J. Dairy Sci. 88:1561–1568

Casein Haplotype Structure in Five Italian Goat Breeds

P. Sacchi,[1] S. Chessa,[2] E. Budelli,[3] P. Bolla,[2] G. Ceriotti,[2] D. Soglia,[1] R. Rasero,[1] E. Cauvin,[1] and A. Caroli[4]

[1]Dipartimento di Produzioni Animali, Epidemiologia ed Ecologia, via Leonardo da Vinci 44, 10095 Grugliasco, Torino, Italy
[2]Dipartimento di Scienze e Tecnologie Veterinarie per la Sicurezza Alimentare, Università degli Studi di Milano, 20134 Milano, Italy
[3]Fondazione Parco Tecnologico Padano, Centro Ricerche e Studi Agroalimentari, 20090 Segrate (Milano), Italy
[4]Dipartimento di Sanità e Benessere Animale, Strada Provinciale per Casamassima km 3, 70010 Valenzano (Bari), Italy

ABSTRACT

The aim of this work was to investigate the genetic structure of the casein gene cluster in 5 Italian goat breeds and to evaluate the haplotype variability within and among populations. A total of 430 goats from Vallesana, Roccaverano, Jonica, Garganica, and Maltese breeds were genotyped at α_{s1}-casein (*CSN1S1*), α_{s2}-casein, (*CSN1S2*), β-casein (*CSN2*), and κ-casein (*CSN3*) loci using several genomic techniques and milk protein analysis. Casein haplotype frequencies were estimated for each breed. Principal component analysis was carried out to highlight the relationship among breeds. Allele and haplotype distributions indicated considerable differences among breeds. The haplotype *CSN1S1*F- CSN1S2*F-CSN3*D* occurred in all breeds with frequencies >0.100 and was the most common haplotype in the Southern breeds. A high frequency of *CSN1S1*0-CSN1S2*C-CSN3*A* haplotype was found in Vallesana population (0.162). Principal component analysis clearly separated the Northern and Southern breeds by the first component. The variability of the caprine casein loci and variety of resulting haplotypes should be exploited in the future using specific breeding programs aiming to preserve biodiversity and to select goat genetic lines for specific protein production.
(**Key words:** goat, casein, polymorphism, haplotype)

Abbreviation key: AS-PCR = allele specific-PCR, ***CSN1S1*** = α_{s1}-casein locus, ***CSN1S2*** = α_{s2}-casein locus, ***CSN3*** = κ-casein locus, ***CSN2*** = β-casein locus, **IEF** = isoelectrofocusing, **PCR-SSCP** = PCR-single strand conformational polymorphism.

INTRODUCTION

It is well known that casein genes are organized as a cluster, as first reported by Grosclaude et al. (1978), including in order α_{s1}-casein (***CSN1S1***), β-casein (***CSN2***), α_{s2}-casein (***CSN1S2***), and κ-casein (***CSN3***) (Ferretti et al., 1990; Threadgill and Womack, 1990; Rijnkels et al., 1997). In goats, the entire casein gene cluster region spans about 250 kb on chromosome 6 (Hayes et al., 1993). Furthermore, *CSN1S1* and *CSN2* are only 12 kb apart and are convergently transcribed (Leroux and Martin, 1996).

In recent years, the genetic polymorphism of goat caseins has raised considerable research interest because goat casein polymorphisms are related to milk quality, composition, and technological properties (Martin et al., 2002). Researchers started with *CSN1S1*, which has 16 known co-dominant alleles. Alleles are associated with different rates of protein synthesis. In the 1980s, 7 genetic variants were known at the protein level: *A*, *B*, *C*, *D*, *E*, *F*, and *0* (Boulanger et al., 1984; Grosclaude et al., 1987; Brignon et al., 1989; Mahé and Grosclaude, 1989). Later on, more variants were identified both at the protein and DNA level and named as 0_2, *G*, B_2, B_3, B_4 (Leroux et al., 1990; Martin and Leroux, 1994; Grosclaude and Martin, 1997); *H*, *I*, *L* (Chianese et al., 1997); *M* (Bevilacqua et al., 2002); and *N* (Ramunno et al., 2002). The *B* allele, now renamed B_1, is considered ancestral for the species. The original *0* allele was renamed 0_1 to differentiate it from the second null allele (0_2) found by genomic studies. The *D* variant was traced back to the *G* variant (Martin and Leroux, 1994), characterized as the *F* allele, by an internal deletion. On the basis of the milk content of α_{S1}-casein, the *CSN1S1* variants can be classed into 4 groups: strong alleles (*A*, B_1, B_2, B_3, B_4, *C*, *H*, *L*, and *M*), producing almost 3.5 g/L of α_{S1}-casein each; intermediate alleles (*E* and *I*; 1.1 g/L); weak alleles (*F* and *G*; 0.45 g/L); and null alleles (0_1, 0_2, and *N*), producing no α_{S1}-casein (Grosclaude and Martin, 1997; Rando et al., 2000).

