Overview and results of PhenoFinlait, a large scale project for milk fat and protein composition analysis

Boichard D.[1], Martin P. [1], Miranda G. [1], Ferrand M.[2], Brunschwig P.[3], Govignon-Gion A. [1], Larroque H.[4], Rupp R. [4], Palhière I. [4], Chantry-Darmon C.[5], Lagriffoul G. [6], Lahalle F. [2,7], Fritz S.[8], Brochard M. [2]

[1] INRA, UMR1313 GABI, 78350 Jouy en Josas, France
[2] Institut de l'Elevage, 149 rue de Bercy, 75595 Paris, France
[3] Institut de l'Elevage, 9 rue André Brouard, 49105 Angers, France
[4] INRA, UR631, SAGA, 31326 Castanet-Tolosan, France
[5] LABOGENA, Domaine de Vilvert,78350 Jouy en Josas, France
[6] Institut de l'Elevage, 31326 Castanet-Tolosan, France
[7] CNIEL, 42 rue de Chateaudun, 75009 Paris, France
[8] UNCEIA, 149 rue de Bercy, 75595 Paris, France

PhenoFinlait (which stands for Milk Fine Phenotyping) is a large scale research and development project aiming at monitoring and controlling composition in milk fatty acids (FA) and proteins in French dairy cattle, sheep, and goat. The central idea is to estimate milk components from mid infra red (MIR) spectra. This convenient approach could be generalized to the whole population by using routinely collected spectra produced by milk analysis laboratories. In the first step, calibration equations of milk FA and proteins were derived from experimental samples selected from their large variability and analyzed both with MIR and a reference method. For FA, accurate estimations were obtained for ~25 FA or combinations. Results were more accurate in cattle (due to a larger calibration sample) and in sheep (due to a higher fat concentration in milk) than in goat. For protein, no universal reference method is recognized. We developed a dedicated method combining liquid chromatography and mass spectrometry. This approach is highly resolutive and makes it possible to determine the nature and to estimate the quantity of each component, including genetic, splicing, or post-traductional variants and degradation products. These reference measurements were also used to derive calibration equations for proteins. Again, results were quite accurate in sheep, due to the high milk concentration. Raw results were disappointing for cattle but were consistently improved for caseins by accounting for proteolysis, leading to acceptable results.

In the second step, large scale sampling was organized in over 1,500 herds and flocks. Over 800,000 spectra, 30,000 milk samples, 20,000 blood samples were collected, whereas information about diet and production system was collected in 8,000 surveys. Additional information was obtained from the national database. A standardization procedure was implemented with reference milk samples analyzed monthly in milk analysis laboratories, in order to adjust for the instrument drift. Equations were applied to these spectra to estimate the corresponding profiles in FA and for proteins.

Phenotypic analyses were carried out, illustrating the strong effect of the diet on milk FA, which will be used to improve dairy herd management. Genetic parameters were estimated in Holstein, Montbéliarde, and Normande cattle and in goats and were fairly consistent across populations. Heritability estimates were higher for components expressed in %milk than in %fat, and for saturated FA (SFA) than for unsaturated (UFA). Genetic correlations were high along the lactation, showing that their genetic determinism is fairly constant except in the beginning of the lactation affected by body fat mobilization. SFA were genetically highly correlated, as well as UFA. A strong opposition was

found between SFA and UFA expressed in %fat, but this opposition was much more limited when traits were expressed in %milk. Individual test-day records were adjusted for environmental effects after a polygenic evaluation and then averaged per cow. These combined phenotypes were used for QTL detection. 8100 cows, 1650 ewes, and 2300 goats were genotyped mainly with Illumina 50k chips (500 cows were also genotyped with the LD chip and then imputed). After quality control, phasing and imputation, QTL detection was carried out within population by Linkage and Linkage Disequilibrium Analysis (LDLA) and association study. In cattle, the main regions were found on already described regions known to affect fat components, milk production, or fatty acid desaturation (on chromosomes 14, 19, or 26). Several dozens of additional QTL were found for each trait and overlapped only very partially across breeds. Most QTL were found to affect several traits simultaneously, in agreement with the genetic correlations and the metabolic pathways. This large database and sample library will be the basic material of many forthcoming analyses.