Congruence of evolutionary relationships inside the *Leuconostoc*–*Oenococcus*–*Weissella* clade assessed by phylogenetic analysis of the 16S rRNA gene, *dnaA*, *gyrB*, *rpoC* and *dnaK*

Ivo M. Chelo, Líbia Zé-Zé and Rogério Tenreiro

Correspondence
Rogério Tenreiro
rptenreiro@fc.ul.pt

Universidade de Lisboa, Faculdade de Ciências, Centro de Genética e Biologia Molecular and Instituto de Ciência Aplicada e Tecnologia, Edifício ICAT, Campus da FCUL, Campo Grande, 1749-016 Lisboa, Portugal

The phylogenetic structure of the *Leuconostoc*–*Oenococcus*–*Weissella* clade was evaluated by comparison of 16S rRNA gene, *dnaA*, *gyrB*, *rpoC* and *dnaK* sequence analysis. Phylogenies obtained with the different genes were in overall good agreement and a well-supported, almost fully resolved phylogenetic tree was obtained when the combined data were analysed in a Bayesian approach. A rapid basal diversification of the three genera is suggested. Evolutionary rates of the 16S rRNA gene in these genera seem to be different and specifically related to the evolution of this group, revealing the importance of this sequence in the constitution of the present taxonomy, but preventing its straightforward use in phylogenetic inference.

INTRODUCTION

The genera *Leuconostoc*, *Oenococcus* and *Weissella* constitute a line of descent of the lactic acid bacteria (LAB) that is commonly known as the leuconostoc group. Bacteria belonging to these genera are Gram-positive, non-spore-forming, round or elongated in shape, being anaerobic or aerotolerant. They usually inhabit nutrient-rich environments such as milk, meat, vegetable products and fermented drinks (Kandler & Weiss, 1986). Lactic acid is the main end product of their characteristic heterofermentative carbohydrate metabolism. The present taxonomic classification of these genera results from a reorganization of the genus *Leuconostoc*, with the inclusion of *Leuconostoc paramesenteroides* and some atypical heterofermentative *Lactobacillus* species in the genus *Weissella* (Collins *et al.*, 1993), and the reclassification of *Leuconostoc oenos* as *Oenococcus oeni* (Dicks *et al.*, 1995). These changes were made taking in consideration previous phylogenetic studies (Martinez-Murcia & Collins, 1990; Collins *et al.*, 1991) and in accordance with earlier nucleic acid hybridization studies (Garvie, 1976, 1981). More recently, *Lactobacillus fructosus*

was also reclassified as *Leuconostoc fructosum*, as suggested in 16S rRNA gene phylogeny (Antunes *et al.*, 2002).

Although the taxonomy of these taxa seems to be well-established, their phylogenetic evaluation has been made almost exclusively from 16S rRNA gene sequences and there are still some controversial issues needing further clarification. These mainly concern the correct branching order of the three genera, as pointed out by Martinez-Murcia *et al.* (1993), and the relative rates of evolution revealed by phylogenies made with *rpoC*, encoding a β' subunit of the DNA-dependent RNA polymerase (Morse *et al.*, 1996). In this latter study, the tachytelic nature of *O. oeni*, as first hypothesized by Yang & Woese (1989) based on comparative analysis of 16S rRNA gene sequences, was questioned, since it did not seem to be reflected in *rpoC* gene-based phylogenies. Recent developments coming from the genome analysis of *O. oeni* PSU-1 (Mills *et al.*, 2005) and comparison with other LAB strains indicate that the leuconostoc group has evolved rapidly in comparison with phylogenetically close groups, with *Oenococcus* as the most divergent genus.

16S rRNA gene sequence analysis is still of paramount importance in taxonomic resolution in bacteria (Stackebrandt *et al.*, 2002), but the use of different genes as phylogenetic markers, particularly protein-coding genes, is now a common approach (Yamamoto & Harayama, 1998; Morse *et al.*, 2002; Adékambi & Drancourt, 2004). In some cases, the use of other genes for phylogenetic inference even becomes essential, since it overcomes the limitation of 16S rRNA gene sequences in the phylogenetic resolution of close taxa (Konstantinidis & Tiedje, 2005).

Abbreviations: ILD, incongruence length difference; *L.*, *Leuconostoc*; LAB, lactic acid bacteria; ML, maximum-likelihood; MP, maximum-parsimony; NJ, neighbour-joining.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are DQ335652–DQ335721, as indicated in Table 1.

Primer details and phylogenetic trees based on amino acid sequences are available as supplementary material in IJSEM Online.

