



Molecular characterization and genetic variability at κ -casein gene (CSN3) in camels

A. Pauciuolo ^{a,*}, E.S. Shuiep ^b, G. Cosenza ^c, L. Ramunno ^c, G. Erhardt ^a

^a Institute for Animal Breeding and Genetics, Justus Liebig University, Ludwigstraße 21 B, 35390 Gießen, Germany

^b Department of Animal Production, Faculty of Veterinary Science, University of Nyala, Nyala, South Darfur, Sudan

^c Department of Soil, Plant, Environment and Animal Production Science, University of Naples "Federico II", via Università, 100–80055 Portici (NA), Italy

ARTICLE INFO

Article history:

Accepted 25 October 2012

Available online 12 November 2012

Keywords:

Camelus dromedarius

Casein

CSN3

Promoter analysis

Interspersed elements

HNF-1

ABSTRACT

κ -casein is a glycosylated protein belonging to a family of phosphoproteins (α s1, β , α s2, κ) that represent the major protein component in mammalian milk. κ -casein plays an essential role in the casein micelle stabilization, determining the size and the specific function. In the present paper, we report for the first time the characterization of the nucleotide sequence of the whole κ -casein-encoding gene (CSN3) plus 1045 nucleotides at the 5' flanking region in *Camelus dromedarius*. The promoter region and the complete cDNA were also provided for the first time in *Camelus bactrianus*. The gene is spread over 9.3 kb and consists of 5 exons varying in length from 33 bp (exon 3) to 494 bp (exon 4), and 4 introns from 1200 bp (intron 3) to 2928 bp (intron 2). Highly conserved sequences, located in the 5' flanking region, have been found. The regulatory regions of camels seems to be more related to equids than to other compared species. 17 polymorphic sites have been detected, one of these (g.1029T>C) is responsible for the creation of a new putative consensus sequence for the transcription factor HNF-1. In general, these SNPs are the first reported in camels for casein *loci*. Finally, seven interspersed repeated elements were also identified at intronic level.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The camel stock is at present estimated to be about 23 million in the world. Somalia and Sudan together hold almost 50% of the whole camel population. In the last forty years (1970–2010), the number of animals has increased of almost 45% (www.faostat.fao.org), and such increase is even more evident in Sudan (+92.4%), where camels are mainly bred for milk production and more rarely for dairy purpose. The daily milk production is estimated to vary between 3 and 10 kg during a lactation period of 12–18 months (Farah et al., 2007), depending on breed, stage of lactation, feeding and management conditions, with an average content of 2.9% and 3.1% of protein and fat respectively (Al-haj and Al Kanhal, 2011).

Milk proteins and the corresponding coding genes have been deeply studied in ruminants, whereas such information is still limited in camels. Caseins (α s1, β , α s2 and κ) are coded by single autosomal genes (CSN1S1,

CSN2, CSN1S2 and CSN3, respectively) clustered in a DNA stretch of about 250 kb and mapped on chromosome 6 in cattle, sheep and goat (Rijnkels, 2002). They have been very well characterized both at DNA and protein level. Goats and cows represent the most polymorphic species, for which many alleles associated with different rates of protein synthesis have been identified (Caroli et al., 2006, 2009; Ramunno et al., 2005). In camels, genetic variants are known for α s1 casein (Kappeler et al., 1998) and recently a new protein variant has been identified for such protein fraction (Shuiep et al., 2012). Among caseins, κ -CN plays an essential role in the casein micelle stabilization (Alexander et al., 1988). It is located predominantly on micellar surface and it is the specific substrate of the chymosin, which is responsible for the hydrolyzation of the κ -CN into the para- κ -CN and the caseino-macropeptide (CMP). At least 16 alleles corresponding to 13 κ -CN variants have been identified in goat (Caroli et al., 2006), and at least 19 alleles corresponding to 14 κ -CN variants so far in cattle (Caroli et al., 2009).

Unlike what has been accomplished in these species, κ -casein gene in camels has not received much attention so far, whereas more information is available from a proteomic point of view. Farah and Farah-Riesen (1985) reported on the first characterization of major components of camel milk casein. Partial sequencing of tryptic digests of κ -casein showed a more stable conformation of the active cleavage site of camel chymosin compared to cow's κ -casein site (Kappeler et al., 1998). A quantitative analysis carried out on camel milk protein showed significantly lower amounts of camel κ -casein compared to the homologous cow's casein (Kappeler et al., 2003), and recently five different isoforms of κ -casein were found in camel milk as a result of a strong glycosylation of such protein (Hinz et al., 2012). To date no κ -casein variants have been

Abbreviations: CSN3, κ -casein-encoding gene; SNP, single nucleotide polymorphism; CN, casein; CMP, caseino-macropeptide; cDNA, DNA complementary to RNA; EDTA, ethylene diamine tetra acetic acid; SDS, sodium dodecyl sulfate; PCR, polymerase chain reaction; MgCl₂, magnesium chloride; dNTP, deoxyribonucleoside triphosphate; TBE, tris boric acid EDTA; C/EBP- α , CCAAT enhancer bind protein- α ; MGF, mammary gland factor; CRE-BP, cAMP response element binding protein; cAMP, cyclic adenosine 3',5'-monophosphate; Oct-1, Octamer bind protein; TBP, tata binding protein; GR, glucocorticoid receptor; PRL, prolactin; SP1, specificity protein 1; NF-1, nuclear factor-1; SRF, serum response factor; HNF-1, hepatocyte nuclear factor-1; HNF-3, hepatocyte nuclear factor-3; SINE, short interspersed element; LINE, long interspersed element; RFLP, restriction fragment length polymorphism.

* Corresponding author. Tel.: +49 6419937672; fax: +49 64137629.

E-mail address: alfredo.pauciuлло@agr.uni-giessen.de (A. Pauciuollo).

detected and the investigation at DNA level is limited to Somali camels' cDNA sequence (Kappeler et al., 1998) and to a comparison of the 5' flanking regions (Kappeler et al., 2003). Excluding these two studies no additional information is available at DNA level and no genetic variability is reported so far.

Keeping these reports in mind, the investigation was undertaken to explore genetic variability at camel κ -CN locus. We report on the first polymorphism detected at CSN3 gene in *Camelus dromedarius*, and we provide the full characterization and an extensive annotation of such gene. A comparison between *C. dromedarius* and *Camelus bactrianus* CSN3 promoter regions and cDNA was also accomplished.

2. Materials and methods

2.1. Animals

Blood samples were collected from 188 Sudanese she-camels (*C. dromedarius*, locally known as *Naga*) reared in five regions of the country (Fig. 1) and belonging to different ecotypes including Kahli, Lahaai and Arabi camels. Due to the lack of records, local experience was used to determine ecotypes in different locations and also individuals within the same herd. Typical phenotypic characteristics for each ecotype were strictly followed. Blood samples were immediately applied to classic filter paper (FTA®Classic Card-Whatman®BioScience, Maidstone, UK), allowed to dry at room temperature and stored until DNA isolation.

