HAL
open science

A specific pattern of splicing for the horse αS1-Casein mRNA and partial genomic characterization of the relevant locus

Dragan Milenkovic, Patrice Martin, Gérard Guérin, Christine Leroux

▶ **To cite this version:**

Dragan Milenkovic, Patrice Martin, Gérard Guérin, Christine Leroux. A specific pattern of splicing for the horse αS1-Casein mRNA and partial genomic characterization of the relevant locus. Genetics Selection Evolution, 2002, 34 (4), pp.509-519. 10.1051/gse:2002020 . hal-00894424

HAL Id: hal-00894424

https://hal.science/hal-00894424v1

Submitted on 11 May 2020

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.

Genet. Sel. Evol. 34 (2002) 509–519
© INRA, EDP Sciences, 2002
DOI: 10.1051/gse:2002020

Original article

A specific pattern of splicing for the horse αS1-Casein mRNA and partial genomic characterization of the relevant locus

Dragan MILENKOVIC, Patrice MARTIN,
Gérard GUÉRIN, Christine LEROUX*

Institut national de la recherche agronomique, Département de génétique animale,
Laboratoire de génétique biochimique et de cytogénétique,
Centre de recherches de Jouy, 78352 Jouy-en-Josas Cedex, France

(Received 7 August 2001; accepted 3 December 2001)

Abstract – Mares' milk has a composition very different from that of cows' milk. It is much more similar to human milk, in particular in its casein fraction. This study reports on the sequence of a 994 bp amplified fragment corresponding to a horse αS1-Casein (*αS1-Cn*) cDNA and its comparison with its caprine, pig, rabbit and human counterparts. The alignment of these sequences revealed a specific pattern of splicing for this horse primary transcript. As in humans, exons 3′, 6′ and 13′ are present whereas exons 5, 13 and 14 are absent in this equine mRNA sequence. BAC clones, screened from a horse BAC library, containing the *αS1-Cn* gene allowed the mapping of its locus by FISH on equine chromosome 3q22.2-q22.3 which is in agreement with the Zoo-FISH results. Genomic analysis of the *αS1-Cn* gene showed that the region from the second exon to the last exon is scattered within a nucleotide stretch nearly 15-kb in length which is quite similar in size to its ruminant and rabbit counterparts. The region between *αS1-* and *β-Cn* genes, suspected to contain *cis*-acting elements involved in the expression of all clustered casein genes, is similar in size (*ca.* 15-kb) to the caprine and mouse intergenic region.

horse / αS1-Casein / cDNA / localisation

1. INTRODUCTION

Milk, the natural food for new-born mammals obviously contains all of the components necessary for the initial phases of postnatal development. With regards to the protein content of milk, marked qualitative and quantitative

* Correspondence and reprints
E-mail: cleroux@clermont.inra.fr
Current address: Institut national de la recherche agronomique, Département d'élevage et de nutrition des animaux, Unité de recherches sur les herbivores, Centre de recherches de Theix, 63122 Saint-Genès-Champanelle, France

(*ca.* 4 $g \cdot L^{-1}$ to 200 $g \cdot L^{-1}$ [10]) differences have been observed between species. Caseins (Cn), the major milk proteins in ruminants, are a family of four acidic phosphoproteins (αS1, αS2, β and κ-Cn; for review see [27]) with relative proportions also varying among species. The amino acid or nucleotide sequences of the four caseins have been determined in a number of species (mainly ruminants and rodents). Multiple alignments of αS1-Cn amino acid sequences from different species, taking into account the exon modular splicing derived from the known structural organisation of the bovine [12], goat [14] and rabbit [11] genes, points to the occurrence of insertion/deletion events [18]. Exon-skipping has been shown to be responsible for such events and for the apparent relatively high structural divergence observed between αS1-Cn from different species as well as for its wide variability in size [19].

Caseins are encoded by four genes physically linked in the order *αS1-*, *β-*, *αS2-* and *κ-Cn*. The overall organisation of the casein locus is fairly conserved between bovine, human and mouse species [6,23–25]. Some differences, however, have been described in the distances separating casein genes and their number (an extra *αS2-Cn*-like gene has been detected in the mouse). Regarding the intergenic region, some differences have also been described: the *αS1-/β-Cn* intergenic region ranges from *ca.* 10 kb in the mouse [25] to less than *ca.* 60 kb in humans [24]. To unravel the mechanisms that underly the co-ordinated expression of all four casein genes, a detailed knowledge of the casein locus organisation is required. The existence of *cis*-acting elements in the *αS1/β* region involved in and required for the expression of all four casein genes has been suggested [23].

