Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the Danish Holstein

G. Gebreyesus,*‡ M. S. Lund,* L. Janss,* N. A. Poulsen,† L. B. Larsen,† H. Bovenhuis,‡ and A. J. Buitenhuis*[1]
*Center for Quantitative Genetics and Genomics, and
†Department of Food Science, Aarhus University, Blichers Allé 20, PO Box 50, DK-8830 Tjele, Denmark
‡Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

ABSTRACT

Genetic parameters were estimated for the major milk proteins using bivariate and multi-trait models based on genomic relationships between animals. The analyses included, apart from total protein percentage, α_{S1}-casein (CN), α_{S2}-CN, β-CN, κ-CN, α-lactalbumin, and β-lactoglobulin, as well as the posttranslational sub-forms of glycosylated κ-CN and α_{S1}-CN-8P (phosphorylated). Standard errors of the estimates were used to compare the models. In total, 650 Danish Holstein cows across 4 parities and days in milk ranging from 9 to 481 d were selected from 21 herds. The multi-trait model generally resulted in lower standard errors of heritability estimates, suggesting that genetic parameters can be estimated with high accuracy using multi-trait analyses with genomic relationships for scarcely recorded traits. The heritability estimates from the multi-trait model ranged from low (0.05 for β-CN) to high (0.78 for κ-CN). Genetic correlations between the milk proteins and the total milk protein percentage were generally low, suggesting the possibility to alter protein composition through selective breeding with little effect on total milk protein percentage.
Key words: genetic parameter, milk protein, multi-trait model, genomic relationship

Short Communication

Milk protein composition plays an important role in the technological properties of milk (Ikonen et al., 1999; Bittante et al., 2012). Changes in relative concentrations of individual milk proteins have a major effect on milk coagulation properties (Bonfatti et al., 2011) and coagulation ability of milk is essential in cheese mak-

ing (Cassandro et al., 2008). The major milk proteins include α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, and β-LG. In addition, several posttranslational modifications of these proteins exist in milk.

Previous studies have shown that considerable genetic variation exists in the composition of milk protein (Bobe et al., 1999; Schopen et al., 2009), presenting the opportunity to alter milk protein composition through selective breeding. Reliable estimates of genetic parameters, including heritability and genetic covariances, are crucial to evaluate the potential for breeding. Quantifying specific milk proteins requires specialized and costly equipment, making it difficult and expensive to measure the traits. As a result, sufficient phenotypic data are not available for reliable estimation of genetic parameters. One effective strategy to deal with such scarcely recorded traits could be implementation of multi-trait models that take advantage of information from correlated traits (Calus and Veerkamp, 2011).

Generally, only a few studies have previously estimated genetic parameters for specific milk proteins (Schopen et al., 2009; Bonfatti et al., 2011) and their posttranslational sub-forms (Bijl et al., 2014). More importantly, none of the previous studies have estimated genetic parameters for milk protein profile using multi-trait analyses.

In this study, we estimated genetic parameters for the major milk proteins (α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA and β-LG), the posttranslational sub-forms (glycosylated κ-CN and α_{S1}-CN-8P, where P = phosphorylated serine), as well as protein percentage using bivariate and multi-trait models with genomic relationships between animals and compared standard errors of the estimated genetic parameters.

Morning milk samples were obtained from 650 cows from 21 herds in Denmark. The cows were in different stages of lactation (d 9 to 481 in milk) and parity 1 to 4. The liquid chromatography/electrospray ionization-mass spectrometry (LC/ESI-MS) methods were used

to profile the milk proteins. Details on screening of samples and quantification of milk proteins were previously described by Jensen et al. (2012).

Of the total cows, 372 were genotyped using the BovineHD Illumina BeadChip. The remaining 278 cows were genotyped with the BovineSNP50 beadchip. Genomic DNA was extracted from ear tissue. Genotypes were subsequently imputed to full sequence in a 2-step procedure. The 278 cows genotyped with the BovineSNP50 chip were first imputed to the BovineHD (777k) level using a multi-breed reference of 3,383 animals including the 372 HD genotyped cows used in this study. The true and imputed HD data for the 2 cow groups were then merged and imputed to the whole-genome sequence level using a multi-breed reference of 1,228 animals from the "1000 bull genomes" project (http://www.1000bullgenomes.com/) and data from Aarhus University using IMPUTE2 v2.3.1 (Howie et al., 2011).