Additional blood samples were collected from three *C. bactrianus* belonging to Wilhelma Zoo (Stuttgart, Germany) and treated according to Spin Blood Mini Kit (Invitex, Germany).

2.2. DNA isolation

The filter paper containing blood samples were soaked (56 °C, overnight) in 500 μ l sodium-tris-EDTA buffer with 10 μ l proteinase K (10 mg/ml) in the presence of sodium dodecyl sulfate (SDS). Then DNA was isolated from the emerging lysis according to the procedure described by Sambrook et al. (1989). The isolated DNA was resuspended in 100 μ l TE buffer pH 7.6 (10 mM Tris, 1 mM EDTA).

DNA concentration and OD_{260/280} ratio of the samples were measured with the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific).

2.3. PCR conditions and sequencing

DNA regions spanning from nucleotide –1045 to +10436 of the camel CSN3 gene were amplified by means of a iCycler (Bio-Rad). A set of 33 primers for amplification and sequencing were designed by means of DNAsis-Max ver. 3.0 software (Hitachi), using as preliminary template the complete sequence of camel cDNA available in gene bank (EMBL ID Y10082) and then the new sequences determined in the course of the research. A typical PCR reaction mix (50 μ l) comprised: 100 ng of genomic DNA, 1X PCR Buffer (Promega), 2.5 mM MgCl₂, 5 pmol of each primer, dNTPs each at 200 μ M, 2.5 U of *Taq* DNA Polymerase (Promega). In Table 1 we report the starting amplicons for the sequencing of the whole gene. PCR was performed under the following conditions: 95 °C (4 min), 35 cycles at 95 °C (60 s), annealing temperatures depending on amplicon (Table 1) (45 s), 72 °C (90 s), final extension at 72 °C (10 min).

PCR products were purified and sequenced. The purification was carried out using MSB®Spin PCRapace kit (Invitex, Berlin, Germany). Sequence was accomplished by ABI 3130 Genetic Analyzer (Applied Biosystem). Targeted fragments were sequenced in both directions using BigDye chemistry (Applied Biosystems).

2.4. Genotyping by *Alu* I PCR-RFLP

PCR reaction mixture and thermal conditions for the amplification of the DNA fragment 488 bp long and spanning from –137 bp of 5'

flanking region to +351 bp of the camel CSN3 gene were accomplished by using the following primers: *forward* 5'-CACAAAGATGACTCTGCTATCG-3' and *reverse* 5'-GCCTCCACATATGCTG-3', according to standard PCR conditions already reported above. Product specificity was confirmed by ethidium-bromide-stained 1.5% agarose gel electrophoresis.

The entire panel of 188 animals was genotyped for the g.1029T>C SNP using a PCR-RFLP method. Digestion of 17 μ l of each PCR amplification was accomplished with 10 U of *Alu*I endonuclease (AG↓CT) (Fermentas) over-night at 37 °C. The digestion products were analysed directly by electrophoresis in 3.5% agarose gel in 1X TBE buffer and stained with ethidium bromide.

2.5. Bioinformatics

The allele frequency and Hardy–Weinberg equilibrium (χ^2 test) were calculated for the investigated population. SNP discovery, homology searches, comparison among sequences, and multiple alignments were accomplished using DNAsis-Max ver. 3.0 software (Hitachi Software, San Bruno, CA), whereas the putative transcription factor binding sites were searched by Transfact 7.0 software. Interspersed elements were found by RepeatMasker Web Server (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>). Estimates of evolutionary divergence between the sequences were conducted using the maximum composite likelihood method by using MEGA version 4.0 software. Phylogenetic tree was constructed using the Neighbor-Joining method by using the same software.

3. Results and discussion

The entire sequence of the *C. dromedarius* CSN3 gene plus the analysis of 5' flanking region are described. The promoter region of CSN3 gene and the complete cDNA is also provided for the first time in *C. bactrianus*. The analysis of 5' flanking region of the camel CSN3 gene provide an important contribution in order to evaluate the role and the importance of factors involved in the regulation of milk protein gene expression and the transcriptional effects of polymorphisms located in such regions. Furthermore the knowledge of the gene structure and the distribution of interspersed elements add new information in a species not investigated so far, as well as the genetic variability detected at camel CSN3 gene will provide new opportunities to start selection programmes also in such species.

3.1. Analysis of the gene structure

The whole gene encoding the camel κ -casein (CSN3) plus 1045 nucleotides at the 5' flanking region were sequenced (EMBL ID: HE863813). The gene extends over 9391 bp including 823 bp of exonic regions and 8568 bp of intronic regions with a total similarity with the corresponding bovine sequence (EMBL ID: AY380228) of about 58.3%. Approximately the same level of homology was also found by the following comparison to the κ -casein gene in other species: 58.0% vs river buffalo (EMBL ID: AM900443); 71.2% vs donkey (EMBL ID: FR922990) and 67.5% vs human (EMBL ID: U51899).

On the whole, the camel CSN3 gene shares a similar organization with the bovine counterpart, with some differences in intronic size. In fact, the analysis of the camel CSN3 gene evidenced an exon/intron size ratio higher (1:10.41) than that observed in cattle (1:14.42). The camel CSN3 gene is also characterized by high A/T content compared to G/C (69.6% vs 30.4%). However, this feature seems to be conserved among the species (Ward et al., 1997). It contains 5 exons, ranging in size from 33 bp (exon 3) to 494 bp (exon 4), and 4 introns from 1200 bp (intron 3) to 2928 bp (intron 2).

The first exon (63 bp) plus 8 bp of the exon 2 are not coding at all. The leader peptide (20 amino acids) is coded by the remaining part of the exon 2 (54 bp) plus the first two triplets of the exon 3. The mature peptide (162 amino acids) is encoded by the rest of exon 3, and almost

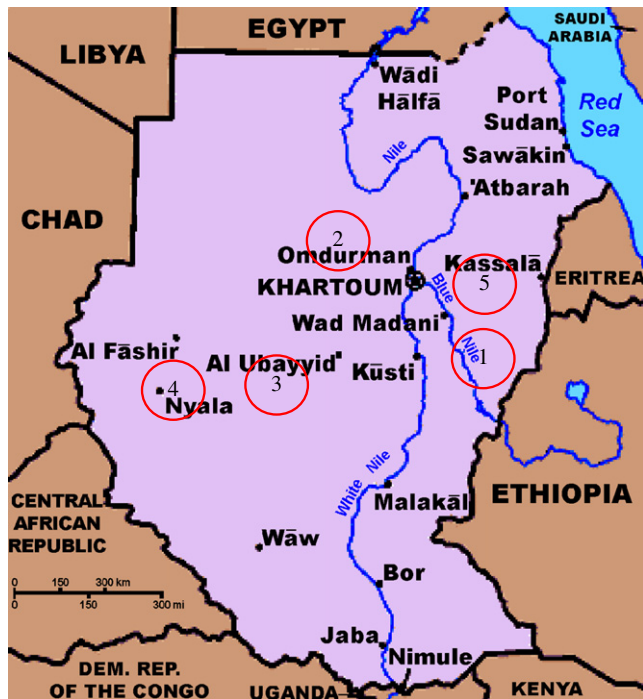


Fig. 1. Locations of sample collection from Sudan: 1. El Shuak (El Gadarif State); 2. West Omdurman (Khartoum State); 3. El Obeid (North Kordofan State); 4. Nyala (South Darfur State); 5. Tamboul (El Butana area).

the entire exon 4. The translation stop codon TAA is realized between the 460th–462th nucleotides of the exon 4, whereas the polyadenylation site is located between the 153rd and 158th nucleotide of the exon 5. All splice junctions follow the 5' GT/3' AG splice rule.

