Characterization of a Novel Integrative Element, ICESt1, in the Lactic Acid Bacterium Streptococcus thermophilus

Vincent Burrus, Yvonne Roussel, Bernard Decaris, Gérard Guédon

To cite this version:
Vincent Burrus, Yvonne Roussel, Bernard Decaris, Gérard Guédon. Characterization of a Novel Integrative Element, ICESt1, in the Lactic Acid Bacterium Streptococcus thermophilus. Applied and Environmental Microbiology, American Society for Microbiology, 2000, 66 (4), pp.1749 - 1753. <hal-01659518>

HAL Id: hal-01659518
https://hal.archives-ouvertes.fr/hal-01659518

Submitted on 8 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Characterization of a Novel Integrative Element, \textit{ICESSt1}, in the Lactic Acid Bacterium \textit{Streptococcus thermophilus}

VINCENT BURRUS, YVONNE ROUSSSEL,‡ BERNARD DECARIS,* AND GÉRARD GUÉDON

Laboratoire de Génétique et Microbiologie, INRA UA952, Faculté des Sciences, Université Henri Poincaré (Nancy 1), 54506 Vandoeuvre-lès-Nancy, France

Received 17 September 1999/Accepted 6 January 2000

The 35.5-kb \textit{ICESSt1} element of \textit{Streptococcus thermophilus} CNRZ368 is bordered by a 27-bp repeat and integrated into the 3' end of a gene encoding a putative fructose-1,6-biphosphate aldolase. This element encodes site-specific integrase and excisionase enzymes related to those of conjugative transposons \textit{Tn5276} and \textit{Tn5252}. The integrase was found to be involved in a site-specific excision of a circular form. \textit{ICESSt1} also encodes putative conjugal transfer proteins related to those of the conjugal transposon \textit{Tn916}. Therefore, \textit{ICESSt1} could be or could be derived from an integrative conjugal element.

Cocultures of various lactic acid bacteria are used during the manufacture of dairy products. Sequence comparisons and hybridizations reveal that horizontal transfers between a large array of species of lactic acid bacteria have occurred, most likely during dairy cocultures (13, 32). This convincing evidence indicates that insertion sequences \textit{IS1191}, \textit{IS981}, \textit{IS1}, and \textit{IS1194} (4, 5, 14, 25) and some open reading frames (ORFs) involved in exopolysaccharide synthesis (6) or in restriction-modification (24) were transferred between the lactic acid bacteria \textit{Streptococcus thermophilus} and \textit{Lactococcus lactis} in cocultures used during cheese manufacture. However, the mechanism of genetic exchange between these two species remains unknown, and no conjugative element has been previously characterized in \textit{S. thermophilus}.

Cloning of \textit{var1C} and localization of its limits. The Sm4 fragment of the \textit{S. thermophilus} CNRZ368 chromosome was previously found to contain the 35-kb variable region \textit{var1C}, which was absent from the corresponding chromosomal fragments of strains A054 and NST2280 (28). A region containing an \textit{IS1191} copy inserted in a truncated \textit{IS981} element (14) was cloned and found to be included in \textit{var1C} (28). Chromosome walking using a \textit{AGEM11} genomic library of CNRZ368 (25) was performed to isolate recombinant \textit{λ} bacteriophages overlapping the \textit{var1C} region. Their inserts were subcloned in pBC KS+ and used as hybridization probes on A054 and NST2280 DNAs. S35, ES27, I132.3, ES13, and SC02 fragments hybridized to A054 and NST2280 DNAs. On the contrary, all of the probes covering the 35.5-kb region (except \textit{IS1191} and \textit{IS981}) and located between the \textit{HindIII} sites \textit{H} and \textit{R} (Fig. 1) did not hybridize to A054 and NST2280 DNAs (data not shown). Furthermore, CNRZ368, A054, and NST2280 showed identical restriction maps in regions located to the left of the \textit{HindIII} site \textit{H} and to the right of the \textit{HindIII} site \textit{R} (Fig. 1). The same fragment from A054 and NST2280, but two different fragments from CNRZ368 are adjacent to each other in strains A054 and NST2280 (Fig. 1).

