Lait (1993) 73, 181-190
© Elsevier/INRA

Genetics – Chromosome mapping

Chromosomal genetic instability in *Streptococcus thermophilus*

M Pébay [1], C Colmin [2], G Guédon [1], JM Simonet [2], B Decaris [1*]

[1] *Laboratoire de Génétique et Microbiologie, Université de Nancy I, Faculté des Sciences, BP 239, 54506 Vandœuvre-lès-Nancy;*
[2] *Université de Paris-Sud, CNRS, URA1354, Bâtiment 400, 91405 Orsay Cedex, France*

Summary — The thermophilic and homofermentative lactic acid bacterium *Streptococcus thermophilus* has a relatively small genome (≈ 1.8 Mb). Nevertheless, several loci which were unstable or exhibiting sequence polymorphism were found. The rRNA gene number varies according to the strain. Whereas the CNRZ368 strain contains 6 *rrn* loci, 5% of its offspring only have 5. This alteration was shown to result from deletions occurring within the 2 close rRNA loci, *rrnD* and *rrnE*. Deletions gave rise to hybrid *rrnD/E* loci and probably originated from homologous recombination events. Intraspecific genetic polymorphism also affects the locus *varA*. The cloned I21 fragment belongs to the *varA* locus and reveals the various structures of this locus. Whereas some strains carry the whole I21 fragment, some others lack either its right side or its left side. Both deletions and insertions of various length could result in such a situation. Moreover, evidence was provided that part of the I21 right side is repeated and dispersed in the genome. Another system of genetic instability affects colony morphology. Four variants of *Streptococcus thermophilus* CNRZ368, differing in size, shape and opacity of colonies have been isolated. Switches between phenotypes occur at different frequencies (10^{-3} to 10^{-1}). The 4 genotypes are different, as demonstrated by the 4 characteristic *Xho*I patterns of the 4 variants. In the case of the NST1403 strain, a CNRZ368 subclone exhibiting the ring phenotype, the mutation was shown to be either an insertion or a duplication. Thus *S thermophilus* genomic instability results from various molecular mechanisms, all of them leading to intraspecific polymorphism.

***Streptococcus thermophilus* / genetic instability / intraspecific polymorphism / rrn locus**

Résumé — Instabilité génétique chromosomique chez *Streptococcus thermophilus*. Plusieurs loci instables ont été mis en évidence chez *Streptococcus thermophilus*. Le nombre de loci *rrn* est variable : 5 % des descendants de la souche CNRZ368, souche à 6 loci *rrn*, ne présentent que 5 loci *rrn*. Cette modification résulte d'une délétion entre 2 loci voisins, *rrnD* et *rrnE*, conduisant à la formation d'un locus hybride *rrnD/E*, vraisemblablement par recombinaison homologue. Le locus *varA* présente un polymorphisme génomique intraspécifique. Alors que la totalité du fragment est présente dans certaines souches, les extrémités gauche ou droite manquent dans d'autres. Ces faits peuvent être interprétés soit par des délétions de taille variée, soit par l'insertion d'un élément dans certaines souches. De plus, l'extrémité droite du locus contient une séquence répétée dispersée sur le génome. Un troisième système d'instabilité affecte le phénotype des colonies de la souche CNRZ368. Quatre clones différant par la taille, la forme et l'opacité des colonies ont été iso-

* Correspondence and reprints

lés. Les fréquences de changement de phénotype sont élevées. Chacun présente un profil de restriction caractéristique. Dans le cas du clone de phénotype auréolé NST1403, cette modification du profil résulte d'une insertion ou d'une duplication. Ainsi, plusieurs mécanismes moléculaires créent un polymorphisme génomique intraspécifique.

Streptococcus thermophilus / *instabilité génétique / polymorphisme intraspécifique / locus rrn*

INTRODUCTION

Intraspecific genetic polymorphism is now well documented in bacteria and results from various mechanisms which lead to the occurrence of spontaneous mutations. These could be point mutations, amplifications, deletions or inversions. All of them, neutral as well as deleterious, play a major part in the evolution of the species by increasing its genetic resources.