Mares' milk has lower protein and fat contents than ruminant milk [4] and its composition is closer to that of human milk (Miranda *et al.*, unpublished). It has been studied for its potential nutritional properties and its effects on human health. This milk and koumiss, a product of its fermentation, are reported to have potential therapeutic benefits and/or dietetic effects [13,16]. Mares' milk can be regarded as an alternative to cows' milk in most children with severe IgE-mediated cows' milk allergy [2].

In contrast to ruminant milk, the mares' milk composition has remained up to now poorly documented, in particular, its caseins composition and structure. Proteomic tools have allowed to characterize the protein fraction of equine milk, clearly demonstrating the presence of αS2-Cn and κ-Cn [21]. Very recently, the primary structure of κ-Cn isolated from mares' milk by RP-HPLC, has also been reported [9] and the *β-Cn* cDNA sequence is now available in data bases (accession number: AF214526). However there is no data regarding the *αS1-Cn* sequence. Similarly, no genomic information is available on the casein locus in the horse. Here, we report an equine *αS1-Cn* cDNA sequence and the comparison with its counterpart from different species as well as the partial casein locus organisation in the horse.

2. MATERIALS AND METHODS

2.1. RNA extraction from mares' milk cells

Mares' milk samples (2 × 50 mL) were freshly collected and total RNA was extracted directly from pelleted cells after centrifugation at 3 500 × g for 30 min at 4 °C essentially according to the procedure described in [3]. Pelleted cells were lysed with 0.5 mL of denaturing solution. Pellets of RNA were stored in ethanol at −20 °C. For RNA analysis, the pellet was dissolved in 50 μL distilled water.

2.2. Reverse-transcription, PCR and cDNA sequence analysis

Total RNA prepared (5 μL) from the cellular pellet was reverse transcribed into first strand cDNA as previously described [18]. PCR primers were designed from conserved regions. Owing to the high conservation of the sequence encoding the signal peptide of calcium-sensitive caseins between species, the BT71 primer was chosen to pair with this sequence. From another highly conserved region detected in the *3′-UTR* after the alignment of known *αS1-Cn* cDNA sequences from various species, the BT56 reverse primer was defined, starting 13 to 15 nt upstream of the polyadenylation signal as previously described [18]. DNA fragments obtained after PCR on reversed transcribed mRNA under standard conditions using the *αS1-Cn* BT56/71 specific primer pair were cloned into pUC18 essentially as previously described [18]. These clones were screened by PCR using the same *αS1-Cn* primer pair, BT56/71. PCR products, 5 μL of each reaction mix, were analysed by electrophoresis. A plasmid showing a *ca.* 1 kb amplified fragment was sequenced after purification with a QIAquick PCR kit (Qiagen), using the Prism ready reaction cycle sequencing kits and an ABI 377 automated sequencer (Applied Biosystems Inc.). A comparison of the resulting sequences was searched using BLAST and Align programmes.

Oligonucleotides were provided by Genosys Biotechnology (United Kingdom).

Forward primer sequences were
BT71: 5′ ATGAA(A/G)IT(T/C)(T/C)TCAT(T/C)(T/C)T(T/C)(A/G)CCTG 3′,
Eq1: 5′ AGCAGCACTTCATCAAGCGAG 3′,
Eq5: 5′ AGACAGCCAGAAATCATTCAG 3′,
Eq7: 5′ CATTCTTGATTTCTCTCCTCA 3′,
Eq*β*1: 5′ TATGTATTATACAGCATTCATCTT 3′,
Eq*β*3: 5′ GACTCAGCCATGAAGATCCT 3′.

Reverse primer sequences were
BT56: 5′ AAATCTGAGT(T/C)ACTGCACACA 3′,

Eq2: 5′ TCAGTGTTAATGGGAACAACTTC 3′,
Eq4: 5′ ACATCAAGGGCCATAATTGTG G 3′,
Eq8: ATGACACTTACAGAAGAAGCA 3′,
Eqβ2: 5′ GACTGGTTAGGAAATAGATTC 3′.

BT71, BT56, Eq1, 5, 7, 2, 4 and 8 were used to amplify cDNA and sequence it. All primers named Eq were used to amplify the genomic region.

The letters in brackets represent the degenerescence at the relevant position.

2.3. Genomic analyses

To characterize the casein locus, BAC clones containing casein genes were isolated from the Inra horse BAC library [20] by PCR screening using *αS1-Cn* (Eq7/Eq8) and *β-Cn* (Eqβ1/β2) specific primer pairs.

