Caprine κ-Casein (*CSN3*) Polymorphism: New Developments in Molecular Knowledge

E.-M. Prinzenberg,[1] K. Gutscher,[1] S. Chessa,[2] A. Caroli,[2,3] and G. Erhardt[1]
[1]Institut für Tierzucht und Haustiergenetik, Justus-Liebig Universität, 35390 Gießen, Germany
[2]Dipartimento di Scienze e Tecnologie Veterinarie per la Sicurezza Alimentare,
Università degli Studi di Milano, 20134 Milan, Italy
[3]Dipartimento di Sanità e Benessere Animale, Università di Bari, 70010 Valenzano, Italy

ABSTRACT

A high degree of polymorphism was recently found at the κ-casein (*CSN3*) locus in the domesticated goat (*Capra hircus*). In the present study, 2 new patterns previously identified by PCR-single-strand conformation polymorphism analysis (SSCP) were characterized. The allele provisionally named "*X*" (GenBank Accession no. AY350425) differs from *CSN3*C* (AF485341) by a (silent) A→G substitution at position 509 of the goat *CSN3* reference sequence (X60763). As this newly identified sequence changes the amino acid sequence, and the already known *CSN3*C* allele (AF485341) has an additional silent mutation, we proposed a change in nomenclature to reflect these changes, indicating the silent mutation with the prime symbol (i.e., ′). The *CSN3*M* allele (provisionally named "*Y*") results in a new protein variant, differing by 2 nonsynonymous mutations from the *CSN3*F* allele. The new variant is characterized by a G→A transition at nucleotide position 384, resulting in the amino acid exchange Asp$_{90}$→Asn$_{90}$, and a C→T transition at position 550, resulting in a Val$_{145}$→Ala$_{145}$ substitution. Thus, the number of alleles identified in the domesticated goat has increased to 16, of which 13 are protein variants and 3 are silent mutations, involving a total of 15 polymorphic sites in *CSN3* exon 4. Data on the distribution of the main alleles in 7 goat breeds of Europe, West Africa, and the Near East show differences in the occurrence and frequency of the alleles between breeds and geographic origin with the highest number of alleles found in goat breeds from the Near East.
(**Key words:** goat, κ-casein polymorphism, single-strand conformation polymorphism, nomenclature)

Abbreviation key: *CSN3* = κ-casein gene, **dsPCR** = double-stranded PCR product, **IEF** = isoelectric focusing, **IP** = isoelectric point, **SSCP** = single-strand confor-

mation polymorphism, **ssPCR** = single-stranded PCR product.

INTRODUCTION

Caseins appear to be a rapidly evolving gene family, presumably due to the minimal structural requirements for function (Bonsing and Mackinlay, 1987). However, in contrast to the calcium-sensitive α_{s1}-, β-, and α_{s2}-caseins, the κ-casein fraction plays a crucial role in the formation, stabilization, and aggregation of the casein micelles and thus affects the technological (Mariani et al., 1976; Aleandri et al., 1990; Lodes et al., 1996; Falaki et al., 1997) and nutritional properties of milk (Mercier et al., 1973, 1976; Malkoski et al., 2001). Therefore, the κ-casein gene (**CSN3**) is unlikely to be completely free of selective constraints (Ward et al., 1997). Interspecies comparisons have demonstrated that *CSN3* possesses the highest degree of conservation among the casein genes, which may be related to its essential function as a stabilizer of casein micelles (Alexander et al., 1988). On the other hand, high intraspecies variability has been reported for *CSN3* in cattle (for a review see Formaggioni et al., 1999) and, more recently, in goats.

Since the discovery of 2 κ-casein protein variants in goats (Di Luccia et al., 1990), successively confirmed at the protein and DNA level by Caroli et al. (2001), additional *CSN3* polymorphisms have been identified by DNA analysis (Yahyaoui et al., 2001; Angiolillo et al., 2002; Yahyaoui et al., 2003; Jann et al., 2004), and different typing methods were developed (Yahyaoui et al., 2001; Chessa et al., 2003; Yahyaoui et al., 2003). A total of 13 polymorphic sites have been identified in the domesticated goat (Jann et al., 2004), allowing for the identification of 14 DNA variants corresponding to 11 protein variants, if one considers the variant first observed in *Capra pyrenaica* (Yahyaoui et al., 2001) and later in domesticated goat breeds (Yahyaoui et al., 2003). The high number of the caprine *CSN3* polymorphisms discovered and sequences available during the last years have led to inconsistency in nomenclature.

Misunderstandings due to different names for the same variants or identical names shared by different alleles have been widely discussed and new nomenclature has been proposed (Chessa et al., 2003; Jann et al., 2004), based primarily on the GenBank chronological order of the variants.

Recently, 2 additional polymorphisms were identified by PCR–single-strand conformation polymorphism (**SSCP**) analysis (Chessa et al., 2003) but not yet characterized. The aims of the present work were 1) the characterization of the mutations responsible for the polymorphisms identified, 2) the development of a simultaneous typing procedure for recently identified *CSN3* variants (Chessa et al., 2003; Jann et al., 2004) by PCR-SSCP analysis, and 3) a modification of the *CSN3* nomenclature considering the order of reports, allele distribution, and sequence evolution. Finally, the distribution of *CSN3* alleles in 7 goat breeds of central Europe, west Africa, and the Asian part of Turkey is presented.