Our study concerns the homofermentative and thermophilic lactic acid bacteria *Streptococcus thermophilus*, whose genome size is relatively small, *eg* 1.8 Mb (Roussel *et al,* 1992). The variability of 2 loci was studied: the instability of the *rrnD* and *rrnE* loci and the variability of the *varA* locus. On the other hand, colony morphology was shown to be heterogeneous within natural populations of *S thermophilus.* We describe the inheritability and the stability of this character in the CNRZ368 strain.

INSTABILITY OF THE NUMBER OF RRN LOCI

rDNA fragments cloned from *S thermophilus* NST1403 were used as probes to detect intraspecific polymorphism and to study the number of *rrn* loci and their organization in this species (Pébay *et al,* 1992a). The organization was found to be 16S-23S-5S, as in most eubacteria. Whereas most of the strains, including the CNRZ368 strain, contained 6 *rrn* loci, the NST1403 strain only had 5. As the latter strain was obtained by subcloning the former, the lack of one *rrn* locus was probably due to a deletion event. An identical situation was observed in the A054 strain progeny (Guédon *et al,* 1992). These observations suggested that variation in the number of *rrn* loci within a strain of *S thermophilus* can be reproducibly observed.

A genealogical analysis of the CNRZ368 strain directly provided evidence for the genetic instability leading to the loss of a *rrn* locus (Pébay *et al,* 1992b). *Hin*dIII digests of the resulting clones were hybridized with the 23S probe I41 located in figure 1 in order to determine their *rrn* loci number. Four independent mutants exhibiting only 5 *rrn* loci were obtained, all of them missing the same 6-kb long frag-

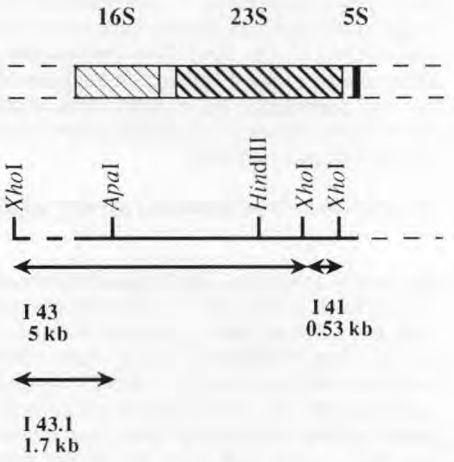

Fig 1. Location of the probes and restriction sites used in this study towards the *rrn* locus map.

ment. The frequency of these deleted mutants is 5 x 10^{-2}. These results demonstrated that a 23S gene was at least partly deleted in these clones.

Deletions were shown to always involve the 2 loci *rrnD* and *rrnE*. Evidence for this was provided by hybridization of the 16S probe I43.1 and of the 23S probe I41 on *Hin*dIII digests of the strain CNRZ368 and of the deleted clones (fig 2). The same 6- kb fragment was revealed in the former strain by both of the 2 probes, whereas it was missing in the deleted clones. As there is one *Hin*dIII site located between the 2 probes, it means that the 6-kb fragment contained the end of a *rrn* locus *(rrnD)* and the beginning of another *(rrnE)*. Consistent results were obtained from experiments carried out with other restriction endonucleases, leading to the conclusion that deletions occurred within 2 closely spaced *rrn* loci. Therefore, they could result from intramolecular homologous recombination events leading to the formation of a hybrid *rrn* locus *(rrnD/E)* (fig 3). Recombination can occur between any point of a locus and the homologous sequence of the other, giving rise to the new