Clones were prepared as described previously [7, 26] and their gene content was ascertained by PCR performed on miniprepped BAC DNA with the same screening primers. PCR fragments were purified with the QIAquick PCR purification kit (Qiagen) and sequenced on an ABI 377 automated sequencer (Applied Biosystems Inc.). Homologies to sequences present in the databases were searched for using the BLAST algorithm (*www.ncbi.nlm.nih.gov*). The BAC insert size was estimated after *Not*I digestion followed by field inverted gel electrophoresis (FIGE). BAC clones containing the casein locus were used for fluorescence *in situ* hybridization (FISH) mapping as described elsewhere [1, 7].

Equine genomic DNA was prepared from peripheral blood, as previously described [7]. The Expand^{TM} long template PCR system of Boehringer-Mannheim (Meylan, France) was used according to the manufacturer's protocol except that annealing was performed at 55 °C. An aliquot (5 μL) of each reaction mix was analysed by electrophoresis, then amplified DNA (10 μL) was digested with one of the following restriction enzymes: *Pst*I, *Pvu*II under conditions specified by the supplier. The size of digested fragments was determined by electrophoresis in 1% agarose gel.

3. RESULTS AND DISCUSSION

3.1. Nucleotide sequence of the cDNA encoding equine *αS1-Cn*

To decipher the equine *αS1-Cn* cDNA primary structure, we took advantage of the presence of mammary epithelial cells in milk to isolate RNA.

After cloning of amplified fragments, recombinant plasmids were screened using *αS1-Cn* specific primers to identify a clone containing a horse *αS1-Cn* cDNA. The sequencing of the identified *αS1-Cn* cDNA insert reveals that the amplified fragment was 1 018 bp in length. The accession number of the obtained *αS1-Cn* cDNA sequence, reported here, is AY049939. Due to the strategy used, the 5′ extremity of the cDNA amplified fragment corresponds

to the sequence encoding the signal peptide and therefore this cDNA lacks 5′ UTR. The coding region stretches over 627 bp coding for a 208 amino acid (aa) residues protein including the 15 aa residues of the signal peptide. Therefore, the mature horse αS1-Cn is 193 aa residues long. Most of the 3′-UTR is included in the amplified cDNA fragment which stretches out 391 bp past the TGA stop codon.

Multiple alignments of the cDNA sequences from different species (Fig. 1), taking into account the exon modular splitting derived from the known structural organisation of genes as proposed by [19] points to constitutive horse specific exon-skipping events. Indeed, exons 5, 13 and 14 (numbering of the caprine gene) are lacking whereas the others including exons 3′, 6′ and 13′ are present. Exons 3′ and 6′ are conserved during splicing as in the rabbit whereas they are absent in ruminant mRNA. Conversely, exon 13′ which was first detected in humans [18] is present in horse *αS1-Cn* mRNA as in the pig. It is worth noting that exon 13, which is a duplication of exon 10 in ruminants, is absent in the horse cDNA sequence reported here as in humans, pigs and rabbits. This event could be due to mutation(s) in the splice site sequence surrounding this exon or to the absence of this duplication in equine species. This horse exon-skipping pattern is more similar to that exhibited by its rabbit counterpart (except for exons 13 and 14) than to its caprine or bovine counterpart. The specificity of the horse *αS1-Cn* pattern of splicing, reported here, resides in exon 5-skipping which is not observed in the caprine, rabbit, pig and human species. However, this work could not provide evidence of the absence of multiple alternative processed transcripts as described in goat [5, 14]. Moreover, due to the lack of polymorphism analysis, we cannot exclude that genetic variability exists. The *αS1-Cn* splicing pattern presented here is the first described in the horse. The present transcript most probably corresponds to the longer of the two, major αS1-Cn proteins identified in the mares' milk (Miranda *et al.*, unpublished). Regarding the multiplicity of alternative splicing described in different species for this gene (see review [19]), we can guess that other patterns could exist in the horse.

Furthermore, it is worth noting that 3 or 6 nucleotide deletions occur at the 5′ or 3′ extremities of the horse exons 4, 6 and 11 (Fig. 1). In goats, it was proposed that such events are due to the usage of cryptic 5′ or 3′ splice sites leading to the elimination of the first or the last codon of exons [14]. This event could be due to inaccurate splicing by promoting selection of cryptic splice sites and/or by equine specific mutation changing this splice site sequence. These phenomena seem to be rather frequent with *αS1-Cn* transcripts occurring in most of the species examined so far. Exon-skipping and cryptic splice site usage result in an interspecies variability of the αS1-Cn structure.