In this study we used partial sequences of *dnaA* (encoding chromosomal replication initiation protein), *gyrB* (encoding DNA gyrase B subunit) and *dnaK* (encoding the 70 kDa heat-shock protein) to clarify the intra- and intergeneric phylogenetic relationships inside the leuconostoc group. 16S rRNA gene and *rpoC* sequences were also used, but with an enlarged set of *Leuconostoc* and *Weissella* strains (11 from *Leuconostoc* and six from *Weissella*) in comparison with previous studies (Morse *et al.*, 1996). Phylogenetic resolution at different taxonomic ranks was also assessed by the additional inclusion of sequences belonging to different strains of the same species in *Leuconostoc*. This study provides (i) an assessment of confidence in the intrageneric resolution of the taxonomic structure of *Leuconostoc* and *Weissella*, (ii) a recognition of the relative timing of the events leading to the separation of *Leuconostoc*, *Oenococcus* and *Weissella* and (iii) an analysis of the adequacy of different genes, including the 16S rRNA gene, for phylogenetic inference in this group.

METHODS

Bacterial strains and growth conditions. All *Leuconostoc* and *Weissella* strains (see Table 1) were grown in MRS medium at 30 °C.

Escherichia coli XL-1 Blue MRF' (Stratagene) was grown at 37 °C in LB broth supplemented with 100 µg ampicillin ml⁻¹ when required for cloning purposes.

DNA extraction, PCR conditions and cloning strategies. DNA extraction followed the method described by Pitcher *et al.* (1989). PCR was performed in a Thermocycler (Stratagene) and the reaction mixtures contained 100–500 ng template DNA, 50 pmol of each primer, 200 µM each dNTP, 1.5 mM MgCl₂ and 2.5 U *Taq* polymerase in the supplied buffer (all from Invitrogen), in a final volume of 50 µl. A list of the primers used for DNA amplification and sequencing can be found in Supplementary Table S1, available in IJSEM Online. When required, cloning of PCR products was done in pBlueScript KSII (Stratagene) cloning vector as described by Marchuk *et al.* (1991).

DNA sequencing. In addition to full-length sequences, partially overlapping PCR products (amplified with internal primers) were also used in the sequencing process. DNA sequences used in this work were obtained using an automated DNA capillary sequencer CEQ 2000-XL (Beckman Coulter) by a dye-labelled dideoxy termination method (DTCS start kit; Beckman Coulter). Sequence accession numbers can be found in Table 1.

Sequence alignment. For phylogenetic inference, six different alignments were created, five corresponding to the alignment of sequences of each gene, 16S rRNA gene, *dnaA*, *gyrB*, *rpoC* and *dnaK*, and one corresponding to the concatenation of all gene sequences, which was used in a Bayesian approach. All alignments were made

Table 1. Strains and GenBank accession numbers of the sequences used in this work

For 16S rRNA gene sequences obtained from complete genome sequences of *L. mesenteroides* subsp. *mesenteroides* ATCC 8293^T (GenBank accession no. AABH00000000), *O. oeni* PSU-1 (AABJ03000000), *Lactobacillus plantarum* WCFS1 (28376974), *Lactobacillus johnsonii* NCC 533 (42518084), *Lactobacillus lactis* II1403 (15671982) and *Enterococcus faecalis* V583 (29374661), locus identification is given by chromosomal localization numbers. In *L. mesenteroides* subsp. *mesenteroides*, numbers refer to locus AABH02000034 and, in *O. oeni*, refer to locus NZ_AABJ03000008. Sequences DQ335652–DQ335712 were generated in this study.