The translation initiation codon (AUG) and the nucleotide immediately upstream (Kozak consensus sequences: GCCGCCCAUGG) are known to play an important role in determining the translation rate and to contribute to the translation process (Kozak, 1991). Camel κ -casein shows 4 nucleotides homologous to Kozak sequence (positions: -6, -5, -3, -1, taking the A nucleotide of the AUG codon as number +1) as also reported in donkey (-6, -5, -3, -2, EMBL ID: FR922990), but differently from ruminants, pig, rabbit and rat (Cosenza et al., 2011) which have only 3 homologous nucleotides (positions: -6, -3, -2). The higher homology with Kozak sequence should guarantee a stronger consensus for camels and donkey compared to ruminants and therefore a higher efficiency of mRNA translation (Kozak, 1991). Further investigations need to confirm this statement, but it opens the way to new studies of expression analysis and translation efficiency in these species – so far – never investigated from this point of view.

A comparison of the sequences of the investigated samples shows only 7 nucleotide differences: 1 polymorphic site in the promoter region

and 6 at intronic level. No polymorphisms were found in the coding regions (Table 2).

The following comparison with the published sequences of the camel cDNA (Kappeler et al., 1998) and promoter region (Kappeler et al., 2003) shows a total of 10 polymorphic sites (Table 2): 5 insertions/deletions and 5 nucleotide substitutions, one of which results in an amino acid substitution at the second triplet of the exon 4 (TGT^{Cys} → TTT^{Phe}). The latter potential polymorphism was characterized by sequencing 40 randomly chosen animals, and it resulted monomorph in our samples.

One single nucleotide polymorphism (c.185C>T) was also detected at the exon 4 of *C. bactrianus* cDNA sequence (EMBL ID: HE863815). Although this mutation is synonymous, it represents the first example of polymorphism in *CSN3* coding region for such species.

To date many genetic variants of κ -casein have been identified at protein or DNA level in many species. Camels were not investigated so far, but the absence of polymorphisms in coding regions and the number of mutations detected in the rest of the *CSN3* gene suggests that the level of genetic variations of camel κ -casein is very low in comparison to other species (Carneiro and Ferrand, 2007; Caroli et al., 2006, 2009; Hobor et al., 2008). The homologous bovine gene, for instance, shows 42 polymorphic sites which differentiate only the allelic variants A and B (Robitaille et al., 2005). Therefore, a larger investigation of camel κ -casein should be performed to detect nucleotide variability, and further studies would be necessary in order to evaluate the potential influence on the quality-quantitative properties of camel milk. For instance, in cattle it is known that cows carrying the κ -casein BB genotype produce milk with a significantly higher protein content (Caroli et al., 2009). This led the dairy farmers to preferentially select these cows in order to have a higher cheese yield. Therefore, we cannot exclude the presence of quantitative alleles also in camels, which might lead similar advantages also for dairy camel breeders.

Furthermore, recently Weimann et al. (2009) showed that in cattle different *CSN3* variants are source of diverse angiotensin I converting enzyme (ACE) inhibitor peptides and demonstrated their potential role for human health. These bio-functional peptides were found also in camel milk (Al-haj and Al Kanhal, 2011). Therefore, it's likely that the genetic variants of the camel κ -casein might also influence its functional role and make the camel milk even more useful for the human nutrition.

3.2. Analysis of the *CSN3* gene promoter

In absence of polymorphisms in the coding regions, we decided to extend the sequencing to the 5'-flanking region and to analyze the putative regulatory regions of the camel κ -casein gene. About 1050 bp of the *CSN3* gene were amplified and sequenced for *C. dromedarius* (EMBL ID: HE863813) and *C. bactrianus* (EMBL ID: HE863814). A high level of homology and similar locations of most putative transcription binding sites were already reported among ovine, caprine, bovine, murine and rabbit κ -promoters (Gerencsér et al., 2002), horse, donkey and zebra *CSN3* 5' flanking region (Hobor et al., 2008). Here we performed a comparative analysis which included the aforementioned sequences in

Table 1
Primer sequences, annealing temperature and amplicon size used for the characterization of *Camelus dromedarius CSN3* gene. Asterisk refer to primer designed on camel *CSN3* promoter sequence (EMBL ID AJ409280), whereas the other primers were designed on cDNA sequence (EMBL ID Y10082).

Amplicon	Region covered	Primers	Sequence	Ta °C	Size (bp)
1	5' UTR/Ex 1	Cd <i>CSN3</i> 5'F* Cd <i>CSN3</i> Ex1R	5'-TGAAACTCTGCCATCTTTCTCT-3' 5'-TTCCCTTGTCCTGTCAG-3'	65.5	1213
2	Ex1/Intr1/Ex2	Cd <i>CSN3</i> Ex1F Cd <i>CSN3</i> Ex2R	5'-GGCAGCTCAACCTACT-3' 5'-AGTCACAAGTGAAGAAGCTC-3'	61.5	2610
3	Ex2/Intr2/Ex3	Cd <i>CSN3</i> Ex2F Cd <i>CSN3</i> Ex3R	5'-AGCTTCTTCCTAGTTGTGAC-3' 5'-GTTTGCACCTCTGCACC-3'	66.0	2993
4	Ex3/Intr3/Ex4	Cd <i>CSN3</i> Ex3F Cd <i>CSN3</i> Ex4R	5'-GGTGACAGGTGCAAAAC-3' 5'-CTGTGTTGCTCTTTG-3'	65.0	1726
5	Ex4/Intr4/Ex5	Cd <i>CSN3</i> Ex4F Cd <i>CSN3</i> Ex5R	5'-CAAATATTTCCCAATTCAGT-3' 5'-GCATCAATTAGCTTTAITA-3'	58.5	2535

Table 2

Polymorphisms detected by the comparison among *Camelus dromedarius* CSN3 gene sequence of the present study (A), with the CSN3 promoter sequence (B) reported by Kappeler et al. (2003) and the cDNA (C) described by Kappeler et al. (1998). Grey cells correspond only to mutations detected in our investigated samples (Y=A/C, W=A/T, M=A/C). Dashes indicate deleted nucleotide, asterisks show unavailable sequence.

