IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Alpha S1-casein polymorphisms in camel (Camelus dromedarius) and descriptions of biological active peptides and allergenic epitopes**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available http://hdl.handle.net/2318/1597874 since 2016-10-04T09:18:54Z

*Published version:*

DOI:10.1007/s11250-016-0997-6

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

12 October 2024

This is the author's final version of the contribution published as:

Erhardt, Georg; Shuiep, El Tahir Salih; Lisson, Maria; Weimann, Christina; Wang, Zhaoxin; El Zubeir, Ibtisam El Yas Mohamed; Pauciullo, Alfredo. Alpha S1-casein polymorphisms in camel (Camelus dromedarius) and descriptions of biological active peptides and allergenic epitopes. TROPICAL ANIMAL HEALTH AND PRODUCTION. 48 (5) pp: 1-9.
DOI: 10.1007/s11250-016-0997-6

The publisher's version is available at:
http://link.springer.com/content/pdf/10.1007/s11250-016-0997-6

When citing, please refer to the published version.

Link to this full text:
http://hdl.handle.net/

This full text was downloaded from iris - AperTO: https://iris.unito.it/

iris - AperTO

University of Turin's Institutional Research Information System and Open Access Institutional Repository

# Alpha S1-casein polymorphisms in camel (*Camelus dromedarius*) and descriptions of biological active peptides and allergenic epitopes

Georg Erhardt [1,*]

Email georg.erhardt@agrar.uni-giessen.de

El Tahir Salih Shuiep

Maria Lisson [1]

Christina Weimann [1]

Zhaoxin Wang [1]

Ibtisam El Yas Mohamed El Zubeir [3]

Alfredo Pauciullo [1,4]

[1] Department of Animal Breeding and Genetics, Justus-Liebig-University Giessen, Ludwigstr. 21B, 35390 Giessen, Germany

[2] Institute of Molecular Biology, University of Nyala, 155 Nyala, Sudan

[3] Department of Dairy Production, University of Khartoum, P.O. Box 321, Khartoum, Sudan

[4] Department of Agricultural, Forest and Food Sciences (DISAFA), University of Torino, Largo Paolo Braccini, 2, 10095 Grugliasco (TO), Italy

## Abstract

Milk samples of 193 camels (*Camelus dromedarius*) from different regions of Sudan were screened for casein variability by isoelectric focusing. Kappa-casein and beta-casein were monomorphic, whereas three protein patterns named $\alpha_{s1}$-casein A, C, and D were identified. The

major allele A revealed frequencies of 0.79 (Lahaoi), 0.75 (Shanbali), 0.90 (Arabi Khali), and 0.88 (Arabi Gharbawi) in the different ecotypes. *CSN1S1**C shows a single G > T nucleotide substitution in the exon 5, leading to a non-synonymous amino acid exchange (p.Glu30 > Asp30) in comparison to *CSN1S1*A* and *D*. At cDNA level, no further single nucleotide polymorphisms could be identified in *CSN1S1* A*, *C*, and *D*, whereas the variants *CSN1S1*A* and *CSN1S1*C* are characterized by missing of exon 18 compared to the already described *CSN1S1*B*, as consequence of DNA insertion of 11 bp at intron 17 which alter the pre-mRNA spliceosome machinery. A polymerase chain-restriction fragment length polymorphism method (PCR-RFLP) was established to type for G > T nucleotide substitution at genomic DNA level. The occurrence and differences of IgE-binding epitopes and bioactive peptides between $\alpha_{s1}$-casein A, C, and D after digestion were analyzed *in silico*. The amino acid substitutions and deletion affected the arising peptide pattern and thus modifications between IgE-binding epitopes and bioactive peptides of the variants were found. The allergenic potential of these different peptides will be investigated by microarray immunoassay using sera from milk-sensitized individuals, as it was already demonstrated for bovine $\alpha_{s1}$-casein variants.

## Keywords

Milk proteins
Genetic polymorphisms
*CSN1S1*
IgE-binding epitopes
Bioactive peptides

This article belongs to the Topical Collection: Camelids. Guest Editor: Bernard Faye

# Introduction

Camel milk plays an important role as protein source for many humans especially for the people living in the arid lands of the world (Konuspayeva et al. 2009). In addition, there is a growing interest in usage of camel milk as a healthy food (Nikkah 2011) and alternative protein source for humans with milk protein allergy (Hinz et al. 2012). However,

the knowledge in milk protein of this species and the genetic variation is very limited. Kappeler et al. (1998) described two genetic variants of $\alpha_{s1}$-CN (*CSN1S1**A; *CSN1S1**B) by protein- and mRNA-sequencing within Somali camel (*Camelus dromedarius*). Shuiep et al. (2013) reported a partial genomic DNA sequence for αs1-casein and could demonstrate genetic variation at protein and DNA level in two camel ecotypes of Sudan, whereas in the same camel population, Pauciullo et al. (2013, 2014) have found genetic variants in the regulatory regions of the kappa- and beta-casein (κ-CN; *CSN3* and β-CN; *CSN2*) genes. In sheep, isoelectric focusing of milk samples led to the identification of further alleles at alphas1-casein ($\alpha_{s1}$-CN; *CSN1S1*) and $\alpha_{s2}$-CN (*CSN1S2*). In addition, in East Friesian Dairy sheep and Lacaune sheep, *CSN1S1* variants were significantly associated with higher protein content (Giambra et al. 2014). Lisson et al. (2014) could demonstrate differences in the immunoreactivity of $\alpha_{s1}$-CN epitopes not only between species but also between genetic variants of cows. In bovine and non-bovine milk, ACE-inhibitory properties could be demonstrated (Weimann et al. 2009, Salami et al. 2011, Moslehishad et al. 2013, Jrad et al. 2014) and proposed to be used as functional food with biological activity.

Therefore, the aim of this study was to investigate further the occurrence of polymorphisms in camel $\alpha_{s1}$-CN especially in Sudan. Furthermore, options to describe the allergenic potential as well as bioactive peptides of those protein variants should be demonstrated.

