Lait 81 (2001) 281–288

Evolution, biodiversity, taxonomy

Polymorphism of *eps* loci involved in exopolysaccharide synthesis of *Streptococcus thermophilus*

Florence CHARRON-BOURGOIN*, Arnaud PLUVINET, Catherine MOREL, Gérard GUÉDON, Bernard DECARIS

Laboratoire de Génétique et Microbiologie, UA INRA 952, Université Henri Poincaré Nancy 1, Faculté des Sciences, BP 239, 54506 Vandœuvre-lès-Nancy, France

Abstract — The nucleotide sequence of two ORFs and five copies of three insertion sequences (IS) types belonging to the *eps* locus of *Streptococcus thermophilus* CNRZ368 involved in exopolysaccharide synthesis are almost identical to ISs and ORFs sequences from *Lactococcus lactis*. Furthermore, sequence comparison of *eps* loci of three *S. thermophilus* strains and hybridization of probes isolated from the *S. thermophilus* CNRZ368 *eps* locus with DNA of 17 *S. thermophilus* strains revealed a very high polymorphism. A small constant region is detected in all the strains whereas a large region is extremely variable. The 17 tested strains could be arranged in six groups according to the presence or absence of a hybridization signal with the different probes tested. A phylogenetic analysis indicated that the ropy NST2280 strain and the non-ropy IP6757 strain are very closely related strains but possess different *eps* loci. The *eps* locus of IP6757 could result from sequence replacement in an IP6757 ancestor. Sequence comparison of different *eps* loci of *S. thermophilus* suggests that the high polymorphism of these loci largely results from sequence replacement following horizontal transfers.

exopolysaccharide / horizontal transfer / *Lactococcus lactis* / polymorphism / *Streptococcus thermophilus*

Résumé — **Polymorphisme des loci *eps* impliqués dans la synthèse d'exopolysaccharides chez *Streptococcus thermophilus*.** L'étude du locus *eps* chez *S. thermophilus* CNRZ368, impliqué dans la synthèse d'exopolysaccharides (EPS), a révélé plusieurs séquences presque identiques à des séquences de *Lactococcus lactis* (5 copies de 3 types d'IS et 2 ORF). Par ailleurs, la comparaison des séquences des loci *eps* de 3 souches de *S. thermophilus* ainsi que l'hybridation de sondes issues du locus *eps* de *S. thermophilus* CNRZ368 sur l'ADN de 17 souches de *S. thermophilus* ont révélé une petite région constante présente chez toutes les souches ainsi qu'une grande région extrêmement variable. Les 17 souches étudiées ont pu être rassemblées en 6 groupes présentant des structures similaires (présence ou non de séquences homologues aux régions testées). Une analyse phylogénétique

* Correspondence and reprints
Tel.: (33) 3 83 91 20 00 (extension 3506); fax: (33) 3 83 91 25 00;
e-mail: Florence.Bourgoin@scbiol.u-nancy.fr

de ces souches indique que *S. thermophilus* NST2280, une souche productrice d'EPS, et IP6757, une souche non productrice, sont très proches phylogénétiquement mais possèdent des loci *eps* différents. Le locus *eps* d'IP6757 résulterait d'un remplacement de séquence qui se serait produit chez un ancêtre d'IP6757. La comparaison des séquences des différents loci *eps* de *S. thermophilus* suggère que des séquences ont été remplacées par des séquences non homologues acquises lors de transferts horizontaux intra- et interspécifiques.

exopolysaccharide / *Lactococcus lactis* / polymorphisme / *Streptococcus thermophilus* / transfert horizontal

1. INTRODUCTION

Cocultures of *Streptococcus thermophilus* and various lactic acid bacteria including *Lactococcus lactis* are used as starters in the production of some cheeses.