Position	Nucleotide	A	B	C
promoter	335-336	-	G	*
	640	T	G	*
	702	T	-	*
	721	T	-	*
	1029	Y	T	*
	1039	T	-	*
Exon 1 (5'UTR)	1055-1056	TT	-G	*
Intron 2	3951	Y	*	*
	5019	Y	*	*
	5033	W	*	*
	5041	M	*	*
	5460	Y	*	*
Exon 4	7862	C ^{Cys}	*	T ^{Cys}
	7864	G ^{Cys}	*	T ^{Phe}
	8018	A ^{Ala}	*	C ^{Ala}
Intron 4	8647	Y	*	*

A: Present study; B: Kappeler et al. (2003); C: Kappeler et al. (1998).

addition to river buffalo (EMBL ID: AM900443), human (EMBL ID: U51899), pig (ENSEMBL ID: ENSSSCG0000009267) and the newly sequenced *C. dromedarius* and *C. bactrianus* κ -promoters. The level of homology differs between the compared sequences. As expected, sequences belonging to species of the same family are closely related each other (Table 3). Camels and equids showed a slightly lower level of divergence (on average 20.8%) compared with ruminants ($\leq 26.6\%$). Phylogenetic data confirmed that camels belong to their own branch (Fig. 2), but their κ -casein promoter sequences appear more related to equids (average homology >80%) than to ruminants, rabbit, mouse, human and pig which showed an average similarity significantly lower at about 60%. For some aspects these data are surprising considering that molecular analysis supported by fibrinogen, from which κ -casein encoding gene is supposed to be originated by a duplication gene event (Jollès et al., 1978), and mitochondrial genome data found *Camelidae* closely related to *Suina* (Geisler and Uhen, 2005). However, our findings are supported by the phenotypic data reported first by Farah and Farah-Riesen (1985) and recently by Hinz et al. (2012). They showed respectively the lack and the low concentration of such

Table 3

Estimates of evolutionary divergence between 13 taxa for κ -casein promoter sequence.

	EMBL ID	1	2	3	4	5	6	7	8	9	10	11	12	13
1. <i>C. dromedarius</i>	HE863813													
2. <i>C. bactrianus</i>	HE863814	0.004												
3. <i>E. asinus</i>	EU429803	0.213	0.211											
4. <i>E. zebra</i>	EU429802	0.202	0.201	0.003										
5. <i>E. caballus</i>	AY579426	0.212	0.210	0.010	0.006									
6. <i>B. taurus</i>	AY380228	0.266	0.266	0.335	0.330	0.339								
7. <i>B. bubalis</i>	AM900443	0.221	0.222	0.319	0.313	0.322	0.032							
8. <i>C. hircus</i>	Z33882	0.212	0.213	0.310	0.307	0.312	0.064	0.044						
9. <i>O. aries</i>	L31372	0.216	0.216	0.311	0.311	0.313	0.067	0.053	0.018					
10. <i>O. cuniculus</i>	AJ309572	0.406	0.416	0.416	0.412	0.418	0.541	0.515	0.544	0.548				
11. <i>M. musculus</i>	AJ309571	0.526	0.536	0.499	0.497	0.504	0.618	0.602	0.613	0.596	0.530			
12. <i>H. sapiens</i>	U51899	0.912	0.908	0.726	0.678	0.712	0.793	0.585	0.599	0.532	0.683	0.844		
13. <i>S. scrofa</i>	ENSSSCG0000009267	1.794	1.761	1.757	1.752	1.744	1.832	1.764	1.739	1.777	2.095	2.055	1.330	

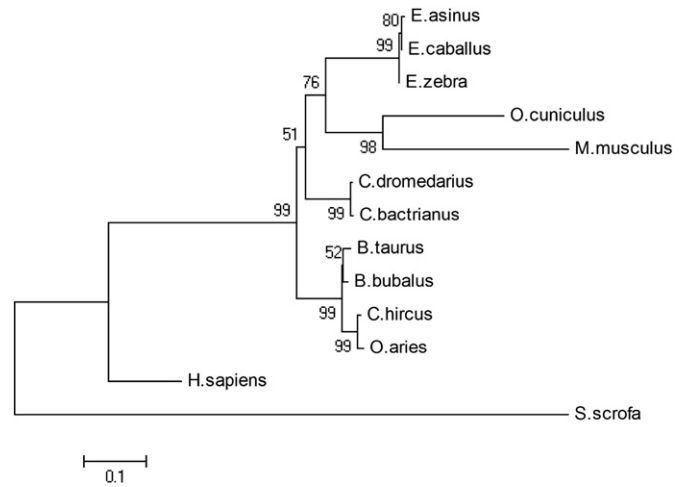


Fig. 2. Evolutionary relationships of 13 taxa for CSN3 gene promoter. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site.

milk protein fraction, as well as already observed in horses (Hinz et al., 2012). The close relationship with equids is also confirmed by cDNA comparison from which the level of similarity is >83%, whereas it is lower than 65% in the other cases.

Similar mechanisms of regulation are supposed for these related species, therefore camels κ -promoter sequences were compared with the homologous sequences of donkey promoter -not characterized so far- and bovine, respectively as non-ruminant and ruminant counterpart. 25 putative consensus sequences for transcription factor binding sites were found and reported in Fig. 3. Briefly, the analysis of the camel promoter showed a TATA-box located, with reference to the first nucleotide of the first exon, at nucleotides -32/-25 which is perfectly conserved among the compared sequences. Particularly interesting are also other 6 putative transcription factor binding sites almost perfectly conserved: one CCAAT enhancer bind protein (C/EBP)-alpha in position -1025/-1016, one Mammary Gland Factor/STAT5 (MGF/STAT5) at nucleotides -1018/-1008, one cAMP response element binding protein (CRE-BP) -900/-891, two Octamer bind protein (Oct-1) in position -307/-299 and -277/-268 respectively and one Tata Binding Protein (TBP) at nucleotides -236/-227.

