

RESEARCH ARTICLE

# Characterization of an Equine $\alpha$ -S<sub>2</sub>-Casein Variant Due to a 1.3 kb Deletion Spanning Two Coding Exons

Julia Brinkmann<sup>1</sup>, Tomas Koudelka<sup>2</sup>, Julia K. Keppler<sup>3</sup>, Andreas Tholey<sup>2</sup>, Karin Schwarz<sup>3</sup>, Georg Thaller<sup>1</sup>, Jens Tetens<sup>1</sup>\*

**1** Institute of Animal Breeding and Husbandry, Christian-Albrechts-Universität Kiel, Kiel, Germany, **2** Systematic Proteomics & Bioanalytics, Institute for Experimental Medicine, Christian-Albrechts-Universität Kiel, Kiel, Germany, **3** Institute of Human Nutrition and Food Science, Division of Food Technology, Christian-Albrechts-Universität Kiel, Kiel, Germany

\* [jtetens@tierzucht.uni-kiel.de](mailto:jtetens@tierzucht.uni-kiel.de)



OPEN ACCESS

**Citation:** Brinkmann J, Koudelka T, Keppler JK, Tholey A, Schwarz K, Thaller G, et al. (2015) Characterization of an Equine  $\alpha$ -S<sub>2</sub>-Casein Variant Due to a 1.3 kb Deletion Spanning Two Coding Exons. PLoS ONE 10(10): e0139700. doi:10.1371/journal.pone.0139700

**Editor:** Benfang Lei, Montana State University, UNITED STATES

**Received:** August 6, 2015

**Accepted:** September 16, 2015

**Published:** October 7, 2015

**Copyright:** © 2015 Brinkmann et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Sequencing data has been submitted to GenBank under submission number 1846887: proteome dataset (PXD002834); transcript sequences (KT368778 and KT368779).

**Funding:** This project was funded by the German Federal Ministry of Education and Research (Bonn, Germany) within the competence network "Food Chain Plus" (FoCus, grant no. 0315539A). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The production and consumption of mare's milk in Europe has gained importance, mainly based on positive health effects and a lower allergenic potential as compared to cows' milk. The allergenicity of milk is to a certain extent affected by different genetic variants. In classical dairy species, much research has been conducted into the genetic variability of milk proteins, but the knowledge in horses is scarce. Here, we characterize two major forms of equine  $\alpha$ <sub>S2</sub>-casein arising from genomic 1.3 kb in-frame deletion involving two coding exons, one of which represents an equid specific duplication. Findings at the DNA-level have been verified by cDNA sequencing from horse milk of mares with different genotypes. At the protein-level, we were able to show by SDS-page and in-gel digestion with subsequent LC-MS analysis that both proteins are actually expressed. The comparison with published sequences of other equids revealed that the deletion has probably occurred before the ancestor of present-day asses and zebras diverged from the horse lineage.

## Introduction

Horses are of minor importance in global dairy production, but mare's milk has traditionally been consumed in Mongolia, Kazakhstan, Kyrgyzstan or Tajikistan [1]. The global amount of production is not exactly known, but it has been estimated that approximately 30 million people worldwide are regularly consuming mare's milk [2]. Also in Europe, especially in Italy, Hungary, The Netherlands and Germany, the production and consumption of mare's milk have gained more and more importance; roughly 1 million kg of mare's milk are produced in Europe [3]. This increased interest is mainly based on positive health effects. The milk of horses and donkeys is e.g. tolerated by the majority of children suffering from cow's milk protein allergy, a condition that affects approximately 2% of infants when nourished with milk replacements on cow milk basis [4,5]. Moreover, positive effects of mare's milk consumption on diseases like atopic dermatitis [6], Morbus Crohn [7] or cardiovascular diseases [8] have been reported.

**Competing Interests:** The authors have declared that no competing interests exist.

The composition of equine milk, and especially the milk protein fraction, is very different from that of cows' milk. It is lower in fat and protein, but has a high lactose content similar to what is found in human milk [9,1]. While in cattle the casein fraction accounts for the major part of the total milk protein, the casein to whey ratio in horses is around 1.1:1, which more closely resembles human milk [9,1]. In fact, it has been reported that the balance between caseins and whey proteins is a major determinant of cow's milk allergenicity [10] possibly giving an explanation for the low allergenic potential of horse milk. However, there is also strong evidence that genetic milk protein variants affect the allergenicity of milk protein based on the presence or absence of particular epitopes [11,12]. While there has been intense research into the genetic variability of milk proteins in ruminants and especially in dairy cows [13], the knowledge about equine milk protein variation is scarce, especially for the caseins. However, in the donkey different variants of  $\alpha$ <sub>S2</sub>-casein have been described, also involving a large deletion exons 4–6 [14].

In the present study, we characterized a major protein variant arising from a 1.3 kb in-frame-deletion covering two exons and proved the protein by means of LC-MS based analytics at the protein level.

## Material and Methods

### Animals and Samples

Genomic DNA was extracted from hair samples of 193 domestic horses from 8 different breeds that are actually used for mare's milk production in Germany applying a modified Miller protocol [15]. The animals were selected to be as unrelated as possible. Hair samples were obtained from 14 different private studs with permission of and in cooperation with the owners by pulling out several hairs from the mane or the tail. Furthermore, individual milk samples were collected from four Haflinger mares with known genotype. These samples were taken by the owners during routine milking of the mares. In concordance with German Animal welfare legislation, these sampling procedures do not require a permission or approval.

