Characterization and evolution of a family of integrative and potentially conjugative or mobilizable elements from Streptococcus thermophilus

Guillaume Pavlovic, Vincent Burrus, Alexandre Toulmay, Frédéric Choulet, Bernard Decaris, Gérard Guédon

Lait 84 (2004) 7–14
© INRA, EDP Sciences, 2003
DOI: 10.1051/lait:2003042

Review

Characterization and evolution of a family of integrative and potentially conjugative or mobilizable elements from *Streptococcus thermophilus*

Guillaume PAVLOVIC, Vincent BURRUS, Alexandre TOULMAY, Frédéric CHOULET, Bernard DECARIS, Gérard GUÉDON*

Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR110, Université Henri Poincaré - Nancy 1, Bd des Aiguillettes, BP 239, 54506 Vandœuvre-lès-Nancy, France

Published online 12 December 2003

Abstract – The integrative and conjugative elements (ICEs) excise by site-specific recombination, self-transfer the resulting circular form by conjugation and integrate into the genome of the recipient bacterium. The 34.7-kb element from *Streptococcus thermophilus* CNRZ368, ICE*St1*, excises and integrates by site-specific recombination. This element also possesses a conjugation module distantly related to that of the conjugative transposon Tn*916* from *Enterococcus faecalis*. Therefore, ICE*St1* is probably an ICE. Four types of elements related to ICE*St1* are integrated into the same location as ICE*St1* in seven other strains of *S. thermophilus*. One of these elements, ICE*St3*, is probably an ICE whereas the three others (CIMEs) would have arisen from ICEs by deletion of the conjugation and recombination modules. These elements also encode functions that are not involved in element maintenance or transfer, such as restriction-modification systems. Each of these elements has a chimerical structure resulting from the acquisition of modules from different origins. Sequence analyses indicate that these elements are involved in horizontal transfers with various species of dairy or pathogenic lactic acid bacteria. ΔCIME308 has exchanged restriction-modification and cadmium resistance modules with plasmids. CIME19258 has acquired a cadmium resistance module by the integration of an ICE related to Tn*916* within the CIME. The site-specific recombination between an internal *attL*-related site and *attR* of ICE*St1* leads to the excision of a circular molecule which could be another ICE, ICE*St2*, suggesting that ICE*St1* has arisen by accretion of a CIME and ICE*St2*. CIME302, ICE*St2* and ICE*St3* would also have arisen by site-specific accretion of CIMEs and ICEs and/or mobilization of CIMEs by ICEs. An ICE would integrate by site-specific recombination in the *attR* site of a CIME; then the CIME-ICE would excise by site-specific recombination and transfer by conjugation.

Site-specific recombination / conjugative transposon / modular evolution / horizontal transfer / genomic island

Résumé – **Structure et évolution d'une famille d'éléments intégratifs potentiellement conjugatifs ou mobilisables de *Streptococcus thermophilus*.** Les éléments intégratifs conjugatifs (ICEs) s'excisent par recombinaison site-spécifique sous forme circulaire, se transfèrent par conjugaison puis s'intègrent dans la bactérie réceptrice. L'élément ICE*St1* de *Streptococcus thermophilus* CNRZ368 (34,7 kb) s'intègre et s'excise par recombinaison site-spécifique. Par ailleurs, il code un

* Corresponding author: guedon@nancy.inra.fr

système de conjugaison putatif apparenté de façon lointaine à celui de Tn*916* d'*Enterococcus faecalis* et constitue donc un ICE putatif. Quatre éléments apparentés à ICE*St1* sont intégrés au même site que lui chez sept souches de *S. thermophilus*. L'un des éléments, ICE*St3* est probablement un ICE. Les trois autres (CIMEs) dériveraient d'un ICE par délétion des modules de conjugaison et de recombinaison. Ces éléments codent des fonctions non impliquées dans le transfert comme des systèmes de restriction-modification. Chacun d'entre eux présente une structure chimérique résultant de l'acquisition de modules d'origines différentes. L'analyse de leur séquence indique qu'ils sont impliqués dans des transferts horizontaux entre *S. thermophilus* et diverses espèces de bactéries lactiques alimentaires ou pathogènes. ΔCIME308 a échangé des modules de restriction-modification et de résistance aux ions Cd^{++} avec des plasmides. CIME19258 a acquis un module de résistance aux ions Cd^{++} par intégration d'un ICE apparenté à Tn*916* dans un CIME. La recombinaison site-spécifique entre un site interne d'ICE*St1* de type *attL* et le site *attR* provoque l'excision de la partie droite d'ICE*St1* qui constituerait elle-même un ICE, ICE*St2*. ICE*St1* serait ainsi apparu par accrétion d'un CIME et d'ICE*St2*. CIME302, ICE*St2* et ICE*St3* seraient également apparus par accrétion site-spécifique d'ICEs et de CIMEs et/ou mobilisation de CIMEs par des ICEs. Un ICE s'intégrerait par recombinaison site-spécifique dans le site *attR* d'un CIME formant ainsi un élément composite (CIME-ICE) par accrétion. Celui-ci s'exciserait par recombinaison site-spécifique puis se transférerait par conjugaison.

