



J. Dairy Sci. 103

<https://doi.org/10.3168/jds.2018-16057>

© 2020, The Authors. Published by FASS Inc. and Elsevier Inc. on behalf of the American Dairy Science Association®.  
This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Genetic analysis of orotic acid predicted with Fourier transform infrared milk spectra

R. M. Zaalberg,<sup>1\*</sup>  A. J. Buitenhuis,<sup>1</sup>  U. K. Sundekilde,<sup>2</sup>  N. A. Poulsen,<sup>3</sup>  and H. Bovenhuis<sup>4</sup> 

<sup>1</sup>Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, DK-8830 Tjele, Denmark

<sup>2</sup>Department of Food Science, Aarhus University, Kirstinebjergvej 10, DK-5792 Årslev, Denmark

<sup>3</sup>Department of Food Science, Aarhus University, Blichers Allé 20, PO Box 50, DK-8830 Tjele, Denmark

<sup>4</sup>Animal Breeding and Genomics, Wageningen University and Research, PO Box 338, 6700AH, Wageningen, the Netherlands

### ABSTRACT

Fourier transform infrared spectral analysis is a cheap and fast method to predict milk composition. A not very well studied milk component is orotic acid. Orotic acid is an intermediate in the biosynthesis pathway of pyrimidine nucleotides and is an indicator for the metabolic cattle disorder deficiency of uridine monophosphate synthase. The function of orotic acid in milk and its effect on calf health, health of humans consuming milk or milk products, manufacturing properties of milk, and its potential as an indicator trait are largely unknown. The aims of this study were to determine if milk orotic acid can be predicted from infrared milk spectra and to perform a large-scale phenotypic and genetic analysis of infrared-predicted milk orotic acid. An infrared prediction model for orotic acid was built using a training population of 292 Danish Holstein and 299 Danish Jersey cows, and a validation population of 381 Danish Holstein cows. Milk orotic acid concentration was determined with nuclear magnetic resonance spectroscopy. For genetic analysis of infrared orotic acid, 3 study populations were used: 3,210 Danish Holstein cows, 3,360 Danish Jersey cows, and 1,349 Dutch Holstein Friesian cows. Using partial least square regression, a prediction model for orotic acid was built with 18 latent variables. The error of the prediction for the infrared model varied from 1.0 to 3.2 mg/L, and the accuracy varied from 0.68 to 0.86. Heritability of infrared orotic acid predicted with the standardized prediction model was 0.18 for Danish Holstein, 0.09 for Danish Jersey, and 0.37 for Dutch Holstein Friesian. We conclude that milk orotic acid can be predicted with moderate to good accuracy based on infrared milk spectra and that infrared-predicted orotic acid is heritable. The availability of a cheap and fast method to

predict milk orotic acid opens up possibilities to study the largely unknown functions of milk orotic acid.

**Key words:** spectroscopy, cattle, orotic acid, breed difference

### INTRODUCTION

Fourier transform infrared spectral analysis is a cheap and fast method for quantification of milk components. Infrared milk spectra are widely used in routine milk recording schemes to predict milk fat, protein, and lactose content (Luinge et al., 1993). Besides major milk components, infrared milk spectra have been proposed for prediction of components that occur in low concentrations, such as fatty acids (Rutten et al., 2009; Soyeurt et al., 2011), milk protein fractions (Rutten et al., 2011), and acetone and BHB (de Roos et al., 2007). Prediction of fat or protein composition is challenging, especially for components with low concentrations, yet the ability to predict acetone and BHB with an accuracy of 0.64 to 0.72 (de Roos et al., 2007) shows that infrared milk spectra have the potential to predict milk components that are present in lower concentrations.

A relatively unknown and not very well studied milk component is orotic acid. Orotic acid was originally classified as vitamin B<sub>13</sub> and serves as an intermediate in the synthesis of pyrimidine nucleotides, and therefore, DNA and RNA synthesis. Cells of the mammary gland are responsible for production and secretion of orotic acid into milk (Ahmed et al., 1978). Levels of orotic acid in milk are higher in ruminants compared with nonruminant species (Larson and Hegarty, 1979). Considerable variation in levels of bovine milk orotic acid has been reported (Jesse et al., 1980; Buitenhuis et al., 2013). This variation can be partly explained by genetics, breed differences, lactation stage, or parity (Jesse et al., 1980; Sun et al., 2017). A genetic cause for increased levels of orotic acid in milk is the metabolic disorder deficiency of uridine monophosphate synthetase (**dUMPS**; Robinson et al., 1983), which is a mono-

Received November 27, 2018.

Accepted December 3, 2019.

\*Corresponding author: [rooszaalberg1990@gmail.com](mailto:rooszaalberg1990@gmail.com)

genic recessive cattle disorder. Embryos homozygous for dUMPS usually do not live past d 40 of pregnancy (Shanks and Robinson, 1989). Cows heterozygous for dUMPS have reduced conversion of orotic acid into pyrimidine nucleotides, and consequently, increased levels of orotic acid in milk (Robinson et al., 1984).

Relatively little is known about the function of orotic acid in milk, its possible effect on human health, calf health, or manufacturing properties of milk. A suggested function of orotic acid in milk is establishment of the microbiome in the calf's stomach (Motyl et al., 1993; Löffler et al., 2016) by serving as a growth factor for lactobacilli (Okonkwo and Kinsella, 1969; Fernandez-Garcia and McGregor, 1994; Østlie et al., 2003).

Orotic acid in milk can be quantified using nuclear magnetic resonance (NMR) spectroscopy (Buitenhuis et al., 2013), but these analyses are expensive and time consuming and therefore less suited for large-scale routine measurements. Accurate prediction of orotic acid using infrared milk spectra will create opportunities for large-scale analyses of milk orotic acid. This will allow for phenotypic and genetic characterization of orotic acid in milk.

A possible relation between milk orotic acid and infrared milk spectra has been suggested based on genome-wide association studies (GWAS). A GWAS for milk orotic acid showed a significant effect for a genomic region on BTA1 close to *UMPS* (Buitenhuis et al., 2013). The same genomic region was identified in a GWAS for infrared wavenumbers (Wang and Bovenhuis., 2018). These results suggest that the infrared spectrum may contain direct information on milk orotic acid and that it might be possible to predict milk orotic acid based on infrared milk spectra.

The aims of this study were (1) to determine if milk orotic acid can be predicted based on infrared milk spectra, and (2) to identify genetic and systemic environmental factors affecting infrared-predicted milk orotic acid.

## MATERIALS AND METHODS

In this study we distinguish populations that were used for building an infrared prediction model for orotic acid (training and validation populations), and populations that were used for a genetic study on infrared-predicted orotic acid (study populations).

### Building the Infrared Prediction Model for Orotic Acid

The infrared prediction model for orotic acid was built based on a training population and validated us-

ing data from an independent validation population. Different training populations were used, which were either single breed or multi-breed. The multi-breed scenario will be presented and discussed. The single breed scenarios can be found in Appendix Table A1.

**Training Population.** To build the infrared prediction model, a multi-breed training population consisting of a Danish Holstein population (Danish Holstein I) and a Danish Jersey population (Danish Jersey) was used. Danish Holstein I consisted of 292 cows that were between 5 and 365 DIM. Cows came from 3 herds and milk samples were collected from May until June 2013. The Danish Jersey population consisted of 361 cows, which were between 130 and 252 DIM. Cows came from 22 herds and milk samples were collected from February until April 2010. For each cow, one morning milk sample was collected.

**Validation Population.** To validate the infrared prediction model, a second Danish Holstein population (Danish Holstein II) was used. Danish Holstein II consisted of 381 cows, which were between 129 and 228 DIM. Cows came from 20 herds, and milk samples were collected from October until December 2009. One morning milk sample was collected for each cow.

**Infrared Milk Spectra for the Training and Validation Population.** For both the training population and validation population, infrared spectral analysis on milk samples was done with the MilkoScan FT2 (Foss, Hillerød, Denmark). Transmittance values were obtained for 1,060 individual wavenumbers of the infrared spectral region 5,008 through 925  $\text{cm}^{-1}$ . Transmittance values of individual wavenumbers were transformed to absorbance values using

$$\text{absorbance} = \log_{10}(1/\text{transmittance}). \quad [1]$$

Two infrared spectral regions that interact with water molecules were removed: wavenumbers 5,008 through 3,008  $\text{cm}^{-1}$ , and 1,700 through 1,600  $\text{cm}^{-1}$ . The remaining 522 wavenumbers were used to build the prediction model. Absorbance values were smoothed with the Savitzky-Golay method with the `prospectr` package in R (version 3.4.1; <http://www.r-project.org>), using a window size of 9, and a first derivative transformation. For detection of outliers, a Mahalanobis distance of 5 was used.

**NMR Orotic Acid for the Training and Validation Population.** For both the training and validation population, orotic acid in milk was quantified using NMR spectroscopy. The procedure is described in detail by Sundekilde et al. (2013). In brief, milk samples were prepared for analysis by mixing and homogenizing the samples. Then, samples were filtered

to remove residual lipids and proteins. In the next step, an internal chemical shift reference was made by combining filtered sample with heavy water ( $D_2O$ ). Finally, samples were randomized, and proton NMR spectroscopy was done at 298 K on a Bruker Avance III 600 spectrometer, with a proton frequency of 600.13 MHz (Bruker BioSpin GmbH, Rheinstetten, Germany). A single 90-degree pulse experiment was performed to obtain standard 1-dimensional spectra. After proton NMR spectroscopy, proton NMR signals were assigned using literature, spike experiments, 2-dimensional NMR spectroscopy, and the Human Metabolome Database (<http://www.hmdb.ca/>). Orotic acid was expressed in grams per liter of milk, which was calculated from units of intensity of the proton NMR signal.

