Comparative sequence analysis of a recA gene fragment brings new evidence for a change in the taxonomy of the Lactobacillus casei group

Giovanna E. Felis,1 Franco Dellaglio,1 Luca Mizzi2 and Sandra Torriani1,3

Author for correspondence: Franco Dellaglio. Tel: +39 045 8027917. Fax: +39 045 8027928.
e-mail: dellaglio@sci.univr.it

The taxonomic positions of species of the Lactobacillus casei group have been evaluated by sequencing and phylogenetic analysis of a 277 bp recA gene fragment. High sequence similarity between strain ATCC 393T, currently designated as the type strain of L. casei, and the type strain of Lactobacillus zeae, LMG 17315T, has been established, while L. casei ATCC 334 and Lactobacillus paracasei NCDO 151T form a single phylogenetic group. The taxonomic status of species and strains at issue is discussed.

Keywords: Lactobacillus casei, recA, taxonomy

Lactobacillus casei and related species include phenotypically and genetically heterogeneous strains able to colonize various natural and man-made environments (Vescovo et al., 1995). Many strains are currently employed as probiotics or mixed cultures for dairy products. Despite its importance for the food industry, the taxonomy of the L. casei group is still unclear. Historically, the systematics of the L. casei group have been problematic and adjustments have been introduced based on new taxonomic methods that have gradually become available. Initially, Hansen & Lessel (1971) designated strain ATCC 393T as the type strain of L. casei subsp. casei on the basis of a few phenotypic traits. However, Dellaglio et al. (1975) found little homology between authentic strains belonging to the subspecies of L. casei and ATCC 393T in DNA–DNA hybridization experiments. Based on such evidence, Collins et al. (1989) reclassified the majority of reference strains of L. casei subsp. casei, L. casei subsp. alactosus and L. casei subsp. pseudoplantarum as Lactobacillus paracasei. Subsequent DNA–DNA hybridization experiments and numerical analysis of total soluble cell protein patterns led to the proposal of designating strain ATCC 334 as the neotype for the species L. casei, with rejection of the name L. paracasei (Dellaglio et al., 1991). The proposal was denied (Wayne, 1994). Dicks et al. (1996) provided new indications in favour of the reclassification of strain ATCC 334 as the neotype of the species L. casei; moreover, they proposed the reclassification of L. casei subsp. casei ATCC 393T and Lactobacillus rhamnosus ATCC 15820 as strains of Lactobacillus zeae. Mori et al. (1997) performed a comprehensive study of the phylogenetic relationships existing between L. casei, L. zeae, L. rhamnosus and L. paracasei on the basis of a comparative sequence analysis of 16S rRNA genes (16S rDNA). High similarity was found between the sequences analysed, with a maximum of 19 variable nucleotide positions out of more than 1520. Thirteen nucleotides at different positions were indicated as sequence signatures and the results supported the classification of L. paracasei and L. casei ATCC 334 in one species. About 1% maximum divergence and only two signatures differed between L. casei ATCC 393T and L. zeae LMG 17315T, fewer than the seven different signatures found between L. casei ATCC 393T and L. casei ATCC 334 that were not a sufficiently strong argument to show clear interspecific differentiation between L. casei ATCC 393T and L. casei ATCC 334.

Nour (1998) proposed that 16S–23S and 23S–5S intergenic spacer region sequences might represent a useful alternative to 16S and 23S rDNA for establishing the phylogeny of closely related bacterial species, because of the higher sequence and fragment length polymorphism. Comparative analysis of the 23S–5S rRNA intergenic spacer regions of L. casei-related strains was performed by Chen et al. (2000). Their observations supported the request of Dellaglio et al. (1991) to replace strain ATCC 393T with strain ATCC 334 as the neotype of L. casei subsp. casei.
However, the nomenclature of L. casei-group species is still ambiguous, as demonstrated by recently published articles that indicate either ATCC 334 (Yoon et al., 2000) or NCDO 161T (= ATCC 393T) (Swezey et al., 2000) as the L. casei type strain.

