Journal of Dairy Research (2004) **71** 188–195. © Proprietors of Journal of Dairy Research 2004
DOI: 10.1017/S0022029904000093 Printed in the United Kingdom

High polymorphism in the κ-casein (*CSN3*) gene from wild and domestic caprine species revealed by DNA sequencing

Oliver C Jann[1], Eva-Maria Prinzenberg[1], Gordon Luikart[2], Anna Caroli[3] and Georg Erhardt[1]*

[1] Institut für Tierzucht und Haustiergenetik, Ludwigstr. 21b, Justus-Liebig Universität Gießen, 35390 Gießen, Germany
[2] Laboratoire de Biologie des Populations d'Altitude, Centre National de la Recherche Scientifique, UMR 5553, Université Joseph Fourier, B.P. 53, F-38041 Grenoble Cedex 9, France
[3] Dipartimento di Sanità e Benessere Animale, University of Bari, Strada Provinciale per Casamassima km 3, Valenzano, Bari, Italy

Received 25 March 2003 and accepted for publication 25 June 2003

We assessed polymorphisms in exon IV of the κ-casein gene (*CSN3*) in ten different breeds of domestic goat (*Capra hircus*) from three continents and in three related wild caprine taxa (*Capra <ibex> ibex*, *Capra <ibex> sibirica* and *Capra aegagrus*). Thirty-five DNA samples were sequenced within a 558 bp fragment of exon IV. Nine polymorphic sites were identified in domestic goat, including four new polymorphisms. In addition to four previously described polymorphic positions, a total of 13 polymorphisms allowed the identification of 13 DNA variants, corresponding to 10 protein variants. Because of conflicting nomenclature of these variants, we propose a standardized allele designation. *CSN3*A*, *CSN3*B*, and *CSN3*D* were found as widely distributed alleles in European goat breeds. Within *Capra ibex* we identified three variants and showed that the sequence of *Capra aegagrus* is identical to the most common *Capra hircus* variant, consistent with *Capra aegagrus* being the wild progenitor of domestic goats. A dendrogram was drawn to represent the molecular network between the caprine *CSN3* variants.

Keywords: *Capra*, κ-casein, polymorphism, nomenclature.

Caseins of domesticated milk-producing species are of growing economic interest because of their direct relationship with milk quality, composition and cheesemaking properties (Aleandri et al. 1990; Lodes et al. 1996; Falaki et al. 1997). Various studies suggest important relationships between casein polymorphisms and milk production traits (Bovenhuis & Weller, 1994; Ikonen et al. 1999; Velmala et al. 1999). κ-Casein plays an important role in the formation, stabilization, and aggregation of the casein micelles thus altering the manufacturing properties and digestibility of milk. Chymosin splits the κ-casein into an insoluble (para-κ-casein: amino acids 1–105) and a soluble glycopeptide (caseinomacropeptide: amino acids 106–171), a crucial process for the production of cheese and also for the nutrition of sucklings (Mercier et al. 1973). Caseinomacropeptide (CMP) fulfils important physiological functions such as increasing digestive efficiency (Mercier et al. 1976) and antibacterial activity (Malkoski et al. 2001).

Owing to these properties, polymorphisms in the *CSN3* gene might affect protein structure, which is strongly connected with production traits and biological fitness, suggesting an important role of selection in its molecular evolution (Ward et al. 1997). Moreover, expression of *CSN3* homologous mRNA has been identified also in non-mammalian species, confirming important physiological properties and an extremely high conservation of the gene throughout evolution (Ottaviani et al. 1999).

Until recently, *CSN3* was little studied and considered to be less polymorphic in goats than in cattle, in which several polymorphisms have been described (Miranda et al. 1993; Erhardt, 1996; Mitra et al. 1998; Prinzenberg et al. 1999). In an unspecified Italian goat breed, Di Luccia et al. (1990) described two κ-casein variants, which were named A and B. Recently, Caroli et al. (2001) and Yahyaoui et al. (2001) described genetic polymorphisms in goat *CSN3* exon IV, causing amino acid exchanges at protein positions 44, 65, 119, 156, and 159, besides a number of silent mutations. However, conflicting nomenclature derives from the recent *CSN3* molecular characterizations.

Molecular evolution of *CSN3* has been the subject of a few studies. Gatesy et al. (1996) compared *CSN3* exon IV of 21 species, and reported similar rates for synonymous

*For correspondence; e-mail: Georg.Erhardt@agrar.uni-giessen.de

and non-synonymous substitutions, and a substitution pattern affecting nucleotides independently of their position within the codon. In contrast, Ward et al. (1997) reported strong directional selection on the caseinomacropeptide between distantly related bovid *taxa*, leading to an accelerated divergence of this peptide within a 34-codon region. In closely related species this pattern was not observed. Consequently the authors predicted low polymorphism of κ-caseinomacropeptide within species, and more extensive polymorphism between species.

