Mosaic structure of large regions of the Lactococcus lactis subsp. cremoris chromosome

C. Delorme, J.-J. Godon, S. D. Ehrlich and P. Renault

INTRODUCTION

The genus Lactococcus is a distinct and homogeneous group of lactic acid bacteria (Collins et al., 1989) isolated from dairy products, plants, and the mucosa and digestive tract of animals. Within the genus, the status of Lactococcus lactis subsp. lactis (Streptococcus lactis, Lister 1873) and Lactococcus lactis subsp. cremoris (Streptococcus cremoris, Orla-Jensen 1919) as separate species or subspecies has been the subject of some debate, since their phenotypic differentiation is difficult. Garvie & Farrow (1982) proposed that both organisms belong to a single species because of similarities in their phenotype and DNA homology, which was reported to be 80% for the type strains and to range between 75 and 100% (Jarvis et al., 1981). This classification is supported by the high degree of homology of L. lactis subsp. lactis and L. lactis subsp. cremoris 16S rRNA (Collins et al., 1989; Salama et al., 1991). However, it has been argued that 16S rRNA sequence identity may not be sufficient to guarantee species identity (Fox et al., 1992) and the lactis/cremoris classification was questioned by the recent evidence that certain genes of one subspecies hybridize only very weakly with those of the other (Godon et al., 1992b).

To extend the previous observations we undertook hybridization analysis of the entire tryptophan, histidine and branched-chain amino acid (BCAA; isoleucine, leucine and valine) operons. Interestingly, a mosaic structure of the operons was found, with alternating conserved and divergent regions.

METHODS

Bacterial strains, plasmids, phage and media. The bacterial strains, plasmids and the phage used in this work were described previously (Delorme et al., 1992; Godon et al., 1992a). Strains earlier misclassified as L. lactis subsp. lactis are here termed L. lactis subsp. cremoris as proposed previously (Godon et al., 1992b). Lactococcal strains were grown in M17 medium (Terzaghi & Sandine, 1975).

Cloning of L. lactis subsp. cremoris NCDO 763 DNA in Escherichia coli. Total DNA from lactococci was isolated by standard procedures (Loureiro Dos Santos & Chopin, 1987). Subcloning of total DNA was carried out in pBluescript plasmid (Stratagene). A ligation mix was used to transform competent E. coli cells, as described by Sambrook et al. (1989). To isolate the 2-5 kb segment, previously detected in L. lactis subsp. cremoris genomic library was prepared. EcoRI DNA fragments ranging from 2.2 to 2.7 kb were purified by gel electrophoresis and ligated to a dephosphorylated pBluescript vector. The relevant clones were identified by colony hybridization at low stringency (25%, v/v, formamide) with pIL704 (Delorme et al., 1992) as probe. Three positive clones harboured the same plasmid, as judged by its size and by restriction analysis. One, designated pIL732, was chosen for further study.
A 1.5 kb chromosomal fragment of the L. lactis subsp. cremoris his operon was amplified by the polymerase chain reaction (PCR) using the primers His1 (CAATAATCAATTCCATG, nucleotides 1809-1790 of NCDO 763) and His2 (ATTTCTGATTGGTTACGGAT, nucleotides 329-348). The amplified fragment was either rendered blunt-ended with T4 polymerase and inserted into pBluescript in the EcoR site or partially cut with Ald and cloned in the M13mp18 vector cleaved with SmaI.

**Southern hybridization analysis.** Standard procedures were used (Sambrook et al., 1989) with 2 µg of chromosomal DNA. DNA segments were transferred to a nylon membrane (Pal Biodyne) and hybridized with probes that were labelled with [32P]dCTP by the random priming procedure (Boehringer). Stringency of hybridization was controlled by varying the formamide concentration as previously described (Godon et al., 1992). The filters were washed twice in 2 × SSC at 42 °C and then twice at 50 °C.

**DNA sequence analysis.** Nested deletions were produced by the action of DNase I on pBluescript plasmids containing the segments to be sequenced. Single-stranded DNA was prepared as described by Vieira & Messing (1987) and sequenced essentially according to the Applied Biosystems protocol accompanying the 370A DNA sequencer. The DNA was used in dideoxynucleotide chain-termination sequencing reactions with Tag DNA polymerase and fluorescent dye-coupled primers (Applied Biosystems). The reported sequence was determined on both DNA strands. Errors arising from gene amplification were resolved by sequencing at least two independently cloned segments.

The DNA and protein sequences were analysed using the BISANCE and the University of Wisconsin GCG packages, implemented at the Centre InterUniversitaire d'Informatique à Orientation Biomédicale, Paris.