# Material and methods

## Milk and blood samples

Milk and blood samples were collected from Lahaoi ($n = 81$), Shanbali ($n = 40$), Arabi Kahli ($n = 51$), and Arabi Gharbawi ($n = 21$) camel (*C. dromedarius*) ecotypes in Sudan. Areas of sampling were El Shuak, El Obied, west Omdurman, and Nyala, respectively (Fig. 1). Fresh milk samples ($n = 5$) of *C. dromedarius* were obtained from Kamelhof Rotfelden (Rotfelden-Ebhausen, Germany) and immediately kept at 4 °C.

**Fig. 1**

Areas of sampling. The locations (1–4) correspond to Table 1

Map No. 4458 UNITED NATIONS
October 2011

## Isoelectric focusing of milk proteins

Simultaneous phenotyping of camel milk protein variability on protein level was done by isoelectric focusing (IEF) according to Erhardt (1989) using a 0.3-mm polyacrylamide gel (T = 5 %, C = 3.75 %, and 7.1 M urea). The following mixture of carrier ampholytes was used: 0.4 % (*w*/*v*) Servalyte pH 3.0–5.0 (Serva Electrophoresis, Heidelberg, Germany), 1.1 % (*w*/*v*) Pharmalyte pH 4.2–4.9 (GE Healthcare Europe GmbH, Freiburg, Germany) and 1.2 % (*w*/*v*) Servalyte pH 5.0–7.0 (Serva Electrophoresis, Heidelberg, Germany). After prefocusing for 14 min (20 mA, 3.000 V), milk samples as well as standard samples were separated for 67 min (40 mA, 3.000 V).

## Molecular analyses: reverse transcription, PCR amplification, and sequencing

DNA was isolated according to Sambrook et al. (1989) from the blood samples on the filter paper (FTA®Classic Card—Whatman®BioScience, Maidstone, UK). For mRNA extraction, somatic cells of the fresh milk samples with $\alpha_{s1}$-CN AA and AC were gained for mRNA extraction by centrifugation (10 min, 2200×*g*), and washed twice with phosphate buffered saline/0.5 M ethylenediaminetetraacetic acid according to Boutinaud et al. (2002). Invisorb® Spin RNA Mini Kit (Invitek GmbH, Berlin, Germany) was used for extraction of total RNA which was reverse transcribed to cDNA using Verso™ cDNA kit (Thermo Fisher Scientific, Waltman, MA, USA).

For sequencing of *CSN1S1* cDNA, a set of primers for amplification were designed by means of DNAsis-Max ver. 3.0 software (Hitachi Software, San Bruno, CA, USA), using first the complete sequences of camel cDNAs reported by Kappeler et al. (1998) and then the new sequences determined. PCR was used to amplify complete cDNA of CSN1S1.

Primers were designed using GenBank sequence NW_006210750.1 in order to analyze the exons of CSN1S1 from animals carrying $\alpha_{S1}$-CN*A and $\alpha_{S1}$-CN*D. Primer pairs used for identifying the causal mutation on DNA level and their relevant annealing temperatures (x°C) are listed in Table 1.

**Table 1**

Forward and reverse primers, size of PCR products, gene regions, and annealing temp
*CSN1S1*. Primers were designed using Genbank Acc. No. NW_006210750.1

| PCR no. | Forward primer 5′-3′ (name, sequence) | Reverse primer 5' (name, sequence |
|---|---|---|
| 1 | CD-CSN1S1-5′UTR-F GGGAATCTTATTGATGTAACAGT | CD-CSN1S1-Intron1-R TACAAATGCATATAAAT |
| 2 | CD-CSN1S1-Intron1-F CACTTCATGGATCAAATGCG | CD-CSN1S1-Intron2-R TGACCAAAATGTGAAGC |
| 3 | CD-CSN1S1-Intron2-Fb ACTTGGTGTCAGACAGATTTAGCT | CD-CSN1S1-Intron3-R TGGCTGTTGTAGAAGGC |
| 4 | CD-CSN1S1-Intron3-F CCTGTGCCTCCCCTCTCTAT | CD-CSN1S1-Intron4-R TCCTTCTTTACCAGACC |

| | | |
|---|---|---|
| 5 | CD-CSN1S1-Ex4-F<br>TGAACCAGACAGCATAGAG | CD-CSN1S1-Ex6-R<br>CTAAACTGAATGGGTGA |
| 6 | CD-CSN1S1-Intron5-F<br>TGTACCTTTGCAAAACACTCATG | CD-CSN1S1-Intron6-R<br>TGCATTAAGTGTCTCCT |
| 7 | CD-CSN1S1-Intron6-Fb<br>GCAAGGAGACACTTAATGCAAG | CD-CSN1S1-Intron7-RZW<br>TCAGTAGAATTCATCAG |
| 8 | CD-CSN1S1-Intron7-F<br>TTTGGATCTGGATTCTCTACCA | CD-CSN1S1-Intron8-R<br>TGAATTAGAGATATCTG |
| 9 | CD-CSN1S1-Intron8-F<br>GGTTATTTACATTGATTTATCACAACA | CD-CSN1S1-Intron9-R<br>GGCTATAATAGTCCTGG |
| 10 | CD-CSN1S1-Intron9-F<br>ATAAAAAGGGCCCCAAGTGT | CD-CSN1S1-Intron10-R<br>TGGGCAGCACAGATAGT |
| 11 | CD-CSN1S1-Intron10-F<br>CCCAGGAATTTGTGGCTAAA | CD-CSN1S1-Intron12-R<br>ACCACGTGCTGCTAAAA |
| 12 | CD-CSN1S1-Intron12-F<br>ATCCTCGGTCAGAAAGCACA | CD-CSN1S1-Intron13-R<br>ACAGCACATTAGTCTCA |
| 13 | CD-CSN1S1-Intron13-F<br>TGCTGGTGTTTAAAGTCCAAG | CD-CSN1S1-Intron16-R<br>TCAGTTGGTTTCTTCAC |
| 14 | CD-CSN1S1-Intron16-FZW<br>TTTTAAATAATGCAATTGGCAA | CD-CSN1S1-Intron17-RZW<br>AATCATTCTTTCAAAGG |
| 15 | CD-CSN1S1-Intron17-FZW<br>TAGCAAGACTTTTCCATGTAAT | CD-CSN1S1-Intron18-RZW<br>ATTCAAGTAATTCTCTA |
| 16 | CD-CSN1S1-Intron18-FZW<br>AGAGTCAATAACTCTTGTGATG | CD-CSN1S1-Intron19-RZW<br>GGTACCAGAAATTATTA |
| 17 | CD-CSN1S1-Intron19-F<br>CCTTTGTGAGCTGCATCGTAAG | CD-CSN1S1-3′UTR-R<br>CTCTGTTCCCACACCTT |
| 18 | CD-CSN1S1-mRNA-F<br>CTTCTTCCCAGTCTTGGGTTC | CD-CSN1S1-mRNA-R<br>ATAGGCGTGGAGGAGA |
| 19 | CD-CSN1S1-Ex15-F<br>GAGCCTGTGAAAGTAGTGA | CD-CSN1S1-Ex17-R<br>CCTTCAGGGGTGTCGTA |
| 20 | CD-CSN1S1-Ex4-F<br>TGAACCAGACAGCATAGAG | CD-CSN1S1-Ex6-R<br>CTAAACTGAATGGGTGA |

