Predicting milk phosphorus content based on genotypic and milk infrared data

H. Bovenhuis[1], I. Jibrila[1] and J. Dijkstra[2]
[1]Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH, Wageningen, the Netherlands
[2]Animal Nutrition Group, Wageningen University & Research, P.O. Box 338, 6700 AH, Wageningen, the Netherlands

Introduction

Concerns about the environmental burden caused by nutrient leaching from agricultural soils into the environment makes it necessary to strive for more efficient phosphorus (**P**) utilization on dairy farms. Improving P efficiency in intensive dairy farming would not only benefit the environment, but is also essential in view of scarcer global mineral P resources during the coming decades. Several surveys revealed that dietary P of the total ration fed to dairy cattle are well in excess regarding the need of dairy cattle (e.g. Van Krimpen et al. 2012; Kebreab et al. 2008). Reasons for high P levels in the diet are partly due to the inability to accurately target the requirements of individual animals, which results in the use of sizeable safety margins. Recommended levels of P in the diet for dairy cattle assume a fixed content of P in milk: the milk P content in NRC (2001) recommendations is 0.90 g/kg milk, and in the Netherlands 0.97 g/kg milk is assumed. However, it was shown that milk P content of heifers is highly variable and ranged from 0.84 to 1.24 g/kg milk (Van Hulzen et al., 2009). Accounting for differences in milk P content would allow farmers to better estimate P requirements of individual cows and in this way improve P efficiency.
Milk P content can be accurately estimated through the use of e.g. Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES, e.g. Soyeurt et al., 2009; Toffanin et al., 2015; Bonfatti et al., 2016). This method is however expensive, and therefore not suitable for large scale application. Some studies estimated milk P content based on milk infrared data (Soyeurt et al., 2009; Toffanin et al., 2015; Bonfatti et al., 2016). However, these studies were based on small sample sizes and therefore need to be confirmed.
Some genomic regions have been identified which are associated with milk P content. They include the DGAT1 K232A polymorphism on BTA14 (Bovenhuis et al., 2016) and on BTA1 SNP rs29019625, which is close to the gene SLC37A1, a phosphorous antiporter (Buitenhuis et al. 2015; Kemper et al., 2016). Therefore genotypic information also might be used to predict milk P content. This study was aimed at predicting milk P content based on routinely recorded milk composition traits, genotypic data and milk infrared spectra.

Material and Methods

Data

Animals used in this study were part of the Dutch Milk Genomics Initiative. One test day morning milk sample from 2,000 first parity Dutch Holstein-Friesian cows, which were between 63 and 282 days in milk, and located on 398 herds throughout the Netherlands, was collected between February and March 2005. All cows had at least 87.5% Holstein-Friesian genes. Details

of the study can be found in Stoop et al. (2008).
The infrared spectra consisted of the transmittance values measured at 1,060 wavenumbers ranging from 925 to 5,008 cm−1. FTIR spectra were recorded in a 10-mL milk sample using the MilkoScan FT 6000 equipment (Foss, Hillerod, Denmark) at the certified laboratory of the Dutch Milk Control Station (Zutphen, the Netherlands). All milk samples used in this study were analyzed on the same MilkoScan FT 6000. More details can be found in Wang et al. (2016). Milk fat, protein and lactose were obtained from routine milk recordings and are based on these infrared spectroscopic data.
The P content of the milk samples (in mg/kg milk) was determined using ICP-AES, as described by van Hulzen et al. (2009). P content of each milk sample was adjusted for the P content of a reference sample, which was analysed across different runs of ICP-AES.
The cows were genotyped for the DGTA1 K232A and rs29019625. These two polymorphisms are now known to have effects on milk P content (Bovenhuis et al., 2016; Kemper et al., 2016). DGTA1 K232A genotypes were obtained as described by Schennink et al. (2007). Genotypes for rs29019625 were determined using an Illumina Bovine 50K SNP chip, as described by Schopen et al. (2011).
Only samples with milk P content, genotypic information and Infrared spectra were included in this study. This resulted in 1389 samples. Outlying observations were identified using the Mahalanobis Distance. Observations with Mahalanobis Distance > 3 SD were removed, and the final dataset consisted of 1379 records.