Three variants of *CNS2* were found to be associated with a normal β-casein content: *A*, *B* (Mahé and Gros-

Received July 13, 2004.
Accepted November 29, 2004.
Corresponding author: P. Sacchi; e-mail: paola.sacchi@unito.it.

claude, 1993), and *C* (Neveu et al., 2002), the last allele differing for a single amino acid substitution ($Ala_{177} \rightarrow Val_{177}$) from the *A* variant. Furthermore, 2 null *CSN2* alleles were identified, both characterized by mutations responsible for premature stop codons in exon 7 (Ramunno et al., 1995: GenBank Accession number AJ011019; Persuy et al., 1999: GenBank Accession number AF172260). The mRNA analysis revealed that the transcript product amounts were almost 10 (Ramunno et al., 1995) and 100 (Persuy et al., 1999) times lower for the null alleles than for the *A* variant.

The genetic polymorphism of *CSN1S2* was first described by Boulanger et al. (1984). At least 8 alleles have been identified. Alleles are associated with 3 synthesis levels; *A*, *B* (Boulanger et al., 1984), *C* (Bouniol et al., 1994), *E* (Lagonigro et al., 2001), *F* (Ramunno et al., 2001a), and *G* (Erhardt et al., 2002) are associated with a normal synthesis level, whereas *D* and *0* are associated with lower and null synthesis levels, respectively (Ramunno et al., 2001a, b).

Since the discovery of 2 caprine *k*-casein variants identifiable by isoelectrofocusing (**IEF**) (Di Luccia et al., 1990) and successively confirmed both at the protein and DNA level by Caroli et al. (2001), further *CSN3* polymorphisms were detected by DNA analysis (Yahyaoui et al., 2001; Angiolillo et al., 2002; Yahyaoui et al., 2003; Jann et al., 2004). A total of 13 polymorphic sites were identified in the domestic goat (Jann et al., 2004), allowing the identification of 14 alleles corresponding to 11 protein variants, if taking into account the *Capra pyrenaica* variant (Yahyaoui et al., 2001), which was later observed in domestic goat breeds (Yahyaoui et al., 2003). Recently, 2 more polymorphisms were demonstrated by PCR-single strand conformational polymorphism (**PCR-SSCP**) analysis (Chessa et al., 2003), but they are not yet characterized. The number of caprine *CSN3* polymorphisms and available sequences has led to some nomenclature problems, but new nomenclature has been proposed (Chessa et al., 2003; Jann et al., 2004), mainly based on the GenBank chronological order of the variants.

Because of the tight association among casein genes, the estimation of the relationship between casein variants and milk production traits can be improved by considering the entire casein haplotype instead of single gene typing. This was first suggested in cattle by Grosclaude et al. (1978) and was further developed in the same species (Aleandri et al., 1990; Ikonen et al., 2001; Boettcher et al., 2004; Caroli et al., 2004). The effect of an allele at a given locus could be confounded with the effect of linked alleles at another locus. Moreover, selection for an allele could automatically increase the frequency of linked alleles that may not result in a favorable effect on the trait of interest. Research focused at the haplotype level is necessary to detect important effects that could be used for the genetic improvement of goat breeds.

The aim of this work was to investigate the genetic structure at the casein gene cluster in 5 Italian goat breeds and to evaluate the casein haplotype variability within and among breeds.

MATERIALS AND METHODS

Samples and Breeds

Blood and milk samples were collected from the following goat breeds: Vallesana (n = 83), Roccaverano (n = 77), Maltese (n = 70), Jonica (n = 110), and Garganica (n = 38).

Vallesana is reared in Northern Italy (Piedmont Region) and was originally imported from Switzerland as Walliser Schwarzhalsziege. The animals are medium sized and have black forequarters and white hindquarters. In 2002, the breed risk status was classified as critical because the total population size was around 100 head (http://dad.fao.org). Vallesana is considered a dual-purpose (milk and meat) breed, and the milk is mainly used for fresh cheese manufacturing. Average lactation milk production is about 260 kg in 210 d.

Roccaverano is a population found in Northern Italy. The animals are medium sized, brown or white, and polled. The breed is well adapted to the local environment and currently has a census of about 1000 individuals (http://dad.fao.org). Average lactation milk production is about 350 kg in 198 d.

Maltese was derived from Italian and North African breeds and is distributed in Southern Italy with an estimated population of 7000 head (http://dad.fao.org). The animals are lightweight and normally polled. Maltese hair is long and white or yellowish. Maltese have black marks on the ears and on both sides of the head and neck. Average lactation milk production is about 350 kg in 200 d.

Garganica goats are spread all over the Southern Apennine, and the breed is well adapted to the hilly environment. The animals are black and have screw-shaped horns. The current census indicates a population of about 19,000 breeding females (http://dad.fao.org). The typical lactation milk production is from 200 to 250 kg in 210 d. Garganica milk is characterized by high protein and fat content.

