PAPER

Identification and characterization of the donkey *CSN1S2* I and II cDNAs

Gianfranco Cosenza,[1] Alfredo Pauciullo,[1] Anna Lidia Annunziata,[1] Andrea Rando,[2] Lina Chianese,[3] Donata Marletta,[4] Gabriella Iannolino,[5] Davide Nicodemo,[1] Dino Di Berardino,[1] Luigi Ramunno[1]

[1]Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali,Università di Napoli "Federico II", Portici, Italy
[2]Dipartimento di Scienze delle Produzioni Animali, Università della Basilicata, Potenza, Italy
[3]Dipartimento di Scienza degli Alimenti, Università di Napoli "Federico II", Portici, Italy
[4]Dipartimento di Scienze Agronomiche, Agrochimiche e delle Produzioni Animali, Università di Catania, Italy
[5]Istituto Sperimentale Zootecnico per la Sicilia, Palermo, Italy

Abstract

The αs2 casein, encoded by the *CSN1S2* gene, is one of the three Calcium sensitive caseins present in the milk of ruminants of zootechnical interest and in the milk of *Equidae* species (horse and donkey). In the present study, we cloned, sequenced and analysed two different donkey *CSN1S2* cDNAs that we called *CSN1S2* I and *CSN1S2* II. The first, which spans over a fragment of 1016 nt, is constituted by 19 exons and encodes for a predicted protein (called αs2-I) of 221 aminoacids; the second, of which we determined the entire sequence (16 exons), encodes for a predicted peptide (called αs2-II) of 168 aminoacids. Alternative splicing and genetic markers are reported for both genes.

Introduction

In the last years, the request for donkey's milk is substantially increased, mainly because it is used in the diet of children affected by allergy to the cow's milk proteins (Businco *et al.*, 2000). Compared to humans, the donkey's milk has a similar content in casein (5.8 g/L and 6.6 g/L, respectively) and a higher content in whey-proteins (7.5 g/L vs. 2.1 g/L, respectively) (Vincenzetti *et al.*, 2008).

Recently, analogously to what already observed for the horse (Ochirkhuyag *et al.*, 2000) and pony (Miranda *et al.*, 2004), the presence of all four casein fractions α_{s1}, α_{s2}, β e κ-CN, has been demonstrated in the donkey's milk (Mauriello *et al.*, 2009). Compared to the caseins, the whey-proteins are more investigated. In particular, the aminoacid sequences of α-lactoalbumin (Giufrida *et al.*, 1992), lysozime (Godovac-Zimmermann *et al.*, 1988b) and β-lactoglobulin I and II (Godovac-Zimmermann *et al.*, 1988a, 1990) have been yet determined.

Concerning the nucleotide sequences, only the cDNA of the serum albumin (EMBL n° AY754333), the αs1-casein encoding gene (*CSN1S1*) (EMBL n° FN386610) and the partial sequence of the k-casein encoding gene (*CSN3*) (EMBL n° EU429803 and EU448385) are available in GenBank.

In the present paper, we report on the identification of the complete sequences of two Ragusana donkey cDNAs of the αs2-casein, named *CSN1S2* I and *CSN1S2* II, as well as on the detection of polymorphisms at these *loci*.

Ragusana breed developed from breeding Pantelleria males to Martina Franca females. Ragusana donkeys are good draught animals, mainly used for mule production. The coat is dark bay with light underparts. The hoofs are strong, the animals are resistant against diseases and have a strong, vigorous character. Since 1990 the breed is registered as endangered autochthonous. Nowadays, in Sicily there are over 1,400 breeding females and 60 active males under natural mating system, distributed in all provinces of the island. There are small groups under selection in other regions of Italy.

Materials and methods

DNA and mRNA samples

For this study, 3 unrelated female donkeys were used, belonging to the Ragusana breed, at comparable age (about 6 years old), type of feed (hay and concentrate), diet, feeding level and lactation stage (about 60 days), kept in individual indoor stalls and reared in the same breeding farm located in South Italy. Genomic DNA and total RNA were extracted from somatic cells present in the milk by using Nucleospin Blood and NucleoSpin® Extract Kits (Macherey-Nagel).

Corresponding author: Prof. Luigi Ramunno, Università degli Studi di Napoli "Federico II", Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, Portici, Napoli, Italy.
Tel. +39.081.2539004 – Fax: +39.081.7762886.
E-mail: ramunno@unina.it

Key words: Donkey, αs2-casein, cDNA, *CSN1S2* I, *CSN1S2* II.

Acknowledgments: the Ministry of Agricultural and Forestry Policies (MiPAF, Italy) financially supported this research (SELMOL project).

Received for publication: 21 September 2009.
Accepted for publication: 28 January 2010.

Italian Journal of Animal Science 2010; 9:e40
doi:10.4081/ijas.2010.e40

Primer design, RT-PCR conditions for amplification and cloning of the donkey *CSN1S2* I e *CSN1S2* II cDNAs

Reverse Transcription (RT) was performed by using the ImProm-IITM Reverse Transcriptase (Promega) and oligo dT for priming.

