J. Dairy Sci. 100:1–5
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.