MATERIALS AND METHODS

DNA Sequencing

The DNA samples of animals showing the "*X*" or "*Y*" PCR-SSCP patterns (Chessa et al., 2003) were chosen for direct sequencing of PCR products. Primers derived from the bovine *CSN3* sequence were used for amplification of the entire coding region in exon 4 (forward: 5′ AGAAATAATACCATTCTGCAT 3′ and reverse: 5′ GT TGTCTTCTTTGATGTCTCCTTAGAG 3′; Prinzenberg et al., 1999). The total reaction volume was 50 μL, containing 10 pmol of each primer, 80 μM of each deoxyribonucleotide, and 1 U of Taq DNA polymerase (Eppendorf AG, Hamburg, Germany) in the buffer supplied by the manufacturer, with $MgCl_2$ added to a final concentration of 2.5 mM. The temperature profile was 90 s for initial denaturation at 94°C, 30 cycles of 45 s each at 94, 56.5, and 72°C, and a final 5-min, 72°C elongation step. The PCR products were purified after excision of a 1% agarose gel using the Montage Gel Extraction kit (Millipore GmbH, Schwalbach/Ts, Germany). A modified Tris-acetate/EDTA buffer for direct use of the eluted DNA was supplied with the kit and cycle sequencing of both strands was done with the Big Dye Terminator Kit 1.1 (Applied Biosystems, Darmstadt, Germany) using 5 pmol of each of the PCR primers. Sequences were separated on an ABI Prism 377 sequencer and data were processed using free Chromas 1.45 (www.technelysium.co.au) and GeneDoc 2.6 (Nicholas and Nicholas, 1997) software.

In Silico Sequence Analyses

Calculation of isoelectric points (**IP**) of the protein corresponding to the DNA sequences of *Capra hircus CSN3* was done with the DAMBE program (Xia and Xie, 2001). The 9 amino acids encoded in exon 3 were assumed identical in all variants for this calculation. The DAMBE software was also used to propose the nucleotide sequence-based maximum parsimony trees. For this purpose, DAMBE uses the algorithm originally implemented in the DNAPARS option of the PHYLIP program package (Felsenstein, 1989). To get consensus trees, the input order was randomized (5 times) and bootstrapping and jackknifing was applied (50, 100, or 1000 datasets tested).

PCR-SSCP Analysis

The PCR and subsequent SSCP analysis of 407 bp of exon 4 was carried out using primers KCN I F (5′ GGTATCCTAGTTATGGACTCAAT 3′) and KCN I R (5′ GTTGAAGTAACTTGGGCTGTGT 3′) according to Chessa et al. (2003). Reference samples for all the *Capra hircus CSN3* variants, except for the *CSN3*F* variant, were used as standards. For the *CSN3*H* allele, a plasmid clone (described in Jann et al., 2004) was used as standard. Additionally, longer PCR products (459 bp) were generated for reference samples using primers Kb1 (5′ TGTGCTGAGTAGGTATCCTAGT-TATGG 3′) and Kb2 (5′ GCGTTGTCCTCTTTGATGTC-TCCTTAG 3′) as described by Yahyaoui et al. (2001). These were also used for a second PCR with only forward (Kb1) or only reverse (Kb2) primer and 3 μL of the double-stranded PCR (**dsPCR**) product as starting template, which generates single-stranded PCR products (**ssPCR**). Analysis by SSCP was done using the same conditions as for dsPCR products.

PCR-RFLP Analysis

The mutation in position 550 of the *CSN3*M* allele creates a *Pst*I recognition sequence and was verified by restriction enzyme digestion of 5 to 7 μL of the 407-bp PCR product with 1.5 U of *Pst*I (MBI Fermentas, St. Leon-Roth, Germany) in 20 μL total volume of the buffer supplied by the manufacturer at 37°C. The expected fragments were 334 + 73 bp for *CSN3*M* and 407 bp for *CSN3*F* and all other alleles and were separated by standard agarose gel electrophoresis (1.5% agarose in 0.5× Tris-borate/EDTA buffer).

DNA Samples for Allele Frequency Determination

Two hundred thirty-three DNA samples from Weisse Deutsche Edelziege (Germany, n = 32), Saanen (Italy,

Table 1. κ-Casein gene (*CSN3*) variants in the domesticated goat according to the modified nomenclature proposed in this paper. The code for isoelectric focusing (IEF) migration pattern, GenBank accession numbers, nucleotide differences compared with GenBank accession no. X60763, and published references are indicated. Synonymous mutations are shown in italics and mutations changing the amino acid sequence are shown in normal font.

CSN3 variant	IEF pattern	GenBank Acc. No.	Nucleotide position															Reference
			170	245	247	274	284	290	298	309	384	385	471	509	550	583	591	
A	A^{IEF}	X60763	C	T	A	A	G	C	A	G	G	A	G	A	T	C	T	Coll et al., 1993
B	A^{IEF}	AF485340											A					Yahyaoui et al., 2001
		AF434988											A					Jann et al., 2004
B′	A^{IEF}	AY166706	*T*										A					Jann et al., 2004
B″	A^{IEF}	AY166707					*T*						A					Jann et al., 2004
C	A^{IEF}	AY350425		*C*			A			A			A			T	C	This paper
C′	A^{IEF}	AF485341		*C*			A			A			A	*G*		T	C	Yahyaoui et al., 2001
D	B^{IEF}	AY027868		*C*	G					A			A				C	Caroli et al., 2001
		AY090465		*C*	G					A			A				C	Yahyaoui et al., 2001
E	B^{IEF}	AF486523										G	A					Angiolillo et al., 2002
F	A^{IEF}	AY090466		*C*									A				C	Yahyaoui et al., 2003
G	A^{IEF}	AY090467		*C*						A			A				C	Yahyaoui et al., 2003
H	A^{IEF}	AF521022				G							A					Jann et al., 2004
I	A^{IEF}	AY166710								A			A					Jann et al., 2004
J	A^{IEF}	AY166711						G					A					Jann et al., 2004
K	B^{IEF}	AY166709			G								A					Jann et al., 2004
L	A^{IEF}	AY166708								A			A				C	Jann et al., 2004
M	B^{IEF}	AY428577		*C*							A		A		C		C	This paper

n = 22), Borno goat (Nigeria, n = 29), Red Sokoto (Cameroon/Nigeria, n = 31), West African Dwarf (Cameroon/Nigeria, n = 26), Hair goat (Anatolia, Turkey, n = 50), and Angora goat (Anatolia, Turkey, n = 43) were genotyped using the PCR-SSCP procedure according to Chessa et al. (2003). Breed characterizations are available at http://dad.fao.org or http://www.ansi.okstate.edu/breeds/goats/. The Turkish Hair goat is also known as Turkish Native and listed in the Oklahoma State University web page as Anatolian Black. The Borno goat, which is not reported in the FAO web site, is an indigenous population reared in Nigeria. Altogether, these breeds from central Europe, western Africa, and the Asian part of Turkey were chosen to represent geographically distant and separated populations. Allele frequencies were determined by direct counting.