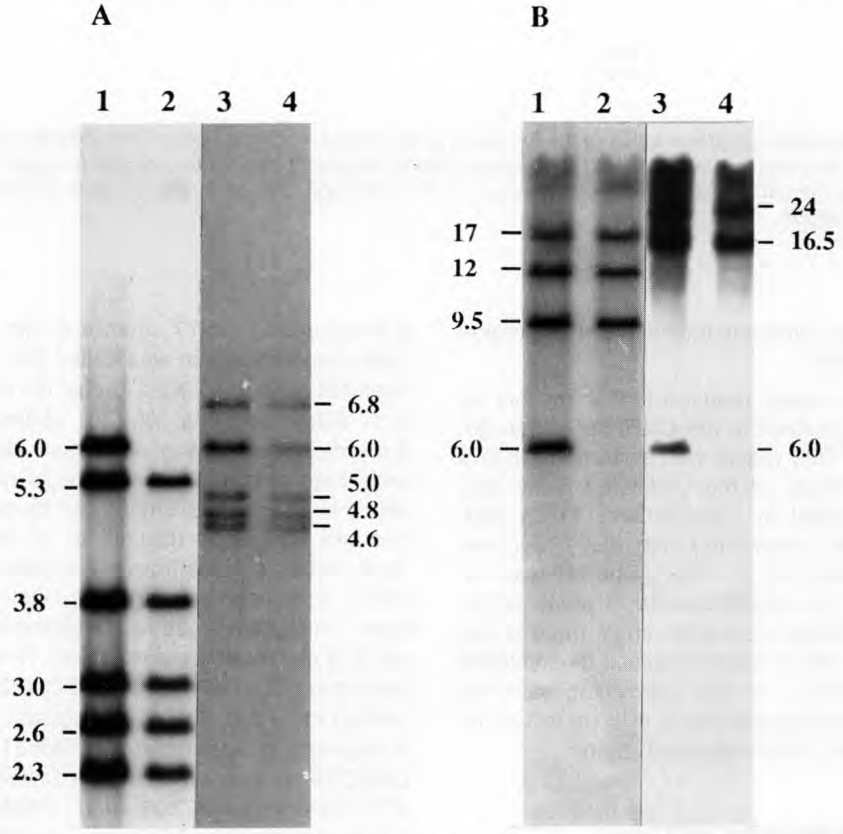

Fig 2. Hybridization patterns of strains CNRZ368 and B210. Fragment sizes are indicated in kilobases. A. *Hin*dIII digests. B. *Apa*I digests. Lane 1: CNRZ368 strain hybridized with I41. Lane 2: strain B210 hybridized with I41. Lane 3: CNRZ368 strain hybridized with I43.1. Lane 4: B210 strain hybridized with I43.1.

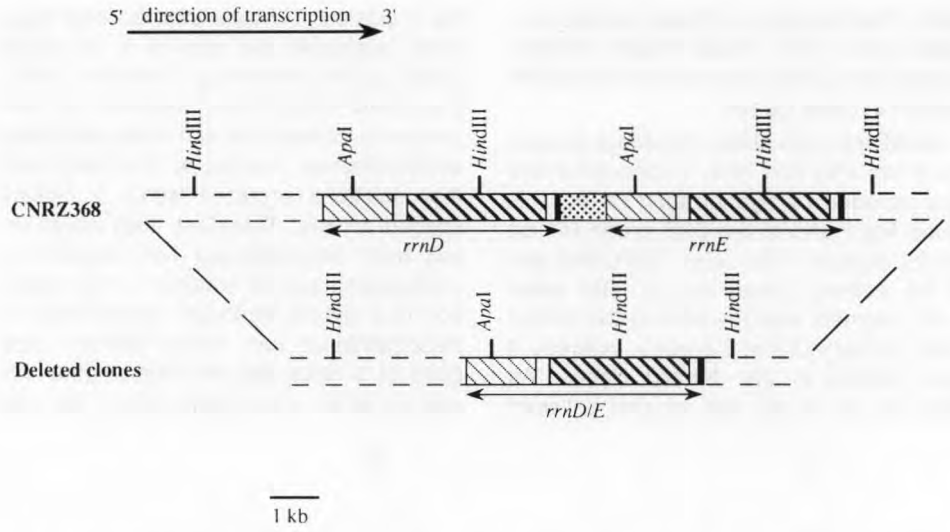

Fig 3. Schematic representation of the formation of the hybrid *rrn* locus. Dotted lines indicate extraoperonic sequences. Upstream and downstream *Hin*dIII fragments remain unchanged in sizes, whereas the internal *Hin*dIII fragment disappears. ▨: 16S gene. ▨: 23S gene. ■: 5S gene. ▩: interoperonic region.

rrn locus, whereas the interoperonic region is deleted.