Sequence comparison and insertion sequence (IS) distribution showed that horizontal transfers of five IS types have occurred spontaneously between *S. thermophilus* and *L. lactis*, probably during cheese manufacture [1, 2, 6–8]. Furthermore, the *citP* genes from *L. lactis* and *Leuconostoc lactis* which encode citrate permease share 99.2% identity, suggesting a recent horizontal transfer between these species [13]. Moreover, the nucleotide sequences of the putative *hsdS* genes of the pCI65st plasmid of *S. thermophilus* and the pIL7 plasmid of *L. lactis* encoding the specificity subunit of a type I restriction-modification system share 94% identity [9]. Many horizontal transfers of IS and other genes have probably occurred between various lactic acid bacteria species used in the dairy industry, contributing to their genetic polymorphism [8].

2. INVOLVEMENT OF HORIZONTAL TRANSFERS IN THE CHIMERIC STRUCTURE OF THE *EPS* LOCUS OF *S. THERMOPHILUS* CNRZ368

A 32.5 kb variable region of the *S. thermophilus* CNRZ368 chromosome, the *eps* locus, contains seven complete or truncated ISs and 25 ORFs or pseudo-ORFs (Fig. 1). The putative products of 17 ORFs, named *epsA* to *epsW*, are related to proteins involved in the synthesis of exopolysaccharides or capsular polysaccharides in various bacteria.

This *eps* locus includes a 15.3 kb region which contains two αIS*S1* copies, one αβIS*S1* copy and two of the four IS*981* copies present in the *S. thermophilus* CNRZ368 genome. These ISs were probably acquired by horizontal transfers from *L. lactis* [7, 8]. This 15.3 kb region also contains *orfB* and *epsL* which share 97.6% identity with *orfY* and *epsL* of the plasmidic *eps* locus of *L. lactis* NIZOB40 [12]. Furthermore, most of the probes isolated from this 15.3 kb region hybridize with closely related sequences of *L. lactis* strains (Fig. 1). This suggests that a large part of this *eps* locus has been transferred from *L. lactis* to *S. thermophilus*.

The 126-nt 5' end of *epsL* of *S. thermophilus* CNRZ368 shares only 55% with *epsL* of *L. lactis* NIZOB40 whereas the other region of *epsL* is almost identical. Therefore, an illegitimate recombination between distantly related sequences could be involved in the integration of a lactococcal sequence into the *S. thermophilus* CNRZ368 chromosome. IS*981*SC could correspond to the other end of the 15.3 kb transferred region and could be involved in the integration. This transferred region probably transported four IS types and genes involved