Because A054 and CNRZ368 are very closely related to each other, but distantly related to NST2280 (28), the absence of \textit{var1C} in A054 and NST2280 probably results from an insertion in CNRZ368 rather than from two independent deletions in the two other strains.

\textit{var1C} is bordered by a direct repeat and encodes an integrative system. Sequencing of the \textit{var1C} limits revealed that the element is bordered by a 27-bp direct repeated sequence (R1) containing a \textit{HindIII} site (Fig. 2). A 362-bp fragment was obtained by PCRs performed with the DNA of \textit{S. thermophilus} A054 by using the convergent primers O132.3 (GGACTACT AAGAGACAT) and O131.2 (TTTGTGATA TACGAA GC) (Fig. 3). The sequence of this fragment revealed a unique R1 copy identical to those found on either side of \textit{var1C} in CNRZ368 (Fig. 2). Sequence comparison indicates that R1 direct repeats correspond to the boundaries of \textit{var1C} (Fig. 2).

Two ORFs, \textit{int} and \textit{xis}, are located within \textit{var1C} near the right copy of R1 (Fig. 1 and Table 1). The putative protein encoded by \textit{int} shows significant similarities to site-specific recombinases belonging to the \textit{βlC3} subgroup of the integrase family (http://members.home.com/domespo/trhome.html). This subgroup includes a large array of integrases of temperate bacteriophages and conjugative transposons of lactic acid bacteria and other gram-positive low-G+C bacteria. The C terminus of \textit{Int} contains the five amino acids which are perfectly conserved in this family (data not shown) (1, 3, 11). Furthermore, \textit{xis}, located to the left of the \textit{int} gene, encodes a small basic protein (pI 9.88) which show significant similarities to excisionases of two conjugative transposons, \textit{Tn5252} of \textit{Streptococcus pneumoniae} and \textit{Tn5276} of \textit{L. lactis} (Table 1). \textit{int} and \textit{xis} are located at comparable positions in many prophages and conjugative transposons.

Therefore, these ORFs probably encode an integrative system which would mediate excision of \textit{var1C} by site-specific recombination between the two R1 copies corresponding to the cores of the left and right attachment sites \textit{attL} and \textit{attR}. The unique R1 sequence found in A054 would be the \textit{att} attachment site used for \textit{var1C} integration. \textit{fda}, which flanks the right of \textit{var1C} (Fig. 1), encodes a putative fructose-1,6-biphosphate aldolase (Table 1). The 3' end of \textit{fda} includes 20
bp of the R1 core of attR (Fig. 2). Thus, var1C integration does not change the sequence of fda. Numerous integrative elements (e.g., prophages or integrative conjuguative elements) integrate into the 3' end of genes encoding tRNAs, their sequences remaining unmodified by the integration (8, 15, 17, 23, 30, 31). Other integrative elements (e.g., most of the conjuguative transposons) integrate into several or numerous sites (19, 26). Only a few elements site specifically integrate into the 3' end of protein-encoding genes. The substitution sequence is then generally only similar to the original one (10, 18).

An imperfect 14-bp inverted repeat, R2, is located 29 bp to the right of the 3' end of the int gene and 21 bp to the left of the R1 core of attR (Fig. 1). The potential stem-loop structure \((\Delta G = -18.8 \text{ kcal} \cdot \text{mol}^{-1})\) (33), preceded by a stretch of A's and followed by a stretch of T's, could be used as a p-independent transcription terminator for both int and fda. A perfect 13-bp inverted repeat, R5 \((\Delta G = -14.8 \text{ kcal} \cdot \text{mol}^{-1})\), preceded by a stretch of A's, is located 2 bp to the left of the core of attL (Fig. 1) and could be used as a transcription termination signal for fda prior to the var1C integration. Therefore, these data suggest that the expression of fda would not be changed after var1C integration.

R3, a perfect 9-bp direct repeat, was found 2 bp downstream from the stop codon of int (Fig. 1). A copy of this 9-bp sequence was also found 148 bp to the right of the R1 core of attL. R6, an imperfect 12-bp inverted repeat, and R4, an imperfect 9-bp inverted repeat, are located 123 and 229 bp to the right of the core of attL, respectively. R2, R3, R4, and R6 could be binding sites for integrase or host-encoded proteins involved in the recombination.