This model requires that 2 *rrn* loci be closely located in the CNRZ368 strain genome. This region was searched for in a gene library of the undeleted clone B40 constructed in LambdaGem 11: a 6-kb fragment hybridizing with the 16S probe I43.1 and with the 23S probe I41 was observed on *Hin*dIII digests of some of the recombinant clones. Whereas there is one *Hin*dIII site in each *rrn* locus, these clones contained 2 *rrn* loci , providing evidence for tandem organization of 2 *rrn* loci in the genome of the undeleted strains.

Variability of the varA locus

When studying intraspecific genetic polymorphism, we have cloned from the *S thermophilus* NST7 strain a 4.2-kb long *Nde*II fragment which was called I21. This fragment was used as a probe on *Hae*III and *Nde*II patterns of 25 strains of *S thermophilus*. Strong and weak signals were then obtained. On the basis of the strong signals, the 25 strains can be classified into 10 groups (Colmin *et al,* 1991). Three sub-fragments from I21, called A, B, and C, were used as probes on *Hae*III patterns. Among the 9 strains presented in figure 4, 6 situations were revealed. Three of them (the NST7, CNRZ160 and CNRZ 385 strains) carry the whole I21 fragment. The A fragment is missing in the IP6631 and CNRZ702 strains, whereas the CNRZ368, ATCC14485, CNRZ308 and CNRZ302 strains lack the B and C fragments. Whether these DNA sequences are present or not results in an intraspecific genetic polymorphism which is revealed by the I21