Assignment of New Allele Names

In the case of a specific amino acid substitution, the letter describing a new allele was chosen in alphabetical order following the last reported allele. Alleles with silent nucleotide substitutions were named with the identical letter as the related protein-associated allele and a prime symbol was added.

Changes in Nomenclature

The original allele names were maintained in all cases of unequivocal assignment of a protein-coding DNA variant. For different alleles with identical names given by 2 authors and in the case of different names for the same alleles, the name reported first was chosen (reassigned) in all variants observed in domesticated goats and not only in wild species so far. Alleles resulting from silent mutations were renamed and marked with a prime symbol according to the guideline above for new allele names.

RESULTS AND DISCUSSION

Two new *CSN3* variants were characterized by DNA sequencing and sequences were entered to the GenBank with accession numbers AY350425 (formerly "*X*", named *C* here) and AY428577 (formerly "*Y*", named *M* here). The nucleotide substitutions in all the domesticated goat variants described until now, including the 2 newly characterized alleles, are summarized in Table 1. In addition, a partially modified version of the nomenclature proposed by Jann et al. (2004) is given. The AY090466 sequence is now indicated as *F* because it was found not only in the wild *Capra pyrenaica* (Yahyaoui et al., 2001) but also in domesticated goats (Yahyaoui et al., 2003). For this reason, the first nomenclature deposited in the GenBank is maintained (reassigned), whereas the later deposited AY166708 variant, which was initially named *F* in the GenBank and in the nomenclature proposed by Jann et al. (2004), is now changed to *L*. Moreover, *CSN3*D*, *D′*, and *D″*, which differ from *CSN3*A* at the protein level only by the substitution Val_{119} (*CSN3*A*) to Ile_{119}, are now renamed *CSN3*B*, *B′*, and *B″*, which takes into account the widespread occurrence of the resulting protein variant in the different breeds analyzed and the evolutionary trend of

these variants (Chessa et al., 2003; Yahyaoui et al., 2003). The *CSN3*B* allele later assigned by Jann et al. (2004) is now renamed *CSN3*D*, according to the earlier published report (Yahyaoui et al., 2003).

The allele corresponding to the "*X*" SSCP pattern described by Chessa et al. (2003) is a variant of *CSN3*C* (GenBank Accession no. AF485341), sharing the identical amino acid sequence. In comparison with the reference sequence *CSN3*A* (Coll et al., 1993; GenBank Accession no. X60763), *C* and "*X*" alleles show 6 common point mutations (Table 1), whereas an additional A→G substitution, not changing the amino acid, occurs in *CSN3*C* (AF485341) at position 509. The "*X*" variant does not include this synonymous mutation and therefore represents a molecular event preceding *CSN3*C* in the phylogenetic evolution of the gene. At the protein level, *CSN3*C* and "*X*" code for the same variant. We therefore propose to change the nomenclature of *CSN3*C* (AF485341) to *CSN3*C′* and use the name *CSN3*C* for the provisionally named "*X*" variant. In doing so, we maintain the rule of indicating silent mutations by using a prime symbol for the silent allele and a regular letter for the allele corresponding to the protein variant. This underlines again the problem in establishing an allele nomenclature meeting current standards, logic, and discovery timeline aspects—especially for genes with rapidly growing allele numbers.

The *CSN3*M* allele, which corresponds to the previously named "*Y*" pattern (Chessa et al., 2003), results in a new protein variant, differing in 2 nonsynonymous mutations from the *CSN3*F* allele (Table 1). A G→A transition at nucleotide position 384, resulting in the amino acid exchange Asp_{90}→Asn_{90}, and an additional C→T transition at position 550, resulting in a Val_{145}→Ala_{145} substitution, characterize the new variant. The first mutation, involving the exchange of an acid Asp to a neutral Asn, changes the IP compared with the *F* variant. Nevertheless, it cannot be identified by isoelectric focusing (**IEF**) of milk samples (Chessa et al., 2003). Calculation of the IP (DAMBE program) for all κ-casein variants found in domesticated goats so far reveals only 2 different IP, and the κ-casein variants cluster in 2 groups: *D*, *E*, *K*, *M* (IP = 5.66) and *A*, *B*, *B′*, *B″*, *C*, *C′*, *F*, *G*, *H*, *I*, *J*, *L* (IP = 5.29). In fact, in IEF of milk samples, only 2 patterns are visible, corresponding to these 2 IP groups. The substitution of Asp to Asn shifts the *M* variant compared with *F* to the IP = 5.66 group, but does not enable specific discrimination by IEF. Therefore, we propose to differentiate the nomenclature at the protein level from the one used at the DNA level, by introducing 2 codes corresponding to the IP groups: A^{IEF}, comprising the DNA variants *A*, *B*, *B′*, *B″*, *C*, *C′*, *F*, *G*, *H*, *I*, *J*, *L*, and B^{IEF}, corresponding to DNA variants *D*, *E*, *K*, and *M* (Table 1, all allele names

correspond to the newly proposed nomenclature in this paper). In this way, conflicts arising from allele names A and B referring to different variants at the DNA level, and in previous studies carried out by milk protein IEF typing (Di Luccia et al., 1990; Chianese et al., 2000; Caroli et al., 2001) will be avoided. Moreover, an interesting difference was suggested by Chianese et al. (2000) between the 2 κ-casein IEF variants, as B^{IEF} seems to be associated with higher casein content in milk than A^{IEF}. This finding is probably because, until now, B^{IEF} had been found only in haplotypes with strong alleles at α_{s1}-casein (*CSN1S1*) and α_{s2}-casein (*CSN1S2*) loci. Only 2 haplotypes have been detected carrying the *CSN3*B* allele, renamed *CSN3*D* in the present paper: *CSN1S1*A- CSN1S2*B- CSN3*D* and *CSN1S1*B-CSN1S2*A- CSN3*D* (Sacchi et al., 2005). The importance of monitoring the occurrence of κ-casein B^{IEF} in goat milk by routine screening at the protein level is therefore evident.

