

Genome, genomics

Characterization and chimeric structure of a family of integrative and potentially conjugative elements from *Streptococcus thermophilus*

Vincent BURRUS, Guillaume PAVLOVIC, Bernard DECARIS,
G rard GU DON*

Laboratoire de G n tique et Microbiologie UA INRA 952, Universit  Henri Poincar  Nancy 1,
1 boulevard des Aiguillettes, BP 239, 54506 Vand uvre-l s-Nancy, France

Abstract — A 34.7-kb element, *ICESt1*, is integrated in the 3' end of *fda* locus from *Streptococcus thermophilus* CNRZ368. *ICESt1* excises by a site-specific recombination between two 27-pb identical sequences flanking the element. It encodes an integrase required for excision. Furthermore, eleven putative proteins encoded by *ICESt1* are related to proteins encoded by various conjugative elements from low G + C Gram positive bacteria. Therefore, *ICESt1* could be a site-specific integrative conjugative element (ICE). Comparison of proteins encoded by *ICESt1* and the sequenced genome of *Bacillus subtilis* 168 revealed a putative 20.5-kb ICE, *ICEBs1*. Sequence comparison of *ICESt1*, *ICEBs1*, Tn916 and Tn5252 revealed exchanges of modules between ICEs, conjugative transposons and prophages. Four types of elements related to *ICESt1* (IEs) were found in seven other strains of *S. thermophilus* and are integrated in the same location as *ICESt1*. One of these elements, IE385, could be an ICE whereas the others do not seem to be integrative and conjugative. Comparison of the various elements and *ICESt1* showed that all of them have a chimerical structure resulting from exchanges of regions from different origins. The left end of IE19258 is identical to an internal recombination site of *ICESt1*, *attL'*, but shares only 57% identity with its left end, *attL*. The site-specific recombination between the cores of *attL'* and of the right end, *attR*, leads to the excision of a circular molecule corresponding to the region flanked by these sites. Therefore, this suggests that *ICESt1* results from the integration of a 28.2-kb ICE, *ICESt2*, in the *attR'* site of an IE element and that *ICESt2* have mobilized the IE.

site-specific recombination / conjugation / chimerical sequence / horizontal transfer / *Streptococcus thermophilus*

R sum  — Caract risation et structure chim rique d'une famille d' l ments int gratifs potentiellement conjugatifs chez *Streptococcus thermophilus*. Un  l ment de 34,7 kb, *ICESt1*, est int gr  dans l'extr mit  3' du locus *fda* de *Streptococcus thermophilus* CNRZ368. *ICESt1* s'excise

* Correspondence and reprints
Tel.: (33) 3 83 91 20 00; fax: (33) 3 83 91 25 00; e-mail: guedon@nancy.inra.fr

par recombinaison site-spécifique entre des séquences identiques de 27 pb flanquant l'élément. *ICESt1* code une intégrase nécessaire à cette excision. Onze des ORF d'*ICESt1* codent des protéines apparentées à celles codées par divers éléments conjugatifs de bactéries Gram positives à bas G + C. *ICESt1* serait donc un élément conjugatif à intégration site-spécifique (ICE). La comparaison des protéines codées par *ICESt1* et le génome séquencé de *Bacillus subtilis* 168 a révélé un ICE de 20,5 kb, *ICEBs1*. L'analyse des séquences d'*ICESt1*, *ICEBs1*, Tn916 et Tn5252 révèle des échanges de modules entre éléments conjugatifs intégratifs, transposons conjugatifs et prophages. Par ailleurs, 4 types d'éléments apparentés à *ICESt1* d'une taille de 12,8 à 26,1 kb (IE) sont intégrés exactement au même site qu'*ICESt1* chez 7 autres souches de *S. thermophilus*. Seul, IE385 serait un ICE. Les autres éléments ne semblent ni conjugatifs ni intégratifs. Les hybridations entre éléments et le séquençage partiel montrent que chacun des éléments possède une structure chimérique complexe associant des régions d'origines différentes. L'extrémité gauche d'IE19258 est identique à un site de recombinaison interne d'*ICESt1*, *attL'*, mais ne présente que 57 % d'identité avec son extrémité gauche. La recombinaison site spécifique entre *attL'* et l'extrémité droite *attR* d'*ICESt1* conduit à l'excision d'une forme circulaire de la région comprise entre les 2 sites. Ceci indique qu'*ICESt1* serait constitué de 2 éléments : un IE compris entre *attL* et *attL'* et un élément conjugatif intégratif de 28,2 kb compris entre *attL'* et *attR*, *ICESt2*, qui se serait intégré à la frontière de l'IE et l'aurait mobilisé.

