J. Dairy Sci. 96:699–709
http://dx.doi.org/10.3168/jds.2012-5908

DNA-based identification of novel bovine casein gene variants

J. L. Gallinat,* S. Qanbari,† C. Drögemüller,‡ E. C. G. Pimentel,†§ G. Thaller,* and J. Tetens*[1]
*Institute of Animal Breeding and Husbandry, Christian-Albrechts-University Kiel, D-24118 Kiel, Germany
†Institute of Animal Breeding and Genetics, University of Göttingen, D-37075 Göttingen, Germany
‡Institute of Genetics, Vetsuisse Faculty, University of Bern, CH-3001 Bern, Switzerland
§Department of Animal Breeding, Faculty of Organic Agricultural Sciences, University of Kassel, D-37213 Witzenhausen, Germany

ABSTRACT

In cattle, at least 39 variants of the 4 casein proteins (α_{S1}-, β-, α_{S2}- and κ-casein) have been described to date. Many of these variants are known to affect milk-production traits, cheese-processing properties, and the nutritive value of milk. They also provide valuable information for phylogenetic studies. So far, the majority of studies exploring the genetic variability of bovine caseins considered European taurine cattle breeds and were carried out at the protein level by electrophoretic techniques. This only allows the identification of variants that, due to amino acid exchanges, differ in their electric charge, molecular weight, or isoelectric point. In this study, the open reading frames of the casein genes *CSN1S1*, *CSN2*, *CSN1S2*, and *CSN3* of 356 animals belonging to 14 taurine and 3 indicine cattle breeds were sequenced. With this approach, we identified 23 alleles, including 5 new DNA sequence variants, with a predicted effect on the protein sequence. The new variants were only found in indicine breeds and in one local Iranian breed, which has been phenotypically classified as a taurine breed. A multidimensional scaling approach based on available SNP chip data, however, revealed an admixture of taurine and indicine populations in this breed as well as in the local Iranian breed Golpayegani. Specific indicine casein alleles were also identified in a few European taurine breeds, indicating the introgression of indicine breeds into these populations. This study shows the existence of substantial undiscovered genetic variability of bovine casein loci, especially in indicine cattle breeds. The identification of new variants is a valuable tool for phylogenetic studies and investigations into the evolution of the milk protein genes.

Key words: casein, genetic variant, *Bos indicus*, *Bos taurus*

INTRODUCTION

The 4 caseins (α_{S1}-, α_{S2}-, β- and κ-CN) account for more than 75% of the whole bovine milk protein. They are encoded by 4 genes mapped to BTA 6 (Hayes and Petit, 1993) in a tightly linked 250-kb cluster. The gene order is *CSN1S1* (α_{S1}-CN-encoding gene), *CSN2* (β-CN-encoding gene), *CSN1S2* (α_{S2}-CN-encoding gene), and *CSN3* (κ-CN-encoding gene; Threadgill and Womack, 1990). Caseins do not only provide the suckling infant with calcium, phosphate, and amino acids, but also have an influence on milk-production traits and cheesemaking properties (Boettcher et al., 2004; Wedholm et al., 2006; Nilsen et al., 2009). Consequently, caseins have already been investigated intensively at the protein as well as at the DNA level, especially in cattle. All bovine casein genes have been shown to be polymorphic (Farrell et al., 2004). To date, 9 protein variants of α_{S1}-CN (A, B, C, D, E, F, G, H, and I), 4 protein variants of α_{S2}-CN (A, B, C, and D), 12 protein variants of β-CN (A^1, A^2, A^3, B, C, D, E, F, G, H^1, H^2, and I), and 13 protein variants (A, B, B^2, C, D, E, F^1, F^2, G^1, G^2, H, I, and J) and 1 synonymous variant (A^I) of κ-CN are known (Caroli et al., 2009). The majority of these casein variants are caused by SNP within the open reading frame resulting in amino acid exchanges in the mature protein (missense mutations). Other variants, such as *CSN1S1*A* and *CSN1S2*D*, show deletions of several amino acids due to nucleotide exchanges that affect splice sites and hence are generated by exon skipping (Bouniol et al., 1993; Mohr et al., 1994).

Up to now, new variants were predominantly identified at the protein level by electrophoretic or isoelectrophoretic techniques. This allows the identification of variants differing in their electric charge, their molecular weight, or their isoelectric point (Caroli et al., 2009). At the DNA level, PCR-single-strand conformation polymorphism (PCR-SSCP) analysis is a common tool to detect and genotype casein variants (Prinzenberg et al., 1999). However, the identification of mutations in fragments larger than 200 bp is difficult (Jaeckel et al., 1998).