The genomic relationship matrix was calculated as described by the first method presented in VanRaden (2008). In total, 3.7 million SNP markers spread over BTA1 to BTA29 were included to calculate the **G** matrix.

The REML approach in DMU was used to estimate genetic parameters and variance components (Madsen and Jensen, 2010). Bivariate and multi-trait analyses were performed and compared using standard errors for the estimated heritability.

The general model used was

$$y_{ijkl} = \mu + \text{parity}_i + \text{herd}_j + b_1 \times \text{DIM}_k + b_2$$
$$\times \exp^{-0.05 \times \text{DIM}}{}_k + \text{animal}_l + e_{ijkl}, \qquad [1]$$

where y_{ijkl} was the observation of animal l, in parity i and herd j; μ was the fixed mean effect; b_1 was the regression coefficient for DIM_k; and DIM_k was a covariate

describing the effect of days k in milk. Wilmink adjustment $(\exp^{-0.05 \times \text{DIM}})$ was used for DIM, b_2 was the regression coefficient for the Wilmink adjustment; animal_l was the random additive genetic effect based on **G** of animal l with distribution $N\left(0, \mathbf{G}\sigma_a^2\right)$, and e_{ijkl} was the random residual effect, which was assumed to be normally distributed with $e \sim N\left(0, \mathbf{I}\sigma_e^2\right)$, where **G** is the genomic relationship matrix, **I** was the identity matrix, σ_a^2 was the genetic variation, and σ_e^2 was the residual variation.

The bivariate analyses were run for each milk protein analyzed in combination with protein percentage. For the multi-trait analysis, all 9 traits were fitted simultaneously. Correlations between traits were based on the multi-trait analyses.

Table 1 summarizes the descriptive statistics for the milk protein profile and the total milk protein percentage. Mean protein content in the sampled milk was 3.38%. The major proteins (α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, and β-LG) made up 83% of the total milk protein fraction. The caseins constituted 72.3% of the total protein, with β-CN and α_{S1}-CN alone contributing to 34.1 and 26.8% of the total milk protein, respectively. The whey proteins constituted 10.8% of the total protein.

The α_{S1}-CN-8P accounted for 19.2% of the total milk protein and 71.6% of the α_{S1}-CN fraction of the total protein percentage. This was comparable to previous findings, in which α_{S1}-CN-8P accounted for 21.3% of the total protein (Bijl et al., 2014) and 74% of the α_{S1}-CN (Heck et al., 2008) in the Dutch Holstein population.

Heritability values and standard errors of estimation from the bivariate and multi-trait models are given in Table 2. Generally, the heritability estimates for the milk proteins were moderate to high except for β-CN, which had the lowest estimates (0.01–0.05). Glycosyl-

Table 1. Descriptive statistics[1] of milk protein profile and the total milk protein percentage

Protein or fraction[2]	Mean (%)	CV (%)	5% quantile	95% quantile
α_{S1}-CN	26.8	9	25.5	28.1
α_{S1}-CN-8P	19.2	11	17.7	20.7
α_{S2}-CN	5.3	20	4.5	5.9
β-CN	34.1	10	31.5	36.7
κ-CN	6.1	18	5.3	6.9
Glycosylated κ-CN	1.7	47	1.2	2.0
α-LA	3.3	19	2.9	3.6
β-LG	7.5	21	6.5	8.4
Total protein (%)	3.38	9	3.20	3.55

[1]Mean = phenotypic mean of the trait.

[2]Protein composition was expressed as percentage fractions of the total milk protein percentage (wt/wt); total protein was expressed as percentage (%) of the total milk yield; individual proteins comprise only the peaks identified as intact proteins and isoforms; that is, α_{S1}-CN (comprises α_{S1}-CN 8P + 9P), α_{S2}-CN (comprises α_{S2}-CN 11P + 12P), β-CN (comprises β-CN 4P + 5P), and κ-CN (comprises κ-CN G + 1P), where P = phosphorylated serine group.