Regarding cDNA sequence comparison, different types of analysis of similarity were performed. A global analysis with the entire sequences gave a

```
                                                                                                         Exon 2
Equine                                                                  ATGAAGGTTTTCATCCTCGCCTGCCTTGTGGCTGTTGCTCTTGCCAGGCTT
Caprine           AACCCAGCTTGCTGCTTCTTCCCAGTCTTGGGTTCAAAGTCTTGACAACC-----AC--C-------TA---T--C---------------------C-
Rabbit   ATCGCCTAGATCATCAACCCAACTTGCCTCTTTTCAGTCGAGTTTAAGGCTTGGACCACC------C--C--------A-T------------AC-------------A-
Human                                                                                       T------------------------C-
Pig      TTCGATCATCAGCTCAGCTTGCTTCTTCTTTCCGGTCTTGGGTTCAAGGTGTTGACAACC-----AC--C-----T-TAT---T----CA--------C----------C-

                   Exon 3                       Exon 3'                              Exon 4
Equine     AAACTTCCTCATAGACAGCCAGAAATCATTCAG   AATGAACAGGACAGTAGAGAG     ***AAAGTCCTCAAAGAAAGAAAGTTTCCCAGTTTTGCT******
Caprine    ----A----ATC-AT--C-A--G-C--TC--CA                             ***G----------T----ATTTAC-CAGGTT-G---TGGCG***
Rabbit     ---T---A-TTAG----C-TGAA-C---C----   G-GC-G-CT--G--C--T---     CAGG--A--T-A------------C-C-T---G----TC******
Human      ----------T----T-C----A-CG-C-----   ---CC-TCA--G--C--T---
Pig        ----C-----TC--G--T-AG-A-CA-C-----   --C--G-CA-----C------     ***G--C-CTTCA---------------T---G---C--GAGGTG

                 Exon 5                          Exon 6                                 Exon 6'
Equine                                   CTAGAGTACATCAATGAACTGAAC***     AGGCAGAGAGAACTTCTGAAAGAAAAACAGAAAGATGAACACAAG
Caprine   CCTTTTCCAGAAGTGTTTAGAAAG       ***---A---------------GTAAG
Rabbit    CAGACAGTACCACTAGAATTAAGA       GAG--A--TGT-------------***     ------------------G-------G----T--G---AT----
Human     CCTATACCATTAGAATCAAGAGAG       ***--A-----G----GTA-----***     -----------CA------G-----------CT------AT----
Pig       CCTTTATTAAGTCAGTTCAGACAG       ***--AAT----------------AGG

                Exon 7                     Exon 8                           Exon 9
Equine     GACACTAGCAATGAATCAACTGAA   GAATATCTCATAGAAGACCCTGAG     CAACAGGAATCTAGCAGCACTTCATCAAGCGAG
Caprine    --T-T-G-G-G------------G   --TC-AGC---G-----TG--A--     ---AT-A--G--G-A----G---G-----T---
Rabbit     -GA-----A------GT------G   ---C--G-TT-G-C-----G----     ***ACA---G------T--GC--------T---
Human      --T-----G-----G--T---C-G   A-C-G-G-TG-G-C---G------     A-GAT------C----T--G------G--T---
Pig                                   A-TC--GG---G----G--A----     ---AG--G--------***G---------T---

                    Exon 10                              Exon 11
Equine          GAAGTTGTTCCCATTAACACTGAG    ***AAGCGCATTCCAAGGGAAGACA*********TGCTCTACCAACACACTCTG
Caprine         ---A---------A--GTG--C--    CAG---TA-----A--A------TGTGCCCTCTGA--GT----TGGGTTA----
Rabbit          ---A---------GC-G---CA--    CAG---TA-G-G------------CTGGCTTACCAA-CT---GTG--GCAG***
Human           ---A-GTC--T--G---GTG--C-
Pig             ---------GG--A--GTG---AG    CAG----A-G---A--AA-----AGATG******---C--C----AG-TA----

                    Exon 12                                Exon 13                     Exon 13'
Equine     GAACAGCTTCGCAGACTGAGCAAATACAACCAACTCCAGCTG                                 CAAGCCATCCATGCCCAG
Caprine    ----------T--------AA---------GTG-C-------      GAAATTGTTCCCAAAAGTGCTGAG
Rabbit     ***-------T----A---AAG--CG-T-----A-----GA-
Human      ------T--T-T-------A-G-------------T------                                 -----TGC----------
Pig        -G---T----AAG------A---------A-TG-G-------                                 G----T--T----A----
```