recombinaison spécifique de site / conjugaison / séquence chimérique / transfert horizontal / *Streptococcus thermophilus*

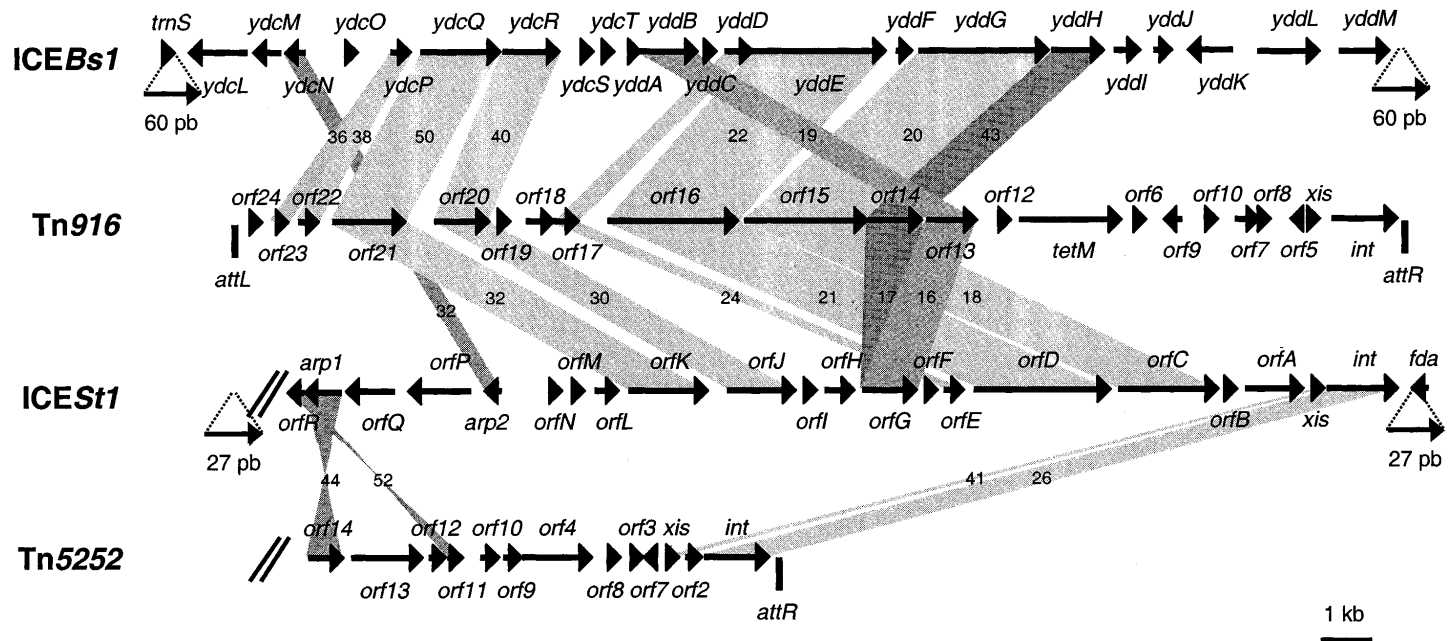
1. INTRODUCTION

Genome analysis of a large array of bacterial species revealed a high intraspecific polymorphism which largely corresponds to the presence of variable regions in some strains and their absence in other strains. At least some of these variable regions were acquired by interspecific and intraspecific horizontal transfers. Numerous variable regions were found by comparison of the chromosomal map of three strains of the lactic acid bacteria *Streptococcus thermophilus* [13]. Sequence comparisons and hybridizations revealed that multiple horizontal transfers have recently occurred between *S. thermophilus* and *Lactococcus lactis*, probably in cocultures used during cheese manufacture [2, 3, 8–10]. However, the mechanism of the genetic exchanges between the two species remains unknown, and, until recently, no conjugative element was reported in *S. thermophilus*.