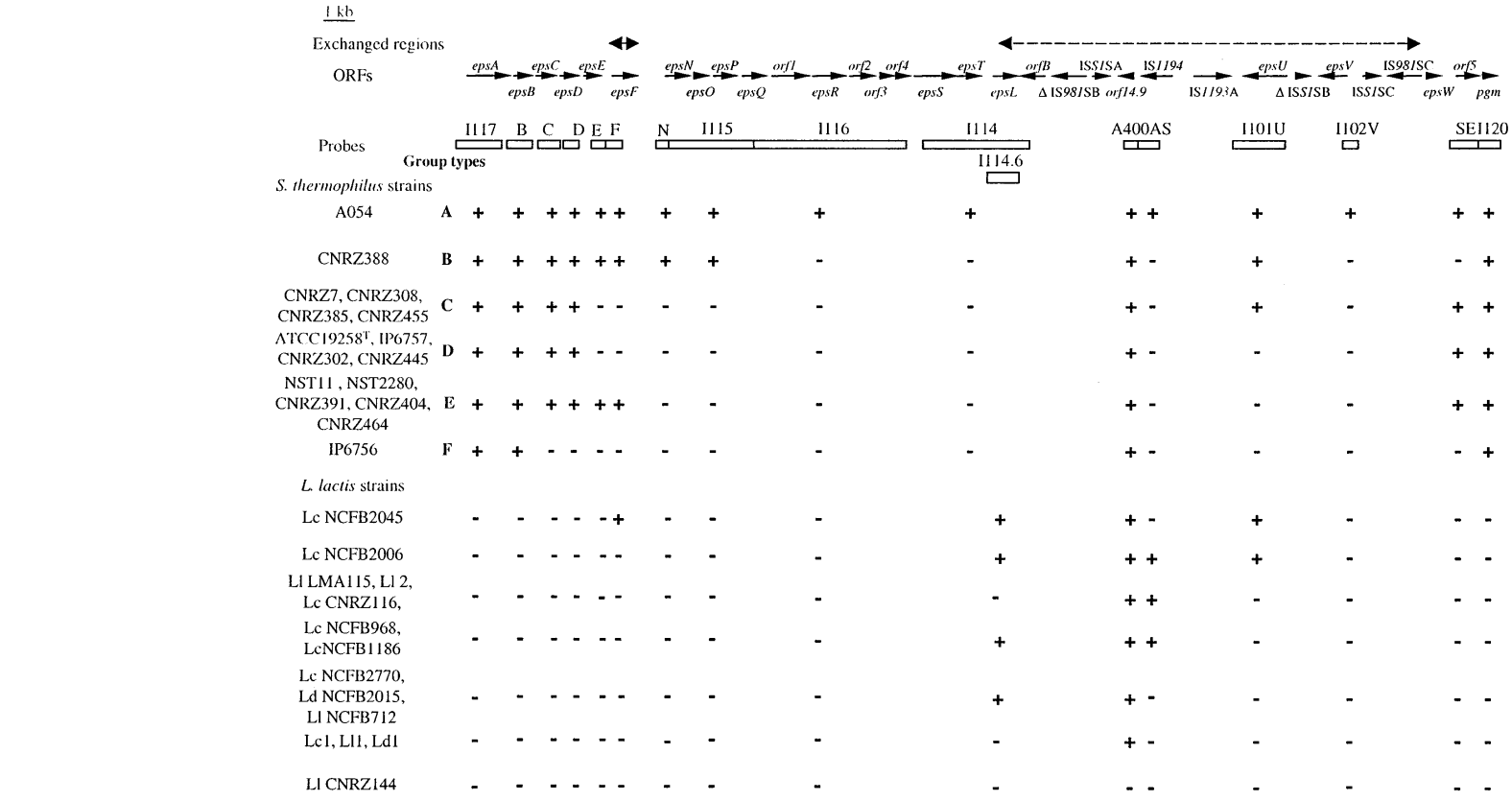

Figure 1. Map of the *eps* locus of *S. thermophilus* CNRZ368 and hybridization of probes with DNA of *S. thermophilus* or *L. lactis* strains. Arrows indicate localization and orientation of the putative genes or pseudo-genes. ΔIS corresponds to a truncated IS. All the ORFs indicated as *eps* encode putative proteins related to proteins involved in polysaccharide synthesis. The probes shown by white boxes were hybridized on *Eco*RI restriction patterns of genomic DNA from *S. thermophilus* or *L. lactis* strains.

in exopolysaccharide synthesis and was subject to rearrangements after its integration.

Furthermore, a specific probe of *epsF* isolated from the *eps* locus of *S. thermophilus* CNRZ368 hybridized with *Lactococcus* DNAs (Fig. 1), suggesting that *epsF* would also be exchanged between *S. thermophilus* and *L. lactis*.

The G + C content of the *eps* locus of *S. thermophilus* CNRZ368 is highly heterogeneous (26 to 42%) while the G + C content of the *S. thermophilus* genome is 37.2–39.8% [4]. This suggests that some regions of this locus could be acquired by horizontal transfer from different origins, perhaps from other lactic acid bacteria used in cocultures.

Hybridization of various probes from the *eps* locus of *S. thermophilus* CNRZ368 on *Eco*RI restriction patterns of 17 *S. thermophilus* strains revealed a very high polymorphism (Fig. 1). Several small regions are present in all the *S. thermophilus* strains tested while the other regions are present in only some strains or are replaced by sequences which do not hybridize to the specific probes of the CNRZ368 *eps* locus. These hybridization results also showed that the various strains tested could be arranged in six groups, named A to F, according to the presence or absence of a hybridization signal with the different probes tested (Fig. 1).