For the amplification of the cDNA of the donkey *CSN1S2* I e *CSN1S2* II genes the following forward primers were used: *CSN1S2*-F2 (5'-GACTAATAACCATGAAGTTC-3') and PonyF (5'-TCCTCCAGTGAGGAA-3'), designed respectively on the basis of the regions encoding for the exon 2 of bovine *CSN1S2* gene (EMBL NM_174528) and the nucleotidic sequence deduced from the N-terminal amino acid sequence of pony αs2-casein (exon 3-4) (Miranda *et al.*, 2004).

For both protocols, the 100 μL PCR reaction mix comprised: 20 μL of RT reaction product, 50 mM KCl, 10 mM Tris–HCl (pH 9.0), 0.1% Triton X-100, 2 mM $MgCl_2$, 10 pmoL of each primer, dNTPs each at 0.2 mM, 5 U of Taq DNA Polymerase (Promega, Madison, WI, USA). The amplification protocol consisted of 39 cycles: the first cycle involved a denaturation step at 97°C for 2 min, an annealing step at 57 or 50°C for 30 s and an extention step at 72°C for 1 min and 30 s. The next 37 cycles were performed under the following conditions: 94°C for 30 s, 57 or 50°C for 30 s and 72 °C for 1 min and 30 s. In the 39th cycle, the extension step was carried out at 72°C for 10 min.

The amplified products were first analysed by electrophoresis on 3% agarose gel in TBE buffer and then cloned in pCR2.1-TOPO plasmid by using the TOPO TA cloning kit (Invitrogen, Pero, Milano, Italy).

Clone screening by PCR and sequencing

The clone screening by PCR was performed using the following pairs of primers: M13 forward (5-CAGGAAACAGCTATGAC-3') and M13 reverse (5-GTAAAACGACGGCCAG-3'), both designed on the plasmid sequence. The amplification program was based on a first step of cell lysis and nuclease inactivation at 94°C for 10 min; subsequently, the following 30 reaction cycles were performed as follows: the first 29 cycles were characterized by denaturation at 94°C for 1 min, primers annealing at 55°C for 30 s and extension at 72°C for 1 min and 30 s. The last cycle involved denaturation at 94°C for 30 s, annealing at 55°C for 30 s and extension at 72°C 10 min long. The reaction took place in 25 μL of mix which included: 0.5 μL of over night fluid cultures of transformed cells, 50 mM KCl, 10 mM Tris–HCl (pH 9.0), 0.1% Triton X-100, 3 mM $MgCl_2$, 5 pmol of each primer, dNTP 200 μM each, 2.5 U of Taq DNA Polymerase, 0.04% BSA. All the amplified fragments were analysed by electrophoresis on 3% agarose gels in TBE buffer 1X and stained with ethidium bromide.

Plasmids from positive colonies were sequenced on both strands at the Researches Centre CEINGE – Advanced Biotechnologies (Napoli, Italy).

Primer design for amplification of the donkey *CSN1S2* I promoter/exon 2 and donkey *CSN1S2* II promoter/exon 3

For amplification of the DNA tract between the promoter and the 2nd exon of the *CSN1S2* I gene, the forward primer *CSN1S2*-5'F (5'-TGTTCTAAATCAACCTGAGT-3') was designed on the basis of the highly conserved region obtained by comparing the promoter region of the gene available in GenBank for the following species: buffalo, EMBL EF066480, cattle, EMBL M94327, goat EMBL AJ249789. The reverse primer *CSN1S2*-R2 (5'-AGAACTTCATGGTTATTAGTC-3') was designed on the obtained donkey *CSN1S2* I cDNA sequence relative to the exon 2.

For amplification of the DNA tract between the promoter and the 3rd exon of the *CSN1S2* II gene, the forward primer *CSN1S2*II-5'F (5'-TCATGTTCTAAATCAAACTG-3') was designed on the basis of the highly conserved region obtained by comparing the promoter region of the gene available in GenBank for the previously mentioned species. The reverse primer *CSN1S2*-R3 (5'-ACCTCACTGGAGGA-3') was designed on the obtained donkey *CSN1S2* II cDNA sequence relative to the exon 3. Additional primers, designed on newly determined intron sequences, were also used for sequencing. A typical 50 μL reaction mix comprised: 100 ng of genomic DNA, 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 3 mM MgCl2, 200 nmol of each primer, dNTPs each at 400 μM, 2.5 U of Taq DNA Polymerase (Promega, Madison, WI, USA), 0.04% BSA.

The amplification programs consisted of 31 cycles. The first one characterised by a denaturation at 97°C for two min, a primers annealing at 53°C for 45 sec and an extension step at 72°C for 2 min. The next 30 cycles involved a denaturation step at 94°C for 45 sec, annealing at 53°C for 45 sec and extension at 72°C for two min with the exception that in the last cycle the extension time was 10 min long. All the amplified fragments were analysed by electrophoresis on 1-2% agarose gels (Biorad) in TBE buffer and stained with ethidium bromide. Before nucleotide sequencing, PCR products were purified with QIAquick columns (QIAGEN).