The first two sites were reported to be very well conserved in more than 9 species (Gerencsér et al., 2002; Hobor et al., 2008; Malewski, 1998). Here we confirm them also in camels increasing the probability

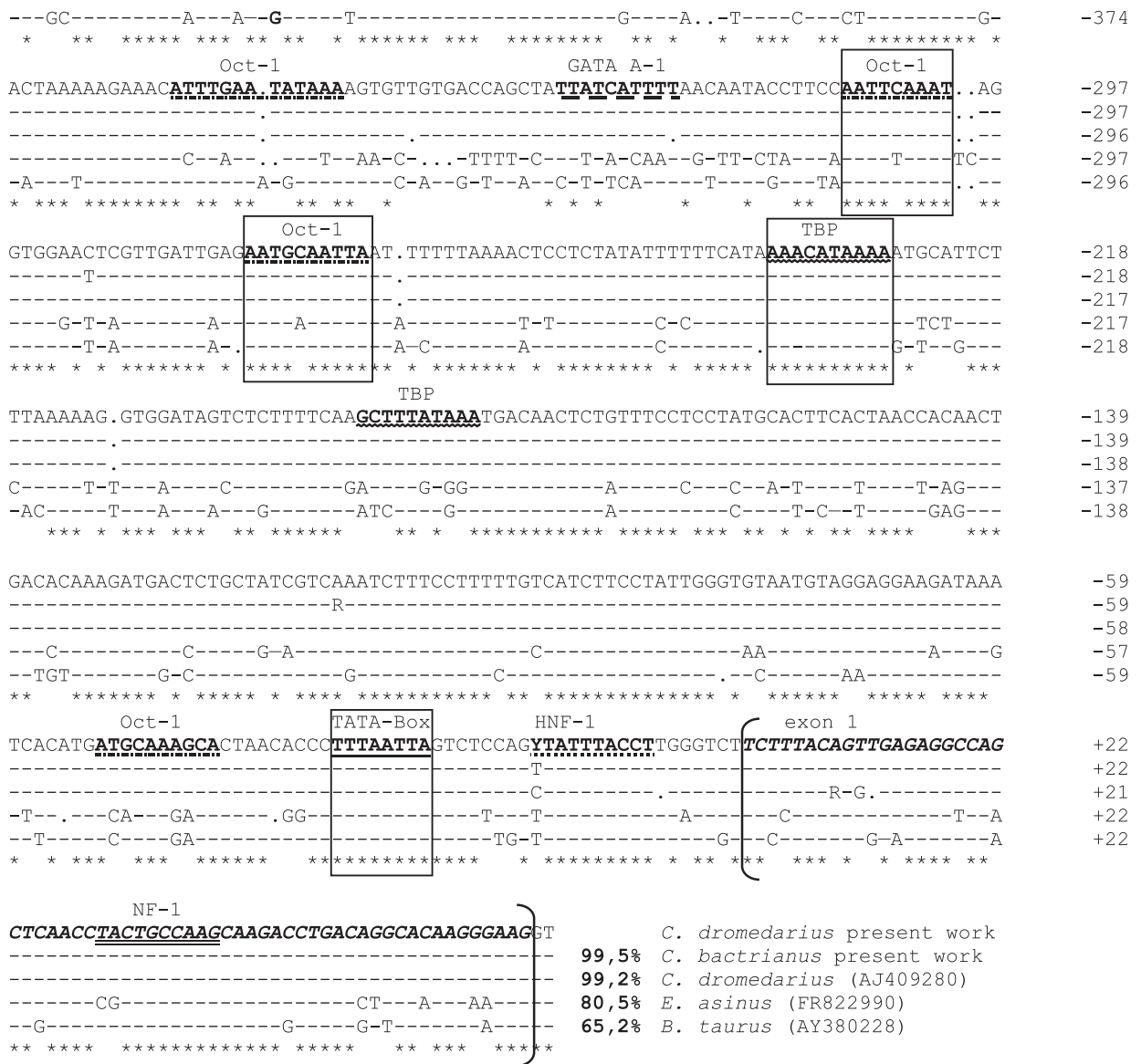


Fig. 3. Alignment and homology level (%) between the nucleotide sequences of the promoter region and exon 1 of *C. dromedarius* and *C. bactrianus* CSN3 gene with the published sequence of the camel (EMBL ID: AJ409280), donkey (EMBL ID: FR822990) and cattle (EMBL ID: AY380228) sequences. Numbering is relative to the first nucleotide of the first exon (+1). Dashes represent identical nucleotides to those in upper lines, whereas asterisk is a conserved nucleotide. Putative consensus sequence for transcription factors are differently underlined, whereas the boxes represent conserved putative binding sites.

that these factors are necessary for the regulation of the gene expression. The CRE binding protein (CRE-BP) is a transcriptional co-activator, which potentiates the activity of several groups of transcription. It interacts directly with STAT 5, reported as a crucial functional activator of casein gene expression (Vanselow et al., 2006); but it functions also as a transcriptional co-activator of the glucocorticoid receptor (GR) and prolactin (PRL) (Pfitzner et al., 1998). Oct-1 is not known to be a strong transcriptional activator by itself, but in conjunction with other co-activators. For instance, Zhao et al. (2002) showed that Oct-1 and STAT5 are both involved in the hormonal induction of casein gene expression. They bind DNA cooperatively and they interact in the stabilization of the transcription complex. However, Oct-1 can also interact with the TBP components (Zwilling et al., 1994). Oct-1 and TBP are putative sites very representative of camel κ -casein promoter being characterized by 7 Oct-1 and 3 TBP putative sites. Therefore, it would be reasonable to suggest that Oct-1 may also mediate stimulation of κ -casein expression in mammary gland through this way.

According to the comparative analysis of Gerencsér et al. (2002) on six different species, also in camels we found wide conserved regions (CB1, CB2 and CB4) apparently not characterized by known transcription factor binding sites. In fact, camel CB1 region is only partially covered by glucocorticoid receptor (GR) consensus sequence, whereas no consensus sites have been found for CB2 and CB4. Analogously, the wide region between nucleotides -180/-60 is completely lacking of known sites, whereas in bovine the same region is characterized at least by two SP1 sites, one NF-1 and one Ying Yang-1 factor, and the donkey sequence showed at least one SRF and one C/EBP- α .

Two polymorphic sites were detected in camel promoters. The first is a transition g.975A>G in position -112 which characterizes the *C. bactrianus* sequence and does not affect any putative regulatory region. The latter, at nucleotide -17, is a transition g.1029T>C realized in *C. dromedarius* promoter just upstream the exon 1. The presence of thymine is responsible for the creation of a extra putative consensus sequence for the hepatocyte nuclear factor-1 (HNF-1). This transcription