Six caprine wild species can be differentiated: *Capra aegagrus*, *Capra ibex*, *Capra caucasica*, *Capra cylindricornis*, *Capra pyrenaica*, and *Capra falconeri*. Four or more species of *Capra ibex* are recognized by some authors (Shackelton, 1997). Cross breeding between all *Capra* species is possible in captivity but is not documented in the wild (Mason, 1984). Takada et al. (1997) provided evidence that the Bezoar (*Capra aegagrus*) is likely to be the maternal ancestor species of domestic goat (*Capra hircus*). Recent studies (Luikart et al. 2001; Mannen et al. 2001) based on mtDNA analysis revealed a multiple maternal origin of domestic goat, probably from different lineages of *Capra aegagrus*. These lineages were apparently domesticated independently, and are widespread among breeds and regional populations, showing weak intercontinental structuring.

The primary aim of the present study was to investigate the occurrence of new, potentially valuable genetic variants at caprine *CSN3* gene as a necessary starting point for more detailed research on the molecular evolution of the gene. A much broader sampling was used than in previous reports. A standardized nomenclature for caprine *CSN3* variants is also proposed. Moreover, data are given on the distribution of some domestic goat variants, and a dendrogram is drawn to represent the molecular network between the caprine *CSN3* alleles.

Materials and Methods

Samples

Samples of wild and domestic species were collected in several geographically distant countries across the Old World. Blood from 30 domestic goats (*Capra hircus*) and frozen AI sperm from one ram were collected. DNA was extracted from leucocytes by standard protocols (Montgomery & Sise, 1990) and from sperm according to Lien et al. (1990) but using a 10-fold smaller scale. Four samples of three wild caprine species (*Capra <ibex> ibex*, *Capra <ibex> sibirica*, and *Capra aegagrus*) were obtained from hunter-killed animals provided by hunters and field biologists (Table 1). For the nomenclature of wild caprine species, we used those recognized by the IUCN (Shackelton, 1997; where '*<ibex>*' means species *v.* subspecies status controversial). To confirm Mendelian transmission, six informative families of Bunte Deutsche Edelziege were analysed.

Table 1. Characterization of the samples by species, breed, origin, and number of sequenced animals (*n*)

Species	Breed	Origin	*n*
Capra hircus	Undefined	Italy	3
Capra hircus	Red Sokoto	Nigeria	1
Capra hircus	Bunte Deutsche Edelziege	Germany	17
Capra hircus	Weisse Deutsche Edelziege	Germany	3
Capra hircus	Toggenburger	Germany	2
Capra hircus	African Dwarf Goat	Nigeria	1
Capra hircus	Angora	Turkey	1
Capra hircus	Hair Goat	Turkey	1
Capra hircus	Alpine	France	1
Capra hircus	Undefined	Malaysia	1
Capra ibex ibex	Wild	Italy	1
Capra ibex sibirica	Wild	Mongolia	2
Capra aegagrus	Wild	Russia	1

PCR conditions

A 558 bp fragment containing exon IV of the goat *κ-casein* gene was amplified by PCR, using primers selected from the corresponding region in cattle (5′-AGAAATAATACCATTCTGCAT-3′ and 5′-GTTGTCTTCTTTGATGTCTCCTTAGAG-3′; Prinzenberg et al. 1999). Each 50-µl reaction contained 1 µl DNA solution (10–100 ng DNA), 1 U *Pwo*-polymerase (Peqlab Biotechnologie, Erlangen Germany), 10 pmol of each primer, 80 µM each dNTP in complete buffer (containing 1·5 mM-$MgCl_2$) supplied with the polymerase. Cycling conditions in an iCycler (Bio-Rad Laboratories, Hercules, CA, USA) were as follows: 5 min at 93 °C, 35 cycles comprising 30 s at 93 °C, 40 s at 54 °C, 50 s at 72 °C, a final elongation at 72 °C for 2 min with subsequent cooling to 4 °C.

Cloning, sequencing, and sequence analysis of PCR fragments

PCR products of the wild caprine species were gel purified, ligated into a pCR®-Blunt vector and transformed into chemically competent *Escherichia coli* TOP10 cells (Invitrogen GmbH, Karlsruhe, Germany). Recombinants were identified by direct colony PCR and plasmid DNA for sequencing was prepared with a GFX Micro kit (Amersham Biosciences, Freiburg, Germany). Cycle sequencing was performed with a CycleReader™ Auto (MBI Fermentas, St. Leon-Rot Germany) sequencing kit and reactions were run on A.L.F. express using a 0·3-mm ReproGel (Amersham Biosciences). Alternatively, sequencing was performed by the Institute for Medical Microbiology, Justus-Liebig University, Giessen, using cycle sequencing and a MegaBace capillary sequencing unit (Amersham Biosciences). Consensus sequences were generated from a minimum of three clones per animal to eliminate polymerase errors during PCR and cycle sequencing. PCR products of domestic goats were directly sequenced on both strands by MWG Biotech (Ebersberg, Germany). For each allele, at least two PCR products were sequenced on both strands, to exclude

https://doi.org/10.1017/S0022029904000093 Published online by Cambridge University Press

polymerase errors. The nucleotide sequences and the deduced amino acid sequences were analysed with the DNASIS® for Windows®-Sequence Analysis Software (Hitachi Software Engineering Co., San Bruno, CA, USA).