Bioinformatics and molecular modeling of the protein

The homology among the aminoacid sequences of the αs2-casein for donkey I (EMBL FM946022), cattle (EMBL NM_174528), buffalo (EMBL FM865618), sheep (EMBL X03238), goat (EMBL X65160), camel (EMBL AJ012629), donkey A (EMBL FM946022), donkey II (EMBL FN298386), rab-

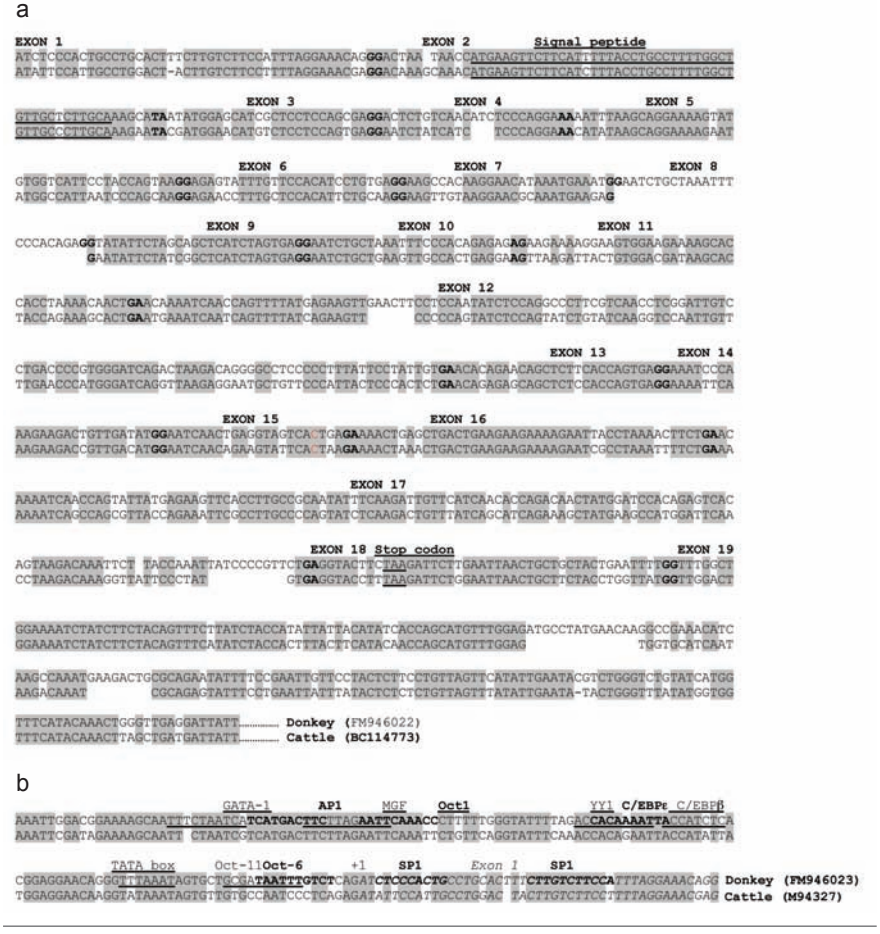

Figure 1. a) Alignment of the cDNA sequences of the s2-caseins from the donkey (s2-IA) and cattle. Signal peptide and stop codon are underlined. The possible sites of exon-exon junctions are shown in bold type. Conserved nucleotides are shown in shade. b) Promoter sequence of the donkey *CSN1S2* I gene and comparison with the bovine counterpart. Congruent and putative factors are underlined, in shaded bold letters and exon 1 is in italics. Alignment was performed using DNAsis pro Software v2.0 (Hitachi) in combination with manual adjustment by eye.

bit A (EMBL NM_001082400), rabbit B (NM_001082401), guinea pig (EMBL X00374), rat A (EMBL NM_001105741), rat B (EMBL NM_001082401) and mouse B (EMBL NM_009973) deduced from the respective mRNA sequences was performed by DNAsis pro Software v2.0 (Hitachi). The same software was used for the homologies ad the DNA level.

The transcription binding site prediction was carried out by AliBaba2 program: http://www.gene-regulation.com/pub/programs/alibaba2/index.html; Pairsim to known sites: 50; Match width in bp: 10; Minimum number of sites: 4; Minimum match Conservation: 75%; Similarity of sequence to match 1%; Factor class level: 4 (e.g. RAR-B').

Multiple sequence aligned and Phylogenetic analysis

Multiple sequence aligned and phylogenetic analysis were performed using the evolutionary distances computed by the maximum composite likelihood method in MEGA 4 (Tamura *et al.* 2007). The evolutionary distances were computed using the Poisson correction model and are in the units of the number of amino acid substitutions per site. A bootstrap analysis based on 1000 iterations was used to evaluate the robustness of the tree. The aminoacid sequences for the αs2-casein were obtained from the EMBL sequence database (see paragraph *Bioinformatics and molecular modeling of the protein*).