The Jonica breed resulted from crosses of the local population and Maltese. Jonica are almost exclusively raised in the southern area of Apulia. Jonica are adapted to live on arid soils. They are either polled or horned and have long, white hair. The number of breeding females is of about 1600 individuals (http://

Table 1. Allelic frequencies in the different breeds (blanks = 0.000).

Locus[1]	Allele	Vallesana	Roccaverano	Maltese	Jonica	Garganica
		(n = 83)	(n = 77)	(n = 70)	(n =110)	(n = 38)
CSN1S1	*A*	0.030	0.227	0.414	0.350	0.276
	B	0.127	0.117	0.157	0.305	0.408
	C					0.026
	E	0.283	0.214	0.057	0.064	
	F	0.386	0.377	0.371	0.282	0.224
	0_1	0.175	0.045			
	N		0.019			
	H					0.066
CSN2	*A**[2]	1.000	1.000	0.964	0.964	0.974
	0			0.036	0.036	0.026
CSN1S2	*A*	0.042	0.078	0.286	0.291	0.382
	B	0.090	0.175	0.086	0.014	
	C	0.530	0.481	0.264	0.355	0.145
	E		0.006	0.107	0.005	0.066
	F	0.337	0.260	0.250	0.332	0.408
	0			0.007	0.005	
CSN3	*A*	0.373	0.253	0.114	0.132	0.066
	B		0.065	0.179	0.091	0.145
	C	0.024	0.032			
	D	0.596	0.604	0.693	0.745	0.776
	X	0.006	0.032	0.007	0.018	0.013
	Y		0.013	0.007	0.014	

[1]*CSN1S1* = α_{s1}-Casein locus, *CSN2* = β-casein locus, *CSN1S2* = α_{s2}-casein locus, and *CSN3* = κ-casein locus.

[2]*A** = *CSN2*A* + *CSN2*C*.

dad.fao.org). Average lactation milk production is about 300 kg in 215 d.

Genotyping

Genotyping was carried out at the protein level and at the DNA level to take into account both the phenotypic expression of the casein genes and the genetic structure of their nucleotide sequences. First, milk samples were typed by IEF according to Caroli et al. (2001). The DNA was extracted from blood or milk by the GFX Genomic Blood DNA Purification Kit (Amersham Biosciences, Uppsala, Sweden) and then typed using the following analyses:

CSN1S1: Allele-specific-PCR (**AS-PCR**) (Jànsa Pérez et al., 1994; Leroux et al., 1992); PCR-RFLP (Ramunno et al., 2000, 2002);
CSN1S2: PCR-RFLP (Ramunno et al., 1999, 2001a; Lagonigro et al., 2001);
CSN2: AS-PCR (Ramunno et al., 1995);
CSN3: PCR-SSCP (Chessa et al., 2003).

Statistical Analyses

Genepop (Raymond and Rousset, 1995) software was used to estimate allelic frequencies and to verify Hardy-Weinberg equilibrium. Significance levels were based on *F*-statistics and were from FSTAT (Goudet, 1995). Casein haplotype frequencies were estimated by using EH (Xie and Ott, 1993), considering only the alleles with frequencies >0.05. Principal component analysis was carried out on haplotype frequencies (SAS, 1990) to highlight the relationship among the breeds on the basis of the casein cluster variability.

RESULTS AND DISCUSSION

Several differences were found in the allelic frequencies of the 5 breeds when considering the casein loci separately (Table 1). At *CSN1S1*, either the allele *A* or the allele *B* was predominant in the Southern breeds, whereas in the Northern breeds, the allele *F* showed the highest frequencies. If considered as a group, alleles ($A + B + C + H$) amounted to 57, 65, and 78% in Maltese, Jonica, and Garganica, respectively. This allelic distribution seems to characterize the goat populations from the Mediterranean area (Grosclaude and Martin, 1997; Tadlaoui Ouafi et al., 2002). On the other hand, the predominance of the allele *F* in the Vallesana and Roccaverano make them more similar to the Swiss breeds than to the French and Spanish breeds (Jordana et al., 1996; Grosclaude and Martin, 1997). The *H* allele, recently identified and differing from the *A* variant by the Arg_1 to Lys_1 substitution (Chianese et al., 1997), was detected at the protein level by IEF only in the

Table 2. Classification of different calcium-sensitive casein haplotypes on the basis of casein level (Rando et al., 2000). The seventh haplotype was found in the present work.

Quantitative haplotype	α_{s1}-casein (g/L)	β-casein (g/L)	α_{s2}-casein (g/L)	Total (g/L)
I	3.5	5	2	10.5
II	3.5	0	2	5.5
III	1.1	5	2	8.1
IV	0.5	5	2	7.5
V	0.5	5	0	5.5
VI	0	5	2	7
VII	0.5	0	2	2.5

Garganica breed. When typing is performed only at the DNA level, *CSN1S1*H* is confounded with *CSN1S1*A* because of the absence of a specific molecular test.