Maximum parsimony analysis of sequences listed in Table 1 generated 379 most parsimonious trees if no resampling but randomized input order was used. This result underlines the complexity of the mutation events. After bootstrapping and jackknifing options were applied, the same resulting consensus tree was always generated with bootstrapping of either 50 or 100 datasets and jackknifing of either 50 or 1000 datasets. This consensus tree, indicating a probable nucleotide sequence based evolution of the alleles known up to now, is shown in Figure 1.

As already shown (Chessa et al., 2003), the 2 newly characterized variants are clearly identified by PCR-SSCP analysis (Figure 2), which allows the simultaneous typing of the *CSN3*B″*, *CSN3*H*, and *CSN3*J* alleles identified by Jann et al. (2004). Because overlapping occurs for *CSN3*D/I/K/L* and for *CSN3*B/B′* (results not shown), 11 different SSCP patterns are identifiable, corresponding to 9 single alleles (*A*, *B″*, *C*, *C′*, *E*, *G*, *H*, *J*, *M*) as well as the 2 just-mentioned variant groups. However, due to the lack of reference DNA samples, the PCR-SSCP pattern of *CSN3*F* could not be demonstrated in the present study. When typing by the single nucleotide primer extension method proposed by Yahyaoui et al. (2003), the possibility of confounding *F* and *M* alleles should be taken into account. This typing procedure does not consider the 2 mutations discriminating between *F* and *M* variants. Sequencing data clearly demonstrated these differences occurring between *F* and *M* alleles and PCR-RFLP analysis with *Pst*I discriminated *CSN3*M* from all other variants (results not shown). By using PCR-RFLP, we were able to verify *M* in all "*Y*" samples and exclude presence of *CSN3*F*. The occurrence of *CSN3*F* in domesticated goats is therefore questionable again.
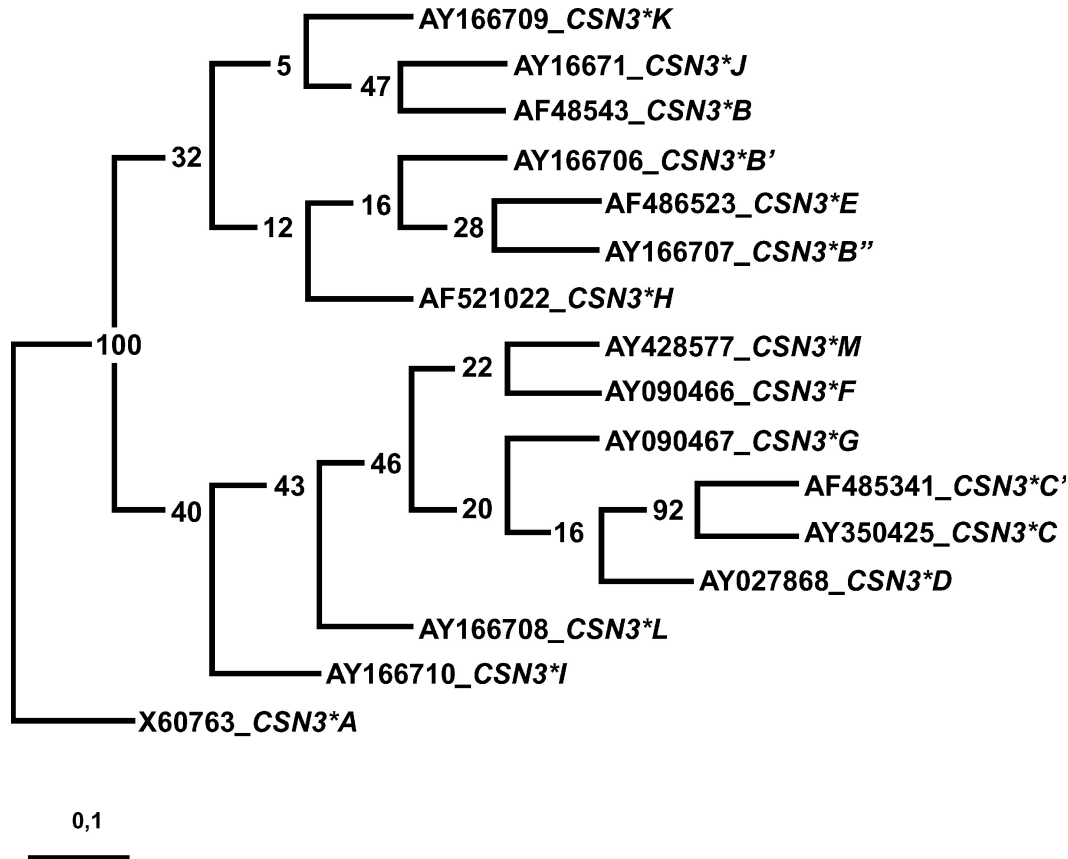
Figure 1. Nucleotide sequence-based maximum parsimony consensus tree indicating relationship and possible evolution of the DNA sequence-based alleles in *Capra hircus CSN3*. This consensus tree was generated with DAMBE, *CSN3*A* / X060763 was set as outgroup, and resampling statistics were applied. The resulting tree shown is the same with 50 or 100 data set bootstrapping or 50 to 1000 data set jackknifing.

