Identification of rare genetic variants of the α_S -caseins in milk from native Norwegian dairy breeds and comparison of protein composition with milk from high-yielding Norwegian Red cows

N. R. Roin,^{1*} L. B. Larsen,¹ I. Comi,² T. G. Devold,² T. I. Eliassen,² R. A. Inglingstad,² G. E. Vegarud,² and N. A. Poulsen¹

¹Department of Food Science, Aarhus University, Agro Food Park 48, DK-8200 Aarhus N, Denmark

²Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1433 Aas, Norway

ABSTRACT

Several factors influence the composition of milk. Among these, genetic variation within and between cattle breeds influences milk protein composition, protein heterogeneity, and their posttranslational modifications. Such variations may further influence technological properties, which are of importance for the utilization of milk into dairy products. Furthermore, these potential variations may also facilitate the production of differentiated products (e.g., related to specific breeds or specific genetic variants). The objective of this study was to investigate the genetic variation and relative protein composition of the major proteins in milk from 6 native Norwegian dairy breeds representing heterogeneity in geographical origin, using the modern Norwegian breed, Norwegian Red, as reference. In total, milk samples from 144 individual cows were collected and subjected to liquid chromatography-electrospray ionization/mass spectrometry-based proteomics for identification of genetic and posttranslational modification isoforms of the 4 caseins (α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN) and the 2 most abundant whey proteins (α -lactalbumin and β -lactoglobulin). Relative quantification of these proteins and their major isoforms, including phosphorylations of α_{S1} -CN and glycosylation of κ -CN, were determined based on UV absorbance. The presence and frequency of genetic variants of the breeds were found to be very diverse and it was possible to identify rare variants of the CN, which, to our knowledge, have not been identified in these breeds before. Thus, α_{S1} -CN variant D was identified in low frequency in 3 of the 6 native Norwegian breeds. In

general, α_{S1} -CN was found to be quite diverse between the native breeds, and the even less frequent A and C variants were furthermore detected in 1 and 5 of the native breeds, respectively. The α_{S1} -CN variant C was also identified in samples from the Norwegian Red cattle. The variant E of κ -CN was identified in 2 of the native Norwegian breeds. Another interesting finding was the identification of α_{S2} -CN variant D, which was found in relatively high frequencies in the native breeds. Diversity in more common protein genetic variants were furthermore observed in the protein profiles of the native breeds compared with milk from the high-yielding Norwegian Reds, probably reflecting the more diverse genetic background between the native breeds.

Key words: milk proteins, mass spectrometry, genetic variation, posttranslational modifications

INTRODUCTION

Conservation of native dairy breeds is important to minimize erosion of genetic diversity. The Nordic countries have committed to conserve these animal genetic resources, where many of the native breeds have been classified as endangered (Kierkegaard et al., 2020). Compared with a high number of studies focusing on genetic diversity within and among these breeds, only a limited number of studies have been dedicated to provide phenotypic characterization and even fewer have studied milk compositional traits (Kierkegaard et al., 2020), which are nevertheless important characteristics in relation to eventual exploitation that could contribute to their preservation.

The major proteins in bovine milk consist of the 4 caseins, α_{S1} -CN, α_{S2} -CN, β -CN, and κ -CN, together with the 2 most abundant whey proteins, α -LA and β -LG. Generally, the genes encoding the 6 major proteins are polymorphic to different degrees, and more than 47 different milk protein variants have been identified within the *Bos* genus (Caroli et al., 2009).

Received March 15, 2021.

Accepted August 23, 2021.

*Corresponding author: nrr@food.au.dk

Genetic polymorphisms within the coding regions for the bovine milk proteins originate from SNP or deletions/insertions of one or more nucleotides, resulting in synonymous or nonsynonymous mutations or even longer sequences of the coding regions being deleted or inserted (Caroli et al., 2009; Ketto et al., 2017). This results in a wide range of protein variants influencing the overall composition of milk, which can potentially affect both technological and nutritional properties of milk and milk-derived products (Hallén et al., 2008; Caroli et al., 2009). Total milk yield and especially cheese production are highly influenced by variations caused by genetic polymorphisms of the major milk proteins (Wedholm et al., 2006; Glantz et al., 2010; Frederiksen et al., 2011; Jensen et al., 2012b; Poulsen et al., 2013, 2017a; Nilsson et al., 2020).

Some of the native Nordic dairy breeds have been subject to profound studies (Kierkegaard et al., 2020). Using DNA extracts from blood samples, Lien et al. (1999) reported large frequency differences in casein genetic variants among 17 native Nordic dairy breeds and 5 modern breeds. Such variation may imprint in different characteristics of the milk and its suitability for utilization into specific dairy products.

The present investigation on genetic variants of the major milk proteins and the resulting protein profile of native Norwegian dairy breeds aims to provide knowledge of the differences between these breeds and compare them with a modern high-yielding dairy breed. A better understanding of the unique characteristics of the protein composition of native Norwegian breeds may provide a basis for potential conservation programs and may lead to better utilization and development of niche dairy products with specific properties. Moreover, this study might also shed light on how the breeding for specific traits resulting in today's modern breeds has resulted in a loss of heterogeneity among the most abundant milk proteins. Hence, this study's objective is to investigate the genetic variation among native and more common Norwegian dairy breeds and how these differences are reflected in the protein profiles.

MATERIALS AND METHODS

Animals and Milk Samples

Milk samples were collected from Telemark (**TF**, $n = 15$), Black-sided Trønder and Nordland (**STN**, $n = 30$), Doela (**DF**, $n = 17$), Western Fjord (**VFF**, $n = 35$), Western Red Polled (**VR**, $n = 13$), Eastern Red Polled (**ORA**, $n = 18$), and Norwegian Red (**NRF**, $n = 16$) cattle from a total of 30 different farms, where both the native and modern cows were milked routinely. Some farms had more than one of the breeds represented in their herd. The number of herds represented by each breed is noted in Table 1. A representative milk sample of 50 mL was collected from each cow during one morning and one evening milking and preserved with one tablet of bronopol (2-bromo-2-nitropropane-1,3-diol; D&F Control Systems Inc.) bronopol. The morning and evening milk samples from each cow were, based on volume, pooled into one sample upon arrival to the laboratory. The milk samples were skimmed (Inglingstad et al., 2014) and stored at -20°C until further analysis. The samples were transported frozen to Aarhus University, Department of Food Science, and stored at -40°C until further analysis.

Sample Preparation and Proteomic Profiling by Liquid Chromatography-Electrospray Ionization/MS

The frozen skim milk samples were thawed at room temperature and 150 μL of each skim milk sample was mixed with 450 μL of working solution (5.37 mM sodium citrate in 100 mM Bis-Tris buffer, 6 M guanidine hydrochloride pH 6.8) and 12 μL of 1 M dithioerythritol. The samples were incubated at room temperature for 1 h and centrifuged at $20,817 \times g$ (Eppendorf centrifuge 5417 R) for 10 min at 7°C . Four hundred microliters of the sample solution was filtered through a 0.45- μm polytetrafluoroethylene filter (Mini-UniPrep, Whatman).

Table 1. Abbreviations and local and English names of the 6 native Norwegian breeds and 1 modern breed (Norwegian Red) subjected to analysis in current study¹

Abbreviation	Number of samples	Number of herds	Local name	English name
TF	15	6	Telemarkfe	Telemark cattle
STN	30	9	Sidet Trønderfe and Nordlandsfe	Black-sided Trønder and Nordland cattle
DF	17	7	Dølafe	Doela cattle
VFF	35	6	Vestlandsk fjordfe	Western Fjord cattle
VR	13	4	Vestlandsk raudkolle	Western Red Polled cattle
ORA	19	3	Østlandsk rødkolle	Eastern Red Polled cattle
NRF	16	5	Norsk Rødt Fe	Norwegian Red cattle

¹Number of samples per breed and herd represented are listed.