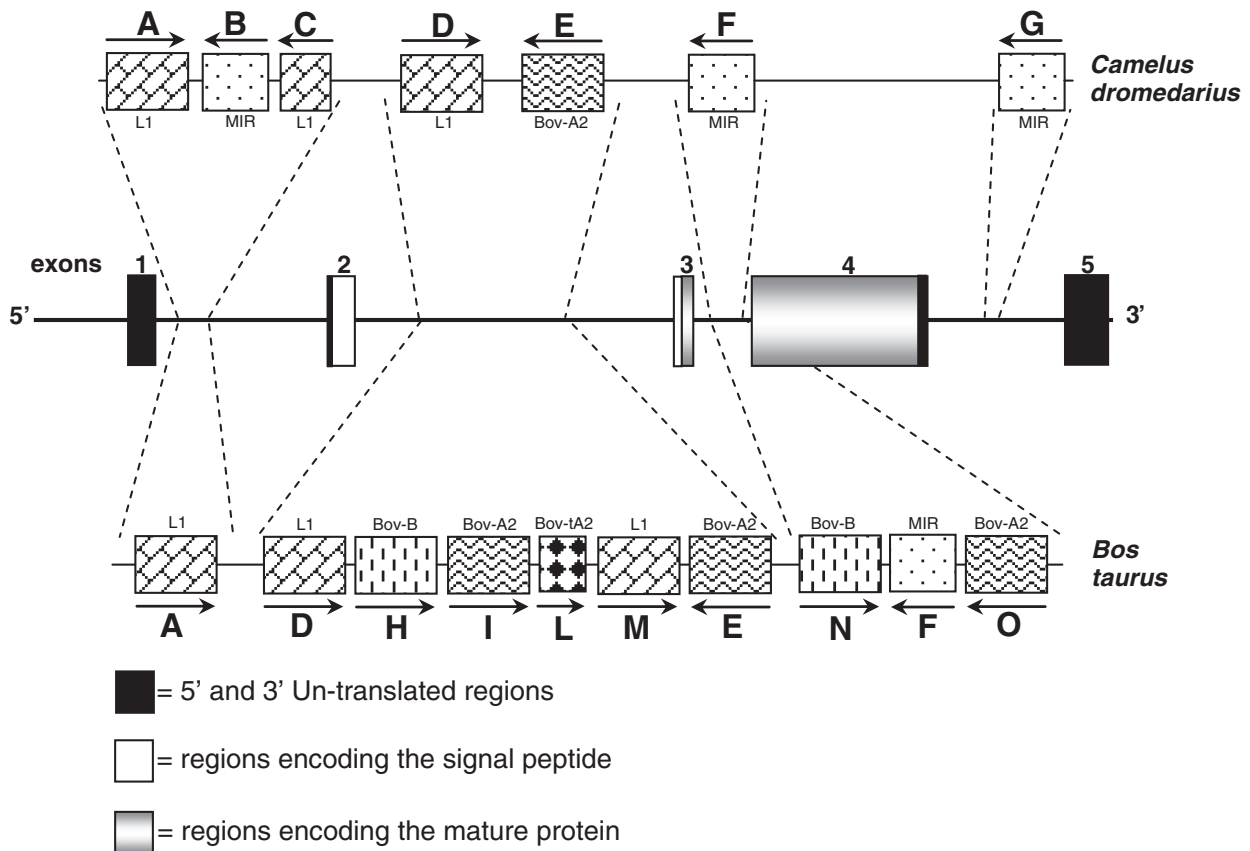


Fig. 4. Schematic representation of the *CSN3* gene and of the interspersed elements observed in camel and cattle. Common interspersed elements are indicated with the same letter.

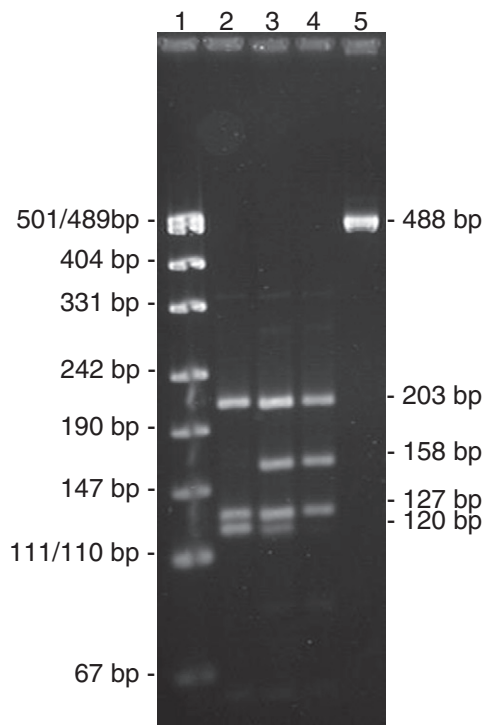


Fig. 5. Genotyping of *C. dromedarius* *CSN3* g.1029T>C SNP by Alu I (AG↓CT) PCR-RFLP. Line 2: CC samples; line 4: TT samples; line 3: CT samples; line 5: undigested sample. In the line 2 and 3 the band 38 bp long is not visible. Line 1 is pUC 19 DNA/MspI (HpaII) Marker, 23 (Fermentas).

factor is reported to regulate a number of genes involved in innate immunity, lipid and glucose transport, metabolism, etc. (Reiner et al., 2009), and to bind the promoters of hepatocyte specific genes including those of the fibrinogen chain (Mendel and Crabtree, 1991). However, conflicting are the information regarding the expression of HNF-1 in the mammary gland. Although Dunn et al. (2003) reported that no HNF-1 transcripts were detected in the human mammary gland, recently expression microarray analysis of pubertal mouse mammary gland showed the presence of HNF-1 transcripts (McBryan et al., 2007), whereas no additional information is reported in other species. However, another member of hepatocyte nuclear factors (HNF-3) is well known to be a potential regulator of casein genes expression (Schild and Geldermann, 1996; Schild et al., 1994), thus further investigation is required to verify also the influence of HNF-1 allelic variant in camel κ -casein gene regulation.

Table 4

Genotyping data and allele frequency of the g.1029T>C SNP at the *CSN3* promoter gene in Sudanese *Camelus dromedarius* population.

Breed	Observed Genotypes				Allelic frequency	
	CC	CT	TT	Total	C	T
Shanbali	9	17	12	38	0.46	0.54
Kahli	7	13	25	45	0.30	0.70
Arabi	4	10	12	26	0.35	0.65
Lahaoi	14	34	31	79	0.39	0.61
Total	34	74	80	188	0.38	0.62

$\chi^2 = 4.97 - p \leq 0.05 - d.o.f. = 1.$

3.3. Interspersed elements

Retroposons are genetic elements that are propagated via a process known as retroposition. Retroposons are gathered in two major groups: Short Interspersed Elements (SINEs; Okada and Ohshima, 1995) and Long Interspersed Elements (LINEs; Eickbush, 1992).

The camel *CSN3* gene sequence is characterised by 7 DNA fragments (boxes from A to G) showing high similarity (on average 80%) to interspersed elements belonging to both groups (Fig. 4).

Three elements (A, B, C) have been found in the intron 1. In particular, two (A and C) located between the nucleotides +1728/+1925 and +3239/+3306 appear to be LINEs, both belonging to the family L1. The C element has a complementary sequence and it seems to be a truncated LINE (68 bp long), flanked upstream by (TA)_n repeat of 78 bp and downstream by an AT rich region of 24 bp. This element appears characteristic of the camel *CSN3* sequence, since it has not been found by the comparison with bovine, buffalo, donkey, and human κ -casein gene.