### DNA Sequencing

Primer pairs were designed to amplify the coding exons contributing of equine *CSNIS2* and adjacent intronic regions using the Primer 3 software [16] based on the genomic reference sequence of the casein gene *CSNIS2* (Acc. No NC\_009146.2). A further Primer pair (Forward: 5'-GGAAAAGATTTGTGAGCCATTTG-3', Reverse: 5'-GCTGGATAATTGCTCAACTCA-3') was designed to specifically amplify the entire region of *CSNIS2* encompassing the deletion. PCR amplification and DNA sequencing were done as previously described [17]. The obtained sequences were analyzed and compared with the genomic reference sequence (Acc. No. NC\_009168.2) using the software Sequencher 4.9 (Gene Codes Corp., Ann Arbor, MI).

### RNA Isolation from Milk Samples and cDNA Synthesis

Individual milk samples were obtained from four mares with known deletion genotype. An aliquot of 40 ml was centrifuged at 6,000 g for 10 minutes. The supernatant including the milk fat layer was discarded and remaining milk fat was thoroughly removed with alcohol wipes. The cell pellet was washed three times with 1x phosphate buffered saline. Cells were homogenized using QIAshredder columns (Qiagen, Hilden, Germany) and total RNA was isolated using the Qiagen RNeasyMini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The isolated RNA was transcribed into cDNA using SuperScript<sup>®</sup> III First-Strand Synthesis SuperMix kit (Invitrogen) with oligo-dT primers. PCR amplification and sequencing

was done with primers located in the untranslated regions (Forward: 5'-TGCCTGCACTTTC TTGTCTTCCA-3', Reverse: 5'-TGCACAGTCTTTCATTTGGCTTGA-3').

## Protein and Peptide Analysis

Individual milk samples of two mares with known genotype were used for protein analysis. These samples were dialyzed to remove lactose, subsequently freeze dried and stored at -18°C for 4 months. Lyophilized milk powder was dissolved in Laemmli buffer (1x) at a concentration of 2 mg/mL and 10 and 20  $\mu$ g of crude protein was loaded onto a 12% SDS-PAGE gel (150V for 85 min). Gel bands were destained, reduced and alkylated and then subsequently in gel-digested overnight with trypsin (60 ng) using standard protocols. Peptides were extracted from the gel, dried down using vacuum centrifugation and resuspended in 3% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA) before being analyzed by LC-MS.

Nano-UHPLC-MS was performed on an UltiMate 3000 RSL Nano/Cap System (Thermo Fisher Scientific, Bremen, Germany) coupled online to an Orbitrap QExactive (Thermo Fisher Scientific). Samples were desalted for 4 minutes (Acclaim PepMap100 C-18, 300  $\mu$ m I.D. x 5 mm, 5  $\mu$ m, 100 Å, Thermo Fisher Scientific) at a flow rate of 30  $\mu$ L/min using 3% ACN, 0.1% TFA. An Acclaim PepMap100 C-18 column (75  $\mu$ m I.D. x 500 mm, 2  $\mu$ m, 100 Å, Thermo Fisher Scientific) was used for analytical separation at a flow rate of 300 nL/min using binary gradients of buffers A (0.05% FA) and B (80% ACN and 0.04% FA). The elution used gradient steps of 5–50% B (4–30 min) and 50–90% B (30–35 min) followed by an isocratic wash (90% B, 35–45 min) and column re-equilibration (5% B, 45–60 min) steps.

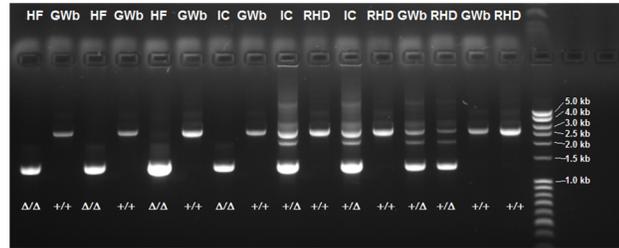
MS scans were acquired in the mass range of 300 to 2,000 m/z at a resolution of 70,000. The ten most intense signals were subjected to HCD fragmentation using a dynamic exclusion of 15 s. MS/MS parameters—minimum signal intensity: 1000, isolation width: 3.0 Da, charge state:  $\geq 2$ , HCD resolution: 15,000, Normalized collision energy of 25. Lock mass (445.120025) was used for data acquired in MS mode.

HCD spectra were searched using Proteome Discoverer 1.4 (1.4.0.288, Thermo Fisher Scientific) with the Sequest-HT search algorithm against the complete reviewed and unreviewed *Equus caballus* database (28,188 sequences, downloaded 2015.07.16) with common contaminants (<ftp://ftp.thegpm.org/fasta/cRAP/>) appended. The following database search settings were used: MS tolerance;  $\pm 10$  ppm, MS2 Tolerance; 0.02 Da, enzyme specificity; trypsin with up to three missed cleavages allowed. Carbamidomethylation on cysteine residues was set as a fixed modification while, oxidation on methionine, and phosphorylation on serine and threonine residues was set as a variable modification. Only peptides which were identified with medium confidence (FDR <5%) were included. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://www.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD002834.

