

Focusing on the Goat Casein Complex

A. Caroli,*¹ F. Chiatti,† S. Chessa,† D. Rignanese,† P. Bolla,† and G. Pagnacco†

*Dipartimento di Scienze Biomediche e Biotecnologie, Università degli Studi di Brescia, Viale Europa 11, 25123 Brescia, Italy

†Dipartimento di Scienze e Tecnologie Veterinarie per la Sicurezza Alimentare, Università degli Studi di Milano, Via Trentacoste 2, 20134 Milano, Italy

ABSTRACT

The analysis of casein polymorphisms in goat species is rather difficult, because of a large number of mutations at each locus, and the tight linkage involving the 4 casein genes. Three goat breeds from Northern Italy, Orobica, Verzasca, and Frisa, were analyzed at the casein complex by milk isoelectrofocusing and analyses at the DNA level to identify the majority of all known polymorphisms. The casein gene structure of the 3 local breeds at α_{S1} -casein (*CSN1S1*), β -casein (*CSN2*), α_{S2} -casein (*CSN1S2*), and κ -casein (*CSN3*) was compared with that of Camosciata, a more widely distributed breed. A new allele was identified and characterized at *CSN2* gene, which seemed to be specific to the Frisa breed. It was named *CSN2*E*, and was characterized by a transversion TCT→TAT responsible for the amino acid exchange Ser₁₆₆→Tyr₁₆₆ in the mature protein. The casein haplotype structure is highly different among breeds. A total of 26 haplotypes showed a frequency higher than 0.01 in at least 1 of the 4 breeds considered, with 12, 3, 5, and 19 haplotypes in Frisa, Orobica, Verzasca, and Camosciata breeds, respectively. Only 13 haplotypes occurred at a frequency higher than 0.05 in at least 1 breed. With the molecular knowledge of each locus, the ancestral haplotype coding for *CSN1S1*B*, *CSN2*A*, *CSN1S2*A*, and *CSN3*B* protein variants can be postulated. A protein evolutionary model considering the whole casein haplotype is proposed.

Key words: casein genotype, goat, proteome, variability

INTRODUCTION

The analysis of caseins in goat species is quite complex, because there are a large number of mutations involving the 4 coding genes. The polymorphisms have been intensively investigated both at the protein and DNA level. The 3 calcium-sensitive caseins, α_{S1} -CN, β -CN, and α_{S2} -CN, are coded by *CSN1S1*, *CSN2*, and

CSN1S2 genes, respectively, whereas κ -CN, which plays an essential role in the casein micelle stabilization (Alexander et al., 1988), is coded by the *CSN3* gene. Initial investigations into goat casein polymorphisms were carried out by Boulanger et al. (1984) with research on *CSN1S1*. Until now, 16 alleles associated with different rates of protein synthesis have been identified. On the basis of the milk content of α_{S1} -CN, the *CSN1S1* variants can be grouped into 4 classes: strong alleles (*A*, *B₁*, *B₂*, *B₃*, *B₄*, *C*, *H*, *L*, and *M*), producing almost 3.5 g/L of α_{S1} -CN each; intermediate alleles (*E* and *I*; 1.1 g/L); weak alleles (*F* and *G*; 0.45 g/L); and null alleles (*O₁*, *O₂*, and *N*) apparently producing no α_{S1} -casein (reviewed by Grosclaude and Martin, 1997; Rando et al., 2000; Chessa et al., 2003a; Ramunno et al., 2005). The evolutionary pathway of goat *CSN1S1*, first proposed by Grosclaude et al. (1994), was slightly modified by Grosclaude and Martin (1997) and Chinese et al. (1997). The *B₁* allele was considered ancestral, and 2 divergent lineages were suggested, one leading to the *A*, *G*, *O₁*, *O₂*, *I*, and *H* variants (lineage A), and the other leading to *B₂*, *B₃*, *B₄*, *L*, *F*, *C*, and *E* variants (lineage B). Later, Bevilacqua et al. (2002), including the *M* allele in the phylogeny, focused on the intragenic recombination event possibly responsible for this new variant, which could be placed in both lineages (A and B) arising from the putative ancestral allele. A similar event was also proposed for the origin of *CSN1S1*N* (Ramunno et al., 2005).

More recently, variation has been described at the other casein genes. For the *CSN2* gene, 3 variants were found to be associated with a normal β -CN content: *A*, *B* (Mahé and Grosclaude, 1993), and *C* (Neveu et al., 2002). The *C* variant was found to be predominant in Italian goat breeds (Chessa et al., 2005). Furthermore, 2 null *CSN2* alleles were identified, both characterized by mutations responsible for premature stop codons in exon 7 (Ramunno et al., 1995; Persuy et al., 1999). The 2 null alleles were named respectively as *CSN2*0'* and *CSN2*0* by Neveu et al. (2002). More recently an *SspI* PCR-RFLP, detecting a silent *CSN2* allele, was described and named *CSN2*A1* (Cosenza et al., 2005).

The alignment among the mature β -CN sequence of different species suggests that *CSN2*A* is the ancestral

Received October 30, 2005.

Accepted February 27, 2006.

¹Corresponding author: anna.caroli@unimi.it

allele in comparison with *CSN2***C* (Chessa et al., 2005). In fact, the presence of alanine at position 177 of the mature protein coded by goat *CSN2***A* has been found also in other ruminant species, whereas *CSN2***C* codes for valine at the same amino acid position.

At least 7 alleles have been identified at *CSN1S2*, associated with 3 synthesis levels. The *A*, *B* (Boulanger et al., 1984), *C* (Bouniol et al., 1994), *E* (Lagonigro et al., 2001), and *F* (Ramunno et al., 2001a) alleles were associated with a normal α_{S2} -CN synthesis level, whereas *D* and *O* were associated with lower and null synthesis levels, respectively (Ramunno et al., 2001a,b). An evolutionary pathway has been proposed (Sacchi et al., 2005) starting from the *A* variant and leading independently to the *B*, *C*, *F* alleles, each characterized by a different amino acid substitution with respect both to *CSN1S2***A* and to the bovine (Swissprot Accession number P02663) and ovine (P04654) sequences. Thus, *CSN1S2***A* may be considered the ancestral variant. Moreover, if goat *CSN1S2***A*, *B*, and *C* (P33049), and goat *CSN1S2***E* (CAC21704) sequences are compared, it can be postulated that *CSN1S2***E* variant derives from *CSN1S2***C*, because the 2 alleles share an Ile, instead of Lys, at position 167 of the mature protein (Lagonigro et al., 2001; Sacchi et al., 2005).

