J. Dairy Sci. 100:6371–6375
https://doi.org/10.3168/jds.2017-12663

Short communication: Genetic parameters for milk protein composition predicted using mid-infrared spectroscopy in the French Montbéliarde, Normande, and Holstein dairy cattle breeds

M. P. Sanchez,*[1] M. Ferrand,† M. Gelé,† D. Pourchet,‡ G. Miranda,* P. Martin,* M. Brochard,† and D. Boichard*
*Génétique Animale et Biologie Intégrative, INRA, AgroParisTech, Université Paris Saclay, F-78350 Jouy-en-Josas, France
†Institut de l'Elevage, F-75012 Paris, France
‡Conseil Elevage 25-90, F-25640 Roulans, France

ABSTRACT

Genetic parameters for the major milk proteins were estimated in the 3 main French dairy cattle breeds (i.e. Montbéliarde, Normande, and Holstein) as part of the PhénoFinlait program. The 6 major milk protein contents as well as the total protein content (PC) were estimated from mid-infrared spectrometry on 133,592 test-day milk samples from 20,434 cows in first lactation. Lactation means, expressed as a percentage of milk (protein contents) or of protein (protein fractions), were analyzed with an animal mixed model including fixed environmental effects (herd, year × month of calving, and spectrometer) and a random genetic effect. Genetic parameter estimates were very consistent across breeds. Heritability estimates (h^2) were generally higher for protein fractions than for protein contents. They were moderate to high for α_{S1}-casein, α_{S2}-casein, β-casein, κ-casein, and α-lactalbumin ($0.25 < h^2 < 0.72$). In each breed, β-lactoglobulin was the most heritable trait ($0.61 < h^2 < 0.86$). Genetic correlations (r_g) varied depending on how the percentage was expressed. The PC was strongly positively correlated with protein contents but almost genetically independent from protein fractions. Protein fractions were generally in opposition, except between κ-casein and α-lactalbumin ($0.39 < r_g < 0.46$) and κ-casein and α_{S2}-casein ($0.36 < r_g < 0.49$). Between protein contents, r_g estimates were positive, with highest values found between caseins ($0.83 < r_g < 0.98$). In the 3 breeds, β-lactoglobulin was negatively correlated with caseins ($-0.75 < r_g < -0.08$), in particular with κ-casein ($-0.75 < r_g < -0.55$). These results, obtained from a large panel of cows of the 3 main French dairy cattle breeds, show that routinely collected mid-infrared spectra could be used to modify milk protein composition by selection.

Received January 30, 2017.
Accepted April 29, 2017.
[1]Corresponding author: marie-pierre.sanchez@inra.fr

Key words: dairy cattle, mid-infrared spectrometry, protein composition, genetic parameters

Short Communication

In cattle, the relative proportions of proteins in milk play a key role in determining the functional properties of milk, such as clotting and cheese yield (Wedholm et al., 2006). Accurate genetic analyses of milk protein composition require large-scale studies, but reference methods such as capillary zone electrophoresis are time consuming and expensive. They have therefore only been applied to small or moderate numbers of milk samples. To date, the 2 most important studies aiming to estimate genetic parameters of milk protein composition traits measured by reference methods included 1,940 Dutch Holstein-Friesian cows (Schopen et al., 2009) and 2,167 Simmental cows (Bonfatti et al., 2011a). More recently, Gebreyesus et al. (2016) used genomic relationships between 650 Danish Holstein cows to estimate genetic parameters for milk protein composition. Mid-infrared (**MIR**) spectrometry has been shown to be useful to predict milk protein composition (De Marchi et al., 2009; Bonfatti et al., 2011b; Rutten et al., 2011; Ferrand et al., 2012; Samore et al., 2012) and offers an alternative method for large-scale analyses.

PhénoFinlait, a major project implemented to study milk in dairy cattle, sheep, and goats (Gelé et al., 2014) aimed, among other objectives, to dissect the genetic architecture of individual milk protein composition. In cattle, MIR predictive equations were derived from 450 reference samples analyzed using reverse-phase (**RP**) HPLC. The equations were applied to the MIR spectra routinely collected in Montbéliarde (**MO**), Normande (**NO**), and Holstein (**HO**) French dairy breeds (Ferrand et al., 2012). Concentrations of the 6 major milk proteins (α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, and β-LG) were predicted with satisfactory accuracy (Sanchez et al., 2016). A genetic analysis of milk protein

Table 1. Milk protein composition: accuracy of mid-infrared (MIR) predictions (g/100 g of milk) and means ± SD as a percentage of milk or as a percentage of proteins in the Montbéliarde (MO), Normande (NO), and Holstein (HO) breeds