2. CHARACTERIZATION OF AN INTEGRATIVE POTENTIALLY CONJUGATIVE ELEMENT, *ICEST1*

2.1. Characterization of a site-specific integrative element

Comparison of their maps and probe hybridizations showed that the chromosome of the strain CNRZ368 contains a 34.7-kb region, *var1C*, which is entirely absent in the closely related strain A054 and the distantly related strain NST2280 [5, 13]. Sequence comparisons of the *var1C* ends revealed the presence of a 27-bp direct repeat (Fig. 1) whereas only one copy of this sequence was found in corresponding regions of A054. The right 27-pb sequence of *var1C* and the unique 27-bp sequence of A054 include the 3' end of *fda* which encodes a putative fructose-1,6-diphosphate aldolase. Comparisons also showed that these 27-bp sequences correspond to the limits of the *var1C* region in CNRZ368 (*attL* and *attR*) whereas the unique 27-bp



Integrative conjugative elements of *S. thermophilus*

Figure 1. Comparison of gene organization of ICESt1, Tn5252, Tn916 and ICEBs1. The ORFs and their orientation are indicated by the arrows. Grey areas join ORFs that code for related proteins; however, only the relationships between integrases belonging to the same subgroup are indicated. The numbers correspond to the identity percent shared by these related proteins. The complete sequences of ICEBs1 and Tn916 and the right regions of ICESt1 are shown. The regions of Tn5252 involved in conjugative transfer were not sequenced.

sequence could be an integration site in A054 (*attB*). Two ORFs were found in the right region of *var1C* (Fig. 1) [5]. *xis* codes for a protein which shares 37–41% identity with the putative excisionases of the conjugative transposons Tn5252 from *S. pneumoniae* [14] and Tn5276 from *L. lactis* [12]. *int* codes for a site-specific integrase related

to those of Tn5252 [14], Tn5276 [12] and numerous temperate phages of lactic acid bacteria.

This data suggested that the 27-bp sequence would be the core sequence of a site-specific recombination system involving the proteins encoded by *xis* and *int*. The recombination between the cores of *attL* and