Sequence comparison and molecular network

Variability of goat *CSN3* exon IV was investigated on the basis of the analysed domestic and wild goat *CSN3* sequences. In addition we compared the obtained results with the sequences of *Capra hircus* published by Coll et al. (1993) (GenBank accession no. X60763), Caroli et al. (2001) (GenBank accession no. AY027868), Yahyaoui et al. (2001) (accession nos AF485341 and AF485340), Angiolillo et al. (2002) (accession no. AF486523), and with the published sequence of *Capra pyrenaica* described by Yahyaoui et al. (2001) (GenBank accession no. AY090466). We included in the analysis two other *Capra hircus* sequences available in the GenBank (accession nos AY090465 and AY090467). The sequences of *Ovis aries* (Furet et al. 1990, GenBank accession no. X51822), *Capricornis crispus* (Chikuni et al. 1995, GenBank accession no. D14376) and *Bos taurus* (Alexander et al. 1988, GenBank accession no. X14908) were also compared with the caprine sequences.

Ovis aries was used as outgroup for drawing a molecular network. Relationships between the different sequences were analysed using the program TREECON (Van de Peer & Wachter, 1994). The distance matrix was calculated according to the Jukes and Cantor one parameter model with the bootstrap option. A total of 1000 bootstraps was done. The cluster analysis was performed by the Neighbour-Joining method.

Results

Nucleotide and deduced amino acid sequences

Tables 2a and b show, respectively, differences between nucleotide and deduced amino acid sequences in domestic and wild goats, both analysed in the present paper and obtained from GenBank. In addition, *Ovis aries, Capricornis crispus,* and *Bos taurus* sequences are reported, referring only to the polymorphisms shared with the other considered sequences. A total of 25 polymorphic nucleotide positions were thus compared. New sequences identified here include seven from domestic goats (GenBank accession numbers AY166705–AY166711, AF521022), one from *Capra aegagrus* (AF521023), and three from *Capra ibex* (A_{ibex}=AF527806, B_{ibex}=AF525023, C_{ibex}=AF527805).

Out of 31 domestic goat samples, we identified nine polymorphic sites corresponding to amino acid positions 18, 43, 44, 53, 58, 61, 65, 119, 159 (Table 2b). Four additional polymorphic positions (56, 90, 131, 156) described by other authors were absent in our sampling. Thus a total of 13 polymorphisms have been identified in domestic goat *CSN3* until now (Table 2a). Five of the 13 substitutions are silent mutations without effect on the primary structure of the protein (18, 43, 56, 58, 131), as shown in Table 2b. Four of them are located in the para-κ-casein (18, 43, 56, 58) and just one (131) in the caseinomacropeptide. Eight exchanges alter the amino acid sequence (positions 44, 53, 61, 65, 90, 119, 156, 159). Five of them are located in the para-κ-casein (44, 53, 61, 65, 90), and three in the caseinomacropeptide (119, 156, 159).

New nomenclature for domestic goat CSN3 *variants*

Because of conflicting nomenclature, it was not possible to maintain the already established allele designations. According to the guidelines of COGNOSAG (Broad et al. 1999), we introduced a standardized nomenclature of the recent allelic variants at caprine *CSN3* (Table 2a). In the old nomenclature, GenBank accession no. AY027868 and GenBank accession no. AY090465 were contradictorily assigned as *CSN3*B* and *CSN3*D* respectively, although representing identical sequences. Owing to the more recent GenBank publishing date, we renamed AY090465 as *CSN3*B*. GenBank accession no. AF485340, previously also named *CSN3*B*, was renamed as *CSN3*D*. We excluded AY090466 from the domestic goat nomenclature, because this allele was observed only in the wild *Capra pyrenaica*, but not in *Capra hircus* (Yahyaoui et al. 2001). The other alleles were renamed in chronological order of the GenBank publishing date, with the exception of *D′* and *D″*, only differing at the DNA level (by synonymous substitution) from *CSN3*D*.

According to the new nomenclature presented here, 13 DNA variants (*A, B, C, D, D′, D″, E, F, G, H, I, J, K*) and 10 protein variants (*A, B, C, D, E, F, H, I, J, K*) have been identified thus far in domestic goat *CSN3* exon 4. Out of the mentioned variants, Mendelian transmission within families was demonstrated for *CSN3*A, B, D,* and *J*.