Statistical analysis
The Partial Least Square Regression procedure of SAS (Proc PLS) was used for developing the prediction equations. Test set validation was used to determine the accuracies of the models. For this the final dataset was divided into a calibration set and a validation set. The training set consisted of 1000 observations and the test set consisted of the remaining 379 observations.
For each subset of explanatory variables, the selected model is the model with the smallest number of latent variables (#L), yet with Root Mean Prediction Residual Sum of Squares (**RMPRESS**) that is insignificantly larger than the smallest possible RMPRESS for that subset, as suggested by van der Voet (1994). Across subsets of the explanatory variables, the models developed were compared using the validation coefficient of determination ($\mathbf{R^2_v}$).

Results and Discussion
Descriptive statistics for the milk constituents are shown in table 1. Milk P content (adjusted for the reference sample) had a mean of 1032 mg/kg and a SD of 93.5 mg/kg. The standard deviation for fat content (0.64) was considerably higher than the standard deviation for protein (0.29) or lactose content (0.13) which is in agreement with other studies (e.g. Samoré et al. 2012).

Table 1. *Descriptive statistics of milk constituents (n=1379)*

Variable	Mean	SD	5th percentile	95th percentile
P content (mg/kg milk)*	1032	93.5	886	1195
Fat content (%)	4.36	0.64	3.37	5.44
Protein content (%)	3.51	0.29	3.04	4.00
Lactose content (%)	4.64	0.13	4.42	4.85

* adjusted for the reference sample

The performance of the prediction models based on different information sources are shown in Table 2. Milk P content can be predicted based on milk fat content with accuracy of 18%. Based on milk protein content an accuracy of 41% was obtained. Milk lactose does not contain information on milk P content. All caseins are phosphorylated and possess one or multiple phosphoserine clusters and about 58% of milk P is found in the casein micelles (e.g. Bijl et al., 2013). This explains the relative high accuracy of predicting milk P content based on milk protein content. Adding fat and lactose content to a model that predicts milk P content based on protein content only marginally increases the accuracy of prediction. The predictive power of milk fat content works mainly indirectly, i.e. through the correlation between fat and protein content.

Table 2. *Prediction of milk P content based on different sets of explanatory variables*

Explanatory variable(s)	R^2_v (%)
Routinley recorded traits	
Fat content	18.4
Protein content	41.1
Lactose content	0.3
Genotypes	
DGAT1 K232A	8.7
BTA1 rs29019625 (BTA1)	4.7
Routinley recorded traits and Genotypes	
Fat-, protein- and lactose content and genotypes	49.4
Infrared wavenumbers	
1060 wavenumbers (excl. water absorption regions)	83.8

Milk P content is predicted with accuracy of 8.7% based on the DGAT1 K232A polymorphism and with accuracy of 4.7% by the SNP on BTA1 (rs29019625). Combining information from all SNP, i.e. genomic prediction, is expected to result in higher prediction accuracy but is bounded by the heritability; 53% of the total variation in milk P content can be explained by genetic factors (van Hulzen et al., 2009).

Milk fat-, protein- and lactose content are predicted based on infrared wavenumbers and therefore reflect part of the information captured by the infrared spectrum. Using the complete spectrum, i.e. all wavenumbers but excluding the water absorption regions, milk P content can be predicted with accuracy of about 84% (Table 2). Soyeurt et al. (2009) reported a prediction accuracy for milk P content of 85%, Toffanin et al. (2015) of 72% and Bonfatti et al. (2016) of 43%. Prediction accuracy for milk P content based on the complete spectra is more than twice as high as based on protein content only. Combining infrared data with genotypic information did not further improve the accuracy. This suggests that the genotypic information on milk P content is already captured by the infrared spectra. Based on a GWAS of individual infrared wavenumbers Wang and Bovenhuis (2017) identified a region on BTA1 containing SNP rs29019625. The region on BTA1 has been associated with milk phosphorus content in studies by Buitenhuis et al. (2015) and Kemper et al. (2016). These studies suggested the solute carrier

family 37 member 1 (SLC37A1) as the most likely candidate gene. SLC37A1 functions as a phosphorus antiporter (Chou et al., 2013), transporting glucose-6-phosphate and phosphorus in opposite directions. The infrared wavenumbers that enabled detection of this region are related to P=O chemical bonds. This indicates that the infrared spectrum contains direct information on milk P content.

