Microbiology, Vol. 71, No. 6, 2002, pp. 707–711. Translated from Mikrobiologiya, Vol. 71, No. 6, 2002, pp. 819–823.

EXPERIMENTAL ARTICLES

Comparison of Genomes in *Streptococcus thermophilus* Strains of Different Origin

S. G. Botina*, A. M. Lysenko, V. V. Sukhodolets*, and M. A. Trenina***

**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 December 27, 2001

Abstract—According to DNA hybridization data, thermophilic streptococci used in Russia as starters in the dairy industry are divided into six different genomovars, with a degree of DNA homology not exceeding 20–50%. The analysis of genomes from these genomovars using *Sma*I restriction endonuclease and pulsed-field gel electrophoresis revealed wide variability of the genome size. In some strains, the genome size considerably exceeded 2000 kbp. Most of the strains studied contained plasmids about 120 kbp in size.

Key words: thermophilic streptococci, genome, *Sma*I restriction fragments, pulsed-field gel electrophoresis, *Streptococcus thermophilus*, plasmids.

Thermophilic streptococci are widely used as starters in the production of yogurts and different varieties of cheese. These streptococci differ from other lactic acid bacteria in that they are able to grow at elevated temperatures (from 44 to 52°C). This phenotypic feature often serves as the only ground to ascribe a lactic acid bacterium to the species *Streptococcus thermophilus.* Our previous works [1–3] showed that about 50 investigated domestic strains of thermophilic streptococci used as starters belong to six groups of DNA homology (designated as groups I–VI), differing in the DNA hybridization level. The latter was no lower than 80–90% between members of the same group and did not exceed 20 to 50% between the members of different groups. However, some strains belonging to different groups showed a relatively high degree of DNA homology (up to 60%), indicating that these strains are genetically related.

There is debate in the literature as to how phenotypically similar but genomically slightly differing groups of prokaryotes should be named [4]. In our opinion, the most appropriate relevant term is "genomovar" [5, 6]. In this work, we will use this term instead of the previously used "DNA homology group." Low levels of DNA reassociation (lower than 60%) between different *S. thermophilus* genomovars may indicate their affiliation to different species, whereas DNA homology levels ranging between 60 and 100% indicate their affinity at the species level [4–6].

In the present work, we comparatively studied the genomes of thermophilic streptococci from different genomovars using *Sma*I restriction endonuclease and pulsed-field gel electrophoresis.

MATERIALS AND METHODS

Experiments were carried out with 17 strains of thermophilic streptococci, including the type strain *S. thermophilus* ATCC 19258. Most of these strains were isolated from the cultures of lactic acid bacteria obtained from the All-Russia Research Institute of the Dairy Industry in Moscow, but the others were isolated from fermented dairy products of different origin [2, 3]. All the strains of genomovar II were isolated from fermented dairy products manufactured in the south of the former Soviet Union (in the cities of Tashkent, Baku, and Krasnodar). Strains 704 and 722 (genomovar IV) were obtained from O.V. Averina (Tashkent), the strains of genomovar VI from V.I. Ganina, and strain IL1704 from S. Kulakauskas (France).

Our previous studies [1–3] showed that the *S. thermophilus* strains used in the present work are divided into seven genomovars: I (strains B3371, B6448, and CK1010), II (strains CK1004 and CK1013), III (strains 5 and B2094), IV (strains 704 and 722), V (strains 6kb, B1364, and CK1023), VI (strains CK1025, CK1026, and CK1028), and T (strains ATCC 19258^T and IL1704). T is the genomovar of the type strain.

Genomic DNA was studied by digesting it with *Sma*I restriction endonuclease and analyzing the fragments obtained by pulsed-field gel electrophoresis [7]. Fresh cultures of lactic acid bacteria grown in M21 medium were centrifuged, and the precipitated cells