The recA gene has been proposed as a useful marker in inferring bacterial phylogeny (Lloyd & Sharp, 1993; Eisen, 1995) and has recently been used successfully to differentiate species of some bacterial genera (Kullen et al., 1997; Gruber et al., 1998; Smith et al., 1999; Maréchal et al., 2000; van Waasbergen et al., 2000). The aim of the present study was to elucidate the phylogenetic relationships between species and strains of the L. casei group as delineated by the recA gene.

The bacterial strains used in this study were L. casei ATCC 334, L. casei ATCC 393T, L. paracasei subsp. paracasei NCDO 151T, L. paracasei subsp. paracasei LMG 9438, L. paracasei subsp. tolerans LMG 9191T, L. rhamnosus LMG 6400T, L. zeae LMG 17315T and Lactobacillus sp. DF1, a strain isolated from a commercial probiotic product. Strains obtained from different culture collections were checked for purity. Strain DF1 was identified biochemically as L. casei by the API 50 CHL kit (bioMérieux).

The strains were grown aerobically in MRS broth for 16 h at 37°C. Genomic DNA of lactic acid bacteria was extracted following the procedure described by Marmur (1961). Three hundred ng chromosomal DNA from each strain was added to a 50 μl PCR mixture composed of 2.5 U Taq DNA polymerase (Sigma-Aldrich) in its buffer, 50 pmol each primer and 100 μM dNTPs. The PCR program was made up of initial denaturation at 94°C for 5 min and 30 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 90 s and extension at 72°C for 1 min. Primers used for the amplification had the following sequences:

- 5′-TTATGAYGAGANCRACAYG-3′ (forward primer, corresponding to amino acids 92–98 of the Escherichia coli RecA protein; Dybvig et al., 1992) and
- 5′-CCWCCWKWGTGTYTNCGG-3′ (reverse primer, corresponding to amino acids 206–211 of E. coli RecA; Duwat et al., 1992). The expected size of the amplicon was 360 bp. Because of the lack of amplification with this primer pair from L. rhamnosus LMG 6400T and strain DF1, we designed an internal reverse primer (5′-ACTTTYCTCMCGMAWCTGAT-3′) based on the sequences obtained for the other strains tested. The amplification program applied with this primer coupled with the above forward primer differed from the former program in the annealing temperature (50°C). The expected size of the amplicon was 318 bp. Fragments of the expected sizes were purified from 2% agarose gel by using the GeneClean kit (Bio 101), ligated in pGEM-T vector (Promega) and electroporated into E. coli XL-1 Blue. Plasmids were extracted from white colonies by the plasmid miniprep kit (QIAGEN) and the inserted fragment length was verified by agarose gel electrophoresis after digestion with AfaI (Takara). Sequencing of inserts was performed at Centro Genoma Vegetale-ENEA CR Casaccia (Rome).

The similarity of the sequences obtained to previously reported recA sequences was determined by a BLASTX search (Altschul et al., 1997) against major molecular databases. Moreover, the presence, in the putative translations of fragments, of Asp-100, Tyr-103, Asp-144 and Ser-145 further confirmed their identity because those residues are considered functionally important (Story & Steitz, 1992). Regions of primer annealing were excluded from the analysis because point mutations were observed due to primer degeneration. Only homologous regions complete in all sequences were included in the analysis. The CLUSTAL W program (Thompson et al., 1994) was used to align both nucleotide and putative amino acid sequences of the recA gene fragments.