Recombinaison site-spécifique / transposon conjugatif / évolution modulaire / transfert horizontal / îlot génomique

1. INTRODUCTION

Horizontal gene transfer between strains and species constitutes a major driving force in the evolution of bacteria [6, 9]. In this way, pathogenic bacteria have acquired antibiotic resistance genes by horizontal transfer of various plasmids or transposons and the acquisition of pathogenicity islands, prophages and/or plasmids has turned non-pathogenic strains from various species into virulent strains. Sequence comparisons revealed that multiple horizontal transfers have recently occurred between *S. thermophilus* and other lactic acid bacteria such as *Lactococcus lactis,* probably in co-cultures from cheese manufacture [5].

2. CHARACTERIZATION OF AN INTEGRATIVE AND POTENTIALLY CONJUGATIVE ELEMENT, ICE*ST1*

2.1. Characterization of a site-specific integrative element

Map comparison and probe hybridizations showed that the chromosome of *S. thermophilus* CNRZ368 contains a 34.7-kb region, ICE*St1*, which is absent in its close relative A054 [1]. ICE*St1* is flanked by a 27-bp direct repeat whereas only one copy of this sequence was found in the corresponding region of A054. The right 27-bp sequence of ICE*St1* and the unique 27-bp sequence of A054 include the 3' end of *fda*, which encodes a putative fructose-1,6-diphosphate aldolase. The ICE*St1* right end (Fig. 1) codes for the protein Xis which shares 37–41% identity with the putative excisionase of the conjugative transposons Tn*5252* from *S. pneumoniae* and Tn*5276* from *L. lactis* [1]. It also encodes an integrase (Int) related to those of Tn*5252*, Tn*5276* and numerous temperate phages of lactic acid bacteria [1].

The site-specific recombination between the two 27-bp sequences, included in the attachment sites *attL* and *attR*, generates a chromosomal integration site, *attB*, identical to that found in A054 and a recombination site, *attI,* carried by a circular molecule corresponding to an excised form of ICE*St1* [1]. *attB* and *attI* were detected by PCR in CNRZ368. The PCR products carrying *attB* or *attI* were not obtained when the *int* gene of ICE*St1* was disrupted by the integration

of a thermosensitive plasmid. The recombination module (i.e. *xis-int-attI*) and the integration module (i.e. *int-attI*) were cloned downstream from a promoter on the thermosensitive plasmid pG$^+$host9. These plasmids integrate by site-specific recombination between their *attI* site and the chromosomal *attB* site. After integration, the plasmid containing the recombination module can excise by recombination between *attL* and *attR,* whereas the plasmid containing the integration module does not excise. Therefore, site-specific excision is catalyzed by both the integrase and excisionase, whereas the integrase but not the excisionase is needed for site-specific integration.

2.2. ICE*St1* carries a putative conjugation module

Seven putative proteins encoded by the right part of ICE*St1* are related to proteins encoded by conjugation modules of some elements from low G+C Gram-positive bacteria [4]. *orfA* codes for a protein related to a transfer protein encoded by the staphylococcal conjugative plasmid, pSK41. The putative products of six ORFs (*orfC*, *orfD*, *orfE*, *orfG*, *orfJ* and *orfK*) are related to transfer proteins encoded by the conjugative transposon Tn*916* from *Enterococcus faecalis* [4]. Furthermore, the right part of ICE*St1* also encodes four putative proteins (OrfF, OrfH, OrfL and OrfM) related to proteins encoded by the conjugation modules of putative conjugative elements which are integrated into the completely sequenced chromosome from various low G+C Gram-positive bacteria (data not shown). Therefore, the right region of ICE*St1* is likely to carry a conjugation module. However, the translation products of two small ORFs belonging to this module, *orfB* and *orfI*, are not significantly related to known or putative proteins.