Null alleles were only found in Vallesana and Roccaverano, with a particularly high frequency of *CSN1S1*0*$_1$ (0.175) in Vallesana. The recently discovered allele *CSN1S1*N* (Ramunno et al., 2002) was detected only in Roccaverano and in the heterozygous condition.

No polymorphism for *CSN2* was observed in Roccaverano and Vallesana. Because milk IEF as well as the DNA typing methods do not allow the detection of the *CSN2*C* allele (Neveu et al., 2002), the *A** allele in Table 1 could include this genetic variant, which is associated with a normal β-casein content. In the Southern breeds, few heterozygous animals carried the null allele identified by Ramunno et al. (1995). No typing was done for the null allele characterized by Persuy et al. (1999), and it was found only in Creole and Pyrenean breeds. Interestingly, a Maltese goat heterozygous for *CSN2*0* was homozygous for the F allele at *CSN1S1* locus. This indicates the occurrence of a *CSN1S1*F-CSN2*0* haplotype, which is associated with a very low casein level even if it is linked to a strong *CSN1S2* allele. The casein content of this haplotype is around 2.5 g/L. This haplotype can be added to the 6 quantitative haplotypes observed at the goat calcium-sensitive casein loci (Table 2) (Rando et al., 2000) on the basis of the approximate casein level (Ramunno et al., 1995; Grosclaude and Martin, 1997; Ramunno et al., 2001a, b). The new haplotype found in Maltese could be exploited in eventual breeding programs aiming at producing milk with specific nutritional properties (i.e., less caloric value) rather than milk used for cheesemaking.

At *CSN1S2*, *C* and *F* variants were predominant, with different trends depending on the breed. The *F* allele, recently identified by Ramunno et al. (2001a), and resulting from a Val to Ile exchange in the seventh amino acid of the mature protein, is not distinguishable from *CSN1S2*A* and *CSN1S2*C* by SDS–PAGE (Ramunno et al., 2001a). Genotyping for the *CSN1S2*F* variant at the DNA level is recommended in goat casein investigations. This allele has a high frequency ranging from 0.250 in Maltese to 0.408 in Garganica. The frequency was also 0.261 (Ramunno et al., 2001a) in an undefined goat population. The PCR-RFLP developed by the same authors for *CSN1S2*0* identification allows the simultaneous detection of the *CSN1S2*D* allele (not found in our study), which is associated with a reduced α_{s2}-casein content. The null allele was found in Maltese and Jonica at a very low frequency (0.007 and 0.005, respectively). The typing at both the protein and DNA level allowed us to assess that the *CSN1S2*E* allele, characterized at the DNA level by Lagonigro et al. (2001), matches with the *G* variant identified at the protein level by Erhardt et al. (2002). In the study at DNA level, only one nonsynonymous mutation was found in *CSN1S2*E*, leading the authors to hypothesize an amino acid substitution Pro_{197} to Arg_{197} in the mature protein (Lagonigro et al., 2001). The IEF migration pattern found in the study at protein level (Erhardt et al., 2002) clearly suggested the non-correspondence between the observed *G* variant and the hypothesized *E* variant. However, the coincidence assessed in the present study between *CSN1S2*G*, typed at the protein level, and *CSN1S2*E*, typed at the DNA level, indicates the necessity to analyze the *CSN1S2*E* protein to identify the modifications responsible for the IEF migration pattern. The usefulness of investigating milk protein loci not only at the DNA level but also taking into account their phenotypic expression is evident.

Extensive polymorphism was observed at *CSN3*. The most common allele was the *CSN3*D* according to the nomenclature proposed by Chessa et al. (2003). Frequencies ranged from 0.596 (Vallesana) to 0.776 (Garganica). The allele was referred to as *B* in Yahyaoui et al. (2003). It will be important to find an international agreement on *CSN3* nomenclature in the future. The *CSN3*A* allele was also found in all breeds. This allele had variable frequencies ranging from 0.066 (Garganica) to 0.373 (Roccaverano). The *CSN3*B*, which is detectable at the protein level by IEF, results from the $Gln_{44} \rightarrow Arg_{44}$ exchange (Caroli et al., 2001) and was observed in the Southern breeds and in Roccaverano. It had the highest frequencies in Maltese (0.179) and Garganica (0.132). Only the Northern populations carried the *CNS3*C* allele. The frequency was <0.05. In addition, 2 new SSCP patterns were detected, and the relative alleles were named *X* and *Y*, according to Chessa et al. (2003). The molecular characterization of the 2 variants is in progress.

