costs will be increased and this is not favourable in comparison with MLSA, i.e. rapid, reliable and inexpensive identification. The 16S rRNA gene is very useful for discriminating the main groups of enterococci, i.e. *E. avium*, *E. casseliflavus*, *E. cecorum*, *E. faecalis* and *E. faecium* species groups, but it fails to discriminate closely related species. One example is the members of the *E. faecium* species group, i.e. *E. faecium*, *E. hirae*, *E. durans*, *E. villorum*, *E. mundtii* and *E. ratti*. The 16S rRNA genes of these species show similarities of 98·8–99·7 % (Devriese *et al.*, 2002), but the highest gene sequence similarities observed for *rpoA*, *pheS* and *atpA* were 97, 86 and 89·9 %, respectively. This also demonstrates the advantage of using several housekeeping genes for species identification studies (Stackebrandt *et al.*, 2002). Consequently, all currently known *Enterococcus* species were clearly differentiated on the basis of *rpoA* and *pheS* gene sequences (Figs 1 and 2).

Both genes provide efficient screening methods for the detection of novel species. At the interspecies level, the simultaneous analysis of *rpoA* and *pheS* gene sequences offers an alternative to DNA–DNA hybridization to differentiate closely related *Enterococcus* species. To evaluate the intraspecies *rpoA* and *pheS* gene sequence similarities, multiple strains of each species and, in particular, 16 well characterized strains of *E. faecium* were included. The results showed that *rpoA* and *pheS* genes had a high degree of homogeneity among strains of the same species. Consequently, this indicated the low discriminatory power of these genes for intraspecies differentiation. Therefore, we could conclude that strains of the same enterococcal species will have at least 99 % *rpoA* and 97 % *pheS* gene sequence similarity, respectively. In comparison, strains of a single *Enterococcus* species have at least 96·3 % *atpA* gene sequence similarity.

Rapid and robust classification using MLSA may be used with a universal set of protein-coding genes that are widely distributed among bacterial genomes and present in single-copy, that show levels of variation below saturation for the group being analysed and that are not unusually prone to recombination (Zeigler, 2003). The fact that different chronometers provide different closest neighbours of a given strain does not hamper their use to unambiguously circumscribe bacterial species. Several factors account for the different topologies determined for different housekeeping genes, i.e. the level of the information content, the different rates of evolution due to different selection forces on various genes and the length of partial sequences compared (Christensen *et al.*, 2004). The use of several housekeeping genes in bacterial taxonomy is best suited for analysis at the species and genus levels as it integrates the information of different molecular clocks around the bacterial chromosome (Lerat *et al.*, 2003; Palys *et al.*, 1997; Stackebrandt *et al.*, 2002; Ventura *et al.*, 2004; Zeigler, 2003). This type of data may aid the development of better species definition for *Enterococcus*.

ACKNOWLEDGEMENTS

S. M. N. acknowledges a PhD scholarship from the Ministry of Education and Higher Education. F. L. T. acknowledges a postdoctoral fellowship from BCCM/LMG Bacteria Collection. J. S. acknowledges grants from of the Fund for Scientific Research (FWO), Belgium. We thank K. Vandemeulebroecke and L. Christiaens for their excellent technical assistance.

REFERENCES

Angeletti, S., Lorino, G., Gherardi, G., Battistoni, F., De Cesaris, M. & Dicuonzo, G. (2001). Routine molecular identification of enterococci by gene-specific PCR and 16S ribosomal DNA sequencing. *J Clin Microbiol* **39**, 794–797.

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

Carvalho, G., Steigerwalt, G., Morey, E., Shewmaker, L., Teixeira, M. & Facklam, R. (2004). Characterization of three new enterococcal species, *Enterococcus* sp. nov. CDC PNS-E1, *Enterococcus* sp. nov. CDC PNS-E2, and *Enterococcus* sp. nov. CDC PNS-E3, isolated from human clinical specimens. *J Clin Microbiol* **42**, 1192–1198.

Christensen, H., Kuhnert, P., Olsen, J. E. & Bisgaard, M. (2004). Comparative phylogenies of the housekeeping genes *atpD*, *infB* and *rpoB* and the 16S rRNA gene within the *Pasteurellaceae*. *Int J Syst Evol Microbiol* **54**, 1601–1609.

Collins, M., Jones, D., Farrow, J., Kilpper-Balz, R. & Schleifer, K. (1984). *Enterococcus avium* nom. rev., comb. nov.; *E. casseliflavus* nom. rev., comb. nov.; *E. durans* nom. rev., comb. nov.; *E. gallinarum* comb. nov.; and *E. malodoratus* sp. nov. *Int J Syst Bacteriol* **34**, 220–223.