Various elements, which excise by site-specific recombination, self-transfer the resulting circular form by conjugation and integrate by recombination between a specific site of this circular form and a site in the genome of their host, have been found in bacteria. All these elements were recently proposed to be grouped in Integrative and Conjugative Elements (ICEs) irrespective of the mechanisms of conjugative transfer or of integration [3, 4]. Therefore, ICE*St1* is an ICE or derives from an ICE.

3. IDENTIFICATION AND STRUCTURE OF ELEMENTS RELATED TO ICE*ST1*

Elements related to ICE*St1* and/or integrated into the same location were searched for in twenty strains of *S. thermophilus* by hybridization with probes corresponding to the various regions of ICE*St1* and by amplifications of the *fda* locus. Four types of elements were found in seven strains. All are integrated into exactly in the same site as ICE*St1*.

ICE*St3* (28.1 kb) has a recombination module (including *attL* and *attR*) almost identical to that of ICE*St1* (Fig. 1) and excises by site-specific recombination. Furthermore, it carries putative conjugative and regulation modules that are almost identical or closely related to those of ICE*St1*. Therefore, this element could be an ICE. The three other elements do not carry conjugation modules and their regulation and recombination modules are truncated (Fig. 1). Therefore, they are neither integrative nor conjugative. The sequence comparisons suggest that these three elements derive from ICEs closely related to ICE*St1* and ICE*St3* by deletion of these modules. Sequences related to *attL* and *attR* of ICE*St1* were found to flank two of these elements, that were named CIMEs for CIs Mobilizable Elements, CIME19258 (15.0 kb) and CIME302 (13.1 kb). Another element, ΔCIME308 (16.8 kb), carries an *attR* site (Fig. 1) and a truncated integrase gene but does not harbor an *attL* site. Sequence comparisons suggest that it derives from a CIME by an IS*1193*-mediated deletion of its left end.

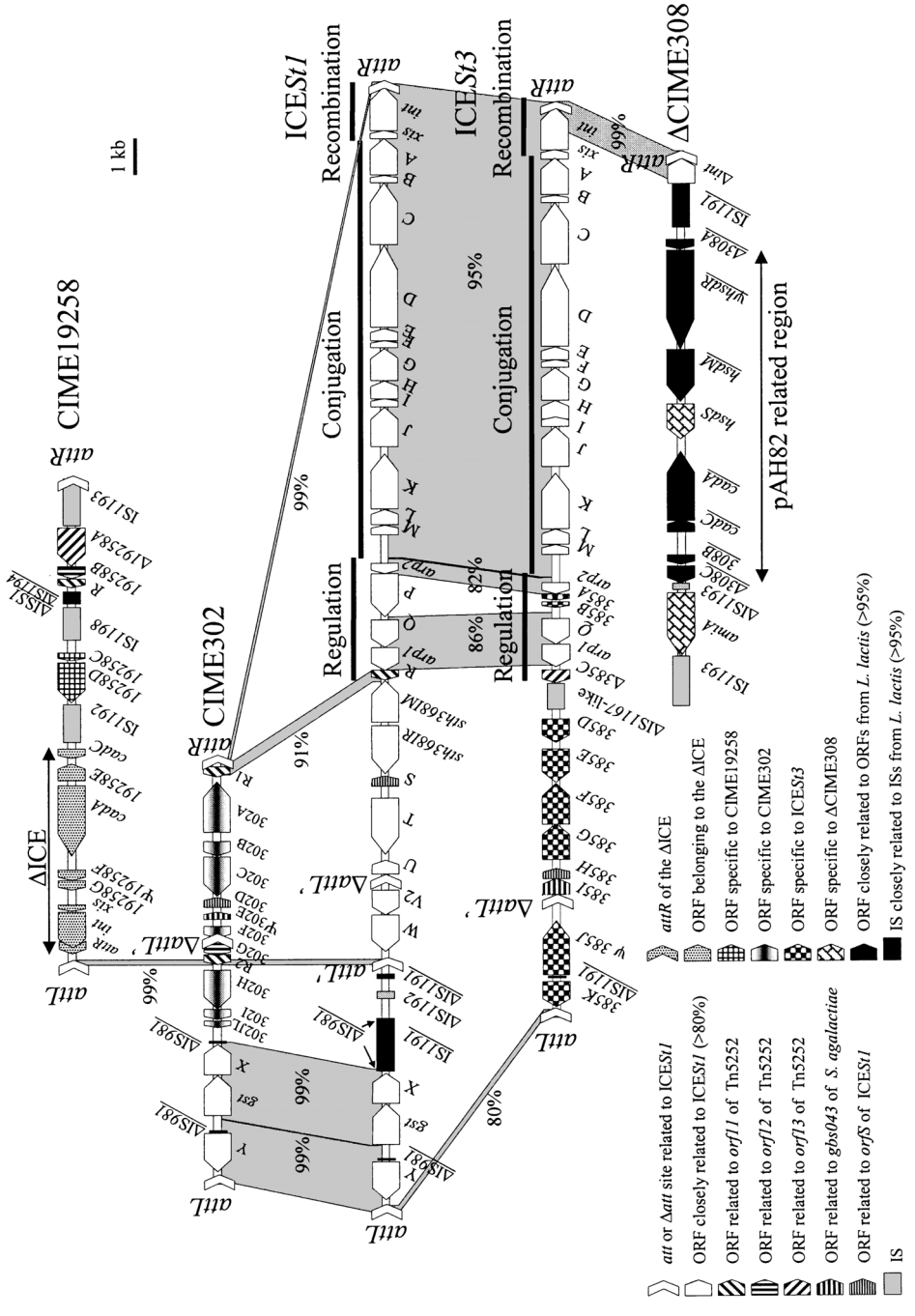
1 kb
CIME19258
attR
attL
ΔICE
CIME302
ICESt1
Recombination
Conjugation
Regulation
ICESt3
ΔCIME308
pAH82 related region
99%
91%
86%
82%
95%
80%
att or Δatt site related to ICESt1
ORF closely related to ICESt1 (>80%)
ORF related to orf11 of Tn5252
ORF related to orf12 of Tn5252
ORF related to orf13 of Tn5252
ORF related to gbs043 of S. agalactiae
ORF related to orf3 of ICESt1
IS
attR of the ΔICE
ORF belonging to the ΔICE
ORF specific to CIME19258
ORF specific to CIME302
ORF specific to ICESt3
ORF specific to ΔCIME308
ORF closely related to ORFs from L. lactis (>95%)
IS closely related to ISs from L. lactis (>95%)