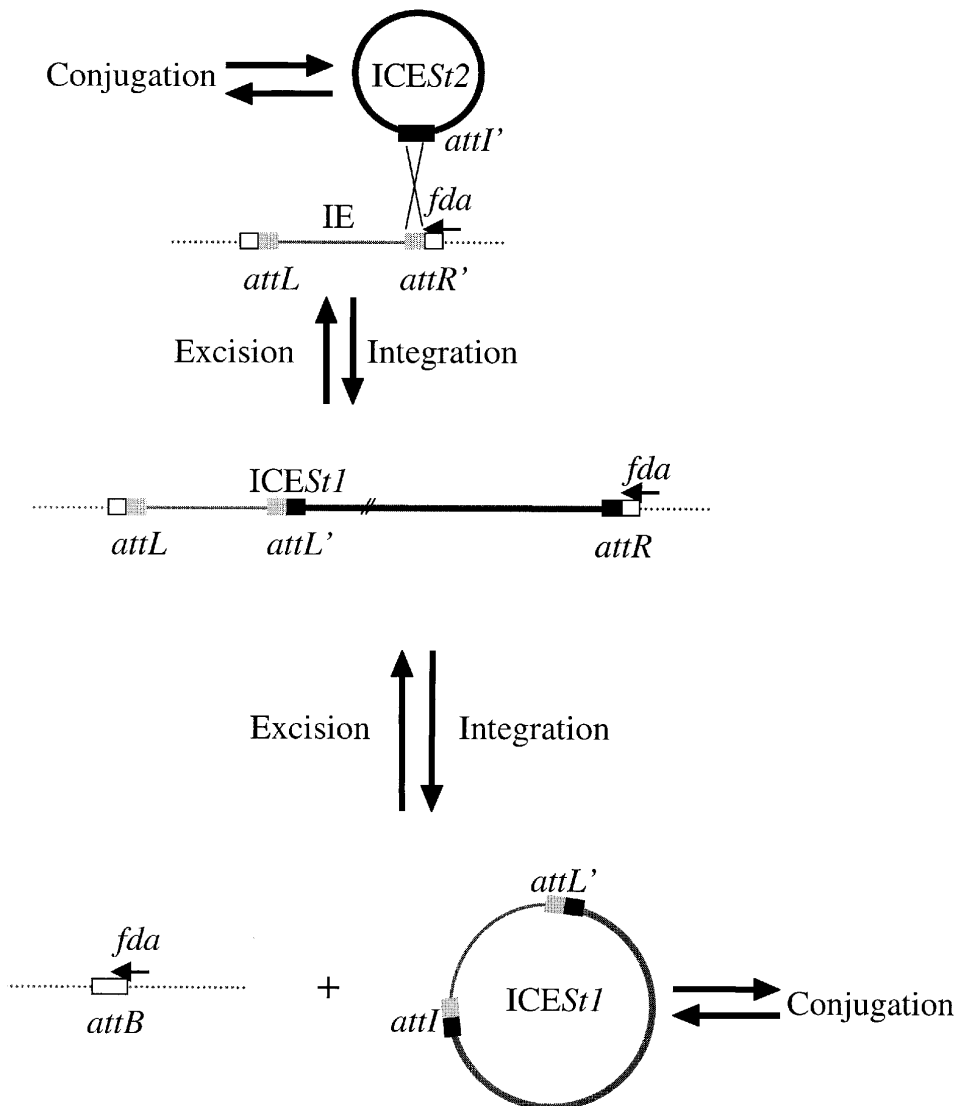


Figure 2. Model of excision and formation of ICES1 (*var1C*).

attR would generate a chromosomal integration site *attB*, identical to the site of A054, and a recombination site *attI* carried by a circular molecule corresponding to an excised form of *varIC* (Fig. 2). A fragment containing an *attB* site was amplified by PCR from CNRZ368 and A054 DNAs. A junction fragment containing an *attI* site was amplified by PCR from CNRZ368 DNA but not from A054 DNA. The sequences of these PCR products were found to be identical to those that would be obtained by recombination between the 27-bp sequences of *attL* and *attR*. Furthermore, PCR products carrying *attB* or *attI* were not obtained when the *int* gene of *varIC* was disrupted by the integration of a thermosensitive plasmid [5]. Therefore, the variable region *varIC* of CNRZ368 is a site-specific integrative element.

2.2. The integrative element encodes conjugative functions

Twenty-one putative proteins are encoded by the right 18 kb of *varIC*. Eleven are related to proteins encoded by various conjugative elements from low G + C Gram positive bacteria. *orfA* (Fig. 1) codes for a protein related to transfer proteins encoded by two staphylococcal conjugative plasmids, pSK41 and pG01 (47% identity) [5]. The putative products of 6 ORF (*orfC*, *orfD*, *orfE*, *orfG*, *orfJ* and *orfK*) are related to transfer proteins encoded by the conjugative transposon Tn916 (Fig. 1) [7]. Furthermore the integrase, the excisionase and two other proteins encoded by *varIC* (*Arp1* and *OrfR*) are related to proteins encoded by the conjugative transposon Tn5252 (Fig. 1) [14]. Therefore, *varIC* encodes an integrative system and probably a conjugative system, the first found in *S. thermophilus*. This suggests that *varIC* is a novel integrative and conjugative element (ICE), called ICE*St1*, which integrates in the 3' end of a gene encoding an aldolase. ICE*St1* excises by site-specific recombination between the cores of its *attL* and *attR* sites (Fig. 2); then,

the excised element could be transferred by conjugation and could integrate by site-specific recombination in the recipient cell.

3. THE DIFFERENT REGIONS OF ICE*St1* HAVE SEPARATE ORIGINS

Eight proteins encoded by the 45°–47° region of the completely sequenced genome of *Bacillus subtilis* 168 [11] are related to putative transfer proteins from ICE*St1* and Tn916 and another protein is related to the putative regulatory protein Arp2 from ICE*St1* (Fig. 1). Therefore, this region probably contains a conjugative element. Furthermore, a 20511-pb region encoding these proteins also codes for the putative integrase YdcL and is flanked by a 60-bp direct repeat. One of these 60-bp sequences corresponds to the 3' part of *trnS* which encodes a tRNA^{Leu}. Therefore, the 20511-pb region is probably another ICE that we have named ICE*Bs1*.