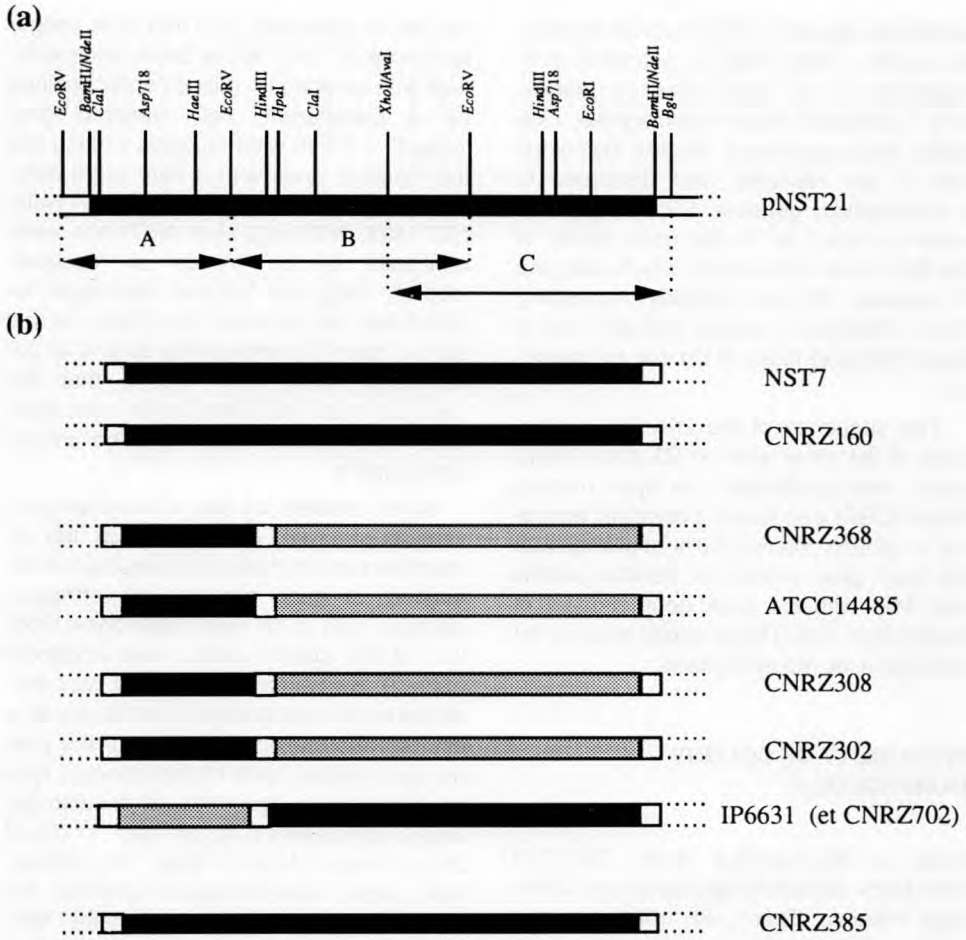

Fig 4. Map of the *varA* locus in 9 strains of *S thermoplihus*.
(a) Restriction map of the I21 fragment (■) cloned in plasmid pNST21 and location of the A, B and C sub-fragments. (b) maps deduced from hybridizations of A, B and C fragments on genomic DNA of various strains. ■: DNA homologous to I21. ▒: DNA fragment missing in the strain. □: limit not exactly determined. ::: : uncharacterized DNA region.

fragment and is more complex than the usually described restriction fragment length polymorphism. This variable chromosomal locus was called *varA*.

A 600-bp long fragment adjacent to I21, on its C-side, was cloned and hybridized on DNA patterns of the CNRZ368 and CNRZ308 strains. A signal was revealed in the CNRZ368 strain DNA but not in the CNRZ308 strain DNA. Thus the polymorphic region missing in the CNRZ308 strain is longer than that missing in the CNRZ368 strain; and thus *varA* extends further than I21.

Besides the strong signals corresponding to the *varA* locus, I21 weakly revealed

additional signals in DNA from all the tested strains. When the A, B and C subfragments of I21 were used as probes, only C revealed these weak signals, indicating that sequences sharing homology with C are repeated and dispersed in *S thermophilus* genome. Such sequences were searched for in the gene library of the B40 clone (see upper), which lacks the C fragment. We are currently sequencing them. Preliminary results indicate that at least 180 nucleotides of I21 can be repeated.

The sequence of the 204 first nucleotides of the other side of I21 (the A fragment), was established. An open reading frame (ORF) was found, potentially encoding a protein sharing 66% similarity with the *pur*A gene product of *Bacillus subtilis* and 44% with the *pur*A gene product of *Escherichia coli.* These genes encode for adenylosuccinate synthetase.

INSTABILITY OF COLONY MORPHOLOGY

When *S thermophilus* strain CNRZ368 from stock collection was spread on TPPY agar medium (Bracquart, 1981), various colony morphologies were observed. Among them, the following 2 phenotypes may be considered as the extremes: on the one hand, white opaque raised circular colonies, on the other, transparent flat irregular colonies. Besides these, many intermediate types were seen: opaque colonies with transparent halos of various widths, transparent flat irregular colonies with one or several opaque papilla, opalescent irregular colonies thicker than transparent ones. By successive subclonings (at least 6) we isolated 1 clone for each of the following types of colony (fig 5), 'opaque': white, opaque, raised, 1-mm wide, circular colonies; 'diffuse': transparent, flat or umbonate, 2–3 mm wide, irregular colonies; 'ring': about 2-mm wide colonies with an opaque, raised center outlined by a transparent, flat, crenated ring; 'edged': ≈ 1 mm wide colonies looking like the 'opaque' ones with a very slight diffusion halo. Strain numbers NST1401, NST1402, NST1403 and NST1404 were attributed to the clones of 'opaque', 'edged', 'ring' and 'diffuse' phenotype respectively. Similar work has been carried out on other *S thermophilus* strains of our collection. It is worth noting that the "opaque" type was obtained in each case whereas the others were isolated only in some strains.

When isolates of the 4 morphological variants of CNRZ368 were plated, the respective variant phenotype persisted in the majority of each progeny. Nevertheless colonies with a different phenotype from that of the parent clone were observed (table I). Switching phenomenon was estimated by the percentage of colonies with a different morphology from that of the plated clone. When cells of the 'opaque' type were spread on agar, 0.3% of the colonies were outlined with a diffuse halo; 11.7% of the colonies obtained from the 'diffuse' type were more opaque, whether the whole colony was thicker, or whether they exhibited opaque papilla. In the latter case the mutation would have occurred after plating. Among 'ring' offspring, 2 types of switching colonies were observed: 7.3% of the colonies had a smaller diffuse or even absent halo than the parent colony, 0.6 % of the colonies had a smaller opaque or even absent center than the parent colony. 0.7% of the colonies of 'edged' progeny had a larger halo than the parent colony. The difference between 'edged' and 'opaque' morphologies was too slight to establish a switching frequency from the former to the latter. These results showed that the 'colony morphology' character was inheritable. Nevertheless, the character is