Hardy-Weinberg equilibrium was verified for almost all loci within populations. Deviations were found for *CSN1S1* and *CSN1S2* (Vallesana), *CSN1S2* (Maltese),

Table 3. Haplotype (*CSN1S1-CSN1S2-CSN3*[1]) frequencies in the different breeds. *CSN2* is monomorphic for the *A** allele. Frequencies >0.05 in at least one breed are reported. Frequencies >0.1 are bolded. Expected frequencies under independence hypothesis are shown in italics.

Haplotype	Vallesana	Roccaverano	Maltese	Jonica	Garganica
	(n = 66)	(n = 63)	(n = 62)	(n = 57)	(n = 30)
A A D			**0.107**	**0.164**	0.017
			0.087	*0.091*	*0.094*
A B B			0.088	0.016	
			0.010	*0.002*	
A C D		0.085	0.011	0.096	**0.100**
		0.069	*0.077*	*0.111*	*0.027*
A F D			0.029	0.025	0.067
			0.066	*0.076*	*0.019*
A E D			0.088		0.067
			0.033		*0.019*
B A A			0.027	0.027	0.050
			0.005	*0.012*	*0.011*
B A B		0.008		0.035	**0.100**
		0.001		*0.009*	*0.025*
B A D		0.024	0.022	0.012	**0.145**
		0.006	*0.028*	*0.068*	*0.130*
B C D	0.033	0.051	0.028	**0.153**	
	0.034	*0.032*	*0.066*	*0.082*	
B F D	0.009	0.011		0.011	**0.105**
	0.021	*0.017*		*0.083*	*0.131*
E B A		0.081			
		0.012			
E B D	0.053	0.027			
	0.018	*0.029*			
E C D	**0.272**	0.067	0.024	0.044	
	0.116	*0.076*	*0.089*	*0.021*	
F A D		0.018	**0.102**	0.023	0.055
		0.023	*0.082*	*0.081*	*0.089*
F C D	0.051	**0.131**	0.078		
	0.124	*0.115*	*0.073*		
F F A	**0.179**	0.069		0.011	
	0.047	*0.025*		*0.009*	
F F D	**0.117**	**0.106**	**0.188**	**0.216**	**0.211**
	0.074	*0.060*	*0.062*	*0.055*	*0.089*
O C A	**0.162**	0.024			
	0.046	*0.007*			

[1]*CSN1S1* = α_{s1}-Casein locus, *CSN1S2* = α_{s2}-casein locus, and *CSN3* = κ-casein locus.

and *CSN3* (Jonica). The F_{ST}-statistics ranged from 0.013 (CSN2) to 0.074, with an overall value of 0.062, and were significantly >0 for all loci. The F_{IS}-values were significantly >0 in 2 cases: for *CSN1S2* in Maltese (0.169) and for *CSN1S2* in Vallesana (0.136).

The casein haplotype frequencies are shown in Table 3. Haplotypes with a frequency >0.05 in at least one breed are reported. The 18 most common haplotypes were identified. When considering each breed separately, frequencies >0.05 were found only for 6 (Vallesana), 7 (Roccaverano), 6 (Maltese), 4 (Jonica), and 9 (Garganica) haplotypes. The expected frequencies under the independence hypothesis are also shown in Table 3. The χ^2 test was significantly different from the null hypothesis (no association) in Vallesana, Maltese, and Jonica, but no differences were found for Roccaverano and Garganica, indicating substantial linkage equilibrium in these breeds. These latter 2 breeds also had a higher haplotype number (7 and 9) compared with the other breeds. The haplotype *FFD* (in the order: *CSN1S1*F-CSN1S2*F-CSN3*D*) occurred in all breeds with a high frequency (>10%) and was the most common haplotype in the Southern breeds. The estimated frequencies of this haplotype were always much higher than expected frequencies. The distribution of the other haplotypes showed variability among the 5 breeds. In Vallesana and Roccaverano, *ECD* and *FCD* haplotypes were most prevalent, respectively. A high frequency of *OCA* haplotype was found in Vallesana (0.162), indicating that the null allele *CSN1S1*0* (frequency of 0.175) is mainly present within the casein cluster. Few haplotypes had expected frequencies higher than 0.1: *BFD* and *BAD* in Garganica, *FCD* and *ECD* in Vallesana, and *ACD* in Jonica.

The connection between the calcium-sensitive haplotypes described by Rando et al. (2000), and the frequencies of the haplotypes considered in our work is shown in Table 4. The first 10 haplotypes of Table 3 correspond

Table 4. Frequencies of the calcium-sensitive haplotypes in the different breeds. Only the haplotypes reported in Table 3 are considered (blanks = 0.000).