Figure 1. Comparison of ICE*St1*, ICE*St3*, CIME302, CIME19258 and ΔCIME308 from *S. thermophilus*. The location and orientation of ORFs, pseudogenes (ψ) or truncated ORFs (Δ) are indicated by arrowed boxes, complete or truncated ISs by rectangles, and complete or truncated *att* sites by arrowheads. Some very short truncated ORFs and the ORFs belonging to ISs are not indicated. The white boxes depict ORFs closely related to ORFs of ICE*St1* (>80%). ORF names beginning with "*orf*" are abbreviated with the corresponding letter and/or number. The gray areas join sequences that are closely related (>80% nucleotide identity); the identity percent is indicated. The close relationship between short non-coding internal regions or ISs is not displayed. The ORFs sharing 40–80% nucleotide identities are represented by boxes with identical symbols. The black boxes and rectangles correspond to ORFs and ISs almost identical to *L. lactis* sequences (>95% nucleotide identity); the names of these ORFs and ISs are underlined. The left limit of the ΔICE integrated into CIME19258 (corresponding to the right end of the DICE in the figure) cannot be defined exactly.

4. FUNCTIONS ENCODED BY ICES AND CIMES

ICEs and CIMEs carry modules that are not involved in the conjugation or the maintenance of the element. They encode phage resistance: ICE*St1* encodes the restriction-modification (RM) system *Sth368I* (*sth368IR* and *sth368IM*) [2] and a putative abortive infection resistance mechanism (*orfY*), CIME302 a putative isoschizomer of the RM systems *Mva*I or *Bcn*I (*orf302A* and *orf302C*), ICE*St3* two methyltransferases which probably belong to an unknown putative RM system (*orf385F* and *orf385G*), and ΔCIME308 a type I RM system which is probably not functional (*hsdS*, *hsdM* and ψ*hsdR*). Furthermore, ΔCIME308 encodes a putative binding subunit of an oligopeptide ABC carrier (*amiA*) and a putative cadmium resistance (*cadA* and *cadC*), CIME19258 a putative cadmium resistance (*cadA* and *cadC*), and ICE*St1* and CIME302 a putative glutamate and/or aspartate carrier (*gst*). Some of these functions would be interesting for applied purposes. Therefore, conjugative transfers of ICEs or CIMEs could be useful for improving industrial strains.