## Results and Discussion

### DNA Sequencing and Mutation Screening

The current annotation of the equine *CSNIS2* gene (GeneID 100327035) is based on the mRNA reference sequence NM\_001170767.2 containing 15 coding exons with an open reading frame of 645 bp. In an attempt to resequence the open reading frame using exon flanking primer pairs, we recognized that the PCR reactions for exons 8 and 9 consistently failed in particular horses. In order to unravel the possible cause for this phenomenon, we amplified a 2.6 kb fragment spanning the entire region. While the expected product was obtained from samples that had been successfully amplified before, the product obtained from initially unsuccessful samples was found to be approximately 1.3 kb shorter (Fig 1). Subsequent Sanger

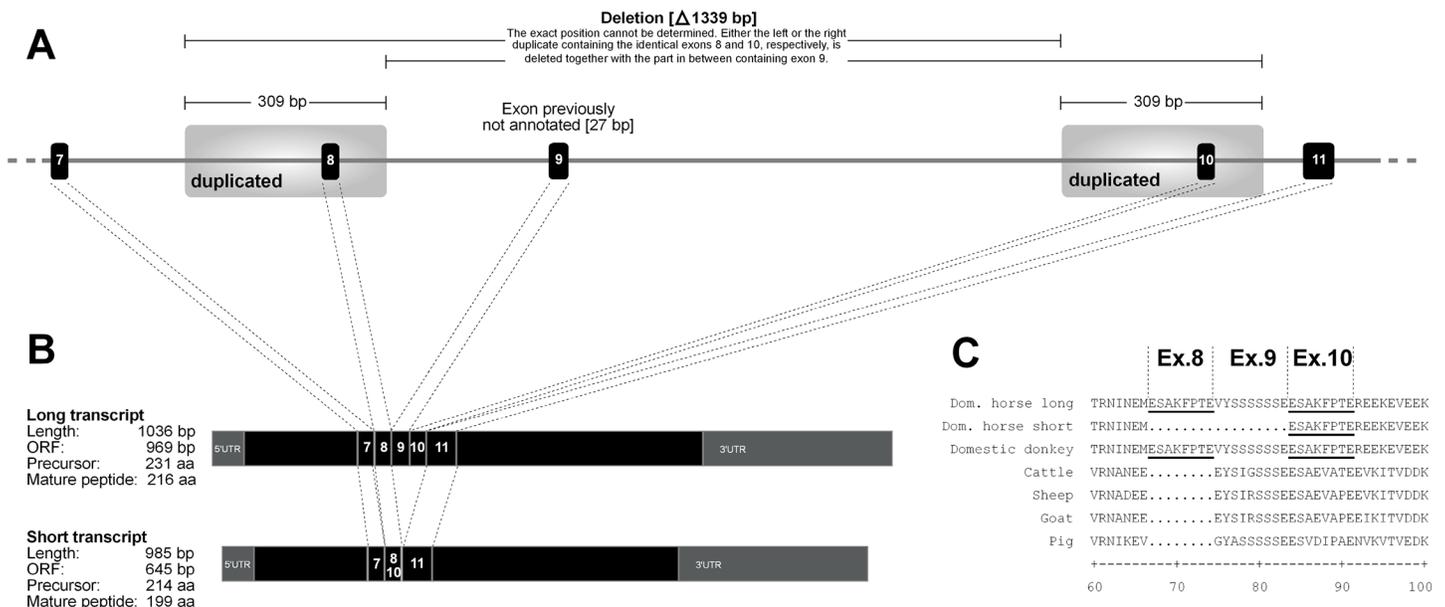


**Fig 1. Agarose gel electrophoresis of PCR products spanning the 1.3 kb deletion.** The upper visible band corresponds to the long variant (denoted as +), the lower one to the short variant containing the deletion (denoted as Δ). Only in heterozygotes, a third band with a size of approximately 2.1 kb is visible, which is possibly arising from asymmetric hybridization of the alleles due to the presence of a duplication. The breeds of the corresponding samples are given above the lanes (RHD = Russian Heavy Draft, GWb = German Warmblood, IC = Icelandic horse, HF = Haflinger).

doi:10.1371/journal.pone.0139700.g001

sequencing of the products revealed the presence of a 1,339 bp deletion in the short variant (Fig 2A), while the long product was found to completely correspond to the genomic reference sequence (NC\_009146.2). Analysis of this sequence revealed the presence of a 309 bp duplication of the region encompassing exon 8 of the gene (Fig 2A). Because this duplication is located exactly at the boundary of the deletion, the exact position cannot be determined, i.e. it cannot be ruled out, whether the upstream or downstream duplicate is involved in the deletion.

A total of 193 horses belonging to 8 breeds (Table 1) were tested for the presence of the deletion by PCR and subsequent agarose gel electrophoresis (Fig 1). The deletion was found to be present in all analyzed breeds; the highest frequencies of 0.36 and 0.25 were observed in Haflinger and Icelandic horses, respectively. Notably, these breeds are common in mare's milk production, especially the Haflinger breed is widely used. This might possibly indicate an effect of the mutation or a certain casein-haplotype on milk yield as this is e.g. the case in cattle [18,19].