Haplotype	Vallesana	Roccaverano	Maltese	Jonica	Garganica
I	0.042	0.179	0.400	0.539	0.651
III	0.325	0.175	0.024	0.044	
IV	0.347	0.324	0.368	0.250	0.266
VI	0.162	0.024			

to the haplotype I of Table 2. This haplotype is characterized by the occurrence of strong alleles at *CSN1S1*, *CSN2*, and *CSN1S2* loci. Moreover, *EBA*, *EBD*, and *ECD* match with haplotype III; *FAD*, *FCD*, *FFA*, and *FFD* with haplotype IV; and OCA corresponds to haplotype VI. Much variability in the frequency of haplotypes occurs among breeds. Haplotype I ranges from 0.042 (Vallesana) to 0.651 (Garganica). Haplotype III occurs in the Northern breeds at rather high frequencies (0.325 in Vallesana and 0.175 in Roccaverano); its frequency is <0.05 in Southern breeds. Haplotype IV has a similar frequency in all breeds, ranging from 0.250 in Jonica to 0.368 in Maltese. Finally, haplotype VI occurs only in Vallesana (0.162) and Roccaverano (0.024).

The principal component analysis (Figure 1) clearly separates the Northern from the Southern breeds on the basis of the first component. This first component accounted for 51% of the total variability. The second component accounted for 20% of variability and distinguishes Maltese and Jonica from Roccaverano and Vallesana. Although Garganica occupies an anomalous position from a geographical point of view, it fits along the second axis of the scatter behind the 2 Northern breeds. Garganica differs from Maltese and Jonica but is similar to the Northern breeds with a low frequency of the *AAD* haplotype (0.017). The 3 Southern breeds show similar expected frequencies under the hypothesis of independence; the estimated haplotype frequencies diverge, indicating the occurrence of different linkage phases depending on the breed.

Finally, the third component from the principal component analysis accounted for 15% of the variability and is not related to the geographical area. It discriminates the 2 groups: Jonica and Roccaverano (negative values) vs. Maltese, Vallesana, and Garganica (positive values). The *BCD* haplotype frequency helps justify this clustering.

The evolutionary pathway of caprine *CSN1S1*, first proposed by Grosclaude et al. (1994), was slightly modified by Grosclaude and Martin (1997) and by Chianese et al. (1997). A putative B_1 allele was considered ancestral, and 2 divergent lineages were suggested, one leading to *A*, *G*, 0_1, 0_2, *I*, and *H* variants (lineage A), and the other leading to *B*, *L*, *F*, *C*, and *E* variants (lineage B). Later on, Bevilacqua et al. (2002) included the *M*

Table 5. Synthesis of the amino acid differences among the protein variants resulting from the 18 casein haplotypes. The bold type indicates the mutation occurrence referring to the possible ancestral haplotype *BAD*. AA = amino acid, Δ = AA deletion, § = missing protein, * = missing AA in the mature protein, and ° = AA position change in the mature protein because of deletion.

	CSN1S1 (AA position)					*CSN1S2* (AA position)				*CSN3* (AA position)			
Haplotype	16	77	100	195	Δ	7	64	167	193	44	65	119	159
AAD	**L**	**Q**	R	T		V	E	K	P	Q	V	I	S
ABB	**L**	**Q**	R	T		V	**L**	K	P	**R**	**I**	I	**P**
ACD	**L**	**Q**	R	T		V	E	**I**	P	Q	V	I	S
AFD	**L**	**Q**	R	T		**I**	E	K	P	Q	V	I	S
AED	**L**	**Q**	R	T		V	E	**I**	**R**	Q	V	I	S
OCA	**L***	**Q***	R*	T*	§	V	E	**I**	P	Q	V	**V**	S
BAA	P	E	R	T		V	E	K	P	Q	V	**V**	S
BAB	P	E	R	T		V	E	K	P	**R**	**I**	I	**P**
BAD	P	E	R	T		V	E	K	P	Q	V	I	S
BCD	P	E	R	T		V	E	**I**	P	Q	V	I	S
BFD	P	E	R	T		**I**	E	K	P	Q	V	I	S
EBA	P	E	**K**	**A**		V	**L**	K	P	Q	V	**V**	S
EBD	P	E	**K**	**A**		V	**L**	K	P	Q	V	I	S
ECD	P	E	**K**	**A**		V	E	**I**	P	Q	V	I	S
FAD	P	E*	R°	T°	**59–95**	V	E	K	P	Q	V	I	S
FCD	P	E*	R°	T°	**59–95**	V	E	**I**	P	Q	V	I	S
FFA	P	E*	R°	T°	**59–95**	**I**	E	K	P	Q	V	**V**	S
FFD	P	E*	R°	T°	**59–95**	**I**	E	K	P	Q	V	I	S