A total of 15 polymorphic sites were identified in domestic goat allowing the identification of 16 *CSN3* alleles corresponding to 13 κ -CN variants, and 3 synonymous mutations. Prinzenberg et al. (2005) proposed a new nomenclature for the alleles, and their subdivision in 2 groups, on the basis of the different isoelectric point, which allows identifying 2 different patterns by isoelectrofocusing (IEF), named as A^{IEF} (isoelectric point = 5.29) and B^{IEF} (isoelectric point = 5.66). A phylogeny for *CSN3* was proposed by Yahyaoui et al. (2003), Jann et al. (2004), and Prinzenberg et al. (2005). Apart from the conflicting nomenclature, the 3 papers agree that *CSN3***A* appeared later in the evolutionary pathway, and that 2 different lineages occurred. The phylogeny is complex, most probably resulting from different intragenic recombination events, similarly to *CSN1S1*.

Furthermore, goat casein genes are closely linked in a complex including in the order *CSN1S1*, *CSN2*, *CSN1S2*, and *CSN3* (Ferretti et al., 1990; Threadgill and Womack, 1990). The entire casein gene complex spans about 250 kb on chromosome 6 (Hayes et al., 1993; Popescu et al., 1996). The 2 first genes of the casein complex, *CSN1S1* and *CSN2*, are only 12 kb apart and convergently transcribed (Leroux and Martin, 1996). Due to the tight linkage among casein genes, the variability of the whole haplotype has to be accounted for when analyzing the goat caseins for biodiversity studies or breeding strategies. A previous work followed this research approach (Sacchi et al., 2005).

However, *CSN2* was not considered in the casein haplotype analysis in that study, because a typing test for *CSN2***C* allele was not yet available.

This paper aimed to focus on the goat casein complex variability, by including *CSN2* in the haplotype analysis of 3 Lombardy local breeds, Orobica, Verzasca, and Frisa, which were compared with a sample of the Camosciata, a more widely distributed breed. A new *CSN2* variant was also identified and characterized.

MATERIALS AND METHODS

Blood and milk samples were collected from local goat breeds (<http://dad.fao.org>): Frisa (n = 70), Orobica (n = 66), and Verzasca (n = 67). Goats were reared in Lombardy, a region in Northern Italy. First, milk samples were typed by IEF according to Caroli et al. (2001). Then, DNA was extracted from blood or milk by standard methods and typed by several methods. Locus *CSN1S1* was analyzed by allele-specific PCR (Jànsa Pérez et al., 1994; Cosenza et al., 2003) and PCR-RFLP (Ramunno et al., 2000); *CSN1S2* was analyzed by PCR-RFLP (Ramunno et al., 2001a,b); *CSN2* was analyzed by PCR-single strand conformation polymorphism (PCR-SSCP; Chessa et al., 2005); *CSN3* was also analyzed by PCR-SSCP (Chessa et al., 2003b). Detailed information about the methods used to detect the different alleles is shown in Table 1. The nomenclature proposed by Prinzenberg et al. (2005) was used for *CSN3* in the upcoming discussion of the results obtained.

Samples of DNA showing previously uncharacterized patterns on SSCP gels during *CSN2* typing were randomly selected for sequencing. Four samples were sequenced, 1 homozygous and 3 heterozygous for the eventual new polymorphism. Primers used for sequencing were the same used for the PCR-SSCP techniques. The PCR products were sequenced by PRIMM Srl (Milano, Italy). The nucleotide sequences and the deduced amino acid sequences were analyzed by Bioedit (Hall, 1999) software.

The Genepop program (Raymond and Rousset, 1995) was used for the evaluation of allele frequencies and deviations from Hardy-Weinberg equilibrium. The casein haplotype frequencies were estimated by the EH program (Xie and Ott, 1993). For EH computation, alleles with frequencies lower than 0.05 were ignored. Allele and haplotype frequencies of the 3 local breeds were compared with a sample of the Camosciata breed reared in different flocks of Northern Italy, analyzed by Budelli et al. (2005).

The structure of the whole casein complex was analyzed in silico at the protein and DNA levels. A multivariate analysis was performed by the average method of the CLUSTER procedure (SAS Institute, 1990) on

Table 1. Analyses applied to the milk and DNA samples for the genotyping

Casein fraction	Level	Method ¹	Identifiable alleles	Reference
<i>CSN1S1</i>	Protein	IEF	<i>A, C, B/E</i>	Caroli et al., 2001
	DNA	PCR-RFLP	<i>F, N A/0₁, B/E</i>	Ramunno et al., 2000
		AS-PCR	<i>E, non-E</i>	Jànsa Pérez et al., 1994
		AS-PCR	<i>0₁, non-0₁</i>	Cosenza et al., 2003
<i>CSN2</i>	Protein	IEF	<i>A, 0'</i>	Caroli et al., 2001
	DNA	PCR-SSCP	<i>A, C, E*, 0</i>	Chessa et al., 2004; *current study
<i>CSN1S2</i>	Protein	IEF	<i>A, B, C, E</i>	Caroli et al., 2001
	DNA	PCR-RFLP	<i>D, 0</i>	Ramunno et al., 2001
		PCR-RFLP	<i>F</i>	Ramunno et al., 2001
		IEF	<i>A^{IEF}, B^{IEF}</i>	Caroli et al., 2001 ²
<i>CSN3</i>	Protein	IEF	<i>A, B/B', B'', C, C', E, D/I/K/L, G, H, J, M</i>	Prinzenberg et al., 2005
	DNA	PCR-SSCP		

¹Methods: IEF = isoelectric focusing; AS-PCR = allele-specific PCR; PCR-SSCP = PCR-single strand conformation polymorphism.

²The nomenclature of the 2 classes of alleles with the IEF superscript was used by Prinzenberg et al. (2005).

the amino acid exchange or deletion classed as 1 (presence of the ancestral amino acid or amino acid sequence) or 0 (absence of the ancestral amino acid or amino acid sequence). Thirteen haplotypes showing a frequency higher than 0.05 were included in the cluster analysis. A phylogeny model was suggested for the whole casein complex.