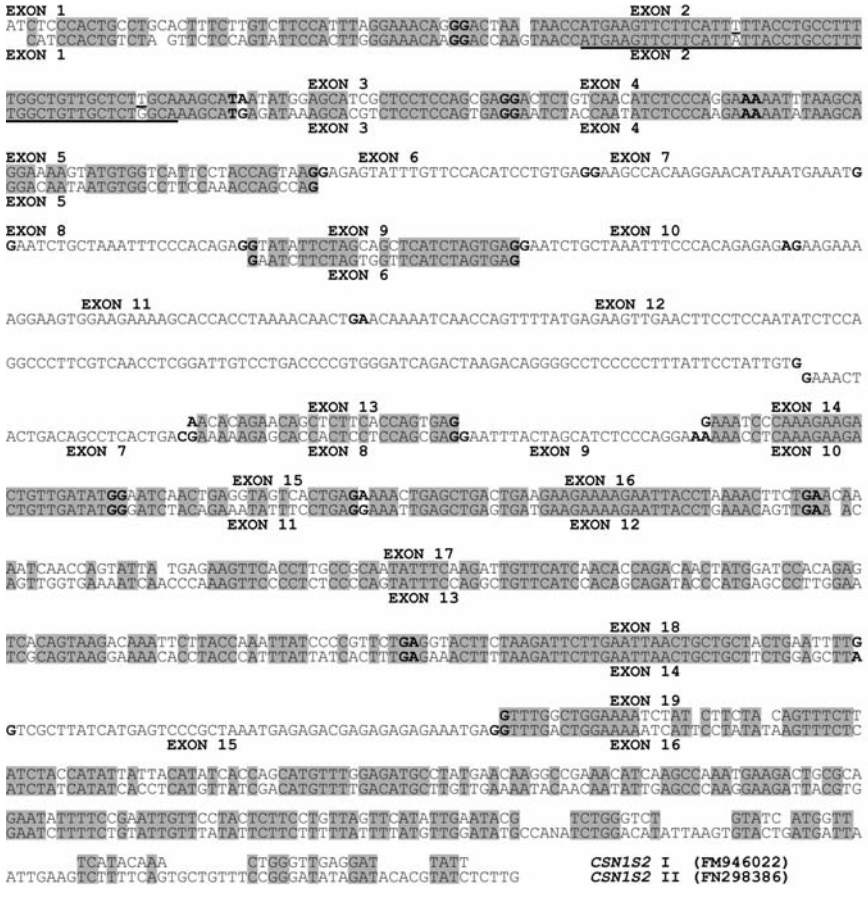

Figure 2. Alignment of the donkey *CSN1S2* I and CSN1S2 II cDNA sequences. Alignment was performed using DNAsis pro Software v2.0 (Hitachi) in combination with manual adjustment by eye. Signal peptide and stop codon are underlined. The possible sites of exon-exon junctions are shown in bold type. Conserved nucleotides are shown in shade.

Results and discussion

Identification and structural analysis of the CSN1S2 I and CSN1S2 II genes

Amplicons obtained by the retrotranscription of three donkey mRNA samples with oligo dT and the CSN1S2-F2 as forward primer were cloned, screened and sequenced.

Length analysis of the cloned fragments was realized on a total of 90 clones (about 30 each sample). Sequencing results of 15 clones (5 each sample) showed the presence of two cDNA populations, respectively 971 bp and 926 bp long.

Multiple sequence alignment and the exon subvision on the basis of the sequence of the bovine *CSN1S2* gene (EMBL no. BC114773) showed that the two cDNAs obtained are attributable to the cDNA of the gene encoding for the αs2-casein, thus we named it *CSN1S2* I.

Analysis of the obtained sequences showed that both transcripts originate from the 1st nucleotide of the exon 2 and that they differ for a deletion event of 45 bp.

By using the overlapping strategy for the PCR fragments obtained both by the amplification of the mRNA (2nd exon – last exon) and the genomic DNA (5' flanking region – 2nd exon), it was possible to establish the whole cDNA structure of the donkey *CSN1S2* I gene (EMBL no. FM946022) as well as the nucleotide sequence of the promoter region (134 bp) apart from the first 630 bp of the 1st intron (EMBL no. FM946023).

On the basis of comparative analysis of genomic and cDNA sequences of cattle and, analogously to what already observed for the *CSN1S2* B in humans (Rijnkels, 2002), the entire donkey *CSN1S2* I cDNA spans over a fragment of 1016 nt and it is constituted by 19 exons, whose size ranges from 24 bp (exons 4, 8 and 15) to 156 bp (exon 19) (Figure 1a).

The 5' untranslated region comprises the whole first exon and the first 11 nt of the secon exon, whereas the last two exons contain the major part of the 3' UTR. The signal peptide (45 bp) is coded from the 12th to the 56th nt of the exon 2, while the stop codon is located at the 10th-12th nt of the 18th exon.