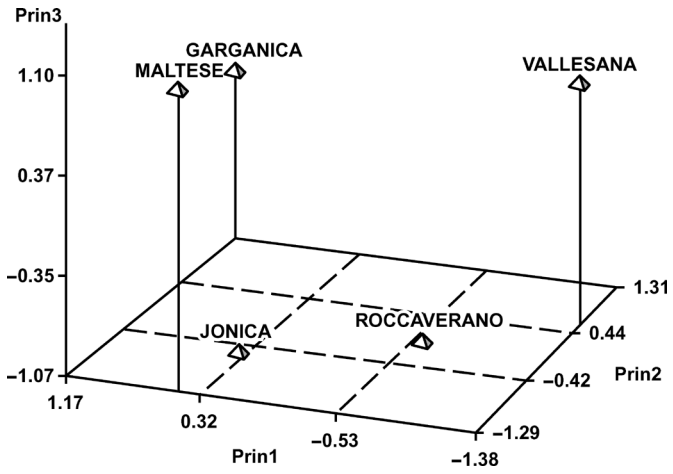

Figure 1. Principal (Prin) component analysis of the haplotype frequencies.

allele in the phylogeny, and they focused on the interallelic recombination event possibly responsible for this new variant.

As far as *CSN1S2* is concerned, an evolutionary pathway can be proposed starting from the *A* variant and leading independently to the *B*, *C*, and *F* alleles, each one 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 as the ancestral variant. If goat *CSN1S2*A*, *B*, and *C* (P33049) and goat *CSN1S2*E* (CAC21704) sequences are compared, one can postulate that *CSN1S2*E* variant derives from *CSN1S2*C*, because the 2 variants share an Ile, instead of Lys, at position 167 of the mature protein.

A phylogeny for *CSN3* was proposed by Yahyaoui et al. (2003) and by Jann et al. (2004). Apart from the conflicting nomenclature, both papers agree that *CSN3*A* appeared later in the evolutionary pathway and that 2 different lineages occurred.

Taking into account the amino acid differences among the protein variants resulting from the 18 casein clusters considered, the most likely ancestral haplotype could be the *BAD* (Table 5), followed by *BFD*, which gives rise to the most common *FFD*. However, strong evidence of recombination among the casein genes as well as of further intragenic recombination events arise if one considers the possible evolutionary patterns involving the 18 haplotypes.

CONCLUSIONS

The variability of the caprine casein loci and the resulting haplotypes should be exploited in future breeding schemes. Specific breeding programs that aim to preserve biodiversity and to select goat genetic lines for specific production situations could be implemented. Casein haplotypes should be used in selection instead of single loci. This approach to selection should help to exploit the effects of the entire casein cluster on milk yield and component traits. However, a complete definition of casein haplotypes is complex in goats because of the genetic variation found at the different casein loci. The frequency of defective alleles should be reduced in some populations but increased in other populations, depending on the destination of the milk (either for cheese making or fresh consumption). Mutations resulting in amino acid substitutions that change the protein isoelectric point, such as $Gln_{44} \rightarrow Arg_{44}$, revealed in *CSN3*, should be identified to clarify the functional importance of these substitutions. A screening of goat casein variability at the protein level is recommended for a more complete picture of the genetic polymorphisms in the casein genes.

ACKNOWLEDGMENT

This work was supported by MURST contract year 2001–prot. 2001077279 and year 2002–prot. 2002072131.

REFERENCES

Aleandri, R., L. G. Buttazoni, J. C. Schneider, A. Caroli, and R. Davoli. 1990. The effects of milk protein polymorphisms on milk components and cheese-producing ability. J. Dairy Sci. 73:241–255.

Angiolillo, A., M. H. Yahyaoui, A. Sanchez, F. Pilla, and J. M. Folch. 2002. Characterization of a new genetic variant in the caprine κ-casein gene. J. Dairy Sci. 85:2679–2680.

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.

Boettcher, P. J., A. Caroli, S. Chessa, E. Budelli, A. Stella, F. Canavesi, S. Ghiroldi, and G. Pagnacco. 2004. Effects of casein haplotypes on production traits in Italian Holstein and Brown Cattle. J. Dairy Sci. 87:4311–4317.

Boulanger, A., F. Grosclaude, and M. F. Mahé. 1984. Polymorphisme des caséines αS1 et αS2 de la chèvre (*Capra hircus*). Genet. Sel. Evol. 16:157–176.

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.

Brignon, G., M. F. Mahé, and F. Grosclaude. 1989. Sequence of caprine α_{S1}-CN and characterization of those of its genetic variants which are synthesized at a high level, α_{S1}-CN *A*, *B* and *C*. Protein Seq. Data Anal. 2:181–188.

Caroli, A., S. Chessa, P. Bolla, E. Budelli, and G. Gandini. 2004. Genetic structure of milk protein polymorphisms and effects on milk production traits in a local dairy cattle. J. Anim. Breed. Genet. 121:119–127.

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, K. Gutscher, A. Caroli, and G. Erhardt. 2003. 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 Milk Protein Polymorphism. International Dairy Federation, Brussels, Belgium.

Di Luccia, A., R. Mauriello, L. Chianese, L. Moio, and F. Addeo. 1990. κ-casein polymorphism in caprine milk. Sci. Tecn. Latt.-. Cas. 41:305–314.