Strain	16S rRNA	<i>dnaA</i>	<i>gyrB</i>	<i>rpoC</i>	<i>dnaK</i>
<i>Leuconostoc mesenteroides</i> subsp. <i>mesenteroides</i> ATCC 8293 ^T	21453–22834	ZP_00063287	ZP_00063283	ZP_00064061	ZP_00062807
<i>Leuconostoc mesenteroides</i> subsp. <i>cremoris</i> CECT 872 ^T	M23034	DQ335698	DQ335681	DQ335669	DQ335653
<i>Leuconostoc mesenteroides</i> subsp. <i>dextranicum</i> CECT 912 ^T	AB023246	DQ335697	DQ335682	DQ335668	DQ335652
<i>Leuconostoc citreum</i> CECT 4018	X53963	DQ335699	DQ335683	DQ335670	DQ335654
<i>Leuconostoc citreum</i> CECT 4025 ^T	AF111948	DQ335700	DQ335684	DQ335671	DQ335655
<i>Leuconostoc gelidum</i> DSM 5578 ^T	AB022921	DQ335701	DQ335685	DQ335672	DQ335656
<i>Leuconostoc pseudomesenteroides</i> CECT 4027 ^T	X95979	DQ335702	DQ335686	X95812	DQ335657
<i>Leuconostoc fallax</i> DSM 20189 ^T	AB023239	DQ335703	DQ335687	DQ335673	DQ335658
<i>Leuconostoc ficulneum</i> DSM 13613 ^T	AF360736	DQ335705	DQ335689	DQ335675	DQ335660
<i>Leuconostoc fructosum</i> DSM 20349 ^T	AF360737	DQ335704	DQ335688	DQ335674	DQ335659
<i>Leuconostoc pseudoficulneum</i> LC51 ^T	AY169967	DQ335706	DQ335690	DQ335676	DQ335661
<i>Oenococcus oeni</i> PSU-1	9941–11314	ZP_00320217	ZP_00320221	ZP_00319072	ZP_00319805
<i>Weissella paramesenteroides</i> DSM 20288 ^T	AB023238	DQ335707	DQ335691	X95813	DQ335662
<i>Weissella hellenica</i> DSM 7378 ^T	X95981	DQ335708	DQ335693	X96470	DQ335663
<i>Weissella confusa</i> DSM 20196 ^T	M23036	DQ335709	DQ335692	DQ335677	DQ335664
<i>Weissella kandleri</i> CECT 4307 ^T	X52570	DQ335712	DQ335696	DQ335678	DQ335665
<i>Weissella halotolerans</i> CECT 573 ^T	AB022926	DQ335711	DQ335694	DQ335679	DQ335666
<i>Weissella viridescens</i> DSM 20410 ^T	AB023236	DQ335710	DQ335695	DQ335680	DQ335667
<i>Lactobacillus plantarum</i> WCFS1	159863–161433	NP_783867	NP_783871	NP_784718	NP_785552
<i>Lactobacillus johnsonii</i> NCC 533	558550–560200	NP_964015	NP_964019	NP_964353	NP_965281
<i>Lactococcus lactis</i> II1403	537561–539108	NP_266157	NP_267061	NP_267956	NP_267110
<i>Enterococcus faecalis</i> V583	248466–249987	NP_813814	NP_813818	NP_816835	NP_815030

using CLUSTAL W program (Thompson *et al.*, 1994). For protein-coding genes, DNA alignments were made based on the corresponding amino acid alignments by using CodonAlign 2.0 (Hall, 2004). Aligned regions with gaps or where the alignment was of low confidence, as a result of the presence of surrounding gaps, were removed before phylogenetic analyses.

Phylogenetic analyses. Phylogenetic analyses including neighbour-joining (NJ), maximum-parsimony (MP) and maximum-likelihood (ML) trees of DNA sequence alignments were conducted in PAUP* 4.0b10 software (Swofford, 2003). Genetic distances used in NJ trees are Kimura two-parameter-model distances with a transition/transversion ratio of 2:1. Analyses of protein sequences were made using several programs present in the PHYLIP software package, version 3.6 (Felsenstein, 2002). Bootstrap analysis was made with 1000 replicates except in ML, where only 100 replicates were generated. For the Bayesian approach, MrBayes software version 3.0b4 (Huelsenbeck & Ronquist, 2001) was used. Each analysis consisted of 2.0×10^7 generations from a random starting tree and four Markov chains (with default heating values) sampled every 100 generations. The first 8000 sampled trees were discarded, resulting in a set of 12 000 analysed trees sampled after stationarity. To prevent a point of only apparent stationarity being reached, two separate runs were made for each analysis. Hierarchical likelihood-ratio tests were conducted with a batch file supplied with ModelTest 3.06 (Posada & Crandall, 1998) to provide the evolutionary models used in ML and Bayesian analysis. *Lactobacillus plantarum* WCFS1, *Lactobacillus johnsonii* NCC 533, *Lactococcus lactis* I1403 and *Enterococcus faecalis* V583 were used as outgroups (note that the abbreviation *L.* is used throughout for *Leuconostoc*). For the comparison of genus rates of evolution, Kimura two-parameter distances were used to calculate 95% confidence limits taken from Student's *t* distributions in *Leuconostoc* and *Weissella*.

RESULTS AND DISCUSSION

Alignment of DNA sequences resulting from the sequencing of PCR-amplified products provided a total of 7672 nucleotide sites for phylogenetic analysis: 1381 from the 16S rRNA gene, 660 from *dnaA*, 1140 from *gyrB*, 2793 from *rpoC* and 1698 from *dnaK*. Cloning was required prior to sequencing of *gyrB* for some strains, because the use of the primers described by Yamamoto & Harayama (1995) (see Supplementary Table S1) often led to the amplification of a second PCR product (similar to *parE* in *Bacillus subtilis*, which encodes a topoisomerase type IV, B subunit) with the same size as *gyrB*. For protein-coding genes (except *dnaK*), the third codon position was included in the analysis of DNA sequences in order to include the maximum phylogenetic information, at the same time minimizing the introduction of sampling errors. This procedure was supported by the facts that (i) transition/transversion rate ratios did not indicate that a saturation plateau was reached for this codon position in any of the analysed protein-coding genes, (ii) incongruence length difference (ILD) tests also indicated that the nucleotides at this position could be included for phylogenetic analysis (Bull *et al.*, 1993) and (iii) ML tree reconstruction methods that optimize several parameters such as variation of substitution rates among sites, which can accommodate differences resulting from the use of different codon positions, were also used in this study (Yang, 1996). For *dnaK* sequences, we chose to remove the third codon position from the analysis since ILD tests

revealed a value of $P \leq 0.001$, indicating heterogeneity of the tested partitions (i.e. codon positions), thus preventing combination of the data.

