Lama glama α_{S1} -casein: Identification of new polymorphisms in the *CSN1S1* gene

A. Pauciullo,*†¹ M. Gauly,‡ G. Cosenza,§ H. Wagner,# and G. Erhardt†

*Department of Agricultural, Forest and Food Sciences, University of Turin, 10095, Grugliasco (TO), Italy

†Institute for Animal Breeding and Genetics, Justus Liebig University, 35390, Gießen, Germany

‡Faculty of Science and Technology, Free University of Bozen, 39100, Bozen, Italy

§Department of Agriculture, University of Naples "Federico II" 80055, Portici (NA), Italy

#Department of Obstetrics, Gynaecology and Andrology of Large and Small Animals with ambulance, Justus Liebig University, 35392, Gießen, Germany

ABSTRACT

South American camelids have been poorly genetically investigated and little information is available in llamas (*Lama glama*) regarding the diversity of the caseins at the protein and gene level. Exon skipping and duplication events previously reported in the α_{S1} -casein gene (*CSN1S1*) led us to investigate the genetic variability at this locus. Seventy-two positive clones for the α_{S1} -casein transcripts were analyzed and randomly sequenced. The comparative analysis of the sequences revealed 2 transitions, c.366A>G and c.690T>C, at the 10th nucleotide of exon 12 and 94 bp of exon 19, respectively. These SNP are responsible for 2 amino acid changes, Ile→Val in position 86 and Tyr→His in position 194 of the mature protein. Both polymorphisms clarify the genetic events behind the protein variants A and B. This result was confirmed by isoelectric focusing analysis of llama milk samples. Quick methods based on PCR-RFLP and allele-specific PCR were set up for allelic discrimination in a population of 128 animals. Based on genotyping results, 4 haplotypes were observed and the estimated frequencies indicated B as the most common haplotype (0.629) in the investigated population. These data add knowledge to the genetic variability of a species little investigated, and open opportunity for new investigation in the field of milk protein for South American camelids, including the possibility, in the future, to select alleles with favorable characteristics.

Key words: *Lama glama*, *CSN1S1*, α_{S1} -casein, haplotype

INTRODUCTION

The origin of the camelids traces back to the Eocene (40–45 million years ago), when the first ancestors of the Camelidae family were found in North America. Llamas (*Lama glama*) belong to the New World camelids (Lamini tribe), which diverged from the Old World camelids (dromedary, Bactrian, and feral camels) between 11 (Kadwell et al., 2001) and 25 million years ago (Ji et al., 2009). The Andean highlands (especially the Altiplano of southeast Peru and western Bolivia) are the natural habitat of llamas, where they are kept as multipurpose animals (mainly for transport, fleece, and meat), representing an important economic source for the rural populations of these countries and having a fundamental sociocultural role in the development of pastoral areas.

In recent years, this species has spread to other continents (North America, Europe, and Australia), where it is reared as source of hypoallergenic fibers for the international textile market and handmade clothes production. However, the llama is used also as pet or trekking animal, contributing to the development of the agro-tourism business of the European alpine zones.

In contrast to the Old World camels (dromedary and Bactrians), there is no historic tradition of milking llamas, and knowledge in this regard is still limited. In fact, whereas dromedary camel milk has been characterized with regard to its protein fractions (Kappeler et al., 1998; Shuiep et al., 2012; Pauciullo et al., 2013, 2014) and biologically active peptides (Erhardt et al., 2016), very little information is available on milk proteins from South American camelids, including llamas. Early studies on llama milk were focused on the composition in fat, protein, lactose, microminerals, and so on (Morin et al., 1995; Riek and Gerken, 2006), whereas only recently a detailed description of llama milk proteins has been provided (Saadaoui et al., 2014) and the genetic basis of casein synthesis has been elucidated (Pauciullo and Erhardt, 2015).

Received August 24, 2016.

Accepted October 17, 2016.