Results from this study confirm the potential of milk infrared data to predict milk P content (Soyeurt et al., 2009; Toffanin et al., 2015; Bonfatti et al., 2016). Milk samples are routinely analyzed using infrared spectroscopy. Therefore, infrared prediction of milk P content can be easily implemented and used by farmers to feed their cows according P requirement. Phosphorus requirement for cows producing 9000 kg of milk per lactation can differ up to 3 kg due to differences in milk P content. We quantified that under Dutch farming conditions, feeding cows according P requirement based on the developed infrared prediction can improve phosphorus efficiency with 17%.

References

Bijl, E., van Valenberg, H. J. F., Huppertz, T., & C. M. van Hooijdonk. 2013. Protein, casein, and micellar salts in milk: current content and historical perspectives. Journal of Dairy Science, 96: 5455–5464.

Bonfatti, V., Degano, L., Menegoz, A., Carnier, P., Afseth, N. K., Martens, H. & M. Cassandro. 2016. Short communication: Mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental. Journal of Dairy Science, 99: 8216–8221.

Bovenhuis, H., Visker, M. H. P. W., Poulsen, N. A., Sehested, J., van Valenberg, H. J. F., van Arendonk, J.A.M, & A.J. Buitenhuis. 2016. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. Journal of Dairy Science, 1: 1–11.

Buitenhuis, B., N. A. Poulsen, L. B. Larsen, and J. Sehested. 2015. Estimation of genetic parameters and detection of quantitative trait loci for minerals in Danish Holstein and Danish Jersey milk. BMC genetics, 16: 52.

Kebreab, E., N.E. Odongo, B.W. McBride, M.D. Hanigan, and J. France. 2008. Phosphorus utilization and environmental and economic implications of reducing phosphorus pollution from Ontario dairy cows. Journal of Dairy Science 91: 241–246.

Kemper, K. E., M. D. Littlejohn, T. Lopdell, B. J. Hayes, L. E. Bennett, R. P. Williams, X. Q. Xu, P. M. Visscher, M. J. Carrick, and M. E. Goddard. 2016. Leveraging genetically simple traits to identify small-effect variants for complex phenotypes. BMC genomics. 17: 858.

NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th revised edition. Natl. Acad. Sci., Washington DC.

Samorè, A. B., 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.

Schennink, A., Stoop, W. M., Visker, M. H. P. W., Heck, J. M. L., Bovenhuis, H., Van Der Poel, J. J.,& J. A. M. Van Arendonk. 2007. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Animal Genetics, 38:467–473.

Schopen, G. C. B., Visker, M. H. P. W., Koks, P. D., Mullaart, E., van Arendonk, J. A. M., & H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy cattle. Journal of Dairy Science, 94: 3148–3158.

Soyeurt, H., Bruwier, D., Romnee, J.-M., Gengler, N., Bertozzi, C., Veselko, D., & P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared

spectrometry. Journal of Dairy Science, 92: 2444–2454.
Stoop, W. M., van Arendonk, J. A. M., Heck, J. M. L., van Valenberg, H. J. F., & Bovenhuis, H. 2008. Genetic parameters for major milk fatty acids and milk production traits of Dutch Holstein-Friesians. Journal of Dairy Science, 91: 385–394.
Toffanin, V., De Marchi, M., Lopez-Villalobos, N., & M. Cassandro. 2015. Effectiveness of mid-infrared spectroscopy for prediction of the contents of calcium and phosphorus, and titratable acidity of milk and their relationship with milk quality and coagulation properties. International Dairy Journal, 41:68–73.
van der Voet, H. 1994. Comparing the predictive accuracy of models using a simple randomization test. Chemometrics and Intelligent Laboratory Systems, 25: 313–323.
van Hulzen, K. J. E., Sprong, R. C., van der Meer, R., & J.A.M. van Arendonk. 2009. Genetic and nongenetic variation in concentration of selenium, calcium, potassium, zinc, magnesium, and phosphorus in milk of Dutch Holstein-Friesian cows. Journal of Dairy Science, 92: 5754-5759.
Van Krimpen, M.M., R.M.A. Goselink, J. Heeres, and A.W. Jongbloed 2012. Phosphorus requirements of dairy cattle, beef cattle, pigs and poultry: a review. Livestock Research Wageningen UR, Report 574.
Wang, Q., Hulzebosch, A., & H. Bovenhuis. 2016. Genetic and environmental variation in bovine milk infrared spectra. Journal of Dairy Science, 99: 6793–6803.
Wang, Q. & H. Bovenhuis. 2017. Genome wide association study for milk Fourier transform infrared wavenumbers. Submitted to Journal of Dairy Science