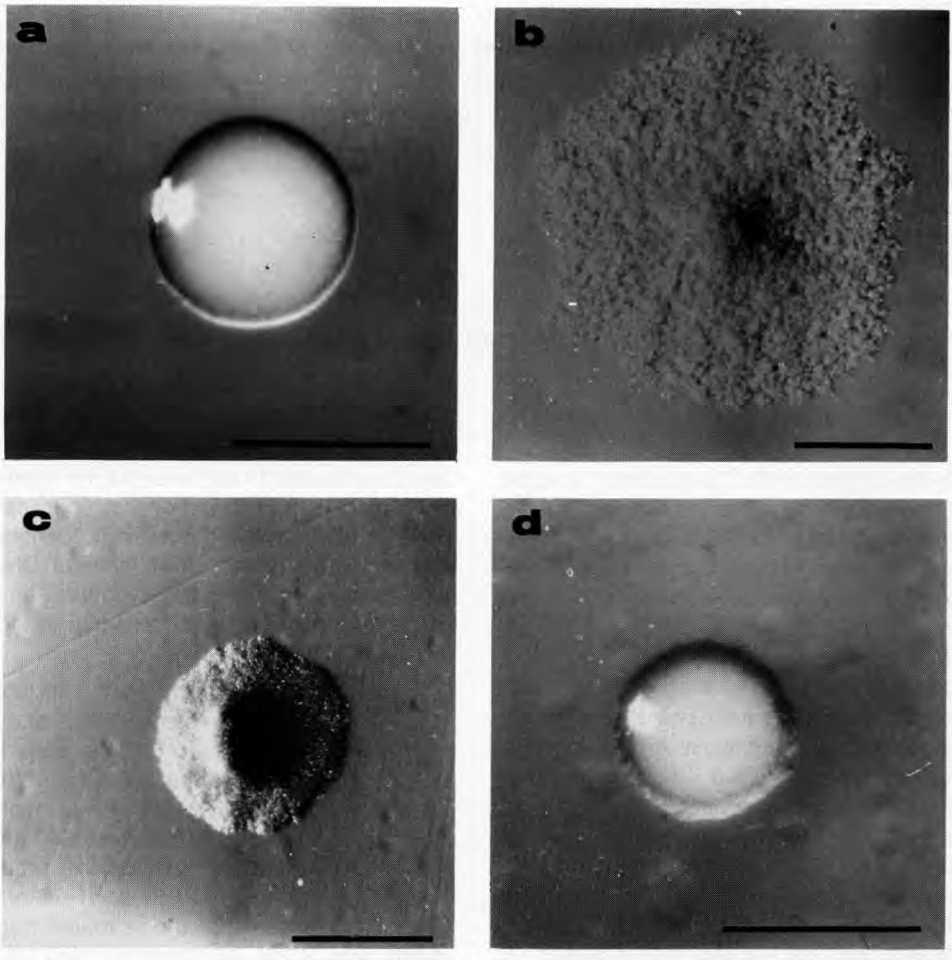

Fig 5. Variant colonies of *S thermophilus* CNRZ368 on TPPY agar medium. (a): 'opaque'. (b): 'diffuse'. (c): 'ring'. (d) : 'edged'. Colonies were grown for 24 h at 42 °C in aerobic conditions. Bars, 1 mm.

unstable, as shown by the appearance of phenotypes different from that of the parent clone. Switching towards opacity was the most frequent, the 'opaque' type looking much more stable than the 'diffuse' type. This may have several explanations: for example, 'opaque' and 'diffuse' mutation frequencies might be different, 'diffuse papilla' on 'opaque' colonies would not be perceptible, 'opaque' mutations might confer a selective advantage.

The genetic origin of morphology variability was then investigated in DNA. No plasmid was found in any of the variants or

Table I. Switching of CNRZ368 variant colonies.