The B element is also complementary and it is situated between the nucleotides +2547/+2670. This is a SINE belonging to the class of mammalian-wide interspersed repeats (MIR-3) and it seems to be absent in bovine and buffalo κ -casein gene sequences, whereas it characterizes the camel, donkey and human *CSN3* gene.

Elements D and E are located in the second intron. The former (from +4972 to +5163) shows a similarity of 87.5% with a LINE 1, whereas the latter (+6394/+6622) has a similarity of 81.1% with a complementary Bov-A2 sequence (Lenstra et al., 1993).

Two additional complementary SINEs (MIR-b) have been found in the camel *CSN3* gene, one (F) is located between the nucleotides +7548/+7633 of the intron 3, whereas the other (G) is positioned at the level of the intron 4 between the nucleotides +9542/+9627. As for the element B, also the retroposon G is not present in bovine and buffalo *CSN3* sequence, however it characterizes also the human and donkey κ -casein gene.

The seven repetitive elements observed at the camel *CSN3* gene represent the 10.4% of the sequence deposited in the EMBL database. This feature raises to 21.4% in the bovine counterpart because of six extra repetitive elements (H, I, L, M, N, O) for a total of 2791 bp observed in this species (Fig. 4). The camel *CSN3* is also characterized by several low complexity repeats. All of them belong to AT rich regions and they cover a total of 321 bp, representing approximately the 3.4% of the gene.

All vertebrate genomes contain repetitive sequences, for instance the 45%, 37% and 55% of the human, mouse and opossum genomes are made up of these elements (Gentles et al., 2007; Lander et al., 2001). Transposable elements have been reported to have a large influence on evolution of the mammalian genome playing a prominent role in species diversification (Bowen and Jordan, 2002). The lower number of repetitive element in camel is an indication that *Tylopoda* diverged from *Ruminantia* before additional retrotransposition events took place at different times during the divergence of in such suborder (Nijman et al., 2002). In such contest, the distinctive retroposon element is typical of camel sequence, and as the extra elements that characterize the bovine sequence, it can be considered a powerful phylogenetic marker for evolution and clustering studies.

3.4. Genotyping

The transition g.1029T>C creates a restriction site for the endonuclease *AluI*, thus a PCR-RFLP protocol was set up for the quick genotyping of the samples. The digestion of the PCR product (488 bp) allows the identification of both alleles (Fig. 5).

The restriction pattern is characterized by 3 fragments of 203 bp, 158 bp and 127 bp for the TT samples, whereas the band 158 bp

long is further restricted into two fragments of 120 bp and 38 bp in the presence of cytosine. The restriction pattern of the heterozygous samples results in 5 fragments. The frequency of cytosine in the sample of 188 camels was 0.38, with a variation among the breeds between 0.30 and 0.46 (Table 4). This value might open the possibility for a rapid directional selection in favor of the C allele, which is responsible for the creation of an extra putative site for the HNF-1 transcription factor. Hardy-Weinberg disequilibrium was detected for the genotype distribution ($\chi^2=4.97$), but this event could be the result of several causes. In particular, the restricted number of genotyped individuals and the mating system. In the last case, even though the mating technique in camel is exclusively natural, we can't exclude inbreeding effect due to the use of the same stallion.

4. Conclusions

1. For the first time, the *CSN3* gene was completely sequenced and characterized in *C. dromedarius* species. We also provided for the first time the complete cDNA sequence of *C. bactrianus*. On the whole, the camel *CSN3* gene shares a similar organization compared to the other species. In particular some differences in intronic size characterize camel sequence. The ratio exon/intron is higher (1:10.41) than that one observed in bovine (1:14.42).
2. Such difference in size is also due to a different number of interspersed elements (7 in camel vs 10 in bovine) which account of 10.4% of the camel sequence vs 21.4% in the bovine counterpart. One LINE truncated appears to be characteristic of the camel *CSN3* sequence, since it has not been found by the comparison with bovine, buffalo, donkey, and human κ -casein gene, and it could be used as genetic marker typical of the specie, as well as for evolution and clustering studies.
3. The promoter sequence was deeply characterized and used as reference to estimate the genetic distance between 13 taxa. The regulatory regions of camels seem to be more related to equids than to ruminants. 25 putative consensus sequences for transcription factors have been identified and 6 of them are very well conserved. Oct-1 and TBP are very representative motifs of camel κ -casein gene promoter, and although further studies are necessary, it would be reasonable to suggest a possible regulation of gene expression through this way.
4. It's also interesting to notice the higher homology of camel translation starting codon with Kozak sequence. This event should guarantee a stronger consensus for camels and therefore an higher efficiency of mRNA translation.
5. A total of 17 SNP have been found, 7 by the comparison of our sequences and 10 by the comparison of the sequences already available in database. The g.1029T>C SNP, located in the promoter region, creates an extra putative site for the transcription factor HNF-1. Further investigations are required to verify the influence of HNF-1 allelic variant in camel κ -casein gene regulation.
6. The polymorphism detected in the promoter region (g.1029T>C) is the first genetic marker for the casein cluster in camels, since no information is available so far. Population data indicate that the frequency of the C allele is 0.38 thus opening the possibility for a rapid directional selection in favor of such allele.

Acknowledgment

The authors thank camel owners in Sudan for providing samples from *C. dromedarius* and Dr. Tobias Knauf-Witzens from Wilhelma Stuttgart Germany for providing samples from *C. bactrianus*. This work was partly funded by DAAD by a scholarship to E.S. Shuiep.