A typical PCR reaction mix (50 µl) comprised 50 ng of gDNA/total cDNA, 1× PCR Buffer (Promega, Madison, WI, USA), 2.5-mM $MgCl_2$, 5 pmol of each primer, dNTPs each at 200 µM, and 1 U of Taq DNA Polymerase (Promega). PCR was performed under the following thermal conditions: 95 °C for 4 min, 35 cycles at 95 °C for 45 s, annealing temperature (x°C)

according to the amplicon (Table 1) for 45 s, 72° C for 60 s, and the final extension at 72 °C for 5 min. The amplified products were analyzed by electrophoresis on 1.5 % agarose gel in 0.5× TBE buffer and stained with Midori Green Advance (Nippon Genetics). The amplified fragments were afterward sequenced using a Big Dye Terminator sequencing kit v.1.1 (Applied Biosystems, Forster City, CA, USA).

## Development of a DNA-based test

In order to screen for an identified non-synonymous nucleotide substitution in exon 5 (c.150G > T GenBank ID JF429138) of camel *CSN1S1* on DNA level, a polymerase chain-restriction fragment length polymorphism (PCR-RFLP) method has been established using *Sml*I (New England Biolabs, München, Germany). PCR product no. 20 (Table 1) containing exon 4 to exon 6 was digested in a 10-µl reaction mixture with *Sml*I (55 °C–180 min). Digested fragments were separated by electrophoresis in a 3 % agarose gel and visualized after staining in Midori Green Advance with ultraviolet light.

## Identification of IgE-binding epitopes and bioactive peptides

To determine peptides containing IgE-binding epitopes and differences between the genetic variants after digestion, potential cleavage sites were predicted using the *in silico* ExPASy tool, Peptide Cutter ( http://web.expasy.org/peptide_cutter/ ). Enzymes chymotrypsin (high-specificity) and trypsin were selected. Resulting peptides were compared with epitopes described in bovine $\alpha_{s1}$-casein (Chatchatee et al. 2001, Lisson et al. 2014). For identification of bioactive peptides, the amino acid sequences of $\alpha_{s1}$-casein A, C, and D were digested *in silico* ( http://www.uwm.edu.pl/biochemia/ ) using the following enzymes: pepsin, trypsin, chymotrypsin A, chymotrypsin C, pancreatic elastase I, pancreatic elastase II, oligopeptidase B, and oligopeptidase II.

## Statistical analyses

Allele and genotype frequencies were calculated using PopGene program v. 1.31 (Yeh et al. 1999). To test the distribution of genotypes on the base of Hardy-Weinberg, chi-square ($c^2$) test was performed.

# Results and discussion

## IEF and allele frequencies

Simultaneous typing of camel milk protein using isoelectric focusing (IEF) revealed polymorphic protein pattern named $\alpha_{s1}$-CN in the most acidic region of the gel (Fig. 2). Three variants, A, C, and D, were observed in this fraction. Each variant consists of one major and two more alkaline-located minor bands. $\alpha_{s1}$-casein A revealed the least acidic apparent isoelectric point (pI), and C was just focused in its acidic side, whereas D revealed the most acidic pI. Nomenclature was done considering Kappeler et al. (1998). Due to this, $\alpha_{s1}$-CN C and $\alpha_{s1}$-CN D seem to be new variants which are clearly separated by IEF. This shows once more that screening of milk samples still offers the possibility to identify further protein variants (Giambra et al. 2010) as starting point for additional milk protein analyses. Screening of all 193 camel milk samples by IEF revealed dominance of $\alpha_{s1}$-CN A in comparison to $\alpha_{s1}$-CN C and $\alpha_{s1}$-CN D. The major allele A revealed frequencies ranging from 0.750 to 0.902 among Shanbali and Arabi Kahali ecotypes, respectively. While $\alpha_{s1}$-CN C was found as 0.111, 0.100, 0.078, and 0.071 in Lahaoi, Shanbali, Arabi Kahali, and Arabi Gharbawi, respectively. Allele D revealed the lowest frequencies in all populations except in Shanbali (Table 2).

**Fig. 2**

Isoelectric focusing of milk samples from *Camelus dromedarius* and the demonstration of different $\alpha_{s1}$-casein (CN) genotypes. Each variant is characterized by a main (marked and used for phenotyping) and up to three minor bands