¹Corresponding author: alfredo.pauciullo@unito.it or alfredo.pauciullo@agrar.uni-giessen.de

The α_{S1} -CN fraction has been deeply studied in ruminants and is well characterized at both the protein and DNA level. Many alleles associated with different rates of protein synthesis have been identified in the corresponding coding gene (*CSN1S1*). At least 17 alleles have been reported in goat (Ramunno et al., 2005, Caroli et al., 2006), whereas 8 genetic variants have been reported in sheep (Amigo et al., 2000; Giambra et al., 2010a,b). Thirteen alleles corresponding to 9 α_{S1} -CN variants have been identified so far in cattle (Caroli et al., 2009), whereas at least 5 alleles have been reported at the same locus in river buffalo (Cosenza et al., 2015). Two genetic variants of α_{S1} -CN have been described in dromedary camels by Kappeler et al. (1998) and recently clarified by Erhardt et al. (2016), considering a new allele identified by Shuiep et al. (2012). Conversely, in llama, protein isoforms have been found for α_{S1} -CN by Saadaoui et al. (2014) and confirmed as an exon-skipping event by Pauciullo and Erhardt (2015), whereas 2 variants (A and B) were reported by Saadaoui et al. (2014).

To contribute to a more detailed knowledge of llama milk proteins, an investigation was undertaken to estimate the genetic diversity at protein and DNA level for α_{S1} -CN. Here we report casein variability by isoelectric focusing (IEF) and the identification of new polymorphisms at the *CSN1S1* gene. We describe methods based on PCR-RFLP and allele-specific (AS) PCR for their detection and we clarify the genetic events behind the protein variants so far detected in *Lama glama*. Furthermore, our investigation sought to establish the frequencies of the alleles in a llama population, representative of the species distributed throughout the world.

MATERIALS AND METHODS

Sample Collection and Nucleic Acid Isolation

Milk samples were collected from 2 unrelated lactating llamas reared in Lower Saxony State (Germany) and belonging to 1 farm. The animals, approximately 5 yr old and at the third lactation, were comparable for the type of diet and lactation stage (mo 4). Total RNA was isolated from milk somatic cells using InviTrap Spin Cell RNA Mini Kit (Stratec Molecular GmbH, Berlin, Germany), whereas the genomic DNA was isolated according to NucleoSpin Tissue (Macherey-Nagel, Düren, Germany).

An additional 64 llama milk samples were collected in German farms for IEF screening; for DNA genotyping, further blood samples were collected from 128 llamas belonging to different flocks in Germany, the Netherlands, Switzerland, and Italy, countries that regularly

import animals from Peru and Chile. Therefore, the structure of the investigated population was considered as representative of the world llama population.

Isolation of DNA was done according to the procedure described by Sambrook et al. (1989). The isolated DNA was resuspended in 100 μ L of TE buffer pH 7.6 (10 mM Tris, 1 mM EDTA). Both RNA and DNA concentrations and optical density 260/280 ratios of the samples were measured with the Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA).

IEF of Milk Proteins

The variability of llama milk protein was investigated by IEF. According to Erhardt (1989), a 0.3 mm polyacrylamide gel (T = 5%, C = 3.75%, and 7.1 M urea) and the following carrier ampholytes were used: 0.4% (wt/vol) Servalyte pH 3.0 to 5.0 (Serva Electrophoresis, Heidelberg, Germany), 1.1% (wt/vol) Pharmalyte pH 4.2 to 4.9 (GE Healthcare Europe GmbH, Freiburg, Germany), and 1.2% (wt/vol) Servalyte pH 5.0 to 7.0 (Serva Electrophoresis). Prefocusing was carried out at 20 mA (3,000 V) for 14 min, followed by milk and standard samples separation at 40 mA (3,000 V) for 67 min.

Reverse Transcription-PCR, Cloning, and Sequencing

The reverse transcription of total RNA was performed by using an oligo dT₁₈ (Microsynth AG, Balgach, Switzerland). The mix was set up in a final volume of 20 μ L using Verso Reverse Transcriptase (Thermo Fisher Scientific Inc.) according to the standard protocol recommended by the firm. The PCR was performed according to Pauciullo and Erhardt (2015) using primers (forward: 5'-AGTTTGCTGCTTCTTCCC-3'; reverse: 5'-ATGGCAGTTACAGGAGAAG-3') corresponding to nucleotides 21 to 38 of the first exon and complementary to the nucleotides 295 to 313 of the 21st exon of the llama *CSN1S1* cDNA (EMBL ID: LK999986).