Figure 1. Sequence alignment of equine (AY049939), caprine (x59836), rabbit (x13042), human (x98084) and pig (x54973) αS1-casein cDNAs. Sequences are split into blocks to visualize the exonic modular structure as deduced from known splice junctions of the relevant genes. Reference sequence in bold is that of the horse. – indicate nucleotides identical to those of the reference sequence. * are inserted gaps introduced to maximize the alignment.

```
                 Exon 14                                                  Exon 15                          Exon 16
Equine                                                          AGAAAGCCTATGAGAGTAGTGAATCAG      GAACAGGCCTACTTCTACCTTGAG
Caprine  GAACAACTTCACAGTATGAAAGAGGGAAACCCTGCCCACCAG             -A-C---------T--C----------      ----T--------------C-C--
Rabbit                                                          ---G-------------G---------      ----T---TC-GC----T---C--
Human    GAGCAAATTCGCAGAATGAATGAAAACAGCCATGTCCAAGTG
Pig      ***GAACTTCACAGAACAAATGAGGACAAGCATACCCAACAG             G--G---------A--G----------      -----------------TT-----

                                          Exon 17
Equine   CCTTTCCAACCATCCTACCAGCTTGATGTCTATCCCTATGCTGCTTGGTTTCATCCTGCGCAAATCATGCAACATGTTGCTTACTCACCATTCCATGACACTGCCA
Caprine  -T----AG--A--T---------G----C------A-C--G---C----ACT-C-T-C-ACT-GG-ACA---T-CAC--A-GC-C-CT-----TC-----TCC-T-
Rabbit   ------G---A-C---------------C-----T-CC----C------ACT--A--C-G---G-G-----GT-----CT--C-C---TG---T-----CTC-TT-
Human    --------G-AGCT-A----A----C--C---C----------T-----ACT----ACAAATC--GCA-T-TGT-CC-TTCCCAC-GTTT-C-G-CAT-T-CA***
Pig      ---C----C-AG-T--------------C-------------A-C----A-T-----C-A*********---T--A-----C--C--TT----ACCA---TCC-TC

                                 Exon 17                                                  Exon 18
Equine   AACTCATTGCCTCTGAGAACTCGGAAAAAACTGACATTATACCAGAGTGG*T          GAAGGATTAAGTGAATTCTCAGAAACTCCACAATTATGGCCCTT
Caprine  -T-C-----G----------AGT-G---G---***AC---G---CT----*-          ----AG-C--------------G----------------T-T--
Rabbit   C--C--G-----T-----GTG-T------------G----T---------T-          ----A-------C---------G---------------A--A--
Human    -T-C--C---TCA---A--T-AT-------A-A--G-C--G-T-C-----*-
Pig      ---C--C----C-------GGGT-G---------G-----G--TC-----*-          ----A-----------------G------------------T--

                                          Exon 19
Equin  GATGTGACTGAAAAT*CCATTCT*TGA*TTTCTCTCCTCAATGTCGATCGTGT*AACCCACCTATCCCAAGGCTTCAACTGTTATTTTAGAGTAGGACAATCCCAAAT
Caprin ----G—T--------T----G-****TT-ACA-G--T-TTCATCTATCATGTCA----ATT------A----------T--C-G-------A----------T-----
Rabbit ---------------T-A-C*--A---A------AT--TTG-T-TATAATCTAA----ACTT-----A---A--****T-------C---GA-----*----T--T--
Human  ---A---T-------TT------C---A------*********CTCTCAAG-AA----AT-T----TG---A--*GG------G-------A---T-A-------T--
Pig    ---------------A----G--*---AA--TCTC-TCTCCATCTACCATGTAG----ATTT----TG---A---TG------C-------AC---GA----G-----

Equin  TGAGGGTAATCTTTCTTCCTGAGTTCTCTTCTGATATT*AGATAGCATAT*CCTCCTTTTCTTTAAGTTAAATTTTCCTGTACAGTT*ATCGTCTGAATTCCAGTTG*
Caprin ---A--C-C---------T---------CA---TAT--T----T-TG--*******ACATCC------G-----G----A*-G--C-TG-TAC—A----------A-T
Rabbit ------CT---------TT----C-A---A--CT-TAAC---C-------CAT-**AC--TC-----CA--------T-AG------T--TA---A--C--------T
Human  ---A--A---TG-----TT--G---A---A--*******TA---------CAT—T--------*---C-----------AG-G----T--T----A----T------T
Pig    C--A-TC------C----T---A----T-A--CTATA-T-----------*AA--------C--TG-CA--G--G----A*------T-GT----A----T------T

Equine  ATTATGCCAGTATGAAGGTCACTGAATCAGAGTG**TTAAAAGTCTTTACT*GGTT**CTATA*TGGGAATTTTGTTGAAAAA*TCTTTAAATTGCTTCTTCTGTAA
Caprine --C-C--TG----A----C------C---A--G-AA---C-*-----C-T-AAA--T*-----*---A--A-**---TTT---GC----G---CA-C---C------
Rabbit  --C-----------G-AAG---CTGG-T--G-G-TA-----*-------G-AA---T*--G--A---A-----C---T-A---G-----GG--G-A--T--------
Human   G-CT-----*----G---G---CT--------G-TA-----*--G------AA---TT---G**---AC--------T-A---G-----G------CAG--------
Pig     --C---------G----AC-------------G-AA---------------AAA--T*-A---*---A---------T-A---GC----G----------C------

Equine  GTGTCATCATTTCAAATAATTGTGTGCAGTAACTGAGATTT
Caprine ---C---------------C---------------------TGTCTTTCTTCTTTTCAATAAATTACATTTTAAGGCACTATT
Rabbit  ---CA-----G-------------------G----------GTTCTTTCTTC***TCAATAAATTACATTTTAAGGCA
Human   ---C-----A--A-----G-
Pig     ---C-----T--------------------G----------TTCTTCCTTCTTTTCAATAAATT
```