Based on the nucleotide sequence data, the PCR-RFLP procedure developed for genotyping of *CSN3*A*, *B*, and *C* (renamed *C′* here) using *Bse*NI and *Alw*44I (Yahyaoui et al., 2001) will enable further discrimination within the *D/I/K/L* group after SSCP analysis. Due to the G in nucleotide position 309, *Bse*NI will cut in *CSN3*K* only and not in *CSN3*D*, *I*, and *L*, which share 309/A. This enzyme consequently discriminates *CSN3*K* from the rest of the *D/I/K/L* group. The second enzyme, *Alw*44I, has recognition sites in *CSN3*D* and *L* (591/C) but not in *CSN3*I* and *K* (591/T). In contrast to the initial observation in French Saanen and in 5 Spanish breeds (Yahyaoui et al., 2001), the mutations in position 309 and 591 are not always linked, as shown in the *CSN3*I* sequence (Jann et al., 2004) and therefore use of both enzymes in PCR-RFLP analysis will provide more information and enable specific discrimination of *CSN3*I* and *K*.

Analysis of ssPCR by SSCP (Figure 3) showed that the set of fast-migrating bands corresponds to the sense strand generated by the forward primer, whereas the slower migrating—and for genotyping purposes, more informative—set of bands represents the antisense strand (revealed by PCR with the reverse primer). A simple explanation for the higher SSCP variability in the antisense strand might be the difference in GC content, which is 44% in the sense strand and 56% in the antisense strand. Higher GC content might increase the frequency and stability of secondary structures here. The total number of bands visible was reduced by more than 50% when ssPCR was analyzed. This is a strong hint that additional conformers are visible in SSCP analysis of dsPCR, which are probably the result of an interaction of both strands. A similar phenomenon was described by Orita et al. (1989) but not, to our knowledge, further examined until now. Differentiation between variants within the *D* group and between *CSN3*B* and *B′* was not improved by ssPCR-SSCP analysis compared with dsPCR-SSCP analysis. Moreover, discrimination of *CSN3*G* and *H* is essentially based on the distance between the major and the secondary band (see Figure 2). This result was reproducible in
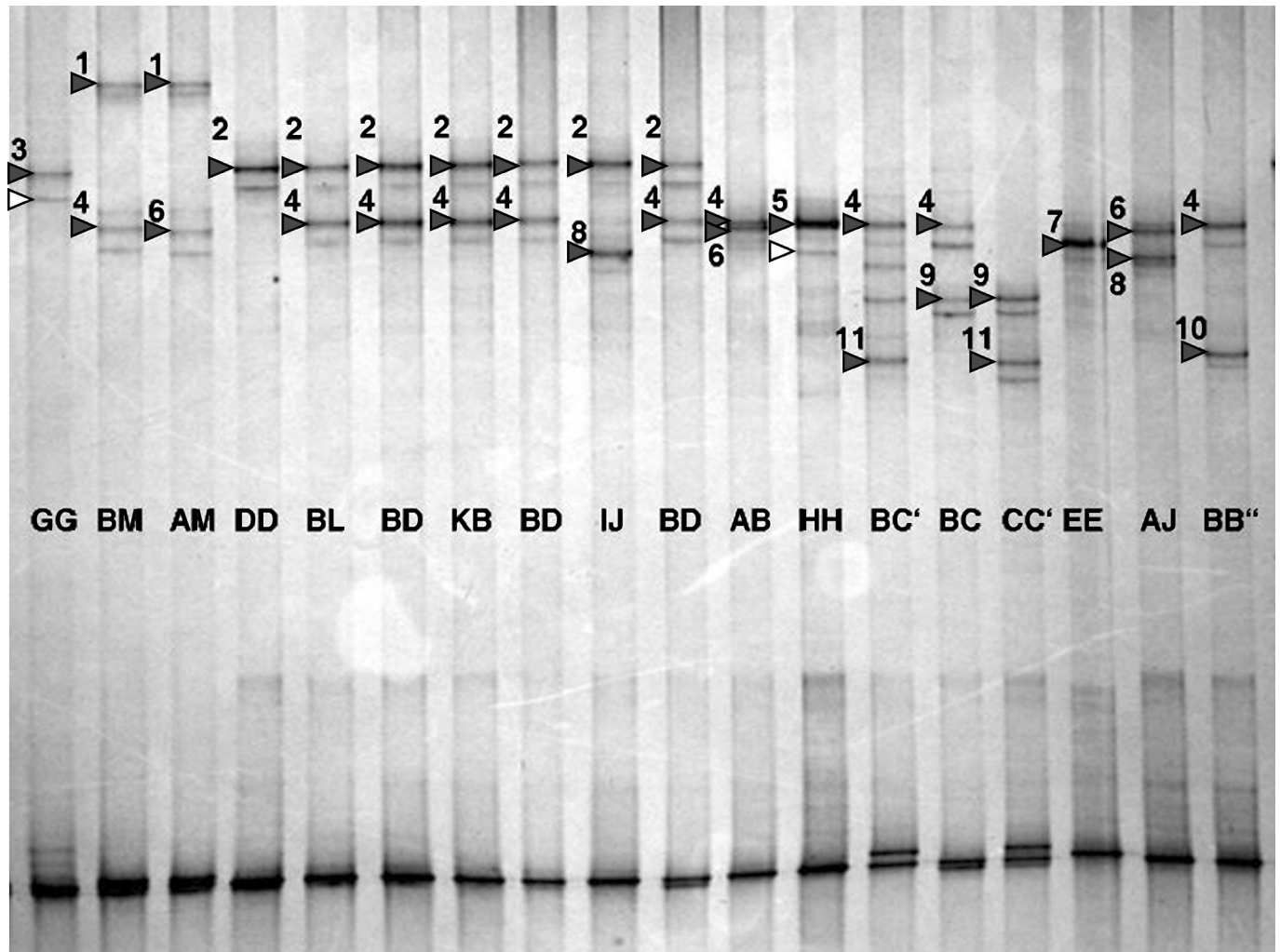
Figure 2. Double-strand PCR single-strand conformation polymorphism analysis of different κ-casein (*CSN3*) standard samples in a 9.25% polyacrylamide gel (29:1) with 0.5% glycerol using the 407-bp fragment. Major bands of the upper strand specific for alleles or allele groups are marked by solid triangles and numbered according to increasing migration: 1 = *M*, 2 = *D/I/K/L*, 3 = *G*, 4 = *B/B′* (*B′* not shown here), 5 = *H*, 6 = *A*, 7 = *E*, 8 = *J*, 9 = *C*, 10 = *B″*, 11 = *C′*. The genotypes resulting from sequencing data are given in the middle of the gel. The open triangles indicate the position of the secondary band in the *G* and *H* variants, which clearly migrated further than those in the *D/I/K/L* and *B/B′* groups, respectively (and proved to be reliable in differentiating these). Formation and migration of the secondary bands varied between different alleles and samples and was very intense in the *CC′* sample here. The bands on the bottom, compressed due to migration with the glycerol front, enabled differentiation of *CSN3*C′* and *E*, but no other alleles.