**Fig 2. Structure of the long and short equine  $\alpha$ <sub>s2</sub>-casein variants.** A. Genomic organization of the respective gene segment. Grey shading indicates the equid specific 309 bp duplication comprising coding exons 8 and 10, respectively. The 1.3 kb in-frame-deletion is indicated above the figure. B. Structures the resulting transcript variants. C. Protein alignment of available ungulate  $\alpha$ <sub>s2</sub>-casein protein sequences.

doi:10.1371/journal.pone.0139700.g002

**Table 1. Frequencies of the 1.3 kb deletion in different horse breeds.**

| Breed                  | n   | ins/ins <sup>a</sup> | ins/del <sup>a</sup> | del/del <sup>a</sup> | Frequency of deletion |
|------------------------|-----|----------------------|----------------------|----------------------|-----------------------|
| Crossbred <sup>b</sup> | 21  | 14                   | 6                    | 1                    | 0.19                  |
| Criollo                | 27  | 24                   | 3                    | 0                    | 0.06                  |
| Fjord Horse            | 3   | 2                    | 1                    | 0                    | - <sup>c</sup>        |
| Haflinger Horse        | 39  | 17                   | 16                   | 6                    | 0.36                  |
| Icelandic Horse        | 24  | 13                   | 10                   | 1                    | 0.25                  |
| Quarter Horse          | 20  | 16                   | 3                    | 1                    | 0.13                  |
| Russian Heavy Draft    | 24  | 20                   | 3                    | 1                    | 0.10                  |
| German Warmblood       | 35  | 33                   | 2                    | 0                    | 0.03                  |
| TOTAL                  | 193 | 139                  | 44                   | 10                   | 0.17                  |

<sup>a</sup> ins = long variant corresponding to the genomic reference NC\_009146.2; del = short variant encompassing the 1,339 bp deletion spanning two coding exons.

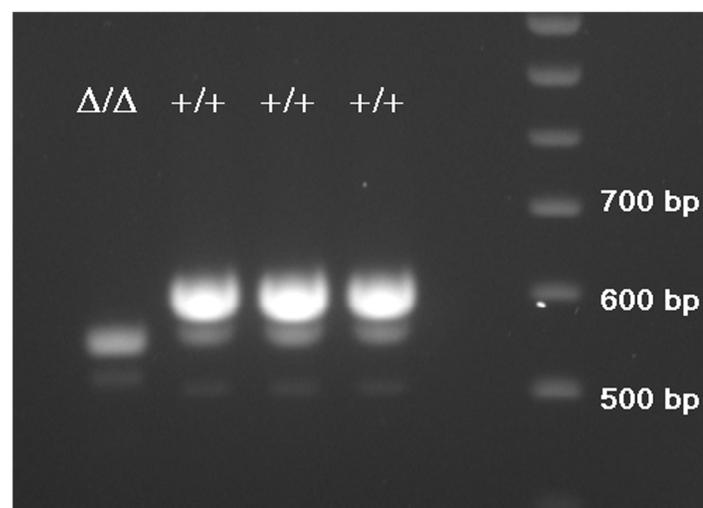
<sup>b</sup> A synthetic cross involving German Riding Pony, Haflinger Horse, Connemara Pony and New Forrest Pony; bred for milk yield.

<sup>c</sup> The frequency in Fjord horses is not reported with respect to the small sample size, but the breed is included in the total values.

doi:10.1371/journal.pone.0139700.t001

### Analysis of Transcripts

The duplicated region within the 1.3 kb deletion contains a coding exon with a length of 24 bp. The two copies were found to be completely identical including intact splice sites. However, only one of the identical exons is present in the current RefSeq transcript NM\_001170767.2. Thus, it was unclear which exons are transcribed and whether both variants are transcribed at all. Therefore, we purified total RNA from the skimmed milk of four mares, three of them being homozygote for the long and one for the short variant, respectively. After reverse transcription, the *CSN1S2* transcripts were amplified using primers located in the untranslated regions. Agarose gel electrophoresis of the PCR products revealed a difference of approximately 50 bp between the alternatively homozygote animals (Fig 3) showing that both, a long and a short transcript, were actually expressed. Subsequently, the open reading frames of both transcripts were sequenced. The difference was found to be due to a 51 bp in-frame insertion/



**Fig 3. Agarose gel electrophoresis of equine *CSN1S2* cDNA.** The RNA was isolated from the milk of a mare being homozygous for the deletion ( $\Delta/\Delta$ ), and three mares homozygous for the long variant ( $+/+$ ).

doi:10.1371/journal.pone.0139700.g003

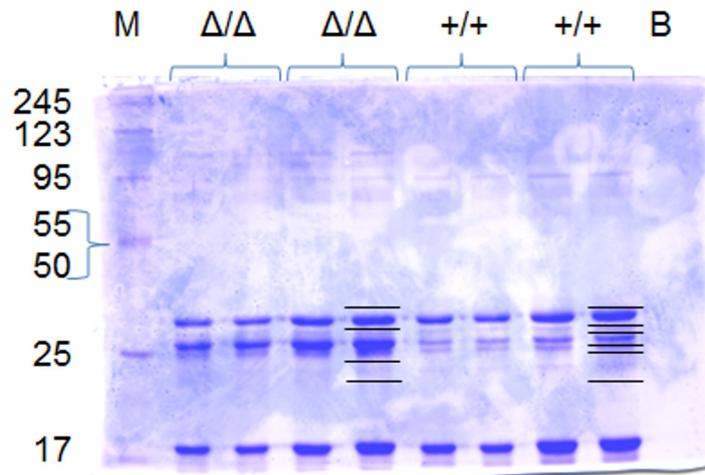
deletion after exon eight encompassing the duplicated exon as well as a previously not annotated exon that perfectly aligns within the 1.3 kb deletion (Fig 2A and 2B). Exon numbering was consequently adapted counting the newly annotated exon as exon 9 and the duplicate of exon 8 as exon 10. Although the genome assembly (EqCab2.0) comprises the long variant, the RefSeq transcript NM\_001170767.2 used for annotation represents the short variant. A BLAST search showed that both transcripts had been reported before (GenBank KP658381.1 and KP658382.1) and both variant transcripts have recently been added to the unreviewed UniProt database (Acc. No. A0A0C5DH76 and D2KAS0), but no further information or publication is available. The transcript sequences from the current study are available under the accession numbers KT368778 and KT368779.