**Training and Validating the Infrared Prediction Model.** A partial least squares regression (PLSR) method was used for predicting NMR orotic acid with infrared milk spectra. The infrared prediction model for orotic acid was built with the PLS package in R (version 3.4.1; <http://www.r-project.org>). Figure 1 shows how Danish Holstein I and Danish Jersey were used to train the infrared prediction model, how the prediction model was externally validated with Danish Holstein II, and how it was internally validated with Danish Holstein I and Danish Jersey.

For selecting the optimal number of components, the residual mean squared error of the prediction (RM-

SEP) and the accuracy ( $R^2$ ) were quantified. Once both the RMSEP and  $R^2$  improved  $\leq 5\%$ , the optimum number of 18 latent variables was selected (Figure 2).

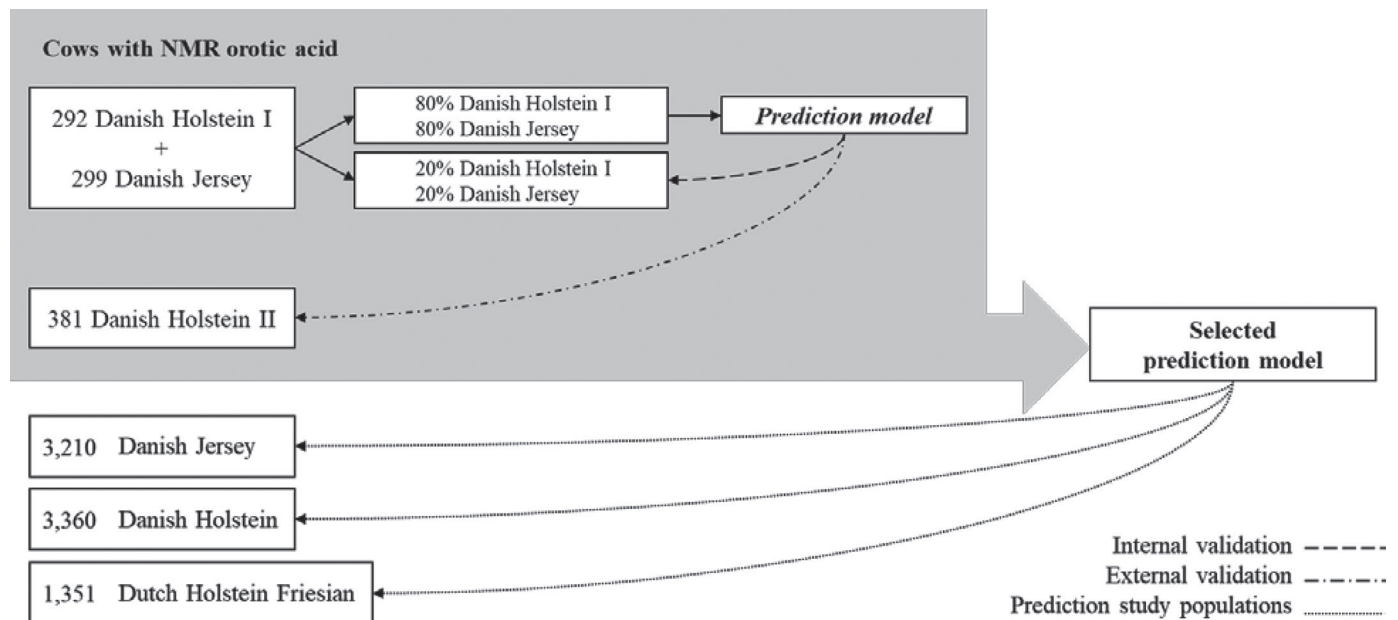
Impact of individual wavenumbers on prediction was determined by inspection of variable importance in projection (VIP; Chong and Jun, 2005) scores. The VIP scores were calculated for each wavenumber individually as

$$VIP_i = \sqrt{\frac{N}{\sum_{m=1}^M \%expl.var_m} \sum_{m=1}^M w_{mi}^2 \%expl.var_m}, \quad [2]$$

where  $N$  is the number of wavenumbers,  $M$  is the number of latent variables,  $\%expl.var_m$  is the proportion of variation for infrared-predicted orotic acid that is explained by latent variable  $m$ , and  $w$  is the weight given to wavenumber  $i$  for latent variable  $m$ .

### Genetic Parameter Estimation for Infrared Predicted Orotic Acid

A genetic analysis was performed on infrared-predicted orotic acid. Genotypes and pedigree information were not available for the training population and validation population. Therefore, 3 additional populations were introduced to this study, which will be referred to as study populations. Two Danish study populations were



**Figure 1.** Process to select the optimal prediction model. The procedure of internal validation (dashed line) was repeated 100 times. After analysis of residual mean squared error (RMSEP) and accuracy ( $R^2$ ) of internal and external validation, one model was selected for infrared prediction of orotic acid for 3 populations of cows without orotic acid values. On these predicted orotic acid values, a statistical analysis was performed. NMR = nuclear magnetic resonance.

used to estimate SNP-based heritability of infrared orotic acid, and for a GWAS in order to identify genes underlying infrared-predicted orotic acid. A Dutch study population was used to estimate pedigree-based heritability and genetic relations between infrared orotic acid and breeding goal traits.

**Danish Study Populations.** Two Danish study populations were used: 3,274 Danish Holstein cows from 308 herds (Danish Holstein study population), and 3,407 Danish Jersey cows from 175 herds (Danish Jersey study population). Morning-milk samples were collected during routine milk recordings in the period of October 2015 through September 2016. For each cow, between 1 and 20 milk samples were available. Average time between milk recordings was 32 days. All milk

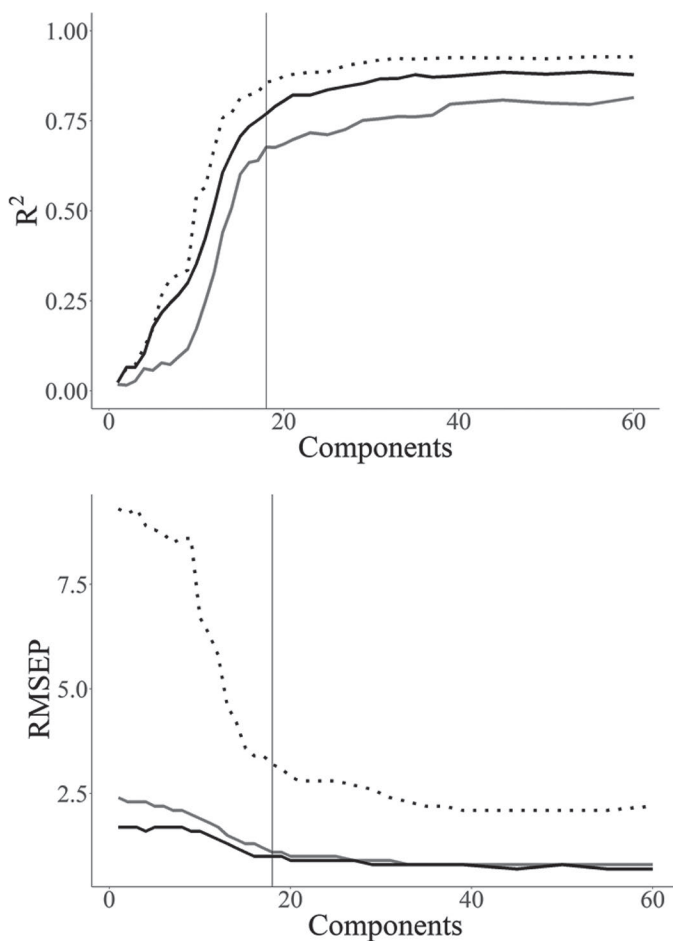
samples were taken between 5 and 365 DIM. If multiple records from the same test day were available for one cow, all records from that particular test day were excluded. Herds with milk records from <5 cows were removed. After data filtering, 3,211 Danish Holstein with 19,144 milk records and 3,360 Danish Jersey with 19,793 milk records were left for analysis.

All cows from the Danish study populations were genotyped with EuroG10K custom SNP chip (10K chip). The 10K chip is a collection of SNP, which were selected based on their association to production traits. The SNP on the 10K chip were selected from whole-genome-sequencing data of bulls from the 1,000-bull-genome project. The 10K genotypes were imputed to the 50K using BEAGLE 4 (Browning and Browning, 2016) and a reference population of animals that were genotyped for Illumina 50K BovineSNP50 v.2 BeadChip (50K chip; Illumina Inc., San Diego, CA). For Danish Holstein, a reference population of 4,000 cows was used, and for Danish Jersey, a reference population of 4,576 cows was used. Reference cows were genotyped for Illumina 50K BovineSNP50 v.2 BeadChip (50K chip, Illumina Inc.). For the genetic analysis, all SNP located on autosomes were selected, SNP that had more than 40% missing genotypes were excluded, and SNP with a MAF <1% were removed. For Danish Holstein 43,807 SNP, and for Danish Jersey 39,235 SNP were left for the genetic analysis.

**Dutch Study Population.** The Dutch study population consisted of 1,748 Dutch Holstein Friesian cows from 371 herds [Dutch (Holstein Friesian) study population]. Milk samples were collected in the period February through March 2005. For each cow, one morning milk sample was available. Milk samples were collected when cows were between 63 and 263 DIM. Herds with milk records from <5 cows were removed. Two milk samples with infrared orotic acid values more than 3 standard deviations from the mean were removed before analysis. After this filtering step, 1,351 Holstein Friesians from 267 herds, with 1,351 milk records were left for the statistical analysis.