Colony morphology of tested clones	*No of switching colonies*	*No of observed colonies*	*Percentage of switching colonies*
Opaque (NST1401)	10[a]	3 405	0.3
Edged (NST1402)	26[a]	3 675	0.7
Ring (NST1403)	18[a] + 229 [b]	3 115	0.6[a] + 7.3 [b]
Diffuse (NST1404)	553 [b]	4 540	11.7

[a] Switching colonies are more diffuse than the parent colony; [b] switching colonies are more opaque than the parent colony.

in the original strain. So restriction analysis of total DNA was carried out, our extraction procedure providing accurate restriction profiles. Numerous restriction enzymes have been used. With *Aval*, *Eco*RI, *Hin*dII, *Hin*dIII, *Bst*EII, *Sac*I, *Sal*I, *Nde*II, *Bcl*I, *Bgl*I, *Eco*RV, *Hpa*I, *Hpa*II, *Msp*I, *Sma*I, *Nae*I, *Asp*718, *Xho*I, *Pvu*II and *Pst*I, enzymes that cleaved DNAs of the variants, standard conditions of electrophoresis (overnight, 0.8–1 V/cm) generated identical restriction patterns: fragment sizes seemed identical for the 4 variants as were their relative intensities. The very strong similarity of DNA restriction patterns for the different variants demonstrated unequivocally that various morphologies did not result from contaminations and that variants belonged to the same strain. When migrations were allowed to run for a long time (40–48 h) with buffer recirculation, as separation of fragments improved, discrete differences between variant patterns became apparent. For example, on *Hpa*II and *Msp*I digest electrophoresis NST1401 lacked a 8.8-kb fragment, instead of which a 7-kb band appeared. After digestion by *Xho*I, patterns of the 4 variants were distinguished: bands of 7.3, 7.6 or 21.5 kb were respectively characteristic of NST1403, NST1404 and NST1401 patterns, as the NST1402 one did not exhibit any specific band. Differences between patterns are presented in table II. Identical results were obtained from several DNA extractions, dismissing the hypothesis that contaminant DNA was responsible for these differences.

Clones hybridizing with the 7.3-kb long *Xho*I fragment characteristic of strain NST1403 were searched for in a gene library of strain B40 (see upper). One of the positive clones, called λNST19 was hybridized on various digests of strains B40 and NST1403. Because of its size, this probe revealed several signals on each pattern. In most cases, one of the revealed fragments was ≈ 200 bp larger on NST1403 patterns than on B40 pattenrs. Some of these results are presented in table III. First of all, it can be noted that all the restriction fragments characteristic of strain NST1403 are revealed by this probe, providing evidence that all of them result from the same mutation. Moreover, whereas all NST1403 mutated fragments are larger than those from B40, these results demonstrated that this mutation was either an insertion or a duplication of a fragment ≈ 200 pb in length. The 0.95 kb *Nde*II fragment was then subcloned in plasmid pBluescriptKS$^+$, giving rise to plasmid pNST46.

Table II. Size (kb) of the fragments specific of some variant patterns.

Restriction endonucleases	*Recognition sequence* [a]	*NST1401*	*NST1402*	*NST1403*	*NST1404*
*Msp*I	C+C°GG	7	8.8	8.8	8.8
*Hpa*II	C+C+GG	7	8.8	8.8	8.8
*Eco*RV	GA + TATC	14.5	14.5	14.8	14.5
*Hpa*I	GTTAA+C°	4.6	4.6	–	4.6
*Xho*I	CTC+GA+G	21	–	7.3	7.6

[a] As indicated by the manufacturer; ° digestion is not influenced by methylation; + digestion is inhibited by methylation; – no specific band.

Table III. Size (kb) of fragments differing between the NST1403 and B40 strains DNA after hybridization with INST19.

Enzyme	*Size of the fragment characteristic of strain NST1403 (kb)*	*Size of the corresponding fragments in strain B40 (kb)*
*Xho*I	7.3	7.0
*Eco*RV	14.8	14.5
*Hpa*I	4.6	4.4
*Cfo*I	2.6	2.4
*Nde*II	1.15	0.95
*Hae*III	0.55	0.35
*Alu*I	0.7	0.5

This cloning enables further characterization of NST1403 mutation.