Phylogenetic analysis of the 16S rRNA gene

Phylogenetic analysis of 16S rRNA gene sequences resulted in the NJ tree (Fig. 1) that is usually regarded as representative of the phylogenetic relationships of 'leuconostocs'. The genus *Weissella* constitutes one line of descent and the other line includes both *Oenococcus* and *Leuconostoc*. In *Leuconostoc*, most of the species cluster in one group, here represented by *Leuconostoc mesenteroides*, *Leuconostoc pseudomesenteroides*, *Leuconostoc citreum* and *Leuconostoc gelidum*. The remaining species form another group (represented by *Leuconostoc ficulneum*, *Leuconostoc pseudo-ficulneum* and *Leuconostoc fructosum*), with the exception of *Leuconostoc fallax*, which constitutes the most peripheral line in the genus. As found previously (Yang & Woese, 1989; Martinez-Murcia & Collins, 1990; Collins *et al.*, 1991), in *Weissella* the first branching pattern leads to the formation of two groups, one with *Weissella paramesenteroides* and *Weissella confusa* (together with *Weissella hellenica*) and the other including *Weissella viridescens*, the type species of the genus, and *Weissella halotolerans* plus *Weissella kandleri*. Also, as reported by Collins *et al.* (1991) but not so clear in the other studies, the three genera have distinct branch lengths. A long branch leading to *O. oeni* reveals its fastest evolutionary rate, which is followed by *Leuconostoc*, although *L. fallax* seems to be remarkably slower than the remaining *Leuconostoc* species.

In this study, as shown in Fig. 1, branch support is given by bootstrap values resulting from three different phylogenetic approaches, namely distance, parsimony and likelihood, given by NJ, MP and ML trees. In ML analysis, the general time reversible (GTR) model (Lanave *et al.*, 1984) with gamma (Γ) variation and a proportion of invariable sites (I) was used. Groups are only considered when bootstrap support was above 70% for the three methods. The rationale for this approach rests on the fact that, if a clade is supported by different methods of tree reconstruction (with their specific advantages and pitfalls), we can be confident that the assessment of monophyly comes from the data under analysis and not from the method in use. Inevitably, this will lead to conservative results in the estimation of topology, but reduces the eventual necessity for a precise interpretation of evolutionary reality, such as rate heterogeneity among sites and lineages. Still, some situations (such as differences in group evolutionary rates associated with extreme disproportion of internal and terminal branch lengths) may lead to the recovery of the same wrong tree or monophyletic group in a well-known phylogenetic artefact called long-branch attraction (Felsenstein, 1978), even when different methods are used (Huelsenbeck & Hillis, 1993; Huelsenbeck, 1995). We can see that, in the 16S rRNA gene phylogeny (Fig. 1), although many of the groups receive bootstrap support above 70% in the NJ analysis, only a few are reproduced in the MP and ML trees. In general, groups

Fig. 1. NJ tree obtained from 16S rRNA gene sequences and Kimura two-parameter model distances. Sequence accession numbers are provided in Table 1. Bootstrap percentages refer to NJ/MP/ML analysis. Only bootstrap values above 50 % are shown. Bar, 10 % sequence divergence.

that have short basal branches are less well supported than groups with longer branches. This is independent of the taxonomic ranking, as in the case of the clear separation of *L. citreum* from *L. gelidum* and the less well-supported separation of *L. mesenteroides* from *L. pseudomesenteroides*, and we believe this to be a reflection of the different degrees of genetic resemblance between different taxa of the same taxonomic rank.

It is noteworthy that, although the monophyly of the *Leuconostoc*–*Oenococcus*–*Weissella* clade is clear, as well as that of two of the main groups in *Leuconostoc*, there is no complete evidence of the monophyly of either *Leuconostoc* or *Weissella*. This is due mainly to the ML approach, which takes into account the existence of rate variation among sites and among lineages, which are estimated from the data. Nevertheless, there is still strong evidence for the evolutionary line leading to *Oenococcus* and *Leuconostoc*. It should be noticed that the position of *L. fallax* as an outgroup of the remaining *Leuconostoc* species is not clear from this analysis.