RESULTS AND DISCUSSION

A New *CSN2* Allele

An unknown PCR-SSCP pattern was found at *CSN2* locus, showing an intermediate migration between *A* and *C* alleles (Figure 1). Direct sequencing of the PCR

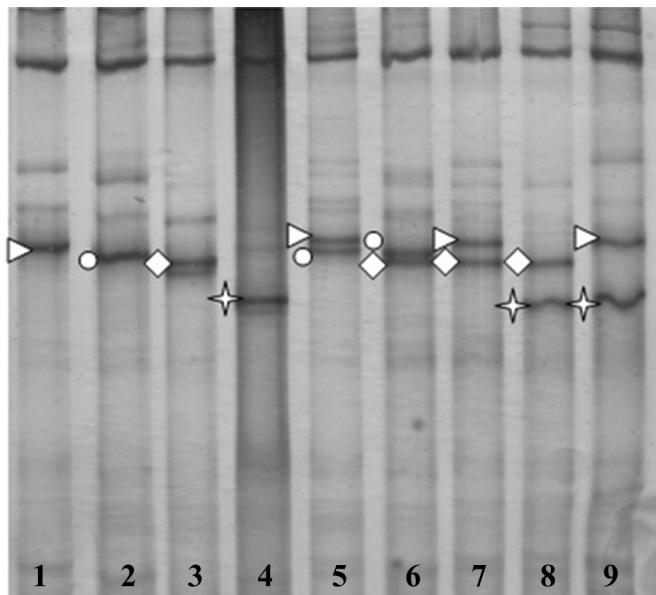


Figure 1. Polymerase chain reaction-single strand conformation polymorphism analysis of goat *CSN2*. The main band of each allele is indicated by a triangle (*CSN2***C*), a circle (*CSN2***E*), a diamond (*CSN2***A*), and a star (*CSN2***O'*). The genotypes of the 9 samples are: *CC* (1), *EE* (2), *AA* (3), *O'O'* (4), *CE* (5), *AE* (6), *AC* (7), *AO'* (8), *CO'* (9).

products allowed the characterization of the new pattern, resulting from a transversion TCT→TAT responsible for the amino acid exchange Ser₁₆₆→Tyr₁₆₆ in the mature protein. The transversion occurs in the codon 124 of exon 7 and involves nucleotide 8,215 of the reference *CSN2* sequence (GenBank Acc. No. AJ011018). The amino acid exchange deduced from the nucleotide sequence in the new variant involved 2 apolar amino acids, and was not detectable by standard protein screening techniques such IEF, in which it shared the same migration pattern with *CSN2***A* and *C*. Typing at the DNA level or by a higher resolution protein approach was therefore necessary to correctly identify the genetic variation of goat β-casein.

The new variant was named *CSN2***E* to avoid misunderstanding with the *D* nonallelic variant described at the proteome level by Galliano et al. (2004). The *D* variant is characterized by the Val₂₀₇ amino acid exchange involving a rather limited amount of the specific protein, and possibly resulting from a translation process mistake consequent to a differential splicing at a premessenger RNA level.

Most probably, *CSN2***E* derived from *CSN2***A*, because its translated sequence corresponds to the *CSN2***A* allele (GenBank Acc. No. AJ011018) except for the substitution described. In fact, both *CSN2***A* and *CSN2***E* share the same codon (GCA, Ala₁₇₇), different from *CSN2***C* (GTA, Val₁₇₇). Moreover, the alignment among the mature β-CN sequences of different species shown by Chessa et al. (2005) suggests that *CSN2***A* is the ancestral allele compared with *CSN2***E* because Ser₁₆₆ occurs in other ruminants.

Allele Frequencies at Each Casein Gene

Table 2 shows the allele frequencies at the casein loci in the 3 local breeds, compared with the more widely distributed Camosciata (Budelli et al., 2005). The genotype frequencies as well as the genotype combinations at the 4 loci are available online at <http://jds.fass.org>.

Table 2. Allele frequencies at the casein loci in Frisa, Orobica, Verzasca breeds, compared with the Camosciata breed (Budelli et al., 2005)

Locus	Allele	Breed			
		Frisa (n = 70)	Orobica (n = 66)	Verzasca (n = 67)	Camosciata (n = 88)
<i>CSN1S1</i>	<i>A</i>	0.129	— ¹	—	0.432
	<i>B</i>	0.007	0.008	0.037	0.199
	<i>C</i>	—	—	—	0.006
	<i>E</i>	0.200	0.008	0.201	0.114
	<i>F</i>	0.557	0.962	0.754	0.233
	<i>O₁</i>	0.107	0.023	0.007	0.011
	<i>N</i>	—	—	—	0.006
<i>CSN2</i>	<i>A</i>	0.200	0.008	0.254	0.341
	<i>C</i>	0.721	0.992	0.746	0.659
	<i>E</i>	0.079	—	—	—
<i>CSN1S2</i>	<i>A</i>	0.250	0.008	0.590	0.568
	<i>B</i>	0.114	—	0.007	0.051
	<i>C</i>	0.014	—	0.067	0.057
	<i>E</i>	0.057	—	—	—
	<i>F</i>	0.564	0.992	0.336	0.324
<i>CSN3</i>	<i>A</i>	0.329	0.083	0.284	0.386
	<i>B</i>	0.657	0.553	0.716	0.614
	<i>C</i>	0.007	0.015	—	—
	<i>D</i>	0.007	0.348	—	—

¹—: allele frequency = 0.

A total of 17, 15, 12, and 12 alleles were found in Frisa, Camosciata, Orobica, and Verzasca, respectively. Genetic equilibrium was generally found at each gene in the different breeds, except for a significant deviation at *CSN1S2* and *CSN2* in Frisa, in which an excess of homozygotes occurred.

In the local breeds, the greatest number of alleles was found at *CSN1S1* (from 4 to 5), even if a greater variability occurred at the same locus in the Camosciata, in which 7 alleles were detected. In Verzasca and Orobica (particularly), a rather unbalanced distribution was found for *CSN1S1* alleles. Only 2 variants (*F* and *E*) occurred at a frequency higher than 0.05 in Verzasca, whereas Orobica was almost monomorphic for *CSN1S1*F*. In the 3 Lombardy breeds, *CSN1S1*B* was a rare variant, unlike the Camosciata. In the Frisa, the null *CSN1S1*O₁* allele had a rather high frequency (0.11). The faint and null variants (*F*, *O₁*, *N*) showed a cumulative frequency of 0.98 in Orobica, 0.76 in Verzasca, 0.66 in Frisa, compared with 0.25 in Camosciata.

At *CSN2*, *C* was the predominant allele, ranging from 0.66 in Camosciata to 0.99 in Orobica, and confirming previous results in Italian breeds (Chessa et al., 2005). No *CSN2*O'* null allele was found, confirming its exclusive presence in Southern Italian goat breeds (Rando et al., 2000; Chessa et al., 2005).

The novel *CSN2*E* allele seems to be breed specific, because it was found only in Frisa in this study, and should also be absent in Camosciata, Saanen, Jonica, Garganica, Maltese, and Cilentana breeds (Budelli et

al., 2005; Chessa et al., 2005) because the same PCR-SSCP method was used for *CSN2* typing.