Wild caprine species sequences

Compared with *Capra hircus CSN3*A* (Coll et al. 1993), 12 polymorphic nucleotide positions (214, 226, 245, 360, 396, 399, 405, 471, 537, 546, 591, 607), were identified within wild caprine *CSN3* exon IV (Table 2a). One of these substitutions is silent (245), whereas 11 nucleotide exchanges alter the amino acid sequence (codons 33, 37, 82, 94, 95, 97, 119, 141, 144, 159, 164; Table 2b). The sequence of *Capra aegagrus* is identical to *CSN3*D* of *Capra hircus*. Three nucleotide substitutions occur in *Capra pyrenaica*. In *Capra ibex*, three extremely divergent variants (A_{ibex}, B_{ibex}, and C_{ibex}) were detected that display one substitution in A_{ibex}, five in B_{ibex}, and nine in C_{ibex} in comparison with *Capra hircus CSN3*A*. Only three of the 12 polymorphic nucleotide positions detected in the wild species were also found in the domestic goat (nucleotide positions 245, 471, 591).

https://doi.org/10.1017/S0022029904000093 Published online by Cambridge University Press

Table 2a. *CSN3* nucleotide differences among domestic goats, wild goats, *Ovis aries*, *Capricornis crispus* and *Bos taurus*. Row and column progressive numbers allow an easier merging with Table 2b. Nucleotide positions are compared with GenBank accession no. X60763. Synonymous mutations are in italics. A grey background shows the new sequences detected in the present paper

	Column number		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
		Nomenclature																									
Row	GenBank		Nucleotide position (para-κ-casein peptide)																	Nucleotide position (caseinomacropeptide)							
no.	acc. no.	New	Old	163	170	214	226	245	247	274	284	290	298	302	309	360	385	396	399	405	471	477	509	537	546	583	591	607
1	X60763	*A*	*A*	G	C	G	G	T	A	A	G	C	A	A	G	A	A	A	C	C	G	G	A	A	A	C	T	C
2	AY027868	*B*	*B*					*C*	G						A						A						C	
3	AY090465	*B*	*D*					*C*	G						A						A						C	
4	AF485341	*C*	*C*					*C*			*A*				A						A		*G*			T	C	
5	AY166705	*D*																			A							
6	AF485340	*D*	*B*																		A							
7	AY166706	*D′*			*T*																A							
8	AY166707	*D″*										*T*									A							
9	AF486523	*E*	*E*														G				A							
10	AY166708	*F*													A						A						C	
11	AY090467	*G*	*G*					*C*							A						A						C	
12	AF521022	*H*								G											A							
13	AY166710	*I*													A						A							
14	AY166711	*J*											G								A							
15	AY166709	*K*							G												A							
16	AF521023	*C. aegagrus*																			A							
17	AY090466	*C. pyrenaica*						*C*													A						C	
18	AF527806	A_{ibex}																										G
19	AF525023	B_{ibex}				A	C	*C*													A						C	
20	AF527805	C_{ibex}						*C*								G		G	A	A	A			G	G		C	
21	X51822	*O. aries*						*C*						G		G		G	A		A			G	G		C	
22	D14376	*C. crispus*	C					*C*											A		A	A				T	C	
23	X14908	*B. taurus*						*C*											A		A	A		G		A	C	T

Table 2b. *CSN3* amino acid differences among domestic goats, wild goats, *Ovis aries*, *Capricornis crispus* and *Bos taurus*. Row and column progressive numbers allow an easier merging with Table 2a. Silent mutations are in Italics. The amino acid positions, deduced on the basis of nucleotide sequences (Table 2a), are compared with GenBank accession no. X60763. A grey background shows the new sequences detected in the present paper

	Column number			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
		Nomenclature		Amino acid position (para-κ-casein peptide)																	Amino acid position (caseinomacropeptide)							
Row no.	GenBank acc. no.	New	Old	16	18	33	37	43	44	53	56	58	61	62	65	82	90	94	95	97	119	121	131	141	144	156	159	164
1	X60763	*A*	*A*	Arg	Phe	Ser	Ser	Tyr	Gln	Asn	Leu	Tyr	Tyr	Ala	Val	Thr	Asp	Thr	Leu	Arg	Val	Ala	Thr	Ile	Thr	Ala	Ser	Ala
2	AY027868	*B*	*B*					*Tyr*	Arg						Ile						Ile						Pro	
3	AY090465	*B*	*D*					*Tyr*	Arg						Ile						Ile						Pro	
4	AF485341	*C*	*C*					*Tyr*			*Leu*				Ile						Ile		*Thr*			Val	Pro	
5	AY166705	*D*																			Ile							
6	AF485340	*D*	*B*																		Ile							
7	AY166706	*D′*			*Phe*																Ile							
8	AY167707	*D″*										*Tyr*									Ile							
9	AF486523	*E*	*E*														Gly				Ile							
10	AY166708	*F*													Ile						Ile						Pro	
11	AY090467	*G*	*G*					*Tyr*							Ile						Ile						Pro	
12	AF521022	*H*								Ser											Ile							
13	AY166710	*I*													Ile						Ile							
14	AY166711	*J*											Cys								Ile							
15	AY166709	*K*							Arg												Ile							
16	AF521023	*C. aegagrus*																			Ile							
17	AY090466	*C. pyrenaica*						*Tyr*													Ile						Pro	
18	AF527806	A_{ibex}																										Gly
19	AF525023	B_{ibex}				Asn	Thr	*Tyr*													Ile						Pro	
20	AF527805	C_{ibex}						*Tyr*								Ala		Ala	Met	Ser	Ile			Val	Ala		Pro	
21	X51822	*O. aries*						*Tyr*						Tyr		Ala		Ala	Met		Ile			Val	Ala		Pro	
22	D14376	*C. crispus*		Thr				*Tyr*											Met		Ile	Thr				Val	Pro	
23	X14908	*B. taurus*						*Tyr*											Met		Ile	Thr		Val		Glu	Pro	Val