Detection of site-specific recombination products. A nested PCR was performed to amplify the putative junction between the var1C termini, which could result from a site-specific recombination event between the R1 cores of attL and attR. Nested-PCR amplification was performed with the O132.5 (GATGAAATTCACATCAGAT-3131.5) (CAGGAAATCGATATTGACA) outer primer pair and the O132.4 (AGTTGAAACTAGACTCAG)-O131.1 (TTCGACATACCGCATATC) inner primer pair (Fig. 3A) according to the method described by Manganelli et al. (21). As expected, no product was identified in strain A054 (Fig. 3B), which does not contain var1C. The sequence of the 536-bp PCR product obtained in CNRZ368 (attL, Fig. 2) is identical to the expected sequence resulting from site-specific recombination between the R1 cores of attL and attR. The PCR product was digoxigenin-labelled and hybridized to EcoRI-digested A054 and CNRZ368 chromosomal DNA. As expected, this probe hybridizes with the two fragments containing the var1C termini in CNRZ368, but not with A054 DNA (data not shown). Site-specific excision of var1C in CNRZ368 should also lead to a junction between sequences flanking var1C, identical to that observed in A054. PCR amplification using the O132.3-O131.2 primer pair (Fig. 3A) was performed to detect this junction. PCR products obtained for A054 and CNRZ368 show the same size (Fig. 3B) and restriction map (data not shown).

Detection of these two junction fragments implies in

\[
\begin{align*}
\text{attL} & \quad \text{tttctggatc tctggtttta cAGTCTAATT AAGCTTTGTG TGCTGACat aggytggattt attc} \\
\text{attR} & \quad \text{cttttatattt caagttata tAGTCTAATT AAGCTTTGTG TGCTGACca aatacagtca tagq} \\
\text{attB} & \quad \text{ctttctggtt cttggtttta cAGTCTAATT AAGCTTTGTG TGCTGACca aatacagtca tagq} \\
\text{attL} & \quad \text{ctttatatattt caagttata tAGTCTAATT AAGCTTTGTG TGCTGACat aggytggattt attc} 
\end{align*}
\]

FIG. 2. Comparison of the nucleotide sequences of the four attachment sites. attL and attR include, respectively, the left and right termini of var1C of strain CNRZ368. attB corresponds to the partial sequence of a PCR product obtained from strain A054 with the primer pairs O132.3 and O131.2. attL corresponds to the partial sequence of a nested-PCR product obtained from strain CNRZ368 with the primer pairs O132.5-O131.5 and O132.4-O131.1 (Fig. 3). R1 sequences are written in upper-case letters. The italic letters correspond to the internal sequence of var1C. Underlined letters indicate the bases that are complementary to the 3' end of the fda gene encoding fructose-1,6-bisphosphate aldolase. Sequences underlined twice correspond to the HindIII restriction sites included in R1.
CNRZ368 the excision of a covalent circular molecule in some cells of the population. The R1 sequences found in the chromosome of A054, in the circular form of var1C, and in the ends of integrated var1C probably constitute the core of the attB, attI, attL, and attR attachment sites: the strand exchange reaction probably takes place by crossover events similar to those involved in λ integration and excision. The length of the core of attachment sites suggests that this element would show very strong insertional site specificity.

Disruption of the int gene prevents var1C excision. The ORF int was disrupted in order to prove its involvement in var1C excision. The thermosensitive plasmid pNST152 was constructed by subcloning the 754-bp HindIII fragment of pNST131.1 containing a fragment of int (region encoding residues 137 to 383 of the integrase) into pG+Host9 (20). pNST152 was used to transform S. thermophilus CNRZ368 by electroporation according to the method of Marciset and Mollet (22). Integration of pNST152 into the int gene was promoted by homologous recombination at a nonpermissive temperature (42°C). The integration site and the number of integrated copies were verified by hybridization of probe I131.1 to PstI patterns of integrants (data not shown). The recombinant strain NST1008 contains two truncated copies of int resulting from the integration of a unique copy of pNST152 within the int gene of CNRZ368. Junction fragments containing attII or attI were not detected in NST1008 by PCR experiments (Fig. 3B), whereas a fragment bearing attR was amplified from NST1008 by using the O131.1 and O131.2 primers (Fig. 3). Therefore, int gene disruption causes the disappearance of the two junction fragments and, therefore, of the covalent circular molecule, showing that this gene is actually involved in var1C excision.

var1C encodes proteins related to those of some conjugative system. The 5,881-bp region located to the left of the var1C element of S. thermophilus belongs to another subfamily whereas the integrase of Tn916 belongs to the LC3 integrase subfamily, was named ICESt1, for integrative conjugative element of S. thermophilus no. 1.