A pedigree was provided by CRV (Cooperative Cattle Improvement Organization, Arnhem, the Netherlands), including 26,300 individuals. The EBV for sires of these 1,351 cows were available for 78 traits and were based on publications by CRV BV. A total of 84 bulls with at least 3 daughters were used for this study.

**Infrared Milk Spectra for Danish and Dutch Study Populations.** For the Danish study populations, infrared spectral analysis of all morning-milk samples was performed at the Eurofins-Steins laboratory (Vejen, Denmark) using the MilkoScan FT+ (Foss, Hillerød, Denmark). For the Dutch Holstein Friesian



**Figure 2.** The accuracy ( $R^2$ ) and residual mean squared error (RMSEP) across different numbers of components. Results are shown for internal validation on Danish Holstein I (solid black), internal validation on Danish Jersey (solid gray), and external validation on Danish Holstein II (dotted black). The vertical line indicates the selected number of latent variables (18). Units of RMSEP are milligrams per liter.

study population, infrared spectral analysis of all milk samples was performed in the Milk Control Station (Qlip, Zutphen, the Netherlands) using the MilkoScan FT6000 (Foss). For all study populations, transmittance values were obtained for 1,060 individual wavenumbers of the infrared spectral region 5,008 through 925  $\text{cm}^{-1}$ . These transmittance values of individual wavenumbers were transformed to absorbance values using model 1. Absorbance values were smoothed with the Savitzky-Golay method with the prospectr package in R (version 3.4.1; <http://www.r-project.org>), using a window size of 9 and a first derivative transformation.

**Infrared-Predicted Orotic Acid for Danish and Dutch Study Populations.** The infrared prediction model for orotic acid was used to predict orotic acid for the Danish study populations and the Dutch study population (Figure 1). For the Dutch Holstein Friesian study population, 2 milk records with extreme infrared orotic acid values were removed, and 1,349 cows each with one milk record were left for analysis. For Danish Holstein and Danish Jersey study population, no milk records were removed based on outlying infrared orotic acid records.

**Model Description for Genetic Analysis on Danish study populations.** For the Danish Holstein study population and the Danish Jersey study population, SNP data were available. A genetic analysis of infrared orotic acid for these 2 Danish study populations was done with Bayz software package (version 2.4, <http://www.bayz.biz/>; Krag et al., 2013). Infrared orotic acid was analyzed with the hierarchical model

$$y_{ijk} = \mu + \text{Parity}_i + \beta_1 \text{DIM}_{ijk} + \beta_2 e^{-0.05 \text{DIM}_{ijk}} + \text{HYS}_j + \text{CowA}_k + \text{CowPE}_k + e_{ijk}, \quad [3]$$

where  $y_{ijk}$  is the infrared orotic acid concentration for one milk record;  $\mu$  is the mean;  $\text{Parity}_i$  is a fixed effect ( $i = 1$  or  $2$ );  $\beta_1 \text{DIM}_{ijk}$  and  $\beta_2 e^{-0.05 \text{DIM}_{ijk}}$  adjust for lactation stage (Wilmink function; Wilmink, 1987), where  $\text{DIM}_{ijk}$  is days in milk expressed in years. For  $\text{Parity}$  and  $\text{DIM}$ , a uniform prior distribution was assumed, where values ranged from  $-\infty$  to  $+\infty$ ;  $\text{HYS}_j$  is a random herd-year-season effect (autumn from October 1, 2015, to December 31, 2015; winter from January 1, 2016, to March 31, 2016; spring from April 1, 2016, to June 30, 2016; and summer from July 1, 2016, to September 30, 2016), for which a normal prior distribution was assumed, where  $\text{HYS} \sim N(\mathbf{0}, \sigma_{\text{HYS}}^2)$  and  $\sigma_{\text{HYS}}^2$  is the variance explained by HYS.  $\text{CowPE}_k$  is the permanent environmental effect for cow  $k$ , for which a normal prior distribution was assumed, where  $\text{CowPE} \sim N(\mathbf{0}, \sigma_{\text{PE}}^2)$  and  $\sigma_{\text{PE}}^2$  is the permanent environmental variance, and

$e_{ijk}$  is the residual variance, for which a normal prior distribution was assumed, where  $e \sim N(\mathbf{0}, \sigma_e^2)$  and  $\sigma_e^2$  is the residual variance.  $\text{CowA}_k$  was modelled using SNP data with the hierarchical model

$$\text{CowA}_k = \sum_m a_m g_{km}, \quad [4]$$

where  $\text{CowA}_k$  is the additive genetic value for cow  $k$ , modelled as the sum of additive SNP effects  $a_m$  of SNP  $m$  times allele dosages  $g_{km}$  for SNP  $m$  of cow  $k$ . Allele dosages were centered. A BayesA model was used for the additive SNP effects with  $a_m \sim N(\mathbf{0}, \sigma_{am}^2)$ , where  $\sigma_{am}^2$  is the additive genetic variation explained by SNP  $m$ . All  $\sigma_{am}^2$  have a common scaled inverse chi-square distribution with 5 degrees of freedom, and a scale parameter estimated from the data with a uniform prior distribution. The total additive genetic variance modeled by all SNP was  $\sigma_A^2 = \text{var}(\text{CowA})$ , where  $\sigma_A^2$  is the additive genetic variance. This parameter was computed by evaluating  $\text{var}(\text{CowA})$  over the samples of the Markov chain Monte Carlo. All SNP variances were estimated simultaneously.

A Metropolis-Hastings sampler was used, with 100,000 cycles, including 30,000 burn-in cycles. Convergence of the model was confirmed by analysis of traceplots, and Markov chain Monte Carlo coefficient of variation (MCMC-CV), which represents the standard error of the mean of the chain. The MCMC-CV values  $< 1\%$  suggest that convergence has been reached.

**Significance Testing.** For the Danish study populations, significance of lactation stage ( $\beta_1 \text{DIM}$  and  $\beta_2 e^{-0.05 \text{DIM}}$ ), herd-year-season, and parity on infrared orotic acid were analyzed by comparing deviance information criterion (DIC) scores between a model with all effects, and a model excluding one effect at the time. The model with the lowest DIC score was considered the best fitting model.

**SNP-Based Heritability.** For the Danish study populations, heritability was estimated as the SNP-based heritability of infrared orotic acid, which was calculated as

$$h_{\text{SNP}}^2 = \frac{\sigma_A^2}{\sigma_{\text{HYS}}^2 + \sigma_A^2 + \sigma_{\text{PE}}^2 + \sigma_e^2}, \quad [5]$$

The total variation explained by herd-year-season was calculated as

$$\text{HYS} = \frac{\sigma_{\text{HYS}}^2}{\sigma_{\text{HYS}}^2 + \sigma_A^2 + \sigma_{\text{PE}}^2 + \sigma_e^2}. \quad [6]$$

The repeatability was calculated as

$$\rho = \frac{\sigma_A^2 + \sigma_{PE}^2}{\sigma_{HYS}^2 + \sigma_A^2 + \sigma_{PE}^2 + \sigma_e^2}, \quad [7]$$

where  $\sigma_A^2$  is the additive genetic variance,  $\sigma_{HYS}^2$  is the herd-year-season variance,  $\sigma_{PE}^2$  is the permanent environmental variance, and  $\sigma_e^2$  is the residual variance.

**GWAS.** For the Danish study populations, associations between genomic regions and infrared orotic acid were analyzed. After SNP variances were estimated simultaneously, within each chromosome SNP were divided into groups of 100 consecutive SNP (Gebreyesus et al., 2017). The grouping procedure was repeated 5 times for each chromosome, starting with counting at SNP 1, 21, 41, 61, or 81 on the chromosome. Between the 5 repeated procedures, SNP groups overlapped, yet SNP groups were never identical. Groups with <80 SNP were excluded from analysis. For each group of 100 SNP, variance of genomic estimated breeding value was calculated with the postanalytical gbayz function of Bayz software (<http://www.bayz.biz/>; Krag et al., 2013). Proportion of total additive genetic variance explained per SNP group was calculated as

$$\% \sigma_{A,i}^2 = \frac{\sigma_{gEBV,i}^2}{\sigma_A^2} \times 100\%, \quad [8]$$

where  $\% \sigma_{A,i}^2$  is the percentage of total additive genetic variance of infrared orotic acid explained by SNP group  $i$ ;  $\sigma_{gEBV,i}^2$  is the variance of genomic EBV (gEBV) for infrared orotic acid of SNP group  $i$ ; and  $\sigma_A^2$  is the total additive genetic variance of infrared orotic acid.

Gene annotation was performed for all SNP in Ensemble (92), using the UMD3.1 assembly, and the variant effect predictor function (McLaren et al., 2016).