CONCLUSION

Whereas *S thermophilus* genome is relatively small in size, one could expect it to be quite compact. Nevertheless, we have shown some dispensable sequences to persist in it. Some chromosomal regions can be deleted in laboratory conditions without resulting lethality. In particular, this is the case of the region lying between the *rrnD* and *rrnE* loci. The high frequency of spontaneously occurring deletions (5%), accurately estimated from a genealogy study, in itself constitutes novel data on the genetic instability affecting *rrn* loci. This high frequency implies firstly that the deletion events frequently occur, and secondly that the deleted mutants are viable. Direct repetitions of *rrn* genes enable intramolecular homologous events to take place, leading to deletion of one *rrn* operon equivalent and to the loss of genetic material lying between the 2 recombining loci. Thus, the probability of losing essential genes increases with the distance between these 2 loci. These considerations explain that all the deleted mutants that we have obtained resulted from crossovers between 2 very close loci. Albeit such deletions of *rrn* loci have been described in *Escherichia coli* (Elwood and Nomura, 1980), they are mostly documented in *Baccilus subtilis* (Loughney *et al,* 1983; Gottlieb *et al,* 1985; Widom *et al,* 1988). Like *S thermophilus*, the latter bacterium belongs to the low GC content subdivision of Gram-

positive bacteria, and some of its *rrn* loci are arranged in clusters.

The *varA* locus is another dispensable locus. In particular, the *purA* gene which belongs to the *varA* locus was shown to be, at least in part, absent from the genome of the strains lacking the C fragment of I21 (the ATCC19258 and the CNRZ702 strains). If this gene is expressed, its loss should result in auxotrophy for AMP. Such an auxotrophy is supplemented in the usual habitat of *S thermophilus*, namely milk, and so should not confer a selective disadvantage.

We have described a high frequency phenotypic switch in the *S thermophilus* CNRZ368 strain. Our study provides evidence that one particular strain of *S thermophilus* can spontaneously exhibit various colony morphologies. The 4 mutants studied present discrete genome modifications which were detected by restriction pattern analysis. Whether these genotypic differences are responsible for the various phenotypes is currently being investigated by cloning and sequencing some of them.

REFERENCES

Bracquart P (1981) An agar medium for the differential enumeration of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* in yoghurt. *J Appl Bacteriol* 51, 303-305

Colmin C, Pébay M, Simonet JM, Decaris B (1991) A species-specific DNA probe obtained from *Streptococcus salivarius* subsp *thermophilus* detects strain restriction polymorphism. *FEMS Microbiol Lett* 81, 123-128

Ellwood M, Nomura M (1980) Deletion of a ribosomal ribonucleic acid operon in *Escherichia coli*. *J Bacteriol* 143, 1077-1080

Gottlieb P, LaFauci G, Rudner R (1985) Alterations in the number of rRNA operons within *Bacillus subtilis* genome. *Gene* 33, 259-268

Guédon G, Pébay M, Colmin C, Decaris B, Simonet JM (1992) The 23S-5S spacer of two rRNA loci of *Streptococcus salivarius* subsp *thermophilus* includes a promoter. *Biochimie* 74, 585-588

Loughney K, Lund E, Dahlberg JE (1983) Deletion of an rRNA gene set in *Bacillus subtilis*. *J Bacteriol* 154, 529-532

Pébay M, Colmin C, Guédon G, De Gaspéri C, Decaris B, Simonet JM (1992a) Detection of intraspecific DNA polymorphism in *Streptococcus salivarius* subsp *thermophilus* by a homologous rDNA probe. *Res Microbiol* 143, 37-46

Pébay M, Roussel Y, Simonet JM, Decaris B (1992b) High-frequency deletion involving closely spaced rRNA gene sets in *Streptococcus thermophilus*. *FEMS Microbiol Lett* 98, 51-56

Roussel Y, Decaris B, Simonet JM (1992) Genome size and organization in *Streptococcus salivarius* subsp. thermophilus. *In: Les Bactéries Lactiques/ Lactic Acid Bacteria*. Centre de Publications de l'Université de Caen, Caen pp 41-55

Widom RL, Jarvis ED, LaFauci G, Rudner R (1988) Instability of rRNA operons in *Bacillus subtilis*. *J Bacteriol* 170, 605-610