our laboratories in Germany and Italy. Therefore, a conventional dsPCR seems to be generally more useful for genotyping by PCR-SSCP in silver-stained gels. On the other hand, knowledge on the migration of bands corresponding to the sense or antisense strand is essential information if analyses on automated sequencers using one fluorescent labeled primer are to be applied. In the case of *CSN3*, labeling of the reverse primer is strongly recommended for fluorescence-based SSCP analysis.

The SSCP technique with silver staining, although a powerful mutation screening method, reaches limitations in simultaneous discrimination of high numbers

of alleles. At present, a combination of SSCP and subsequent RFLP analyses of the same PCR product for unresolved groups would provide the most precise results, but this approach is time consuming. Higher resolution methods should be implemented to get a complete view of caprine *CSN3* variability and, in all cases, the influence of the diagnostic tools on allele discrimination should be considered when allele frequencies are determined.

The analysis of the *CSN3* variation in 7 goat breeds from Europe, Africa, and the Asian part of Turkey indicates that only a few alleles occur at a rather high frequency. Only 7 out of 11 *CSN3* SSCP patterns de-
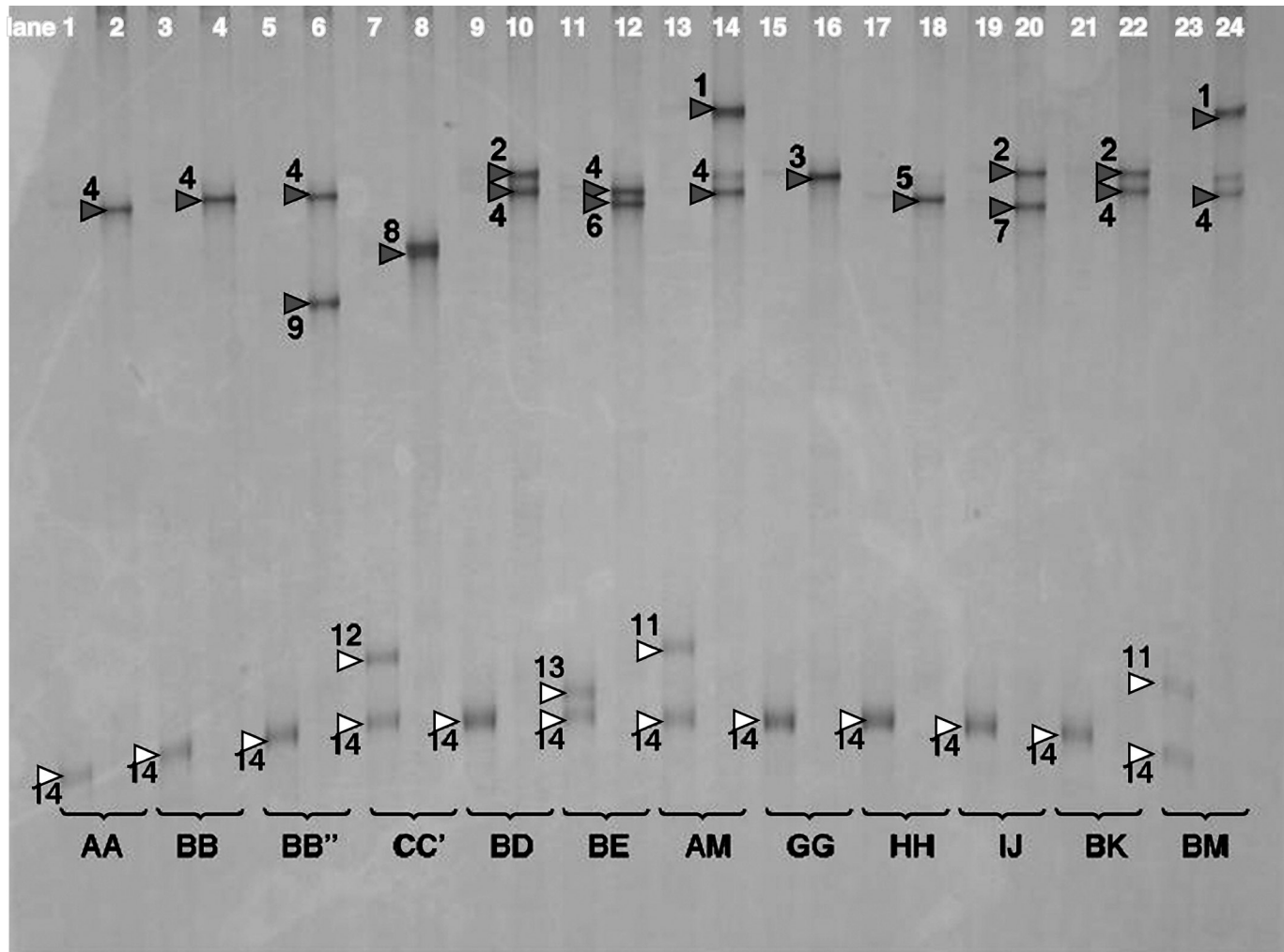
Figure 3. Single-strand PCR (ssPCR) single-strand conformation polymorphism analysis of all goat κ-casein (*CSN3*) variants, except *B′* and *F*. The fragment used was 459 bp long and samples were run on a 9.25% polyacrylamide gel (29:1) with 1% glycerol at 5°C. Single-strand PCR products of the sense strand (generated by forward primers) were in odd lanes and marked with open triangles. They showed faster migration and less variation than the ssPCR generated by reverse primers (in even lanes, marked with solid triangles). Numbering was according to increasing mobility: 1 to 9 for the antisense strand and 11 to 14 for the sense strand. The antisense strand discriminated 1 = *M*, 2 = *D/I/K/L*, 3 = *G*, 4 = *A/B/B′* (*B′* not shown), 5 = *H*, 6 = *E*, 7 = *J*, 8 = *C/C′*, 9 = *B″*. The sense strand enabled discrimination of single alleles 12 = *C′*, 13 = *E*, and 11= *M*, whereas 14 = *A*, B, *B′*, *B″*, *C*, *D*, *G*, *H*, *I*, *J*, *K*, *L* comprised a large group of alleles. Compared with Figure 1, the larger fragment used here did not allow specific separation of *CSN3*A* and *B/B′*. Discrimination of band 6 and 7 was weak, but *CSN3*E* was clearly identified by band 13 in the sense strand.