Phylogenetic analysis of protein-coding genes

Fig. 2(a–d) shows NJ phylogenies made with nucleotide data from the different protein-coding genes. Phylogenetic

analysis based on protein sequence trees led to the same overall conclusions, and these trees are provided as Supplementary Figs S1–S12 in IJSEM Online. ML trees were made using the TrN (Tamura & Nei, 1993) model for *dnaA* sequences and GTR model for *gyrB*, *rpoC* and *dnaK* sequences, with gamma rate variation and a proportion of invariable sites. NJ trees from the different genes resulted in the formation of the same phylogenetic groups when there was good bootstrap support. The only exception refers to the clustering of *W. viridescens* with *W. paramesenteroides*, *W. hellenica* and *W. confusa* with bootstrap values of 99/84/85 % in *rpoC* analysis (Fig. 2c), in disagreement with its clustering with *W. halotolerans* in the *dnaK* tree (Fig. 2d) with bootstrap values of 100/100/96 %.

While some clades are well defined in every gene tree, such as *L. mesenteroides* or *L. citreum* and the supraspecific groupings of *L. fructosum*–*L. ficulneum*–*L. pseudoficulneum* or *W. paramesenteroides*–*W. hellenica*, many other are gene-specific. The existence of many clades common to those of the 16S rRNA gene phylogeny reveals a general consistency of the data, although differences in clade support and branching order deserve more attention. Regarding *Weissella*, it is worth mentioning that the monophyly of the genus receives high support in *rpoC* analysis. The

Fig. 2. NJ trees (Kimura two-parameter model distances) obtained from sequences of protein-coding genes *dnaA* (a), *gyrB* (b), *rpoC* (c) and *dnaK* (d). Sequence accession numbers are provided in Table 1. Bootstrap percentages refer to NJ/MP/ML analysis. Only bootstrap values above 50 % are shown. Bars, 10 % sequence divergence.

possible grouping of *W. confusa* with *W. paramesenteroides* and *W. hellenica* becomes evident in the *dnaK* phylogeny (Fig. 2d). The clustering of *W. viridescens* and *W. halotolerans* is suggested not only in 16S rRNA gene analysis but also in the *gyrB* (Fig. 2b) and *dnaK* phylogenies; however, only in this last case do the bootstrap proportions become relevant. The relative phylogenetic position of *W. kandleri* remains uncertain, although it may be the result of an early divergent line.

Another uncertainty revealed by the analysis of protein-coding genes concerns the evolutionary line leading to *O. oeni*. In the majority of cases, *O. oeni* appears as an outgroup to both *Leuconostoc* and *Weissella*, although never with strong support; the highest bootstrap values (77/77/69 %) come from the *dnaA* phylogeny (Fig. 2a). In this aspect, only the *gyrB* phylogeny is in agreement with the 16S rRNA gene tree, but still unsupported. As expected in such uncertain cases, the branch leading to the ingroup in question (*Leuconostoc* and *Weissella* or *Leuconostoc* and *O. oeni*) is distinguished by being of short length, revealing the proximity of the two branching events.

Another discrepancy of the phylogenies obtained with the 16S rRNA gene and protein-coding genes concerns the recognition of the most divergent line in *Leuconostoc*. This can be identified in cases where there is strong support in the monophyly both for the genus *Leuconostoc* as a whole and the intrageneric clade that does not include the divergent line. In all protein-coding gene phylogenies, the clade of *L. fructosum*, *L. ficulneum* and *L. pseudoficulneum* appears as the first divergent line, although this becomes evident only in the *dnaK* phylogeny, with bootstrap values of 100/91/99 % supporting the monophyly of the other *Leuconostoc* species (the *gyrB* phylogeny does not support a *Leuconostoc* monophyly indisputably). Still, as can be seen in the *dnaA* and *rpoC* phylogenies, *L. fallax* constitutes the next divergent line in *Leuconostoc*. It should also be noted that the marked differences observed in evolutionary rates in the 16S rRNA gene phylogeny are no longer clear. In this respect, the lengths of branches of *Leuconostoc* and *Weissella* species are more alike, and the difference from *O. oeni* is reduced and even inverted in the *dnaA* phylogeny.

Simultaneous analysis of all genes

Since there seems to be a general agreement of the different gene phylogenies, we combined all genes in one single analysis in order to obtain a tree taking into account all of the data. By using MrBayes software and the data partitioned in sets concerning the different gene sequences, we were aiming to obtain posterior probability values to support clades, taking in account gene-specific evolutionary rates and among-site rate patterns (Castoe *et al.*, 2004). Fig. 3 shows the result of this combined approach. It should be stated that, although Bayesian posterior probabilities are not the same as bootstrap values in confidence evaluation, they can be the result of an even more sensitive process in the evaluation of the phylogenetic signal (Alfaro *et al.*, 2003).

