Microbiology, Vol. 74, No. 4, 2005, pp. 448–452. Translated from Mikrobiologiya, Vol. 74, No. 4, 2005, pp. 520–525.

EXPERIMENTAL ARTICLES

Elucidation of the Taxonomic Status of Industrial Strains of Thermophilic Lactic Acid Bacteria by Sequencing of 16S rRNA Genes

S. G. Botina*, A. M. Lysenko,[1], and V. V. Sukhodolets***

**State Research Institute of Genetics and Selection of Industrial Microorganisms, Pervyi Dorozhnyi proezd 1, Moscow, 113545 Russia*

***Institute of Microbiology, Russian Academy of Sciences, pr. 60-letiya Oktyabrya 7, k. 2, Moscow, 117312 Russia*

Received May 5, 2004; in final form, June 28, 2004

Abstract—Both phenotypic characteristics and results of PCR tests for the presence of species-specific genes indicate that a number of strains of thermophilic lactic acid bacteria previously considered as belonging to *Streptococcus thermophilus* are actually closely related to enterococci. In the present study, partial (over 500 nucleotides) sequencing of 16S rRNA genes from 12 strains of thermophilic lactic acid bacteria used as starters for manufacturing sour milk products on the territory of the Commonwealth of Independent States (CIS) has been performed. According to the results of the sequencing, seven of the strains have been classified with *Enterococcus durans.* The earlier classification (based on PCR tests) of two of the strains as *S. thermophilus* and three of the strains as *E. faecium* has been confirmed. The data obtained demonstrate that the enterococci *E. durans* and *E. faecium* are widely used as thermophilic starters for manufacturing sour milk products on the territory of the CIS.

Key words: DNA–DNA hybridization, 16S rRNA gene sequencing, thermophilic lactic acid bacteria, streptococci, enterococci.

It has previously been demonstrated that at least six genomovars of thermophilic lactic acid bacteria are used as starters for manufacturing sour milk products on the territory of the Commonwealth of Independent States (CIS) [1–3]. The term *genomovar* is not strictly defined; in practice, it may correspond either to different species for which phenotypic differences have not been determined or to different populations within a single species [4]. In our previous studies [1–3], lactic acid bacteria were classified into different genomovars on the basis of DNA–DNA hybridization data obtained using the method of optical reassociation of DNA in a solution. At least 80% of DNA reassociates in the case of strains belonging to the same genomovar, whereas strains of different genomovars usually show low levels of DNA hybridization (20–40%).

Of the 52 strains of thermophilic streptococci collected on the territory of the CIS, 16, 7, 4, 2, 14, and 9 strains were classified with genomovars I, II, III, IV, V, and VI, respectively [3]. Genomovars I, III, V, and VI contained cultures originating from the central regions of the European part of Russia and from Ukraine, while the strains originating from the southern regions of the CIS (Krasnodar, Baku, and Tashkent) constituted genomovar II. The two strains constituting genomovar IV also originated from Tashkent.

Previously, it was believed that most of the strains of thermophilic lactic acid bacteria used as starters belong to the thermophilic streptococcal species *Streptococcus thermophilus.* According to our DNA–DNA hybridization data, the majority of the strains of the genomovars that we revealed are indeed relatively close to type strain *S. thermophilus* ATCC 19258 (DNA hybridization level of about 50%). The strains of genomovar II are an exception. Their level of DNA hybridization with strain ATCC 19258 and with the rest of the genomovar strains does not exceed 35% [3, 5]. According to these data, the strains of genomovar II should be considered a separate species of thermophilic streptococci. However, we have demonstrated that both strains of genomovar II and a number of strains of the other genomovars of thermophilic lactic acid bacteria are close to enterococci [5].

In that study, the application of PCR tests allowed us to reveal that among the 25 strains previously classified with different genomovars, 9 carry the *Enterococcus faecium ddlM* gene and only three strains carry the *S. thermophilus lacZ* gene. The *Enterococcus faecium ddlM* gene is mostly found in the strains of genomovars II, III, and IV. These results have been confirmed in the present study: we found that, on the basis of their 16S

[1] Corresponding author; e-mail: amlysenko@rambler.ru

0026-2617/05/7404-0448

rRNA gene sequences, the strains with the *S. thermophilus lacZ* gene can be classified with this species and the strains of genomovars II, III, and IV can be classified with *E. faecium*. It was also shown that the majority of the strains from genomovars I, V, and VI can be classified with *E. durans* based on their 16S rRNA gene sequences.