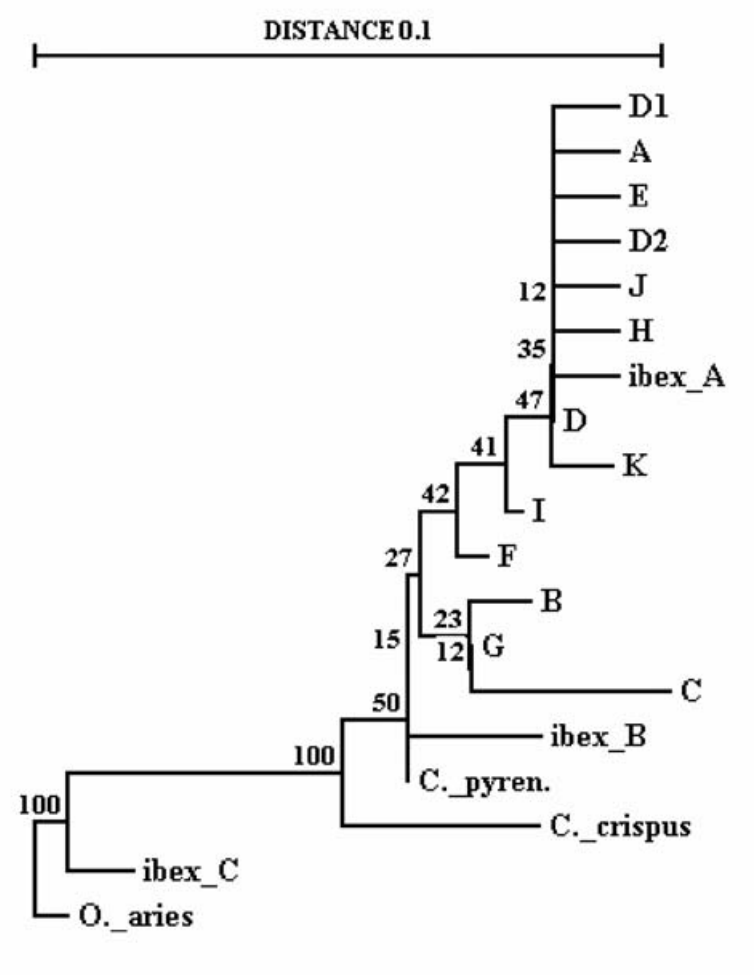

Fig. 1. Dendrogram of the caprine *CSN3* sequences. *Ovis aries* (O._aries) and *Capricornis crispus* (C._crispus) were included in the analysis. Bootstrap values are expressed as percentage of times each node appeared in 1000 bootstrapped samples.
D1 = *CSN3*D′*. D2 = *CSN3*D″*. D = *CSN3*D/Capra aegragus*. C._pyren. = *Capra pyrenaica*.

Distribution of polymorphisms within sequences and taxa

Most alleles share an identical CMP amino acid sequence. CMP of the variants *CSN3*D, D′, D″, E, H, I, J, K*, and *Capra aegagrus* differ at just one position from the CMP shared by *CSN3*B, F, G, Capra pyrenaica*, and *Capra ibex ibex* (B_{ibex}). Interestingly, the caseinomacropeptide of *Capra ibex sibirica* (C_{ibex}) is identical to that of *Ovis aries*.