Table 2. Heritability values and standard errors of heritability estimates

	Bivariate model		Multi-trait model	
Trait	h^2	SE	h^2	SE
α_{S1}-CN-8P[1]	0.13	0.10	0.15	0.09
α_{S1}-CN	0.14	0.10	0.14	0.09
α_{S2}-CN	0.36	0.10	0.36	0.10
β-CN	0.01	0.07	0.05	0.08
Glycosylated κ-CN	0.44	0.10	0.44	0.09
κ-CN	0.79	0.09	0.78	0.08
α-LA	0.22	0.11	0.25	0.10
β-LG	0.56	0.11	0.54	0.10
Total protein (%)	0.53	0.11	0.59	0.10

[1]Where P = phosphorylated serine.

ated κ-CN (0.44) and α_{S2}-CN (0.36) had moderate heritability values estimated using the 2 models; κ-CN had the highest heritability estimates in both models (0.78–0.79).

Generally, standard errors were lower for heritability estimates using the multi-trait model for all traits except β-CN. The standard errors of estimation from the multi-trait model in the current study (0.08–0.10) were also lower compared with previous studies, including that of Schopen et al. (2009; 0.08–0.12), Buitenhuis et al. (2014; 0.12–0.21), and Bobe et al. (1999; 0.07–0.12). Given the observed medium to high genetic correlations between the proteins, the multi-trait analyses might have benefitted from use of information from correlated traits.

Heritability values estimated with the multi-trait model were comparable to estimates from previous studies. Intraherd heritability of 0.66 for total protein and 0.64 for κ-CN estimated by Schopen et al. (2009) were comparable to the current estimates of 0.59 and 0.78, respectively. The estimates for α_{S2}-CN (0.36) in the current study were also comparable to the earlier reported value of 0.30 (Ikonen et al., 1997) but lower than estimates of Schopen et al. (2009; 0.73). β-Casein had the lowest heritability estimate (0.05) in this study.

This was in agreement with previous estimates for β-CN by Buitenhuis et al. (2014; 0.05), but considerably lower than estimates by Ikonen et al. (1997; 0.33 to 0.40) and Schopen et al. (2009; 0.25).

Genetic and phenotypic correlations estimated using multi-trait analysis are given in Table 3. Generally, the milk protein compositions had low to medium genetic correlation with total protein (−0.02–0.32). Higher genetic (0.75) as well as phenotypic (0.61) correlations were also observed between the 2 whey proteins (β-LG and α-LA).

The genetic correlation between α_{S1}-CN-8P and β-LG was low (−0.01). Bijl et al. (2014) have previously reported a distinct effect of β-LG protein variants and β-LG concentration on α_{S1}-CN-8P concentration. Nonetheless, according to Bijl et al. (2014), the mechanism behind the established association between β-LG protein variants, as well as β-LG concentration and α_{S1}-CN-8P concentration remains unclear, calling for further study.

Genetic correlations between the milk proteins and the total milk protein percentage were generally low, except for α_{S1}-CN-8P (0.38) and β-LG (0.27).

Our results suggest that genetic parameters can be estimated with high accuracy for scarcely recorded

Table 3. Genetic (above diagonal) and phenotypic (below diagonal) correlations[1]

Trait[2]	α_{S1}-CN-8P	α_{S1}-CN	α_{S2}-CN	β-CN	Glyc κ-CN	κ-CN	α-LA	β-LG	Total protein (%)
α_{S1}-CN-8P		0.80	−0.06	−0.49	−0.39	−0.19	−0.43	−0.01	0.38
α_{S1}-CN	0.86		−0.45	−0.18	−0.33	−0.33	−0.32	−0.09	0.20
α_{S2}-CN	0.29	0.12		−0.70	−0.07	−0.21	0.17	0.08	0.16
β-CN	0.42	0.48	0.04		0.28	0.38	0.06	0.04	−0.10
Glyc κ-CN	−0.04	0.03	0.02	0.05		0.81	0.17	−0.04	−0.03
κ-CN	0.20	0.16	0.05	0.22	0.68		0.01	−0.02	0.12
α-LA	0.10	0.09	0.09	0.09	−0.02	0.06		0.75	0.06
β-LG	0.01	0.01	0.04	−0.01	−0.06	−0.01	0.61		0.27
Total protein (%)	0.17	0.02	0.12	−0.05	0.09	0.18	−0.04	0.14	

[1]SE (0.06–0.31).