MATERIALS AND METHODS

Bacterial strains. The study was performed with the following strains of thermophilic lactic acid bacteria, which had previously been assigned [1, 3] to different *Streptococcus thermophilus* genomovars on the basis of DNA–DNA hybridization (the genomovar number is given in parentheses): B3371, CK1002, and CK1010 (I); CK1013 (II); 5 and B2095 (III); 722 (IV); 6kb (V); and CK1025 and CK1026 (VI).These strains are employed as starters in the manufacture of sour milk products in different regions of the CIS [1–3]. The present study also used strains B3166 and B3165, which we had not previously studied and which we obtained from the culture collection of the State Research Institute of Genetics and Selection of Industrial Microorganisms, where they are stored as representatives of *S. thermophilus.*

For comparison, nucleotide sequences of 16S rRNA genes of the following enterococcal and streptococcal species from the NCBI database were used: *S. thermophilus* AY188354 and X68418; *E. durans* AJ276354 and AJ 420801; *E. faecium* AJ420800 and AY172570; and *E. faecalis* AJ420803, AF515223, and AB022693. The GenBank accession numbers for the corresponding 16S rRNA sequences were used by us for strain designation.

Cultivation. Bacteria were grown at 42°C on an agarized M21 medium supplemented with glucose carbon source.

Primer design was performed based on the available data on 16S rRNA sequences for different species of streptococci and enterococci. The following software was employed: NCBI Blast2 was used to assess primer affinity to the rDNA sequences; Oligo 6.31 was used to determine the main parameters of PCR, to assess the possibility of duplex and hairpin formation by the primers, and to find possible sites of nonspecific priming on the templates; and Oligonucleotide Properties Calculator was used to determine the primer G+C composition and annealing temperature.

Determination of 16S rRNA gene nucleotide sequences. Amplification of the 16S rRNA genes was performed using the specially designed primers 1F (5'-gag ttt gat cct ggc tca gga cga-3') and 1R (5'-cgc acc ttc cga tac ggg cta cct-3'). The time–temperature profile of the PCR was as follows: 94°C for 5 min; then, 35 cycles of 94°C for 30 s (DNA denaturation), 55°C for 30 s (primer annealing), and 72°C for 1 min (extension); then, 72°C for 4 min (final polymerization). The amplificates were stored at 4°C. The DNAs of the cultures studied were used as templates in the PCR. The PCR was performed in a Perkin Elmer amplifier. The primers were synthesized by Syntol (Moscow, Russia), and the PCR reagents were purchased from Fermentas (Vilnius, Lithuania).

The PCR products were analyzed by electrophoresis in 1% agarose–TBE gel. Isolation of the 16S rRNA gene fragments from the agarose gel and their purification were performed using a Sigma kit for amplified DNA purification according to the manufacturer's recommendations. Analysis of the quality of purification and determination of the sample concentrations were performed in 1% agarose (Helicon)–TBE gel at 15 V/cm with the use of Fermentas (Vilnius, Lithuania) molecular weight markers.

Sequencing of the purified 16S rRNA gene fragments was performed by the Sanger method using an automatic sequencer.

Analysis of the nucleotide sequences of the 16S rRNA genes was performed using online BLAST software (http://www.ncbi.nlm.nih.gov/blast). Multiple alignment of the sequences was performed using Clustal X. Phylogenetic analysis, dendrogram construction, and calculation of phylogenetic distances were performed using online Mega2 software (http://www.megasoftware.net).

RESULTS

To determine the species affiliation of the 12 investigated strains of thermophilic lactic acid bacteria, we sequenced 530-bp fragments of their 16S rRNA genes, located approximately between *E. coli* positions 50 and 580. The sequences obtained were then compared with the GenBank 16S rRNA sequences of the three species *Streptococcus thermophilus, Enterococcus faecium*, and *E. durans*. Two strains of each of these species were used (see Materials and Methods). All of the investigated strains of thermophilic lactic acid bacteria were found to exhibit a high sequence similarity with one of these three species. The obtained sequences were also compared with the *E. faecalis* 16S rRNA gene; however, no pronounced similarity was found.