Figure 1. *Continued.*

degree of similarity with caprine, rabbit, human and pig full-length cDNA of *ca.* 60%. Due to the exon-skipping events, this global comparison does not reflect biological reality. The comparison of common exon sequences reveals higher degrees (*ca.* 75%) of similarity between the horse and other species. Furthermore, sequence comparisons taking into account the exon modular structure of the cDNA show a similarity ranging between 61 to 93%

depending on the exons and the species compared. In particular, the equine exon 9 sequence which encodes the highly conserved multiple phosphorylation site of the protein has around 75% similarity with its caprine, rabbit and pig counterpart. Surprisingly, similarity with human exon 9 is lower (less than 65%). It is worth noting that equine exons 15 and 18 show a high similarity (> 80%) with the other species (Fig. 1).

3.2. Partial characterization of the casein gene locus

Using PCR to screen the Inra horse BAC library, the *αS1-Cn* (Eq7/Eq8) primers identified 12 clones likely to contain the relevant gene. One BAC (593-G-3) clone was studied in more detail. The insert size of this clone, estimated after *Not*I digestion followed by field inverted gel electrophoresis (FIGE) analysis, was *ca.* 90 kb. The presence of the *αS1-Cn* gene in this clone was confirmed by sequencing of the PCR product obtained. A specific hybridisation signal obtained by FISH using this clone as a probe was observed on the equine chromosomal region 3q22.2-q22.3 (ECA3) after examination of the metaphase spread. This result is in agreement with Zoo-FISH analyses [22] and comparative mapping data [20]. Indeed, the casein locus was localised on the 6q31 bovine chromosomal region [28], on caprine 6q32 [8] and on human 4q21.1 [17], which are orthologous of ECA3.

With respect to the regulation of expression of the four casein genes, the *αS1/β* region suspected to contain *cis*-acting elements was studied. Since the *β-Cn* gene was not detected in BAC 593-G-3, a second screening of the BAC library was performed using primers specific for *β-Cn* gene (Eq*β*1/ Eq*β*2). Six clones were shown to contain the *β-Cn* gene and one (577-C-4) of them included both the *αS1-Cn* and the *β-Cn* genes. The size of the insert was estimated by FIGE to be *ca.* 82-kb. Given the close proximity of the caprine, bovine and mouse *αS1-Cn* and *β-Cn* loci [15,23,25], the size of this intergenic region in the horse was evaluated using long-range PCR. Electrophoresis of PCR products was obtained using Eq7 (matching the 5′ of the last exon of the *αS1-Cn* gene) and Eq*β*1 (matching of the exon 8 sequence of the *β-Cn* gene; accession number AF214526) revealed an amplified fragment, the size of which was estimated to be *ca.* 11.2-kb after *Pst*I, *Pvu*II digestion and gel electrophoresis. This result (Fig. 2) demonstrates that these two genes are also convergently transcribed as it was first shown in caprine [15], bovine [23], mice [25] and human [24]. Interspecies comparison of the *αS1/β-Cn* region revealed its conservation during evolution. Only the distance separating these two casein genes is quite different. In horses, the size of this intergenic region is between that of its mouse counterpart (10 kb, [25]) and its ruminant (12 kb in caprine [15] and 20 kb in bovine [23]) counterparts. In humans this