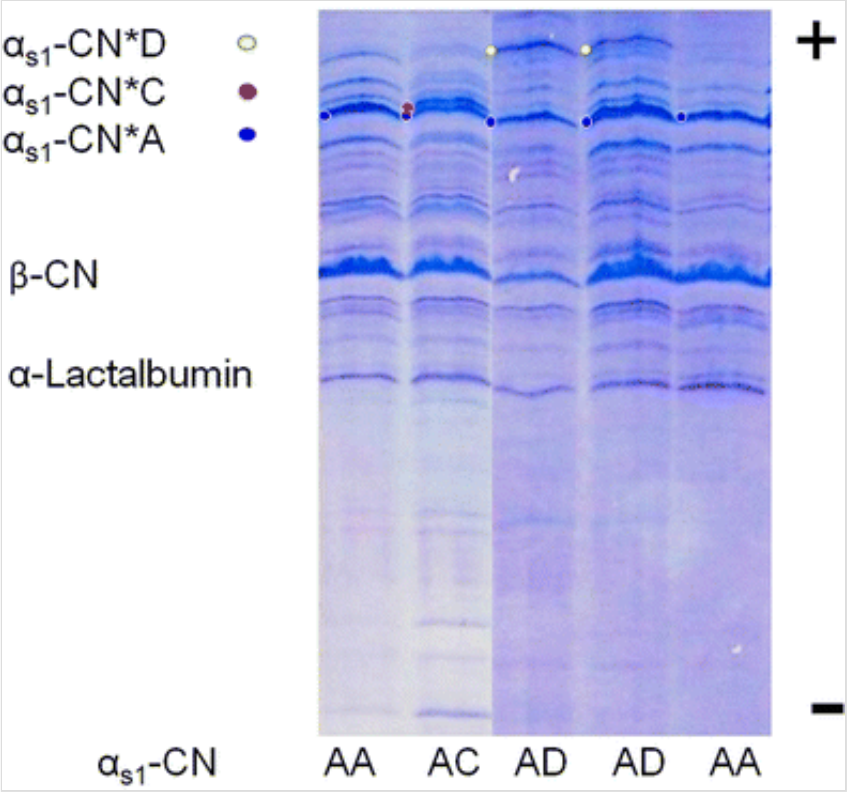



**Table 2**

Genotypes and allele frequencies of camel $\alpha_{s1}$-CN in different ecotypes from differe of Sudan

| No. | Location | Ecotype | *N* | Genotype frequencies (%) | | | Allele frequencies | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | AA | AC | AD | A | C | D |
| 1 | El Shuak | Lahaoi | 81 | 58.0 | 22.2 | 19.8 | 0.790 | 0.111 | 0.099 |
| 2 | El Obied | Shanbali | 40 | 50.0 | 20.0 | 30.0 | 0.750 | 0.100 | 0.150 |
| 3 | Omdurman | Arabi Khali | 51 | 80.4 | 15.7 | 3.9 | 0.902 | 0.078 | 0.020 |
| 4 | Nyala | Arabi Gharbawi | 21 | 76.2 | 14.3 | 9.5 | 0.881 | 0.071 | 0.048 |

## Sequencing of cDNA and genomic DNA

In this study, we found in total 20 exons confirming the results of Pauciullo and Erhardt (2015). Sequencing of the complete coding sequence of camel *CSN1S1* of IEF-pretyped $\alpha_{s1}$-CN A and D DNA samples showed full sequence similarity to $\alpha_{s1}$-CN A of Kappeler et al. (1998). Samples with $\alpha_{s1}$-CN C are characterized by a non-synonymous G > T-

SNP (c.150G > T; GenBank ID JF429138) resulting in the amino acid substitution p.Glu30 > Asp30 in the mature protein. *CSN1S1*C* and $\alpha_{s1}$-CN B (Kappeler et al. 1998) are both characterized by p.Glu30 > Asp30 in the deduced mature protein sequence and are differing by missing or non-missing of exon 18, respectively. Therefore, we postulate a new genetic variant *CSN1S1*C*, coding for 207 AAs like *CSN1S1*A*, and not for 215 AAs like $\alpha_{s1}$-CN B.

Removal of intron sequences from pre-mRNA is carried out by the spliceosome machinery, which recognizes specific sites (donor splice site, branch point, polypyrimidine tract, and acceptor splice site) in a complex molecular mechanism. Any deviation from consensus can result in an overall decreased affinity for the spliceosome (Clark and Thanaraj 2002). In order to verify the presence of these sites, the PCR-fragments containing genomic DNA sequence of exon 17 to exon 19 was sequenced. Sequence analyses showed the presence of exon 18 on genomic level in individuals carrying AA, AD, and AC genotypes. No mutation at the splicing donor and acceptor sites in the sequenced animals was identified; however, the following comparison with the genome sequence available in Genbank (acc. no. NW_011591251 ) showed an insertion of 11 bp (ATTGAATAAAA) at the intron 17 between the branch sequence and the polypyrimidine tract just upstream the exon 18 (Fig. 3).

**Fig. 3**

Comparison of the DNA sequence spanning from the exon 17 to exon 19 of the dromedary camel *CSN1S1* gene from the genome sequence (*up*) and the same DNA fragment belonging to the variants *CSN1S1* A, C, and D (*down*). Numbering is relative to the gene sequence submitted with acc. no. JF937195 . *Dots* indicate identical nucleotides, *dashes* show missing base pairs. Branch sequence is *underlined*, whereas polypyrimidine tract is *double underlined*. Junction of splice are reported in *bold underlined*

```
ttaggacattcttaattagcttatttatgtgttaaaattagactttttattaatatagtattgtagtgaaaataacttga   +400
.......................................................................................   +400

gttttcagtctgagatctaacaccttctaatcatatgaccttgagcaaattgtgttgttctctttgaaaactgctttttc   +480
.......................................................................................   +480

atttgtgaaattaatatggaactatcaatttatttcctccttaaactcctatatcctgatttttgcagtgttcttctctt   +560
.......................................................................................   +560

gtgtcttcactttttagctgtagttaactgtgcattcagctcagttgctttggtctattgataattcaacatatatatgg   +640
.......................................................................................   +640

catgctaattttttaaaaaagtatgtaaagaaaaccatagaattttctttataattatgtgattagggagaaaaaaaaaa   +720
.......................................................................................   +720

cctttcctcaagatttaaaacatcgtagtcatcaatcagttttatgataatgattcctcactgtgcatcagattttactc   +800
.......................................................................................   +800

tcaatttcttctgagattctctaattactgcttgaccaaaggaaaaagaagaaaaaaaaaaaaaa-caattactctcttc   +879
.......................................................................a...............   +880

atttgagctccaaagagtacattagcaagacttttccatgtaattaaaaaaaaaagcaagaatgataattataatcttaa   +959
.......................................................-...............................   +959

                          +988   +994                +1013   +1014        +1024
                            ↓      ↓                    ↓       ↓           ↓
agaaagtatttggtagtaataatttagtctatcatggtatttgatgggatttaa-----------attgaaactattttt   +1028
.......................................................attgaataaaa...............   +1039
                                                      +1013/+1024          +1035
                        Exon 18
   +1036      +1043              +1066
      ↓       E  Q  A  Y  F  H  L  E
tcctctctgttaagGAACAGGCATACTTCCATCTTGAGgtaaattgactttattatatttaattttttatctaaatgac   +1108
.......................................................................................   +1119

      ↑       ↑                    ↑
   +1047      +1054              +1077

ttaataaagagaaaatcccaaataccatatttattataatctttacatgaaattctttcattttttctaattccatcaa   +1188
.......................................................................................   +1199

agaacacataacttccacttaccagtatagagaattacttgaattcatagaaacaagtaaaatataatagattaatagtg   +1268
.......................................................................................   +1279

gcagctgttccaaaatttcagaaaggaaaataatgctttgttaacaaaactgtatgattaggaaggagcactgggctagg   +1348
.......................................................................................   +1359

aattggtgaattattttctccttattcctggcatttctgaaagccagctgtattgacccttgggtgggaggtccttcctt   +1428
.......................................................................................   +1439

cccttcctttgagctcagtatcaccatgaaaaattaaactgtagaattaaatggcaaaaattatcctttctagtactaat   +1508
.......................................................................................   +1519

acccatgagtttatacattatgtaataattaaattcacacaactactttcctctgtactcaggggattgaacccctgtta   +1588
.......................................................................................   +1599

gtaatgcaggatcttggaattcaatactgactattgaccttaaagaaatgatttactctgataccctaaaaaagcccttt   +1668
.......................................................................................   +1679

ccaaaagttacttcaccatgatccaccaaagggaagagtttgatttctcctttttcctcagagtcaataactcttgtgat   +1748
.......................................................................................   +1759

gccactggccatgttggaaatgaatgaagtgattcattttcagcagttcaaaacagcagtttctcatacattgttgcttt   +1828
.........................a.............................................................   +1839

                                Exon 19
                        P  F  P  Q  F  F  Q  L  G  A
ttcaatggtctttctctctagCCTTTCCCACAATTCTTCCAGCTTGGTGCC        EMBL Acc. No. NW_011591251
.......................................................      EMBL Acc. No. JF937195
```