High posterior probability values provide confidence in almost all of the presented clades. Only the intraspecific grouping in *L. mesenteroides* has a lower value, indicating difficulty in distinguishing groups at this low degree of divergence. We consider this general improvement in branch support to be the result of an increase in the number of data available, added to the consistency provided by the use of different partitions (Brandley *et al.*, 2005). The majority of groups presented are in agreement with previous analyses. In the *Weissella* clade, we can also see that *W. kandleri* represents a line resulting from the first divergence within this genus. As suggested in the phylogenies of protein-coding genes, the clade of *L. ficulneum*, *L. pseudoficulneum* and *L. fructosum* constitutes the most peripheral group in *Leuconostoc*. In contrast, the formation of the clade of *Leuconostoc* and *Oenococcus* is in agreement with the 16S rRNA gene phylogeny, suggesting the greater value of this gene in the resolution of this branching order.

To assess further the adequacy of the Bayesian tree to explain the evolution of each gene, we performed SH tests with RELL optimization (Shimodaira & Hasegawa, 1999; see also Goldman *et al.*, 2000). In this test, different tree topologies are compared and their adequacy to fit the data is tested based upon the comparison of their log-likelihood values. Usually, a value of $P \leq 0.05$ is used for the rejection of a given tree topology, representing a hypothesis that might explain the data in the form of an alignment. Beyond the topology given by the Bayesian tree, the set of trees tested included the ML tree of each gene and the NJ trees presented in Figs 1 and 2. The Bayesian tree was rejected only when the test was made to the 16S rRNA gene data (result not shown). This might be due to the different phylogenetic position of *L. fallax* in the combined and 16S rRNA gene phylogenies, since its removal from the analysis was enough to prevent the rejection of the Bayesian tree (whereas this did not happen with the removal of any other strain). These results indicate that the phylogenetic tree presented in Fig. 3 provides a good explanatory hypothesis for the evolution of the genes under study and that the observed differences are the result of a lack of phylogenetic information needed for their resolution.

Evaluation of evolutionary rates

The genetic distances from the three genera under study to the four outgroups were plotted (Fig. 4) in order to obtain relative rates of group evolution that are unbiased by phylogenetic reconstruction. Genetic distances of *Leuconostoc* and *Weissella* from outgroup species were considered as Student's *t* distributions resulting from individual species distances. In all but one case (distances based on *dnaA*), *Oenococcus* appears as the most distant group, supporting its aforementioned fast evolutionary nature. Distances based on *dnaA* (Fig. 4b) show that this gene presents the highest rate of divergence, which could preclude its use for this comparison. *Leuconostoc* and *Weissella* are in general at the same distance from the outgroup species, although *Leuconostoc* sometimes presents

Fig. 3. MrBayes tree using concatenated gene sequences from the 16S rRNA gene, *dnaA*, *gyrB*, *rpoC* and *dnaK*. Sequence accession numbers are provided in Table 1. Posterior probability values provide branch support at each node. Bar, 10% sequence divergence.

greater distances. A particular case refers to the 16S rRNA gene distances, in which there is no overlapping of the 95 % confidence intervals. This fact, allied to the highest rate differences shown by *O. oeni*, suggests uniqueness of the evolutionary rate of the 16S rRNA gene in this species, indicating that care should be taken in conclusions based on this gene. In fact, the greater distances presented by *O. oeni* and *Leuconostoc* associated with the existence of a short internal branch (as suggested in phylogenies from protein-coding genes) could lead to their incorrect grouping in 16S rRNA gene trees, as a result of long-branch attraction. It has been shown that high (> 95 %) posterior probability values could be incorrectly assigned in low-rate trees with short branch lengths (Alfaro *et al.*, 2003), which could justify the high support of *Leuconostoc* and *Oenococcus* clustering in MrBayes tree. The difference in branch- or group-specific evolutionary rates could also explain the position of *L. fallax* in 16S rRNA gene phylogenies, as the result of a relatively low rate of evolution in comparison with the other *Leuconostoc* species (see Fig. 1).

Evolutionary and taxonomic implications

In agreement with assertions made previously (Martinez-Murcia *et al.*, 1993), our results indicate that the emergence of the three evolutionary lines leading to the genera

Leuconostoc, *Oenococcus* and *Weissella* must have been the result of a fast diversifying process, in which the order of separation is difficult, if not impossible, to resolve. With respect to 16S rRNA gene evolution, this study shows that any assertion should always take into account the differences found in its apparent rate of evolution. As already mentioned, the presence of higher rates of mutation of this gene in *O. oeni* and *Leuconostoc* relative to *Weissella* is the most likely reason for the high support given to the monophyly of *Oenococcus* and *Leuconostoc*, as a result of long-branch attraction. In this group, this phylogenetic artefact is most likely enhanced by the presence of a slow-evolving lineage, the *Weissella* group.