0026-2617/02/7106-0707$27.00

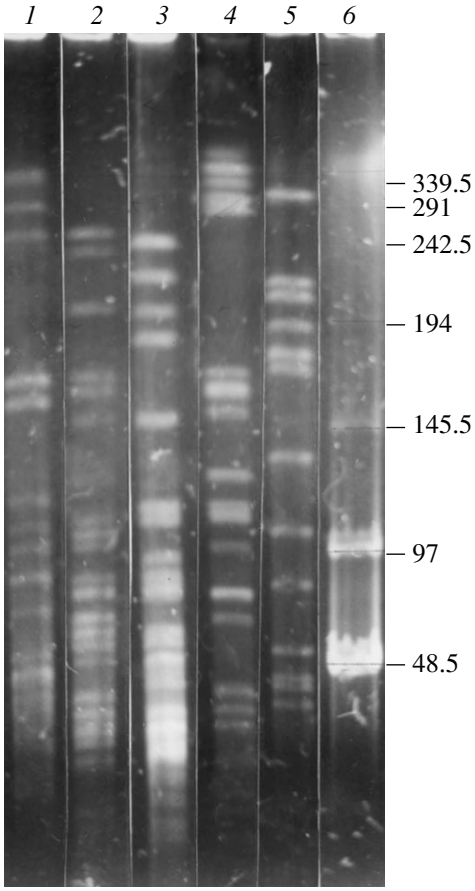

Fig. 1. The pulsed-field gel electrophoresis of the *Sma*I-digested genomic DNA of *S. thermophilus* strains from different genomovars. Lanes: *1*, B3371 (genomovar I); *2*, CK1004 (genomovar II); *3*, strain 5 (genomovar III), *4*, strain 6kb (genomovar V); *5*, CK1026 (genomovar VI); *6*, concatemeric phage λ.

were washed with 50 mM EDTA, suspended in 50 mM EDTA with 10 mg/ml lysozyme, and incubated at 37°C for 1 h. Then, the cells with adsorbed lysozyme were precipitated by centrifugation and suspended in 50 mM EDTA with 1% low-melting-point agarose (Sigma). The suspension was placed in a refrigerator at 4°C for 15 min to allow it to solidify. The agarose blocks were incubated at 60°C in 0.5 M EDTA containing 10% SDS and 0.4 mg/ml proteinase K, washed with distilled water, placed in a sterile flask with 0.5 M EDTA, and stored at 4°C.

To cleave bacterial DNA, a piece of the agarose block with a width corresponding to that of the wells in the agarose gel for pulsed-field electrophoresis was washed with distilled water and digested by *Sma*I restriction endonuclease (MBI Fermentas, Lithuania) according to the manufacturer's instructions.

Pulsed-field gel electrophoresis was carried out in 1% type A agarose (Sigma) in 0.57 × TBE buffer at a field intensity of 6 V/cm, using a Bio-Rad CHEF-DRIII apparatus. The pulse length was gradually (over a period of 24 h) increased from 5 to 15 s. After the separation of the *Sma*I restriction fragments, the gel was stained with ethidium bromide and photographed.

Plasmids were isolated by the method of Anderson and McKay [8].

RESULTS AND DISCUSSION

The pulsed-field gel electrophoresis of the *Sma*I-digested bacterial DNA was primarily used for the biotyping of thermophilic streptococci from different genomovars, which seems to be a necessary procedure for strains lacking genetic markers and subcultured for a long time [9]. For genetically related thermophilic streptococci, such an analysis made with the use of rare-cleaving restriction enzymes (for instance, *Sma*I) allows the similarity of genomes to be evaluated from the number of identical fragments in DNA digests [10].

In our case, genomes belonging to the same genomovar gave a sufficiently large number of identical restriction fragments upon digestion by *Sma*I , whereas the *Sma*I-digested DNA profiles of strains from different genomovars considerably differed (Fig. 1). Based on the path length of various DNA fragments in the agarose gel, we calculated the size of all the *Sma*I restriction DNA fragments of the 17 investigated strains of thermophilic streptococci and evaluated the size of their genomes (table). The table summarizes data for 11 strains, including 5 whose *Sma*I-digested DNA profiles are presented in Fig. 1. Data for the other 6 strains are not presented in the table, as they were obtained with a small number of determinations (less than three).

The error in the determination of the size of DNA fragments depends on their size. The error is maximum in the case of large fragments containing more than 300 kbp, since their paths in the gel are very short and cannot be measured with sufficient accuracy. According to our estimates, the error in the determination of the size of DNA fragments containing > 300, 250–300, 100–250, and 20–100 kbp is equal to > 10, 6, 4, and 2 kbp, respectively. With allowance made for the respective error, the tabulated data can be used to calculate the number of *Sma*I-derived DNA fragments of the same size in particular genomes and their fraction in the total number of fragments. From these data, the degree of genome relatedness can be evaluated [10]. Calculations showed that the fraction of equal DNA fragments in the strains of a particular genomovar was < 50% (in many cases, it was equal to 30–40%), which did not allow the genetic divergence of the strains to be unambiguously estimated. The only exception was the strains CK1004 and CK1013 of genomovar II, which exhibited a large fraction of equal DNA fragments (59%). These strains, however, exhibited almost the same fraction (58%) of equal DNA fragments as did strain 5 from genomovar III.

