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 Vancanneyt2 and Jean Swings1,2

1,2Laboratory of Microbiology1 and BCCM/LMG Bacteria Collection2, Ghent University, K. L. Ledeganckstraat 35, Ghent 9000, Belgium
3Bioinformatics and Evolutionary Genomics, Ghent University/VIB, Technologiepark 927, Ghent 9052, Belgium

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 rpoA 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 (Godfere 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 bacteremia (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.


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 phenylalanine-tRNase synthase α-subunit (*pheS*) as alternative identification tools for all enterococcal 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 CO₂-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, *rpoS*-21-F/*rpoS*-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* (V883), *Lactobacillus plantarum* (WCFS1), *Lactococcus lactis* subsp. *lactic* (IL1403), *Streptococcus pneumoniae* (TIGR4 and R6), *Streptococcus agalactiae* (NEM316 and 2603 V/R), *Streptococcus pyogenes* (MGA8823, SSI-1, MGA3315 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, AL035263, AE005176, AE005672, AE007317, AL732565, AE009948, AE009949, BA000324, 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 SEQmix 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>
<thead>
<tr>
<th>Species name</th>
<th>Strain no.</th>
<th>Place of isolation</th>
<th>Year of isolation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. asini</em></td>
<td>LMG 18727T</td>
<td>Sweden</td>
<td>1995</td>
<td>Donkey (<em>Equus asinus</em>), caecum</td>
</tr>
<tr>
<td><em>E. avium</em></td>
<td>LMG 16309</td>
<td>Sweden</td>
<td>1995</td>
<td>Human, sacral wound</td>
</tr>
<tr>
<td></td>
<td>LMG 10744T</td>
<td>Sweden</td>
<td>1995</td>
<td>Human, faeces</td>
</tr>
<tr>
<td><em>E. canis</em></td>
<td>LMG 12316T</td>
<td>Belgium</td>
<td>1991</td>
<td>Dog, anus</td>
</tr>
<tr>
<td><em>E. casseliflavus</em></td>
<td>LMG 10745T</td>
<td>Plant material</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. cecorum</em></td>
<td>LMG 12902T</td>
<td>Belgium</td>
<td>1995</td>
<td>44-year-old woman</td>
</tr>
<tr>
<td><em>E. columbae</em></td>
<td>LMG 11740T</td>
<td>Belgium</td>
<td>1995</td>
<td>Pigeon</td>
</tr>
<tr>
<td><em>E. dispar</em></td>
<td>LMG 13521T</td>
<td>USA</td>
<td>1995</td>
<td>Human, synovial fluid</td>
</tr>
<tr>
<td><em>E. durans</em></td>
<td>LMG 10746T</td>
<td>Belgium</td>
<td>1999</td>
<td>Farm, human faeces</td>
</tr>
<tr>
<td></td>
<td>LMG 16199</td>
<td>Irish</td>
<td>1999</td>
<td>Human, blood</td>
</tr>
<tr>
<td><em>E. faecalis</em></td>
<td>LMG 7937T</td>
<td>Belgium</td>
<td>1999</td>
<td>Human with septic infection, synovia</td>
</tr>
<tr>
<td><em>E. faecium</em></td>
<td>LMG 11423T</td>
<td>Sweden</td>
<td>1968</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LMG 16198</td>
<td>Sweden</td>
<td>1995</td>
<td>Human newborn, umbilical catheter</td>
</tr>
<tr>
<td></td>
<td>LMG 16270</td>
<td>Belgium</td>
<td>1995</td>
<td>Dog</td>
</tr>
<tr>
<td><em>E. flavescens</em></td>
<td>LMG 13518T</td>
<td>Italy</td>
<td>1985</td>
<td>Human suffering from septicemia</td>
</tr>
<tr>
<td><em>E. gallinarum</em></td>
<td>LMG 13129T</td>
<td>Belgium</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LMG 12904</td>
<td>Sweden</td>
<td>1995</td>
<td>Human, blood</td>
</tr>
<tr>
<td></td>
<td>LMG 16204</td>
<td>Sweden</td>
<td>1995</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LMG 12290</td>
<td>Belgium</td>
<td>1995</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LMG 16197</td>
<td>Belgium</td>
<td>1995</td>
<td></td>
</tr>
</tbody>
</table>
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 (dS/dN) and Sawyer’s test were calculated using the software package START obtained from http://pubmlst.org/software/analysis/start/ (Jolley et al., 2001).

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 (dS/dN) and Sawyer’s test were calculated using the software package START obtained from http://pubmlst.org/software/analysis/start/ (Jolley et al., 2001).

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 (dS/dN) and Sawyer’s test were calculated using the software package START obtained from http://pubmlst.org/software/analysis/start/ (Jolley et al., 2001).
RESULTS AND DISCUSSION
The rpoA (533 nt) and pheS (455 nt) partial gene sequences had a GC content of 39.3 ± 1.3 mol% and 41.5 ± 2.0 mol%, respectively. The rpoA and pheS dS/dN 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 PhcS 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 Enterococcus avium species group, i.e. E. avium, E. casseliflavus, E. faecalis, E. ecaeorum 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 10745T and E. flavescens LMG 15318T 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 21727T (Vancanneyt et al., 2004), E. italicus LMG 22039T (Fortina et al., 2004) and Enterococcus CDC PNS-E1 (=LMG 22681T) (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. herman niensis (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 RNA 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 21727T, Enterococcus CDC PNS-E1 (=LMG 22681T) and E. italicus LMG 22039T (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 distinctly related to E. villorum (85% pheS gene sequence similarity), E. ratti and Enterococcus CDC PNS-E2 (=LMG 22682T) (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 22683T) (Carvalho et al., 2004) (82%), E. herman niensis (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 sequence 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. cas seliflavus (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 (Poyart et al., 2000), dcd (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, 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.
The 16S rRNA genes of these species show similarities of 97, 86 and 89% atpA, pheS and atpA gene sequence similarities, confirming 16S rRNA data. Consequently, this indicates that E. saccharominimus LMG 21727T (Vancanneyt et al., 2004), E. italicus LMG 22039T (Fortina et al., 2004) and Enterococcus CDC PNS-E1 (=LMG 22681T) (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 21727T (Vancanneyt et al., 2004), Enterococcus CDC PNS-E1 (=LMG 22681T) (Carvalho et al., 2004) and E. italicus LMG 22039T (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 19487T and E. moraviensis LMG 19486T 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 fulfill 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 interspecies level nor provided sufficient evidence at deeper phylogenetic branches. To fulfill 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 interspecies 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 chromometers 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 the Fund for Scientific Research (FWO), Belgium. We thank K. Vandemeulebroecke and L. Christiaens for their excellent technical assistance.

REFERENCES