**Model Description for Genetic Analysis on Dutch Study Population.** For the Dutch study population, infrared-predicted orotic acid was analyzed using the following model:

$$y_{ijkl} = \mu + Date_i + Sirecode_j + \beta_1 DIM_{ijkl} + \beta_2 e^{-0.05 DIM_{ijkl}} + HYS_k + A_l + e_{ijkl}, \quad [9]$$

where  $y_{ijkl}$  is the infrared orotic acid concentration for one milk record;  $\mu$  is the mean;  $Date_i$  is a fixed effect of date of infrared analysis of the milk sample ( $i = 1$  through 17);  $Sirecode_j$  is a fixed effect that corrects for

possible genetic differences between groups of proven bull daughters and young bull daughters ( $j = 1$  through 3);  $\beta_1 DIM_{ijkl}$  and  $\beta_2 e^{-0.05 DIM_{ijkl}}$  adjust for lactation stage (Wilmink function; Wilmink, 1987), where  $DIM_{ijkl}$  is days in milk expressed in years. For  $Date$ ,  $Sirecode$ , and  $DIM$ , a uniform prior distribution was assumed, where values ranged from  $-\infty$  to  $+\infty$ ;  $HYS_k$  is a random herd-year-season effect (summer from June 1, 2004, to August 31, 2004; autumn from September 1, 2004, to November 30, 2004; and winter from December 1, 2004, to January 31, 2005), for which a normal prior distribution was assumed, where  $HYS \sim N(\mathbf{0}, \sigma_{HYS}^2)$ ;  $A_l$  is the additive genetic effect for animal  $l$ , where  $A \sim N(\mathbf{0}, \sigma_A^2)$ ; and  $e_{ijkl}$  is the residual variance, for which a normal prior distribution was assumed, where  $e \sim N(\mathbf{0}, \sigma_e^2)$ . Parameters were estimated using the Bayz software package (<http://www.bayz.biz/>; Krag et al., 2013).

A Metropolis-Hastings sampler was used, with 1,100,000 cycles, including 100,000 burn-in cycles. Convergence of the model was assumed with the MCMC-CV, which represents the standard error of the mean of the chain. The MCMC-CV values <1% suggest that convergence has been reached.

**Pedigree-Based Heritability.** For the Dutch study population, heritability was estimated as the pedigree-based heritability of infrared orotic acid, which was calculated as

$$h_{ped}^2 = \frac{\sigma_A^2}{\sigma_{HYS}^2 + \sigma_A^2 + \sigma_e^2}. \quad [10]$$

The total variation explained by herd was calculated as

$$HYS = \frac{\sigma_{HYS}^2}{\sigma_{HYS}^2 + \sigma_A^2 + \sigma_e^2}, \quad [11]$$

where  $\sigma_A^2$  is the additive genetic variance,  $\sigma_{HYS}^2$  is the herd-year-season variance, and  $\sigma_e^2$  is the residual variance.

**Calo's Genetic Correlations.** For the Dutch Holstein Friesian study population, genetic correlations between EBV of infrared-predicted orotic acid and EBV of in total 78 other traits were estimated using Calo's method (Calo et al., 1973). The Calo's correlation method gives an estimate of the genetic correlation. Standard errors of the genetic correlations are not provided (Mitchell et al., 2005). The Dutch study population consisted of 5 large paternal half-sib families from proven sires (98 through 196 daughters per sire), 50 small paternal half-sib families from test

sires (8 through 23 daughters per sire), and 168 cows descending from 44 other proven sires. Calo's genetic correlations were calculated as

$$r_g = \frac{\sqrt{\sum_{i=1}^n \rho_{i,EBV} \times \sum_{i=1}^n \rho_{i,EBV'}}}{\sum_{i=1}^n \rho_{i,EBV} \times \rho_{i,EBV'}} \times r_{EBV, EBV'}, \quad [12]$$

where  $\rho_{i,EBV}$  is the reliability of the EBV for infrared orotic acid for the bull  $i$ ,  $\rho_{i,EBV'}$  is the reliability of the EBV for 1 of the 78 production traits, and  $r_{EBV,EBV'}$  is the Pearson correlation between the 2 EBV.

#### Standardization of Infrared Milk Spectra.

Three different types of spectrometers were used for the infrared spectral analysis of the milk samples from the 3 different populations. For the training and validation population the Milkoscan FT2 was used, for the Danish study populations the Milkoscan FT+ was used, and for the Dutch Holstein Friesian study population the Milkoscan FT6000 was used.

The type of spectrometer influences the shape of the milk spectra (Appendix Figure A1). For directly comparing infrared-predicted orotic acid, standardization of milk spectra is necessary (Grelet et al., 2015). The spectrometer on which most milk samples were analyzed was the Milkoscan FT+. Milk samples were, therefore, standardized towards this spectrometer. The following equation was used for standardizing absorbance values for one individual wavenumber:

$$wvn_{ij}^* = \left( \frac{wvn_{ij} - \mu_{1,i}}{\sigma_{1,i}} \right) \times \sigma_{2,i} + \mu_{2,i}, \quad [13]$$

where  $wvn_{ij}^*$  is the standardized absorbance value for wavenumber  $i$  of individual  $j$ ,  $wvn_{ij}$  is the unstandardized absorbance value for wavenumber  $i$  of individual  $j$ ,  $\mu_{1,i}$  and  $\sigma_{1,i}$  are the unstandardized mean and standard deviation of the absorbance values of wavenumber  $i$  for the population to be standardized,  $\mu_{2,i}$  and  $\sigma_{2,i}$  are the mean and standard deviation of the absorbance values of wavenumber  $i$  for the reference population (milk samples measured on Milkoscan FT+). Mean absorbance values before and after standardization are shown in Appendix Figure A1.

For the Danish Holstein I, Danish Holstein II, and Dutch Holstein Friesian study population, absorbance values of individual wavenumbers were standardized according to means and standard deviations of absorbance values for cows from the Danish Holstein study population. For the Danish Jersey population (training), absorbance values of individual wavenumbers were standardized according to means and standard

deviations of absorbance values for cows from the Danish Jersey study population. For calculating  $\mu_{2,i}$  and  $\sigma_{2,i}$  only milk samples from the same lactation stage were selected.

The infrared prediction model for orotic acid that was built with standardized milk spectra will be referred to as the standardized prediction model. Standardization always took place before the prediction model was built and before the analyses were performed. The standardized prediction model was only used to compare infrared orotic acid for study populations across lactation. For the all genetic analyses, the unstandardized prediction model was used.

## RESULTS

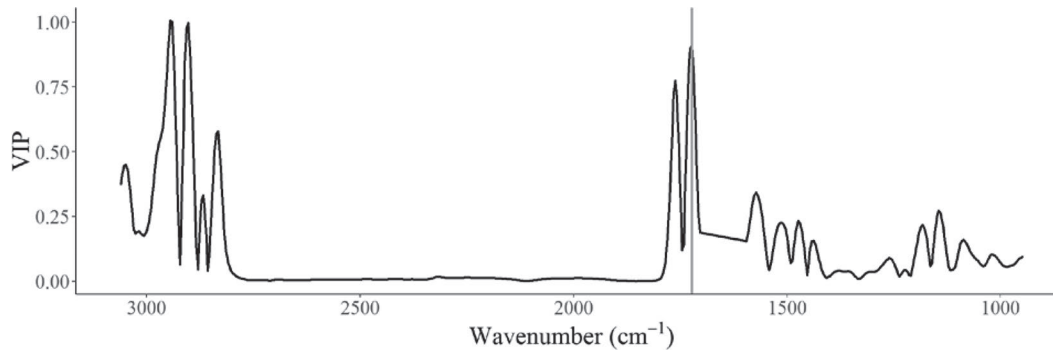
### Infrared Prediction Model for Orotic Acid

**NMR Orotic Acid.** Means and standard deviations of the NMR orotic acid for the Danish Holstein I (training), Danish Holstein II (validation), and Danish Jersey (training) population are presented in Table 1. Mean NMR orotic acid concentration was 14.5 mg/L in the Danish Holstein I population, and 19.9 mg/L in the Danish Holstein II population, and 14.5 mg/L in the Danish Jersey population.

**Infrared Prediction.** The performance of the infrared prediction models for orotic acid built with unstandardized and standardized milk spectra are presented in Table 2. Figure 2 shows the  $R^2$  and RMSEP for the prediction models with 1 through 70 latent variables. A model with 18 latent variables was selected (vertical line in Figure 2). Accuracy varied from 0.68 to 0.86 for the unstandardized prediction model, and from 0.60 to 0.79 for the standardized prediction model. The RMSEP varied from 1.0 to 3.7 for the unstandardized prediction model, and from 3.2 to 5.5 for the standardized prediction model. Performance for prediction models based on a single breed training population is shown in Appendix Table A1.

The VIP scores for individual wavenumbers are plotted in Figure 3. Results were similar for the scenario that used unstandardized milk spectra and the scenario that used standardized milk spectra. Wavenumbers with a high VIP score have a greater effect on the prediction model than those wavenumbers with a low VIP score. Wavenumbers with a VIP score  $>1$  were VIP scores for wavenumbers 3,061 and 3,057  $\text{cm}^{-1}$ .

**Infrared Orotic Acid.** When using the unstandardized prediction model on the study populations, infrared orotic acid of the study populations was lower (8.9 to 14.8 mg/L) compared with the NMR orotic acid of the training and validation population (14.4 to 19.8 mg/L). When using the standardized prediction model,



**Figure 3.** Plot with variable importance in projection (VIP) scores for individual wavenumbers of a prediction model with 18 latent variables.

however, all 3 study populations had infrared orotic acid values that were similar to the training and validation populations (12.5 to 14.5 mg/L).

Figure 4 shows infrared orotic acid across lactation for the Danish study populations and the Dutch study population. Means were calculated for intervals of 10 d, with on average 82 records per interval for Dutch Holstein Friesian, 522 records for Danish Holstein, and 539 records for Danish Jersey. Results are presented for both the unstandardized prediction model and the standardized prediction model.

### Genetic Parameter Estimation for Infrared Predicted Orotic Acid

**Significance Testing.** The DIC scores of the model containing all variables, and models excluding one variable at the time are shown in Appendix Table A2. The DIC scores suggest that for Danish Holstein all variables have an effect on orotic acid. For Danish Jersey, on the other hand, parity has no effect on infrared orotic acid.