## Comparative Analysis

Translations of the long and short transcript, respectively, were aligned to available protein sequences of domestic donkey (Acc. No. CAV00691.1 [20]), cattle (Acc. No. NP776953.1), sheep (Acc. No. NP\_001009363.1), goat (Acc. No. NP\_001272514.1) and pig (Acc. No. NP\_001004030.1). From Fig 2C it can be seen that the duplication of exon 8/10 is unique to donkey and horse, while the equine exon 9 appears to have a homologous sequence in other ungulates. A BLAST search against the genome assembly of Przewalski's horse (Burgud assembly, CGF0000696695.1), a species that separated from the ancestral population of domesticated horses 38–72 kyr BP ago [21,22], also revealed the presence of the long variant including the duplicated exon. Furthermore, the small read archive data sets of a Middle Pleistocene horse, the “Thistle Creek horse” sequenced by Orlando and colleagues [22] (BioProject Accession PRJNA205517), as well as of several asses and zebras sequenced by Jónsson and colleagues ([23], BioProject Accession PRJEB7446) were checked for reads either falling into the deleted region (indicating the long variant) or being split at the boundaries (indicating the short variant). Only a single read almost perfectly aligning within the deletion was found in the Middle Pleistocene horse, which might point to the presence of the long variant. In the Somalian wild ass (*E. asinus somalicus*), we only identified the long variant, while both alleles were found in the Tibetan Kiang (*E. kiang*). The sequenced Onager (*E. hemionus onager*) was found to be homozygous for the deletion. These findings indicate that the duplication event giving rise to an additional coding exon as well the deletion might be specific to equids and must both have occurred before the ancestor of present-day asses and zebras dispersed into the Old World 2.1–3.4 Mya [23]. However, all analyzed zebras [23] (Hartmanns Mountain zebra, *E. zebra hartmannae*; Grevy zebra, *E. grevyi*; Böhm's plains zebra, *E. quagga boehmi* as well as the extinct Quagga, *E. q. quagga*) were found to be homozygous for the complete deletion. It seems also possible that the deletion initially occurred in horses and represents the result of a gene flow between horses and ass species. It has been shown that this has played a significant role in equid evolution [23].

## Protein and Peptide Analysis

Lyophilized milk powder from two mares being homozygote for the long and short variant, respectively, were analyzed using SDS-PAGE resulting in different patterns of distinct bands (Fig 4). In-gel trypsin digestion and analysis by LC-MS revealed the presence of unique peptides only for the long form of  $\alpha$ <sub>s2</sub>-casein (A0A0C5DH76) in milk of the animal homozygous for the insertion, i.e., peptides FPTEVYSSSSSEESAK, FPTEVYSSSSSEESAKFPTER, FPTEVYSSSSSEESAKFPTEREEK and NINEMESAKFPTEVYSSSSSEESAK. Interestingly, these peptides were identified in both phosphorylated (singly phosphorylated at different residues) and non-phosphorylated forms. Evidence for multiple phosphorylations on these peptides was

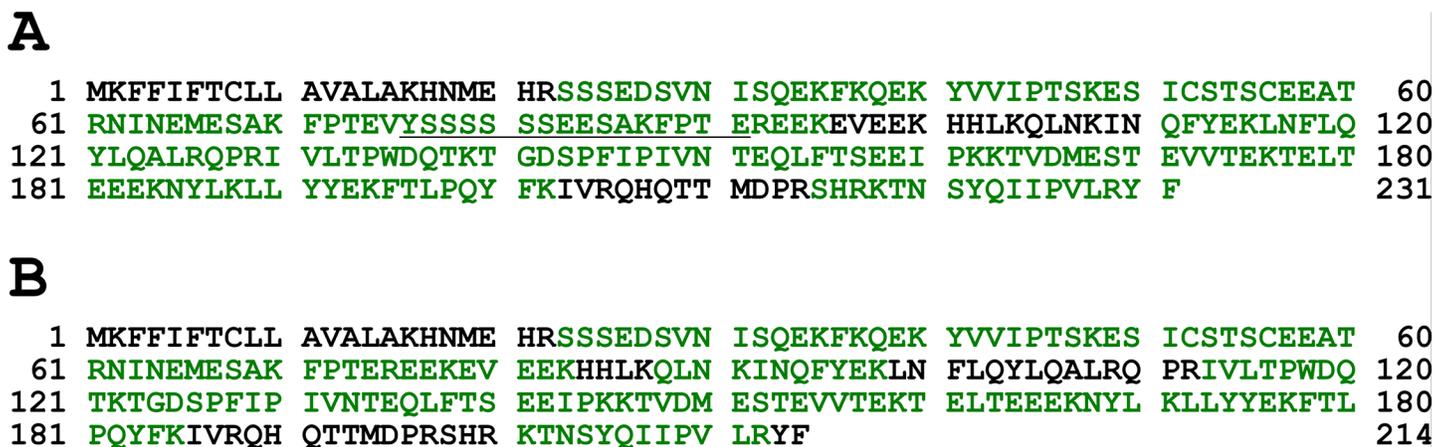


**Fig 4. SDS-PAGE of crude mare's milk being homozygous for the deletion ( $\Delta/\Delta$ ) and homozygous for the long variant ( $+/+$ ).** Ten and 20  $\mu$ g of crude milk from both mares was loaded in duplicate. Excised bands that were in-gel digested with trypsin and later analyzed by LC-MS are indicated. M (Roti<sup>®</sup>-Mark Bicolor protein standard); B (Blank, Laemmli buffer).