The experimental setup for liquid chromatography (LC)-electrospray ionization (ESI)/MS analysis and the subsequent data analysis are based on previous studies (Bobe et al., 1998; Bonfatti et al., 2008; Bonizzi et al., 2009) with minor modifications (Frederiksen et al., 2011). The major proteins from the skim milk samples were separated using a reverse-phase LC-based method, according to their hydrophobic properties. The separation was performed on an Agilent 1260 Infinity II series instrument (Agilent Technologies), equipped with a binary pump including degasser (G7112B), a vial sampler (G7129A), and a column thermostat (G7116A). The system was coupled to a diode array detector (G7117C) for protein identification and relative quantification. The column used for separation was Jupiter C4 (250 × 2.00 mm, 5 µm particle size, 300 Å pores; Phenomenex). The running temperature was 40°C and UV absorbance was detected at 214 nm. An attached Agilent LC/MSD XD (G6135B) was used for MS. The different genetic variants and their post-translational modifications (PTM) were identified by ESI/MS. The running flow was 0.3 mL/min and the injection volume was 6 µL. The elution gradient started at 33% solvent B (0.05% trifluoroacetic acid in acetonitrile) mixed with solvent A (0.05% TFA in MilliQ water) and increased linearly to 50% solution B over 25 min by a decrease in solution B from 50 to 33% for 1 min. Each milk sample was analyzed once.

Identification and Quantification of Variants of the Major Milk Proteins

Data from HPLC was analyzed by the ChemStation software (Rev.C.01.10, Agilent Technologies). The major milk proteins were eluted in the following order: κ -CN, α_{S2} -CN, α_{S1} -CN, β -CN, β -LG, and α -LA, where retention times of eluted proteins varied, depending on genotype and degree of PTM (Frederiksen et al., 2011; Jensen et al., 2012b). The deconvolution tool in ChemStation was used for identification of genetic variants of the major isoforms of κ -CN (A, B, E), α_{S2} -CN (A), α_{S1} -CN (B, C), β -CN (A¹, A², B, I), β -LG (A, B), and α -LA (A), according to Jensen et al. (2012b). An in-house custom-made database of known masses and retention times was used for identification of the possible genetic variants and isoforms of the 6 major milk protein. In addition, low abundant genetic variants, not previously identified in these Norwegian breeds, were identified. These included α_{S2} -CN (D) and α_{S1} -CN (A, D), which have been reported previously, though in other breeds (Grosclaude et al., 1979; Eberhardt, 1993; Mohr et al., 1994; Jann et al., 2002; Ibeagha-Awemu et al., 2007).

Relative quantification of major proteins was calculated based on UV absorbance at 214 nm of the 4

caseins (α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN) and the 2 major whey proteins (β -LG, α -LA) to ascertain relative protein composition per breed. The relative quantification was based on integrated individual peak areas relative to the sum of total integrated area, calculated as the sum of peaks representing all genetic variants of α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN, α -LA, and β -LG and their major PTM isoforms. More specifically, κ -CN was identified in both glycosylated (κ -CN G) and unglycosylated (κ -CN UG) forms, including separation into its minor molecular glycosylation forms, as outlined by Visser et al. (1991). It was further possible to distinguish and separately quantify the 2 major α_{S1} -CN phosphorylation isoforms α_{S1} -CN 8P and 9P (for B and C variants) by integrating their separate peaks, as outlined by Jensen et al. (2012b). For the D variant of α_{S1} -CN it was possible to distinguish the 9P and 10P forms (Visser et al., 1991). Based on this, the phosphorylation degree (PD) of α_{S1} -CN was calculated as α_{S1} -CN 9P/total α_{S1} -CN for both B and C variants. For the D variant of α_{S1} -CN PD was calculated as 10P/total α_{S1} -CN. The peak representing the A variant contained masses corresponding to both 8P and 9P isoforms of this variant of α_{S1} -CN. Hence, the phosphorylation of the A variant did not contribute to the PD in the sample containing this variant of α_{S1} -CN. When defining the peaks for integration in the assessment of the protein compositions, proteolytic degradation products were not included, which means that, for example, the γ_3 -CN fragment of β -CN was not included in integration of the β -CN peaks or for determining the total protein, even though it was identified by its mass (being 11,557 Da). The peak representing the γ_3 -CN fragment was, however, used as a quality parameter for the milk samples. It was verified that none of the included milk samples contained more than 4% of γ_3 -CN, as this could indicate an increased proteolytic activity.

Statistical Analysis

One-way ANOVA was used to compare mean values of the relative abundance of each protein, as well as PD and glycosylation degree (GD) among breeds. Pairwise comparisons for all tests were done using Tukey's pairwise post hoc test. The statistical analysis were all conducted in R (version 4.0.0; <http://www.r-project.org>).

RESULTS

The sample set contained 144 milk samples from individual cows collected from 30 different farms in Norway analyzed as individual skim milk samples by LC-ESI/MS Single Q.

Figure 1. Full chromatogram of the elution profile of one sample from the sample set. The figure shows the entire elution profile with identified proteins and genetic variants indicated by arrows.

Identification of Rare α_{S1} - and β -CN Genetic Variants

Figure 1 represents a full UV chromatogram of the elution profile of one of the milk samples. The figure shows a clear separation of the major milk proteins and some of their genetic variants. This separation is used to identify the protein profile and the presence of the genetic variants that can be identified solely by LC.

Figure 2 displays representative UV chromatograms of milk samples from individual cows representing genetic variations in α_{S1} -CN. The chromatograms are enlarged in the interval from 12 to 19 min, representing α_{S1} -CN eluting first followed by β -CN (included as a reference). From this figure it is clear that different genetic variants of α_{S1} -CN show different elution profiles based on the LC only. Mass spectrometry is then included to identify and confirm masses of variants that were not identified by LC only. Based on obtained deconvoluted masses, the identified variants of α_{S1} -CN are presented in Table 2 along with the detected major phosphorylation isoforms, their elution times, as well as their detected and theoretical molecular masses. The α_{S1} -CN variant B 8P serves as the reference variant. The difference in the peak height of the 8P and 9P isoforms of α_{S1} -CN B variant and hence peak area indi-

cates the relative proportion of each isoform (Figure 2). The α_{S1} -CN C variant elutes at the same retention time as the B variant, and displays a chromatogram, where the 8P and 9P isoforms are also shown as a double peak, like the B variant. The difference between the B and the C variants is a substitution of only one AA residue (Glu²⁰⁷ → Gly), which apparently does not affect retention time of the protein isoforms significantly ($P < 0.05$) in this system. Thus, milk samples representing BB, BC, and CC show similar elution profiles (Figure 2), and the 2 variants are recognized from each other only by their deconvoluted masses (Table 2).

Furthermore, α_{S1} -CN variant D was also identified, which represents a substitution of Ala⁶⁸ → Thr⁶⁸P, and introducing a potential extra phosphorylation site. This results in a slightly higher elution time for the α_{S1} -CN variant D isoforms, as the major isoform of the α_{S1} -CN D 9P variant elutes at the approximately same position, as the α_{S1} -CN B and C 8P isoforms, whereas the α_{S1} -CN D 10P isoform elutes a bit later. This results in a triple peak in Figure 2 for BD genotypes of α_{S1} -CN. In contrast, the triple peak in α_{S1} -CN DD milk is due to the 9P isoforms splitting into a double peak. These findings illustrate the degree of information obtained by coupling LC with MS as the genetic variants and

Figure 2. Elution profile of samples with detection of different masses of α_{s1} -CN, corresponding to different genetic variants of the α_{s1} -CN. The figure represents milk from AB, BB, BC, BD, CC, and DD cows. The figure is an enlargement of the chromatogram in the interval of 12 to 19 min showing the peaks of both α_{s1} -CN and β -CN. The genotypes of α_{s1} -CN are noted at the right side of each chromatogram.

their differences in phosphorylation sites are very hard to identify only by using LC.

Using the experimental setup with the LC-ESI/MS we were able to detect 6 different genotypes of α_{s1} -CN across the analyzed samples, including genotypes with the rare A variant and the low abundant variant D (represented by the genotypes AB, BB, BC, BD, CC, and DD). The α_{s1} -CN peaks are followed by the peaks representing different β -CN isoforms. These are represented either by a single peak or present as double peaks, depending on the retention times of the actual genetic variants present in each of the analyzed individual milk samples. The well-known variants A¹, A², B, and I of β -CN were identified in the sample set of 144 samples. The most common β -CN variants A1 and A2 were identified in all of the breeds, whereas the I and B variants were only found in 2 of the native breeds each.