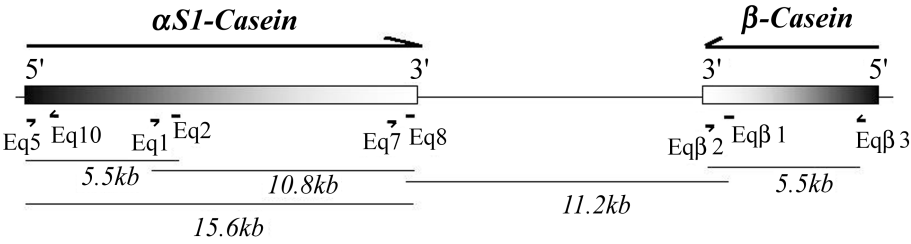

Figure 2. Relative position and orientation of αS1- and β-casein genes. The genes are represented as rectangles, orientated by arrows on the top. Names, position and orientation of primers as well as size of amplified fragments are also given.

region seems to be larger since the *αS1-* and *β-Cn* gene appeared to be 60 kb apart [24].

In the same way, the size of the *αS1-Cn* gene has been determined by long-range PCR using primer pairs Eq5 (matching exon 2)/Eq2 (matching exon 10); Eq1 (matching exon 9)/Eq8 (matching the last exon) and Eq1/Eq2 (Fig. 2). The sizes of the amplified fragments estimated after *Pst*I, *Pvu*II digestion and gel electrophoresis were *ca.* 5.5-, 10.8- and 0.7-kb, respectively. These results suggested that the size of the *αS1-Cn* gene is at least 15-kb without the region spanning exons 1 to 2. The size of this segment is quite similar to those determined for ruminants and rabbits (*ca.* 15-kb).

Lastly, the size of the *β-Cn* locus was determined by long-range PCR using the primers Eq*β*3 (matching with exon 2) and Eq*β*2 (matching with the 3′ extremity of the gene). Gel electrophoresis of the PCR product revealed that the size of this region, 5.5 kb in the horse, is at least 1 kb shorter than its ruminant counterpart. The characterisation of this gene is under investigation to determine the deletion site.

4. CONCLUSION

The results presented here are in agreement with the hypothesis that exon-skipping and cryptic splice site usage outcome in an interspecies variability of the *αS1-Cn* [19]. We have described here a horse specific pattern of *αS1-Cn* primary transcript splicing. Nevertheless, we cannot exclude the existence of other patterns due to the extensive alternative splicing of this gene described in other species. We have also analysed the genomic organisation of a part of the casein locus which presents a large similarity with the loci of the other species described so far. This knowledge provides tools to study the regulation of the casein gene cluster in equine species and the homologous cDNA sequence leads us to study the *αS1-Cn* alternative splicing in the horse.

ACKNOWLEDGEMENTS

We thank Mr Dinard for providing us with mares' milk. We express our gratitude to Anne Oustry-Vaiman for FISH experimentation; Corinne Giraud-Delville and Séad Taourit for their technical assistance in sequencing and Johann Laubier for his help in preparing PCR fragments. We would like to thank Fergal Lawless for revising the English version of the manuscript.

REFERENCES

[1] Bahri-Darwich I., Vaiman D., Olsaker I., Oustry A., Cribiu E.P., Assignment of bovine synteny groups U27 and U8 to R-banded chromosome 12 and 27, respectively, Hereditas 120 (1994) 261–265.

[2] Businco L., Giampietro P.G., Lucenti P., Lucaroni F., Pini C., Di Felice G., Iacovacci P., Curadi C., Orlandi M., Allergenicity of mare's milk in children with cow's milk allergy, J. Allergy Clin. Immunol. 105 (2000) 1031–1034.

[3] Chomczynski P., Sacchi N., Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, Anal. Biochem. 162 (1987) 156–159.

[4] Doreau M., Le lait de jument, INRA Prod. Anim. 4 (1991) 297–302.

[5] Ferranti P., Addéo F., Malorni A., Chianese L., Leroux C., Martin P., Differential splicing of pre-messenger RNA produces multiple forms of mature caprine αS1-casein, Eur. J. Biochem. 249 (1997) 1–7.

[6] Fujiwara Y., Miwa M., Nogami M., Okumura K., Nobori T., Suzuki T., Ueda M., Genomic organization and chromosomal localization of the human casein gene family, Hum. Genet. 99 (1997) 368–373.

[7] Godard S., Schibler L., Oustry A., Cribiu E.P., Guérin G., Construction of a horse BAC library and cytogenetical assignment of 20 type I and type II markers, Mamm. Genome 9 (1998) 633–637.