The possible conjugative system of ICESt1 is related to that of Tn916, but not to the system encoded by Tn5252. On the contrary, the ICESt1 excisionase is related only to those of Tn5276 and Tn5252. Moreover, the integrases of ICESt1, Tn5276, and Tn5252 belong to the 4bLC3 integrase subfamily, whereas the integrase of Tn916 belongs to another subfamily (http://members.home.net/domespo/trhome.html). Furthermore, differences in G + C content between the xis and int genes (about 34%) and orfABCD (about 42%) of ICESt1 also suggest that the integration-excision system and the possible conjugative system have different origins or have undergone very different evolutions. A similar structure is observed in Tn916 (about 36% G + C for the xis and int genes versus about 40% G + C for the conjugative system). This suggests that ICESt1 and Tn916 possess a modular structure which results from exchanges or acquisitions of sequences from different sources. This modular structure and evolution are similar to
TABLE 1. Characteristics of the sequenced ORFs and encoded proteins examined in this study

| ORF | % G+C | Start | Stop | RBS | Dist1 | Dist2 | Length (no. of amino acids) | Related protein | Origin | % Identity (no. of amino acids) |
|-----|-------|-------|------|-----|-------|-------|-----------------------------|----------------|--------|-------------------------------|
| fla | 38.3  | ND    | TAA  | ND  | ND    | ND    | ND                          | Fruuctose-1,6-biphosphate aldolase Fba (AJ005697) | Chromosome of S. pneumoniae | 80 (77) |
| int | 33.6  | ATG   | TAA  | TAAGGAGG | 7 | −1 | 448 | Integrase Int (U93688) | Pathogenicity island SaPlI of S. aureus | 26 |
|     |       |       |      |     |       |       |                             | Integrase Int (M27965) | Phage ϕL54a of S. aureus | 28 |
|     |       |       |      |     |       |       |                             | Integrase Int (M62697) | Phage ϕadl of Lactobacillus gasseri | 32 (227) |
|     |       |       |      |     |       |       |                             | Integrase Int (L27649) | Conjugative transposon Tn5276 of L. lactis | 24 |
|     |       |       |      |     |       |       |                             | Integrase Int (L29324) | Conjugative transposon Tn5252 of S. pneumoniae | 26 (266) |
| xis | 35.7  | ATG   | TAA  | AAAGGAGT | 5 | +13 | 82 | Excisionase Xis (L29324) | Conjugative transposon Tn5252 of S. pneumoniae | 41 |
|     |       |       |      |     |       |       |                             | Excisionase Xis (L27649) | Conjugative transposon Tn5276 of L. lactis | 41 |
| orfA | 43.3 | ATG   | TAG  | AAAGGAGA | 4 | +25 | 370 | Putative transfer protein TraG (AR051917) | Conjugative plasmid pSK41 of S. aureus | 30 |
|     |       |       |      |     |       |       |                             | Putative transfer protein TraG (L11998) | Conjugative plasmid pGO1 of S. aureus | 30 |
|     |       |       |      |     |       |       |                             | Immunogenic secreted protein Isp (U31811) | Chromosome of Streptococcus pyogenes D471 | 57 (325) |
| orfB | 37.8 | ATG   | TAA  | AGAGGA | 5 | +1 | 74 | No similarity | ND | ND |
| orfC | 41.5 | ATG   | TAG  | TTAGGAGG | 7 | +11 | 626 | Putative membrane protein Orf15 (U09422) | Conjugative transposon Tn916 of E. faecalis | 18 (267) |
| orfD | 42.1 | ATG   | TAG  | AAAGGAGG | 4 | ND | 834 | Putative transfer protein Orf16 (U09422) | Conjugative transposon Tn916 of E. faecalis | 21 |
|     |       |       |      |     |       |       |                             | Unknown protein YddE (AB001488) | Chromosome of B. subtilis 168 | 30 |