The greatest number of alleles at *CSN1S2* occurred in Frisa, in decreasing frequency order: *F*, *A*, *B*, *E*, *C*, whereas 4 alleles were found in Camosciata and Verzasca (*A* > *F* > *C* > *B*). Orobica was almost monomorphic for *CSN1S2*F*, with the *A* allele occurring at a very low frequency (0.01).

In the Frisa, a high frequency was found for *CSN1S2*B* (0.11), which was associated with *CSN1S1*O₁* as already demonstrated (Grosclaude et al., 1987). Interestingly, *CSN1S2*B* is characterized by an amino acid exchange leading to a further Lys residue in the protein variant compared with the other *CSN1S2* alleles associated with a normal α_{S2} -content (Bouniol, 1993). Because goat milk is particularly rich in the essential amino acid lysine (Salvadori Del Prato, 1998), an adaptive mechanism might be responsible for the casein complex structure occurring in Frisa breed. In fact, the lack of α_{S1} -casein and its essential amino acids might be balanced by a genetic variant in the casein complex providing a larger amount of Lys, such as *CSN1S2*B*, to increase the individual fitness. However, the association between *CSN1S1*O₁* and *CSN1S2*B* was not found in Orobica, in which *CSN1S2*B* did not occur. Two of the 3 Orobica goats carrying *CSN1S1*O₁* were informative for *CSN1S1-CSN1S2* linkage phase, revealing association between *CSN1S1*O₁* and *CSN1S2*F* in this breed. The adaptive mechanism possibly involving *CSN1S1* and *CSN1S2* should be further investigated.

Table 3. Haplotype frequencies at the casein loci in Frisa, Orobica, and Verzasca, compared with Camosciata breed (Budelli et al., 2005)¹

Casein loci				Frisa (n = 65)	Orobica (n = 58)	Verzasca (n = 61)	Camosciata (n = 84)	Overall ²
<i>CSN1S1</i>	<i>CSN2</i>	<i>CSN1S2</i>	<i>CSN3</i>					
A	A	F	B	— ³	—	—	0.0238	0.0075
A	C	A	A	0.0429	—	—	0.1436	0.0554
A	C	A	B	—	—	—	0.1682	0.0527
A	C	B	A	—	—	—	0.0357	0.0112
A	C	C	B	—	—	—	0.0146	0.0046
A	C	E	B	0.0586	—	—	—	0.0142
A	C	F	A	0.0152	—	—	—	0.0037
A	C	F	B	0.0188	—	—	0.0364	0.0160
B	A	A	B	—	—	—	0.1450	0.0454
B	A	C	B	—	—	—	0.0152	0.0048
B	A	F	A	—	—	—	0.0182	0.0057
B	C	F	A	—	—	—	0.0181	0.0057
B	C	F	B	—	—	—	0.0117	0.0037
E	A	A	A	—	—	—	0.0224	0.0070
E	A	A	B	0.1135	—	0.2129	0.0527	0.0925
E	A	F	B	0.0783	—	—	0.0368	0.0305
F	A	C	B	—	—	—	0.0119	0.0037
F	C	A	A	0.0196	—	—	—	0.0048
F	C	A	B	0.0648	—	0.3269	0.0391	0.1024
F	C	C	A	—	—	—	0.0179	0.0056
F	C	C	B	—	—	0.0656	—	0.0149
F	C	F	A	0.1604	0.0689	0.2366	0.0942	0.1372
F	C	F	B	0.2427	0.5775	0.1077	0.0632	0.2282
F	C	F	D	—	0.3534	—	—	0.0765
F	E	F	B	0.0534	—	—	—	0.0130
0 _I	C	B	A	0.0923	—	—	—	0.0224

¹Frequencies were estimated by EH program, excluding allele frequencies lower than 0.01. Haplotypes with frequencies higher than 0.05 in at least one breed are bolded.

²Overall = average frequencies across sampled animals.

³—: indicates haplotype frequencies lower than 0.01.

Only 2 *CSN3* alleles (A and B) were found in Verzasca and Camosciata with rather similar frequencies. The *CSN3*B* allele was predominant in all 4 breeds, with frequency ranging from 0.55 (Orobica) to 0.72 (Verzasca). Two additional alleles, C and D, occurred in Frisa and Orobica. In the Orobica, the low genetic variation shown at the calcium-sensitive caseins seems to be balanced by a higher variation at κ -CN. The high frequency of *CSN3*D* (0.35) in the breed is particularly important. Among the 4 identified alleles, *CSN3*D* is the only one belonging to the B^{IEF} group, according to the subdivision of *CSN3* alleles suggested by Prinzenberg et al. (2005). The presence of *CSN3*D* in the casein haplotype modifies the electric charge of κ -casein and, consequently, the physicochemical properties of the casein micelle. The favorable effect of the B^{IEF} variant on milk casein content (0.5 g/L more than A^{IEF} variant) was recently observed by Chiatti et al. (2005). The high frequency of the *CSN3*D* allele in Orobica might be explained as an adaptive mechanism because in this breed, alleles associated with faint or null α_{S1} -casein content showed a high incidence (0.98). The high frequency of *CSN3*D* variant in the breed might balance the unfavorable *CSN1S1* genetic effect on specific pro-

tein expression by increasing the milk content of κ -casein fraction.

Haplotype Distribution

Haplotype frequencies at *CSN1S1-CSN2-CSN1S2-CSN3* loci are reported in Table 3. A total of 26 haplotypes showed a frequency higher than 0.01 in at least 1 of the 4 breeds considered. We found 12, 3, 5, and 19 haplotypes with frequencies higher than 0.01 in Frisa, Orobica, Verzasca, and Camosciata, respectively. Only 13 haplotypes occurred at a frequency higher than 0.05 in at least one breed. High linkage disequilibrium was found in Frisa, Verzasca, and Camosciata, as indicated by the χ^2 test for association performed by EH. Linkage disequilibrium was not tested in Orobica, because the observations used for the haplotype frequency evaluation were monomorphic at *CSN1S1*, *CSN2*, and *CSN1S2*.