The following thermal conditions were applied for PCR amplification: 95°C for 4 min, 35 cycles at 95°C for 45 s, 61°C for 45 s, 72°C for 60 s, and the final extension at 72°C for 5 min. The amplified products were first analyzed by electrophoresis on 1.5% agarose gel in 0.5× Tris-borate-EDTA buffer and then cloned into pGEM-T Easy Vector (Promega, Madison, WI). The ligation products were transformed into JM109 High-Efficiency Competent Cells (Promega) following the manufacturers' guidelines. White recombinant clones were randomly chosen and screened by PCR using standard vector primers M 13. All PCR products were purified using MSB Spin PCRapace kit (Stratec,

Birkenfield, Germany) and sequenced in both directions using BigDye chemistry (Applied Biosystems, Foster City, CA) by ABI 3130 Genetic Analyzer (Applied Biosystems).

Genotyping of Exon 12 c.366A>G by MaeII PCR-RFLP

The entire panel of 128 animals was genotyped for the SNP c.366A>G using the PCR-RFLP method. A DNA fragment 153 bp long (estimated according to EMBL Acc. N. KN269544.1) spanning from bp 2 of exon 11 to the 36th nucleotide of exon 12 of the llama *CSN1S1* gene was amplified by using the primers (forward) 5'-AAGTTGTTTCCAGTACCAC-3' and (reverse) 5'-TTGGGAGGGCATATCTT-3', as well as the same PCR mixture conditions reported by Pauciullo and Erhardt (2015). Thermal conditions were 95°C for 4 min, 35 cycles at 95°C for 30 s, 60°C for 30 s, 72°C for 20 s, with the final extension at 72°C for 5 min. Product specificity was confirmed by Midori Green Advance (Nippon Genetics, Dürren, Germany) stained by 1.5% agarose gel electrophoresis.

Ten microliters of each PCR amplification was digested with 5 U of MaeII endonuclease (5'-A↓CGT-3'; New England Biolabs, Ipswich, MA) overnight at 65°C. The digested products were analyzed by electrophoresis in 3.5% agarose gel in 0.5× Tris-borate-EDTA buffer and stained with Midori Green Advance (Nippon Genetics).

Genotyping of Exon 19 c.690T>C by AS-PCR

A multiplex PCR-based assay was set up for the efficient discrimination of the SNP c.690T>C in the llama population aforementioned. A fragment of 327 bp was amplified in AS reaction using the following common reverse primer 5'-GCCATCATCAGAATGAATTCACCT-3' (complementary to nucleotides 218–242 of intron 19, EMBL acc. No. KN269544.1) and 2 forward primers: 5'-TTGCTCACCCCTCATCCT-3', specific for the thymine, and 5'-TTGCTCACCCCTCATCCC-3', specific for the cytosine. In the same reaction tube, a further PCR amplification (496 bp long) was accomplished for the *CSN3* gene (κ-CN) with the primers (forward) 5'-CACAAAGATGACTCTGCTATCG-3' and (reverse) 5'-AGAAGTTAGCCCTCACAT-3' (Pauciullo et al., 2013). The latter PCR product served as positive control.

The PCR mixture was prepared according to Pauciullo and Erhardt (2015), whereas the thermal program was identical to the aforementioned genotyping conditions with the exception of the annealing temperature set up at 66.5°C. The amplified products were then

analyzed by electrophoresis on 2.5% agarose gel and stained with Midori Green Advance (Nippon Genetics).

Bioinformatic and Statistical Analysis

Homology searches, comparison among sequences, and multiple alignments were achieved by DNAsis-Pro (Hitachi Software Engineering Co., San Bruno, CA). The same software was used to estimate the molecular mass of the mature proteins and their isoelectric points. Allelic frequencies and Hardy-Weinberg equilibrium (χ^2 test) were evaluated for both SNP. Haplotype frequencies and the background recombination rate (ρ) were estimated by PHASE ver.2.1 (Li and Stephens, 2003).