The donkey *CSN1S2* I cDNA sequence shows an identity of about 74% with the corresponding bovine sequence. The main differences are due to: i) the presence of an additional exon of 24 bp inserted between the 7th and the 8th exon, which is the result of a perfect duplication of the exon 10; ii) the insertion of a triplette between the nucleotide 12 and 13 of the 4th exon; iii) the insertion of two triplettes between the nucleotide 29 and 30 of the 10th exon. Furthermore, the exon 17 is characterized by a deletion of a single nucleotide in position 108. Such event is responsible for a

frameshift of the sequence, which is restored 8 nucleotides down by an insertion of 10 bp. Other differences are located in the 3'UTR (exon 19) (Figure 1a). The second population of the observed mRNAs, whose incidence was around 10%, is characterized by a deletion of the sequence relative to the exon 11.

The RT-PCR products of the same three mRNA samples obtained by using as forward the primer PONYF have been cloned and sequenced. The screening of the positive clones was realized on a total of 120 clones (approximately 40 for animal). The sequencing of the insert of 24 clones (8 each animal) evidenced 4 populations of cDNA of 756, 732, 729 and 711 bp, respectively.

Multiple sequence alignment and the exon subdivision on the basis of the sequence of the donkey *CSN1S2* I gene showed that the four fragments are attributable to cDNA of the gene coding for a second αs2-casein enconding gene, called *CSN1S2* II.

The four *CSN1S2* I cDNA populations detected originated from the 16th nt of 3rd exon and differ from each other for deletions of 24, 27 and 45 bp.

Analogously to what has been done for the *CSN1S2* I gene, in order to determine the complete exonic sequence of the *CSN1S2* II, we amplified the DNA tract from the promoter to the 3rd exon (nearly 3000 bp).

The analysis of the obtained sequences allowed us to determine the sequence of the 1, 2 and 3 exons of such a gene, as well as the partial sequences of the 1 and 2 introns.

The *CSN1S2* II cDNA spans over a tract of 876 nt, with an homology of nearly 52 % to the corresponding cDNA *CSN1S2* I.

Analogously to the human *CSN1S2* A (Rijnkels, 2002), the *CSN1S2* II gene is characterized by the presence of 16 exons ranging in size from 24 bp (exons 7, 9 and 11) to 257 bp (exon 16). Compared to the *CSN1S2* I cDNA, this gene is also characterized by the lack of sequences that in the form I correspond to exons 6, 7, 8, 10, 11, 12, by the insertion up and downstream in the 13th exon (8th exon for *CSN1S2* II) of two sequences of 24 bp, the second of which (*CSN1S2* II exon 9) would seem a duplicate of the 4th exon, and by an extra exon (*CSN1S2* II exon 15) between the *CSN1S2* I 18th and 19th exon which is characterized by a GA microsatellite sequence. Moreover, as for the *CSN1S2* I form, the stop codon (TAA) is localized between the 10th and the 12th nucleotide of the 14th exon (corresponding to 18th exon in form I) (Figure 2).

The remaining three identified transcripts are also characterized by the deletion of exons 8, 10 and 11, respectively.