Erhardt, G., S. Jäger, E. Budelli, and A. Caroli. 2002. Genetic polymorphism of goat α_{S2}-casein (*CSN1S2*) and evidence for a further allele. Milchwissenschaft 57:137–140.

Ferretti, L., P. Leone, and V. Sgaramella. 1990. Long range restriction analysis of the bovine casein genes. Nucleic Acids Res. 18:6829–6833.

Goudet, J. 1995. FSTAT, Version 1.2, A computer program to calculate F-statistics. J. Hered. 86:485–486.

Grosclaude, F., P. Joudrier, and M. F. Mahé. 1978. Polymorphisme de la caséine α_{S2} bovine: étroite liaison du locus α_{S2}-*Cn* avec les loci α_{S1}-*Cn*, β-*Cn* et κ-*Cn*, mise en évidence d'un délétion dans le variant α_{S2}-CnD. Ann. Genet. Sel. Anim. 10:313–327.

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. Casein polymorphisms in the goat. Pages 241–253 in Milk Protein Polymorphism. 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.

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.

Ikonen, T., H. Bovenhuis, M. Ojala, O. Ruottinen, and M. Georges. 2001. Associations between casein haplotypes and first lactation milk production traits in Finnish Ayrshire cows. J. Dairy Sci. 84:507–514.

Jann, O., E. M. Prinzenberg, G. Luikart, G. 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 alphas1-casein gene associated with a reduced level of protein synthesis. Gene 147:179–187.

Jordana, J., M. Amills, E. Diaz, C. Angulo, J. M. Serradilla, and A. Sanchez. 1996. Gene frequencies of caprine α_{S1}-casein polymorphism in Spanish goat breeds. Small Rumin. Res. 20:215–221.

Lagonigro, R., E. Pietrola, M. D'Andrea, C. Veltri, and F. Pilla. 2001. Molecular genetic characterization of the goat s2-casein E allele. Anim. Genet. 32:391–393.

Leroux, C., and P. Martin. 1996. The caprine alpha s1 and beta–casein genes are 12-kb apart and convergently transcribed. Anim. Genet. 27(Suppl. 2):93.

Leroux, C., P. Martin, M. F. Mahé, H. Levéziel, and J. C. Mercier. 1990. Restriction fragment length polymorphism identification of goat α_{S1}-casein alleles. A potential tool in selection of individuals carrying alleles associated with a high level protein synthesis. Anim. Genet. 21:341–351.

Leroux, C., N. Mazure, and P. Martin. 1992. Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat alpha$_{s1}$-casein transcripts. Structural organization of the relevant gene. J. Biol. Chem. 267:6147–6157.

Mahé, M. F., and F. Grosclaude. 1989. α_{S1}-CnD, another allele associated with a decreased synthesis rate at the caprine α_{S1}-casein locus. Genet. Sel. Evol. 21:127–129.

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.

Martin, P., and C. Leroux. 1994. Characterization of a further a$_{S1}$-casein variant generated by exon-skipping. Proc. Int. Soc. Anim. Genet. Conf., Prague, The Netherlands. Anim. Genet. 25(Suppl. 2):86.

Martin, P., M. Szymanowska, L. Zwierzchowski, and C. Leroux. 2002. The impact of genetic polymorphisms on the protein composition of ruminants milks. Reprod. Nutr. Dev. 42:433–459.

Neveu, C., D. Mollé, J. Moreno, P. Martin, and J. Léonil. 2002. Heterogeneity of caprine beta-casein elucidated by RP-HPLC/MS: Genetic variants and phoshorylation. 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. 30:444–451.

Ramunno, L., G. Cosenza, D. Gallo, R. Illario, A. Rando, and P. Masina. 2002. Un nuovo allele al locus *CSN1S1* di capra: *CSN1S1* N. Page 220 in Proc. Atti XV Conv. SIPAOC, Cagliari, Italy.

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., E. Longobardi, G. Cosenza, P. Di Gregorio, A. Rando, and P. Masina. 1999. A *Pst*I PCR-RFLP at the goat *CSN1S2* gene. Anim. Genet. 30:242.

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 alphas2-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. ASPA, Grado, 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.

Rijnkels, M., P. M. Kooiman, H. A. de Boer, and F. R. Pieper. 1997. Organisation of the bovine casein gene locus. Mamm. Genome 8:148–152.

SAS User's Guide: Statistics, Version 6, 4th Edition. 1990. SAS Inst., Inc., Cary, NC.

Tadlaoui Ouafi, A., J.-M. Babilliot, C. Leroux, and P. Martin. 2002. Genetic diversity of the two main Moroccan goat breeds: Philogenetic relationships with four breeds reared in France. Small Rumin. Res. 45:225–233.

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.

Yahyaoui, M. H., A. Coll, A. Sanchez, and J. M. Folch. 2001. Genetic polymorphism of the caprine kappa casein gene. J. Dairy Res. 68:209–216.