In order to avoid misleading phylogenetic conclusions, different reference organisms were selected to be used as outgroups from those for which recA gene sequences have already been determined (Stackebrandt & Ludwig, 1994): Acholeplasma laidlawii (accession number M81465), Acidothermus cellulolyticus (AJ006705), Anaebena variabilis (M29680), Arthrobacter globiformis (AF214780), Bacillus subtilis (X52132), Bifidobacterium breve (AF094756), Clostridium perfringens (U61497), Frankia alni (AJ006707), Geodermatophilus obscurus (AJ006706), Mycoplasma mycoides (L22073), Ruminococcus albus (U30293), Spirulina platensis (U33924), Staphylococcus aureus (L25893), Streptococcus pyogenes (U21934), Synechococcus sp. (M29495) and Lactococcus lactis (M88106). Trees were constructed by calculating distance matrices with the DNADIST program in the PHYLIP software package (Felsenstein, 1993) on nucleotide sequences and PROTDIST on putative amino acid sequences with the default models. Trees were inferred from the matrices using the NEIGHBOR program in PHYLIP, which implements the neighbour-joining method of Saitou & Nei (1987). In order to improve the accuracy of phylogenetic estimation with this outgroup, we produced trees with different methods (Kim, 1993). A phylogenetic tree from gene sequences was also constructed with the DNAML (maximum-likelihood) and DNAPARS (parsimony) programs in PHYLIP (Felsenstein, 1993). Results obtained with all the tree methods were in agreement. Moreover, since it is known that the order of sequence alignment can influence the tree topology (Lake, 1991), we always selected the ‘jumble’ option in program menus to test different sequence orders. Finally, in order to prove the reliability of tree topologies, a bootstrap confidence analysis on 1000 replicates was performed (Felsenstein, 1985), generating a consensus tree. All the different methods tested produced the same topology of the tree. The tree obtained with Bacillus subtilis as the outgroup is reported in Fig. 1. The same classification scheme was obtained with nucleotide and amino acid sequences, but phylogenetic distances for the amino acid tree were shorter (data not shown), due to the
Fig. 1. Neighbour-joining tree obtained with recA gene sequences. Bootstrap values are reported for a total of 1000 replicates. Bootstrap values for the L. paracasei cluster range between 853 and 1000. Trees generated by the maximum-likelihood and parsimony methods have the same topology (not shown). Bar, 10% sequence divergence. In this tree topology, the phylogenetic distance between organisms is the sum of the horizontal segments.

masking of synonymous nucleotide substitution after translation.

Fig. 1 shows that the phylogenetic distance between L. casei ATCC 393T and L. casei ATCC 334 is greater than that between ATCC 393T and the type strain of L. zeae. In contrast, the distances between strains of L. paracasei and L. casei ATCC 334 are virtually zero. These data are strongly supported by the bootstrap values reported.

Because of the similar G + C content of the organisms considered, the differences in sequences are likely due to real evolutionary divergence; we therefore focused our interest on recA nucleotide sequences.

K values (Hori & Osawa, 1979) calculated for recA sequences and K values for 16S rDNA as calculated by Mori et al. (1997) for the type strains are reported in Table 1. It can be noted that, except for the K values between L. casei ATCC 393T and L. zeae LMG 17315T, the K values calculated for recA sequences are at least 30-fold greater than those calculated for 16S rDNA sequences. From these data, it appears that intraspecific K values are at least one logarithmic unit smaller than those between species: for example, L. casei ATCC 334 and L. paracasei strains share K values between 0 and 0.021, while they are clearly distinguished from L. rhamnosus by K values between 0.265 and 0.279. Similarly, they are distinct from L. zeae LMG 17315T, with values between 0.372 and 0.390. From this point of view, it is evident that L. casei ATCC 393T, sharing a K value of 0.0358 with strain LMG 17315T, forms an intraspecific group with L. zeae, while the values it shares with L. casei ATCC 334 and related strains (0.367–0.390) indicate an interspecific relationship. We cannot yet propose a numerical value for recA homology to be used as a taxonomic discriminator among species, however, since L. paracasei and L. rhamnosus, sharing K values between 0.265 and 0.279, are validly considered distinct species; the same status should be accepted for L. casei ATCC 334 and L. casei ATCC 393T, which show higher K values.