RESULTS AND DISCUSSION

Seventy-two positive clones were screened by PCR (36 for each of the investigated llamas) and randomly chosen for sequencing after the evaluation of their length by gel electrophoresis. The comparative analysis of the correctly assembled cDNA sequences of the llama *CSN1S1* gene (53 out of 72, representing the 73.6% of the analyzed clones) showed 2 transitions, A→G at the 10th nucleotide of exon 12 and T→C at the 94th nucleotide of exon 19. Taking as reference the complete cDNA sequence published by Pauciullo and Erhardt (2015), these mutations (c.366A>G and c.690T>C) are responsible for 2 amino acids replacements of the mature protein Ile→Val in position 86 and Tyr→His at position 194.

The first amino acid change has no effect on the polarity of the mature protein, whereas the second slightly shifts the predicted isoelectric point (pI) from 4.78 to 4.85 as consequence of the different chemical properties of the involved amino acid, from neutral (tyrosine) to basic (histidine). As pI and charge of milk proteins are important for their interactions and properties (Khaldi and Shield, 2011), protein variants were submitted to the NetSurfP ver. 1.1 software (<http://www.cbs.dtu.dk/services/NetSurfP/>) to predict the secondary structure and detect potential modifications of the surface accessibility for the investigated amino acid. Higher absolute surface accessibility was detected for the Ile⁸⁶ (29.12) and Tyr¹⁹⁴ (54.96) variants versus the Val⁸⁶ (18.17) and His¹⁹⁴ (46.75) counterparts; however, no changes in the amino acid class assignment came out (all buried) and no secondary structure alteration was predicted. This result suggests that these mutations do not affect the protein function.

On the other side, the analysis of whole llama milk by IEF from 66 samples showed polymorphic protein patterns corresponding to the α_{S1} -CN migration area (Figure 1). According to the position in the gel, the variants

were named from 1 to 4, with 1 being the most anodic. Three homozygous (1/1, 2/2, 3/3) and corresponding heterozygous (1/2, 2/3) variants could be observed and clearly separated in addition to variant 2/4. Estimation of frequencies resulted in 0.023, 0.757, 0.212, and 0.008 for the variants 1, 2, 3, and 4, respectively. The missing of some possible IEF patterns based on the 4 variants was mainly caused by the low frequency of profiles 1 and 4 and the low number of samples analyzed. In comparison with Saadaoui et al. (2014), our variants 2 and 3 were shown by IEF to correspond to alleles B and A, respectively. This demonstrates that further research is necessary, including exchange of samples to avoid confusion in nomenclature. In addition, we tried to link the protein diversity to the mutations (c.366A>G and c.690T>C) evidenced at mRNA level.

Milk samples from IEF showed different genotypes at DNA level for both SNP (c.366A^{Ile}>G^{Val} and c.690T^{Tyr}>C^{His}). Whereas the first mutation at exon 12 does not lead to a relevant change in terms of pI of protein, and therefore it was not possible to establish a link with IEF results, the second SNP at exon 19 was clearly linked to IEF variants. In particular, all the IEF samples carrying the allele B (variant 2) were homozygous for the thymine (c.690T) at DNA level, whereas the IEF variant 3, corresponding to allele A, showed the genotype homozygous CC.

At the level of exon 19, IEF variant 1 showed the same genotype of allele A (c.690C), whereas variant 4

showed the same genotype of allele B (c.690T). Therefore, the different IEF pattern of variants 1 and 4 might be explained as an out-spliced form of allele A and B (8 amino acids EQAYFHLE corresponding to exon 18), respectively. In fact, IEF is a powerful method that shows all protein isoforms with an effect on pI, including variants as a result of constitutive and alternative mRNA transcripts.

The polymorphisms found at exon 12 and 19 have not been described before. In fact, Saadaoui et al. (2014) reported only on 2 long variants, A (25,188.92 Da) and B (25,230.05 Da), but they did not indicate any possible amino acid change. However, such a difference in the molecular mass is explained by the amino acid replacements caused by the SNP found at the DNA level in the present work. Furthermore, the information reported by Saadaoui et al. (2014) was very useful to establish the correct haplotype of both variants (Table 1). In fact, the estimated molecular mass (Da) of the dephosphorylated α_{S1} -CN (IEF variant 3 of the present work) carrying the amino acids Val (at exon 12) and His (at exon 19) is 25,188.98, whereas the estimated molecular mass for IEF variant 2 (Ile at exon 12 and Tyr at exon 19) is 25,229.03. Both protein masses agree with the data of Saadaoui et al. (2014) for the long variants A and B (Table 1).