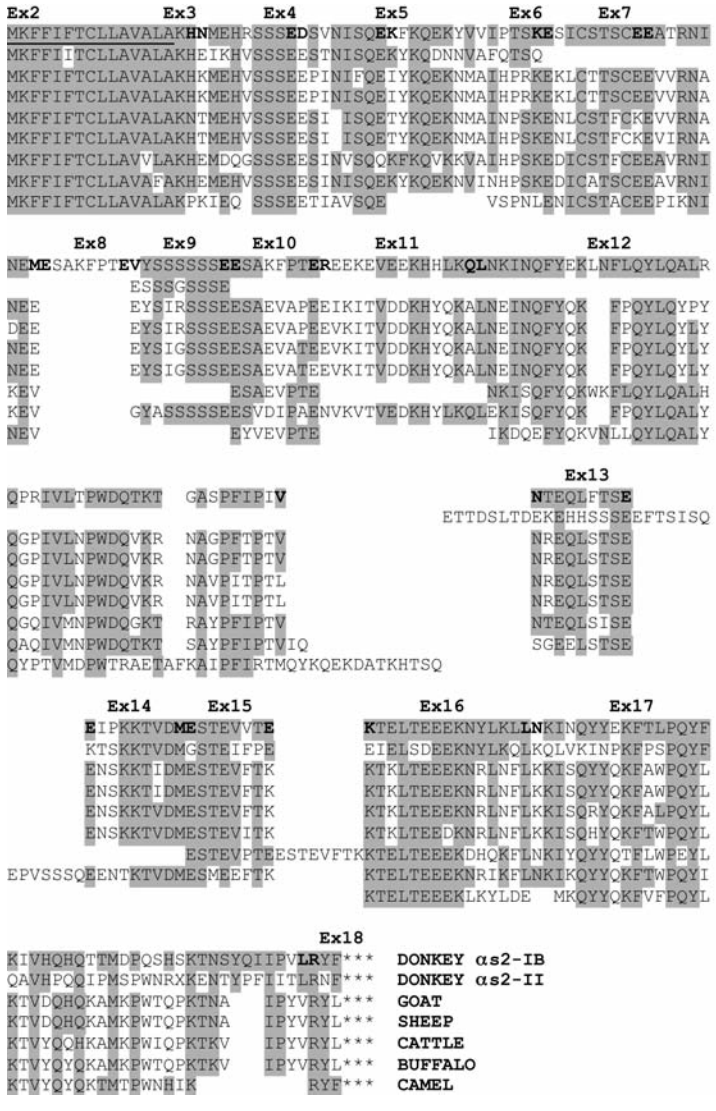

Figure 3. Multiple alignment of the predicted protein-coding regions of donkey αs2-IB and αs2-II with the αs2-caseins of the goat, sheep, cattle, buffalo, camel, pig and rabbit. The putative positions of the exon-exon splice junctions are shown in bold type. Signal peptide is underlined. Conserved aminoacids are shown in shade.

Examples of duplicated genes encoding for milk proteins are well known. In mouse, rat, rabbit and humans, comparative analysis identified two α-s2-like genes (called A and B) at conserved positions (Rijnkels, 2002; Dawson *et al.*, 1993). This finding contrasts with the bovine, ovine and caprine gene cluster that only harbors one *CSN1S2*-like gene (Rijnkels, 2002). Moreover, inactive copies (pseudogene) of the β-lactoglobulin encoding gene (Folch *et al.*, 1996; Passey and Mackinlay, 1995; Mercier and Vilotte, 1993) have been also characterized in cattle, goat and sheep, while in the horse (Conti *et al.*, 1984, Godovac-Zimmermann *et al.*, 1985) and donkeys they are expressed (Cunsolo *et al.*, 2007). A non-coding pseudogene also characterizes the locus for the bovine α-lactoalbumin (Vilotte *et al.*, 1993).

Genetic markers at *CSN1S2* I and II *loci*

The analysis of trancript sequences produced by the three subjects examined showed two mutations at the *CSN1S2* I locus. The first one, of miss-sense type, took place at the 119th nt of the 12th exon (T→C transition) (observed

genotype: 1 T/T, 1 C/C, 1 T/C) and it is responsible for the Ile→Thr aminoacidic change in 117 position of the mature protein. The second one is a substitution of silent type at the 12th nt of the 14th exon (G→A transition) (1 A/A, 1 G/G, 1 A/G). In the investigated subjects the two mutations were not found to be associated to each other. By considering only the missense mutation it's possible, therefore, to hypothesize the presence of at least two alleles at the *CSN1S2* I locus in the donkey that we named *CSN1S2* IA (Thr) and *CSN1S2* IB (Ile).

Such a mutation would seem of particular interest because the presence of the Ile in position 117 would characterize the pig, rabbit and camelidae species, whereas the presence of Thr would be specific for cattle, river buffalo, sheep and goat species.

Concerning the *CSN1S2* II gene, the sequence comparison did not show any difference at exonic level and only a (A→G) nucleotide substitution at the 197th nt of the 1st intron (1 A/A, 2 G/G).

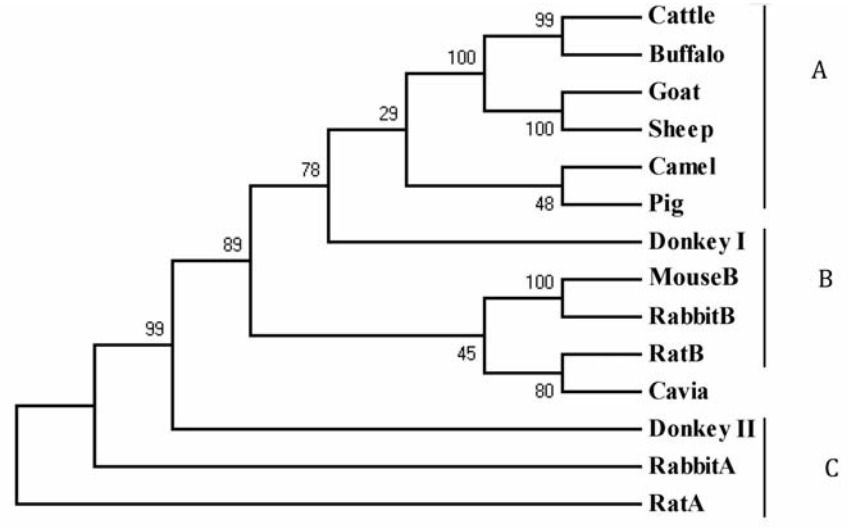

Figure 4. Phylogenetic relationship of αs2-casein aminoacid sequences. Phylogenetic tree of selected αs2-casein sequences. The MEGA v4.0 program was used to reconstruct a phylogenetic tree according to the Neighbour Joining method. Bootstrap values are indicated.