**RESULTS**

**Heterogeneity of the L. lactis subsp. cremoris chromosome**

In a previous study, certain L. lactis subsp. lactis DNA segments were found to hybridize only very weakly with the chromosomal DNA of L. lactis subsp. cremoris (Godon et al., 1992). To determine whether the level of divergence between L. lactis subsp. lactis and L. lactis subsp. cremoris DNA varies significantly along the chromosome we undertook the following Southern hybridization analysis. Segments of two previously sequenced L. lactis subsp. lactis chromosomal regions of ~30 and ~8 kb, covering a total of 15% of the chromosome and containing his (Delorme et al., 1992), leu-ilv-ald (Godon et al., 1992a) and trp (Bardowski et al., 1992) operons (Fig. 1), were hybridized with L. lactis subsp. cremoris DNA at a stringency low enough to detect homology of > 80%. Hybridization was observed with some, but not all, probes (Fig. 1). Control experiments indicated that all probes hybridized with L. lactis subsp. cremoris DNA at a lower stringency (25% formamide, data not shown). We conclude that L. lactis subsp. lactis and L. lactis subsp. cremoris diverge considerably in certain chromosomal regions and less in others.

**Sequence of a DNA segment from the L. lactis subsp. cremoris his operon**

To determine more precisely the level of divergence between the poorly hybridizing segments, the sequence of the 2.5 kb segment from the L. lactis subsp. cremoris his operon, carried on pIL732 (third box in Fig. 1, see Methods), was determined and compared with the sequence of the corresponding segment of L. lactis subsp.
The sequence of this fragment was previously compared in two *L. lactis* subsp. *lactis* strains and was found to be more than 98% identical (Delorme et al., 1993). The *L. lactis* subsp. *cremoris* sequence is ~200 bp longer, due to an insertion extending from position 1384 to 1597. Excluding the insertion, alignment of the two sequences is possible, the mean divergence being ~27%. Surprisingly, divergence is much greater in certain regions than in others, as shown by comparing 70 bp intervals of the two sequences (Fig. 3). It is relatively high, up to 65% with a mean of ~45%, in the central region (positions 480–1700), with only one better conserved interval (15% divergence, see below for possible explanation). In contrast, divergence is relatively low, 10–30% with a mean of ~15%, at the 3' end (beyond position 1750). Some short divergent stretches are interspersed at the conserved 3' end (up to position 480). Interestingly, the G+C content of the entire divergent central region of the *L. lactis* subsp. *cremoris* segment is quite high, 48 mol%, whereas that of the corresponding region of *L. lactis* subsp. *lactis* is only 36 mol% G+C, which is close to the 38 mol% found for total *L. lactis* DNA (Garvie & Farrow, 1982). In contrast, the G+C content of both 5' and 3' conserved regions is not significantly different.

The insertion present in the highly divergent central region of the *L. lactis* subsp. *cremoris* bis segment is composed of (i) three quasi-perfect repeats of 59 bp (position 1400–1582) and (ii) two flanking non-repeated and non-coding stretches, of 22 and 28 bp, respectively (Fig. 2). A remarkable feature of this insertion is its high G+C content (54 mol%), which is higher than that of the flanking region (48 mol%).

**Analysis of the coding sequences**

Variations in the extent of divergence found in the bis operon could be due to rapid accumulation of mutations or to acquisition of DNA from phylogenetically distant species by horizontal gene exchange. In the first case, the divergent regions would almost certainly conserve codon usage (assuming that the gene was functional after it diverged) whilst in the second the chances are that the codon usage would be different. These considerations led us to compare the coding properties of the previously sequenced *L. lactis* subsp. *lactis* and *cremoris* segments. The former was previously shown to carry the 3' end of bisC, the entire orf3 and bisG and the 5' end of bisD (Fig. 2; Delorme et al., 1992), which encode enzymes involved in histidine biosynthesis (HisC, HisG, HisD) and a protein homologous to histidyl tRNA synthetases (Orf3). The same proteins are encoded by the *L. lactis* subsp. *cremoris* segment, except that frameshift mutations are present in the four genes (at positions 117, 1043, 1945 and 2424). The four gene products are therefore probably inactive, which is expected from the observation that all the *L. lactis* subsp. *cremoris* strains are auxotrophic for histidine (17 strains tested, Delorme et al., 1993). When corrected for the frameshifs, the potential *L. lactis* subsp. *cremoris* gene products share 78, 51, 83 and 73% homology, respectively, with their *L. lactis* subsp. *lactis* counterparts. Interestingly, codon usage in the better-conserved terminal regions of the *L. lactis* subsp. *cremoris* sequence is as expected for lactococci, whereas in the highly divergent central region, which encodes ~60% of Orf3 (the carboxy end) and ~20% of HisG (the amino end), codon usage is quite atypical (Table 1; only the significantly deviant codons are shown). For instance, codons GTG for valine, GCG for alanine, GAG for glutamine, and CGG and CGC for arginine are preferentially used in this region whereas they occur rarely in other regions of the *Lactococcus* chromosome (Chopin, 1993). Unusual codon usage, together with the high G+C content and the presence of additional DNA (see above) suggest that the central *L. lactis* subsp. *cremoris* region might originate from a taxonomically distant bacterial species.