In accordance with these data, the majority of the strains (seven) were classified with *E. durans.* Three of them (B2095, CK1025, and CK1026) exhibited no differences whatsoever in the 530-nucleotide fragment of their 16S rRNA gene from the *E. durans* strains AJ420801 and AJ276354 taken from the GenBank database. Three of the strains (6kb, B3165, and B3166) differed from the above strains by only one nucleotide substitution, and strain B3371 differed by two substitutions in the central part of the compared gene fragment. Since strains 6kb, B3165, and B3166, as well as strain B3371, differed from the other strains of *E. durans* by the same nucleotide substitution (T instead of C) in the compared

Homology levels among the 16S rRNA genes* of strains of enterococci and a thermophilic streptococcus

Strain or group of strains**	Homology, %										
	No.	1	2	3	4	5	6	7	8	9	10
E. f. group	1										
722	2	98.6									
CK1013	3	99.8	98.4								
E. d. group	4	99.4	98.0	99.6							
6kb group	5	99.2	97.8	99.4	99.8						
B3371	6	99.0	97.6	99.2	99.6	99.8					
AY188354***	7	80.4	78.6	80.4	80.4	80.4	80.2				
CK1010	8	79.6	78.4	79.6	79.7	79.6	79.4	99.2			
CK1002	9	80.7	78.9	80.7	80.7	80.7	80.4	99.8	99.0		
X68418	10	79.9	78.1	79.9	79.9	79.9	79.7	99.4	98.6	99.2	

* According to a comparison of 530-bp nucleotide sequences (see text).

** Composition of the groups of strains. E. f. group: *Enterococcus faecium* AJ420800, AY172570, and strain 5; E. d. group: *E. durans* AJ420801, AJ276354, B2095, CK1025, and CK1026; and 6kb group: 6kb, B3165, and B3166.

*** Type strain *S. thermophilus* ATCC 19258.

16S rRNA gene fragment (data not shown), they can be assigned to a special taxon within this species.

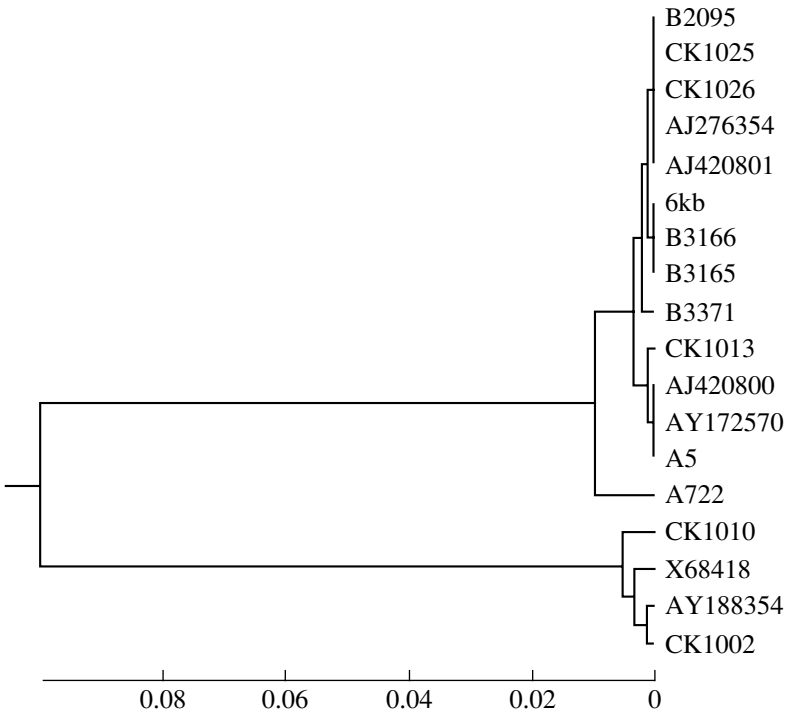
Strains CK1013, 5, and 722, which had previously been classified with *E. faecium* [5] due to PCR test results (see above), exhibited an affinity with this species in respect to the nucleotide sequence of the compared fragment of the 16S rRNA gene as well. Strain 5 has the same nucleotide sequence for this gene fragment as the *E. faecium* strains AJ420800 and AY172570 taken from the GenBank database. Strain CK1013 differs by one nucleotide substitution, while strain 722 has seven unique nucleotide substitutions and five mononucleotide insertions. It is noteworthy that strain 722 originates from Uzbekistan and, according to DNA–DNA hybridization data, represents a well separated genomovar (IV) [1]. Since the PCR species-specific test [5] revealed the presence of an *E. faecium ddlM* gene in this strain, it may represent a geographical variant of *S. faecium* or a separate species very closely related to it.