	Accuracy of MIR predictions[1]				g/100 g of milk			g/100 g of protein		
Trait	R^2_{val}	RE	RMSEP	RPD	MO	NO	HO	MO	NO	HO
PC[2]	1.00	0.73	0.025	14.1	3.4 ± 0.4	3.6 ± 0.4	3.3 ± 0.4	—	—	—
α-LA	0.59	14.4	0.020	1.6	0.14 ± 0.02	0.15 ± 0.02	0.14 ± 0.02	4.07 ± 0.28	4.16 ± 0.36	4.27 ± 0.42
β-LG	0.74	11.7	0.044	2.0	0.28 ± 0.05	0.28 ± 0.05	0.28 ± 0.05	8.25 ± 1.12	7.94 ± 1.03	8.46 ± 1.17
α_{S1}-CN	0.88	4.7	0.046	2.9	0.94 ± 0.10	0.99 ± 0.10	0.92 ± 0.11	27.8 ± 0.55	27.8 ± 0.68	27.9 ± 0.69
α_{S2}-CN	0.82	7.5	0.024	2.4	0.32 ± 0.04	0.35 ± 0.04	0.32 ± 0.04	9.53 ± 0.30	9.89 ± 0.33	9.69 ± 0.39
β-CN	0.92	3.7	0.044	3.5	1.24 ± 0.11	1.29 ± 0.11	1.20 ± 0.13	36.6 ± 0.88	36.2 ± 1.2	36.2 ± 1.2
κ-CN	0.80	8.4	0.038	2.2	0.33 ± 0.05	0.35 ± 0.05	0.31 ± 0.05	9.75 ± 0.60	9.87 ± 0.48	9.43 ± 0.58

[1]R^2_{val} = coefficient of determination; RE = relative error; RMSEP = root mean squared error of prediction; and RPD = ratio of prediction to deviation, calculated on MIR predictions in the validation set (n = 133) as g/100 g of milk.
[2]PC = total milk protein.

composition was therefore carried out in the 3 French dairy cattle breeds using a very large data set and for the first time in MO and NO breeds.

We herein report the estimation of genetic parameters for the 6 major milk proteins, using 133,592 test-day records in first lactation from 8,477 MO, 6,253 NO, and 5,734 HO cows.

The MIR spectra of 848,068 milk samples from 156,660 MO, NO, and HO cows were collected between November 2009 and August 2012 during the Phéno-Finlait program using MIR spectrometry with defined routine Fourier transform MIR analyses (MilkoScan FT6000, Foss Electric A/S, Hillerød, Denmark). The samples were distributed across 1,043 herds, covering a broad range of geographical locations (16 small regions) and production systems (grass or maize silage, high- or low-input, conventional or organic, and so on).

A total of 450 cow milk reference samples of the 3 breeds were analyzed using RP-HPLC. Equations were derived from these samples to predict total milk protein content (**PC**) and milk protein composition (Ferrand et al., 2012). Outliers were removed from reference data using the Grubbs test (Grubbs, 1969). Samples were then randomly assigned to either the calibration (70%) or validation (30%) set. Only the wavelengths not spoiled by water molecules were used (i.e., 446 wavelengths following the recommendations of the MilkoScan FT600 manufacturer), and the most informative wavelengths were selected using genetic algorithms (Ferrand-Calmels et al., 2014). Moreover, to improve the robustness of equations, calibration samples with a studentized residual greater than 2.58 were considered as outliers and deleted. Final calibration and validation sets contained 311 and 133 samples, respectively. Individual protein contents were predicted for the 6 main milk proteins: α-LA and β-LG whey proteins; and α_{S1}-CN, α_{S2}-CN, β-CN, and κ-CN, and expressed as grams per 100 g of milk (protein contents). Individual protein fractions, as grams per 100 g of protein, were then calculated by dividing predicted protein contents by PC. Means and standard deviations, as well as prediction accuracies obtained for milk protein contents of the validation set, are detailed in Table 1. As expected, PC was well predicted ($R^2 = 1$). The accuracy of content traits was high for caseins ($0.80 \leq R^2 \leq 0.92$) and moderate for α-LA ($R^2 = 0.59$) and β-LG ($R^2 = 0.74$). A total of 13 traits were therefore analyzed: PC, the 6 protein contents, and the 6 protein fractions.