The haplotype distribution was highly different among breeds. In the Camosciata, the predominant haplotype was A-C-A-B (0.17), followed by B-A-A-B (0.15), A-C-A-A (0.14), F-C-F-A (0.09), and F-C-F-B (0.06). In the Frisa, the frequency of the 5 most common

Table 4. Amino acid differences among the *CSN1S1-CSN2-CSN1S2-CSN3* haplotypes¹

Haplotype	Amino acid position in the mature protein														
	<i>CSN1S1</i>					<i>CSN2</i>		<i>CSN1S2</i>				<i>CSN3</i>			
	16	77	100	195	Δ	166	177	7	64	167	193	43	65	119	159
<i>BAAB</i>	Pro	Glu	Arg	Thr		Ser	Ala	Val	Glu	Lys	Pro	Gln	Val	Ile	Ser
<i>EAAB</i>			Lys	Ala											
<i>EAFB</i>			Lys	Ala				Ile							
<i>FCAB</i>		*	°	°	a		Val								
<i>FCCB</i>		*	°	°	a		Val			Ile					
<i>FCFB</i>		*	°	°	a		Val	Ile							
<i>FCFA</i>		*	°	°	a		Val	Ile						Val	
<i>FCFD</i>		*	°	°	a		Val	Ile				Arg	Ile		Pro
<i>FEFB</i>		*	°	°	a	Tyr		Ile							
<i>ACAB</i>	Leu	Gln					Val								
<i>ACAA</i>	Leu	Gln					Val							Val	
<i>ACEB</i>	Leu	Gln					Val			Ile	Arg				
<i>O₁CBA[§]</i>		<i>Gln</i> *	—	—	b		Val		Lys					Val	

¹Δ = Amino acid deletion (a = 59-95; b = 110-199); * = amino acid missing and deduced from the nucleotide sequence; ° = amino acid position changed in the mature protein due to the deletion; § = all the protein is missing; — = the nucleotide sequence is missing and we cannot deduce the amino acid which should occur in the mature protein. Haplotypes with frequencies higher than 0.05 in at least one breed are shown.

haplotypes was as follows: *F-C-F-B* (0.24), *F-C-F-A* (0.16), *E-A-A-B* (0.11), *O₁-C-B-A* (0.09), and *E-A-F-B* (0.08). Only 5 haplotypes showed a frequency higher than 0.01 in the Verzasca: *F-C-A-B* (0.33), *F-C-F-A* (0.24), *E-A-A-B* (0.21), *F-C-F-B* (0.11), and *F-C-C-B* (0.07). In the Orobica, only 3 haplotypes were found, differing only at *CSN3*: *F-C-F-B* (0.58), *F-C-F-D* (0.35), and *F-C-F-A* (0.07).

Among the 25 haplotypes carrying *CSN3**A or *CSN3**B, *B-A-A-B* may be considered the ancestral haplotype based on the evolutive hypotheses discussed in the introduction, which assume a parsimony evolution model at each casein gene. The results obtained in this study for *CSN2**E are also in agreement with the parsimony model suggesting the origin of *CSN2**E from *CSN2**A. The assumption of *B-A-A-B* as the ancestral haplotype will be the basis for the following discussion.

The *B-A-A-B* haplotype had a high incidence in the Camosciata, whereas it was not found in the 3 local breeds. Four haplotypes carried an allele found only in a breed at a frequency higher than 0.05: *O₁-C-B-A*, *F-E-F-B*, *A-C-E-B* in Frisa, and *F-C-F-D* in Orobica.

The amino acid exchanges and deletions in the casein variants affecting the 13 most common haplotypes are shown in Table 4. It is evident that Ala₁₇₇, which characterizes *CSN2**A, is usually associated with *CSN1S1**B and *CSN1S1**E, with an overall cumulative frequency of 0.19 compared with 0.01 for *CSN2**C associated with the same *CSN1S1* alleles (Table 3). Alternatively, Val₁₇₇, characterizing *CSN2**C, is generally associated with *CSN1S1**A and *CSN1S1**F (overall cumulative frequency of 0.73 compared with 0.01 for *CSN2**A associated with the same *CSN1S1* alleles). However, the

association *CSN1S1**F-*CSN2**C is related to breeds containing a very few number of haplotypes; i.e., Orobica and Verzasca, and needs to be confirmed in other goat populations.

A less clear evolutive trend concerns the remaining casein cluster. When considering *CSN2* and *CSN1S2*, the most common combination is *C-F* (cumulative frequency = 0.47), followed by *C-A* (0.21), and by the ancestral *A-A* (0.15). The last part of the casein haplotype involves *CSN1S2* and *CSN3* loci. The highest cumulative frequency was found for *F-B* (0.30) followed by the ancestral combination *A-B* (0.29), and by *F-A* (0.15).

Figure 2 shows the tree obtained from the cluster analysis carried out on the thirteen haplotypes having a frequency higher than 0.05. For this analysis, amino acid exchanges or deletions at the different loci shown in Table 4 had been classified as a matrix of 0 or 1. As an example, the presence of Pro or Leu at position 16 of the *CSN1S1* mature protein was coded as 1 or 0, whereas the occurrence or the deletion of the 59th-95th amino acid sequence at the same protein was classified as 1 or 0, respectively.

The 2 main clusters correspond to the 2 evolutive lineages of *CSN1S1*: lineage A (right panel of Figure 2, with *CSN1S1**A) and B (left panel of Figure 2, with *CSN1S1**B, *E* and *F*). The 2 clusters overlap for *CSN2* locus. In fact, *CSN2**C always occurred in the A cluster, whereas, in the B cluster, *CSN2**A was associated with *CSN1S1**B or *CSN1S1**E, and *CSN2**C was associated with *CSN1S1**F. Some haplotypes representing exceptions to this trend were also found in the Camosciata (*A-A-F-B*, *B-C-F-A*, *B-C-F-B*, *F-A-C-B*) but were not included in the cluster analysis due to the low frequency