**Divergence is not due to the loss of function**

Sequence analysis (see above) and phenotype tests (Delorme et al., 1993) show that the bis operon of *L. lactis* subsp. *cremoris* NCDO 763 is inactive. The question arises whether the divergence might be due to a rapid drift of genes which are no longer used, notwithstanding the atypical codon usage and G+C content of the central region. In this case, the corresponding region from different *L. lactis* subsp. *cremoris* strains might be poorly conserved. To test this possibility a 1.4 kb segment from two different *L. lactis* subsp. *cremoris* strains, IL182 and AM2, was amplified by PCR and its sequence determined (Fig. 4). The two coding sequences differ by less than 0.8% and differ only slightly more, ~3%, from the sequence of the NCDO 763 strain. Moreover, most of the differences between NCDO 763 and the two other strains are either silent (22 and 20 for IL182 and AM2, respectively), or conservative (5 and 7), as expected for functional diverging genes. In addition, all the changes which affect the protein sequence (10 and 9 for IL182 and AM2, respectively) are present at positions which are not well conserved, suggesting that they do not affect protein function. Finally, the frameshift found in position 1043 of orf3 from NCDO 763 is not present in the other two strains, which implies that this change occurred relatively recently when the environmental conditions allowed the fixation of this mutation. These results suggest that the region was functional over a long period of time since strain NCDO 763 diverged from IL182 and AM2, and argue against its rapid drift due to the loss of gene function.

It should be noted that the domains of the highly divergent central region which are better conserved than the adjacent domains (see, for instance, the domain close to position 600, which diverges only ~15%, Fig. 3), code either for the motifs which are typical for class II tRNA synthetases, to which Orf3 belongs (M1 and M2), or for the signature regions (SR1, SR2a and b), specific for histidyl tRNA synthetases (Figs 2 and 3; Raben et al., 1992; Eriani et al., 1990). It is possible that the constraints of preserving protein function might have limited the nucleotide sequence divergence of these domains.

---

**L. lactis subsp. cremoris chromosome structure**

---
Fig. 2. For legend see facing page.
DISCUSSION

The level of divergence between _L. lactis_ subsp. _lactis_ and _L. lactis_ subsp. _cremoris_ varies considerably along a large region of the chromosome. It is low (10–20%) in conserved regions, which agrees with comparisons of the 16S rRNA sequences (Salama _et al._, 1991) and with global DNA–DNA homology measurements (Jarvis _et al._, 1981), but is very high (35–65%) in variable regions. The conserved and variable regions are interspersed, which suggests that the chromosomal region studied has a mosaic structure.

The divergence appears not to be due to an abnormally rapid genetic drift, as indicated by analysis of a region of the _his_ operon which is highly divergent between _L. lactis_ subsp. _lactis_ and _cremoris_, but is remarkably well conserved among different _L. lactis_ subsp. _cremoris_ strains. Several features of this region in _L. lactis_ subsp. _cremoris_ are not representative of the genus _Lactococcus_, such as a high G+C content, unusual codon bias and the presence of a repeated DNA element, suggesting that it might have been acquired by horizontal gene transfer from distantly related bacteria. This interpretation is supported by the existence of a rather sharp boundary between this atypical region and the adjacent lactococcus-like 3' region. Genes acquired by horizontal transfer often have unusual G+C content and codon bias, as shown by extensive analysis of _E. coli_ sequences (Médigue _et al._, 1991). Comparable results have also been found in other bacterial genera (Groisman _et al._, 1992; Shanley _et al._, 1994). It is not known at present from which species the divergent _L. lactis_ subsp. _cremoris_ region was derived.