For each trait, the phenotype of each cow was defined as the average test-day measures during the first lactation per cow. The variance components were estimated within-breed by REML with the procedure described by Meyer (1985) and implemented as in Boichard et al. (1989) with the following animal model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e}, \quad [1]$$

where $\mathbf{y}$ is the vector of phenotypes; $\boldsymbol{\beta}$ a vector of fixed effects; $\mathbf{a} \sim N(0, \mathbf{A} \otimes \mathbf{G}_0)$ is the vector of random genetic effects; $\mathbf{e} \sim N(0, \mathbf{I} \otimes \mathbf{R}_0)$ is the vector of random residual effects. $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices common to all traits. $\mathbf{A}$ is the relationship matrix among individuals, $\mathbf{I}$ is the identity matrix, $\mathbf{G}_0$ is the 13 × 13 matrix of genetic covariances, and $\mathbf{R}_0$ is the matrix of residual covariances.

Only first lactation records with at least 7 test-day measurements in MO and at least 3 test-day measurements in NO and HO were included in the analyses. They corresponded to 72,561, 31,189, and 29,842 test-day record data from 8,477 MO, 6,253 NO, and 5,734 HO cows, respectively. Fixed effects included in the model were herd (944 in MO, 398 in NO, and 390 in HO), year × month of calving (12 in MO, 15 in NO, and 14 in HO), and spectrometer (1 in MO, 3 in NO, and 4 in HO). Pedigrees were traced over 3 generations and contained 23,956 individuals in MO, 17,376 in NO, and 15,895 in HO.

For all milk composition traits analyzed, heritability values were medium to high and generally higher for protein fractions than for protein contents (Table 2). Only one previous study compared heritability values of milk protein fractions and contents measured by RP-HPLC (Bonfatti et al., 2011a) and the authors also observed that protein fractions were more heritable than protein contents. For each trait, heritability estimates were similar in NO and HO and ranged from 0.27 to 0.72. Estimates were higher in MO (0.61–0.86) because of the higher test-day number in this breed (8.6) than in NO (5) and HO (5.2), resulting in a lower residual variance. In the 3 breeds, β-LG was the most heritable trait, with estimates ranging from 0.71 to 0.79 for β-LG fraction and from 0.61 to 0.86 for β-LG content. These results are consistent with those (0.80) obtained in Dutch Holstein Friesian as a percentage of protein (Schopen et al., 2009) and higher than those (0.54) found in Danish Holstein as a percentage of protein (Gebreyesus et al., 2016) or in Simmental (Bonfatti et al., 2011a) as a percentage of protein (0.34) and as a percentage of milk (0.37). Generally, heritability values obtained in our study from MIR spectra were similar or even higher than values estimated from reference analyses, except for α_{S2}-CN. For this casein, Schopen et al. (2009) have found a high heritability value (0.73), whereas in our study, estimations were weak to moderate (from 0.25 to 0.58, depending on the breed) but comparable or higher than values found by Bonfatti et al. (2011a) or Gebreyesus et al. (2016). It should be emphasized, however, that estimates based on reference measures were obtained with one unique measure, whereas our results are on a lactation level and used several measures per lactation. The loss of accuracy due to MIR prediction is compensated for by repeated measures.

Genetic correlations were similar across breeds (Table 3). Phenotypic correlations, not shown here, were comparable to the genetic correlations. As expected, PC was strongly genetically correlated with protein contents (g/100 g of milk). Higher genetic correlations were observed between PC and casein contents than between PC and whey protein contents. Conversely, PC was relatively genetically independent from protein fractions (g/100 g of protein). Only a moderate negative genetic correlation, ranging from −0.34 to −0.42 according to the breed, was observed between PC and α-LA. Protein contents were always positively correlated. Highest estimates were obtained between caseins (0.83–0.98), and the smallest estimates were found between β-LG and the other proteins. In contrast, protein fractions showed genetic correlations close to zero or negative in most cases. Only κ-CN was positively correlated with α-LA (0.39–0.46) and α_{S2}-CN (0.36–0.49). In the 3 breeds, β-LG was negatively correlated with all caseins, moderately with α_{S1}-CN, α_{S2}-CN, and β-CN (from −0.08 to −0.33), and more strongly with κ-CN (from −0.55 to −0.75). These results suggest that β-LG could share genetic regulatory pathways with caseins and in particular with κ-CN. Bonfatti et al. (2011a) consistently found positive genetic correlations between protein contents and negative ones between protein fractions. However, with MIR predictions, we generally found stronger genetic correlations than those reported by Bonfatti et al. (2011a) with milk protein composition measured by RP-HPLC. It should also be emphasized that protein fractions sum to 1 and this automatically tends to generate negative covariances between the main components.

