Gene 547 (2014) 159–168

Contents lists available at ScienceDirect

Gene

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

The β-casein in camels: molecular characterization of the *CSN2* gene, promoter analysis and genetic variability

A. Pauciullo [a,b,*], I.J. Giambra [a], L. Iannuzzi [b], G. Erhardt [a]

[a] Institute for Animal Breeding and Genetics, Justus Liebig University, Ludwigstraße 21 B, 35390 Gießen, Germany
[b] ISPAAM, Laboratory of Animal Cytogenetics and Gene Mapping, National Research Council, via Argine 1085, 80147 Naples, Italy

ARTICLE INFO

Article history:
Received 9 May 2014
Received in revised form 13 June 2014
Accepted 24 June 2014
Available online 25 June 2014

Keywords:
Camel
β-casein
Promoter analysis
Interspersed elements
Bio-peptides
microRNA

ABSTRACT

The β-casein is the most abundant protein in camel milk and its encoding gene (*CSN2*) is considered in other species a 'major' gene for the presence of alleles associated to different level of expression. In the present paper, we report for the first time the characterization of the nucleotide sequence of the whole β-casein-encoding gene (*CSN2*) plus 2141 bp at the 5′-flanking region in *Camelus dromedarius*. The promoter region and the complete cDNA are also provided for the first time in *Camelus bactrianus*. The gene is spread over 7.8 kb and consists of 9 exons varying in length from 24 bp (exon 5) to 519 bp (exon 7), and 8 introns from 95 bp (intron 5) to 1950 bp (intron 1). The composite response element (CoRE) region was identified in the promoter, whereas the presence of mature microRNA sequences improves the knowledge on the factors putatively involved in the gene regulation. A total of 46 polymorphic sites have been detected. The transition g.2126A>G falls within the TATA-box of dromedary *CSN2* promoter with a putative influence on the transcription factor binding activity. The frequency of the G allele is 0.35 in a population of 180 she-camels belonging to 4 different ecotypes. In the same population, a conservative SNP (g.4175C>A) was found at the codon 7 of the signal peptide, whereas a comparative analysis with a cDNA sequence available in the database evidenced a missense SNP (g.4180T^{Leu}>G^{Arg}) at exon 2. Four SNPs were found in the bactrian camel. The SNP c.666G>A is responsible for the amino acid change Met^{201} → Ile and it represents the first missense allele at the β-casein in camels. Finally, five interspersed repeated elements were identified at intronic level, whereas the presence of putative bio-functional peptides belonging to ACE-inhibitor and anti-oxidative families confirms the potential protective role of the camel milk for the human nutrition.

1. Introduction

Camel's milk has played and plays a key role for the outliving of the Bedouin, nomad and pastoral households in the hard arid and semi-arid regions of the world, supporting migrant populations since their domestication happened millennia ago. The Camelidae family consists of six species (*Camelus dromedarius*, *Camelus bactrianus*, *Lama glama*, *Lama guanicoe*, *Vicugna pacos*, *Vicugna vicugna*) among which the dromedary camel is certainly the most economical important domesticated one. According to FAO statistic (2011), most of the dromedary camels live in the Horn of Africa mainly gathered in Somalia, Sudan, Ethiopia and Kenya, which together hold almost 60% of the entire population estimated to be about 24.5 million in the world. Despite their potential to survive on marginal resources in severe environment, camels have not been exploited as an important food source in the aforementioned countries. For instance, only 10% of the total milk produced in these rural regions is of camel origin (Faye and Konuspayeva, 2012). Conversely, in the countries of the Gulf, intensive camel milk production in high scale modernized unit has been already realized (Faye et al., 2002) and genetic improvement programs for the milk production have been implemented (Hermas, 1998; Nagy et al., 2013).

The daily milk production of dromedary camels 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).

As for the other mammals, the main component of camel milk proteins is caseins, among which the β-casein is the most abundant component (~65%) (El-Agamy, 2006; Kappeler et al., 2003). This amount is

Abbreviations: *CSN2*, β-casein-encoding gene; SNP, single nucleotide polymorphism; CN, casein; 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; RFLP, restriction fragment length polymorphism; pI, isoelectric point; CoRE, composite response element; C/EBP, CCAAT enhancer bind protein; MGF, Mammary Gland Factor; GR, glucocorticoid receptor factor; YY1, Ying Yang 1; Oct-1, Octamer bind protein; TBP, Tata Binding Protein; NF-1, Nuclear factor-1; STR, short tandem repeat; SINE, short interspersed element; LINE, long interspersed element; MIR, mammalian-wide interspersed repeats; miR, microRNA.

* Corresponding author at: Institute for Animal Breeding and Genetics, Justus Liebig University, Ludwigstraße 21 B, 35390 Gießen, Germany.
E-mail addresses: alfredo.pauciullo@agrar.uni-giessen.de, alfredo.pauciullo@cnr.it (A. Pauciullo).

http://dx.doi.org/10.1016/j.gene.2014.06.055
0378-1119/

definitively higher than the 45% reported in bovine milk (Farrell et al., 2004) and, since β-CN is more sensitive to peptic hydrolysis than αs-CN (El-Agamy et al., 2009), the percentage of this casein fraction could reflect its higher digestibility rate and the lower incidence of allergy in the gut of infants nourished with camel milk (El-Agamy et al., 2009).