AQ2

The genome sequence might be likely considered as the variant *CSN1S1*B*. In fact, it is characterized by the presence of a thymine at the exon 5 like the variant *CSN1S1*C*, but it does not show the 11-bp insertion at the intron 17. In order to verify whether the 11-bp insertion could affect the secondary structure of the pre-mRNA, both sequences (from the genome and the allele object of the present study) underwent computation analysis by RNAfold server ( http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi ).

The result evidenced that the insertion of 11 bp negatively affect the secondary structure of the pre-mRNA for the investigated alleles (Fig. 4). In fact, the polypyrimidine tract and the following exon 18 appear both extremely hairpin looped and coiled (Fig. 4), whereas the same regions in the pre-mRNA of the putative allele B are well extended (Fig. 5).

**Fig. 4**

**a** Secondary structure prediction of the pre-mRNA generated by the transcription of the DNA fragment spanning from the exon 17 to exon 19 of the dromedary camel *CSN1S1* gene alleles A, C, and D. **b** Detail of the structure with relative indication of branch point and 11-bp insertion. The polypyrimidine tract and exon 18 appear hairpin looped and coiled

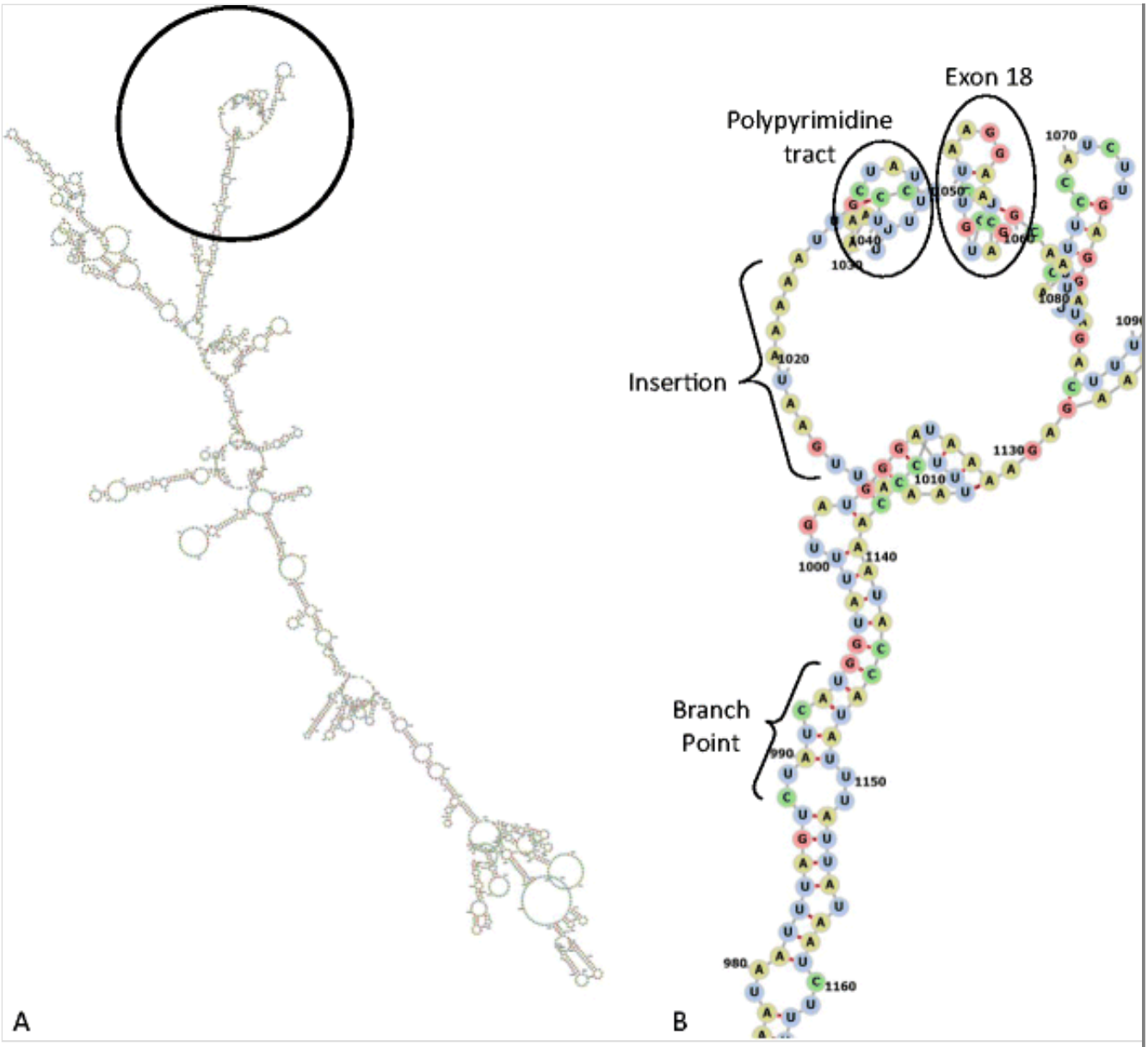
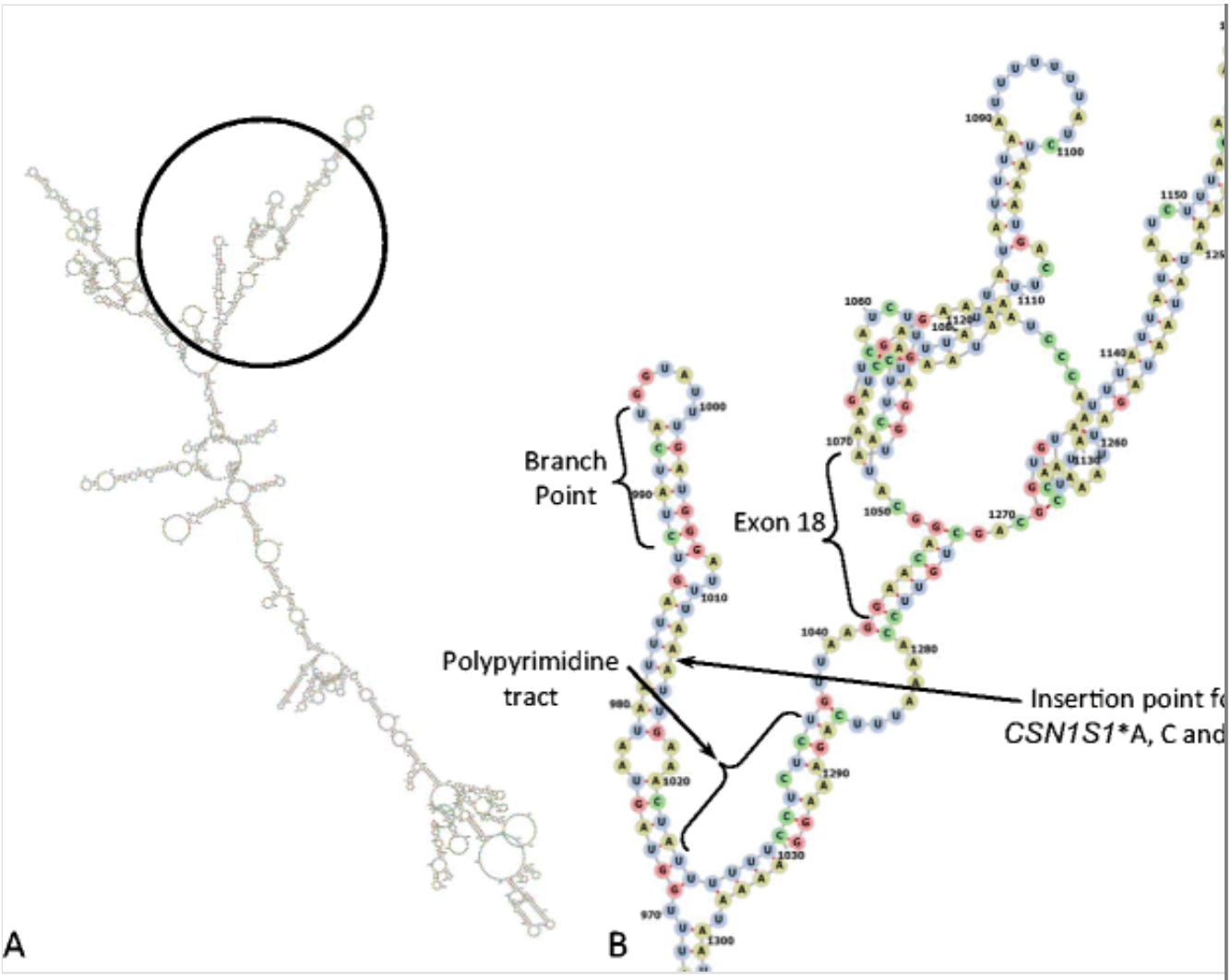



**Fig. 5**

**a** Secondary structure prediction of the pre-mRNA generated by the transcription of the DNA fragment spanning from the exon 17 to exon 19 of the allele B for the dromedary camel *CSN1S1* gene. **b** Detail of the structure with relative indication of branch sequence and insertion point. The polypyrimidine tract and exon 18 appear extended




Alterations of the polypyrimidine tract have been shown to inhibit the 5′ cleavage reaction, binding of the splicing factor U2AF and the U2snRNP, and splicing complex assembly (Frendeway and Keller 1985). Specifically, exon skipping as a consequence of the formation of mRNA secondary structure has been already suggested for the goat $\alpha_{s1}$-casein (Leroux et al. 1992) and human β-casein (Martin and Leroux 1992). Of particular interest is the proposal of Leroux et al. (1992) which is similar to that described here. They suggest that the skipping of exon 9 of the goat $\alpha_{s1}$-casein gene is caused by an 11-bp insertion within intron 9 base-pairing with the donor site. Therefore, it seems likely that also the variants *CSN1S1**A, *CSN1S1**C, and *CSN1S1** D (all characterized by this 11-bp insertion) were affected by the same alteration of the spliceosome machinery and, therefore, preferentially splice-out of the exon 18.