### Heritability, Repeatability, and Herd Effect.

Table 1 shows an overview of heritability for infrared orotic acid. The SNP-based heritability of infrared-predicted orotic acid was 0.18 in the Danish Holstein study population, and 0.09 in the Danish Jersey study population. Pedigree-based heritability was 0.37 in the Dutch Holstein Friesian study population.

**GWAS.** Percentages of additive genetic variance of infrared orotic acid explained by genomic regions are presented in Figure 5 in a Manhattan format. For both Danish Holstein and Danish Jersey, most genetic variation was explained by a genomic region on BTA1 harboring *UMPS*. For Danish Holstein, a region on BTA14 that harbors *DGAT1* explained >1% of the additive genetic variation.

**Calo's Genetic Correlations to (Non-)Production Traits.** Estimated Calo's genetic correlations between EBV of infrared orotic acid and a subset of the 78 traits are presented in Table 3. Traits in Table 3 had a Calo's correlation with infrared orotic acid >0.15 or <-0.15. Calo's correlation between EBV of

**Table 1.** Means and SD for nuclear magnetic resonance (NMR) orotic acid for the training and validation populations, and for infrared orotic acid for all the populations used in the study<sup>1</sup>

Breed	n <sup>2</sup>	Records	DIM	NMR	Mean orotic acid (SD) in mg/L				
					Infrared		h <sup>2</sup>	ρ <sup>3</sup>	HYS <sup>4</sup>
					UnSt	St			
Danish Holstein I	292		9–364	14.4 (5.4)	14.4 (4.5)	14.5 (4.7)	—	—	—
Danish Holstein II	381		129–228	19.8 (7.7)	18.9 (6.1)	16.4 (5.5)	—	—	—
Danish Jersey	299		130–252	14.5 (3.4)	14.7 (3.4)	14.6 (3.2)	—	—	—
Danish Holstein	3,210	19,144	5–365	—	10.3 (4.4)	14.5 (6.3)	0.18 <sup>5</sup>	0.23	0.15
Danish Jersey	3,360	19,793	5–365	—	8.9 (4.7)	12.5 (6.8)	0.09 <sup>5</sup>	0.11	0.22
Dutch Holstein Friesian	1,349	1,349	63–263	—	14.8 (3.5)	14.5 (6.4)	0.37 <sup>6</sup>	—	0.07

<sup>1</sup>Infrared-predicted orotic acid is presented for the unstandardized prediction model (UnSt) and standardized prediction model (St).

<sup>2</sup>Numbers after removal of outliers.

<sup>3</sup>Repeatability.

<sup>4</sup>Herd-year-season effect, SE = 0.01.

<sup>5</sup>SNP-based heritability, SE = 0.01.

<sup>6</sup>Pedigree-based heritability, SE = 0.12.



**Table 2.** Prediction accuracy ( $R^2$ ) and residual mean squared error (RMSEP) for the scenario with unstandardized milk spectra (UnSt) and standardized milk spectra (St)<sup>1</sup>

Validation population	Type of validation	$R^2$		RMSEP	
		UnSt	St	UnSt	St
Danish Holstein I	Internal	0.77	0.76	1.1	4.7
Danish Holstein II	External	0.86	0.79	3.2	5.5
Danish Jersey	External	0.68	0.60	1.0	3.2

<sup>1</sup>Both prediction models used 18 components.

selected traits and fat, protein, or lactose % was  $<0.2$  and  $>-0.2$ , and were always smaller than the genetic correlation to infrared orotic acid.

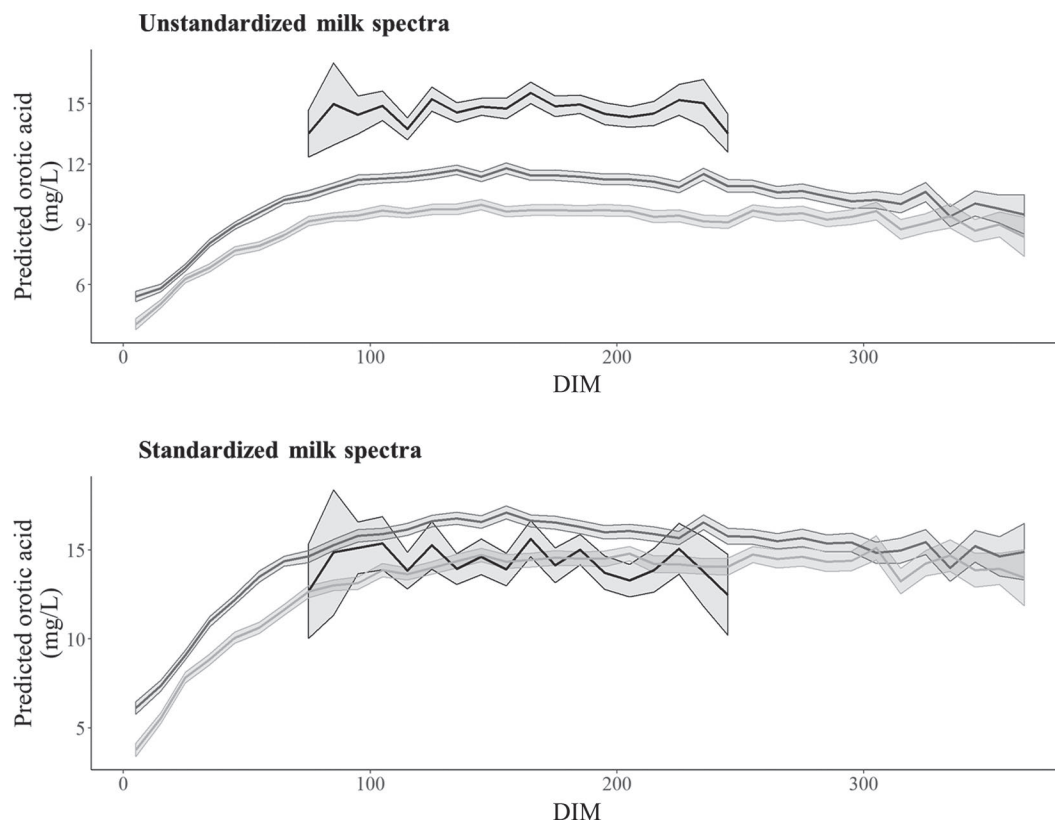
## DISCUSSION

The aims of our study were to determine if milk orotic acid can be predicted from Fourier transform infrared milk spectra, and to perform a genetic analysis based on infrared orotic acid. Various methods have been suggested for determining orotic acid in milk, such as HPLC (Counotte, 1983; Saidi and Warthesen, 1989) and NMR (Wevers et al., 1999). This is, to the best of our knowledge, the first time that milk orotic acid was

predicted from Fourier transform infrared milk spectra. Other novelties of our study were the comprehensive analysis on systemic environmental factors affecting milk orotic acid and estimation of Calo's correlations between EBV of infrared orotic acid and (non-)production traits.

### Infrared Prediction Model for Orotic Acid

**Variation in NMR Orotic Acid.** To build the prediction model, the “gold standard” phenotype NMR orotic acid was used. Table 1 reveals that mean NMR orotic acid varied between the Danish Holstein I (14.4 mg/L), Danish Holstein II (19.8 mg/L), and Danish



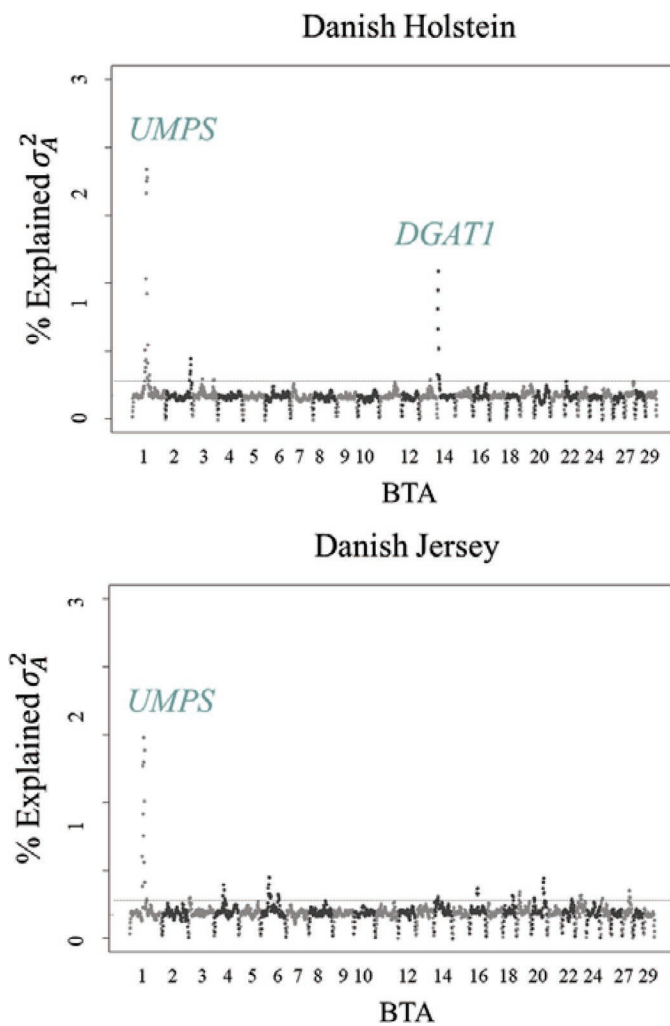
**Figure 4.** Mean predicted orotic acid with confidence interval ( $P = 0.05$ ) across lactation for 3 study populations: Danish Holstein (dark gray), Danish Jersey (light gray), and Dutch Holstein Friesian (black).