The module grouping the putative transfer genes have similar organization in ICE*St1*, Tn916 and ICE*Bs1* (Fig. 1) suggesting that the transfer modules of the three elements have a common ancestor. The integrases of the three elements belong to different subgroups of the integrase family. The ICE*St1* integrase belongs to the ϕ LC3 group which includes those of Tn5252, Tn5276 and numerous bacteriophages of lactic acid bacteria (<http://members.home.net/domespo/trhome.html>). However, the integrase of ICE*Bs1* belongs to the ϕ 11 group which also includes those of two bacteriophages. Furthermore, the integrase of Tn916-like elements constitute a third group, the Tn916 group. This comparison shows that exchange of modules (integration or transfer) have occurred between ICEs, conjugative transposons and phages.

The G + C percent of ICE*St1* is highly variable. The G + C percent of the integration

module is 34.2% whereas that of the putative transfer module is 42.2%. Furthermore, G + C percent of the other ORFs (excluding insertion sequences IS1191 and IS981) varies from 25.8 to 35.9%. The G + C percent of the various regions of ICES*t1* is different from the mean G + C percent of *S. thermophilus*, i.e. 37.2–39.8% [6]. Therefore, the various regions or modules of ICES*t1* probably have separate origins.

Furthermore, an internal sequence of ICES*t1*, *attL'*, shares 61.3% identity with *attL*. The recombination between the core sequence of *attL'* and *attR* generates a chromosomal recombination site *attR'* and a recombination site carried by a circular molecule, *attI'* (Fig. 2). *attI'* was detected by PCR in CNRZ368 DNA. PCR product carrying *attI'* was not obtained when the integrase gene was disrupted. Therefore, the 28.2-kb region included between the cores of *attL'* and *attR* could be another ICE, ICES*t2* whose origin could be separate from that of *attL-attL'* region.

4. IDENTIFICATION AND CHIMERIC STRUCTURE OF ELEMENTS RELATED TO ICES*t1*

Elements related to ICES*t1* and/or integrated in the same location were searched in 22 strains of *S. thermophilus* by hybridization with probes corresponding to the various regions of ICES*t1* and by Long Range PCR amplifications of *fda* locus. Elements (12.8–26.1 kb) were found in seven strains (Fig. 3). All are integrated exactly in the same location as ICES*t1*. Each of the four element types (IEs) possesses sequences (1.5–20 kb) hybridizing to at least one of the other types of element. Sequences related to *attL* and *attR* of ICES*t1* were found in all elements except in IE308 which does not seem to possess an *attL* site. Sequences hybridizing to all regions involved in integration and putative transfer of ICES*t1* were identified in IE385. Furthermore, IE385

excision was detected by PCR. Therefore, IE385 could be an integrative conjugative element which was called ICES*t3*. Most of the sequences involved in integration and transfer were not found in the other elements. Therefore, they are probably neither integrative nor conjugative.

Each type of element possesses sequences (8–15 kb) which do not hybridize to the others at high stringency (Fig. 3). Partial sequences of some of these regions are distantly related to corresponding sequences of ICES*t1* (15–44% divergence) whereas most are not related to ICES*t1* sequences. In this way, almost all sequences of IE19258 (except the insertion sequence IS1193) do not hybridize to the other elements. The *attR* sequences of ICES*t1* and IE19258 share 67% identity and their *attL* sequences 56%. However, the *attL* sequence of IE19258 is identical to *attL'* of ICES*t1*. Furthermore, the sequences located to the right of IE19258 *attL* are not related to ICES*t1* sequences. However, they share 64% identity with the end of Tn916 encoding the integrase, suggesting that a Tn916-type element is integrated in IE19258. Therefore, each type of element (including ICES*t1*) has a complex chimerical structure including regions which have separate origins and were exchanged by horizontal transfers.

