J. Dairy Sci. 99:1–12
http://dx.doi.org/10.3168/jds.2015-10652

Genetic variability of the equine casein genes

J. Brinkmann,* V. Jagannathan,†‡ C. Drögemüller,†‡ S. Rieder,‡§ T. Leeb,†‡ G. Thaller,* and J. Tetens*[1]
*Institute of Animal Breeding and Husbandry, Christian-Albrechts-University Kiel, D-24098 Kiel, Germany
†Institute of Genetics, University of Bern, CH-3001 Bern, Switzerland
‡Swiss Competence Center of Animal Breeding and Genetics, University of Bern, Bern University of Applied Sciences HAFL and Agroscope, CH-3001 Bern, Switzerland
§Agroscope, Swiss National Stud Farm, CH-1580 Avenches, Switzerland

ABSTRACT

The casein genes are known to be highly variable in typical dairy species, such as cattle and goat, but the knowledge about equine casein genes is limited. Nevertheless, mare milk production and consumption is gaining importance because of its high nutritive value, use in naturopathy, and hypoallergenic properties with respect to cow milk protein allergies. In the current study, the open reading frames of the 4 casein genes *CSN1S1* (α_{S1}-casein), *CSN2* (β-casein), *CSN1S2* (α_{S2}-casein), and *CSN3* (κ-casein) were resequenced in 253 horses of 14 breeds. The analysis revealed 21 nonsynonymous nucleotide exchanges, as well as 11 synonymous nucleotide exchanges, leading to a total of 31 putative protein isoforms predicted at the DNA level, 26 of which considered novel. Although the majority of the alleles need to be confirmed at the transcript and protein level, a preliminary nomenclature was established for the equine casein alleles.
Key words: mare milk, caseins, isoforms, genetic variability

INTRODUCTION

There is a longstanding tradition of mare milk consumption in countries of the central Asian steppes, such as Mongolia or Kazakhstan (Uniacke-Lowe et al., 2010). In several European countries, mainly Italy, Hungary, the Netherlands, and Germany, consumption of mare milk is gaining more importance and it has been estimated that approximately 1 million kilograms are produced in Europe (Fox and Uniacke, 2010). The exact global production is unknown, but it has been reported that 30 million people worldwide regularly consume horse milk (Martuzzi and Vaccari Simonini, 2010). The increasing interest in mare milk is to a large extent driven by reports on positive health effects. The milk of horses is, for example, tolerated by children suffering from cow milk protein allergies, which affects approximately 2% of infants nourished with milk replacements based on cow milk (Businco et al., 2000; Curadi et al., 2001).

Despite this importance, knowledge about equine milk proteins and especially their genetic variability is still very limited. The genetic diversity of milk protein genes in typical dairy species, such as cattle (Caroli et al., 2009) and goat (Selvaggi et al., 2014), has been considered in numerous studies and distinct genetic variants have been described at both the protein and the DNA level.

Due to their ability to form micelles, caseins are important for the supply of the neonate with calcium, phosphate, and AA (Lenasi et al., 2003; Uniacke-Lowe et al., 2010). Whereas the casein fraction in bovine milk accounts for about 75% of the whole milk protein (Martin and Grosclaude, 1993), caseins only make up half of the equine milk protein fraction (Malacarne et al., 2002). The equine casein fraction is divided in α_{S1}-CN (18%), β-CN (79%), α_{S2}-CN (1.5%), and κ-CN (1.5%; Malacarne et al., 2002; Miranda et al., 2004; Inglingstad et al., 2010). The genes encoding these proteins are located on equine chromosome 3 in a tightly linked 290-kb gene cluster. The order is *CSN1S1* (encoding α_{S1}-CN), *CSN2* (encoding β-CN), *CSN1S2* (encoding α_{S2}-CN), and *CSN3* (encoding κ-CN; Egito et al., 2002; Milenkovic et al., 2002; Lenasi et al., 2003, 2005; Miranda et al., 2004; Girardet et al., 2006; Miclo et al., 2007; Selvaggi et al., 2010). Current knowledge about the individual caseins and their genetic variability is summarized in Table 1. The aim of our study was to provide extended knowledge about the genetic variability of equine casein genes and to identify putative protein variants at the DNA level by analyzing and comparing different horse breeds.