[2]P = phosphorylated serine; Glyc = glycosylated.

traits using multi-trait analysis with genomic relationships between animals. Lower genetic correlations between the milk proteins with total protein percentage reported in this study also suggest that altering milk protein compositions through selective breeding might have little or no effect on the total protein percentage.

ACKNOWLEDGMENTS

REFERENCES

Bijl, E., H. J. F. van Valenberg, T. Huppertz, A. C. M. van Hooijdonk, and H. Bovenhuis. 2014. Phosphorylation of α_{S1}-casein is regulated by different genes. J. Dairy Sci. 97:7240–7246.

Bittante, G., M. Penasa, and A. Cecchinato. 2012. Invited review: Genetics and modeling of milk coagulation properties. J. Dairy Sci. 95:6843–6870.

Bobe, G., D. C. Beitz, A. E. Freeman, and G. L. Lindberg. 1999. Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. J. Dairy Sci. 82:2797–2804.

Bonfatti, V., A. Cecchinato, L. Gallo, A. Blasco, and P. Carnier. 2011. Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. J. Dairy Sci. 94:5183–5193.

Buitenhuis, A. J., N. A. Poulsen, and L. B. Larsen. 2014. Estimation of genetic parameters for the protein profile in Danish Holstein milk. Proc. 10th World Congr. Genet. Appl. Livest. Prod. https://asas.org/docs/default-source/wcgalp-posters/614_paper_9289_manuscript_543_0.pdf?sfvrsn=2.

Calus, M. P. L., and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43:26.

Cassandro, M., A. Comin, M. Ojala, R. Dal Zotto, M. De Marchi, L. Gallo, P. Carnier, and G. Bittante. 2008. Genetic parameters of milk coagulation properties and their relationships with milk yield and quality traits in Italian Holstein cows. J. Dairy Sci. 91:371–376.

Heck, J. M. L., C. Olieman, A. Schennink, H. J. F. van Valenberg, and M. H. P. W. Visker. 2008. Estimation of variation in concentration, phosphorylation and genetic polymorphism of milk proteins using capillary zone electrophoresis. Int. Dairy J. 18:548–555.

Howie, B., J. Marchini, and M. Stephens. 2011. Genotype imputation with thousands of genomes. G3 (Bethesda) 1:457–470. http://dx.doi.org/http://dx.doi.org/10.1534/g3.111.001198.

Ikonen, T., K. Ahlfors, R. Kempe, M. Ojala, and O. Ruottinen. 1999. Genetic parameters for the milk coagulation properties and prevalence of noncoagulating milk in Finnish dairy cows. J. Dairy Sci. 82:205–214.

Ikonen, T., M. Ojala, and E.-L. Syväoja. 1997. Effects of composite casein and beta-lactoglobulin genotypes on renneting properties and composition of bovine milk by assuming an animal model. Agric. Food Sci. Finl. 6:283–294.

Jensen, H. B., N. A. Poulsen, K. K. Andersen, M. Hammershoj, and H. D. Poulsen. 2012. 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.

Madsen, P., and J. Jensen. 2010. A User's Guide to DMU. Version 6, Release 5.0. University of Aarhus, Faculty Agricultural Sciences (DJF), Department of Genetics and Biotechnology, Research Centre Foulum, Tjele, Denmark.

Schopen, G. C., J. M. Heck, H. Bovenhuis, M. H. Visker, H. J. van Valenberg, and J. A. van Arendonk. 2009. Genetic parameters for major milk proteins in Dutch Holstein-Friesians. J. Dairy Sci. 92:1182–1191.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.