Our results show that genetic parameters for milk protein composition predicted from MIR spectra were globally consistent with previously reported results from reference methods. We have found medium to high heritabilities for all protein contents/fractions with maximal values obtained for β-LG in spite of its medium MIR prediction accuracy ($R^2 = 0.74$). The MIR prediction accuracies are therefore high enough to undertake genetic investigations. In addition, this study, carried out at a large scale in the 3 main French dairy cattle breeds, shows that (1) the total milk protein content is relatively genetically independent from whey protein and casein fractions, (2) β-LG is highly heritable, and (3) β-LG and casein contents are positively correlated but the corresponding fractions exhibit negative genetic

Table 2. Heritability[1] values for milk protein contents (g/100 g of milk) and fractions (g/100 g of protein) in the Montbéliarde (MO), Normande (NO), and Holstein (HO) breeds

Item	Breed	PC[2]	α-LA	β-LG	α_{S1}-CN	α_{S2}-CN	β-CN	κ-CN
g/100 g of milk	MO	0.57	0.57	0.86	0.54	0.54	0.66	0.48
	NO	0.41	0.42	0.61	0.42	0.38	0.43	0.42
	HO	0.27	0.31	0.61	0.30	0.29	0.27	0.32
g/100 g of protein	MO	—	0.72	0.79	0.67	0.58	0.42	0.61
	NO	—	0.53	0.72	0.57	0.25	0.39	0.55
	HO	—	0.44	0.71	0.53	0.31	0.39	0.54

[1]Standard errors ranged from 0.02 to 0.03.

[2]PC = total milk protein.

Table 3. Genetic correlation values[1] for milk protein contents (g/100 g of milk, above diagonal) and fractions (g/100 g of protein; below diagonal) in the Montbéliarde (MO), Normande (NO), and Holstein (HO) breeds

Item	Breed	PC[2]	α-LA	β-LG	α_{S1}-CN	α_{S2}-CN	β-CN	κ-CN
PC	MO		0.72**	0.52**	0.98***	0.97***	0.99***	0.85***
	NO		0.70**	0.61**	0.99***	0.99***	0.98***	0.90***
	HO		0.73**	0.45*	0.98***	0.98***	0.98***	0.87***
α-LA	MO	−0.34*		0.18	0.69**	0.72**	0.70**	0.80***
	NO	−0.42*		0.27*	0.71**	0.69**	0.68**	0.74**
	HO	−0.38*		−0.01	0.72**	0.71**	0.75**	0.80***
β-LG	MO	0.04	−0.27*		0.51**	0.46*	0.47*	0.10
	NO	0.10	−0.33*		0.59**	0.57**	0.58**	0.31*
	HO	−0.08	−0.52**		0.39*	0.39*	0.39*	0.07
α_{S1}-CN	MO	0.04	0.01	−0.13		0.94***	0.96***	0.84***
	NO	0.25*	0.11	−0.10		0.98***	0.95***	0.89***
	HO	−0.18	−0.01	−0.19		0.96***	0.94***	0.87***
α_{S2}-CN	MO	0.00	0.05	−0.08	−0.33*		0.95***	0.88***
	NO	−0.17	−0.04	−0.14	−0.07		0.96***	0.92***
	HO	−0.26*	−0.09	−0.22*	0.04		0.95***	0.88***
β-CN	MO	−0.08	0.06	−0.33*	−0.26*	−0.37*		0.83***
	NO	−0.09	0.01	−0.25*	−0.44*	−0.51**		0.85***
	HO	0.07	0.25*	−0.20*	−0.48*	−0.37*		0.84***
κ-CN	MO	0.03	0.45*	−0.75**	0.02	0.38*	−0.11	
	NO	0.06	0.39*	−0.55**	0.02	0.49*	−0.36*	
	HO	0.17	0.46*	−0.66**	0.14	0.36*	−0.31*	

[1]Absolute genetic correlation values ($|r_g|$) are indicated by asterisks as follows: ***($|r_g| \geq 0.80$), **($0.50 \leq |r_g| < 0.80$), or *($0.20 \leq |r_g| < 0.80$); SE ranged from 0.04 to 0.08.

[2]PC = total milk protein.

correlations. Several options appear to be possible for selection (e.g., increase concentration of all proteins; increase concentration of caseins; or increase the casein/total protein ratio, leading to a decrease in the proportion of β-LG). The PC is already included in the breeding objectives of the 3 breeds. Reorienting the selection effort to casein concentration would be sound and easy to implement, at least in the breeds with a dominant cheese orientation. The interest of increasing the casein fraction and, therefore, decreasing the β-LG fraction (which is the major and most variable whey protein) should be investigated in more detail. Whatever the definition of the breeding goal, applying prediction equations to routinely collected MIR spectra provides new opportunities to modify milk protein composition and select for better cheese-making abilities.