doi:10.1371/journal.pone.0139700.g004

also observed. Unique peptides for the short form of  $\alpha$ <sub>S2</sub>-casein (D2KAS0) were only identified in milk from the mare homozygous for the deletion, i.e., NINEMESAKFPTER, NINEMESAKFPTEREEK, NINEMESAKFPTEREEKEVEEK (Fig 5, Table 2). As commonly observed in MS based protein analytics, a 100% sequence coverage was not reached; however, the proteotypic peptides identified allowed clearly to distinguish the two equine  $\alpha$ <sub>S2</sub>-casein variants. Therefore, it can be concluded that both protein variants differing in length by 17 aa are expressed. The comparative analysis has shown that the long variant is probably the equid specific ancestral variant, but the deletion also seems to have been present before zebras and asses diverged from horses. Thus, we propose to term the long variant CSNIS2\* A and the short variant CSNIS2\* B.

Generally, the milk proteome is very complex both due to the presence of genetic variants and posttranslational modifications [24]. Recent proteomic studies [25,26] have demonstrated



**Fig 5. Combined sequence coverage (all bands excised) of  $\alpha$ <sub>S2</sub>-casein in mare with  $+/+$  genotype (A) and  $\Delta/\Delta$  genotype (B).** From the  $+/+$  genotype only unique peptides were identified for the longer  $\alpha$ <sub>S2</sub>-casein form (Accession no. A0A0C5DH76, 231 AAs). Sequence which is unique to A0A0C5DH76 is underlined. For the  $\Delta/\Delta$  genotype only unique peptides were identified for the shorter alphaS2-casein form (Accession no. D2KAS0, 214 AAs). Sections in green represents parts of the protein which were identified by the sequest-HT algorithm.

doi:10.1371/journal.pone.0139700.g005

**Table 2. Bands excised from a mare with +/- genotype (upper part) and  $\Delta/\Delta$  genotype (lower part), respectively, and major milk proteins identified are shown (only those with > 20 PSMs).** AlphaS<sub>2</sub>-casein with accession no. A0A0C5DH76 is the long form (231 AA) while alphaS<sub>2</sub>-casein with accession no. D2KAS0 is the shorter form of alphaS<sub>2</sub>-casein (214 AAs). Unique peptides were only identified for the long form of alphaS<sub>2</sub>-casein (A0A0C5DH76) in the +/--mare, while unique peptides for the short form of alphaS<sub>2</sub>-casein (D2KAS0) were identified only in milk from the  $\Delta/\Delta$ -mare.

| Band  | Accession  | Description  | Coverage (%) | #Unique Peptides | #Peptides | #PSMs |
|---|------------|--|--------------|------------------|-----------|-------|
| <b>Mare homozygous for long variant (+/+)</b>                         |            |  |              |                  |           |       |
| 1   | Q9GKK3     | $\beta$ -casein  | 35.68        | 9                | 9         | 87    |
|   | D2KAS0     | $\alpha$ <sub>S2</sub> -casein                               | 49.53        | 0                | 13        | 35    |
|   | A0A0C5DH76 | CSN1S2 protein   | 45.89        | 0                | 13        | 35    |
| 2   | Q9GKK3     | $\beta$ -casein  | 32.78        | 8                | 8         | 45    |
|   | A0A0C5DH76 | $\alpha$ <sub>S2</sub> -casein                               | 72.73        | 2                | 17        | 45    |
|   | D2KAS0     | $\alpha$ <sub>S2</sub> -casein                               | 64.02        | 0                | 15        | 43    |
|   | P82187     | $\kappa$ -casein   | 23.78        | 7                | 9         | 40    |
|   | Q95KZ7     | $\alpha$ <sub>S1</sub> -casein                               | 41.35        | 1                | 12        | 31    |
| 3   | Q9GKK3     | $\beta$ -casein  | 35.68        | 11               | 11        | 138   |
|   | Q95KZ7     | $\alpha$ <sub>S1</sub> -casein                               | 45.67        | 2                | 15        | 47    |
|   | A0A0C5DH76 | $\alpha$ <sub>S2</sub> -casein                               | 58.01        | 1                | 14        | 31    |
|   | D2KAS0     | $\alpha$ <sub>S2</sub> -casein                               | 54.67        | 0                | 13        | 30    |
|   | P82187     | $\kappa$ -casein   | 23.78        | 4                | 6         | 21    |
| 4   | Q95KZ7     | $\alpha$ <sub>S1</sub> -casein                               | 45.67        | 2                | 15        | 67    |
|   | Q9GKK3     | $\beta$ -casein  | 35.68        | 10               | 10        | 35    |
|   | A0A0C5DH76 | $\alpha$ <sub>S2</sub> -casein                               | 60.17        | 2                | 16        | 31    |
|   | D2KAS0     | $\alpha$ <sub>S2</sub> -casein                               | 54.67        | 0                | 14        | 29    |
| 5   | A0A0A1E470 | Immunoglobulin lambda light chain variable region (fragment) | 46.46        | 2                | 12        | 62    |
|   | A0A0C5DH76 | $\alpha$ <sub>S2</sub> -casein                               | 71.86        | 3                | 20        | 59    |
|   | D2KAS0     | $\alpha$ <sub>S2</sub> -casein                               | 63.08        | 0                | 17        | 52    |
| <b>Mare homozygous for short variant (<math>\Delta/\Delta</math>)</b> |            |  |              |                  |           |       |
| 1   | Q9GKK3     | $\beta$ -casein  | 35.68        | 10               | 10        | 80    |
| 2   | Q9GKK3     | $\beta$ -casein  | 35.68        | 12               | 12        | 139   |
|   | Q95KZ7     | $\alpha$ <sub>S1</sub> -casein                               | 45.67        | 2                | 16        | 75    |
|   | Q8SPR1     | $\alpha$ <sub>S1</sub> -casein                               | 63.68        | 0                | 18        | 65    |
|   | D2KAS0     | $\alpha$ <sub>S2</sub> -casein                               | 73.36        | 2                | 20        | 60    |
|   | A0A0C5DH76 | $\alpha$ <sub>S2</sub> -casein                               | 58.44        | 0                | 18        | 58    |
|   | P82187     | $\kappa$ -casein   | 29.73        | 3                | 6         | 34    |
| 3   | D2KAS0     | $\alpha$ <sub>S2</sub> -casein                               | 73.36        | 1                | 24        | 82    |
|   | Q9GKK3     | $\beta$ -casein  | 35.68        | 11               | 11        | 80    |
|   | A0A0C5DH76 | $\alpha$ <sub>S2</sub> -casein                               | 67.97        | 0                | 23        | 78    |
|   | Q95KZ7     | $\alpha$ <sub>S1</sub> -casein                               | 31.25        | 2                | 9         | 24    |