MATERIALS AND METHODS

Animals and Samples

Genomic DNA was extracted from hair samples of 198 horses using a modified version of the Miller et al.

Received November 17, 2015.
Accepted March 9, 2016.
[1]Corresponding author: jtetens@tierzucht.uni-kiel.de

Table 1. Current knowledge about equine casein genes and their genetic variability

Casein	Gene symbol	Length of the open reading frame	Location EquCab2.0 Chrom. 3 NC_009146.2	Comment
α_{S1}	*CSN1S1*	612 bp	64,954,285... 64,970,471	Full-length cDNA sequence: Lenasi et al. (2003). Two variants due to exon skipping (Miranda et al., 2004). Genomic (NC_009146.2) and mRNA (NM_ 001081883.1) reference sequences differ (c.406C>A).
β	*CSN2*	699 bp	64,938,110... 64,946,489	Full-length cDNA sequence: Lenasi et al. (2003) and Girardet et al. (2006). Two smaller variants reported (Miranda et al., 2004).
α_{S2}	*CSN1S2*	642 bp	64,795,317... 64,811,812	First described in 2000 (Egito et al., 2001; Egito et al., 2002; Miranda et al., 2004; Ochirkhuyag et al., 2000). Two major variants (*CSN1S2*A*, *CSN1S2*B*) due to a genomic 1.3-kb deletion covering 2 coding exons (Brinkmann et al., 2015).
κ	*CSN3*	555 bp	64,683,856... 64,694,148	First described in 2001 (Iametti et al., 2001; Miranda et al., 2004). Full-length cDNA sequence: Lenasi et al. (2003). Two putative variants described at the DNA level by Hobor et al. (2006; 2008): Ile383Lys and Thr173Ala.

(1988) protocol. The horses belonged to 8 breeds that are currently used for mare milk production in Germany. The samples were taken in 2012 on 10 different farms across northern, southern, and eastern Germany. In most cases, actual dairy mares were sampled, but in some instances male relatives of mares that had left the herd or natural service stallions were included in the study. The animals were selected to be as unrelated as possible in a way that only one animal was chosen from groups of relatives such as sisters or dam and offspring. Additionally, individual whole-genome sequence variant calling data of a total 55 horses from 10 different breeds available from other studies were incorporated in the analyses. This was an independent sample and the horses were not related to animals from the initial sample. The animals were sequenced to a mean coverage of 15.8×; bioinformatic details were reported previously (Drögemüller et al., 2014; Frischknecht et al., 2014). In total, 253 horses belonging to 14 different breeds or populations were analyzed (Table 2).

DNA Sequencing

A total of 37 primer pairs (Supplemental Table S1; http://dx.doi.org/10.3168/jds.2015-10652) were designed to amplify all exons contributing to the open reading frames of the genes and adjacent intronic regions. Primer 3 software (Rozen and Skaletsky, 2000)

Table 2. Animals used in the sequence analysis of the equine casein genes (n = 253)

Breed	Acronym	SEQ[1]	WGS[2]	Total
Akhal-Teke	AK	—	1	1
Dairy Crossbreed[3]	CB	21	—	21
Argentine Criollo Horse	CR	27	—	27
Fjord Horse	FJ	3	—	3
Franches-Montagnes	FM	—	29	29
Haflinger	HF	39	1	40
Icelandic Horse	IC	25	1	26
Dutch Warmblood (KWPN)	WB_{NL}	—	1	1
Quarter Horse	QH	22	3	25
Russian Heavy Draft	RU	24	—	24
Shetland pony	SP	—	2	2
Swiss Warmblood	WB_{CH}	—	3	3
UK Warmblood	WB_{UK}	—	2	2
German Warmblood	WB_{D}	37	12	49
Total		198	55	253

[1]Data from DNA Sanger sequencing.

[2]Data from whole-genome sequencing.

[3]Breeds that are crossed include German Riding Pony, Haflinger Horse, Connemara Pony, New Forest Pony, and further pony breeds to achieve a preferable high milk yield.