References

- Alexander, L.J., Stewart, A.F., Mackinlay, A.G., Kapelinskaya, T.V., Tkach, T.M., Gorodetsky, S.I., 1988. Isolation and characterization of the bovine kappa-casein gene. *Eur. J. Biochem.* 178, 395–401.
- Al-haj, O.A., Al Kanhal, H.A., 2011. Compositional, technological and nutritional aspects of dromedary camel milk. *Int. Dairy J.* 20, 811–821.
- Bowen, N.J., Jordan, I.K., 2002. Transposable elements and the evolution of eukaryotic complexity. *Curr. Issues Mol. Biol.* 4, 65–76.
- Carneiro, M., Ferrand, N., 2007. Extensive intragenic recombination and patterns of linkage disequilibrium at the CSN3 locus in European rabbit. *Genet. Sel. Evol.* 39, 341–352.
- Caroli, A., Chiatti, F., Chessa, S., Rignanese, D., Bolla, P., Pagnacco, G., 2006. Focusing on the goat casein complex. *J. Dairy Sci.* 89, 3178–3187.
- Caroli, A.M., Chessa, S., Erhardt, G.J., 2009. Milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92, 5335–5352.
- Cosenza, G., et al., 2011. Translational efficiency of casein transcripts in Mediterranean river buffalo. *J. Dairy Sci.* 94, 5691–5694.
- Dunn, C.A., Medstrand, P., Mager, D.L., 2003. An endogenous retroviral long terminal repeat is the dominant promoter for human β 1,3-galactosyltransferase 5 in the colon. *PNAS* 100, 12841–12846.
- Eickbush, T.H., 1992. Transposing without ends: the non-LTR retrotransposable elements. *New Biol.* 4, 430–440.
- Farah, Z., Farah-Riesen, M., 1985. Separation and characterization of major components of camel milk casein. *Milchwissenschaft* 40, 669–671.
- Farah, Z., Mollet, M., Younan, M., Dahir, R., 2007. Camel dairy in Somalia: limiting factors and development potential. *Livest. Sci.* 110, 187–191.
- Geisler, J.H., Uhen, M.D., 2005. Phylogenetic relationships of extinct cetartiodactyls: results of simultaneous analyses of molecular, morphological, and stratigraphic data. *J. Mamm. Evol.* 12, 145–160.
- Gentles, A.J., et al., 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.* 17, 992–1004.
- Gerencsér, A., Barta, E., Boa, S., Kastanis, P., Bosze, Z., Whitelaw, C.B., 2002. Comparative analysis on the structural features of the 5' flanking region of k-casein genes from six different species. *Genet. Sel. Evol.* 34, 117–128.
- Hinz, K., O'Connor, P.M., Huppertz, T., Ross, R.P., Kelly, A.L., 2012. Comparison of the principal proteins in bovine, caprine, buffalo, equine and camel milk. *J. Dairy Res.* 79, 185–191.
- Hobor, S., Kunej, T., Dovc, P., 2008. Polymorphisms in the kappa casein (CSN3) gene in horse and comparative analysis of its promoter and coding region. *Anim. Genet.* 39, 520–530.
- Jollès, P., Loucheux-Lefebvre, M.H., Henschen, A., 1978. Structural relatedness of k-casein and fibrinogen γ -chain. *J. Mol. Evol.* 11, 271–277.
- Kappeler, S., Farah, Z., Puhani, Z., 1998. Sequence analysis of *Camelus dromedarius* milk caseins. *J. Dairy Res.* 65, 209–222.
- Kappeler, S.R., Farah, Z., Puhani, Z., 2003. 5'-Flanking regions of camel milk genes are highly similar to homologue regions of other species and can be divided into two distinct groups. *J. Dairy Sci.* 86, 498–508.
- Kozak, M., 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* 266, 19867–19870.
- Lander, E.S., et al., 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lenstra, J.A., Van Boxtel, A.F., Zwagstra, K.A., Schwerin, M., 1993. Short interspersed nuclear element (SINE) sequences of the Bovidae. *Anim. Genet.* 24, 33–39.
- Malewski, T., 1998. Computer analysis of distribution of putative cis- and trans regulatory elements in milk protein gene promoters. *Biosystems* 45, 29–44.
- McBryan, J., Howlin, J., Kenny, P.A., Shioda, T., Martin, F., 2007. ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. *Oncogene* 26, 6406–6419.
- Mendel, D.B., Crabtree, G.R., 1991. HNF-1, a member of a novel class of dimerizing homeodomain proteins. *J. Biol. Chem.* 266, 677–680.
- Nijman, I.J., van Tessel, P., Lenstra, J.A., 2002. SINE retrotransposition during the evolution of the Pecoran ruminants. *J. Mol. Evol.* 54, 9–16.
- Okada, N., Ohshima, K., 1995. Evolution of tRNA-derived SINEs. In: Marais, R.J. (Ed.), The impact of short interspersed elements (SINEs) on the host genome. R.G. Landes, Austin, Tex., pp. 61–79.
- Pfützner, E., Jähne, R., Wissler, M., Stoecklin, E., Groner, B., 1998. p300/CREB-binding protein enhances the prolactin-mediated transcriptional induction through direct interaction with the transactivation domain of Stat5, but does not participate in the Stat5-mediated suppression of the glucocorticoid response. *Mol. Endocrinol.* 12, 1582–1593.
- Ramunno, L., et al., 2005. Comparative analysis of gene sequence of goat CSN1S1 F and N alleles and characterization of CSN1S1 transcript variants in mammary gland. *Gene* 345, 289–299.
- Reiner, A.P., et al., 2009. Common coding variants of the HNF1A gene are associated with multiple cardiovascular risk phenotypes in community-based samples of younger and older European-American adults: the coronary artery risk development in young adults study and the cardiovascular health study. *Circ. Cardiovasc. Genet.* 2, 244–254.
- Rijnkels, M., 2002. Multi specie comparison of the casein gene loci and evolution of casein gene family. *J. Mamm. Gland Biol.* 7, 327–345.
- Robitaille, G., Britten, M., Morisset, J., Petitclerc, D., 2005. Polymorphism in the bovine k-casein (CSN3) gene and the 5'-flanking region: sequence analysis of CSN3 A and B alleles. *Anim. Genet.* 36, 184–185.
- Shuieip, E.S., Giambra, I.J., El Zubeir, I.E.Y.M., Erhardt, G., 2012. Biochemical and molecular characterization of polymorphisms of α s1-casein in Sudanese camel (*Camelus dromedarius*) milk. *Int. Dairy J.* <http://dx.doi.org/10.1016/j.idairyj.2012.09.002>.
- Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. *Molecular Cloning*, ed. Cold Spring Harbor, New York.
- Schild, T.A., Geldermann, H., 1996. Variants within the 5'-flanking regions of bovine milk-protein-encoding genes. III. Genes encoding the Ca-sensitive caseins α s1, α s2 and β . *Theor. Appl. Genet.* 93, 887–893.
- Schild, T.A., Wagner, V., Geldermann, H., 1994. Variants within the 5'-flanking regions of bovine milk protein genes: I. κ -casein-encoding gene. *Theor. Appl. Genet.* 89, 116–120.
- Vanselow, J., et al., 2006. DNA-remethylation around a STAT5-binding enhancer in the alphaS1-casein promoter is associated with abrupt shutdown of alphaS1-casein synthesis during acute mastitis. *J. Mol. Endocrinol.* 37, 463–477.
- Ward, T.J., Honeycutt, R.L., Derr, J.N., 1997. Nucleotide sequence evolution at the k-casein locus: evidence for positive selection within the family bovidae. *Genetics* 147, 1863–1872.
- Weimann, C., Meisel, H., Erhardt, G., 2009. Bovine κ -casein variants result in different angiotensin I converting enzyme (ACE) inhibitory peptides. *J. Dairy Sci.* 92, 1885–1888.
- Zhao, F.Q., Adachi, K., Oka, K., 2002. Involvement of Oct-1 in transcriptional regulation of β -casein gene expression in mouse mammary gland. *Biochim. Biophys. Acta* 1577, 27–37.
- Zwilling, S., Annweiler, A., Wirth, T., 1994. The POU domains of the Oct-1 and Oct-2 transcription factors mediate specific interaction with TBP. *Nucleic Acids Res.* 22, 1655–1662.