Genotyping methods were set up for both SNP and haplotype evaluation was carried out. In particular, the transition c.366A>G at the llama *CSN1S1* exon

Figure 1. Protein variants demonstrated at the α_{S1} -CN by isoelectric focusing of whole llama milk. Lanes 1–3 denote homozygous genotypes and lanes 4–6 are heterozygous samples. Homozygous samples are characterized by 2 major and 2 more anodically located minor bands. Color version available online.

12 creates a restriction site for the endonuclease *MaeII* (5'-A↓CGT-3'). Therefore, the digestion of a PCR product of 153 bp spanning from bp 2 of exon 11 to the 36th nucleotide of exon 12 allowed carriers for the c.366G to be identified. Therefore, the PCR product, undigested in the presence of adenine, is restricted into 2 fragments of 125 and 28 bp (Figure 2). On the contrary, the transition c.690T>C did not alter any restriction site; therefore, a genotyping method based on AS-PCR was set up to allow the identification of the different genotypes (Figure 3).

The allelic frequencies of the SNP c.366A>G and c.690T>C at *CSN1S1*, determined in 128 llamas belonging to several European farms, representative of the world population, is reported in the Table 2. Based on genotypes detected at exons 12 and 19, 4 haplotypes were observed. Haplotype B (main component of the protein variant B) accounted for the majority of the population, with a frequency of 0.629; other haplotypes (alleles A, C, and D) and their frequencies are reported in Table 1.

The estimation of the haplotype frequencies at DNA level confirmed the data of the IEF screening and protein variant B (haplotypes B and D) as the most spread variant (0.771). Conversely, the low frequency (0.227) of protein variant A (haplotypes A and C) might open the possibility for rapid directional selections and whether this variant might be associated with favorable milk protein properties. However, future studies will provide this information. Although the frequency of the haplotype A (minor component of the protein variant A) is rather low, the observation of all 4 possible haplotypes (with double homozygous animals detected in the investigated population) is an indication that the observed polymorphisms likely took place independently.

Using the general model for recombination rate variation (Li and Stephens, 2003), we estimated the background population recombination rate (ρ) at this locus as 0.166. A chance of crossovers producing recom-

bination between the alleles of the *CSN1S1* gene cannot be a rare event, especially in ruminants. For instance, allele N in goats is considered the result of an inter-allelic recombination event between *CSN1S1* A and *CSN1S1* F (Ramunno et al., 2005); likewise, allele M is generated by an interallelic recombination between the A- and B-type alleles (Bevilacqua et al., 2002; Cosenza et al., 2008). It is therefore very interesting to observe similar events also in llamas, confirming the importance of casein genes also as model for evolutionary studies.

The presence of the adenine at exon 12 and the occurrence of the thymine at exon 19 (haplotype B) of the llama *CSN1S1* might represent the ancestral condition of the gene, as they have also been found in other camelids, such as dromedary (EMBL ID: NM_001303566.1) and feral camels (XM_006188668.1). The same nucleotides have also been found in the corresponding exons 11 and 17 of ruminants such as cattle (EMBL ID: X59856), sheep (JN560175), goat (AJ504710), river buffalo (AY948385), and bison (EU862373).

With the exception of these mutations, no other differences have been found in llamas and no further polymorphisms at the DNA level are known for this gene in South American camelids. On the contrary, genetic variants for the α_{S1} -CN fraction were reported in Sudanese camels for the identification of allele C (Shuiep et al., 2012) and allele D (Erhardt et al., 2016) or Somali camels, where variants of the same casein fraction were reported by Kappeler et al. (1998). Indeed, among the other species, cattle and goat were described as the most polymorphic for the *CSN1S1* locus, and different genetic events characterized homologous exons 11 and 17 in these species. In particular, in cattle the transversion A→T at the 21st nucleotide of exon 11 characterizes the allele I responsible for the amino acid replacement Glu→Asp in position 84 (Caroli et al., 2009), whereas, in goats, the first triplet (CAG) of exon 11 is alternatively recognized as cryptic site of splice and competitively eliminated from the mRNA