The postulated horizontal gene transfer is certainly not restricted to the _his_ operon, since interspersion of conserved and divergent regions was also detected in the BCAA and the _trp_ operons. The sequence of these regions of the _L. lactis_ subsp. _lactis_ chromosome is known (Delorme _et al._, 1992) and its analysis did not reveal any distortion in the G+C content or codon bias from other lactococcal genes. Our results do not permit the conclusion that all _L. lactis_ subsp. _cremoris_ divergent regions deviate from the lactococcal model (as does the divergent _his_ region) but point to the mosaic structure of the _L. lactis_ subsp. _cremoris_ chromosome. The genes analysed here are grouped in two regions. The _his_ and the BCAA operons are adjacent and ~ 200 kb from the _trp_ operon (Tulloch _et al._, 1991; Le Bourgeois _et al._, 1992). Future work should establish whether the _L. lactis_ subsp. _cremoris_ chromosome is mosaic only in these two regions or whether the mosaicism extends further along.

DNA acquired by heterospecific gene transfer may be associated with extrachromosomal elements. For instance, sequences showing a high similarity with the _incFII-oriT_ region of plasmid R100 and with a region of the 145 kb plasmid harbouring the gene for the heat-stable toxin (estAI) were found in an _E. coli_ strain 350 and 400 bp upstream of the _phaN_ and _sinR_ loci, respectively (Groisman _et al._, 1992, 1993). Similarly, the _bin_ and _estAI_
regions are associated with a transposon-like element (Simon et al., 1980) or an insertion sequence (Groisman et al., 1993). Interestingly, a conjugative element has been mapped at a position less than 50 kb away from the his operon carrying the mosaic genes in L. lactis subsp. cremoris chromosome (P. Le Bourgeois, personal communication). Furthermore, the L. lactis subsp. cremoris his region contains an element that is reminiscent of prokaryotic repetitive DNA sequences such as REP and ERIC (Higgins et al., 1988; Hulton et al., 1991; Gilson et al., 1991). It is therefore possible that the mosaicism is restricted to only a portion of the L. lactis subsp. cremoris chromosome, which would permit reconciliation of the reported overall DNA homology (Jarvis et al., 1981) with the high local divergence observed in the present work.

Interestingly, the operons resulting from the postulated gene transfer may all have been functional. The trp operon is known to be active (Reiter & Oram, 1962) and the his operon was presumably active, as deduced from the conservative base change pattern among different L. lactis subsp. cremoris strains. It has been reported that intra-specific gene exchange may generate functional genes, such as those involved in penicillin resistance of Neisseria (Spratt et al., 1992; Maynard Smith et al., 1991) or Streptococcus (Martin et al., 1992; Dowson et al., 1989). However, the exchange that occurred in these cases, as well as the extensive recombination in the rfb O-antigen locus in Salmonella enterica (Reeves, 1993) were generated under intense selection. Our results represent the first indication that a hetero-specific gene exchange might

![Fig. 3. Divergence between the L. lactis subsp. lactis NCDO 2118 and L. lactis subsp. cremoris NCDO 763 sequences. The squares connected by the solid line represent the mean divergence over a 70 bp sequence and the horizontal dashed line represents the mean of the region covered. Nucleotide positions in the L. lactis subsp. cremoris NCDO 763 sequence are indicated on the abscissa. M and SR denote conserved motifs and signature regions present in orf3 (indicated in Fig. 2) and the arrows show their position. The ORFs indicated at the top are delimited by vertical dashed lines; DIR, direct repeat.](image)

### Table 1. Codon usage in lactococci

<table>
<thead>
<tr>
<th>Codon*</th>
<th>Total codon number</th>
<th>Codon proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Central region†</td>
<td>Terminal regions‡</td>
</tr>
<tr>
<td>TTA (Leu)</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>TTG (Leu)</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>GTT (Val)</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>GTG (Val)</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>GCT (Ala)</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>GCG (Ala)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>CAA (Gln)</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>CAG (Gln)</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>AAA (Lys)</td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td>AAG (Lys)</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>GAA (Glu)</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td>GAG (Glu)</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>CGG (Arg)</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>GCC (Arg)</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>AGA (Arg)</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

* Only the codons showing significant differences between central and terminal regions are shown.
† Positions 481–1356 and 1600–1740 from L. lactis subsp. lactis NCDO 763.
‡ Positions 1–480 and 1741–2478 from L. lactis subsp. lactis NCDO 763.
§ From Chopin (1993).
yield functional biosynthetic pathway genes, and this presumably not under selective pressure since bacterial phenotype should not be affected.

ACKNOWLEDGEMENTS

We thank A. Khodja for technical assistance, J. Bardowski for providing trp probes and C. Anagnostopoulou for critical reading of the manuscript. This work was supported in part by BRIDGE contract BIOTCT91-0263 of the Commission of the European Communities.

REFERENCES


Received 11 April 1994; revised 30 June 1994; accepted 12 July 1994.