5. ORIGIN OF MODULES AND MECHANISMS OF MODULE ACQUISITION

5.1. Exchanges with plasmids or unrelated ICEs and horizontal transfers

A large part of ΔCIME308 (i.e. ψ*hsdR*, *hsdM*, *cadA*, *cadC*, *orf308B* and Δ*orf308C*) (Fig. 1) is almost identical to a region of the plasmid pAH82 from *L. lactis* [10]. Furthermore, some ORFs belonging to this region and the adjacent ORF Δ*orf308A* are almost identical to ORFs or genes found in other plasmids from *L. lactis*, *S. thermophilus* and *Listeria innocua* (data not shown). This suggests that genes were exchanged between CIMEs (or ICEs) and plasmids. The sequence analyses also reveal complete or truncated copies of insertion sequences (ISs) which are almost identical to complete or truncated ISs from other bacterial species. These species include lactic acid bacteria used in food fermentation such as *L. lactis* (IS*1191*, IS*981*, IS*S1* α, IS*S1* β and IS*1194*) or *Leuconostoc mesenteroides* (IS*S* α), pathogenic lactic acid bacteria such as *Enterococcus faecalis* (IS*S1* α), *Enterococcus hirae* (IS*1194*) or *Streptococcus pneumoniae* (IS*1194*), and *Listeria innocua* (IS*S1* α). Therefore, various sequences found in ICEs and CIMEs of *S. thermophilus*, and perhaps complete CIMEs or ICEs, have undergone horizontal transfers with other species of bacteria.

The region located at the right of the CIME19258 *attL* site is related (61% identity) to the recombination module of Tn*916*, including its *attR* site, its integrase gene, *int*, and excisionase gene, *xis*. The region located at the right of this module (*orf19258G*, ψ*orf19258F*, *cadA*, *orf19258E* and *cadC*) is related (83% identity) to the regulation module (*lmo1098-lmo1099*) and the cadmium resistance module (*cadA-lmo1102*) of a putative ICE found in the completely

sequenced genome of *Listeria monocytogenes* EGD-e (*lmo1097-lmo1115*), ICE*Lm1*. However, CIME19258 does not possess any sequences related to the conjugation modules of Tn*916* or ICE*Lm1* and to their *attL* site. Therefore, CIME19258 bears a truncated ICE related to Tn*916* and ICE*Lm1*. This suggests that CIME19258 has acquired a cadmium resistance module by the integration of an ICE within the CIME.

5.2. Evolution of ICEs and CIMEs by site-specific accretion

The *attR* sequences of ICE*St1* and CIME19258 (102 bp) share 67% identity, and their *attL* sequences (202 bp) 60%. A plasmid bearing the *xis* and *int* genes from ICE*St1* was introduced into the strain harboring CIME19258 (ATCC19258^T). In this recombinant strain, the *attB* and *attI* sites resulting from the excision of CIME19258 by site-specific recombination were detected by PCR, indicating that the *att* sites of CIME19258 are functional.

Furthermore, the CIME19258 *attL* sequence is almost identical (99% identity) to an internal sequence of ICE*St1* (*attL'*). The site-specific recombination between *attL'* and *attR* of ICE*St1* leads to the formation of a CIME flanked by two recombination sites (*attL* and *attB'*) and to the excision of a 27 205-bp circular molecule carrying an *attI'* site. The sites resulting from the excision (i.e. *attB'* and *attI'*) were detected by PCR in CNRZ368. The PCR products carrying *attB'* or *attI'* were not obtained when the *int* gene of ICE*St1* was disrupted by the integration of a thermosensitive plasmid. The excised molecule includes the conjugation and recombination modules of ICE*St1*. Therefore, the sequences flanked by *attL'* and *attR* (i.e. the central and right regions of ICE*St1*) correspond to a novel putative ICE, ICE*St2*. This suggests that ICE*St1* results from the accretion of an ICE and a CIME.