Donkey *CSN1S2* I promoter analysis

Comparative analysis of the proximal 5′-flanking region of cattle *CSN1S2* (EMBL no. M94327) and the newly sequenced donkey *CSN1S2* I (EMBL no. FM946023) genes revealed a homology of nearly 80%. In particular, within the analyzed region (167 bp), different congruent (the TATA Box, TTTAAATA) and putative binding sites for transcription factors were found in the donkey: one GATA factor (GATA-1), 2 CCAAT/enhancer-binding protein (C/EBPβ and C/EBPε), one mammary gland factor (MGF), one activating protein-1 (AP-1), two Simian virus 40 promoter factor 1 (Sp1), one Yin and Yang factor 1 (YY1) and three Octamer factor (OCT1, OCT6, OCT11). All these sequences are underlined and displayed by shaded bold letters in Figure 1b.

In silico analysis of the aminoacid sequence coded by the *CSN1S2* I and *CSN1S2* II genes

The donkey *CSN1S2* I cDNA encodes for a predicted mature protein (named αs2-I) consisting of 221 aminoacids vs. 207 in cattle (Groenen *et al.*, 1993), with a predicted MW of 26383.84 KD. The protein sequence has a discrete degree of similarity with pig (64%), sheep (60%), goat (59%), cattle (59%), buffalo (59%), camel (56%), donkey form II (43%), rabbit B (41%) and A (33%), guinea pig (39%) mouse B (36%) and rat B (29%). The peptide leader sequence was found to be identical to that of cattle, buffalo, sheep and goat (Figure 3). The deduced aminoacid sequence coded by the *CSN1S2* form II would correspond to a peptide of 168 aminoacids (that we called αs2-II) with a predicted MW of 18356.46 KD that presents an homology of about 43% with that deduced by the corresponding *CSN1S2* I cDNA sequence. The peptide leader sequence was found to be identical, with the exception of a single aminoacidic substitution, to that of the counterpart form I (Figure 3). Compared to the remaining species, the deduced aminoacid sequence showed the highest degree of homology with the sequence of the pig (43%) and a higher homology with the A form of the αs2 of rat and rabbit (34% and 41% respectively) compared to the B forms (25% e 31%, respectively).

The analysis of the deduced aminoacid sequences coded by the donkey *CSN1S2* I gene shows that the protein would be characterized by the presence of predicted major P-sites encoded by three exons homologous to exon 3, 8 and 12 in cattle. Same situation would characterize the aminoacidic sequence of the αs2-II form.

In order to show the phylogenetic relationship of the deduced aminoacidic sequences of the donkey *CSN1S2* I e II genes with those of other sequences known in other mammalian species we analyzed 14 sequences: 2 of donkey (I and II), 5 of ruminant species (cattle, river buffalo, sheep, goat and camelid), 4 of rodents (A and B forms of rat, B form of mouse, guinea pig), 2 of rabbit (A and B forms), 1 of pig.

The phylogenetic analysis showed 3 clades: A, C and B. Clade A includes 4 embranchments: sheep/goat embranchment, cattle/river buffalo embranchment, pig/camel embranchment and donkey I embranchment. The B clade includes 2 embranchments: mouse B/rabbit B embranchment and rat B/guinea pig B. These data indicate, as expected, a greater phylogenetic affinity between goat, sheep, cattle and river buffalo compared to that found between the pig and camel. The constructed phylogenetic tree also indicates that donkey αs2-II, rat A and rabbit A sequences are clustered together (Clade C) and form a separate group (Figure 4).

Such a condition is coherent with the assumption that the A forms of these rodents represent the *CSN1S2* pseudogene at such a *locus*.

Furthermore, compared to cattle, river buffalo, goat and sheep, the donkey αs2-I is characterized by a higher number of potential phosphorylation sites. In particular, three potentially phosphorylated extra Serines are found at the donkey's exon 9 (Figure 3). Same situation would characterize the partial aminoacidic sequence relative to the αs2-II form. Such a finding is in contrast with a previous report by Rijnkels (2002) who states that species with one α-s2-like gene with 2–3 P-sites (at least including E3 and E12) have only one gene, while species with two α-s2-like genes have one gene coding for two P-sites and the other coding for one P-site.

These last considerations, in conjunction with the detection of specific transcripts for the *CSN1S2* II gene, would lead to hypothesize that such a gene in the donkey, analogously to the *CSN1S2* I gene, might be functional and that it could lead to an effective translation of a protein, contrarily to what happens in

humans at the same locus (Rijnkels, 2002) and analogously to what has been already observed for the donkey β-lactoglobulin I and II (Godovac-Zimmermann *et al.*, 1988a, 1990).

Further analysis at proteomic level will be necessary in order to confirm or not the postulated hypothesis.

Conclusions

In conclusion, this research represents the first contribution to the analysis of genes coding for the milk proteins in the donkey. In particular, the main result of the present study is the identification and characterization of two forms of the CSNIS2 gene (*CSN1S2* I and II) in this species.