Sequence comparison of the *eps* locus from *S. thermophilus* CNRZ368, Sfi6 [10] and NCFB2393 [5] revealed almost identical regions, *epsABCDEF* and *orf14.9* in Sfi6 and *epsABCD* in NCFB2393 (Fig. 2). However, the nature and number of genes involved in exopolysaccharide synthesis present in variable regions are very different among the strains. Generally, specific genes of a strain are unrelated or very distant from genes of other strains. Sequences related to *epsE* of CNRZ368 and Sfi6 were not identified in NCFB2393 (Fig. 2). In this strain, it is replaced by a completely different gene involved in exopolysaccharide synthesis. Furthermore, most of the other regions of the *eps* locus of *S. thermophilus* CNRZ368 are replaced by unrelated sequences in Sfi6 (Fig. 2), suggesting that these sequences could have been acquired by horizontal transfer(s).

These results suggest that *eps* loci of *S. thermophilus* strains have undergone numerous rearrangements leading to chimeric loci. Some of the rearrangement points could be precisely located: 5 nt in the 3' end of *epsD*, 26 nt from the 3' end of *epsF* and 8 nt in the 3' end of *orf14.9* (Fig. 2).

3. COMPARISON OF *EPS* LOCI FROM *S. THERMOPHILUS* NST2280 AND IP6757 STRAINS

Phylogenetic relationships between the various *S. thermophilus* strains tested have been determined by hybridization of 12 specific probes of other variable regions, IS and genes encoding rRNA (Fig. 3). Hybridization patterns of the various strains showed a high polymorphism. However, hybridization patterns of NST2280 and IP6757 are almost identical, indicating that these strains are very closely related. However, the ropy strain NST2280 possesses a type E *eps* locus while the non-ropy strain IP6757 possesses a type D *eps* locus (Fig. 1). The phylogenetic analysis suggests that a common ancestor of a group of six strains, including NST2280 and IP6757, possessed a type E *eps* locus. This suggests that an IP6757 ancestor acquired a new *eps* locus type by sequence replacement following horizontal transfer while the other strains kept the ancestral *eps* locus.

The restriction map of the NST2280 *eps* locus is identical to that of Sfi6 deduced from its sequence [10] (data not shown). Furthermore, two PCR products were obtained in NST2280 by using primers located in the specific sequence of Sfi6 *eps* locus (in *epsH* and *epsI*, and in *epsL* and *epsM*). These products have the same size as in Sfi6 (data not shown). Therefore, the