Jersey (14.5 mg/L). A cause for these differences could be lactation stage during which milk samples were collected. Milk samples of Danish Holstein II had the highest NMR orotic acid and were collected from mid-lactation cows (129 to 228 DIM). Danish Holstein I cows, on the other hand, were from across the entire lactation (5 to 365 DIM) and had lower NMR orotic acid. This is in line with Figure 4, which shows that infrared orotic acid was highest in mid lactation. However, the subset of Danish Holstein I cows that were in the same lactation stage as Danish Holstein II cows (129 to 228 DIM) still had a lower mean for NMR orotic acid (15.7 mg/L). A more likely cause would be batch effects on milk NMR analysis, since milk samples were not analyzed on the same day. This day-of-analysis sen-

sitivity should be studied more thoroughly, if one would like to use infrared prediction for not only accurate but also precise prediction of orotic acid.

**Spectrometer Bias and Standardization of Milk Spectra.** In the current study, 3 different spectrometers were used to analyze milk samples. Using a prediction model for orotic acid that was trained with milk spectra from one type of spectrometer will introduce uncertainty to the prediction when applying it on milk spectra from a different type of spectrometer (Grelet et al., 2015). This problem can be solved by standardization of milk spectra. Several standardization methods have been proposed, yet many of these methods demand that at least some milk samples are analyzed on multiple spectrometers (Wang et al., 1991; Grelet et al., 2015). An alternative standardization would be the method presented in Bonfatti et al. (2017), who simulated standardized milk spectra for 2 spectrometers by using the percentiles of the distribution of the absorbance at individual wavenumbers. For future research, it would be interesting to use this method and compare it with the current method.

**Limitations of Standardization Method.** Prediction performance was lower when standardized milk spectra were used to train the prediction model than when unstandardized milk spectra were used. Prediction accuracy dropped by 0.01 to 0.08, and the prediction error for external validation increased by 2.2 to 3.6 mg/L (Table 2). This reduction in prediction performance could be linked to the large number of latent variables used by the prediction model. When many latent variables are used for prediction, it is likely that indirect information from milk spectra is used for prediction. This indirect information could be linked to major milk components. Detailed milk composition, for example, depended on indirect information hidden in covariance structures related to major milk components (Eskildsen et al., 2014, 2016). With the standardization method used in the current study, the standard deviations of individual wavenumbers were altered, which could have disrupted covariance structures that were initially used as an information source by the unstandardized prediction model. One could argue that the standardization method used in the current study may have corrected for more than just spectrometer, but also for factors that influence milk composition, such as breed (Maurice-Van Eijndhoven et al., 2011; Gustavsson et al., 2014), diet (Palmquist and Beaulieu, 1993), or lactation stage (Wilmink, 1987). We expect that by including information from multiple spectrometers for standardizing milk spectra, as has been done by Grelet et al. (2015), covariance structures could remain, and prediction of orotic acid will be improved. To prove this point, it is necessary to compare such multi-spectrom-



**Figure 5.** Percentage of explained additive genetic variation ( $\sigma_A^2$ ) for infrared orotic acid for groups of 100 consecutive SNP. Name labels indicate candidate genes, which are located within the genomic region of the peak SNP group.

eter methods with the method that was used in the current study.

### Genetic Parameter Estimation for Infrared-Predicted Orotic Acid

Heritability of orotic acid has only been studied by Buitenhuis et al. (2013), estimating SNP-based heritability for NMR orotic acid at 0.83. This is much higher than the SNP heritability of our infrared-predicted orotic acid (0.09–0.18). An explanation for this difference could be that Buitenhuis et al. (2013) used a relatively small number of samples ( $n = 371$ ), and the standard error of this estimate was high (0.21).

**BTA1 and UMPS.** The genomic region on BTA1 contains the gene *UMPS*, which has been strongly associated with NMR milk orotic acid (Buitenhuis et al., 2013). A signal from this same genomic region on BTA1 is visible for both Danish Holstein and Danish Jersey (Figure 5). The protein UMPS converts orotic acid during pyrimidine biosynthesis (Löffler et al., 2015). A known mutation in *UMPS* is responsible for the genetic defect dUMPS. This mutation results in transcription of nonfunctional UMPS protein, and consequently, the

conversion of orotic acid is blocked, which results in increased orotic acid excretion in milk (Robinson et al., 1983).

**Other Genomic Regions Associated with Orotic Acid.** Two other QTL for NMR milk orotic acid have been identified by Buitenhuis et al. (2013), in the genomic region of 89.9 Mbp on BTA5, and 1.4 Mbp on BTA19. These 2 regions were not observed in the current study. A reason for the absence of a signal could be that genetic variation within these genomic regions was smaller in the Danish study populations compared with the population used by Buitenhuis et al. (2013). Additionally, the most significant SNP from the genomic regions on BTA5 and BTA11 mentioned by Buitenhuis et al. (2013) were not present on the SNP chip used in the current study. Another reason could be that the infrared prediction model for orotic acid did not capture the variation from the milk spectra, which could ultimately be linked to the genomic regions on BTA5 and BTA19.

**Calo's Genetic Correlations to Production Traits.** Not much is known about the role of orotic acid in milk. The approximate genetic correlations calculated in the current study might provide an introduc-

**Table 3.** Calo's genetic correlations between EBV of a selection of production traits and EBV for fat, protein, or lactose %, and orotic acid<sup>1</sup>

Trait <sup>2,3</sup>	$r_g$ to EBV of trait			
	Fat %	Protein %	Lactose %	Infrared orotic acid
Milk composition				
Fat %	1.00	0.68	0.24	-0.23
Protein %	0.68	1.00	0.12	-0.11
Lactose %	0.24	0.12	1.00	0.11
Milk yield				
Milk yield	-0.71	-0.67	-0.20	0.28
Fat yield	0.33	-0.04	0.02	0.06
Protein yield	-0.49	-0.21	-0.20	0.29
Lactose yield	-0.67	-0.65	0.05	0.31
Feet and legs				
Index (feet and legs)	-0.04	0.00	-0.01	0.23
Rear leg set side view	0.12	-0.11	-0.13	-0.28
Rear legs rear view	-0.03	0.02	0.00	0.15
Foot angle	-0.10	0.06	0.04	0.28
Digital dermatitis	0.07	0.16	-0.07	0.31
Locomotion	-0.05	0.00	-0.03	0.26
Udder				
Index (udder)	0.08	0.10	-0.02	-0.20
Udder depth	0.14	0.10	0.05	-0.20
Fore udder attachment	0.07	0.05	-0.02	-0.21
Other				
Growth	-0.11	-0.03	-0.04	-0.20

<sup>1</sup>Selected traits had a genetic correlation ( $r_g$ ) with infrared orotic acid  $>0.15$  or  $<-0.15$ . With the exception of milk yield traits, the genetic correlation between selected traits and fat, protein, or lactose % was  $<0.2$  and  $>-0.2$ , and always  $<r_g$  for infrared orotic acid. Results are presented for scenario that used unstandardized milk spectra to train the prediction model.

<sup>2</sup>For a detailed explanation of traits, see <https://www.cooperatie-crv.nl/downloads/english/e-chapters/>.

<sup>3</sup> $n = 84$  bulls.

tory insight in the function of the high levels of orotic acid in ruminant milk and may also sketch its potential as an indicator trait. Many of the observed correlations were weak or were caused by strong correlation with milk composition, such as fat, protein, or lactose percentage. There seems to be a link between infrared orotic acid and foot and leg traits. However, the biological link between infrared orotic acid and these traits is not clear.

### Future Prospects

Our study shows that milk orotic acid can be predicted based on Fourier transform infrared milk spectra. Standardization of milk spectra with a robust method, such as piecewise direct standardization (Grelet et al., 2015), would be essential for precise and accurate application of infrared prediction models. Furthermore, to confirm the interesting genetic relations to production traits, it would be helpful to estimate genetic correlations based on a larger dataset.

A possible application of infrared-predicted orotic acid could be to study fermented dairy products, such as yogurt. During the production of yogurt, milk orotic acid is fermented by lactobacilli (Okonkwo and Kinsella, 1969; Østlie et al., 2003; Álvarez-Martín et al., 2008). The extent of fermentation of milk orotic acid depends on the starter culture used (Haggerty et al., 1984; Venica et al., 2014). Organic acids in milk, such as orotic acid, are known to be important for the flavor of fermented dairy products (Guzel-Seydim et al., 2000). Predicting milk orotic acid before fermentation could be used to select milk with desired orotic acid concentration or to select a suitable starter culture.

### CONCLUSIONS

Milk orotic acid in cows can be predicted from infrared milk spectra. Genetics underlying infrared orotic acid revealed weak to moderate heritability, and confirmed association to a genomic region that was previously associated with orotic acid in milk. The results of this study suggest that standardization of milk spectra is essential for precise prediction of orotic acid, but also that is necessary to use a robust standardization method that maintains the information from the original milk spectra. Overall, the results of the current study provide a deeper insight in the possible function and characteristics of orotic acid in ruminant milk.

### ACKNOWLEDGMENTS

We acknowledge our funding projects, “FT-IR spektre i mælk: Genetisk variation og effekt på sundhed,

frugtbarhed og energibalance” of the Milk Levy Fund, Denmark (2015–2017), and Robust and Efficient Dairy Cows-REFFICO, project number: 34009-14-0848 of Grønt Udviklings-og Demonstrations Program (GUDP, Copenhagen, Denmark; 2015–2018), and the financial support from The Graduate School of Science and Technology (GSST) of Aarhus University (Aarhus, Denmark) for Roos M. Zaalberg. We also acknowledge the Dutch Milk Genomics Initiative and the project “Melk op Maat,” funded by Wageningen University (Wageningen, the Netherlands) for providing the data of the Dutch Holstein Friesian study population. The authors have not stated any conflicts of interest.