5. MODEL OF ICES*t1* FORMATION

Structure comparison of the various elements suggests a multistep model for ICES*t1* formation (Fig. 2). (i) In the first step, an IE is integrated in the 3' end of the *fda* locus. It could have been generated by deletions of sequences encoding the integration and transfer of an ICE. (ii) In the second step, ICES*t2* transfers by conjugation to the strain possessing the IE and integrates by site-specific recombination in the *attR'* site of the IE. (iii) In the third step, a site-specific recombination between *attL* and *attR* generates the circular form of ICES*t1* which could be transferred to other strains and

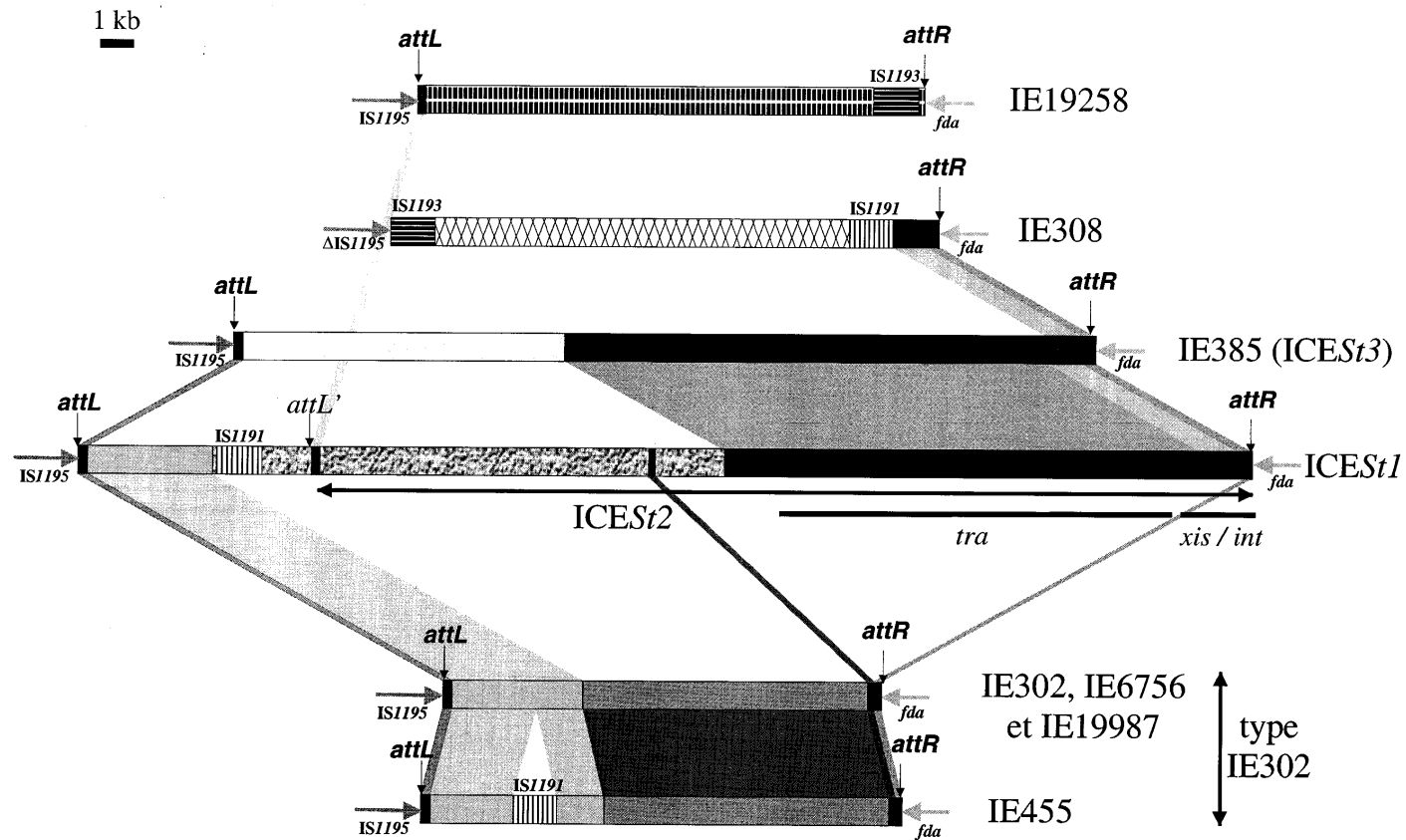


Figure 3. Comparison of the elements integrated in *fda* locus of *S. thermophilus*. Grey areas join unique sequences which hybridized at high stringency and/or are almost identical. Arrows indicate the location and orientation of the ORF *fda* and of the IS1195 copy which flank all the elements. The location of ICESt2 and of the sequences involved in the putative transfer (*tra*) or excision/integration (*xis/int*) of ICESt1 are indicated.

could integrate in the 3' end of *fda* of the recipient strain. In this model, ICE*St1* results from the mobilization of an IE element by ICE*St2*. Such mechanism of mobilization has never been described in other conjugative elements (ICEs, transposons or plasmids).

