Gene 513 (2013) 22–30

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

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

A. Pauciullo [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 glycosilated protein belonging to a family of phosphoproteins (αs1, β, αs2, 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.

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 (αs1, β, αs2 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 αs1 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 glycosilation 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; $MgCl_2$, magnesium chloride; dNTP, deoxyribonucleoside triphosphate; TBE, tris boric acid EDTA; C/EBP-α, CCAAT enhancer bind protein-alpha; 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.pauciullo@agrar.uni-giessen.de (A. Pauciullo).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.10.083

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, Lahaoi 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 (Invitek, Germany).

2.2. DNA isolation

The filter paper containing blood samples were soaked (56 °C, overnight) in 500 μl sodium-tris-EDTA buffer with 10 μl 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 μl 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 μl) comprised: 100 ng of genomic DNA, 1X PCR Buffer (Promega), 2.5 mM $MgCl_2$, 5 pmol of each primer, dNTPs each at 200 μM, 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 (Invitek, 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′-CACAAAGATGACTCTGCTAT CG-3′ and *reverse* 5′-GCCCTCCACATATGTCTG-3′, according to standard PCR conditions already reported above. Product specifity 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 μl of each PCR amplification was accomplished with 10 U of *AluI* 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 ($\chi 2$ 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