The occurrence of different levels of phosphorylation of the β-casein are reported to affect the availability and distribution of calcium and to influence the micelle stability (Amigo et al., 2000), therefore the β-CN plays an essential role both for the nutrition aspects and for the impact on the technological properties of the milk and dairy products.

The β-casein fraction has been deeply studied in ruminants and well characterized both at protein and DNA levels. Many alleles associated with different rates of protein synthesis have been identified in the corresponding coding gene (*CSN2*). At least 6 genetic variants have been reported in sheep (Ceriotti et al., 2004; Chessa et al., 2010; Chianese, 1997), whereas at least 8 alleles corresponding to 7 β-CN variants have been identified in goat (Caroli et al., 2006; Cosenza et al., 2005, 2007), and at least 17 alleles corresponding to 12 β-CN variants have been identified so far in cattle (Caroli et al., 2009). Conversely, in camels no genetic variants have been reported so far for the β-casein, and all the information we have – to date – has been determined at protein level (El-Agamy et al., 2009; Farah and Farah-Riesen, 1985; Hinz et al., 2012; Kappeler et al., 1998).

Genetic variants in camel milk are known for the αs-1 casein fraction (Shuiep et al., 2013), whereas Kappeler et al. (1998) reported on non-allelic variants of the same casein fraction and described an amino acid exchange in position 30 (p.Glu>Asp) of the mature protein, which is not mentioned in the UniProt-database O97943. The κ-casein gene (*CSN3*) has been also investigated in camels and 17 polymorphic sites have been recently detected by Pauciullo et al. (2013).

On the contrary, the investigation at DNA level of camel β-casein gene (*CSN2*) is limited to the cDNA sequence (Kappeler et al., 1998) and to a comparison of the 5′ flanking regions (Kappeler et al., 2003). With the exception of these examples, no further information is available.

Considering the growing interest that camel milk is globally receiving especially for the potential health benefits obtained through a number of bioactive components of camel milk (for a review Al haj and Al Kanhal, 2011), a deep investigation was undertaken to explore genetic variability at the *C. dromedarius CSN2*. A full characterization and an extensive annotation of the *CSN2* gene and promoter region are provided. We report on the first polymorphisms detected at such locus for the dromedary camel and we describe the first allele at the *CSN2* for the bactrian camel.

2. Materials and methods

2.1. Animals

Blood samples were collected from 180 Sudanese she-camels (*C. dromedarius*, locally known as *Naga*) reared in five regions of the country and belonging to different ecotypes including Shanbali, 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 (Pauciullo et al., 2013). Typical phenotypic characteristics for each ecotype were strictly followed. Blood samples were immediately applied to a 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 3 *C. bactrianus* (bactrian camel) belonging to Wilhelma Zoo (Stuttgart, Germany), whereas hair samples were collected from further 5 bactrian camels at the Frankfurt Zoo (Frankfurt, Germany). These samples were treated according to a Spin Blood Mini Kit (Invitek, Germany) and a Gen-ial All Tissue DNA Kit (Gen-ial, Germany), respectively.

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 amplification conditions and sequencing

Four test samples (one for each ecotype) were chosen for the sequencing of the whole camel *CSN2* gene. DNA regions spanning from nucleotides −2141 to +7845 were amplified by means of a iCycler (Bio-Rad). A set of 31 primers for amplification and sequencing was designed by means of DNAsis-Max ver. 3.0 software (Hitachi), using as preliminary template the complete sequence of camel cDNA available in GenBank (EMBL acc. no. AJ012630) and then the new sequences were determined in the course of the research. A typical PCR reaction mix (50 μl) comprised: 100 ng of genomic DNA, 1× PCR buffer (Promega), 2.5 mM of $MgCl_2$, 5 pmol of each primer, dNTPs each at 200 μM, and 2.5 U of *Taq* DNA Polymerase (Promega). In Table 1 we report the 10 starting overlapping 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), and final extension at 72 °C (10 min).

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

SNP discovery was accomplished by the re-sequencing of additional 40 DNA samples (10 for each ecotype) for all the exons including the corresponding flanking regions and the gene promoter.

2.4. Genotyping by HphI PCR-RFLP

PCR reaction mixture and thermal conditions for the amplification of the DNA fragment 659 bp long and spanning from −428 bp of 5′ flanking region to +231 bp of the camel *CSN2* gene were accomplished by using the following primers: *forward* 5′-GTTTCTCCATTACAGCATC-3′ and *reverse* 5′-TCAAATCTATACAGGCACTT-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 180 animals was genotyped for the g.2126A>G SNP using a PCR-RFLP method. Digestion of 17 μl of each PCR amplification was accomplished with 5 U of *HphI* endonuclease (5′…$GGTGAN_8$↓…3′) (New England Biolabs) over-night at 37 °C. The digestion products were analyzed directly by electrophoresis in 2.5% agarose gel in 0.5× TBE buffer and stained with Midori Green Advance (Nippon Genetics).

2.5. Bioinformatics

The allele frequencies and Hardy–Weinberg equilibrium (χ^2 test) were calculated for the 180 camels. SNP discovery, homology searches, comparison among sequences, and multiple alignments were accomplished using DNAsis-Max ver. 3.0 software (Hitachi Software), whereas the putative transcription factor binding sites were searched by Transfact 7.0 software. Interspersed elements were found by a RepeatMasker Web Server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). Introns were also analyzed for potential microRNA sequences by using the bovine miRBase database (http://www.mirbase.