Table 1. *CSN1S1* haplotypes (exon 12 to 19), frequencies, and relative characteristics of the corresponding proteins (isoelectric point and isoelectric focusing profile) ordered according to the increase in the predicted molecular mass (Da) of the mature α_{S1} -CN (EMBL Acc. No. LK999986) in *Lama glama*

Position	cDNA nucleotide amino acid	<i>CSN1S1</i> haplotype			
		A	C	D	B
Exon 12	c.366	G	A	G	A
	86	Val	Ile	Val	Ile
Exon 19	c.690	C	C	T	T
	194	His	His	Tyr	Tyr
	Molecular mass (Da)	25,188.98	25,203.00	25,215.01	25,229.03
	Isoelectric point	4.85	4.85	4.78	4.78
	Isoelectric focusing profile (protein variant)	3 (A)	3 (A)	2 (B)	2 (B)
	Haplotype frequency $\pm$ SE	0.020 $\pm$ 0.003	0.207 $\pm$ 0.009	0.142 $\pm$ 0.003	0.629 $\pm$ 0.010

Figure 2. Genotyping of *CSN1S1* c.366A>G SNP by *MaeII* (5'-A↓CGT-3') PCR-RFLP. Lane 1: AA homozygous samples; lane 3: GG homozygous samples; lane 2: heterozygous samples. In lanes 2 and 3, the 28-bp band is not visible. Lane 4 is the PCR undigested control, whereas lane L is GeneRuler 50 bp DNA ladder (Thermo Fisher Scientific, Waltham, MA).

Figure 3. Genotyping of *CSN1S1* c.690T>C SNP by allele-specific PCR. The fragment of 495 bp is common for all samples and it represents the positive control. Lane 1 shows amplicons obtained with the specific forward primer for thymine, lane 2 is products obtained with the specific forward primer for cytosine. Lane L is GeneRuler 50 bp DNA ladder (Thermo Fisher Scientific, Waltham, MA).

Table 2. Genotyping data, allele frequencies, and Hardy-Weinberg (HW) equilibrium (χ^2 test, $P < 0.05$, $df = 1$) for the SNP c.366A>G and c.690T>C at the *CSN1S1* gene in a *Lama glama* population

Position	Genotyped SNP									
	c.366A>G					c.690T>C				
	Allele frequency					Allele frequency				
	AA	AG	GG	Total		TT	CT	CC	Total	HW equilibrium, χ^2
Exon 12	Observed	28	7	128		79	21	16	116	5.2484
	Expected	35.11	3.45			69.05	40.89	6.05		
Exon 19	Observed	89.45	0.836	0.164		0.772	0.228			27.4501
	Expected									

(Ramunno et al., 2005). The latter event also characterizes llama exon 12, whose first codon is a glutamine residue (Glu83), and, therefore, the canonic site at the end of the 11th intron (tttttttag/CAG) might be alternatively recognized as splicing site (Pauciullo and Erhardt, 2015). In cattle, exon 17 might be carrier of the allele C as result of the transition A→G at 131st nucleotide, which changes a glutamine in glycine in position 192 (Caroli et al., 2009). The same exon is also polymorphic in goats, where the alleles B4, C, and E are characterized by the transition A→G at 139th nucleotide responsible for the amino acid replacement Ala→Thr in position 196 (Cosenza et al., 2008).

Many genetic variants of caseins have been identified so far at the DNA or protein level in domestic animals and analyzed for different aspects including allergenicity to humans (Lisson et al., 2014). The identification of new alleles besides A and B variants in the llama α_{S1} -CN gene (*CSN1S1*) and additional variation at the α_{S1} -CN protein level adds further knowledge on South American camelids and increases interest for one of the most represented milk proteins in Camelidae.

CONCLUSIONS

The identification of genetic diversity at the casein loci represents essential information for dairy production, as it is crucial for providing milk with different protein properties. The α_{S1} -CN gene (*CSN1S1*) has been deeply investigated in ruminants, whereas little information is available in South American camelids. The detection of novel alleles in llama *CSN1S1* adds knowledge to a research field almost completely unexplored, providing useful information not only for milk protein biodiversity itself but also for genotyping and possible future linkage analysis with other traits of economic interest in the llama. The allele frequency detected in the investigated population is an indication of a possible and rapid directional selection in favor of protein variant A. However, further investigations are required to verify the influence of these allelic variants for the casein properties.