The results of the comparison of the experimentally determined 16S rRNA gene sequences and sequences taken from GenBank are presented in the table, which shows percent identities of the 530-bp fragments of the 16S rRNA genes (the length of the entire gene is 1542 bp). It can be seen that strains belonging to the same species exhibit a high 16S rDNA similarity (98.4–99.8%). The degree of similarity between the strains of the closely related species *E. faecium* and *E. durans* is almost the same (97.6–99.6%).

Previously, all of the strains of thermophilic lactic acid bacteria under study were considered to be representatives of *S. thermophilus* [1–3]; however, in the present investigation, the 16S rRNA gene sequencing revealed that only two of the strains, CK1002 and CK1010, can be classified with this species. The affiliation of these strains with *S. thermophilus* has previously been demonstrated by the PCR test for the *lacZ* gene [5].

Unlike the closely related species *E. faecium* and *E. durans*, the difference between the strains of these species and the strains of *S. thermophilus* were found to be more substantial, with the 16S rDNA homology varying from 78.4 to 80.7% (see the table).

Genetic distances between the strains can be more conveniently illustrated by a phylogenetic tree (see the figure). Among the strains classified with *E. durans* (in the upper part of the figure), strain B3371 and the above-mentioned group of strains 6kb, B3166, and B3165 form separate branches. It should be noted that strains B3371 (I), 6kb (V), and CK1025 and CK1026 (VI), which had previously been [3] classified into different genomovars (indicated in parentheses), appeared to be distinguished by a small number of nucleotide substitutions and to belong to the *E. durans* branch of the phylogenetic tree. The position of strain B2095 (III) within the *E. durans* branch deserves attention, since, according to the results of the present study and the data from the PCR species-specific test carried out in [5], the majority of the strains previously classified with genomovar III proved to belong to *E. faecium.* The phylogenetic position of this strain requires further analysis; probably, the DNA–DNA hybridization data obtained for this strain are insufficiently accurate, since the results of the PCR test did not confirm the affiliation of this strain with *E. faecium.* Strain 722, which originates from Uzbekistan and represents genomovar IV, forms an isolated branch in the dendrogram. This strain, with its high number of nucleotide substitutions (see above), possibly represents a separate species closely related to *E. faecium.*



Dendrogram of the phylogenetic relationships between thermophilic streptococci and enterococci based on nucleotide sequence homologies of 530-bp 16S rRNA gene fragments. A5 and A722 are strains 5 and 722, respectively. The 0.01 value on the scale bar at the bottom of the figure corresponds to a 1% genetic distance or to one nucleotide substitution per 100 nucleotides.

DISCUSSION

The results of the 16S rRNA gene sequencing that we performed for a number of industrial strains of thermophilic lactic acid bacteria demonstrate that the majority of these strains belong to enterococci. Thus, enterococci are widely used on the CIS territory as starters in the manufacture of sour milk products. Both the fact of the wide use of enterococcal starters, especially for home manufacture of sour milk products, and the possible risk of the emergence of pathogenic variants of these bacteria due to horizontal gene transfer are widely known. Eaton and Gasson [6], for example, distinguish the following strains of enterococci (*E. faecium* and *E. faecalis*): (1) food strains used as starters, (2) strains contaminating foodstuffs, and (3) clinical isolates. The strains of the first group usually contain no or very few genes responsible for virulence, whereas clinical isolates usually carry complete sets of such genes. The study cited demonstrated the possibility of emergence of pathogenic transconjugants of food strains via horizontal transfer of the genes responsible for virulence from pathogenic enterococcal strains.

The strains of thermophilic lactic acid bacteria used in this study represent the seven genomovars previously identified based on DNA–DNA hybridization data [1, 3] (including the genomovar of the type strain of *S. thermophilus*). Although we performed no additional DNA–DNA hybridization studies in the present investigation, two facts are worth consideration: strains of one species can belong to different genomovars and thus have relatively low levels of DNA homology, and genomovar I, apart from enterococci, was reported by us to include *S. thermophilus* strains CK1002 and CK1010 [3]. However, we in fact assigned these strains to genomovar I conditionally [3], since the relatively high values of DNA homology (72–77%) obtained experimentally still required confirmation. Nevertheless, the possibility of formation of chromosomes of similar composition in bacterial species occupying the same ecological niche due to horizontal gene transfer cannot be ruled out.