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.

Devriese, L. A., Vancanneyt, M., Descheemaeker, P., Baele, M., Van Landuyt, H. W., Gordts, B., Butaye, P., Swings, J. & Haesebrouck, F. (2002). Differentiation and identification of *Enterococcus durans*, *E. hirae* and *E. villorum*. *J Appl Microbiol* **92**, 821–827.

Domig, K. J., Mayer, H. K. & Kneifel, W. (2003). Methods used for the isolation, enumeration, characterisation and identification of *Enterococcus* spp. 2. Pheno- and genotypic criteria. *Int J Food Microbiol* **88**, 165–188.

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.

Dutka-Malen, S., Evers, S. & Courvalin, P. (1995). Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR. *J Clin Microbiol* **33**, 24–27; erratum 1434.

Dzidic, S. & Bedekovic, V. (2003). Horizontal gene transfer – emerging multidrug resistance in hospital bacteria. *Acta Pharmacol Sin* **24**, 519–526.

Fernandez-Guerrero, M. L., Herrero, L., Bellver, M., Gadea, I., Roblas, R. F. & de Gorgolas, M. (2002). Nosocomial enterococcal endocarditis: a serious hazard for hospitalized patients with enterococcal bacteraemia. *J Intern Med* **252**, 510–515.

Fortina, M. G., Ricci, G., Mora, D. & Manachini, P. L. (2004). Molecular analysis of artisanal Italian cheeses reveals *Enterococcus italicus* sp. nov. *Int J Syst Evol Microbiol* **54**, 1717–1721.

Franz, C. M., Holzapfel, W. H. & Stiles, M. E. (1999). Enterococci at the crossroads of food safety? *Int J Food Microbiol* **47**, 1–24.

Franz, C. M., Stiles, M. E., Schleifer, K. H. & Holzapfel, W. H. (2003). Enterococci in foods – a conundrum for food safety. *Int J Food Microbiol* **88**, 105–122.

Gelsomino, R., Vancanneyt, M., Cogan, T. M., Condon, S. & Swings, J. (2002). Source of enterococci in a farmhouse raw-milk cheese. *Appl Environ Microbiol* **68**, 3560–3565.

Gevers, D., Huys, G. & Swings, J. (2001). Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. *FEMS Microbiol Lett* **205**, 31–36.

Godfree, A. F., Kay, D. & Wyer, M. D. (1997). Faecal streptococci as indicators of faecal contamination in water. *Soc Appl Bacteriol Symp Ser* **26**, 110S–119S.

Homan, W. L., Tribe, D., Poznanski, S., Li, M., Hogg, G., Spalburg, E., Van Embden, J. D. & Willems, R. J. (2002). Multilocus sequence typing scheme for *Enterococcus faecium*. *J Clin Microbiol* **40**, 1963–1971.

Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.

Jackson, C. R., Fedorka-Cray, P. J. & Barrett, J. B. (2004). Use of a genus- and species-specific multiplex PCR for identification of enterococci. *J Clin Microbiol* **42**, 3558–3565.

Jett, B. D., Huycke, M. M. & Gilmore, M. S. (1994). Virulence of enterococci. *Clin Microbiol Rev* **7**, 462–478.

Jolley, K. A., Feil, E. J., Chan, M. S. & Maiden, M. C. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**, 1230–1231.

Kirschner, C., Maquelin, K., Pina, P. & 11 other authors (2001). Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. *J Clin Microbiol* **39**, 1763–1770.

Klein, G. (2003). Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. *Int J Food Microbiol* **88**, 123–131.

Knijff, E., Dellaglio, F., Lombardi, A., Andrighetto, C. & Torriani, S. (2001). Rapid identification of *Enterococcus durans* and *Enterococcus hirae* by PCR with primers targeted to the *ddl* genes. *J Microbiol Methods* **47**, 35–40.

Lerat, E., Daubin, V. & Moran, N. A. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-proteobacteria. *PLoS Biol* **1**, 101–109.

Michel, M. & Gutmann, L. (1997). Methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococci: therapeutic realities and possibilities. *Lancet* **349**, 1901–1906.

Morrison, D., Woodford, N. & Cookson, B. (1997). Enterococci as emerging pathogens of humans. *Soc Appl Bacteriol Symp Ser* **26**, 89S–99S.

Murray, B. E. (1990). The life and times of the *Enterococcus*. *Clin Microbiol Rev* **3**, 46–65.