Furthermore, it has been possible to identify genetic markers at such loci which represent the first examples of markers known at the casein *loci* in the donkey. Such markers, prior assessment of their level of polymorphism in different donkey breeds/populations, could be useful instruments for the identification of animals, detection of errors in the paternity tests as well as differences or similarities among the animal populations, evaluation of the modifications of the genetic structure determined by the selective processes and breeding plans, utilization and conservation of animal genetic resources as well as estimation of the genetic variability within the populations. The limited knowledge about the structure of the genes coding for the donkey milk proteins as well as the reduced variability observed at proteomic and genomic levels should be attributed mainly to the paucity of the studies carried out-in so far- on this topic, which represents the true limit for the genetic improvement of this species.

References

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

Conti, A, Godovac-Zimmermann, J, Liberatori, J., Braunitzer, G., 1984. The primary structure of monomeric beta-lactoglobulin I from horse colostrum (Equus caballus, Perissodactyla). Hoppe Seylers Z. Physiol. Chem. 365:1393.

Cunsolo, V., Saletti, R., Muccilli, V., Foti, S., 2007. Characterization of the protein profile of donkey's milk whey fraction. J. Mass Spectrom. 42:1162-1174.

Dawson, S. P, Wilde, C. J., Tighe, P. J., Mayer R. J., 1993. Characterization of two novel casein transcripts in rabbit mammary gland. Biochem. J. 296:777-784.

Folch, J.M., Coll, A., Hayes, H.C., Sànchez, A., 1996. Characterization of a caprine beta-lactoglobulin pseudogene, identification and chromosomal localization by in situ hybridization in goat, sheep and cow. Gene 177:87-91

Giufrida, M.G., Cantisani, A., Napolitano, L., Conti, A., Godovac-Zimmermann, J., 1992. The amino-acid sequence of two isoforms of alpha-lactalbumin from donkey (Equus asinus) milk is identical. Biol. Chem. Hoppe-Seyler 373:931-935.

Godovac-Zimmermann, J., Conti, A., James, L., Napolitano, L., 1988a. Microanalysis of the amino-acid sequence of monomeric beta-lactoglobulin I from donkey (Equus asinus) milk. The primary structure and its homology with a superfamily of hydrophobic molecule transporters. Biol. Chem. Hoppe-Seyler 369:171-179.

Godovac-Zimmermann J., Conti, A., Liberatori J., Braunitzer, G., 1985. The aminoacid sequence of beta-lactoglobulin II from horse colostrum (Equus caballus, Perissodactyla). Structural basis for binding the retinol. Biol. Chem. Hoppe-Seyler 366:601-608.

Godovac-Zimmermann, J., Conti, A., Napolitano, L., 1988b. The primary structure of donkey (Equus asinus) lysozyme contains the Ca(II) binding site of alpha-lactalbumin. Biol. Chem. Hoppe-Seyler 369:1109-1115.

Godovac-Zimmermann, J., Conti, A., Sheil, M., Napolitano, L., 1990. Covalent structure of the minor monomeric beta-lactoglobulin II component from donkey milk. Biol. Chem. Hoppe-Seyler 371:871-879.

Groenen, M.A., Dijkhof, R.J., Verstege, A.J., van der Poel, J.J., 1993. The complete sequence of the gene encoding bovine alpha s2-casein. Gene 123:187–193.

Mauriello, R., Calabrese, M.G., Ferranti, P., Garro, G., Quarto, M., Chianese L., 2009. Chiriotti Ed. Caratterizzazione molecolare della caseina di asina. Nota 1., pp 14-17 in Proc. 9th CISETA Nat. Congr., Milano, Italy.

Mercier, J.C., Vilotte, J.L., 1993. Structure and function of milk protein genes. J. Dairy Sci. 76:3079-3098.

Miranda, G., Mahé, M.F, Leroux, C., Martin, P., 2004. Proteomic tools to characterize the protein fraction of Equidae milk. Proteomics 4:2496–2509.

Ochirkhuyag, B., Chobert, J.M., Dalgalarrondo, M., Haertlé, T., 2000. Characterization of mare caseins. Identification of αs1- and αs2-caseins. Lait 80:223-235.

Passey, R. J., Mackinly, A.G., 1995. Characterisation of a second, apparently inactive, copy of the bovine beta-lactoglobulin gene. Eur. J. Biochem. 233:736-743.

Rijnkels, M., 2002. Multispecies comparison of the casein gene loci and evolution of casein gene family. J. Mammary Gland. Biol. Neoplasia 7:327-345.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. 24:1596-1599.

Vilotte, J. L., Soulier, S., Mercier, J. C., 1993. Complete sequence of a bovine α-lactalbumin pseudogene: the region homologous to the gene is flanked by two directly repeated LINE sequences. Genomics 16:529-532.

Vincenzetti, S., Polidori, P., Mariani, P., Cammertoni, N., Fantuz, F., Vita, A., 2008. Donkey's milk protein fractions characterization. Food Chem. 106:640-649.