Out of the 31 domestic goats analysed, 59 sequences (out of 62 chromosomes) could be exactly determined: *CSN3*A* (12), *CSN*B* (6), *CSN3*D* (29), *CSN3*D′* (1), *CSN3*D″* (1), *CSN3*F* (4), *CSN3*H* (1), *CSN3*I* (1), *CSN3*J* (3) and *CSN3*K* (1). *CSN3*A, CSN3*B*, and *CSN3*D* alleles were widely distributed among European goat breeds, and *CSN3*B* was absent in non-European samples. Both Turkish goat samples carried *CSN3*F*, one as a homozygous genotype. *CSN3*D′* and *CSN3*H* were found in one African dwarf goat and one Malaysian goat respectively. Out of three samples from *Capra ibex*, three alleles were identified. Two *Capra ibex sibirica* samples originating from Mongolia, carried *CSN3*A*$_{ibex}$ and *CSN3*C*$_{ibex}$. A further *Capra ibex ibex*, sampled in Italy, carried *CSN3*B*$_{ibex}$.

Molecular network of goat CSN3 *variants*

The dendrogram (Fig. 1) based on the *CSN3* exon IV sequences shows high bootstrap values (100%) for the first two nodes, allowing the separation of *Ovis aries*, C_{ibex}, and *Capricornis crispus*. *Capra pyrenaica and* B_{ibex} are successively separated with a (not significant) bootstrap value of 50%. Lower bootstrap values were found for the other nodes. A group of domestic goat variants is located in an intermediate position of the dendrogram: *CSN3*G, CSN3*B*, and *CSN3*C*. A third group of sequences, introduced by the variants *CSN3*F, CSN3*I* and *CSN3*K*, includes the other domestic goat variants, stemming from the *CSN3*D/Capra aegragus* sequence. In this group, the third A_{ibex} sequence also is located.

https://doi.org/10.1017/S0022029904000093 Published online by Cambridge University Press

Discussion

Sequence analysis uncovered seven new DNA and four new protein-coding variants in the domestic goat *CSN3* not described previously (Coll et al. 1993; Caroli et al. 2001; Yahyaoui et al. 2001) or available in GenBank (AY090467). Another *CSN3* allele was recently deposited in GenBank (accession no. AF434988), with a sequence corresponding to the *D* variant of the new nomenclature.

The *CSN3*D* allele, which also occurs in *Capra aegagrus*, was the most widely distributed *CSN3* variant in our sampling, observed in 22 out of the 31 domestic goat samples analysed. In contrast, Caroli et al. (2001) describe *CSN3*A* as the most widely distributed allele. However, the allele frequencies were determined at protein level using isoelectric focusing (IEF), which does not allow the separation of most variants (*CSN3*C*, *D*, *D′*, *D″*, *F*, *G*, *H*, and *I*) from *CSN3*A*. *CSN3*A*, *CSN3*B*, and *CSN3*D* were widely distributed alleles in our sampled European goat breeds.

CSN3 is possibly the most polymorphic among the four goat caseins, which have been widely studied (for reviews see Martin, 1993; Grosclaude et al. 1994; Martin et al. 1999; Rando et al. 2000). In cattle, *CSN3* is also considered to be a high polymorphic locus within the casein cluster (Prinzenberg et al. 1999). Among the newly sequenced goat variants, the silent mutations *CSN3*D′* and *D″* as well as the protein variants *CSN3*K* and *I* were detected in only one animal each, whereas no further family samples were available. Therefore the transmission of these alleles could not be demonstrated. Nevertheless, our repeated sequencing is reasonable proof that the alleles exist, although the detection of the mentioned variants in just one animal suggests a limited distribution of these mutations.

Most domestic goat variants (*D*, *D′*, *D″*, *E*, *H*, *I*, *J*, and *K*) share the same CMP, which is different from the variant *A* at amino acid position 119 (Table 2b). Other variants (*B*, *F*, and *G*) share the codon for Ile in position 119, but differ at amino acid position 159 from that group and variant A (Table 2b). Referring again to the domestic goat, we found four polymorphic nucleotide sites within the CMP-coding region and nine within the para-κ-casein coding region. Based on relative sequence proportions (66 amino acids within CMP, 105 amino acids within para-κ-casein, with only 90 codons included in our study), mutations in CMP seem slightly underrepresented (4/66 = 6% *v.* 9/90 = 10%), but a valid comparison would require sequence analyses of the whole coding region as well as a more extensive phylogenetic study. Anyway, we can affirm that different *CSN3* variants share the same CMP. This fact should be carefully evaluated in further analyses of the physiological properties of the different variants, owing to the already mentioned important functions fulfilled by CMP (Mercier et al. 1976; Malkoski et al. 2001).

Comparison with *Ovis aries* and *Capricornis crispus* sequences in the dendrogram suggests that *Capra ibex sibirica* (C_{ibex}) is more ancestral and more closely related to *Ovis aries*. This agrees with mtDNA phylogenetic data (Manceau et al. 1999). *Capra ibex ibex* (B_{ibex}) and *Capra pyrenaica* appear to be phylogenetically older than the observed domestic goat variants. A group of domestic goat *CSN3* variants (*G*, *B*, *C*) are closer to *Ovis aries* and *Capricornis crispus* than the other variants, most of them stemming from *CSN3*D*/*Capra aegragrus* sequence.