Furthermore, the strains CK1026 and CK1028 of genomovar VI were found to have 67% of equal *Sma*I

Approximate sizes of fragments in the *Sma*I-digested DNA of various *S. thermophilus* strains and the estimated sizes of their genomes

Fragment rank	Fragment size*										
	B3371 (I)	CK1004 (II)	CK1013 (II)	5 (III)	722 (IV)	6kb (V)	CK1025 (VI)	CK1026 (VI)	CK1028 (VI)	IL1704 (T)	19258 (T)
1	400	258	275	252	252	435	400	320	213	400	400
2	315	242	252	225	233	380	238	222	213	274	277
3	264	206	220	207	208	305	205	213	202	252	192
4	170	171	210	193	187	286	171	197	193	242	179
5	160	164	173	158	171	177	151	182	173	183	173
6	124	153	168	125	134	169	144	176	170	157	172
7	113	114	121	118	119	168	122	138	138	126	167
8	104	108	117	107	116	160	116	111	122	112	108
9	92	104	107	101	108	135	108	87	116	96	101
10	79	90	97	96.5	99	124	84.5	63	86	86	75.5
11	62	83	92	94	92	121	66	52	63	84	57
12	53	74	77.5	73	89	109	56.5	49	52	79	54
13	53	69	70.5	67	74	92	51	41.5	49	69	38
14	50	62.5	63.5	61	71	90	43		41.5	60	
15	44	56	43	55.5	66	81	39			58	
16	39	43	36	45	58	53	30			50	
17	24	36	29.5	40	51	44	19			45	
18	22	29.5	25	38	47	39					
19	13	25		30	41	24					
20		17		26	34	11					
21					25						
Σ	2181	2105	2177	2112	2275	3003	2044	1851.5	1849.5	2373	1993.5

* Fragment sizes are given in kbp. Genomovars are parenthesized. See text for further details.

restriction fragments, including those containing 63, 52, 49, and 41.5 kbp (table). However, these strains (especially strain CK1026) turned out to be dissimilar to the third strain of genomovar VI (CK1025). This suggests that, within genomovar VI, strains CK1026 and CK1028 mutually diverged later than their common precursor diverged from strain CK1025. It should be noted that it is impossible to derive an analogous conclusion on the divergence of particular genomovars, since almost all strains in the genomovars strongly differ in sets of homologous DNA fragments close in size.

The aforementioned size similarity of the *Sma*I restriction fragments of strains CK1004 and CK1013 from genomovar II and strain 5 from genomovar III may be accidental and unrelated to their homology. In general, the size similarity of DNA fragments obtained with the use of restriction endonucleases cannot be treated as convincing evidence of DNA homology. It should be noted in this regard that DNA–DNA hybridization data, which are much more confident than the size similarity of DNA fragments, indicate only a 25% genetic similarity of genomovars II and III [3]. Thus, the comparison of bacterial genomes by the size of their restriction fragments is likely applicable only to the evaluation of the relatedness of genetically close strains. Only in this case may the size similarity of restriction fragments, provided that their general patterns are also similar, reflect the similarity of the respective genomes.

As can be seen from the table, the genome sizes of most of the strains studied exceed 2000 kbp, being as large as ~3000 kbp in strain 6kb (genomovar V). It is possible that the size of the latter genome is actually smaller than this estimate, if we assume that the four largest fragments in the *Sma*I-digested DNA of this strain (Fig. 1, lane *4*) result from a polymorphism of only two fragments. It should be noted that the genomes of strains B1364, CK1033 (genomovar V), B6448, CK1010 (genomovar I), and B2094 (genomovar III), which are not presented in the table, as well as of the tabulated strains IL1704 (genomovar T) and 722 (genomovar IV), also have relatively large sizes (2617,

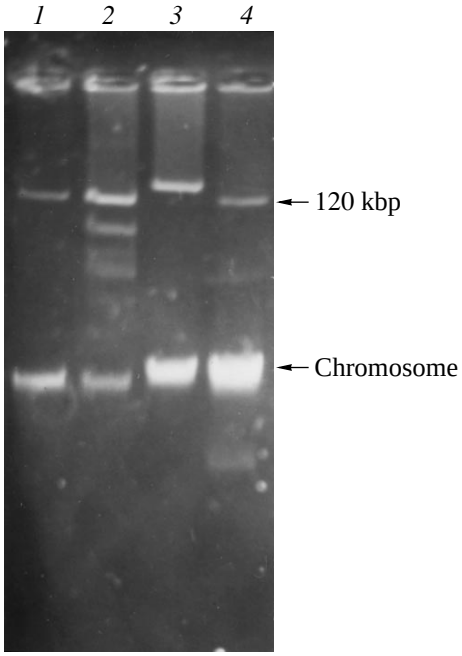

Fig. 2. The plasmid profiles of some strains of thermophilic streptococci. Lanes: *1*, B3371; *2*, CK1028; *3*, CK1026; *4*, strain 6kb.

2140, 2350, 2335, 1960, 2373, and 2275 kbp, respectively).