### REFERENCES

- Ahmed, A. A., G. A. Porter, and R. D. McCarthy. 1978. Isolation, quantification, and biosynthetic origin of orotic acid in milk. *J. Dairy Sci.* 61:39–43. [https://doi.org/10.3168/jds.S0022-0302\(78\)83548-6](https://doi.org/10.3168/jds.S0022-0302(78)83548-6).
- Álvarez-Martín, P., A. Belén Flórez, A. Hernández-Barranco, and B. Mayo. 2008. Interaction between dairy yeasts and lactic acid bacteria strains during milk fermentation. *Food Control* 19:62–70.
- Bonfatti, V., A. Fleming, A. Koeck, and F. Miglior. 2017. Standardization of milk infrared spectra for the retroactive application of calibration models. *J. Dairy Sci.* 100:2032–2041. <https://doi.org/10.3168/jds.2016-11837>.
- Browning, B. L., and S. R. Browning. 2016. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98:116–126. <https://doi.org/10.1016/j.ajhg.2015.11.020>.
- Buitenhuys, A. J., U. K. Sundekilde, N. A. Poulsen, H. C. Bertram, L. B. Larsen, and P. Sørensen. 2013. Estimation of genetic parameters and detection of quantitative trait loci for metabolites in Danish Holstein milk. *J. Dairy Sci.* 96:3285–3295. <https://doi.org/10.3168/jds.2012-5914>.
- Calo, L. L., R. E. McDowell, L. D. VanVleck, and P. D. Miller. 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. *J. Anim. Sci.* 37:676–682. <https://doi.org/10.2527/jas1973.373676x>.
- Chong, I.-G., and C.-H. Jun. 2005. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* 78:103–112. <https://doi.org/10.1016/j.chemolab.2004.12.011>.
- Counotte, G. H. M. 1983. Determination of orotate in ruminant milk by high-performance liquid chromatography. *J. Chromatogr.* 276:423–426. [https://doi.org/10.1016/S0378-4347\(00\)85111-7](https://doi.org/10.1016/S0378-4347(00)85111-7).
- de Roos, A. P., H. J. Van Den Bijgaart, J. Hørlyk, and G. de Jong. 2007. Screening for subclinical ketosis in dairy cattle by Fourier transform infrared spectrometry. *J. Dairy Sci.* 90:1761–1766. <https://doi.org/10.3168/jds.2006-203>.
- Eskildsen, C. E., M. A. Rasmussen, S. B. Engelsen, L. B. Larsen, N. A. Poulsen, and T. Skov. 2014. Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables. *J. Dairy Sci.* 97:7940–7951. <https://doi.org/10.3168/jds.2014-8337>.
- Eskildsen, C. E., T. Skov, M. S. Hansen, L. B. Larsen, and N. A. Poulsen. 2016. Quantification of bovine milk protein composition and coagulation properties using infrared spectroscopy and chemometrics: A result of collinearity among reference variables. *J. Dairy Sci.* 99:8178–8186. <https://doi.org/10.3168/jds.2015-10840>.
- Fernandez-Garcia, E., and J. U. McGregor. 1994. Determination of organic acids during the fermentation and cold storage of yogurt. *J. Dairy Sci.* 77:2934–2939. [https://doi.org/10.3168/jds.S0022-0302\(94\)77234-9](https://doi.org/10.3168/jds.S0022-0302(94)77234-9).
- Gebreyesus, G., M. S. Lund, B. Buitenhuys, H. Bovenhuis, N. A. Poulsen, and L. G. Janss. 2017. Modeling heterogeneous (co)variances from adjacent SNP groups improves genomic prediction for milk protein composition traits. *Genet. Sel. Evol.* 49:89.

- Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. *J. Dairy Sci.* 98:2150–2160. <https://doi.org/10.3168/jds.2014-8764>.
- Gustavsson, F., A. J. Buitenhuis, M. Johansson, H. P. Bertelsen, M. Glantz, N. A. Poulsen, H. Lindmark Månsson, H. Stålhmar, L. B. Larsen, C. Bendixen, M. Paulsson, and A. André. 2014. Effects of breed and casein genetic variants on protein profile in milk from Swedish red, Danish Holstein, and Danish Jersey cows. *J. Dairy Sci.* 97:3866–3877.
- Guzel-Seydim, Z., A. C. Seydim, and A. K. Greene. 2000. Organic acids and volatile flavor components evolved during refrigerated storage of kefir. *J. Dairy Sci.* 83:275–277. [https://doi.org/10.3168/jds.S0022-0302\(00\)74874-0](https://doi.org/10.3168/jds.S0022-0302(00)74874-0).
- Haggerty, R. J., L. O. Lueddecke, C. W. Nagel, and L. K. Massey. 1984. Effect of selected yogurt cultures on the concentration of orotic acid, uric acid and a hydroxymethylglutaric-like compound in milk after fermentation. *J. Food Sci.* 49:1194–1195. <https://doi.org/10.1111/j.1365-2621.1984.tb10426.x>.
- Jesse, B. W., C. R. Anderson, and J. L. Robinson. 1980. Bovine milk orotate: Differences between cows and changes during lactation. *J. Dairy Sci.* 63:235–242. [https://doi.org/10.3168/jds.S0022-0302\(80\)82919-5](https://doi.org/10.3168/jds.S0022-0302(80)82919-5).
- Krag, K., N. A. Poulsen, M. K. Larsen, L. B. Larsen, L. L. Janss, and B. Buitenhuis. 2013. Genetic parameters for milk fatty acids in Danish Holstein cattle based on SNP markers using a Bayesian approach. *BMC Genet.* 14:79. <https://doi.org/10.1186/1471-2156-14-79>.
- Larson, B. L., and H. M. Hegarty. 1979. Orotic acid in milks of various species and commercial dairy products. *J. Dairy Sci.* 62:1641–1644. [https://doi.org/10.3168/jds.S0022-0302\(79\)83474-8](https://doi.org/10.3168/jds.S0022-0302(79)83474-8).
- Löffler, M., E. A. Carrey, and E. Zameitat. 2015. Orotic acid, more than just an intermediate of pyrimidine de novo synthesis. *J. Genet. Genomics* 42:207–219. <https://doi.org/10.1016/j.jgg.2015.04.001>.
- Löffler, M., E. A. Carrey, and E. Zameitat. 2016. Orotate (orotic acid): An essential and versatile molecule. *Nucleosides Nucleotides Nucleic Acids* 35:566–577. <https://doi.org/10.1080/15257770.2016.1147580>.
- Luinge, H. J., E. Hop, E. T. G. Lutz, J. A. van Hemert, and E. A. M. de Jong. 1993. Determination of the fat, protein and lactose content of milk using Fourier transform infrared spectrometry. *Anal. Chim. Acta* 284:419–433. [https://doi.org/10.1016/0003-2670\(93\)85328-H](https://doi.org/10.1016/0003-2670(93)85328-H).
- Maurice-Van Eijndhoven, M. H. T., S. J. Hiemstra, and M. P. L. Calus. 2011. Short communication: Milk fat composition of 4 cattle breeds in the Netherlands. *J. Dairy Sci.* 94:1021–1025.
- McLaren, W., L. Gil, S. E. Hunt, H. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The Ensembl variant effect predictor. *Genome Biol.* 17:122. <https://doi.org/10.1186/s13059-016-0974-4>.
- Mitchell, R. G., G. W. Rogers, S. D. Dechow, J. E. Vallimont, J. B. Cooper, U. Sander-Nielsen, and J. S. Clay. 2005. Milk urea nitrogen concentration: Heritability and genetic correlations with reproductive performance. *J. Dairy Sci.* 88:4434–4440.
- Motyl, T., M. Wincenciak, S. Blachowski, W. Kukulska, K. Grzelkowska, M. Podgurniak, M. Kasterka, J. Krzeminski, E. Bartnikowska, E. Skalska, F. Kobrynczuk, M. Kupczynska, M. Katkiewicz, and W. Holska. 1993. Metabolic effect of orotic acid in calves. *J. Vet. Med. A Physiol. Pathol. Clin. Med.* 40:676–689. <https://doi.org/10.1111/j.1439-0442.1993.tb00684.x>.
- Okonkwo, P. O., and J. E. Kinsella. 1969. Orotic acid in food milk powders. *Am. J. Clin. Nutr.* 22:532–534. <https://doi.org/10.1093/ajcn/22.5.532>.
- Østlie, H. M., M. H. Helland, and J. A. Narvhus. 2003. Growth and metabolism of selected strains of probiotic bacteria in milk. *Int. J. Food Microbiol.* 87:17–27. [https://doi.org/10.1016/S0168-1605\(03\)00044-8](https://doi.org/10.1016/S0168-1605(03)00044-8).
- Palmquist, D. L., and A. D. Beaulieu. 1993. Feed and animal factors influencing milk fat composition. *J. Dairy Sci.* 76:1753–1771.
- Robinson, J. L., D. B. Dombrowski, G. W. Harpestad, and R. D. Shanks. 1984. Detection and prevalence of UMP synthase deficiency among dairy cattle. *J. Hered.* 75:277–280. <https://doi.org/10.1093/oxfordjournals.jhered.a109932>.
- Robinson, J. L., M. R. Drabik, D. B. Dombrowski, and J. H. Clark. 1983. Consequences of UMP synthase deficiency in cattle. *Proc. Natl. Acad. Sci. USA* 80:321–323. <https://doi.org/10.1073/pnas.80.2.321>.
- Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. 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>.
- Rutten, M. J. M., H. Bovenhuis, K. A. Hetingga, H. J. F. van Valenberg, and J. A. M. van Arendonk. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. *J. Dairy Sci.* 92:6202–6209. <https://doi.org/10.3168/jds.2009-2456>.
- Saidi, B., and J. J. Warthesen. 1989. Analysis and stability of orotic acid in milk. *J. Dairy Sci.* 72:2900–2905. [https://doi.org/10.3168/jds.S0022-0302\(89\)79440-6](https://doi.org/10.3168/jds.S0022-0302(89)79440-6).
- Shanks, R. D., and J. L. Robinson. 1989. Embryonic mortality attributed to inherited deficiency of uridine monophosphate synthase. *J. Dairy Sci.* 72:3035–3039. [https://doi.org/10.3168/jds.S0022-0302\(89\)79456-X](https://doi.org/10.3168/jds.S0022-0302(89)79456-X).
- Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J. Dairy Sci.* 94:1657–1667. <https://doi.org/10.3168/jds.2010-3408>.
- Sun, H.-Z., K. Shi, X. H. Wu, M. Y. Xue, Z. H. Wei, J. X. Liu, and H. Y. Liu. 2017. Lactation-related metabolic mechanism investigated based on mammary gland metabolomics and 4 biofluids' metabolomics relationships in dairy cows. *BMC Genomics* 18:936. <https://doi.org/10.1186/s12864-017-4314-1>.
- Sundekilde, U. K., N. Poulsen, L. B. Larsen, and H. C. Bertram. 2013. Nuclear magnetic resonance metabolomics reveals strong association between milk metabolites and somatic cell count in bovine milk. *J. Dairy Sci.* 96:290–299. <https://doi.org/10.3168/jds.2012-5819>.
- Vénica, C. I., M. C. Perotti, and C. V. Bergamini. 2014. Organic acids profiles in lactose-hydrolyzed yogurt with different matrix composition. *Dairy Sci. Technol.* 94:561–580.
- Wang, Q., and H. Bovenhuis. 2018. Genome wide association study for milk infrared wavenumbers. *J. Dairy Sci.* 101:2260–2272. <https://doi.org/10.3168/jds.2017-13457>.
- Wang, Y., D. J. Veltkamp, and B. K. Kowalski. 1991. Multivariate instrument standardization. *Anal. Chem.* 63:2750–2756. <https://doi.org/10.1021/ac00023a016>.
- Wevers, R. A., U. F. H. Engelke, S. H. Moolenaar, C. Brautigam, J. G. N. de Jong, R. Duran, R. A. de Abreu, and A. H. van Gennip. 1999. <sup>1</sup>H-NMR spectroscopy of body fluids: Inborn errors of purine and pyrimidine metabolism. *Clin. Chem.* 45:539–548.
- Wilmink, J. B. M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16:335–348. [https://doi.org/10.1016/0301-6226\(87\)90003-0](https://doi.org/10.1016/0301-6226(87)90003-0).