ACKNOWLEDGMENTS

The PhénoFinlait program has received financial support from Agence Nationale de la Recherche (ANR-08-GANI-034 Lactoscan), APIS-GENE, Compte d'Affectation Spéciale Dévelopment Agricole et Rural, Centre National Interprofessionnel de l'Economie Laitière, FranceAgriMer, France Génétique Elevage, and the French Ministry of Agriculture (all from Paris, France). The authors thank the breeders who participated in PhénoFinlait; colleagues from Institut de l'Elevage and INRA who designed and coordinated farm samples and data collection; the partners of the project, laboratories, manufacturers, and DHI organizations that provided data; and the members of the PhénoFinlait scientific committee who advised and managed this program.

REFERENCES

Boichard, D., N. Bouloc, G. Ricordeau, A. Piacere, and F. Barillet. 1989. Genetic-parameters for 1st lactation dairy traits in the Alpine and Saanen goat breeds. Genet. Sel. Evol. 21:205–215. https://doi.org/10.1186/1297-9686-21-2-205.

Bonfatti, V., A. Cecchinato, L. Gallo, A. Blasco, and P. Carnier. 2011a. Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. J. Dairy Sci. 94:5183–5193. https://doi.org/10.3168/jds.2011-4297.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011b. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. J. Dairy Sci. 94:5776–5785. https://doi.org/10.3168/jds.2011-4401.

De Marchi, M., V. Bonfatti, A. Cecchinato, G. Di Martino, and P. Carnier. 2009. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. Ital. J. Anim. Sci. 8(S2):399–401. https://doi.org/10.4081/ijas.2009.s2.399.

Ferrand, M., G. Miranda, S. Guisnel, H. Larroque, O. Leray, F. Lahalle, M. Brochard, and P. Martin. 2012. Determination of protein composition in milk by mid-infrared spectrometry. Pages 41–45 in Proc. International Strategies and New Developments in Milk Analysis. VI ICAR Reference Laboratory Network Meeting, Cork, Ireland.

Ferrand-Calmels, M., I. Palhiere, M. Brochard, O. Leray, J. Astruc, M. Aurel, S. Barbey, F. Bouvier, P. Brunschwig, H. Caillatt, M. Douguet, F. Faucon-Lahalle, M. Gele, G. Thomas, J. Trommenschlager, and H. Larroque. 2014. Prediction of fatty acid profiles

in cow, ewe, and goat milk by mid-infrared spectrometry. J. Dairy Sci. 97:17–35. https://doi.org/10.3168/jds.2013-6648.

Gebreyesus, G., M. S. Lund, L. Janss, N. A. Poulsen, L. B. Larsen, H. Bovenhuis, and A. J. Buitenhuis. 2016. Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the Danish Holstein. J. Dairy Sci. 99:2863–2866. https://doi.org/10.3168/jds.2015-10501.

Gelé, M., S. Minery, J. M. Astruc, P. Brunschwig, M. Ferrand, G. Lagriffoul, H. Larroque, J. Legarto, P. Martin, G. Miranda, I. Palhière, P. Trossat, and M. Brochard. 2014. Phénotypage et génotypage à grande échelle de la composition fine des laits dans les filières bovine, ovine et caprine. Prod. Anim. 27:255–268.

Grubbs, F. 1969. Procedures for detecting outlying observations in samples. Technometrics 11:1–21.

Meyer, K. 1985. Maximum-likelihood estimation of variance-components for a multivariate mixed model with equal design matrices. Biometrics 41:153–165.

Rutten, M., H. Bovenhuis, J. Heck, and J. van Arendonk. 2011. Predicting bovine milk protein composition based on Fourier transform infrared spectra. J. Dairy Sci. 94:5683–5690. https://doi.org/10.3168/jds.2011-4520.

Samore, A., F. Canavesi, A. Rossoni, and A. Bagnato. 2012. Genetics of casein content in Brown Swiss and Italian Holstein dairy cattle breeds. Ital. J. Anim. Sci. 11:196–202. https://doi.org/10.4081/ijas.2012.e36.

Sanchez, M., A. Govignon-Gion, M. Ferrand, M. Gele, D. Pourchet, Y. Amigues, S. Fritz, M. Boussaha, A. Capitan, D. Rocha, G. Miranda, P. Martin, M. Brochard, and D. Boichard. 2016. Whole-genome scan to detect quantitative trait loci associated with milk protein composition in 3 French dairy cattle breeds. J. Dairy Sci. 99:8203–8215. https://doi.org/10.3168/jds.2016-11437.

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. https://doi.org/10.3168/jds.2008-1281.

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.