The lack of the exon 18 is the usual form within camel *CSN1S1* mRNA and $\alpha_{s1}$-CN protein, as frequencies of the alleles without this exon seems to be higher in *C. dromedarius*. A similar event also characterizes the $\alpha_{s1}$-CN of South American camelids (Pauciullo and Erhardt 2015). In particular, the

llama αs1-cDNA showed two variants, with and without the exon 18 coding for the same octapeptide (EQAYFHLE) which differentiates the variants A and B found by Kappeler et al. (1998). However, in llama, the shorter variant represents only 26.4 % of the $\alpha_{s1}$-cDNA transcripts, whereas in the present study, there was no simultaneous occurrence of cDNA containing and missing the exon 18. Therefore, further studies need to clarify such a difference between old world and new world camelids, at least in regard of the efficiency of transcription for short and full-length mRNAs.

AQ3

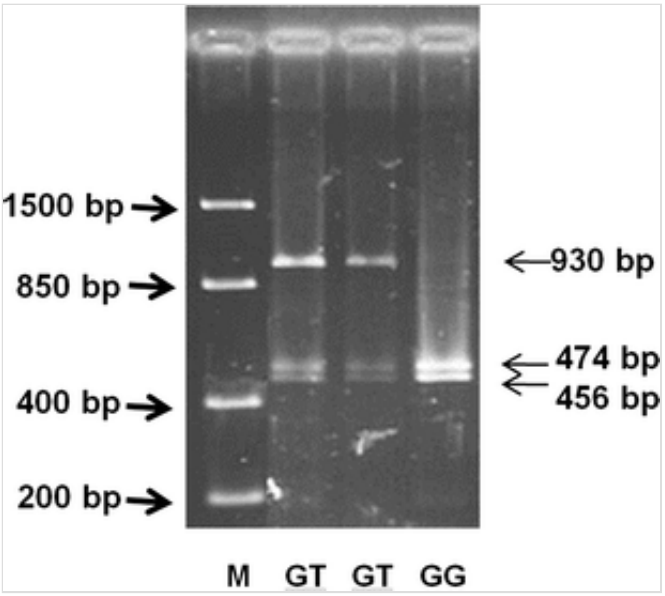
For *CSN1S1*D*, we could finally clarify the same genetic event characterizing the alleles *CSN1S1**A and C; however, no further investigation was possible as the necessary mRNA sample was not available.

## DNA-based test

Using genomic DNA, a PCR-RFLP test was established for typing the g.942G > T SNP (GenBank ID JF429140) characteristic for the new *CSN1S1*C* variant (Fig. 6) The test developed confirmed the IEF and can therefore be used for typing camel *CSN1S1* variability independent of age, gender, and lactation stage. This can be useful for broader analysis of camel milk protein variability and is a requirement for further analyses concerning associations between milk protein variability and milk performance traits in camel, as already done for cattle, goat, and sheep (Martin et al. 2002; Caroli et al. 2009; Giambra et al. 2011).

**Fig. 6**

PCR-RFLP for genotyping camel *CSN1S1**A and C. *M* = marker, CSN1S1 AC (GT 930, 474, and 456 bp) CSN1S1 AA (GG, 474 and 456 bp)




## IgE-binding epitopes and bioactive peptides

A comparison of the identified peptides after *in silico* digestion of $\alpha_{s1}$-casein A, B, and C with IgE-binding epitopes described in bovine $\alpha_{s1}$-casein showed that IgE-binding epitopes f24-43 (NKRKILELAVVSPIQFRQEN) and f153-168 (TQEQAYFHLEPFQQFF) did not resist digestion in intact form and that the amino acid substitutions and deletion (underlined) influence the resulting peptide pattern. Thus, peptide f28-40 (ILELAVVSPIQFR) occurred in all three variants. Peptide f151-159 (VVTQPFQQF) was only identified in variants A and C, whereas peptides f151-158 (VVTQEQAY) and f160-167 (HLEPFQQF) were found in variant B. These peptides comprise major parts of the IgE-binding epitope f153-168. To determine the allergenic potential of these peptides, microarray immunoassay using synthesized peptides and sera from humans with cow milk allergy are currently in progress. A total of 83 bioactive peptides with ACE inhibitory (53), vasoactive/glucose uptake stimulating (8), antioxidative (12), inhibitory (9), and hypotensive (1) activities were found in $\alpha_{s1}$-casein A, B, and C (Table 3). Due to the deletion of eight amino acids, two further ACE-inhibitory and antioxidative peptides (fAY and fHL) were exclusively identified in $\alpha_{s1}$-casein B. This confirms that genetic variants of camel casein are a source of different bioactive peptides as already described in cattle (Weimann et al. 2009) and reveals the additional potential of camel milk protein variants for human health.

**Table 3**

Predictive bioactive peptides from $\alpha_{s1}$-casein variants A, B, C, and D of camel (in parenthesis is the number of identified peptides) according to the BIOPEP database

(Minkiewicz et al. 2008)

| Effect | Peptide sequence | αs1-casein variant |
|---|---|---|
| ACE inhibitor (53) | VMP, RL, IR, RY, LY, VF, FP, PR, VSP, YL, YP, PL, AW…. | A, C, D, |
| Vasoactive/glucose uptake stimulating (8) | SSS, VL, IL, LI, LL, EE, SE, YL | A, C, D |
| Antioxidative (12) | LH, LY, AH, EL, WY, WYY, LHR, VKL, VKV, KD, IR, LK | A, C, D |
| DPP IV inhibitor (9) | PP, MP, VA, LA, FP, LL, VV, IR, LA | A, C, D |
| Hypotensive (1) | IR | A, C, D |
| ACE inhibitor (2), antioxidative (2) | AY, HL | B |

## Acknowledgment

This work was partly funded by DAAD by a scholarship to E.S. Shuiep and by German Research Foundation (DFG ER 122/13-2). We thank the camel owners in Sudan and Mr. Wilhelm Breitling Kamelhof Rotfelden (Germany) for providing milk and blood samples.

### Compliance with ethical standards

*Statement of animal rights* The manuscript does not contain clinical studies or patient data.

*Conflict of interest* The authors declare that they have no competing interests.

# References

Boutinaud, M., Rulquin, H., Keisler, D. H., Djiane, J., Jammes, H., 2002. Use of somatic cells from goat milk for dynamic studies of gene expression in the mammary gland, Journal of Animal Science, 80, 1258–1269.

Caroli, A. M., Chessa, S., Erhardt, G.J., 2009. Invited review: milk protein polymorphisms in cattle: effect on animal breeding and

nutrition, Journal of Dairy Science, 92, 5335–5352.