Furthermore, a truncated *attL'* internal site (71 bp, 54% identity with *attL* of ICE*St1*) was also found in CIME302 (Fig. 1). This putative Δ*attL'* site of CIME302 is associated with two ORFs, *orfR2* and *orf302G*, which are related to the ORFs located at the left end of the complete or truncated regulation modules of ICE*St1* (*orfR*), CIME302 (*orfR1*) and CIME19258 (*orfR* and *orf19258B*). This suggests that the region *orfR2*-Δ*attL'* results from the deletion of a region which initially included complete regulation, conjugation and recombination modules. Moreover, putative internal Δ*attL'* sites are also found in ICE*St2*, i.e. in the putative ICE corresponding to the central and right regions of ICE*St1*, (71 bp, 69% identity) and in ICE*St3* (44 bp, 93% identity). Therefore, ICE*St1* and the related elements have evolved by accretions of ICEs and CIMEs.

In the proposed model (Fig. 2), some ICEs have lost their conjugation and recombination modules by deletion but have retained *attL* and *attR*, leading to CIMEs. Strains that harbor a CIME or an ICE (ICE1) have acquired a related ICE by conjugation (ICE2 in Fig. 2). Then, ICE2 integrated by recombination between its *attI* site and the *attR* site flanking the CIME (or of ICE1) leading to structures such as *attL*-CIME-*attI'*-ICE2-*attR'* (or *attL*-ICE1-*attI'*-ICE2-*attR'*). The resulting internal *attI'* site of the CIME-ICE2 structure or the regulation, conjugation and recombination modules (including the *attI'* site) of ICE1 found in the ICE1-ICE2 structure were partially deleted, leading to an *attL'* or a truncated *attL'* (Δ*attL'*). The whole structure (ICE3 in Fig. 2) could excise by recombination between *attL* and *attR'* (generating *attI''*) and transfer to a new host bacterium, leading to composite ICEs such as ICE*St1* or ICE*St3*. Alternatively, the regulation, conjugation and recombination modules could be deleted, leading to a composite CIME such as CIME302. In this model, the ICEs could acquire novel modules by the site-specific accretion of CIMEs and the ICEs could mobilize CIMEs.

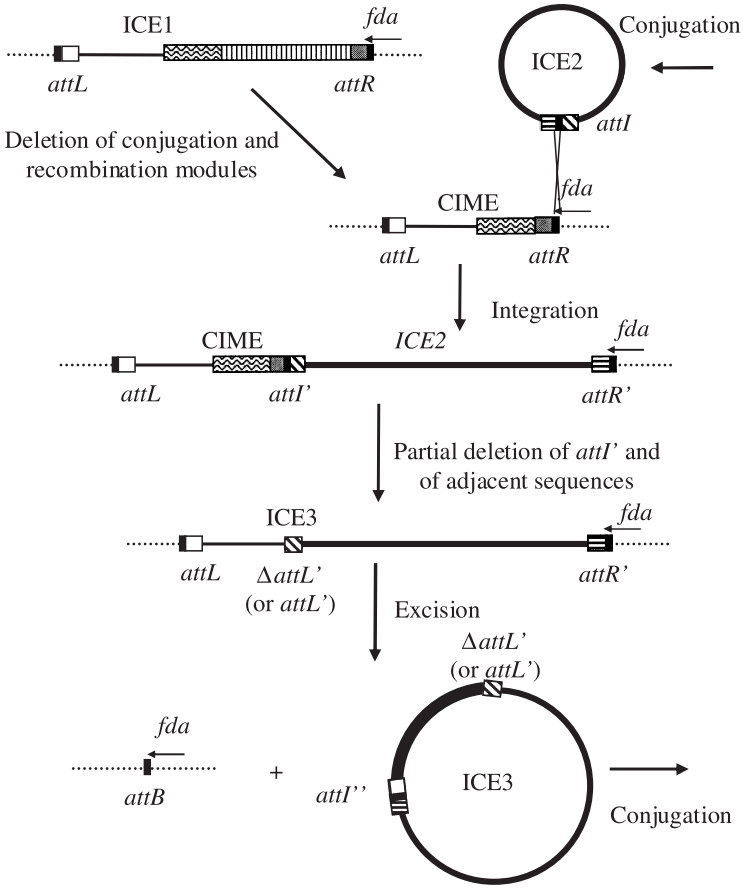

Figure 2. Model of evolution of ICEs and CIMEs by site-specific accretion, deletion and *cis*-mobilization. The black rectangles correspond to the 27-bp sequences (or the homologous sequences) found in all attachment sites of ICE*St1*-related elements. The deletion of the internal *attI* attachment site created by the site-specific accretion of the CIME and of the ICE can lead, not only to a functional *attL'*, but also to a truncated *attL'* as indicated in the figure or to the loss of the site (not shown). The *attL*-CIME-Δ*attL'*-ICE-*attR'* could not only arise by partial deletion of the *attI'* from the structure *attL*-CIME-*attI'*-ICE-*attR'* but also by the tandem integration of two related ICEs followed by deletion of the regulation, conjugation and recombination modules (including a part of *attI'*) of the left ICE.