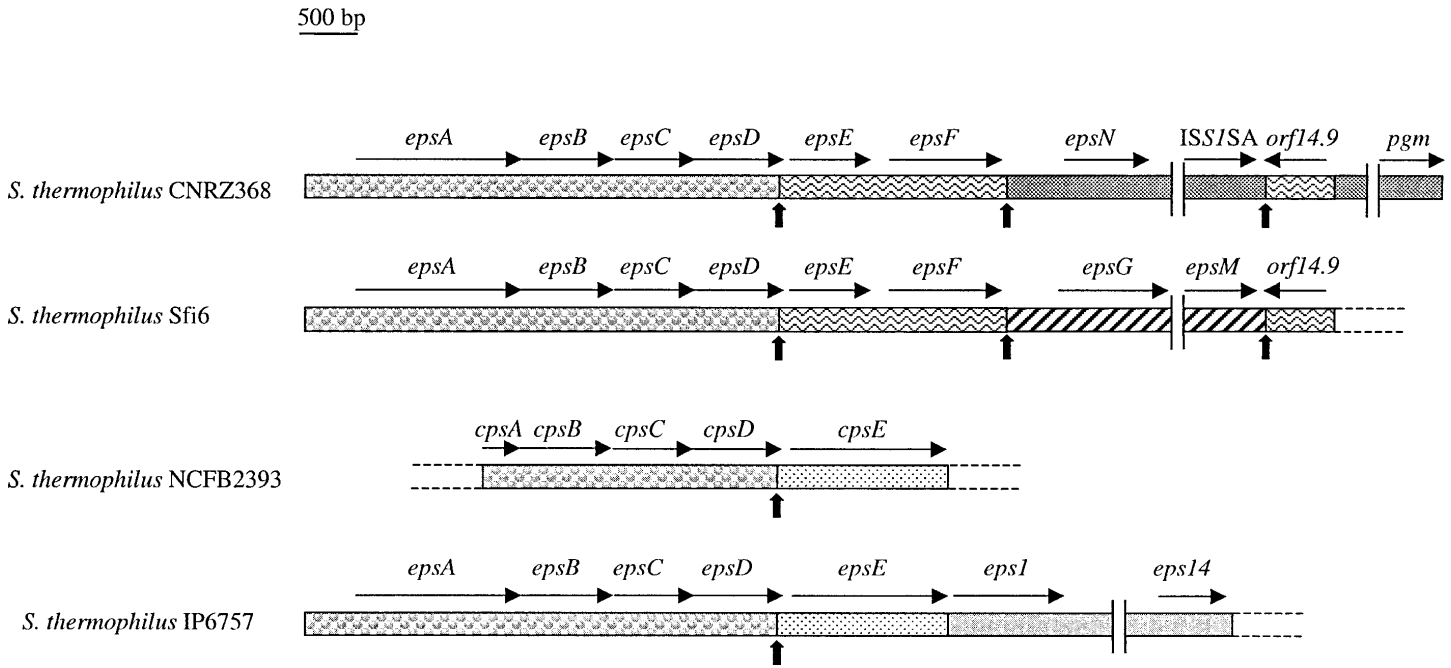

Figure 2. Comparison of the *eps* loci from *S. thermophilus* strains. Closely related nucleotide sequences (more than 90% identity) are shown by the same symbols. Black arrows indicate localization of rearrangement points. Partially sequenced genes or loci are shown by broken lines.

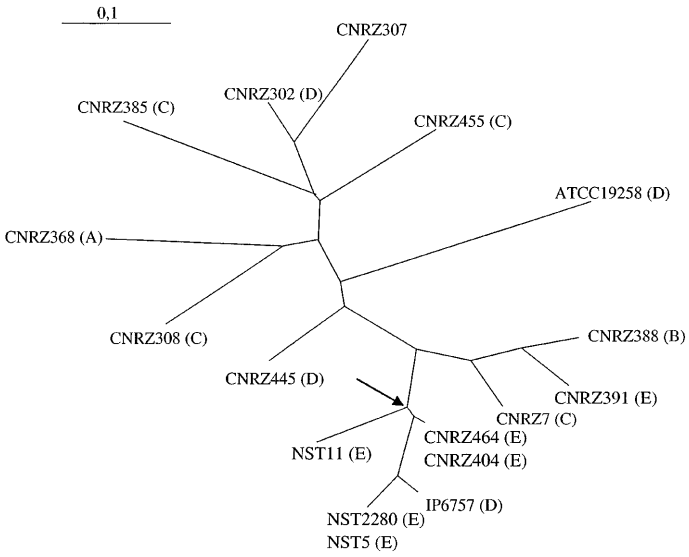

Figure 3. Dendrogram of *S. thermophilus* strains. 12 specific probes of *rrn* loci, IS*1191*, IS*1193* and variable loci were hybridized on DNAs digested by various restriction endonucleases. These hybridization patterns were used to produce a dendrogram by the neighbor-joining method using PAUP program [11]. Arrow indicates the common ancestor of the group of six strains which possess a type E *eps* locus except IP6757.

NST2280 *eps* locus is identical or very closely related to the *eps* locus of the ropy strain Sfi6.

The sequenced *epsABCD* region of the IP6757 *eps* locus is identical to that of Sfi6 (Fig. 2). This region is followed by *epsE* which shares 95% identity with *cpsE* of NCFB2393 and by a region containing 14 ORFs or pseudo-ORFs, named *eps1* to *eps14*. These ORFs are unrelated to those of CNRZ368, Sfi6 and therefore of NST2280. However, the putative products of these ORFs share significant similarities with proteins involved in polysaccharide synthesis of various bacteria. This sequence comparison also showed a rearrangement point at 5nt in the 3' end of *epsD* (Fig. 2).