Received July 5, 2012.
Accepted September 13, 2012.
[1]Corresponding author: jtetens@tierzucht.uni-kiel.de

Table 1. Animals used in the sequencing of casein genes (n = 356)

Breed	Acronym	Origin	n[1]
Bos taurus (n = 285)			
Angler	AN	Germany	20
Eringer	ER	Switzerland	13
German Red Pied	GR	Germany	14
German Yellow	GY	Germany	25
Highland Cattle	HL	United Kingdom	27
Hinterwälder	HW	Germany	21
Hungarian Grey Steppe	HG	Hungary	20
Jersey	JE	United Kingdom	20
Limpurger	LP	Germany	25
Pinzgauer	PG	Germany	17
Retinta	RE	Spain	20
Sarabi	SA	Iran	22
Shorthorn	SH	Germany	16
Vorderwälder	VW	Germany	25
Bos indicus (n = 71)			
Gir	GI	Brazil	25
Golpayegani	GO	Iran	22
Sistani	SI	Iran	24

[1]The number of successfully sequenced animals differed between the genes; details are given in Table 3.

Over the past years, sequencing has become a feasible alternative to PCR-SSCP as costs have been reduced significantly. Nevertheless, this technique has often only been used for characterization of already known variants at the protein level (Ibeagha-Awemu et al., 2007; Lühken et al., 2009) or DNA level (Jann et al., 2002).

Furthermore, studies dealing with the identification of new casein variants often considered only European taurine cattle breeds, even if different variants were described and characterized in *Bos indicus* (Mahé et al., 1999; Ceriotti et al., 2003; Chen et al., 2008; Lühken et al., 2009).

The aim of this study was the identification of new variants of the 4 caseins α_{S1}-CN, α_{S2}-CN, β-CN, and κ-CN at the DNA level. Therefore, we sequenced the open reading frame of the respective genes *CSN1S1*, *CSN1S2*, *CSN2*, and *CSN3* in samples from both taurine and indicine cattle breeds. Furthermore, allele frequencies were calculated and analyzed for all examined breeds.

MATERIALS AND METHODS

Animals and Samples

The casein genes of 356 unrelated individuals of 17 breeds from 7 countries spanning 3 continents were analyzed in this study. Thirteen of the 17 investigated breeds belonged to the European *Bos taurus* cattle species and 3 to the Indian *Bos indicus* group. The native Iranian breed Sarabi has been assigned to the Indian *Bos taurus* group (Nassiry et al., 2008). The breeds are summarized in Table 1. The DNA was extracted either from full blood or semen samples, applying a modified protocol according to Miller et al. (1988). The semen samples were subjected to an additional treatment with dithiothreitol to break the disulfide bonds of nucleoprotamines.

DNA Sequencing

For sequence analysis, the genomic GenBank (http://www.ncbi.nlm.nih.gov/genbank) sequence NC_007304.4 of bovine *CSN1S1*, *CSN2*, *CSN1S2*, and *CSN3* and the PRIMER 3 software (Rozen and Skaletsky, 2000) were used to generate primer pairs for PCR amplification. Primer pairs were created only for exons that represent the open reading frame of each gene. Each primer pair spanned up to 3 exons and the 5′ and 3′ flanking intronic sequences, resulting in a total of 42 PCR products with length from 434 to 888 bp (Supplemental Table 1, available online at http://www.journalofdairyscience.org/). The PCR amplifications with their respective primer pairs were performed in a 12-μL reaction volume including 20 ng of genomic DNA, 0.2 μ*M* concentration of each primer, 200 μ*M* deoxyribonucleotide triphosphate (dNTP), and 5 U of Taq DNA polymerase (Invitek GmbH, Berlin, Germany) in the reaction buffer supplied by the manufacturer. Cycling conditions using a MJ Research PTC-200 thermal cycler (Global Medical Instrumentation Inc., Ramsey, MN) were as follows: initial denaturation at 94°C for 5 min, 35 cycles each of 94°C for 30 s, 62°C for 60 s, 72°C for 90 s, and a final extension step at 72°C for 10 min. After purification of the PCR products with thermosensitive alkaline phosphatase (FastAP; Fer-