The low bootstrap values for the nodes separating most caprine sequences, are probably due to the limited differences among the sequences, which are not surprising given the closely related taxa and the relatively low polymorphism typical of short coding sequences. Low bootstraps might also be due to some retromutation or recombination events, which should be taken into account when considering the molecular evolution of the different variants.

The new wild goat sequences yielded interesting results. The sequence of *Capra aegagrus* was identical to *CSN3*D* of *Capra hircus*, probably owing to genetic exchange or shared ancestry with domestic goat. Another striking result was the occurrence of three extremely divergent alleles within the *Capra ibex sibirica* taxon from Mongolia. Among them, $CSN3*A_{ibex}$, showing one substitution compared with *CSN3*A*, was the only sequence presenting the same nucleotide as *CSN3*A* at position 471.

For future research, DNA-based tests are needed to allow the complete *CSN3* typing for the screening of goat breeds. Among them, PCR-SSCP has been used already with good results allowing the simultaneous detection of the more common goat alleles besides the identification of further variants. Other tests, such as minisequencing, could also be successfully employed to give a more exhaustive picture of *CSN3* polymorphism as well as of casein cluster variability in goats. Another useful application might be to use goat *CSN3* polymorphism for breed characterization and traceability in products (e.g., cheese and milk).

Until now, goat breeding programmes have concentrated on *CSN1S1* variants because of the known effects on milk composition and technological qualities. In future, it will be necessary to take into account the entire casein cluster, with particular regard to *CSN3* owing to the high genetic polymorphism detected. This variability, resulting in several single nucleotide polymorphisms (SNPs), should be carefully investigated from different points of view. The first one, for which this paper is a fundamental basis, is to perform an accurate molecular phylogenetic analysis aimed at a better understanding of goat phylogenesis, and to highlight the physiological and functional properties of the different variants. Moreover, the functional properties should be evaluated by analysing the relationships between *CSN3* variants and quantitative traits. Such studies are completely lacking in goats. This evaluation should include alleles from wild *Capra* because wild taxa are proven genetic resources for improvement of domestic breeds (e.g., Yerxat, 1995).

We thank K Al Samaraie, Betriebsgemeinschaft Kirchhof, EM Ibeagha, C Özbeyaz, Y Zagdsuren, Suchthilfe Hof Fleckenbühl, R Unger, P Weinberg, Y Zagdsuren, S Bolormaa and CAPRI-IA for providing samples, and NATO and the VIGONI programme for financial support.

References

Aleandri R, Buttazoni LG, Schneider JC, Caroli A & Davoli R 1990 The effects of milk protein polymorphisms on milk components and cheese-producing ability. *Journal of Dairy Science* **73** 241–255

Alexander LJ, Stewart AF, Mackinlay AG, Kapelinskaya TV, Tkach TM & Gorodetsky SI 1988 Isolation and characterization of the bovine kappa-casein gene. *European Journal of Biochemistry* **178** 395–401

Angiolillo A, Yahyaoui MH, Sanchez A, Pilla F & Folch JM 2002 Characterization of a new genetic variant in the caprine κ-casein gene. *Journal of Dairy Science* **85** 2679–2680

Bovenhuis H & Weller JI 1994 Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. *Genetics* **137** 267–280

Broad T, Dolling CHS, Lauvergne J-J & Millar P 1999 Revised COGNOSAG guidelines for gene nomenclature in ruminants. *Genetics, Selection, Evolution* **31** 263–268

Caroli A, Jann O, Budelli E, Bolla P, Jäger S & Erhardt G 2001 Genetic polymorphism of goat kappa-casein (*CSN3*) in different breeds and characterization at DNA level. *Animal Genetics* **32** 226–230

Chikuni K, Mori Y, Tabata T, Saito M, Monma M & Kosugiyama M 1995 Molecular phylogeny based on the kappa-casein and cytochrome b sequences in the mammalian suborder *Ruminantia*. *Journal of Molecular Evolution* **41** 859–866

Coll A, Folch JM & Sanchez A 1993 Nucleotide sequence of the goat κ-casein cDNA. *Journal of Animal Science* **71** 2833

Di Luccia A, Mauriello R, Chianese L, Moio L & Addeo F 1990 κ-Casein polymorphism in caprine milk. *Scienza e Tecnica Lattiero-Casearia* **41** 305–314

Erhardt G 1996 Detection of a new kappa-casein variant in milk of Pinzgauer cattle. *Animal Genetics* **27** 105–107

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

Furet JP, Mercier JC, Soulier S, Gaye P, Hue-Delahaie D & Vilotte JL 1990 Nucleotide sequence of ovine kappa-casein cDNA. *Nucleic Acid Research* **18** 5286

Gatesy J, Hayashi C, Cronin MA & Arctander P 1996 Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Molecular Biology and Evolution* **13** 954–963