* ORFs are listed from the right to the left of the map.
* The RBS consensus sequence of the gram-positive low-G+C bacterium Bacillus subtilis is AAAGGAGG.
* Dist1, distance between the RBS and the start codon.
* Dist2, distance between the start codon of an ORF and the stop codon of the previous ORF on the map (Fig. 1). A negative value indicates an overlapping of two ORFs.
* Functions of proteins and GenBank accession numbers (in parentheses) of nucleotide sequences encoding proteins related to the product of the ORFs sequenced in this study are indicated.
* Identities stretch over the entire length of each of the amino acid sequences of proteins encoded by ICESt1, except when indicated in parentheses.
* ND, not determined.
* Many other related integrases were found in databases, but in this table, we have only indicated a selection of the ones more related to the integrase encoded by varIc.
those of bacteriophages (9, 16) and enterobacterial plasmids (7).

The large size of ICESt1 (35 kb) suggests that this element, like Tn5276, which encodes nisin synthesis (26), could carry industrially attractive genes. The ICESt1 element contains a complete copy of IS191, an insertion sequence probably transferred from S. thermophilus to L. lactis, and a truncated copy of IS981, which was probably transferred from L. lactis to S. thermophilus, most likely in cocultures of these species used during the manufacture of cheese (14). Furthermore, conjugative transposons related to ICEst1, like Tn916 of Enterococcus faecalis and Tn5252 of S. pneumoniae, are broad-host-range elements (12, 34). Therefore, ICESt1 or elements related to ICESt1 could be involved not only in intraspecific but also in interspecific horizontal transfers between S. thermophilus and other lactic acid bacteria.

Nucleotide sequence accession numbers. The GenBank accession numbers of the nucleotide sequences reported in this paper are AJ243105 (left terminus of var1C) and AJ243106 (right terminus of var1C).

We thank E. Maguin for providing the thermosensitive plasmid pG−H19.

This work was supported by grants from the Institut National de la Recherche Agronomique, the University of Nancy I, and the Ministère de l’Education Nationale, de la Recherche et de la Technologie, France.

REFERENCES

1. Abremski, K. E., and R. H. Hoess. 1992. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. Protein Eng. 5:87–91.
2. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.
3. Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Hagbard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L. Narayana, S. L. Pierson III, N. Sternberg, and J. M. Leong. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. 5:433–440.
4. Bourgoin, F., G. Guédon, B. Gintz, and B. Decaris. 1998. Characterization of a novel insertion sequence, IS1914, in Streptococcus thermophilus. Plasmid 40:34–49.
5. Bourgoin, F., G. Guédon, M. Pébay, Y. Roussel, C. Panis, and B. Decaris. 1996. Characterization of a mosaic IS51 element and evidence for the recent horizontal transfer of two different types of IS51 between Streptococcus thermophilus and Lactococcus lactis. Gene 178:15–23.
6. Bourgoin, F., A. Plovine, B. Gintz, B. Decaris, and G. Guédon. 1999. Are horizontal transfers involved in the evolution of the Streptococcus thermophilus exopolysaccharide synthesis loci? Gene 233:151–161.
7. Boyd, E. F., C. W. Hill, S. M. Rich, and D. L. Hartl. 1996. Mosaic structure of plasmids from natural population of Escherichia coli. Genetics 143:1091–1100.
8. Brown, D. P., K. B. Idler, and L. Katz. 1990. Characterization of the genetic elements required for site-specific integration of plasmid pSE211 in Saccharopolyspora erythraea. J. Bacteriol. 172:1877–1888.
9. Brüssow, H., A. Bruttin, F. Desiere, S. Lucchini, and S. Foley. 1998. Molecular ecology and evolution of Streptococcus thermophilus bacteriophages—a review. Virus Genes 16:95–109.
10. Campbell, A. M. 1992. Chromosomal insertion sites for phages and plasmids. J. Bacteriol. 174:7495–7499.
11. Caroll, D., M. A. Kehoe, D. Cavanagh, and D. C. Coleman. 1995. Novel organization of the site-specific integration and excision recombination functions of the Staphylococcus aureus serotype F virulence-converting phages d13 and d42. Mol. Microbiol. 16:877–893.