Chatchatee, P., Järvinen, K. M., Bardina, L., Beyer, K., Sampson, H., A., 2001. Identification of IgE- and IgG-binding epitopes on alpha(s1)-casein: Differences in patients with persistent and transient cow's milk allergy. Journal of Allergy and Clinical Immunology, 107, 379–383

Clark, F., Thanaraj, T.A., 2002. Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Human Molecular Genetics,11,451-464.

Erhardt, G., 1989. κ-Kaseine in der Rindermilch—Nachweis eines weiteren Allels (κ-$CN^E$) in verschiedenen Rassen. Journal of Animal Breeding and Genetics, 106, 225–231.

Frendeway, D., Keller, W., 1985. The stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell, 42, 355–367.

Giambra, I. J., Jäger, S., Erhardt, G., 2010. Isoelectric focusing reveals additional casein variants in German sheep breeds. Small Ruminant Research, 90, 11–17.

Giambra, I. J., Brandt, H., Erhardt, G., 2014. Milk protein variants are highly associated with milk performance traits in East Friesian Dairy and Lacaune sheep, Small Ruminant Research, 121, 382–394

Giambra, I. J., Brandt, H., Erhardt, G., 2011. Milk protein variants and their associations to milk performance traits in East Friesian Dairy sheep. In Proceedings of the 62th Annual Meeting of the European Association for Animal Production, pp. 46, Stavanger, Norway

~~Green, M. R., 1986. Pre-mRNA splicing. Annual Review of Genetics, 20, 671–708.~~

AQ4

~~Harris, N., Senapathy, P., 1990. Distribution and consensus of branch point signals in eukaryotic genes: a computerised statistical analysis. Nucleic Acids Research, 18, 3015–3019.~~

AQ5

Hinz, K., O'Connor, P. M., Huppertz, T., Ross, P. M., Kelly, A. L., 2012. Comparison of the principal proteins in bovine, caprine, buffalo, equine and camel milk, Journal of Dairy Research, 79, 185–191.

Jrad, Z., El Hatmi, H., Adt, I., Girardet, J.M., Cakir-Kiefer, C., Jardin, J., Degraeve, P., Khorchani, T., Oulahal, N., 2014. Effect of digestive enzymes on antimicrobial, radical scavenging and angiotensin I-converting enzyme inhibitory activities of camel colostrum and milk proteins. Dairy Science and Technology, 94, 205–224.

Kappeler, S., Farah, Z., Puhan, Z., 1998. Sequence analysis of *Camelus dromedarius* milk caseins. Journal of Dairy Research, 65, 209–222.

Konuspayeva, G., Faye, B., Loiseau, G., 2009. The composition of camel milk: A meta-analysis of the literature data. Journal of Food Composition and Analysis, 22, 95–101.

Leroux, C., Mazure, N., Martin, P., 1992. Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat αsl-casein transcripts, Journal of Biological Chemistry, 267, 6147–6157.

Lisson, M., Novak., N., Erhardt, G., 2014. Immunoglobulin E epitope mapping by microarray immunoassay reveals differences in immune response to genetic variants of caseins from different ruminant species. Journal of Dairy Science, 97, 1939–1954.

Martin, P. and Leroux, C., 1992. Exon skipping is responsible for the 9 amino acid residue deletion occurring near the N-terminus of human β-casein. Biochemical and Biophysical Research Communications, 183, 750–757.

Martin, P., Szymanowska, M., Zwierzchowski, L., Leroux, C., 2002. The impact of genetic polymorphisms on the protein composition of ruminants' milks, Reproduction Nutrition Development, 42, 433–459.

Minkiewicz P., Dziuba J., Iwaniak A., Dziuba M., Darewicz M., 2008.

BIOPEP database and other programs for processing bioactive peptide sequences. Journal of AOAC International, 91, 965–980.

Moslehishad, M., Ehsani, M.R., Salami, M., Mirdamadi, S., Ezzatpanah, H., Naslaji, A.N., Moosavi-Movahedi, A.A., 2013. The comparative assessment of ACE-inhibitory and antioxidant activities of peptide fractions obtained from fermented camel and bovine milk by *Lactobacillus rhamnosus* PTCC 1637. International Dairy Science, 29, 82–87.

Nikkah, A., 2011. Equidae, camel, and yak milks as functional foods: a review. Journal of Nutrition and Food Sciences, 1, 116. doi: 10.4172/2155-9600.1000116 .

Pauciullo, A., Erhardt, G., 2015. Molecular Characterization of the Llamas (Lama glama) Casein Cluster Genes Transcripts (*CSN1S1, CSN2, CSN1S2, CSN3*) and Regulatory Regions, PlosOne 10(4):e0124963

Pauciullo, A, Shuiep, E. S., Cosenza, G., Ramunno, L., Erhardt, G., 2013. Molecular characterization and genetic variability at κ-casein gene (*CSN3*) in camels. Gene, 513, 22–30

Pauciullo, A., Giambra, I. J., Iannuzzi, L., Erhardt, G., 2014. The β-casein in camels: molecular characterization of the *CSN2* gene, promoter analysis and genetic variability, Gene, 547, 159–168

Salami, M., Moosavi-Movahedi, A. A., Moosavi-Movahedi, F., Ehsani, M. R., Yousefi, R., Farhadi, M., et al., 2011. Biological activity of camel milk casein following enzymatic digestion, Journal of Dairy Research, 78, 471–478.

Sambrook, J., Fritz, E., Maniatis, T., 1989. Molecular cloning: A laboratory manual. (2nd ed.), (Cold Spring Harbor Laboratory Press, New York, NY, USA)

Shuiep, E. S., Giambra I. J., El Zubeir I. E., Erhardt G., 2013, Biochemical and molecular characterization of polymorphisms of $\alpha_{s1}$-casein in Sudanese camel (*Camelus dromedarius*) milk, International

Dairy Journal, 28, 88–93.

Weimann, C., Meisel, H., Erhardt, G., 2009. Bovine κ-casein variants result in different ACE inhibitory peptides. Journal of Dairy Science, 92, 1885–1888.

Yeh, F. C., Yang, R. C., Boyle, T., 1999. PopGene Version 1.31. Microsoft Windows based freeware for population genetics analysis. ftp://ftp.microsoft.com/Softlib/HPGL.EXE