Identification of Rare α_{s2} - and κ -CN Variants

Figure 3 displays UV chromatograms from the individual milk samples as an enlargement of the area between 7 and 17 min, including peaks of κ -CN (B variant) and α_{s2} -CN. The broad peak of α_{s2} -CN is due to different phosphorylation isoforms as outlined in Table 3. The detected variants, their phosphorylation isoforms, AA substitutions, elution times, and detected and theoretical masses are presented in Table 3. The α_{s2} -CN variant A was identified in the phosphorylation forms 11 (reference variant), 10, 12, and 13P, all eluting approximately at the same retention time, resulting in one rather broad peak. The α_{s2} -CN variant D is the result of a deletion of AA residues 51 to 59, which comprises 3 phosphorylation sites (Caroli et al., 2009), and the major isoform therefore contains only 8 phosphorylations (Table 3). Despite this, the elution times of the α_{s2} -CN A and the D variants are almost identical, with both variants being represented by only one peak, appearing with some slightly different peak shapes. Apart from these known variants, we also identified some unknown masses in some of the milk samples. These unknown masses were identified in the peak representing the α_{s2} -CN variants but could not be assigned to any of the other known α_{s2} -CN variants. Hence, the elution time corresponded to the known α_{s2} -CN variants but the masses did not. One of these samples is included in Figure 3, where the chromatogram displays a changed elution profile compared with the other samples. In this particular sample, there was a minor peak (molecular weight = 24,279 Da) just before the peak usually representing the α_{s2} -CN. The masses identified in the peak usually representing α_{s2} -CN were also different in these particular samples (molecular weight = 25,198

Table 2. Identified genetic variants of α_{S1} -CN including their phosphorylation isoforms, AA substitution, elution time, and detected and theoretical molecular weights¹

Variant	Phosphorylation	AA substitution ²	ET ³ (min)	Detected M _w ⁴ (Da)	Theoretical M _w ⁵ (Da)
A	8P	Deletion Glu ²⁹ -Ala ⁴¹	13.1	22,066	22,069
	9P		13.1	22,146	22,149
B	8P	Reference variant	15.4	23,612	23,615
	9P		15.8	23,692	23,695
C	8P	Glu ²⁰⁷ → Gly	15.3	23,540	23,542
	9P		15.8	23,620	23,622
D	9P	Ala ⁶⁸ → ThrP	15.6	23,722	23,725
	10P		16	23,802	23,805

¹The variants were identified using liquid chromatography-electrospray ionization/mass spectrometry in milks from individual cows.

²Information about AA substitutions is based on Caroli et al. (2009) and UniProt (www.uniprot.org).

³Elution time (ET) of each of the peaks representing the given protein variants detected during the separation of the milk proteins.

⁴The detected molecular weight (M_w) of each of the identified peaks corresponding to the protein variants measured in dalton (Da).

⁵Molecular weights measured in dalton (Da) on the mature protein with listed modifications. They are calculated using ExPASy (www.expasy.org).

and 25,278 Da). The κ -CN was found in 3 variants: A, B, and E, where E was only detected in 2 of the native Norwegian breeds (VFF cattle).

Variation in Genetic Variant Frequencies Between Breeds

The frequencies of the genetic variants of the 4 caseins and β -LG among the 7 Norwegian breeds are given in Table 4. As stated above, for α_{S1} -CN the A, B, C, and D variants were detected, for variant α_{S2} -CN variants A, D, as well as an unknown variant as mentioned above. For β -CN A¹, A², B, and I were present, and for κ -CN A, B, and E variants were present. Of the whey proteins both A and B variants of β -LG were present, whereas for α -LA all was found exclusively as the B variant, and is therefore not shown in the table, with B variant then comprising 100% of all identified α -LA. The majority of the less frequent alleles were found as heterozygotes, which was expected due to the limited sample size included in the study.

For α_{S1} -CN, the predominant form is the B variant, with frequencies varying from 68.6% in VFF to 96.2% in VR (Table 4). From Table 5 it is clear that the predominant genotype of α_{S1} -CN is the BB (Table 5). The α_{S1} -CN C variant was found in all breeds except TF cattle, and had a relatively high frequency in both NRF (15.6%) and in VFF cattle (28.6%). Only one sample from VFF cattle represented a homozygous cow of α_{S1} -CN variant C (Table 5). The A variant of α_{S1} -CN was identified only in one sample belonging to the TF cattle breed. The D variant of α_{S1} -CN was detected in only 3 of the 7 breeds, and all in relatively low frequencies with VFF (2.9%), TF (3.3%), and DF cattle (14.7%),

and thus mainly found as BD heterozygotes (Tables 4 and 5).

For α_{S2} -CN, the predominant variant was the A variant, which was found in frequencies ranging from 73.1% (VR cattle) to 100% in both the reference breed, NRF, and in ORA cattle. The α_{S2} -CN D variant was found in low frequencies in STN (1.7%) and in VFF cattle (7.1%), but in relatively high frequencies in TF (16.7%) and in VR cattle (19.2%). As outlined in Tables 4 and 5 and in Figure 3, in some (n = 3) of the samples we were unable to identify masses corresponding to any known variants of α_{S2} -CN. These samples with no identified α_{S2} -CN masses belonged to DF (n = 2) and to VR (n = 1) cattle.

The frequencies of the β -CN variants indicate a very diverse distribution among the different breeds (Table 4). For both NRF, DF and VFF cattle, the A² is the predominant allele, but found at quite varying frequencies (87.5%, 52.9%, and 52.9%, respectively). On the other side, for TF, STN, and VR cattle, the most abundant allele is the A¹, found at frequencies ranging from 53.9% to 73.3%. The ORA cattle showed a 50/50 distribution of the β -CN A¹ and A² variants. The B allele variant was found only in STN, and in VFF cattle at relatively low frequencies (10% and 4.3%, respectively), whereas the I variant was found only in DF and VFF cattle (14.7% and 1.4%, respectively). From Table 5 it is clear that these differences in presence and frequencies of genetic variants manifest that genotype frequencies vary a lot among these breeds. The frequency of the homozygous genotype A¹A¹ is highest in TF cattle, where it is accountable for nearly half (46.7%) of the β -CN genotypes (Table 5). The remaining part is constituted by heterozygous A¹A². The reference breed,

Table 3. Identified genetic variants of α_{S2} -CN including the different phosphorylation forms, AA substitution, elution time, and detected and theoretical molecular weights¹

Variant	Phosphorylation isoform	AA substitution ²	ET ³ (min)	Detected M_w ⁴ (Da)	Theoretical M_w ⁵ (Da)
A	11P	Reference variant	10.3	25,225	25,228
	10P		10.3	25,144	25,148
	12P			25,305	25,308
	13P			25,385	25,388
D	8P	Deletion Glu ⁵¹ → Glu ⁵⁹	10.4	24,045	24,049
	7P		10.8	23,965	23,969
	9P			24,126	24,129
	10P			24,205	24,209
Unknown		Unknown	10.4	25,198/25,278	—

¹The variants were identified using liquid chromatography-electrospray ionization/mass spectrometry in milks from individual cows.

²Information about AA substitutions is based on Caroli et al. (2009) and UniProt (www.uniprot.org).

³Elution time (ET) of each of the peaks representing the given protein variants detected during the separation of the milk proteins.

⁴The detected molecular weight (M_w) of each of the identified peaks corresponding to the protein variants measured in dalton (Da).

⁵Molecular weights measured in dalton (Da) on the mature protein with modifications. They are calculated using ExPASy (www.expasy.org).

NRF, showed the opposite tendency of being comprised of 75% A²A² and 25% A¹A². For the remaining breeds the homozygous forms A¹A¹ and A²A² are also present,

but the majority of the samples were characterized by the genotype A¹A². The homozygous form II was identified in only one sample of the DF cattle breed. Doela, STN, and VFF cattle were the 3 breeds expressing the β -CN I and B variants. The samples representing the other 4 breeds expressed only A¹ and A² variants of β -CN. The β -CN A³ and F variants were not identified in any of the 144 samples representing the native Norwegian cattle breeds and common NRF.

For κ -CN, most breeds showed a slightly higher frequency of the A compared with the B variant, but for Norwegian cattle the A variant was found to be predominant (94%), with a corresponding very low frequency of the B variant. Western Red Polled cattle showed an opposite tendency compared with the other breeds, with a higher proportion of the B variant. The reference breed, NRF, was mainly composed of AA homozygotes (87.5%), whereas most of the native breeds had higher frequencies of the heterozygous variant AB. Western Red Polled showed relatively high proportion of homozygous variant BB (38.5%) and low frequency of variant AA (15.4%). The variant E of κ -CN was found only in STN ($n = 1$) and in ORA cattle ($n = 2$).