6. ARE ICE AND IE INVOLVED IN INTERSPECIFIC HORIZONTAL TRANSFERS?

Multiple horizontal transfers of various insertion sequences have recently occurred from *S. thermophilus* to *L. lactis* and from *L. lactis* to *S. thermophilus* [2, 3, 8–10]. However, their mechanisms are unknown. Copies of two transferred insertion sequences, IS1191 (Fig. 3) and IS981 (data not shown), were found in ICE*St1*, IE308, IE302 and IE455. In the same way, partial sequencing of three regions of IE308 reveals sequences almost identical to sequences of various plasmids of *L. lactis*. Furthermore the conjugative transposons related to ICE*St1*, i.e. Tn916, Tn5252 and Tn5307, a Tn5276-type element, can transfer between distantly related species (including transfer of Tn916 between lactic acid bacteria and enterobacteria) [1, 4, 14]. Therefore, some ICEs or IEs could be involved in interspecific horizontal transfers.

REFERENCES

- [1] Bertram J., Stratz M., Durre P., Natural transfer of conjugative transposon Tn916 between Gram-positive and Gram-negative bacteria, *J. Bacteriol.* 173 (1991) 443–448.
- [2] Bourgoin F., Guédon G., Pébay M., Roussel Y., Panis C., Decaris B., Characterization of a mosaic ISSI element and evidence for the recent horizontal transfer of two different types of ISSI between *Streptococcus thermophilus* and *Lactococcus lactis*, *Gene* 178 (1996) 15–23.
- [3] Bourgoin F., Pluvinet A., Gintz B., Decaris B., Guédon G., Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci?, *Gene* 233 (1999) 151–161.
- [4] Broadbent J.R., Sandine W.E., Kondo J.K., Characteristics of Tn5307 exchange and intergeneric transfer of genes associated with nisin production, *Appl. Microbiol. Biotechnol.* 44 (1995) 139–146.
- [5] Burrus V., Roussel Y., Decaris B., Guédon G., Characterization of a novel integrative element, ICE*St1*, in the lactic acid bacteria *Streptococcus thermophilus*, *Appl. Environ. Microbiol.* 66 (2000) 1749–1753.
- [6] 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.
- [7] Flannagan S.E., Zitzow L.A., Su Y.A., Clewell D.B., Nucleotide sequence of the 18-kb conjugative transposon Tn916 from *Enterococcus faecalis*, *Plasmid* 32 (1994) 350–354.
- [8] Guédon G., Bourgoin F., Decaris B., Does gene horizontal transfer occur in lactic acid bacteria co-cultures?, *Lait* 78 (1998) 53–58.
- [9] Guédon G., Bourgoin F., Pébay M., Roussel Y., Colmin C., Simonet J.M., Decaris B., Characterization and distribution of two insertion sequences, IS1191 and iso-IS981, in *Streptococcus thermophilus*: does intergeneric transfer of insertion sequences occur in lactic acid bacteria co-cultures?, *Mol. Microbiol.* 16 (1995) 69–78.
- [10] 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.
- [11] Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Danchin A., et al. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, *Nature* 390 (1997) 249–256.
- [12] Rauch J.G., de Vos W.M., Identification and characterization of genes involved in excision of the *Lactococcus lactis* conjugative transposon Tn5276, *J. Bacteriol.* 176 (1994) 2165–2171.
- [13] Roussel Y., Bourgoin F., Guédon G., Pébay M., Decaris B., Analysis of the genetic polymorphism between three *Streptococcus thermophilus* strains by comparing their physical and genetic organization, *Microbiology* 143 (1997) 1335–1343.
- [14] Vijayakumar M.N., Ayalew S., Nucleotide sequence analysis of the termini and chromosomal locus involved in site-specific integration of the streptococcal conjugative transposon Tn5252, *J. Bacteriol.* 175 (1993) 2713–2719.