[8] Hayes H., Petit E., Bouniol C., Popescu P., Localization of the alpha-S2-casein gene (CASAS2) to the homoeologous cattle, sheep, and goat chromosomes 4 by *in situ* hybridisation, Cytogenet. Cell Genet. 64 (1993) 281–285.

[9] Iametti S., Tedeschi G., Oungre E., Bonomi F., Primary structure of κ-casein isolated from mares' milk, J. Dairy Res. 68 (2001) 53–61.

[10] Jenness R., Holt C., Casein and lactose concentrations in milk of 31 species are negatively correlated, Experientia. 43 (1987) 1015–1018.

[11] Jolivet G., Devinoy E., Fontaine M.L., Houdebine L.M., Sequence of rabbit αs1-casein cDNA, Nucleic Acids Res. 16 (1992) 11813–11813.

[12] Koczan D., Hobom G., Seyfert H.M., Genomic organization of the bovine αs1-casein gene, Nucleic Acids Res. 18 (1991) 5591–5596.

[13] Kurmangalinov S.M., Aïdarkhanov B.B., Sharmanov A.T., Taspolatov B.K., Chuvakova T.K., Saparov A.S., Immunomodulating properties of woman, cow and mare milks, Pediatrija 4–6 (1992) 65–68.

[14] Leroux C., Mazure N., Martin P., Mutations away from splice sites recognition sequences might cis-modulate alternative splicing of goat αs1-casein transcripts. Structural organization of the relevant gene, J. Biol. Chem. 267 (1992) 6147–6157.

[15] Leroux C., Martin P., The caprine αs1- and β-casein genes are 12-kb apart and convergently transcribed, Anim. Genet. 27 (1996) 93.
[16] Lozovitch S., Medical uses of whole and fermented mare milk in Russia, Cultured Dairy Products J. 30 (1995) 18–21.
[17] McConkey E.H., Menon R., Williams G., Baker E., Sutherland G.R., Assignment of the gene for beta-casein (CSN2) to 4q13 → q21 in humans and 3p13 → p12 in chimpanzees, Cytogenet. Cell Genet. 72 (1996) 60–62.
[18] Martin P., Brignon G., Furet J.P., Leroux C., The gene encoding αs1-casein is expressed in human mammary epithelial cells during lactation, Lait 76 (1996) 523–535.
[19] Martin P., Ferranti P., Leroux C., Addéo F., Non-bovine caseins: quantitative variability and molecular diversity, in: Fox P.F. (Ed.), Advanced Dairy Chemistry: Proteins, 3rd ed., Kluwer Academic/Plenum Publishers, New York, 2002, in press.
[20] Milenkovic D., Oustry-Vaiman A., Lear T.L., Billault A., Mariat D., Piumi F., Schibler L., Cribiu E.P., Guérin G., Cytogenetic localization of 136 genes in the horse: comparative mapping with the human genome, Mamm. Genome 13 (2002) in press.
[21] Miranda G., Mahé M.F., Leroux C., Martin P., Proteomic tools to characterise the protein fraction of equine milk, in proceedings of the Milk protein conference, Vinstra, Norvège, March 30–April 2, 2000.
[22] Raudsepp T., Frönicke L., Scherthan H., Gustavssin I., Chowdhary B.P., Zoo-FISH delineates conserved chromosomal segments in horse and man, Chromosome Res. 4 (1996) 218–225.
[23] Rijnkels M., Kooiman P.M., de Boer H.A., Pieper F.R., Organisation of the bovine casein gene locus, Mamm. Genome 8 (1997a) 148–152.
[24] Rijnkels M., Meershoek E., de Boer H.A., Pieper F.R., Physical map and localization of the human casein gene locus, Mamm. Genome 8 (1997b) 285–286.
[25] Rijnkels M., Wheeler D.A., de Boer H.A., Pieper F.R., Structure and expression of the mouse casein gene locus, Mamm. Genome 8 (1997c) 9–15.
[26] Schibler L., Vaiman D., Oustry A., Guinec N., Dangy-Caye A.L., Billault A., Cribiu E.P., Construction and extensive characterization of a goat Bacterial Artificial Chromosome library with threefold genome coverage, Mamm. Genome 9 (1998) 119–124.
[27] Swaisgood H.E., Review and update of casein chemistry, J. Dairy Sci. 76 (1993) 3054–3061.
[28] Threadgill D.W., Womack J.E., Genomic analysis of the major bovine casein genes, Nucleic Acids Res. 18 (1990) 6935–6942.