Strain DF1, identified biochemically as L. casei, is characterized by a particular position in the phylogenetic tree: it groups together with L. zeae LMG 17315T and L. casei ATCC 393T, but their K values are greater than those obtained for the strains of the L. paracasei group. Therefore, we cannot unequivocally assign this strain to L. zeae.

Table 1. K values for recA internal sequences of Lactobacillus strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. L. casei ATCC 393T</td>
<td>–</td>
<td>0.0358</td>
<td>0.367</td>
<td>0.367</td>
<td>0.378</td>
<td>0.390</td>
<td>0.378</td>
<td>0.162</td>
</tr>
<tr>
<td>2. L. zeae LMG 17315T</td>
<td>0.0040</td>
<td>0.0073</td>
<td>0.0073</td>
<td>0.0073</td>
<td>0.0073</td>
<td>0.0086</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3. L. paracasei subsp. paracasei LMG 9438</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0126</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4. L. paracasei subsp. paracasei NCDO 151T</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5. L. paracasei subsp. tolerans LMG 9191T</td>
<td>0.265</td>
<td>0.383</td>
<td>0.265</td>
<td>0.383</td>
<td>0.265</td>
<td>0.383</td>
<td>0.265</td>
<td>0.383</td>
</tr>
<tr>
<td>6. L. casei ATCC 334</td>
<td>0.275</td>
<td>0.407</td>
<td>0.275</td>
<td>0.407</td>
<td>0.275</td>
<td>0.407</td>
<td>0.275</td>
<td>0.407</td>
</tr>
<tr>
<td>7. L. rhamnosus LMG 6400T</td>
<td>0.367</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>8. Lactobacillus sp. DF1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Analyses based on the recA gene or RecA protein were in general agreement with those performed with 16S rRNA (e.g. Eisen, 1995), thus validating the use of recA as a taxonomic marker; on the other hand, it is plausible that the recA gene, which unlike 16S rRNA encodes a protein, can accumulate a greater degree of variability at the nucleotide level without modifying the product of translation substantially. Therefore, the recA gene can be proposed as a reliable marker for inferring phylogenetic/taxonomic relationships between closely related species because it is not completely free of mutation; though the RecA protein has a fundamental role in the cell and mutations must not alter its function, point mutations that do not alter gene product activity are tolerated. Moreover, while the 16S rRNA is a unique analysable sequence, recA has a double potential: the amino acid sequence can be used to infer relationships between distant taxa without the disadvantages of rRNAs (e.g. Collins et al., 1991; Fox et al., 1992; Hasegawa & Hashimoto, 1993; Sogin et al., 1993; Cilia et al., 1996; Ueda et al., 1999; Gupta, 2000) and the nucleotide sequences can be analysed to separate closely related species.

The observations on recA \( K_{\text{mut}} \) values and phylogenetic inference complement the data on 23S–5S rRNA intergenic spacer regions (Chen et al., 2000), 16S rDNA \( K_{\text{mut}} \) values and sequence signatures (Mori et al., 1997), DNA–DNA relatedness (Dellaglio et al., 1975) and numerical analysis of total soluble cell proteins patterns (Dellaglio et al., 1991) in demonstrating that \( L. \) casei ATCC 393\(^T \) is not representative of the \( L. \) casei species.

The high genetic similarity between \( L. \) casei ATCC 393\(^T \) and \( L. \) zeae LMG 17315\(^T \) indicates that they are members of the same species. Moreover, \( L. \) casei ATCC 334 and the type strain of the species \( L. \) paracasei form a single phylogenetic group.

In an accompanying Request for an Opinion (Dellaglio et al., 2002), we propose that the Judicial Commission consider that the strain ATCC 393\(^T \) is scientifically unsuitable as the type strain of the species \( L. \) casei. For strains belonging to the same taxon as strain ATCC 334 and NCDO 151, the species name \( L. \) casei has to be retained on the basis of its priority over \( L. \) paracasei. Consequently, the latter species, phylogenetically indistinguishable from many representative strains of \( L. \) casei, should be rejected.

References