In the majority of the breeds, the most abundant variant of β -LG was the B variant with the highest frequency found in TF (80%; Table 4). Only VR cattle showed a higher frequency of β -LG variant A over B (69.2%).

Figure 3. Elution profile of samples with detection of different masses of α_{S2} -CN. The figure represents milk from AA, AD, and DD cows and a sample with masses not corresponding to any theoretical known masses (noted “not identified”). The figure shows a cutout of the elution time 5 to 15 min to visualize the difference between the peaks of α_{S2} -CN. The first peak at approximately 6 min elution time and the peak just before α_{S2} -CN represent κ -CN and serve as a reference.

Variations in the Relative Abundance of the Major Milk Proteins and Their Isoforms Between Breeds

Relative protein concentrations of major milk proteins and their specific isoforms as determined by LC-ESI/MS UV profiles are given in Table 6 for each breed. The proportion of α_{S1} -CN is seen to be relatively constant

Table 4. Variant frequencies (%) of the major milk proteins: α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN, and β -LG^{1,2}

Protein	Variant	TF (n = 15)	STN (n = 30)	VFF (n = 35)	DF (n = 17)	ORA (n = 19)	VR (n = 13)	NRF (n = 16)
α_{S1} -CN	A	3.3	—	—	—	—	—	—
	B	93.3	95.0	68.6	82.4	91.7	96.2	84.4
	C	—	5.0	28.6	2.9	8.3	3.9	15.6
	D	3.3	—	2.9	14.7	—	—	—
α_{S2} -CN	A	83.3	98.3	92.9	88.2	100.0	73.1	100.0
	D	16.7	1.7	7.1	—	—	19.2	—
	Unknown	—	—	—	11.8	—	7.7	—
β -CN	A1	73.3	56.7	41.4	32.4	50.0	53.9	12.5
	A2	26.7	33.3	52.9	52.9	50.0	46.2	87.5
	B	—	10.0	4.3	—	—	—	—
κ -CN	I	—	—	1.4	14.7	—	—	—
	A	56.7	55.0	55.7	70.6	63.9	38.5	93.8
	B	43.3	43.3	44.3	29.4	30.6	61.5	6.3
β -LG	E	—	1.7	—	—	5.6	—	—
	A	20.0	21.7	24.3	26.5	41.7	69.2	34.4
	B	80.0	78.3	75.7	73.5	58.3	30.8	65.6

¹Genetic variants were identified in milks from individual cows representing 7 Norwegian breeds using liquid chromatography-electrospray ionization/MS.

²TF = Telemark cattle; STN = Black-sided Trønder and Nordland cattle; NRF = Norwegian Red cattle; DF = Doela; VFF = Western Fjord; VR = Western Red Polled; ORA = Eastern Red Polled cattle.

between breeds, but with samples from STN (33.8%) cattle containing a significantly higher proportion of α_{S1} -CN than VR (31.9%) cattle (Table 6). The PD was found to be significantly different among the breeds,

ranging from 20.7% to 24.4% (Table 6). The lowest PD was found in ORA (20.7%) and with the highest in DF (24.4%), STN (24.4%), and VFF (23.7%) cattle (Table 6). The PD could not be calculated for α_{S1} -CN

Table 5. Genotype frequencies (%) of the major milk proteins: α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN, and β -LG^{1,2}

Protein	Genotype	TF (n = 15)	STN (n = 30)	VFF (n = 35)	DF (n = 17)	ORA (n = 19)	VR (n = 13)	NRF (n = 16)
α_{S1} -CN	AB	6.7	—	—	—	—	—	—
	BB	86.7	90.0	48.7	70.6	83.3	92.3	68.8
	BC	—	10.0	40.0	5.9	16.7	7.7	31.3
	BD	6.7	—	5.7	17.7	—	—	—
	CC	—	—	5.7	—	—	—	—
	DD	—	—	—	5.9	—	—	—
α_{S2} -CN	AA	66.7	96.7	85.7	88.2	100.0	61.5	100.0
	AD	33.3	3.3	14.3	—	—	23.1	—
	DD	—	—	—	—	—	7.7	—
	Unknown	—	—	—	11.8	—	7.7	—
β -CN	A ¹ A ¹	46.7	26.7	17.1	5.9	22.2	23.1	—
	A ¹ A ²	53.3	46.7	45.7	47.1	55.6	61.5	25.0
	A ¹ B	—	6.7	2.9	—	—	—	—
	A ¹ I	—	—	—	5.9	—	—	—
	A ² A ²	—	10.0	25.7	23.5	22.2	15.4	75.0
	A ² B	—	10.0	5.7	—	—	—	—
	A ² I	—	—	2.9	11.8	—	—	—
	II	—	—	—	5.9	—	—	—
κ -CN	AA	26.7	20.0	31.4	52.9	38.9	15.4	87.5
	AB	60.0	70.0	51.4	35.3	44.4	46.2	12.5
	AE	—	—	—	—	5.6	—	—
	BB	13.3	6.7	17.1	11.8	5.6	38.5	—
	BE	—	3.3	—	—	5.6	—	—
β -LG	AA	—	6.7	2.9	5.9	16.7	46.2	12.5
	AB	40.0	26.7	42.9	41.2	50.0	46.2	43.8
	BB	60.0	66.7	54.3	52.9	33.3	7.7	43.8

¹Genetic variants were identified in milks from individual cows representing 7 Norwegian breeds using liquid chromatography-electrospray ionization/MS.

²TF = Telemark cattle; STN = Black-sided Trønder and Nordland cattle; NRF = Norwegian Red cattle; DF = Doela; VFF = Western Fjord; VR = Western Red Polled; ORA = Eastern Red Polled cattle.

Table 6. Protein profile based on the relative proportions of each trait extracted from the UV signal using liquid chromatography-electrospray ionization/MS^{1,2}

Trait	DF	NRF	ORA	STN	TF	VFF	VR
α_{S1} -CN	33.2 $\pm$ 1.6 ^{ab}	33.7 $\pm$ 1.6 ^{ab}	32.8 $\pm$ 1.4 ^{ab}	33.8 $\pm$ 1.7 ^b	32.2 $\pm$ 1.8 ^{ab}	33.1 $\pm$ 1.6 ^{ab}	31.9 $\pm$ 1.9 ^a
α_{S1} -CN PD	24.4 $\pm$ 2.9 ^b	22.3 $\pm$ 3.5 ^{ab}	20.7 $\pm$ 3.3 ^a	24.4 $\pm$ 3.3 ^b	23.5 $\pm$ 3.8 ^{ab}	23.7 $\pm$ 2.6 ^b	23.7 $\pm$ 3.0 ^{ab}
α_{S2} -CN	6.9 $\pm$ 1.1 ^a	8.5 $\pm$ 1.5 ^b	7.5 $\pm$ 1.4 ^{ab}	7.3 $\pm$ 0.9 ^a	8.0 $\pm$ 1.1 ^{ab}	7.6 $\pm$ 1.2 ^{ab}	8.0 $\pm$ 1.4 ^{ab}
β -CN	40.4 $\pm$ 2.0 ^{ab}	39.6 $\pm$ 2.1 ^{ab}	39.3 $\pm$ 1.4 ^a	38.9 $\pm$ 2.1 ^a	41.4 $\pm$ 1.8 ^b	39.9 $\pm$ 1.6 ^{ab}	39.7 $\pm$ 2.1 ^{ab}
κ -CN	8.3 $\pm$ 1.5 ^{ab}	7.4 $\pm$ 0.8 ^a	8.7 $\pm$ 1.3 ^{ab}	9.2 $\pm$ 1.4 ^b	9.1 $\pm$ 1.9 ^b	8.9 $\pm$ 1.4 ^b	9.3 $\pm$ 2.0 ^b
κ -CN GD	34.0 $\pm$ 5.2 ^a	29.0 $\pm$ 4.1 ^a	30.0 $\pm$ 6.6 ^a	33.5 $\pm$ 7.0 ^a	33.0 $\pm$ 7.9 ^a	35.5 $\pm$ 6.7 ^a	33.5 $\pm$ 5.5 ^a
α -LA	3.4 $\pm$ 0.6 ^a	3.5 $\pm$ 0.4 ^a	3.4 $\pm$ 0.4 ^a	3.4 $\pm$ 0.5 ^a	3.1 $\pm$ 0.6 ^a	3.2 $\pm$ 0.3 ^a	3.3 $\pm$ 0.6 ^a
β -LG	7.0 $\pm$ 1.8 ^{ab}	6.9 $\pm$ 1.5 ^{ab}	7.7 $\pm$ 1.5 ^b	6.5 $\pm$ 1.5 ^{ab}	6.0 $\pm$ 0.8 ^a	6.4 $\pm$ 1.1 ^a	7.4 $\pm$ 1.5 ^{ab}

^{a,b}Values with different superscripts within each trait differ significantly from each other ($P < 0.05$).