Grosclaude F, Ricordeau G, Martin P, Remeuf F, Vassal L & Bouillon J 1994 [From gene to cheese: the caprine alpha-s1 casein polymorphism, its effects and its evolution]. *INRA Productions Animales* **7** 3–19

Ikonen T, Ojala M & Ruottinen O 1999 Associations between milk protein polymorphism and first lactation milk production traits in Finnish Ayrshire cows. *Journal of Dairy Science* **82** 1026–1033

Lien S, Rogne S, Brovold MJ & Aleström P 1990 A method for isolation of DNA from frozen (A.I.) bulls semen. *Journal of Animal Breeding and Genetics* **107** 74

Lodes A, Buchberger J, Aumann J & Klostermeyer H 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

Luikart G, Gielly L, Excoffier L, Vigne JD, Bouvet J & Taberlet P 2001 Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of USA* **98** 5927–5932

Malkoski M, Dashper SG, O'Brien-Simpson NM, Talbo GH, Macris M, Cross KJ & Reynolds EC 2001 Kappacin, a novel antibacterial peptide from bovine milk. *Antimicrobial Agents and Chemotherapy* **45** 2309–2315

Mannen H, Nagata Y & Tsuji S 2001 Mitochondrial DNA reveal that domestic goat (*Capra hircus*) are genetically affected by two subspecies of bezoar (*Capra aegagurus*). *Biochemical Genetics* **39** 145–154

Martin P 1993 [Genetic polymorphism of caprine milk proteins]. *Lait* **73** 511–532

Martin P, Ollivier-Bousquet M & Grosclaude F 1999 Genetic polymorphism of caseins: a tool to investigate casein micelle organization. *International Dairy Journal* **9** 163–171

Mason IL 1984 Goat. In *Evolution of Domesticated Animals*, pp. 85–99 (Ed. IL Mason). New York: Longman Inc

Manceau V, Boursot P, Després L & Taberlet P 1999 Systematics of the genus *Capra* inferred from mitochondrial DNA sequences. *Molecular Phylogenetic Evolution* **13** 504–510

Mercier JC, Brignon G & Ribadeau-Dumas B 1973 Primary structure of bovine kappa B casein. Complete sequence. *European Journal of Biochemistry* **35** 222–235

Mercier JC, Chobert JM & Addeo F 1976 Comparative analysis of the amino acid sequences of caseinomacropeptides from seven species. *FEBS Letters* **72** 208–214

Miranda G, Anglade P, Mahé MF & Erhardt G 1993 Biochemical characterization of the bovine genetic kappa-casein C and E variants. *Animal Genetics* **24** 27–31

Mitra A, Krause I, Blusch J, Werner T, Balakrishnan CR & Pirchner F 1998 Kappa-casein polymorphisms in Indian dairy cattle and buffalo: a new genetic variant in buffalo. *Animal Biotechnology* **9** 81–87

Montgomery GW & Sise JA 1990 Extraction of DNA from sheep white blood cells. *New Zealand Journal of Agricultural Research* **33** 437–441

Ottaviani E, Franchini A, Prinzenberg EM, Erhardt G & Jolles P 1999 Detection of casein fragments in an invertebrate and in a vertebrate using *in situ* hybridization. *Life Science* **65** 1707–7114

Prinzenberg EM, Krause I & Erhardt G 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, A_1). *Animal Biotechnology* **10** 49–62

Rando A, Ramunno L & Masina P 2000 Mutations in casein genes. *Zootecnica e Nutrizione Animale* **26** 105–114

Shackleton DM (Ed.) **and the IUCN/SSC Caprinae Specialist Group 1997** *Wild sheep and goats and their relatives. Status survey and conservation action plan for caprinae*. Gland, Switzerland and Cambridge, UK: International Union for the Conservation of Nature and Natural Resources

Takada T, Kikkawa Y, Yonekawa H, Kawakami S & Amano T 1997 Bezoar (*Capra aegagrus*) is a matriarchal candidate for ancestor of domestic goat (*Capra hircus*): evidence from the mitochondrial DNA diversity. *Biochemical Genetics* **35** 315–326

Van de Peer Y & De Wachter R 1994 TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer Applications in the Biosciences* **10** 569–570

Velmala RJ, Vilkki HJ, Elo KT, de Koning DJ & Maki-Tanila AV 1999 A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. *Animal Genetics* **30** 136–143

Ward TJ, Honeycutt RL & Derr JN 1997 Nucleotide sequence evolution at the kappa-casein locus: evidence for positive selection within the family *Bovidae*. *Genetics* **147** 1863–1872

Yahyaoui MH, Coll A, Sanchez A & Folch JM 2001 Genetic polymorphism of the caprine kappa-casein gene. *Journal of Dairy Research* **68** 209–216

Yerxat MY 1995 Application of wild goats in cashmere breeding. *Small Ruminant Research* **15** 287–291