doi:10.1371/journal.pone.0139700.t002

considerable microheterogeneity for equine caseins, especially  $\kappa$ -casein [25]. However, these studies did not report any findings regarding  $\alpha$ <sub>S2</sub>-casein, probably due to its very low concentration in horse milk [1]. However, Ochirkhuyag et al. [27] reported the presence of two distinct bands for this protein. In our study, both genetic variants as well as differently phosphorylated peptides have been detected for  $\alpha$ <sub>S2</sub>-casein.

### Conclusion

Within the current study, we have characterized two major variants of equine  $\alpha$ <sub>S2</sub>-casein, which we named CSN1S2\**A* and CSN1S2\**B*. The variation is due to a 1.3 kb in-frame deletion involving two coding exons corresponding to 17 amino acid residues. One of those exons has

arisen from a duplication that is probably specific to the equid lineage. We verified both genomic variants at the transcript as well as the protein level and were able to demonstrate that these variants are also segregating in asses, meaning that they are likely to have occurred before the first ancestor of present-day asses and zebras dispersed into the Old World 2.1–3.4 Mya.

## Acknowledgments

This project was funded by the German Federal Ministry of Education and Research (Bonn, Germany) within the competence network “Food Chain Plus” (FoCUS, grant no. 0315539A). The authors would like to thank all the mare’s milk producers for providing hair samples, Haflinger stud Seraphin for providing milk samples, Julia Tetens for her help with sample collection and Gabriele Ottzen-Schirakow for expert technical assistance.

## Author Contributions

Conceived and designed the experiments: JT GT AT KS. Performed the experiments: JB TK JKK. Analyzed the data: JB TK JT. Contributed reagents/materials/analysis tools: GT KS AT. Wrote the paper: JB JT TK JKK.

## References

1. Uniacke-Lowe T, Huppertz T, Fox PF (2010) Equine milk proteins: Chemistry, structure and nutritional significance. *International Dairy Journal* 20 (9): 609–629.
2. Martuzzi F, Vaccari Simonini F (2010) Advances on equine milk and derivatives for human consumption. Book of Abstracts of the 61st Annual Meeting of the European Association for Animal Production. Heraklion—Crete Island, Greece, 23–27 August 2010: Wageningen Academic Pub. p. 40.
3. Fox PF, Uniacke T (2010) Chemical and physico-chemical properties of equid milk. Book of Abstracts of the 61st Annual Meeting of the European Association for Animal Production. Heraklion—Crete Island, Greece, 23–27 August 2010: Wageningen Academic Pub. p. 39.
4. Curadi MC, Giampetro PG, Lucenti P, Orlandi M, editors (2001) Use of mare milk in pediatric allergology.
5. Businco L, Giampietro PG, Lucenti P, Lucaroni F, Pini C, Di Felice G et al. (2000) Allergenicity of mare’s milk in children with cow’s milk allergy. *Journal of Allergy and Clinical Immunology* 105 (5): 1031–1034. PMID: [10808187](#)
6. Foekel C, Schubert R, Kaatz M, Schmidt I, Bauer A, Hipler UC et al. (2009) Dietetic effects of oral intervention with mare’s milk on the Severity Scoring of Atopic Dermatitis, on faecal microbiota and on immunological parameters in patients with atopic dermatitis. *Int J Food Sci Nutr* 60 Suppl 7: 41–52. doi: [10.1080/09637480802249082](#) PMID: [19462320](#)
7. Schubert R, Kahle C, Kauf E, Hofmann J, Hubert I, Gruhn B et al. (2009) Dietetic efficacy of mare’s milk for patients with chronic inflammatory bowel diseases—clinical study. *Ernährung* 33 (7/8): 314–321.
8. Chen Y, Wang Z, Chen X, Liu Y, Zhang H, Sun T (2010) Identification of angiotensin I-converting enzyme inhibitory peptides from koumiss, a traditional fermented mare’s milk. *J. Dairy Sci.* 93 (3): 884–892. doi: [10.3168/jds.2009-2672](#) PMID: [20172208](#)
9. Malacarne M, Martuzzi F, Summer A, Mariani P (2002) Protein and fat composition of mare’s milk: some nutritional remarks with reference to human and cow’s milk. *International Dairy Journal* 12 (11): 869–877.
10. Lara-Villoslada F, Olivares M, Xaus J (2005) The balance between caseins and whey proteins in cow’s milk determines its allergenicity. *Journal of dairy science* 88 (5): 1654–1660. PMID: [15829656](#)
11. Lisson M, Lochnit G, Erhardt G (2013) Genetic variants of bovine beta- and kappa-casein result in different immunoglobulin E-binding epitopes after in vitro gastrointestinal digestion. *Journal of dairy science* 96 (9): 5532–5543. doi: [10.3168/jds.2013-6684](#) PMID: [23871370](#)
12. 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 (4): 1939–1954. doi: [10.3168/jds.2013-7355](#) PMID: [24485684](#)
13. Caroli AM, Chessa S, Erhardt GJ (2009) Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92 (11): 5335–5352. doi: [10.3168/jds.2009-2461](#) PMID: [19841193](#)