¹The values represent the mean $\pm$ SD as proportions relative to total protein within each breed.

²TF = Telemark cattle; STN = Blacksided Trønder and Nordland cattle; NRF = Norwegian Red cattle; DF = Doela; VFF = Western Fjord; VR = Western Red Polled; ORA = Eastern Red Polled cattle.

BD samples and these samples were thus excluded from the PD calculations (4 DF samples, 1 TF sample, and 2 VFF samples).

For α_{S2} -CN, the lowest proportions were found in DF (6.9%) and STN (7.3%) cattle, being significantly different from the proportion found in the highest amount in the NRF (8.5%) cattle. For β -CN, TF (41.4%) expressed a significantly higher proportion of β -CN compared with ORA (39.3%) and STN (38.9%) cattle (Table 6). With respect to κ -CN, its proportion varied significantly among breeds (Table 6), with the highest percentage of κ -CN found in VR (9.3%), being significantly higher than relative proportion found in NRF cattle, representing the lowest (7.4%). Also, VFF (8.9%), TF (9.1%), and STN (9.2%) had significantly higher κ -CN percentage than the actual NRF cattle. There was no significant difference in GD among the breeds. The GD was around 30% for all breeds. The proportion of α -LA did not vary significantly among the breeds, with all containing between 3.1 and 3.5% of α -LA. For β -LG, the proportion was found to be significantly higher in ORA cattle (7.7%) compared

with proportions found in TF (6.6%) and VFF cattle (6.4%; Table 6).

Relative protein proportions in relation to specific genotypes of α_{S1} - and α_{S2} -CN are displayed in Tables 7 and 8, respectively. The genotypes AB and DD of α_{S1} -CN were both identified only in one sample, as the variants A and D are rare isoforms in this sample set. Hence, the values representing these genotypes are based on only one sample, and these were accordingly not included in the statistical analysis.

The proportion of α_{S1} -CN varied significantly among its different genotypes, with the highest proportion found for BC (34.0%), and the lowest proportion in BB (32.8%), being the most common genotype. The PD was found not to vary significantly among α_{S1} -CN genotypes (Table 7). It was remarkable, however, that the lowest PD was identified in the one sample representing the DD variant (17.3%), which was not included in the statistical analysis. Since the 9P forms of the B and D variants elute simultaneously, resulting in only one peak, the PD of samples from genotype BD were also not included in the statistical analysis. The

Table 7. Protein profile based on the proportions relative to total protein of each trait extracted from the UV signal when using liquid chromatography-electrospray ionization/MS relative to α_{S1} -CN genotypes (mean $\pm$ SD)

Trait	AB n = 1	BB n = 106	BC n = 27	BD n = 6	CC n = 3	DD n = 1
α_{S1} -CN	32.7	32.8 $\pm$ 1.7 ^a	34.0 $\pm$ 1.4 ^b	32.9 $\pm$ 1.3 ^{ab}	36.1 $\pm$ 0.7	29.3
α_{S1} -CN PD	NA ¹	23.5 $\pm$ 3.5 ^a	22.8 $\pm$ 2.8 ^a	NA	22.9 $\pm$ 2.4	17.3
α_{S2} -CN	7.2	7.8 $\pm$ 1.2 ^{ab}	7.3 $\pm$ 1.1 ^a	8.7 $\pm$ 1.9 ^b	6.3 $\pm$ 0.4	6.2
β -CN	39.5	39.7 $\pm$ 2.0 ^a	40.1 $\pm$ 1.7 ^a	39.1 $\pm$ 1.5 ^a	40.2 $\pm$ 6.5	42.6
κ -CN	10.1	8.9 $\pm$ 1.6 ^a	8.7 $\pm$ 1.2 ^a	8.0 $\pm$ 1.2 ^a	7.2 $\pm$ 1.4	7.5
κ -CN GD	29.1	33.4 $\pm$ 6.8 ^a	33.0 $\pm$ 6.9 ^a	33.5 $\pm$ 3.0 ^a	37.4 $\pm$ 3.2	27.5
α -LA	3.8	3.3 $\pm$ 0.5 ^a	3.2 $\pm$ 0.3 ^a	3.1 $\pm$ 0.3 ^a	3.2 $\pm$ 0.2	3.6
β -LG	5.8	6.8 $\pm$ 1.5 ^a	6.3 $\pm$ 1.1 ^a	7.5 $\pm$ 1.6 ^a	6.4 $\pm$ 0.8	9.6

^{a,b}Values with different superscripts within each trait differ significantly from each other ($P < 0.05$). AB, CC, and DD genotypes are not included in the statistical test due to low sample size.

¹NA = not available.

Table 8. Protein profile based on the relative proportions of each trait extracted from the UV signal when using liquid chromatography-electrospray ionization/MS1 relative to α_{S2} -CN genotype¹

Trait	Unknown n = 3	AA n = 126	AD n = 14	DD n = 1
α_{S1} -CN	32.9 $\pm$ 2.2	33.4 $\pm$ 1.6 ^b	30.8 $\pm$ 1.3 ^a	31.0
α_{S1} -CN PD	22.2 $\pm$ 4.4	23.1 $\pm$ 3.4 ^a	23.9 $\pm$ 2.6 ^a	8.6
α_{S2} -CN	7.3 $\pm$ 3.1	7.5 $\pm$ 1.2 ^a	9.1 $\pm$ 0.7 ^b	8.5
β -CN	40.0 $\pm$ 0.7	39.8 $\pm$ 2.0 ^a	40.2 $\pm$ 1.9 ^a	38.9
κ -CN	7.4 $\pm$ 2.0	8.7 $\pm$ 1.5 ^a	10.0 $\pm$ 1.5 ^b	8.7
κ -CN GD	34.6 $\pm$ 6.2	33.5 $\pm$ 6.8 ^a	32.0 $\pm$ 5.4 ^a	30.3
α -LA	3.9 $\pm$ 0.5	3.30 $\pm$ 0.5 ^a	3.2 $\pm$ 0.4 ^a	3.4
β -LG	8.1 $\pm$ 1.1	6.8 $\pm$ 1.4 ^a	6.1 $\pm$ 1.2 ^a	9.5

^{a,b}Values with different superscripts within each trait differ significantly from each other ($P < 0.05$). Unknown and DD genotypes are not included in the statistical test due to low sample size.

¹The values represent the mean $\pm$ SD as relative proportions within each genotype of α_{S2} -CN.

PD of the homozygote DD is based only on one sample, and is calculated as the area of the 10P peak divided by the total area of α_{S1} -CN. The proportion of α_{S2} -CN was significantly different between α_{S1} -CN genotypes with the highest proportion found in BD (8.7%) and the lowest in BC (7.3%). The proportions of the other traits did not vary significantly between α_{S1} -CN genotypes. Table 8 displays proportions of the major proteins, including PD and GD relative to α_{S2} -CN genotypes. The proportion of α_{S1} -CN varied significantly between genotype AA (33.4%) and AD (30.8%) of α_{S2} -CN. The α_{S1} -CN PD did not vary significantly among genotypes, though the highest proportion was interestingly identified in the sample identified as the DD homozygote (28.6%). The proportion of α_{S2} -CN was found to vary significantly between genotypes of α_{S2} -CN, with the highest proportion found in the α_{S2} -CN heterozygous AD (9.1%) and the lowest proportion in α_{S2} -CN homozygous AA (7.5%). The relative concentrations of β -CN did not vary significantly among genotypes. The proportion of κ -CN was highest in the samples of α_{S2} -CN AD genotype (10.0%) and lowest in the “unknown” (7.4%) and AA (8.7%) α_{S2} -CN type samples. The GD of κ -CN, and the proportions of α -LA and β -LG, did not vary significantly among different genotypes of α_{S2} -CN.