ACKNOWLEDGMENTS

This research was financially supported by the Institute for Animal Breeding and Genetics, Justus-Liebig University of Giessen (Giessen, Germany) and by Agricola Project (PA_RIC_LOC_15_01), Department of Agricultural, Forest and Food Sciences, University of Torino, Grugliasco (Italy). We are very grateful to all llama breeders who made samples of their animals in addition available for research. The authors also thank

Sylvia Jäger, Anja Scheuermann, and Zhaoxin Wang (Institute for Animal Breeding and Genetics, Justus Liebig University, Gießen, Germany) for the excellent technical assistance and Patrice Martin (Institut National de la Recherche Agronomique, Jouy-en-Josas, France) for exchanging milk samples for standardization.

REFERENCES

- Amigo, L., I. Recio, and M. Ramos. 2000. Genetic polymorphism of ovine milk proteins: Its influence on technological properties of milk - A review. *Int. Dairy J.* 10:135–149. [https://doi.org/10.1016/S0958-6946\(00\)00034-0](https://doi.org/10.1016/S0958-6946(00)00034-0).
- Bevilacqua, C., P. Ferranti, G. Garro, C. Veltri, R. Lagonigro, C. Leroux, E. Pietrola, F. Addeo, F. Pilla, L. Chianese, and P. Martin. 2002. Interallelic recombination is probably responsible for the occurrence of a new α_{S1} -casein variant found in the goat species. *Eur. J. Biochem.* 269:1293–1303. <https://doi.org/10.1046/j.1432-1033.2002.02777.x>.
- Caroli, A., F. Chiatti, S. Chessa, D. Rignanese, P. Bolla, and G. Pagnacco. 2006. Focusing on the goat casein complex. *J. Dairy Sci.* 89:3178–3187. [https://doi.org/10.3168/jds.S0022-0302\(06\)72592-9](https://doi.org/10.3168/jds.S0022-0302(06)72592-9).
- Caroli, A. M., S. Chessa, and G. J. Erhardt. 2009. Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *J. Dairy Sci.* 92:5335–5352. <https://doi.org/10.3168/jds.2009-2461>.
- Cosenza, G., A. Pauciullo, D. Gallo, L. Colimoro, A. D'avino, A. Mancusi, and L. Ramunno. 2008. Genotyping at the *CSN1S1* locus by PCR-RFLP and AS-PCR in a Neapolitan goat population. *Small Rumin. Res.* 74:84–90. <https://doi.org/10.1016/j.smallrumres.2007.03.010>.
- Cosenza, G., A. Pauciullo, N. P. P. Macciotta, E. Apicella, R. Steri, A. La Battaglia, L. Jemma, A. Coletta, D. Di Berardino, and L. Ramunno. 2015. Mediterranean river buffalo *CSN1S1* gene: search for polymorphisms and association studies. *Anim. Prod. Sci.* 55:654–660. <https://doi.org/10.1071/AN13438>.
- Erhardt, G. 1989. κ -Kaseine in Rindermilch—Nachweis eines weiteren Allels (κ -KnE) in verschiedenen Rinderrassen. *J. Anim. Breed. Genet.* 106:225–231.
- Erhardt, G., E. S. Shuiep, M. Lisson, C. Weimann, Z. Wang, I. E. M. El Zubeir, and A. Pauciullo. 2016. Alpha s1-casein polymorphisms in camel (*Camelus dromedarius*) and descriptions of biological active peptides and allergenic epitopes. *Trop. Anim. Health Prod.* 48:879–887. <https://doi.org/10.1007/s11250-016-0997-6>.
- Giambra, I. J., L. Chianese, P. Ferranti, and G. Erhardt. 2010a. Short communication: Molecular genetic characterization of ovine α_{S1} -casein allele H caused by alternative splicing. *J. Dairy Sci.* 93:792–795. <https://doi.org/10.3168/jds.2009-2615>.
- Giambra, I. J., L. Chianese, P. Ferranti, and G. Erhardt. 2010b. Genomics and proteomics of deleted ovine *CSN1S1**I. *Int. Dairy J.* 20:195–202. <https://doi.org/10.1016/j.idairyj.2009.09.005>.
- Ji, R., P. Cui, F. Ding, J. Geng, H. Gao, H. Zhang, J. Yu, S. Hu, and H. Meng. 2009. Monophyletic origin of domestic Bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). *Anim. Genet.* 40:377–382. <https://doi.org/10.1111/j.1365-2052.2008.01848.x>.
- Kadwell, M., M. Fernandez, H. F. Stanley, R. Baldi, J. C. Wheeler, R. Rosadio, and M. W. Bruford. 2001. Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proc. Biol. Sci.* 268:2575–2584. <https://doi.org/10.1098/rspb.2001.1774>.
- Kappeler, S., Z. Farah, and Z. Puhani. 1998. Sequence analysis of *Camelus dromedarius* milk caseins. *J. Dairy Res.* 65:209–222.
- Khaldi, N., and D. C. Shield. 2011. Shift in the isoelectric-point of milk proteins as a consequence of adaptive divergence between the milks of mammalian species. *Biol. Direct* 6:40. <https://doi.org/10.1186/1745-6150-6-40>.