14. Saletti R, Muccilli V, Cunsolo V, Fontanini D, Capocchi A, Foti S (2012) MS-based characterization of  $\alpha$  (s2)-casein isoforms in donkey's milk. *J Mass Spectrom* 47 (9): 1150–1159. doi: [10.1002/jms.3031](https://doi.org/10.1002/jms.3031) PMID: [22972783](https://pubmed.ncbi.nlm.nih.gov/22972783/)
15. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16 (3): 1215. PMID: [3344216](https://pubmed.ncbi.nlm.nih.gov/3344216/)
16. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365–386. PMID: [10547847](https://pubmed.ncbi.nlm.nih.gov/10547847/)
17. Gallinat JL, Qanbari S, Drögemüller C, Pimentel ECG, Thaller G, Tetens J. (2013) DNA-based identification of novel bovine casein gene variants. *J. Dairy Sci.* 96 (1): 699–709. doi: [10.3168/jds.2012-5908](https://doi.org/10.3168/jds.2012-5908) PMID: [23102962](https://pubmed.ncbi.nlm.nih.gov/23102962/)
18. McLean DM, Graham ER, Ponzoni RW, McKenzie HA (1984) Effects of milk protein genetic variants on milk yield and composition. *Journal of Dairy Research* 51 (04): 531–546.
19. Ikonen T, Bovenhuis H, Ojala M, Ruottinen O, Georges M (2001) Associations between casein haplotypes and first lactation milk production traits in Finnish Ayrshire cows. *Journal of dairy science* 84 (2): 507–514. PMID: [11233036](https://pubmed.ncbi.nlm.nih.gov/11233036/)
20. Chianese L, Calabrese MG, Ferranti P, Mauriello R, Garro G, De Simone C et al. (2010) Proteomic characterization of donkey milk "caseome". *J Chromatogr A* 1217 (29): 4834–4840. doi: [10.1016/j.chroma.2010.05.017](https://doi.org/10.1016/j.chroma.2010.05.017) PMID: [20541767](https://pubmed.ncbi.nlm.nih.gov/20541767/)
21. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A et al. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U.S.A.* 111 (52): E5661–9. doi: [10.1073/pnas.1416991111](https://doi.org/10.1073/pnas.1416991111) PMID: [25512547](https://pubmed.ncbi.nlm.nih.gov/25512547/)
22. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M et al. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499 (7456): 74–78. doi: [10.1038/nature12323](https://doi.org/10.1038/nature12323) PMID: [23803765](https://pubmed.ncbi.nlm.nih.gov/23803765/)
23. Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M et al. (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the United States of America* 111 (52): 18655–18660. doi: [10.1073/pnas.1412627111](https://doi.org/10.1073/pnas.1412627111) PMID: [25453089](https://pubmed.ncbi.nlm.nih.gov/25453089/)
24. Matéos A, Girardet J-M, Mollé D, Corbier C, Gaillard J-L, Miclo L (2010) Identification of phosphorylation sites of equine  $\beta$ -casein isoforms. *Rapid Comm Mass Spect.* 24(11): 1533–1542.
25. Uniacke-Lowe T, Chevalier F, Hem S, Fox PF, Mulvihill DM (2013) Proteomic comparison of equine and bovine milks on renneting. *Journal of agricultural and food chemistry* 61 (11): 2839–2850. doi: [10.1021/jf3045846](https://doi.org/10.1021/jf3045846) PMID: [23414207](https://pubmed.ncbi.nlm.nih.gov/23414207/)
26. Hinz K, O'Connor PM, Huppertz T, Ross RP, Kelly AL (2012) Comparison of the principal proteins in bovine, caprine, buffalo, equine and camel milk. *The Journal of dairy research* 79 (2): 185–191. doi: [10.1017/S0022029912000015](https://doi.org/10.1017/S0022029912000015) PMID: [22365180](https://pubmed.ncbi.nlm.nih.gov/22365180/)
27. Ochirkhuyag B, Chobert J, Dalgalarrrondo M, Haertlé T (2000) Characterization of mare caseins. Identification of  $\alpha$ s1- and  $\alpha$ s2- caseins. *Lait* 80 (2): 223–235.