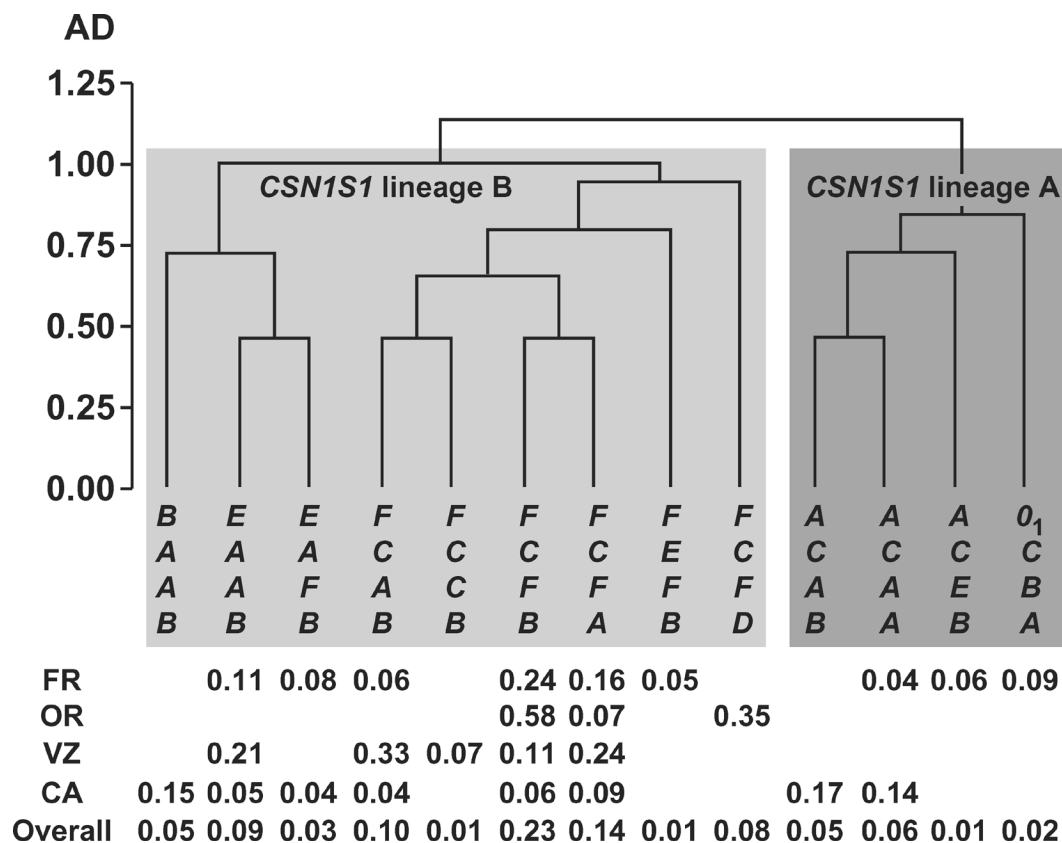


Figure 2. Tree from the cluster analysis of the most common *CSN1S1*-*CSN2*-*CSN1S2*-*CSN3* haplotypes (frequency >0.05 in at least one breed); AD = average distance. Haplotype frequencies in the different breeds (FR = Frisa, OR = Orobica, VZ = Verzasca, CA = Camosciata) and overall are shown below the tree.

(<0.05) observed. The relationships among the other genes of the casein complex are less clear, possibly because of recombination events confounding the evolutionary pathway.

A phylogeny model can be suggested for the whole haplotype starting from the ancestral *CSN1S1***B*-*CSN2***A*-*CSN1S2***A*-*CSN3***B* haplotype (Figure 3). The origin of *E*-*A*-*A*-*B* is self-explanatory, whereas the evolution of the other haplotypes is more complex. Two different pathways may be suggested, one leading to *F*-*C*-*A*-*B* and the other to *A*-*C*-*A*-*B*. It is most probably that a common *B*-*C*-*A*-*B* haplotype led to both pathways, although it was not found in the breeds considered. This fact might cause the overlapping between the 2 *CSN1S1* lineages when considering *CSN2* (Figure 2), and suggests that the *CSN2***A* to *C* exchange occurred before other mutations at *CSN1S1* locus. Different recombination events involved *CSN1S2* and *CSN3* genes in the haplotype evolution.

With respect to the haplotype carrying the *CSN3***D*, it is remarkable that this allele belongs to an evolutionary lineage different from *CSN3***B* and *CSN3***A* (Yahyaoui

et al., 2003; Prinzenberg et al., 2005). Thus, it cannot be assessed when its development occurred with respect to the most common *CSN3***B* allele. In all cases, no ancestral alleles occurred at the other casein genes in the *F*-*C*-*F*-*D* haplotype, which may be considered subsequent to the *B*-*A*-*A*-*B* in the casein complex phylogeny, possibly because of chromosome rearrangement mechanisms. More data are necessary to investigate the haplotype phylogeny involving *CSN3***D* allele.

Most probably, the *F*-*E*-*F*-*B* and *A*-*C*-*E*-*B* haplotypes derived from further recombination events, on the basis of the supposed origin of *CSN2***E* and *CSN1S2***E* from *CSN2***A* and *CSN1S2***C*, respectively. Moreover, *E*-*A*-*F*-*B* might derive from recombinant events if *CSN1S2***F* originated from *F*-*C*-*A*-*B*, as suggested in the evolutionary model. Although recombination among the casein genes is essential in explaining the haplotype variability, a strong linkage disequilibrium condition was found in the breeds, resulting in an unbalanced distribution of the haplotypes described before. The fact that only a few haplotypes occur with a high frequency in the breeds considered indicates that selection has

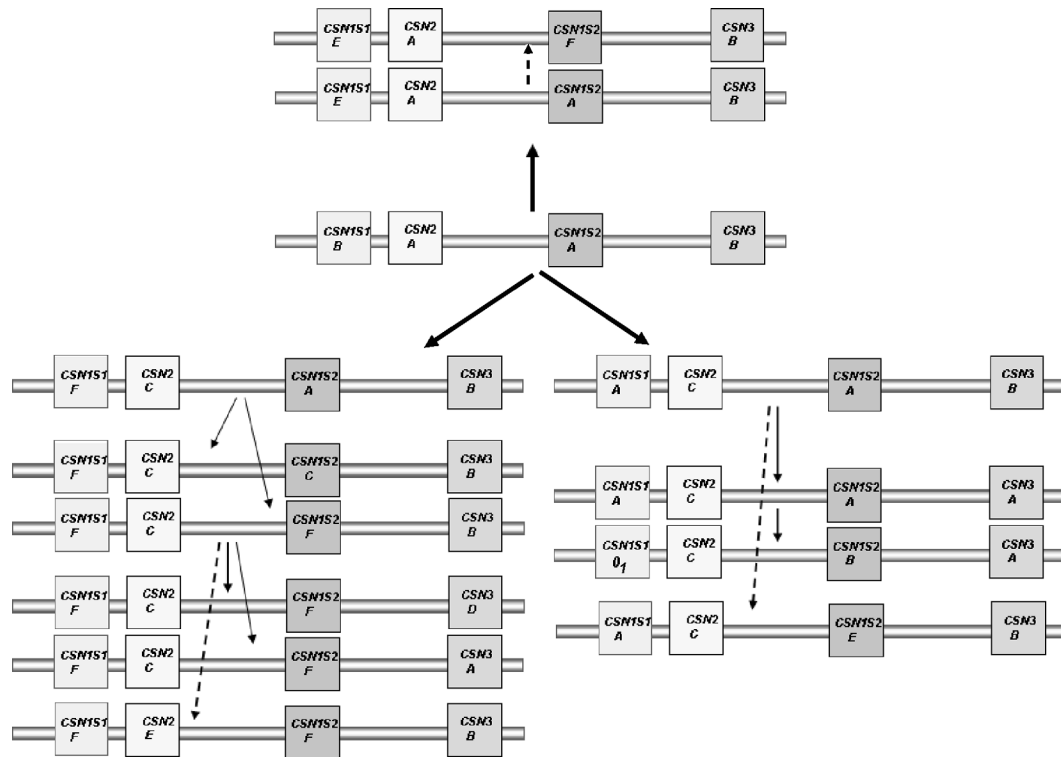


Figure 3. A proposal for the evolution of goat casein haplotype. The 13 most common haplotypes are considered. The broken lines symbolize possible recombinant events discussed in the text.

strongly reduced the casein complex variability. A balance mechanism among the genetic variants in the casein micelles, as suggested by this study, might explain the high linkage disequilibrium found in Frisa, Camosciata, and Verzasca.