scribed here were found among these breeds (Table 2), as the *CSN3*B″*, *E*, *H*, and *J* variants were not observed. The prevalent SSCP pattern was *CSN3*B/B′*, with frequencies ranging from 0.260 (Hair goat) to 0.674 (Angora goat). The second most common allele was *CSN3*A*, with frequencies ranging from 0.151 (Angora goat) to 0.414 (Borno goat). No additional new SSCP patterns were detected. Generally, the African goats showed the lowest variation for *CSN3* with only the 2 major alleles detectable in Borno goat and West African Dwarf and additionally *CSN3*M*, which was found only in the Red Sokoto. The "*D* group" (*D/I/K/L*) pattern

occurred in only 3 breeds, at a low frequency (0.047) in the Weisse Deutsche Edelziege, slightly higher in Angora goats (0.116), and at an intermediate frequency in the Hair goats (0.250). A similar breed distribution was found for *CSN3*C* (provisionally "*X*"), which was present at low frequencies (0.03 to 0.035) in the Weisse Deutsche Edelziege and the 2 Turkish breeds, whereas the silent allele *CSN3*C′* was found rarely in Weisse Deutsche Edelziege, Saanen, and Hair goats (0.01 to 0.09), but not in the Angora goat. The *CSN3*G* allele was most frequent in the Hair goat, which was the most diverse breed among those investigated here, with 6

Table 2. Allele frequencies for the κ-casein (*CSN3*) locus in 7 breeds analyzed. Allele names refer to the nomenclature as proposed in this paper.

CSN3 SSCP[2] pattern	Allele frequency per breed[1]						
	WDE (n = 32)	SA (n = 22)	AG (n = 43)	HG (n = 50)	BG (n = 29)	RS (n = 31)	WAD (n = 26)
A	0.344	0.250	0.151	0.310	0.414	0.387	0.365
B/B′	0.547	0.660	0.674	0.260	0.586	0.597	0.635
C	0.031	—	0.035	0.030	—	—	—
C′	0.031	0.090	—	0.010	—	—	—
D/I/K/L	0.047	—	0.116	0.250	—	—	—
G	—	—	0.023	0.140	—	—	—
M	—	—	—	—	—	0.016	—

[1]WDE = Weisse Deutsche Edelziege; SA = Saanen; AG = Angora goat, HG = Hair goat, BG = Borno goat; RS = Red Sokoto; WAD = West African Dwarf.

[2]SSCP = Single-strand conformation polymorphism.

SSCP patterns. Another marked difference in the Hair goat from Turkey was the predominance of *CSN3*A* instead of *CSN3*B/B′*. An effect of the higher sample number in this breed (n = 50) cannot be excluded, but such an effect is not very likely as all rare variants (*C*, *C′*, *G*, and *M*) were also detected in breeds with smaller sample sizes. The *CSN3*E* allele, until now reported only in the Montefalcone breed (Angiolillo et al., 2002), was not detected in our analysis, and this further stresses the possibility that the allele is breed-specific. The *CSN3*C′* frequency of 9% in Saanen is in accordance with the earlier report of Yahyaoui et al. (2001), who detected 11% *CSN3*C′* in the same breed and very low (1%) frequencies in 2 Spanish breeds. To our knowledge, this work is the first report of *CSN3* allele frequencies in goats from Turkey and Africa. High numbers of alleles were detected in the Hair goat and Angora goat from Anatolia, whereas the African goat breeds and Saanen goats showed rather low variability. The Near East is believed to be one of the domestication centers of cattle, and breeds of this region continue to show higher genetic variability than those found in Europe or Africa (Loftus et al., 1999; Luikart et al., 2001). Allelic variation detected by SSCP analysis at the *CSN3* locus indicates a similar situation for domesticated goats.