In cases of low rates of evolution, there is a tendency for the slow-evolving taxa to resemble outgroup sequences, leading to their phylogenetic 'attraction' (Philippe *et al.*, 2005). There is a strong indication that, in this case, this is due to the proximity of *Weissella* and *Lactobacillus* species. In Table 3 of Yang & Woese (1989), we can see that, of the nine signature positions of lactobacilli (all differing from *Leuconostoc* species), only two differ in the *L. paramesenteroides* group (*Weissella*). Furthermore, in the study of Collins *et al.* (1991), where more than 40 other *Lactobacillus* species were used, the discrepancy of branch lengths of *Weissella* and *Leuconostoc* was greater than in the

Fig. 4. Radar charts showing Kimura two-parameter model distances of *Leuconostoc* (dashed lines), *Oenococcus* (solid lines) and *Weissella* (dotted lines) to outgroup species. Distances are based on sequences of the 16S rRNA gene (a), *dnaA* (b), *gyrB* (c), *rpoC* (d) and *dnaK* (e). Ninety-five per cent confidence limits given by the respective Student's *t* distribution are plotted for *Leuconostoc* and *Weissella*.

study of Martinez-Murcia & Collins (1990), where most of the outgroup sequences belonged to other bacterial genera. Our statement is further corroborated by the work of Morse *et al.* (1996) in which, in contradiction to many other studies, a monophyly of *Weissella* and *Leuconostoc* is suggested, albeit with low bootstrap values (65%). In this last case, most of the outgroup sequences belong to phylogenetically distant species. In spite of the misleading information in the 16S rRNA gene data, this does not mean that the suggested monophyly is wrong, but that it should continue to be treated as a so-far unsupported hypothesis.

The genus-specific rate of evolution of the 16S rRNA gene prevents its use in the determination of the precise branching order and in the evaluation of rates of evolution at genomic level. However, this does not preclude its use in intragenetic phylogenetic evaluation, as shown by the general agreement with other gene-based phylogenies. Another outcome of the specific evolution of the 16S rRNA gene is its possible association with the underlying cladogenic process, even if its meaning is unclear. As to the presence of tachytely in *Oenococcus*, this study partially corroborates the findings of Morse *et al.* (1996), as there seems to be no exceptionally fast nature in the evolution of this genus. Although we found it to be generally more distant than *Leuconostoc* and *Weissella* from the outgroups used, the difference is never as large as in 16S rRNA gene-based phylogenies. In addition, the fact that *O. oeni* is the sole species in this genus presents the drawback of reduced taxon sampling, the effect of which on the accuracy of estimation of the evolutionary rate is still unknown.

ACKNOWLEDGEMENTS

I. M. C. and L. Z.-Z. are the recipients of research grants from FCT SFRH/BD/10675/2002 and SFRH/BPD/3653/2000, respectively.