The results obtained in the current study can be partially compared with a previous investigation by Sacchi et al. (2005), in which *CSN1S1-CSN2-CSN3* haplotype variability was analyzed in 5 Italian local breeds (Vallesana, Roccaverano, Jonica, Garganica, and Maltese). Among the 18 haplotypes considered, an evolutive pathway was proposed starting from the ancestral *B-A-B* haplotype, followed by *B-F-B*, which was suggested to give rise to the most common *F-F-B*. In the Camosciata, *F-F-B* haplotype had a lower frequency (0.07) than in the local breeds both analyzed by Sacchi et al. (2005) and in this study, whereas *F-F-A* was more common (0.10), as in the Vallesana. Interestingly, strong evidences of recombination events, not only among but also within the casein genes, were found for the 18 haplotypes taken into account by Sacchi et al. (2005).

CONCLUSIONS

The low frequency of the strong *CSN1S1* variants in the Lombardy breeds indicates that selection for these

alleles should be an important breeding objective, to improve milk composition and cheese-making aptitude. On the other hand, the high incidence of null and faint alleles might be exploited for the production of milk with specific nutritional properties.

At the *CSN2* locus, the predominance of the *C* allele in Italian breeds was confirmed. Moreover, a new mutation was identified resulting in an amino acid exchange not detectable by standard protein screening techniques. The novel variant, named *CSN2*E*, was found in only one breed.

Adaptive mechanisms were suggested that could balance the effects of alleles at different loci. The *CSN1S2*B* and *CSN3*D* alleles, which were found prevalently associated with null and faint alleles at *CSN1S1*, lead to a higher content, respectively, in lysine and casein than other *CSN1S2* and *CSN3* variants.

Further studies should aim to explain how the goat casein complex evolution led to all the existing mutations. An answer to this problem might provide not only interesting elements from the point of view of the phylogenesis of the goat species, but also useful indications for the selection of the animals on the basis of functional and adaptive aspects concerning the whole casein complex.

An evolutive model has been suggested, starting from the haplotype *CSN1S1*B-CSN2*A-CSN1S2*A-CSN3*B*, and resulting in several protein variants that differently assemble the casein micelle. Both milk composition and technological properties can be strongly affected by the casein micelle structure, which is the result of “genome to proteome” relationships involving the goat caseins.

The proposed phylogeny was an attempt to explain goat casein cluster evolution on the basis of the subset of population, which was taken into account in the present paper. Alternative evolutionary pathways could be traced if considering further goat breeds, or fitting different phylogeny models; that is, not based on the parsimony assumption. Nevertheless, the proposed phylogeny pathway could be the basis for better understanding the biological mechanisms that differentiated the casein complex so deeply in the different goat populations because of natural and artificial selection.

ACKNOWLEDGMENTS

Research supported by CARIPLO foundation and by PRIN 2005. The authors wish to thank Chiara Ghilardi of ARAL (Crema) and the personnel of the breeder associations for the help in collecting milk samples.