CONCLUSIONS

Using DNA sequence analysis, 2 miscellaneous SSCP patterns were assigned to new caprine *CSN3* allelic states. The number of alleles identified in the domesticated goat has increased to 16, of which 13 are protein variants and 3 are synonymous alleles detectable in DNA only. With the 2 nucleotide exchanges found in the *M* allele, the total number of polymorphic sites in *CSN3* exon 4 increased from 13 to 15. Several conflicts in current nomenclature became obvious and changes were, and might again become, necessary. At present,

no procedure for simultaneous genotyping of all known genetic variants in caprine *CSN3* is available. The SSCP analysis, although powerful in screening, reaches limitations in fine resolution of more than 10 alleles. Due to the lack of a sequenced reference DNA sample, it was not possible to describe the PCR-SSCP pattern of the *CSN3*F* allele. Depending on the diagnostic test used, groups of confounding alleles are detected and will need further verification with other tests. In summary, the goat *CSN3* exon 4 becomes a highly complex genetic "mosaic", with a greater number of possible combinations resulting from the different mutation sites available. If it is unlikely for *CSN3* to be completely free of selective constraints due to its important properties, one can wonder if the functional properties themselves may lead to such high biodiversity within the coding sequence of the gene in domesticated goats. Occurrence of at least 6 *CSN3* alleles in the Hair goat from Anatolia points to a high number of ancient alleles and excludes the possibility that recent selection in production breeds was a major reason for *CSN3* variation.

REFERENCES

Aleandri, R., L. G. Buttazzoni, 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.

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.

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

Bonsing, J., and A. G. Mackinlay. 1987. Recent studies on nucleotide sequences encoding the caseins. Dairy Res. 3:447–461.

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. Short communication: Simultaneous identification of five kappa-casein (CSN3) alleles in domesticated goat by polymerase chain reaction single strand conformation polymorphism (PCR-SSCP). J. Dairy Sci. 86:3726–3729.

Chianese, L., B. Portolano, and E. Troncone. F. Pizzolongo, P. Ferranti, F. Addeo, M. L. Alicata, F. Pilla, and G. Calcagna. 2000. The quality of Girgentana goat milk. Pages 946–949 in Proc. 7th International Conference on Goats, Tours, France. International Goat Association (IGA), Little Rock, AR.

Coll, A., J. M. Folch, and A. Sanchez. 1993. Nucleotide sequence of the goat kappa-casein cDNA. J. Anim. Sci. 71:2833.

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.

Falaki, M., A. Prandi, C. Corradini, M. Sneyers, N. Gengler, S. Massart, U. Fazzini, A. Burny, D. Portetelle, and R. Renaville. 1997. Relationships of growth hormone gene and milk protein polymorphisms to milk production traits in Simmental cattle. J. Dairy Res. 64:47–56.

Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166.

Formaggioni, P., A. Summer, M. Malacarne, and P. Mariani. 1999. Milk protein polymorphism: Detection and diffusion of the genetic variants in Bos Genus. Ann. Fac. Med. Vet. Univ. Parma 19:127–165.

Jann, O. C., E.-M. Prinzenberg, G. Luikart, A. Caroli, and G. Erhardt. 2004. High polymorphism in the kappa-casein (CSN3) gene from wild and domestic caprine species revealed by DNA sequencing. J. Dairy Res. 71:188–195.

Lodes, A., J. Buchberger, J. Aumann, and H. Klostermeyer. 1996. The influence of genetic variants of milk proteins on the compositional and technological properties of milk. 1. Casein micelle size

and the content of non-glycosylated kappa-casein. Milchwissenschaft 51:368–373.

Loftus, R. T., O. Ertugrul, A. H. Harba, M. A. A. El-Barody, D. E. MacHugh, S. D. E. Park, and D. G. Bradley. 1999. A microsatellite survey of cattle from a centre of origin: The Near East. Mol. Ecol. 8:2015–2022.

Luikart, G., L. Gielly, L. Excoffier, J.-D. Vigne, J. Bouvet, and P. Taberlet. 2001. From the Cover: Multiple maternal origins and weak phylogeographic structure in domestic goats. Proc. Natl. Acad. Sci. USA 98:5927–5932.

Malkoski, M., S. G. Dashper, N. M. O'Brien-Simpson, G. H. Talbo, M. Macris, K. J. Cross, and E. C. Reynolds. 2001. Kappacin, a novel antibacterial peptide from bovine milk. Antimicrob. Agents Chem. 45:2309–2315.

Mariani, P., G. Losi, V. Russo, G. B. Castagnetti, L. Grazia, D. Morini, and E. Fossa. 1976. Prove di caseificazione con latte caratterizzato dalle varianti A e B delle κ-caseina nella produzione del formaggio Parmigiano-Reggiano. Sci. Tecn. Latt.-Cas. 27:208–227.

Mercier, J. C., G. Brignon, and B. Ribadeau-Dumas. 1973. Primary structure of bovine kappa B casein. Complete sequence. Eur. J. Biochem. 35:222–235.

Mercier, J. C., J. M. Chobert, and F. Addeo. 1976. Comparative analysis of the amino acid sequences of caseinomacropeptides from seven species. FEBS Lett. 72:208–214.

Nicholas, K. B., and H. B. Nicholas, Jr. 1997. GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the author. www.psc.edu/biomed/genedoc.

Orita, M., Y. Suzuki, T. Sekiya, and K. Hayashi. 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874–879.

Prinzenberg, E.-M., I. Krause, and G. Erhardt. 1999. SSCP analysis at the bovine CSN3 locus discriminates six alleles corresponding to known protein variants (A, B, C, E, F, G) and three new DNA polymorphisms (H, I, AI). Anim. Biotechnol. 10:49–62.

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.

Ward, T. J., R. L. Honeycutt, and J. N. Derr. 1997. Nucleotide sequence evolution at the kappa-casein locus: evidence for positive selection within the family Bovidae. Genetics 147:1863–1872.

Xia, X., and Z. Xie. 2001. DAMBE: Software package for data analysis in molecular biology and evolution. J. Hered. 92:371–373.

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.