DISCUSSION

Presence of Rare α_{S1} - and α_{S2} -CN Genetic Variants and PTM Isoforms in Native Breeds

This study documents great variation both in terms of the variants present and their variable frequencies of major milk proteins from native Norwegian dairy breeds, both among the breeds and in comparison with

the current high-yielding NRF cattle. It was possible to identify the presence of rare genetic variants in some of the native breeds, especially of the α_{S1} - and α_{S2} -CN in TF, VFF, DF, and VR cattle as well as some unidentified potential new variants of α_{S2} -CN. These variants had in some instances large implications on their PTM patterns. Within α_{S1} -CN, a total of 4 variants were demonstrated, representing A, B, C, and D variants. The A (present in TF) and D (present in TF, VFF, and DF cattle) variants of α_{S1} -CN have not been identified previously in these native breeds, and were furthermore not represented in the current NRF samples analyzed here. This study also illustrates how the usage of LC coupled to MS provides a more detailed protein profile compared with only using LC for protein profiling. Especially the phosphorylations of the genetic variants of α_{S1} -CN were identified by a combination of LC and MS. Lien et al. (1999) carried out a comparable study of protein allele frequencies in Nordic cattle breeds, though at the DNA level using blood samples. Among others, they included the same native Norwegian breeds as here, but, relying on primer selection, did not allow for identification of these A and D variants of α_{S1} -CN. Interestingly, we found the A variant of α_{S1} -CN to elute earlier than the other α_{S1} -CN variants during the HPLC separation, probably caused by the deletion of AA and buried phosphorylations (Table 2). The α_{S1} -CN A variant has been identified earlier in various breeds, including Holstein Friesians, Red Holsteins, and German Red cattle (Ng-Kwai-Hang et al., 1984; Farrell et al., 2004; Caroli et al., 2009), but is usually found in relatively low frequencies in comparison with the more common α_{S1} -CN B and C variants (Ng-Kwai-Hang et al., 1984). The finding of the rare α_{S1} -CN A variant among the native Norwegian breeds, but only in TF, is interesting from a conservation perspective. Additionally, we identified the, also rare, α_{S1} -CN D variant in 3 of the native breeds (TF, VFF, and DF). It was, not surprisingly, detected primarily in heterozygotes, but is interesting due to its extra phosphorylation site at the Thr⁶⁸ position, resulting in 9P and 10P isoforms. Our data revealed one sample from DF being homozygote α_{S1} -CN DD. It would be interesting to further study the CN micelle size and structural features in this sample, and furthermore to isolate the α_{S1} -CN D variant for comparison of its characteristics, including position and degree in use of exact phosphorylation sites, with other variants of α_{S1} -CN. The chromatogram of this sample (Figure 2) also shows an interesting shape of a triple peak, probably due to the difference in phosphorylation sites, as the extra potential phosphorylation site also facilitates possibilities for heterogeneities.

For α_{S2} -CN, the experimental setup permitted the identification of the rare D deletion variant (Table 3).

As the elution time of α_{S2} -CN D is the same as for the more common variant A, the D variant was identified based solely on its deconvoluted phosphorylation isoform masses (Table 3) of 24,045 and 23,965 Da. The α_{S2} -CN variant D has previously been identified in other breeds. It was first reported in the French Vosgienne and Montbéliarde cattle breeds, by gel electrophoresis (Grosclaude et al., 1979), and has also been found in German breeds by Eberhardt (1993). Ibeagha-Awemu et al. (2007) further identified the α_{S2} -CN D variant in several of the African *Bos indicus* cattle breeds by DNA sequencing of blood samples. They identified the α_{S2} -CN D variant in relatively low allele frequencies (1.1–8.8%), compared with the more common A and B variants (Ibeagha-Awemu et al., 2007).

In this study, both the A and D variants of α_{S2} -CN were found in 4 different phosphorylation forms. The A variant was found as 10P, 11P, 12P, and 13P. Jensen et al. (2012b) were able to detect the varying degrees of phosphorylation of the A variant of α_{S2} -CN by the elution of 2 partly separated peaks. This was not the case in the current study, as the elution of α_{S2} -CN A did not result in 2 separated peaks. We were only able to identify one peak representing α_{S2} -CN A variant (Figure 3) and were not able to separate the α_{S2} -CN A on the basis on phosphorylation forms. The deletion characterizing the α_{S2} -CN D variant (Table 3) implies the loss of 3 phosphorylation sites (Glu⁵¹–Glu⁵⁹), and in accordance with this the D variant was found as 7P, 8P, 9P, and 10P forms. The finding of this variant in homozygous form in the VR breed would permit the potential isolation of this variant and its isoforms for further studies of their peculiar molecular features. Furthermore it would be interesting also to study the micelle size and other characteristics of this variant in fresh (without cooling or freezing history) milk from α_{S2} -CN DD VR animals. Apart from identifying the A and D variants of α_{S2} -CN in the sample set, we also found some samples with masses not corresponding to any of the known variants of α_{S2} -CN. These masses (25,198 and 25,278 Da) were identified in a peak with the same elution time and shape as the known variants of α_{S2} -CN. The difference between the 2 masses is 80 Da corresponding to the mass of a phosphate unit, which is very interesting and could suggest a homozygous variant. It would be interesting to analyze these findings on fresh milk samples and to isolate according to α_{S2} -CN protocols.

Differences in Frequencies of the Genetic Variants Between Breeds

Our results documented large differences in the genotype frequencies of the major milk proteins among the

breeds included in this study. In line with Lien et al. (1999), α_{S1} -CN B is still the predominant variant in all breeds studied here, but VFF cattle had a relatively high frequency of α_{S1} -CN variant C (28.6%). This variant was also relatively high in NRF cattle, suggesting a similarity between these breeds on this point. The C variant of α_{S1} -CN was here reported in DF cattle for the first time. Likewise, α_{S1} -CN variant A and D together with α_{S2} -CN variant D have not been reported before in these breeds. The D variant of α_{S1} -CN was mainly detected in relatively low frequencies, but with a frequency of 14.7% in the DF cattle. Neither the A nor the D variant of α_{S1} -CN were identified in the samples representing the high-yielding NRF cattle, as only the more common α_{S1} -CN B and C variants were found within this breed. However, for all breeds α_{S1} -CN B variant was the most common, which is also the most common variant in European cattle breeds (Lien et al., 1999).

The current study identified the rare D variant of α_{S2} -CN in several breeds. Telemark and VR cattle expressed relatively high proportions of the D variant, 16.7% and 19.2%. One sample of VR cattle even represented the α_{S2} -CN homozygote DD. Phosphorylations of the caseins are very important for technological properties of the milk as they are directly related to the binding capacity of calcium phosphate in the micelles and may therefore have implications for the coagulation properties (Frederiksen et al., 2011; Le et al., 2017). Hence, degree of phosphorylation of especially α_{S1} -CN and α_{S2} -CN and the loss of phosphorylation sites found in α_{S2} -CN D and extra phosphorylation site found in α_{S1} -CN D may thus affect milk coagulation properties due to the importance of casein phosphorylations in casein micelle formation and stability.

With respect to β -CN, in the study of Lien et al. (1999), variant B had similar high frequencies in STN, whereas this variant has not been reported previously for VFF cattle. Likewise, identification of β -CN variant I in DF and VFF is novel. Further compared with the study of Lien et al. (1999), a much higher allele frequency of β -CN variant A² was found in our study (87.5% vs. 48.7%), whereas STN was found to have a higher frequency of β -CN variant A¹ (56.7%) than reported earlier (28.1%), compared with the slight shifts in A¹ and A² frequencies observed when comparing the other breeds. In this study, we found a very high frequency of β -CN A²A² in NRF compared with the native breeds. This is interesting from a technological perspective, as A² has been found to be associated with higher milk yield but also a lower κ -CN concentration and thereby potentially poorer rennet coagulation (Hallén et al., 2007, 2008; Frederiksen et al., 2011; Jensen et al., 2012a). However, it is argued that it is hard to dis-

tinguish a real effect of isolated protein variants like for instance the A² variant of β -CN, as it is often associated with specific genetic variants of the other caseins due to linkage disequilibrium within the casein gene cluster. Sanchez et al. (2020) studied the haplotype frequencies in 12 French cattle breeds and found them to be very diverse across breeds. They found the association of the A² variant of β -CN to be mainly associated with the B variant of α_{S1} -CN and the A variant of κ -CN in Holstein. For Jersey the A² variant of β -CN was oppositely mainly associated with the C variant of α_{S1} -CN and the B variant of κ -CN (Sanchez et al., 2020).