REFERENCES

- Alexander, L. J., A. F. Stewart, A. G. Mackinlay, T. V. Kapelinskaya, T. M. Tkach, and S. I. Gorodetsky. 1988. Isolation and characterization of the bovine kappa-casein gene. *Eur. J. Biochem.* 178:395–401.
- Bevilacqua, C., P. Ferranti, G. Garro, C. Veltri, R. Lagonigro, C. Leroux, E. Pietrolà, 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 in the goat species. *Eur. J. Biochem.* 269:1293–1303.
- Boulangier, A., F. Grosclaude, and M. F. Mahé. 1984. Polymorphisme des caséines α_{S1} and α_{S2} de la chèvre (*Capra hircus*). *Genet. Sel. Evol.* 16:157–176.
- Bouniol, C. 1993. Sequence of the goat alpha s2-casein-encoding cDNA. *Gene* 125:235–236.
- Bouniol, C., G. Brignon, M. F. Mahé, and C. Printz. 1994. Biochemical and genetic analysis of variant C of caprine α_{S2} -casein. *Anim. Genet.* 25:173–177.
- Budelli, E., S. Chessa, F. Chiatti, P. Bolla, and A. Caroli. 2005. Variabilità degli aplotipi caseinici nella razza caprina Camosciata. *Sci. Tecn. Latt. Cas.* 56:205–217.
- Caroli, A., O. Jann, E. Budelli, P. Bolla, S. Jäger, and G. Erhardt. 2001. Genetic polymorphism of goat κ -casein (*CSN3*) in different breeds and characterization at DNA level. *Anim. Genet.* 32:226–230.
- Chessa, S., E. Budelli, P. Bolla, and A. Caroli. 2003a. Il polimorfismo della caseina caprina: Uno straordinario “puzzle” genetico dalle svariate potenzialità applicative. *Sci. Tecn. Latt. Cas.* 54:343–358.
- Chessa, S., E. Budelli, F. Chiatti, A. M. Cito, P. Bolla, and A. Caroli. 2005. Predominance of β -casein (*CSN2*) C allele in goat breeds reared in Italy. *J. Dairy Sci.* 88:1878–1881.
- Chessa, S., E. Budelli, K. Gutscher, A. Caroli, and G. Erhardt. 2003b. Simultaneous identification of five κ -casein (*CSN3*) alleles in domestic goat by PCR-SSCP. *J. Dairy Sci.* 86:3726–3729.
- Chianese, L., P. Ferranti, G. Garro, R. Mauriello, and F. Addeo. 1997. Occurrence of three novel α_{S1} -casein variants in goat milk. Pages 259–267 in *Proc. IDF Milk Protein Polymorphism Seminar II*. International Dairy Federation, Brussels, Belgium.
- Chiatti, F., A. Caroli, S. Chessa, P. Bolla, and G. Pagnacco. Relationships between goat κ -casein (*CSN3*) polymorphism and milk composition. Pages 163–164 in *Proc. FAO, The role of biotechnology*. FAO, Rome, Italy.
- Cosenza, G., R. Illario, A. Rando, P. Di Gregorio, P. Masina, and L. Ramunno. 2003. Molecular characterization of the goat *CSN1S101* allele. *J. Dairy Res.* 70:237–240.
- Cosenza, G., A. Paucillo, D. Gallo, D. Di Bernardino, and L. Ramunno. 2005. A *SspI* PCR-RFLP detecting a silent allele at the goat *CSN2* locus. *J. Dairy Res.* 72:456–459.
- Ferretti, L., P. Leone, and V. Sgaramella. 1990. Long range restriction analysis of the bovine casein genes. *Nucleic Acids Res.* 18:6829–6833.
- Galliano, F., R. Saletti, V. Cunsolo, S. Foti, D. Marletta, S. Bordonaro, and G. D’Urso. 2004. Identification and characterization of a new β -casein variant in goat milk by high-performance liquid chromatography with electrospray ionization mass spectrometry and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 18:1972–1982.
- Grosclaude, F., M. F. Mahé, G. Brignon, L. Di Stasio, and R. Jeunet. 1987. A Mendelian polymorphism underlying quantitative variations of goat α_{S1} -casein. *Genet. Sel. Evol.* 19:399–412.
- Grosclaude, F., and P. Martin. 1997. Palmerston North, New Zealand. Pages 241–253 in *Proc. IDF Milk Protein Polymorphism Seminar II*. International Dairy Federation, Brussels, Belgium.
- Grosclaude, F., G. Ricordeau, P. Martin, F. Remeuf, L. Vassal, and J. Bouillon. 1994. Du gène au fromage: Le polymorphisme de la caséine α_{S1} caprine, ses effets, son évolution. *INRA Prod. Anim.* 7:3–19.
- Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95–98.
- Hayes, H., E. Petit, C. Bouniol, and P. Popescu. 1993. Localisation of the alpha-S2-casein gene (*CASAS2*) to the homologous cattle, sheep and goat chromosome 4 by in situ hybridisation. *Cytogenet. Cell Genet.* 64:282–285.
- Jann, O., E. M. Prinzenberg, G. Luikart, A. Caroli, and G. Erhardt. 2004. High polymorphism in the κ -casein (*CSN3*) gene from wild and domestic caprine species revealed by DNA sequencing. *J. Dairy Res.* 71:188–195.
- Jànsa Pérez, M., C. Leroux, A. Sanchez Bonastre, and P. Martin. 1994. Occurrence of a LINE element in the 3’UTR of an allelic form of the goat α_{S1} -casein gene associated with a reduced level of protein synthesis. *Gene* 147:179–187.
- Lagonigro, R., E. Pietrola, M. D’Andrea, C. Veltri, and F. Pilla. 2001. Molecular genetic characterization of the goat α_{S2} -casein E allele. *Anim. Genet.* 32:391–393.
- Leroux, C., and P. Martin. 1996. The caprine α_{S1} and beta-casein genes are 12-kb apart and convergently transcribed. *Anim. Genet.* 27:93.
- Mahé, M. F., and F. Grosclaude. 1993. Polymorphism of β -casein in the Creole goat of Guadeloupe, evidence for a null allele. *Genet. Sel. Evol.* 25:403–408.
- Neveu, C., D. Mollè, J. Moreno, P. Martin, and J. Lèonid. 2002. Heterogeneity of caprine beta-casein elucidated by RP-HPLC/MS: Genetic variants and phosphorylation. *J. Protein Chem.* 21:557–567.
- Persuy, M. A., C. Printz, J. F. Medrano, and J. C. Mercier. 1999. A single nucleotide deletion resulting in a premature stop codon is associated with marked reduction of transcripts from a goat β -casein null allele. *Anim. Genet.* 30:444.
- Popescu, C. P., S. Long, P. Riggs, J. Womack, S. Schmutz, R. Fries, and D. S. Gallagher. 1996. Standardization of cattle karyotype nomenclature: Report of the committee for the standardization of the cattle karyotype. *Cytogenet. Cell Genet.* 74:259–261.
- Prinzenberg, E. M., K. Gutscher, S. Chessa, A. Caroli, and G. Erhardt. 2005. Caprine κ -casein (*CSN3*) polymorphism: New developments of the molecular knowledge. *J. Dairy Sci.* 88:1490–1498.

- Ramunno, L., G. Cosenza, M. Pappalardo, E. Longobardi, D. Gallo, N. Pastore, P. Di Gregorio, and A. Rando. 2001a. Characterization of two new alleles at the goat *CSN1S2* locus. *Anim. Genet.* 32:264–268.
- Ramunno, L., G. Cosenza, M. Pappalardo, N. Pastore, D. Gallo, P. Di Gregorio, and P. Masina. 2000. Identification of the goat *CSN1S1F* allele by means of PCR-RFLP method. *Anim. Genet.* 31:342.
- 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.
- Ramunno, L., E. Longobardi, M. Pappalardo, A. Rando, P. Di Gregorio, G. Cosenza, P. Mariani, N. Pastore, and P. Masina. 2001b. An allele associated with a non detectable amount of α_{S2} -casein in goat milk. *Anim. Genet.* 32:19–26.
- Ramunno, L., P. Mariani, M. Pappalardo, A. Rando, M. Capuano, P. Di Gregorio, and G. Cosenza. 1995. Un gene ad effetto maggiore sul contenuto di caseina β nel latte di capra. Pages 185–186 in Proc. XI Conv. Associazione Scientifica di Produzione Animale (ASPA). Ed. Università degli Studi di Udine, Udine, Italy.
- Rando, A., L. Ramunno, and P. Masina. 2000. Mutations in casein genes. *Zoot. Nutriz. Anim.* 26:105–114.
- Raymond, M., and F. Rousset. 1995. Genepop (version 1.2): Population genetics software for exact test and ecumenicism. *J. Hered.* 86:248–249.
- Sacchi, P., S. Chessa, E. Budelli, P. Bolla, G. Ceriotti, D. Soglia, R. Rasero, E. Cauvin, and A. Caroli. 2005. Casein haplotype structure in five Italian goat breeds. *J. Dairy Sci.* 88:1561–1568.
- Salvadori Del Prato, O. 1998. Trattato di tecnologia casearia. Edagricole, Bologna, Italy.
- SAS Institute. 1990. SAS/Stat User's Guide. Version 6, 4th ed. SAS Institute Inc., Cary, NC.
- Threadgill, D. W., and J. E. Womack. 1990. Genomic analysis of the major bovine milk proteins genes. *Nucleic Acids Res.* 18:6935–6942.
- Xie, X., and J. Ott. 1993. Testing linkage disequilibrium between a disease gene and marker loci. *Am. J. Hum. Genet.* 53:1107.
- Yahyaoui, M. H., A. Angiolillo, F. Pilla, A. Sanchez, and J. M. Folch. 2003. Characterization and genotyping of the caprine kappa casein variants. *J. Dairy Sci.* 86:2715–2720.