Naser, S., Thompson, F. L., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Thompson, C. C., Vancanneyt, M. & Swings, J. (2005). Phylogeny and identification of enterococci using *atpA* gene sequence analysis. *J Clin Microbiol* **43**, 2224–2230.

Navarro, F. & Courvalin, P. (1994). Analysis of genes encoding D-alanine-D-alanine ligase-related enzymes in *Enterococcus casseliflavus* and *Enterococcus flavescens*. *Antimicrob Agents Chemother* **38**, 1788–1793.

Niemi, R. M., Niemela, S. I., Bamford, D. H., Hantula, J., Hyvarinen, T., Forsten, T. & Raateland, A. (1993). Presumptive fecal streptococci in environmental samples characterized by one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Appl Environ Microbiol* **59**, 2190–2196.

Palys, T., Nakamura, L. K. & Cohan, F. M. (1997). Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* **47**, 1145–1156.

Patel, R., Piper, K. E., Rouse, M. S., Steckelberg, J. M., Uhl, J. R., Kohner, P., Hopkins, M. K., Cockerill, F. R., 3rd, & Kline, B. C. (1998). Determination of 16S rRNA sequences of enterococci and application to species identification of nonmotile *Enterococcus gallinarum* isolates. *J Clin Microbiol* **36**, 3399–3407.

Pompei, R., Berlutti, F., Thaller, M., Ingianni, A., Cortis, G. & Dainelli, B. (1992). *Enterococcus flavescens* sp. nov., a new species of enterococci of clinical origin. *Int J Syst Bacteriol* **42**, 365–369.

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.

Ratanasuwan, W., Iwen, P. C., Hinrichs, S. H. & Rupp, M. E. (1999). Bacteremia due to motile *Enterococcus* species: clinical features and outcomes. *Clin Infect Dis* **28**, 1175–1177.

Rybak, M. J. & Coyle, E. A. (1999). Vancomycin-resistant *Enterococcus*: infectious endocarditis treatment. *Curr Infect Dis Rep* **1**, 148–152.

Saitou, N. & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Santos, S. R. & Ochman, H. (2004). Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* **6**, 754–759.

Saxena, S., Prakash, S. K., Malik, V. K. & Mathur, M. D. (2003). Vancomycin resistant *Enterococcus* in nosocomial urinary tract infections. *Ind J Pathol Microbiol* **46**, 256–258.

Sedgley, C. M., Lennan, S. L. & Clewell, D. B. (2004). Prevalence, phenotype and genotype of oral enterococci. *Oral Microbiol Immunol* **19**, 95–101.

Stackebrandt, E. (2003). The richness of prokaryotic diversity: there must be a species somewhere. *Food Technol Biotechnol* **41**, 17–22.

Stackebrandt, E., Frederiksen, W., Garrity, G. M. & 10 other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043–1047.

Svec, P., Devriese, L. A., Sedlacek, I., Baele, M., Vancanneyt, M., Haesebrouck, F., Swings, J. & Doskar, J. (2001). *Enterococcus haemoperoxidus* sp. nov. and *Enterococcus moraviensis* sp. nov., isolated from water. *Int J Syst Evol Microbiol* **51**, 1567–1574.

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.

Vancanneyt, M., Snauwaert, C., Cleenwerck, I. & 8 other authors (2001). *Enterococcus villorum* sp. nov., an enteroadherent bacterium associated with diarrhoea in piglets. *Int J Syst Evol Microbiol* **51**, 393–400.

Vancanneyt, M., Lombardi, A., Andrighetto, C. & 11 other authors (2002). Intraspecies genomic groups in *Enterococcus faecium* and their correlation with origin and pathogenicity. *Appl Environ Microbiol* **68**, 1381–1391.

Vancanneyt, M., Zamfir, M., Devriese, L. A. & 7 other authors (2004). *Enterococcus saccharominimus* sp. nov., from dairy products. *Int J Syst Evol Microbiol* **54**, 2175–2179.

Ventura, M., Canchaya, C., van Sinderen, D., Fitzgerald, G. F. & Zink, R. (2004). *Bifidobacterium lactis* DSM 10140: identification of the *atp* (*atpBEFHAGDC*) operon and analysis of its genetic structure, characteristics, and phylogeny. *Appl Environ Microbiol* **70**, 3110–3121.

Willey, B. M., Jones, R. N., McGeer, A., Witte, W., French, G., Roberts, R. B., Jenkins, S. G., Nadler, H. & Low, D. E. (1999). Practical approach to the identification of clinically relevant *Enterococcus* species. *Diagn Microbiol Infect Dis* **34**, 165–171.

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int J Syst Evol Microbiol* **53**, 1893–1900.