Furthermore, hybridization of various probes from the *eps* locus of *S. thermophilus* IP6757 on *Eco*RI restriction patterns of 17 *S. thermophilus* strains revealed a small constant region, the same as identified by hybridization of probes isolated from the CNRZ368 *eps* locus, and a large variable region which hybridizes only with IP6757 DNA (Fig. 4). Sequences related to this region are not present in NST2280 nor in the other strains of the D group previously defined. This D group should be divided into at least two groups, D (IP6757) and D' (ATCC19258, CNRZ302 et CNRZ445).

The G + C content of the *eps* locus of *S. thermophilus* IP6757 is highly variable (20 to 43%). Particularly, the G + C content of the large *eps2-eps14* region which is detected only in the IP6757 strain is 30% whereas the G + C content of the *S. thermophilus* genome is 37.2–39.8% [4]. This suggests that this region could be acquired by horizontal transfer from different origins, perhaps from other lactic acid bacteria used in cocultures.

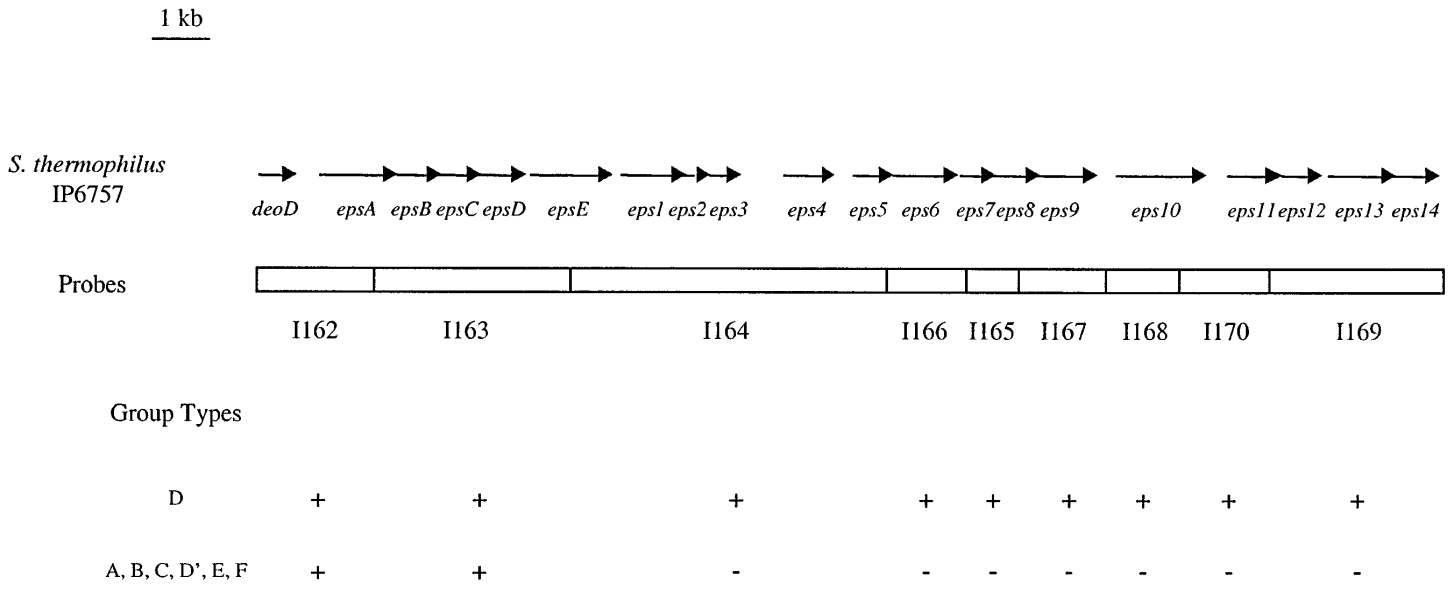

Figure 4. Map of the partially sequenced *eps* locus of *S. thermophilus* IP6757 and hybridization of probes with DNA of *S. thermophilus* strains. Arrows indicate localization and orientation of the putative genes or pseudo-genes. All the ORFs indicated as *eps* encode putative proteins related to proteins involved in polysaccharide synthesis. The probes shown by white boxes were hybridized on *Eco*RI restriction patterns of genomic DNA from *S. thermophilus* strains. A: CNRZ368, A054; B: CNRZ388; C: CNRZ7, CNRZ308, CNRZ385, CNRZ455; D: IP6757; D': ATCC19258, CNRZ302, CNRZ445; E: NST2280, NST11, CNRZ391, CNRZ404, CNRZ464; F: IP6756.