- Li, N., and M. Stephens. 2003. Modelling linkage disequilibrium, and identifying recombination hotspots using SNP data. *Genetics* 165:2213–2233.
- Lisson, M., N. Novak, and G. Erhardt. 2014. Immunoglobulin E epitope mapping by microarray immunoassay reveals differences in immune response to genetic variants of caseins from different ruminant species. *J. Dairy Sci.* 97:1939–1954. <https://doi.org/10.3168/jds.2013-7355>.
- Morin, D. E., L. L. Rowan, W. L. Hurley, and W. E. Braselton. 1995. Composition of milk from llamas in the United States. *J. Dairy Sci.* 78:1713–1720.
- Pauciullo, A., and G. Erhardt. 2015. Molecular characterization of the llamas (*Lama glama*) casein cluster genes transcripts (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*) and regulatory regions. *PLoS One* 10:e0124963 <https://doi.org/10.1371/journal.pone.0124963>.
- Pauciullo, A., I. J. Giambra, L. Iannuzzi, and G. Erhardt. 2014. The β -casein in camels: Molecular characterization of the *CSN2* gene, promoter analysis and genetic variability. *Gene* 547:159–168. <https://doi.org/10.1016/j.gene.2014.06.055>.
- Pauciullo, A., E. S. Shuiep, G. Cosenza, L. Ramunno, and G. Erhardt. 2013. Molecular characterization and genetic variability at κ -casein gene (*CSN3*) in camels. *Gene* 513:22–30. <https://doi.org/10.1016/j.gene.2012.10.083>.
- Ramunno, L., G. Cosenza, A. Rando, A. Pauciullo, R. Illario, D. Gallo, D. Di Berardino, and P. Masina. 2005. Comparative analysis of gene sequence of goat *CSN1S1* F and N alleles and characterization of *CSN1S1* transcript variants in mammary gland. *Gene* 345:289–299. <https://doi.org/10.1016/j.gene.2004.12.003>.
- Riek, A., and M. Gerken. 2006. Changes in llama (*Lama glama*) milk composition during lactation. *J. Dairy Sci.* 89:3484–3493. [https://doi.org/10.3168/jds.S0022-0302\(06\)72387-6](https://doi.org/10.3168/jds.S0022-0302(06)72387-6).
- Saadaoui, B., L. Bianchi, C. Henry, G. Miranda, P. Martin, and C. Cebo. 2014. Combining proteomic tools to characterize the protein fraction of llama (*Lama glama*) milk. *Electrophoresis* 35:1406–1418. <https://doi.org/10.1002/elps.201300383>.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning*. Cold Spring Harbor, New York, NY.
- Shuiep, E. S., I. J. Giambra, I. E. Yas Mohamed El Zubeir, and G. Erhardt. 2012. Biochemical and molecular characterization of polymorphisms of α_{s1} -casein in Sudanese camel (*Camelus dromedarius*) milk. *Int. Dairy J.* 28:88–93. <https://doi.org/10.1016/j.idairyj.2012.09.002>.