For κ -CN, variant E has not previously been reported in STN cattle, and generally higher frequencies of κ -CN B were found for all breeds except NRF cattle, where it was slightly lower and no κ -CN variant E was detected (Lien et al., 1999). For β -LG, variant A was also reported as predominant in VR cattle earlier, but at a lower frequency, whereas β -LG variant B was predominant in the others breeds. Geographically, STN cattle are located in the Northern part of Norway, whereas the other native breeds have generally been kept in the mid and southern area of Norway, apart from VFF- and Red Polled cattle, which are breeds originating from the Norwegian Westland. Based on the Lien et al. (1999) study on protein variants, STN and VFF cattle were classified into a group of Nordic breeds with Northern location, whereas the other native breeds were assigned to a group of breeds with southern location and NRF cattle together with modern breeds. Black-sided Trønder and Nordland and VFF cattle also display very similar variant frequencies here supporting this grouping, but other genetic studies have also suggested VFF to be more closely related to VR (Kantanen et al., 2000; Tapio et al., 2006).

Implications of Variation in Protein Composition Between the Breeds

In relation to detailed protein composition, relative amount of α_{S1} -CN varied significantly among breeds. Highest proportion was found in STN (33.8%) and the lowest in VR (31.9%). Other studies have found different associations between the β -CN variants (A¹ and A²) and the compositional traits. Heck et al. (2009) and Bonfatti et al. (2010) both reported the A¹ variant of β -CN to be associated with higher relative proportions of both α_{S1} -CN and κ -CN, as well as lower concentrations of β -CN and α_{S2} -CN (Heck et al., 2009; Bonfatti et al., 2010). A more recent study found the A¹ variant of β -CN to be associated with a higher relative concentration of β -CN (Poulsen et al., 2017b). In the current study, TF cattle showed the highest frequency of A¹ β -CN (73.3%) and had a significantly higher rela-

tive concentration of β -CN than STN and ORA cattle (Table 6). In line with this, Nilsson et al. (2020) found a significantly higher relative concentration of β -CN in poor and noncoagulating milk.

The κ -CN proportion varied between the native Norwegian breeds and NRF cattle, whereas NRF holds the lowest κ -CN variant B frequency (Table 6) and had significant lower proportion of κ -CN compared with VR, which also had the highest frequency of κ -CN variant B (Table 6). This suggests that the previous association between κ -CN variant B and the proportion of κ -CN in milk is also observed here, whereas no difference in κ -CN GD is observed among the breeds. This could suggest that compared with the native breeds, NRF cattle could have larger casein micelles, which in turn is reflected in lower coagulation properties compared with the native breeds. Ketto et al. (2017) studied the coagulation properties of milk from NRF cows related to the genetic variants and found that κ -CN variant BB was associated with improved coagulation properties (Ketto et al., 2017).

CONCLUSIONS

The genetic variants and important PTM of the major milk proteins, α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN, α -LA, and β -LG, can be identified by LC/ESI-MS from individual cow milk. This study identified variants of α_{S1} -CN and α_{S2} -CN that have not previously been identified in the native Norwegian breeds included in this study. The very rare A variant of α_{S1} -CN was identified in TF cattle, whereas the rare D variant was identified in TF, VFF, and DF cattle. Additionally, we were able to detect a α_{S1} -CN DD homozygote in one of the samples of the DF cattle. In general, the native Norwegian breeds showed more diversity in the genetic variants and the protein profile of the major proteins compared with the high-yielding NRF cattle. In this study, the rare A variant of α_{S1} -CN was identified in TF cattle only. The D variant of α_{S1} -CN was found in both TF, VFF, and DF cattle. These variants have not been identified in these breeds before and were also not found in the NRF. For α_{S2} -CN we identified the rarer D variant as well as the common A variant. The D variant was identified in 4 of the native breeds and not in the NRF cattle. It was identified in homozygous in one sample from VR cattle. The protein profiles based on relative peak areas also showed differences among breeds, as well as the relative protein distribution according to α_{S1} -CN and α_{S2} -CN genotypes. The relative proportion of α_{S1} -CN varied significantly between breeds, being highest in STN and lowest in VR cattle. Norwegian Red cattle showed significantly lower relative amounts of κ -CN and also a very low frequency of the κ -CN B variant compared

with the native breeds, which may cause implications for the technological properties of the milk.

ACKNOWLEDGMENTS

The study was financed by Aarhus University, the Danish Milk Levy Fund, and the Fund for Organic Agriculture, Denmark. The authors thank Gitte Hald Kristiansen (Department of Food Science, Aarhus University, Denmark) for excellent technical assistance, and Nina Sæther and Anna C. Holene [Norwegian Genetic Resource Centre, Norwegian Institute of Bioeconomy Research (NIBIO), Aas, Norway] for assisting in selection of cattle included in this study. The authors have not stated any conflicts of interest.