These results show that the *eps* locus of *S. thermophilus* IP6757 contains several regions from different origins. The *epsABCD* region was inherited from its last common ancestor with NST2280 while the other regions were acquired by horizontal transfer(s).

4. CONCLUSION

These results indicate that *eps* loci of *S. thermophilus* have a chimeric structure resulting from recombination between unrelated or distantly related sequences acquired by spontaneous intra- and interspecific transfers. One of the mechanisms involved in these horizontal transfers could be conjugation since an integrative, potentially conjugative element (ICE) was found in *S. thermophilus* [3] and numerous conjugative elements were identified in *L. lactis*.

REFERENCES

[1] Bourgoin F., Guédon G., Pébay M., Roussel Y., Panis C., Decaris B., Characterization of a mosaic IS*S1* element and evidence for the recent horizontal transfer of two different types of IS*S1* between *Streptococcus thermophilus* and *Lactococcus lactis*, Gene 178 (1996) 15–23.

[2] Bourgoin F., Guédon G., Gintz B., Decaris B., Characterization of a novel insertion sequence, IS*1194*, in *Streptococcus thermophilus*, Plasmid 40 (1998) 44–49.

[3] Burrus V., Roussel Y., Decaris B., Guédon G., Characterization of a novel integrative element, ICE*St*1, in the lactic acid bacteria *Streptococcus thermophilus*, Appl. Environ. Microbiol. 66 (2000) 1749–1753.

[4] Farrow J.A.E., Collins M.D., DNA base composition, DNA-DNA homology and long chain fatty acid studies on *Streptococcus thermophilus* and *Streptococcus salivarius*, J. Gen. Microbiol. 130 (1984) 357–362.

[5] Griffin A.M., Morris V.J., Gasson M.J., The *cps ABCDE* genes involved in polysaccharide production in *Streptococcus salivarius* ssp. *thermophilus* NCFB2393, Gene 183 (1996) 23–27.

[6] Guédon G., Bourgoin F., Pébay M., Roussel Y., Colmin C., Simonet J.M., Decaris B., Characterization and distribution of two insertion sequences, IS*1191* and iso-IS*981*, in *Streptococcus thermophilus*: does intergeneric transfer of ISs occur in lactic acid bacteria co-cultures?, Mol. Microbiol. 16 (1995) 69–78.

[7] Guédon G., Bourgoin F., Decaris B., Does gene horizontal transfer occur in lactic acid bacteria co-cultures?, Lait 78 (1998) 53–58.

[8] Guédon G., Bourgoin F., Burrus V., Pluvinet A., Decaris B., Implication of horizontal transfers in genetic polymorphism of lactic acid bacteria, Sci. Aliments 20 (2000) 85–95.

[9] O'Sullivan T., van Sinderen D., Fitzgerald G., Structural and functional analysis of pCI65st, a 6.5 kb plasmid from *Streptococcus thermophilus* NDI-6, Microbiology 145 (1999) 127–134.

[10] Stingele F., Neeser J.R., Mollet B., Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6, J. Bacteriol. 178 (1996) 1680–1690.

[11] Swofford D.L., PAUP: Phylogenetic analysis using parsimony, version 3.0s., Illinois Natural History Survey, Champaign, Illinois (1991).

[12] van Kranenburg R., Marugg J.D., van Swam I.I., Willem N.J., de Vos W.M., Molecular characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*, Mol. Microbiol. 24 (1997) 387–397.

[13] Vaughan E.E., David S., Harrington A., Daly C., Fitzgerald G.F., de Vos W.M., Characterization of a plasmid-encoded citrate permease (*citP*) gene from *Leuconostoc* species reveals high sequence conservation with the *Lactococcus lactis citP* gene, Appl. Environ. Microbiol. 61 (1995) 3172–3176.