## ORCID

- R. M. Zaalberg  <https://orcid.org/0000-0002-2609-3458>  
 A. J. Buitenhuis  <https://orcid.org/0000-0002-4953-3081>  
 U. K. Sundekilde  <https://orcid.org/0000-0003-4849-0996>  
 N. A. Poulsen  <https://orcid.org/0000-0001-9983-9663>  
 H. Bovenhuis  <https://orcid.org/0000-0002-9074-5334>

## APPENDIX

**Table A1.** Accuracy of prediction ( $R^2$ ) and residual mean squared error of the prediction (RMSEP) for 4 prediction models using different training populations, and trained with unstandardized (UnSt) or standardized (St) milk spectra<sup>1</sup>

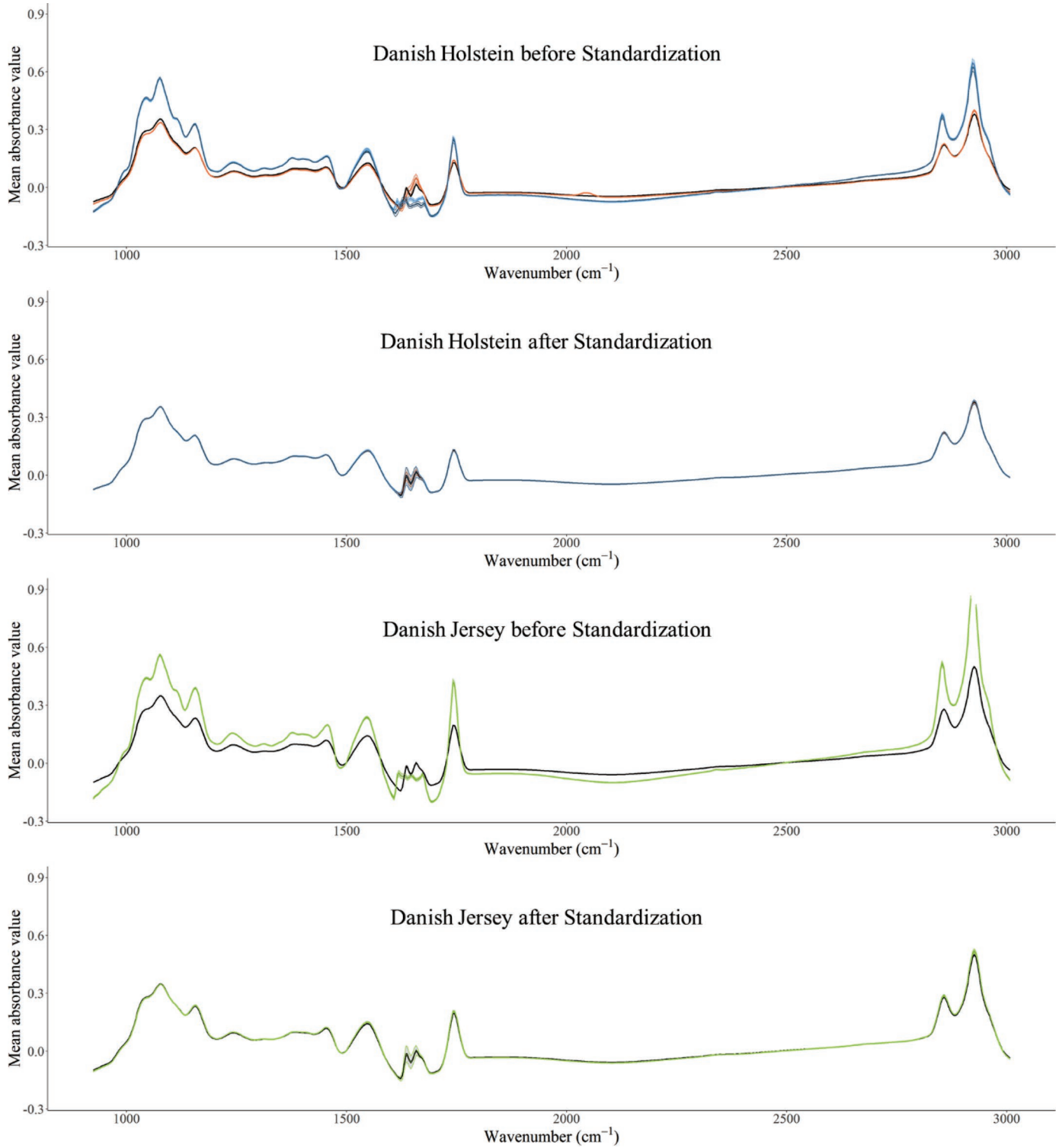
Training population	Validation population	Validation type	$R^2$		RMSEP	
			UnSt	St	UnSt	St
Danish Holstein I	Danish Holstein I	Internal	0.74	0.71	2.2	2.2
	Danish Holstein II	External	0.78	0.72	4.3	5.4
	Danish Jersey	External	0.40	0.46	9.3	10.2
Danish Holstein II	Danish Holstein I	External	0.66	0.44	2.3	7.8
	Danish Holstein II	Internal	0.87	0.89	4.7	2.5
	Danish Jersey	External	0.58	0.48	4.0	11.2
Danish Jersey	Danish Holstein I	External	0.63	0.55	5.3	14.1
	Danish Holstein II	External	0.83	0.74	4.3	10.5
	Danish Jersey	Internal	0.66	0.65	1.8	1.9
Danish Holstein I + Danish Jersey	Danish Holstein I	Internal	0.77	0.76	1.1	4.7
	Danish Holstein II	External	0.86	0.79	3.2	5.5
	Danish Jersey	Internal	0.68	0.60	1.0	3.2

<sup>1</sup>Infrared prediction models trained with one breed used the same procedure as shown in Figure 1, where 80% was used for training and 20% for internal validation of the model.

**Table A2.** Deviance information criterion (DIC) scores for models including all variables, or leaving out one variable at the time

Item	DIC score	
	Danish Holstein	Danish Jersey
Complete model <sup>1</sup>	11,576	9,439
Complete model without $\beta_1 \text{DIM}_k + \beta_2 e^{-0.05 \text{DIM}_k}$	11,837	11,574
Complete model without parity	11,612	9,437
Complete model without herd-year-season (HYS)	12,727	11,103

<sup>1</sup>For the complete model and the model components, see Model 3.



**Figure A1.** Unstandardized and standardized absorbance values for wavenumbers from 3,008 to 925 cm<sup>-1</sup>. Black indicates the Danish Holstein study population (top 2 graphs) and the Danish Jersey study population (bottom 2 graphs). The Danish Holstein I and Danish Holstein II populations are dark and light blue, the Dutch Holstein Friesian population is orange, and the Danish Jersey validation population is green.