Horizontal gene transfer may have no effect on the divergence of ribosomal RNA genes, which belong to the group of informational genes. This type of gene has had almost no involvement in horizontal transfer for the last 2 billion years or even longer [7, 8]. Due to the relative stability of ribosomal RNA genes as compared to the majority of other, operational [7], genes, it is the latter that play the major role in speciation processes in bacteria, which is mainly determined by horizontal gene transfer [9]. Divergence and new species formation with relatively minor changes in the informational genes, including 16S rRNA genes, must have been frequent events. It is not accidental that the level of divergence of ribosomal RNA genes is considered more important for differentiation between genera rather than species of bacteria [10].

Diversity in DNA homology levels within taxa characterized by representatives with identical or similar rRNA gene sequences is not surprising in this context. Such diversity was also found in this study: strains 5 (III), CK1013 (II), and 722 (IV), previously classified with different genomovars [3], possibly represent one species: *E. faecium* (it cannot, however, be ruled out that strains CK1013 and 722 do represent individual species closely related to *E. faecium*). The present study also showed that, according to the results of the 16S rRNA gene sequencing, the species *E. durans* comprises strains earlier assigned to different genomovars; in fact, the strains studied proved to belong to this species, and it is possibly the enterococcal species most widely used for manufacturing sour milk products.

The nucleotide sequences of 16S RNA genes determined in this study were deposited with GenBank under the accession numbers AY683829 (strain CK1002), AY683830 (strain CK1010), AY683831 (strain CK1013), AY683832 (strain 722), AY683833 (strain 5), AY683834 (strain 6kb), AY683835 (strain B3371), and AY683836 (strain CK1025).

ACKNOWLEDGMENTS

We thank I.V. Manukhov for his consultations and help in performing this study.

This work was supported in part by the President of the Russian Federation under the program "Leading Scientific Schools" (grant no. NSh-978.2003.4) and by the Russian Foundation for Basic Research (grant no. 03-04-48247).

REFERENCES

1. Averina, O.V., Lysenko, A.M., Ermakova, L.M., Ogai, D.K., and Sukhodolets, V.V., DNA Homology among Strains of Thermophilic and Mesophilic Lactic Streptococci Obtained from Different Sources, *Mikrobiologiya*, 1998, vol. 67, no. 6, pp. 792–798.
2. Lysenko, A.M., Karpushina, S.G., and Sukhodolets, V.V., Divergence of *Streptococcus thermophilus* Strains Determined by DNA–DNA Hybridization, *Mikrobiologiya*, 1999, vol. 68, no. 4, pp. 514–518.
3. Lysenko, A.M., Botina, S.G., Ganina, V.I., and Sukhodolets, V.V., DNA Relatedness, Divergence, and Sibling Species of the Lactic Acid Bacterium *Streptococcus thermophilus, Mikrobiologiya*, 2001, vol. 70, no. 1, pp. 70–76.
4. Rossello-Mora, R. and Amann, R., The Species Concept for Prokaryotes, *FEMS Microbiol. Rev.*, 2001, vol. 25, pp. 39–67.
5. Botina, S.G., Lobanov, A.O., Lysenko, A.M., and Sukhodolets, V.V., Genetic Diversity of Strains of Thermophilic Lactic Acid Bacteria on the Territory of the Commonwealth of Independent States, *Biotekhnologiya*, 2004, no. 2, pp. 3–12.
6. Eaton, T.J. and Gasson, M.J., Molecular Screening of *Enterococcus* virulence Determinants and Potential for Genetic Exchange between Food and Medical Isolates, *Appl. Environ. Microbiol.*, 2001, vol. 67, pp. 1628–1635.
7. Rivera, M.C., Jain, R., Moore, J.E., and Lake, J.A., Genomic Evidence for the Two Functionally Distinct Gene Classes, *Proc. Natl. Acad. Sci. USA*, 1998, vol. 95, pp. 6239–6244.
8. Jain, R., Rivera, M.C., and Lake, J.A., Horizontal Gene Transfer among Genomes: The Complexity Hypothesis, *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, pp. 3801–3806.
9. Lan, R. and Reeves, P.R., Gene Transfer Is a Major Factor in Bacterial Evolution, *Mol. Biol. Evol.*, 1996, vol. 13, pp. 47–55.
10. Stackebrandt, E. and Goebel, B.M., Taxonomic Note: A Place for DNA–DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology, *Int. J. Syst. Bacteriol.*, 1994, vol. 44, pp. 846–849.