6. KNOWN ICES AND CIMES: THE TIP OF THE ICEBERG

Few ICEs have been described. However, a recent genomic search revealed 17 putative ICEs in completely or incompletely sequenced genomes from 23 various low G+C bacteria [4]. This suggests that ICEs are widespread in bacteria. Therefore, like the conjugative plasmids and the prophages, ICEs could be one of the main types of elements responsible for horizontal gene transfer.

Elements which, like CIMEs, are related to ICEs, and lack conjugation and recombination modules but are flanked by

attL and *attR*, have not previously been reported. Various pathogenicity islands (PAIs) encode their own excision and integration by site-specific recombination but the mechanism of their transfer remains unknown. Other PAIs have lost the integrative functions by inactivation of the integrase by point mutation or deletion [6]. A PAI from *Salmonella enterica* is a putative ICE [11] and the putative ICE EfaD2 from *Enterococcus faecalis* is a PAI [3, 4, 12]. Some PAIs are flanked by direct repeats or are associated with tRNA genes but do not possess functional recombination modules. Like the CIMEs of *S. thermophilus*, some of them could derive from ICEs.

Accretion of CIMEs and ICEs has not previously been described. However, the ICE Tn*5481* from *L. lactis* bears an internal 187-bp sequence sharing 88% identity with its *attL* site [8]. Furthermore, the ICE SXT from *Vibrio cholerae* was found to site-specifically integrate into the *attL* or *attR* sites of the related element R391 [7]. Therefore, site-specific accretion could be a major evolution mechanism of ICEs.

REFERENCES

[1] Burrus V., Roussel Y., Decaris B., Guédon G., Characterization of a novel integrative element, ICE*St1*, in the lactic acid bacterium *Streptococcus thermophilus*, Appl. Environ. Microbiol. 66 (2000) 1749–1753.

[2] Burrus V., Bontemps C., Decaris B., Guédon G., Characterization of a novel type II restriction-modification system, *Sth*368I, encoded by the integrative element ICE*St1* of *Streptococcus thermophilus* CNRZ368, Appl. Environ. Microbiol. 67 (2001) 1522–1528.

[3] Burrus V., Pavlovic G., Decaris B., Guédon G., Conjugative transposons: the tip of the iceberg, Mol. Microbiol. 46 (2002) 601–610.

[4] Burrus V., Pavlovic G., Decaris B., Guédon G., The ICE*St1* element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration, Plasmid 48 (2002) 77.

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

[6] Hacker J., Kaper J.B., Pathogenicity islands and the evolution of microbes, Annu. Rev. Microbiol. 54 (2000) 641–679.

[7] Hochhut B., Beaber J.W., Woodgate R., Waldor M.K., Formation of chromosomal tandem arrays of the SXT element and R391, two conjugative chromosomally integrating elements that share an attachment site, J. Bacteriol. 183 (2001) 1124–1132.

[8] Immonen T., Wahlstrom G., Takala T., Saris P.E., Evidence for a mosaic structure of the Tn*5481* in *Lactococcus lactis* N8, DNA Seq. 9 (1998) 245–261.

[9] Ochman H., Lawrence J.G., Groisman E.A., Lateral gene transfer and the nature of bacterial innovation, Nature 405 (2000) 299–304.

[10] O'Sullivan D., Ross R.P., Twomey D.P., Fitzgerald G.F., Hill C., Coffey A., Naturally occurring lactococcal plasmid pAH90 links bacteriophage resistance and mobility functions to a food-grade selectable marker, Appl. Environ. Microbiol. 67 (2001) 929–937.

[11] Pembroke J.T., MacMahon C., McGrath B., The role of conjugative transposons in the *Enterobacteriaceae*, Cell Mol. Life Sci. 59 (2002) 2055–2064.

[12] Shankar N., Baghdayan A.S., Gilmore M.S., Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*, Nature 417 (2002) 746–750.