REFERENCES

- Adékambi, T. & Drancourt, M. (2004). Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. *Int J Syst Evol Microbiol* **54**, 2095–2105.
- Alfaro, M. E., Zoller, S. & Lutzoni, F. (2003). Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol Biol Evol* **20**, 255–266.
- Antunes, A., Rainey, F. A., Nobre, M. F., Schumann, P., Ferreira, A. M., Ramos, A., Santos, H. & da Costa, M. S. (2002). *Leuconostoc ficulneum* sp. nov., a novel lactic acid bacterium isolated from a ripe fig, and reclassification of *Lactobacillus fructosus* as *Leuconostoc fructosum* comb. nov. *Int J Syst Evol Microbiol* **52**, 647–655.
- Brandley, M. C., Schmitz, A. & Reeder, T. W. (2005). Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst Biol* **54**, 373–390.
- Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L. & Waddell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Syst Biol* **42**, 384–397.
- Castoe, T. A., Doan, T. M. & Parkinson, C. L. (2004). Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. *Syst Biol* **53**, 448–469.
- Collins, M. D., Rodrigues, U., Aguirre, M., Farrow, J., Martinez-Murcia, A., Phillips, B., Williams, A. & Wallbanks, S. (1991). Phylogenetic analysis of the genus *Lactobacillus* and related lactic acid bacteria as determined by reverse transcriptase sequencing of 16S rRNA. *FEMS Microbiol Lett* **77**, 5–12.
- Collins, M. D., Samelis, J., Metaxopoulos, J. & Wallbanks, S. (1993). Taxonomic studies on some leuconostoc-like organisms from fermented sausages: description of a new genus *Weissella* for the *Leuconostoc paramesenteroides* group of species. *J Appl Bacteriol* **75**, 595–603.
- Dicks, L. M. T., Dellaglio, F. E. & Collins, M. D. (1995). Proposal to reclassify *Leuconostoc oenos* as *Oenococcus oeni* gen. nov., comb. nov. *Int J Syst Bacteriol* **45**, 395–397.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* **27**, 401–410.
- Felsenstein, J. (2002). PHYLIP (phylogenetic inference package), version 3.6a. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA.
- Garvie, E. I. (1976). Hybridization between the deoxyribonucleic acids of some strains of heterofermentative lactic acid bacteria. *Int J Syst Bacteriol* **26**, 116–122.
- Garvie, E. I. (1981). Sub-divisions within the genus *Leuconostoc* as shown by RNA/DNA hybridization. *J Gen Microbiol* **127**, 209–212.
- Goldman, N., Anderson, J. P. & Rodrigo, A. G. (2000). Likelihood-based tests of topologies in phylogenetics. *Syst Biol* **49**, 652–670.
- Hall, B. G. (2004). *Phylogenetic Trees Made Easy: a How-To Manual for Molecular Biologists*, 2nd edn. Sunderland, MA: Sinauer Associates.
- Huelsenbeck, J. P. (1995). The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol Biol Evol* **12**, 843–849.
- Huelsenbeck, J. P. & Hillis, D. M. (1993). Success of phylogenetic methods in the four-taxon case. *Syst Biol* **42**, 247–264.
- Huelsenbeck, J. P. & Ronquist, F. R. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.
- Kandler, O. & Weiss, N. (1986). Regular, nonsporing gram-positive rods. In *Bergey's Manual of Systematic Bacteriology*, vol. 2, pp. 1208–1234. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore: Williams & Wilkins.
- Konstantinidis, K. T. & Tiedje, J. M. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**, 6258–6264.
- Lanave, C., Preparata, G., Saccone, C. & Serio, G. (1984). A new method for calculating evolutionary substitution rates. *J Mol Evol* **20**, 86–93.
- Marchuk, D., Drumm, M., Saulino, A. & Collins, F. (1991). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res* **19**, 1154.
- Martinez-Murcia, A. J. & Collins, M. D. (1990). A phylogenetic analysis of the genus *Leuconostoc* based on reverse transcriptase sequencing of 16S rRNA. *FEMS Microbiol Lett* **70**, 73–84.
- Martinez-Murcia, A. J., Harland, N. M. & Collins, M. D. (1993). Phylogenetic analysis of some leuconostocs and related organisms as determined from large-subunit rRNA gene sequences: assessment of congruence of small- and large-subunit rRNA derived trees. *J Appl Bacteriol* **74**, 532–541.
- Mills, D. A., Rawsthorne, H., Parker, C., Tamir, D. & Makarova, K. (2005). Genomic analysis of *Oenococcus oeni* PSU-1 and its relevance to winemaking. *FEMS Microbiol Rev* **29**, 465–475.

- Morse, R., Collins, M. D., O'Hanlon, K., Wallbanks, S. & Richardson, P. T. (1996).** Analysis of the beta subunit of DNA-dependent RNA polymerase does not support the hypothesis inferred from 16S rRNA analysis that *Oenococcus oeni* (formerly *Leuconostoc oenos*) is a tachytelic (fast-evolving) bacterium. *Int J Syst Bacteriol* **46**, 1004–1009.
- Morse, R., O'Hanlon, K. & Collins, M. D. (2002).** Phylogenetic, amino acid content and indel analyses of the β subunit of DNA-dependent RNA polymerase of Gram-positive and Gram-negative bacteria. *Int J Syst Evol Microbiol* **52**, 1477–1484.
- Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N. & Delsuc, F. (2005).** Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol* **5**, 50.
- Pitcher, D. G., Sanders, N. A. & Owen, R. J. (1989).** Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* **8**, 151–156.
- Posada, D. & Crandall, K. A. (1998).** Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
- Shimodaira, H. & Hasegawa, M. (1999).** Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**, 1114–1116.
- Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kämpfer, P., Maiden, M. C. J., Nesme, X., Rosselló-Mora, R., Swings, J. & 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.
- Swofford, D. L. (2003).** PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), version 4.0b10. Sunderland, MA: Sinauer Associates.
- Tamura, K. & Nei, M. (1993).** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512–526.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Yamamoto, S. & Harayama, S. (1995).** PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* **61**, 1104–1109.
- Yamamoto, S. & Harayama, S. (1998).** Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int J Syst Bacteriol* **48**, 813–819.
- Yang, Z. (1996).** Maximum-likelihood models for combined analyses of multiple sequence data. *J Mol Evol* **42**, 587–596.
- Yang, D. & Woese, C. R. (1989).** Phylogenetic structure of the “leuconostocs”: an interesting case of a rapidly evolving organism. *Syst Appl Microbiol* **12**, 145–149.