According to data available in the literature, the *S. thermophilus* strains used in Western Europe as starters in the dairy industry have genomes with estimated sizes close to 1850 kbp [11, 12]. This is just the estimated size of the genomes of two strains from genomovar VI (see table), which possess the most valuable properties as starters. In the degree of DNA–DNA hybridization, the strains of this genomovar are very close to the type strain ATCC 19258 [3], whose genome has a size of about 2000 kbp (table). Unfortunately, to the best of our knowledge, there are no data in the literature as to the size of this genome to prove the correctness of our estimation.

The possibility cannot be excluded that the large sizes of the genomes that we estimated in this work are due, at least partially, to the presence of plasmids in the *Sma*I-digested DNA. To verify this suggestion, we tested 27 strains of thermophilic streptococci from different genomovars for the presence of plasmids and found that most of them contained a large plasmid about 120 kbp in size. Some of these strains also contained smaller plasmids. We failed to reveal any correlation between the presence of plasmids in strains and the distribution of these strains over genomovars. Among the 17 strains studied in the present work, only three strains—B6448 (genomovar I), CK1023 (genomovar V), and CK1025 (genomovar VI)—contain no plasmids. Moreover, both of the strains with the smallest genomes (CK1026 and CK1028) bear plasmids (the latter strain, at least two plasmids) (Fig. 2). Consequently, the presence of plasmids could hardly lead to an overestimation of the genome size.

Of great interest is the fact that only 2 strains of the 17 studied in the present work have genomes of the same size (1850 kbp) as the strains used in Western Europe in the manufacture of yogurt and different types of cheese. These strains are most likely carefully maintained in pure cultures, which excludes the possibility of contamination of their genomes with foreign genes through horizontal gene transfer [13]. In contrast, most domestic industrial and, especially, household strains used as starters are maintained in the form of mixed cultures containing different lactic acid bacteria. Under such conditions, the possibility cannot be excluded that some strains have received foreign genes through their horizontal transfer and have included them into the genomes.

ACKNOWLEDGMENTS

This work was supported by grants nos. 00-04-48147 and 00-15-97855 from the Russian Foundation for Basic Research.

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. Vandamme, P., Pot, V., Gilis, M., De Vos, P., Swings, J., Polyphasic Taxonomy, a Consensus Approach to Bacterial Systematics, *Microbiol. Rev.*, 1996, vol. 60, pp. 407–438.
5. Ursing, J.B., Rossello-Mora, R.A., Garcia-Valdes, E., and Lalucat, J., Taxonomic Note: A Pragmatic Approach to the Nomenclature of Phenotypically Similar Genomic Groups, *Int. J. Syst. Bacteriol.*, 1995, vol. 45, p. 604.
6. Rossello-Mora, R. and Amann, R., The Species Concept for Prokaryotes, *FEMS Microbiol. Rev.*, 2001, vol. 25, pp. 39–67.
7. Nechaeva, A.A. and Sukhodolets, V.V., The Genetic Study of the Industrial Strains of *Lactococcus lactis*: The Detection of Transmissible Plasmids by the Ability

to Ferment Lactose, *Genetika*, 1996, vol. 32, no. 2, pp. 218–227.

8. Anderson, D.G. and McKay, L.L., Simple and Rapid Method for Isolating Large Plasmid DNA from Lactic Streptococci, *Appl. Environ. Microbiol.*, 1983, vol. 46, pp. 549–552.
9. Salzano, G., Moschetti, G., Villani, F., and Coppola, S., Biotyping of *Streptococcus thermophilus* Strains by DNA Fingerprinting, *Res. Microbiol.*, 1993, vol. 144, pp. 381–387.
10. Boutrou, R., Thuault, D., and Bourgeois, C.M., Identification and Characterization of *Streptococcus thermophilus* Strains by Pulsed-Field Gel Electrophoresis, *J. Appl. Bacteriol.*, 1995, vol. 79, pp. 454–458.
11. Roussel, Y., Bourgoin, F., Guedon, G., Pebay, M., and Decaris, B., Analysis of the Genetic Polymorphism between Three *Streptococcus thermophilus* Strains by Comparing Their Physical and Genetic Organization, *Microbiology*, 1997, vol. 143, pp. 1335–1343.
12. O'Sullivan, T.F. and Fitzgerald, G.F., Comparison of *Streptococcus thermophilus* Strains by the Pulsed-Field Gel Electrophoresis of Genomic DNA, *FEMS Microbiol. Lett.*, 1998, vol. 168, pp. 213–319.
13. Prozorov, A.A., The Horizontal Transfer of Genes in Bacteria, *Usp. Sovrem. Biol.*, 2000, vol. 120, no. 6, pp. 515–528.