REFERENCES

- Bobe, G., D. C. Beitz, A. E. Freeman, and G. L. Lindberg. 1998. Separation and quantification of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J. Agric. Food Chem.* 46:458–463. <https://doi.org/10.1021/jf970499p>.
- Bonfatti, V., G. Di Martino, A. Cecchinato, D. Vicario, and P. Carnier. 2010. Effects of β - κ -casein (CSN2–CSN3) haplotypes and β -lactoglobulin (BLG) genotypes on milk production traits and detailed protein composition of individual milk of Simmental cows. *J. Dairy Sci.* 93:3797–3808. <https://doi.org/10.3168/jds.2009-2778>.
- Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier. 2008. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. *J. Chromatogr. A* 1195:101–106. <https://doi.org/10.1016/j.chroma.2008.04.075>.
- Bonizzi, I., J. N. Buffoni, and M. Feligini. 2009. Quantification of bovine casein fractions by direct chromatographic analysis of milk. Approaching the application to a real production context. *J. Chromatogr. A* 1216:165–168. <https://doi.org/10.1016/j.chroma.2008.11.045>.
- Caroli, A. M., S. Chessa, and G. J. Erhardt. 2009. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *J. Dairy Sci.* 92:5335–5352. <https://doi.org/10.3168/jds.2009-2461>.
- Eberhardt, G. 1993. Allele frequencies of milk proteins in German cattle breeds and demonstration of alpha(S2)-casein variants by isoelectric focusing. *Arch. Tierzucht* 36:145–152.
- Farrell, H. M. Jr., R. Jimenez-Flores, G. T. Bleck, E. M. Brown, J. E. Butler, L. K. Creamer, C. L. Hicks, C. M. Hollar, K. F. Ng-Kwai-Hang, and H. E. Swaisgood. 2004. Nomenclature of the proteins of cows' milk—Sixth revision. *J. Dairy Sci.* 87:1641–1674. [https://doi.org/10.3168/jds.S0022-0302\(04\)73319-6](https://doi.org/10.3168/jds.S0022-0302(04)73319-6).
- Frederiksen, P. D., K. K. Andersen, M. Hammershøj, H. D. Poulsen, J. Sørensen, M. Bakman, K. B. Qvist, and L. B. Larsen. 2011. Composition and effect of blending of noncoagulating, poorly coagulating, and well-coagulating bovine milk from individual Danish Holstein cows. *J. Dairy Sci.* 94:4787–4799. <https://doi.org/10.3168/jds.2011-4343>.
- Glantz, M., T. G. Devold, G. E. Vegarud, H. Lindmark Månsson, H. Stålhammar, and M. Paulsson. 2010. Importance of casein micelle size and milk composition for milk gelation. *J. Dairy Sci.* 93:1444–1451. <https://doi.org/10.3168/jds.2009-2856>.
- Grosclaude, F., P. Joudrier, and M.-F. Mahé. 1979. A genetic and biochemical analysis of a polymorphism of bovine α S2-casein. *J. Dairy Res.* 46:211–213. <https://doi.org/10.1017/S0022029900017052>.
- Hallén, E., T. Allmere, J. Näslund, A. Andrén, and A. Lundén. 2007. Effect of genetic polymorphism of milk proteins on rheology of chymosin-induced milk gels. *Int. Dairy J.* 17:791–799. <https://doi.org/10.1016/j.idairyj.2006.09.011>.
- Hallén, E., A. Wedholm, A. Andrén, and A. Lundén. 2008. Effect of β -casein, κ -casein and β -lactoglobulin genotypes on concentration of milk protein variants. *J. Anim. Breed. Genet.* 125:119–129. <https://doi.org/10.1111/j.1439-0388.2007.00706.x>.
- Heck, J. M. L., A. Schennink, H. J. F. Van Valenberg, H. Bovenhuis, M. H. P. W. Visker, J. A. M. Van Arendonk, and A. C. M. Van Hooijdonk. 2009. Effects of milk protein variants on the protein composition of bovine milk. *J. Dairy Sci.* 92:1192–1202. <https://doi.org/10.3168/jds.2008-1208>.
- Ibeagha-Awemu, E. M., E. M. Prinzenberg, O. C. Jann, G. Lühken, A. E. Ibeagha, X. Zhao, and G. Erhardt. 2007. Molecular characterization of bovine CSN1S2*B and extensive distribution of zebu-specific milk protein alleles in European cattle. *J. Dairy Sci.* 90:3522–3529. <https://doi.org/10.3168/jds.2006-679>.
- Inglingstad, R. A., H. Steinshamn, B. S. Dagnachew, B. Valenti, A. Criscione, E. O. Rukke, T. G. Devold, S. B. Skeie, and G. E. Vegarud. 2014. Grazing season and forage type influence goat milk composition and rennet coagulation properties. *J. Dairy Sci.* 97:3800–3814. <https://doi.org/10.3168/jds.2013-7542>.
- Jann, O., G. Ceriotti, A. Caroli, and G. Erhardt. 2002. A new variant in exon VII of bovine B-casein gene (CSN2) and its distribution among European cattle breeds. *J. Anim. Breed. Genet.* 119:65–68. <https://doi.org/10.1046/j.1439-0388.2002.00318.x>.
- Jensen, H. B., J. W. Holland, N. A. Poulsen, and L. B. Larsen. 2012a. Milk protein genetic variants and isoforms identified in bovine milk representing extremes in coagulation properties. *J. Dairy Sci.* 95:2891–2903. <https://doi.org/10.3168/jds.2012-5346>.
- Jensen, H. B., N. A. Poulsen, K. K. Andersen, M. Hammershøj, H. D. Poulsen, and L. B. Larsen. 2012b. Distinct composition of bovine milk from Jersey and Holstein-Friesian cows with good, poor, or noncoagulation properties as reflected in protein genetic variants and isoforms. *J. Dairy Sci.* 95:6905–6917. <https://doi.org/10.3168/jds.2012-5675>.
- Kantanen, J., I. Olsaker, L. E. Holm, S. Lien, J. Vilkkilä, K. Brusgaard, E. Eythorsdottir, B. Danell, and S. Adalsteinsson. 2000. Genetic diversity and population structure of 20 North European cattle breeds. *J. Hered.* 91:446–457. <https://doi.org/10.1093/jhered/91.6.446>.
- Ketto, I. A., T. M. Knutsen, J. Øyaas, B. Heringstad, T. Ådnøy, T. G. Devold, and S. B. Skeie. 2017. Effects of milk protein polymorphism and composition, casein micelle size and salt distribution on the milk coagulation properties in Norwegian Red cattle. *Int. Dairy J.* 70:55–64. <https://doi.org/10.1016/j.idairyj.2016.10.010>.
- Kierkegaard, L. S., L. F. Groeneveld, A. Kettunen, and P. Berg. 2020. The status and need for characterization of Nordic animal genetic resources. *Acta Agric. Scand. A Anim. Sci.* 69:2–24. <https://doi.org/10.1080/09064702.2020.1722216>.
- Le, T. T., H. C. Deeth, and L. B. Larsen. 2017. Proteomics of major bovine milk proteins: Novel insights. *Int. Dairy J.* 67:2–15. <https://doi.org/10.1016/j.idairyj.2016.11.016>.
- Lien, S., J. Kantanen, I. Olsaker, L. E. Holm, E. Eythorsdottir, K. Sandberg, B. Dalsgaard, and S. Adalsteinsson. 1999. Comparison of milk protein allele frequencies in Nordic cattle breeds. *Anim. Genet.* 30:85–91. <https://doi.org/10.1046/j.1365-2052.1999.00434.x>.
- Mohr, U., D. Koczan, D. Linder, G. Hobom, and G. Erhardt. 1994. A single point mutation results in A allele-specific exon skipping in the bovine α S1-casein mRNA. *Gene* 143:187–192. [https://doi.org/10.1016/0378-1119\(94\)90095-7](https://doi.org/10.1016/0378-1119(94)90095-7).
- Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes. 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. *J. Dairy Sci.* 67:835–840. [https://doi.org/10.3168/jds.S0022-0302\(84\)81374-0](https://doi.org/10.3168/jds.S0022-0302(84)81374-0).
- Nilsson, K., L. Buhelt Johansen, D. J. de Koning, S. I. Duchemin, M. Stenholdt Hansen, H. Stålhammar, H. Lindmark-Månsson, M. Paulsson, W. F. Fikse, and M. Glantz. 2020. Effects of milk proteins and posttranslational modifications on noncoagulating milk from Swedish Red dairy cattle. *J. Dairy Sci.* 103:6858–6868. <https://doi.org/10.3168/jds.2020-18357>.
- Poulsen, N. A., H. P. Bertelsen, H. B. Jensen, F. Gustavsson, M. Glantz, H. Lindmark Månsson, A. Andrén, M. Paulsson, C. Ben-

- dixen, A. J. Buitenhuis, and L. B. Larsen. 2013. The occurrence of noncoagulating milk and the association of bovine milk coagulation properties with genetic variants of the caseins in 3 Scandinavian dairy breeds. *J. Dairy Sci.* 96:4830–4842. <https://doi.org/10.3168/jds.2012-6422>.
- Poulsen, N. A., M. Glantz, A. K. Rosenggaard, M. Paulsson, and L. B. Larsen. 2017a. Comparison of milk protein composition and rennet coagulation properties in native Swedish dairy cow breeds and high-yielding Swedish Red cows. *J. Dairy Sci.* 100:8722–8734. <https://doi.org/10.3168/jds.2017-12920>.
- Poulsen, N. A., A. K. Rosenggaard, B. D. Szekeres, V. R. Gregersen, H. B. Jensen, and L. B. Larsen. 2017b. Protein heterogeneity of bovine β -casein in Danish dairy breeds and association of rare β -casein F with milk coagulation properties. *Acta Agric. Scand. A Anim. Sci.* 66:190–198. <https://doi.org/10.1080/09064702.2017.1342858>.
- Sanchez, M. P., S. Fritz, C. Patry, A. Delacroix-Buchet, and D. Boichard. 2020. Frequencies of milk protein variants and haplotypes estimated from genotypes of more than 1 million bulls and cows of 12 French cattle breeds. *J. Dairy Sci.* 103:9124–9141. <https://doi.org/10.3168/jds.2020-18492>.
- Tapio, I., S. Värvi, J. Bennewitz, J. Maleviciute, E. Fimland, Z. Grislis, T. H. E. Meuwissen, I. Miceikiene, I. Olsaker, H. Viinalass, J. Vilkki, and J. Kantanen. 2006. Prioritization for conservation of northern European cattle breeds based on analysis of microsatellite data. *Conserv. Biol.* 20:1768–1779. <https://doi.org/10.1111/j.1523-1739.2006.00488.x>.
- Visser, S., C. J. Slangen, and H. S. Rollema. 1991. Phenotyping of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J. Chromatogr.* 548:361–370. [https://doi.org/10.1016/S0021-9673\(01\)88619-2](https://doi.org/10.1016/S0021-9673(01)88619-2).
- Wedholm, A., L. B. Larsen, H. Lindmark-Månsson, A. H. Karlsson, and A. Andrén. 2006. Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *J. Dairy Sci.* 89:3296–3305. [https://doi.org/10.3168/jds.S0022-0302\(06\)72366-9](https://doi.org/10.3168/jds.S0022-0302(06)72366-9).

ORCIDS

- N. R. Roin <https://orcid.org/0000-0002-5523-001X>
 L. B. Larsen <https://orcid.org/0000-0001-9674-0107>
 T. G. Devold <https://orcid.org/0000-0001-6824-2397>
 R. A. Inglingstad <https://orcid.org/0000-0001-6413-7557>
 N. A. Poulsen <https://orcid.org/0000-0001-9983-9663>