Genomic analyses for predicted milk fatty acid composition throughout lactation in North American Holstein cattle

P. H. F. Freitas,^{1,2} H. R. Oliveira,¹ F. F. Silva,² A. Fleming,³ F. Miglior,⁴ F. S. Schenkel,⁴ and L. F. Brito^{1*}

¹Department of Animal Sciences, Purdue University, West Lafayette, IN 47907

²Department of Animal Sciences, Federal University of Viçosa, Viçosa, Minas Gerais, 36570-000, Brazil

³Lactanet Canada, Guelph, ON, N1K 1E5, Canada

⁴Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, N1G 2W1, Canada

ABSTRACT

Milk fat composition has important implications in the nutritional and processing properties of milk. Additionally, milk fat composition is associated with cow physiological and health status. The main objectives of this study were (1) to estimate genetic parameters for 5 milk fatty acid (FA) groups (i.e., short-chain, medium-chain, long-chain, saturated, and unsaturated) predicted from milk infrared spectra using a large data set; (2) to predict genomic breeding values using a longitudinal single-step genomic BLUP approach; and (3) to conduct a single-step GWAS aiming to identify genomic regions, candidate genes, and metabolic pathways associated with milk FA, and consequently, to understand the underlying biology of these traits. We used 629,769 test-day records of 201,465 first-parity Holstein cows from 6,105 herds. A total of 8,865 genotyped (Illumina BovineSNP50K BeadChip, Illumina, San Diego, CA) animals were considered for the genomic analyses. The average daily heritability ranged from 0.24 (unsaturated FA) to 0.47 (medium-chain and saturated FA). The reliability of the genomic breeding values ranged from 0.56 (long-chain fatty acid) to 0.74 (medium-chain fatty acid) when using the default τ and ω scaling parameters, whereas it ranged from 0.58 (long-chain fatty acid) to 0.73 (short-chain fatty acid) when using the optimal τ and ω values (i.e., $\tau = 1.5$ and $\omega = 0.6$), as defined in a previous study in the same population. Relevant chromosomal regions were identified in *Bos taurus* autosomes 5 and 14. The proportion of the variance explained by 20 adjacent single nucleotide polymorphisms ranged from 0.71% (saturated FA) to 15.12% (long-chain FA). Important candidate genes and pathways were also identified. In summary, our results contribute to a better understanding of the genetic architecture of predicted milk FA in dairy cattle

and reinforce the relevance of using genomic information for genetic analyses of these traits.

Key words: longitudinal traits, random regression, single-step genomic BLUP, GWAS

INTRODUCTION

The proportion of fat (i.e., lipids) has great economic and nutritional value in milk composition, and it is directly related to flavor and chemical-physical characteristics of the milk and milk products and to cheese-making properties (Baer, 1991; Bergamaschi et al., 2016). In this regard, fatty acids (FA) are important components of milk fat, comprising 90% of its weight (Samková et al., 2012). The FA vary in chain length (number of carbon atoms), saturation, and arrangement. Milk FA originating from de novo synthesis in the mammary gland are usually classified as short-chain FA (SCFA; from 4 to 8 or 10 carbons) and medium-chain FA (MCFA; 10 or 12 to 14 or 16 carbons) chain, whereas FA from the bloodstream and diet are usually classified as long-chain FA (LCFA; 16 to 18 carbons; Cozma et al., 2013; Fleming et al., 2017). Fatty acids can also be classified as saturated (i.e., FA that have no double bonds between the individual carbon atoms) and unsaturated (i.e., FA that have at least one double bond). In brief, SFA account for 70 to 75% of total fat in cow milk and comprise FA containing from 4 to 18 carbons. The most common SFA are palmitic (C16:0), stearic (C18:0), and myristic (C14:0) acids (MacGibbon and Taylor, 2006). The UFA account for 25 to 30% of the total FA in milk and are characterized by the presence of a *cis*-double bond between carbons 9 and 10, in a carbon chain length that ranges from 10 to 18 (Ntambi, 1995). The lower amount of UFA in milk compared with SFA is due, in particular, to biohydrogenation that occurs in the rumen (Sauer et al., 1998).

Consumer choices today are based not only on the nutritional aspects of the food, but also on products known to promote better health or prevent disease. In this regards, the proportion of SFA in milk is a concern

Received September 23, 2019.

Accepted March 12, 2020.

*Corresponding author: britol@purdue.edu

(Bilal et al., 2014). This is because SFA are associated with increased risk of cardiovascular diseases, obesity, and weight gain, as they can increase levels of total and low-density lipoprotein cholesterol in the blood (Temme et al., 1996; Haug et al., 2007; Zong et al., 2016; Briggs et al., 2017). In contrast, some FA have antiallergic, antimicrobial, anticarcinogenic and anti-inflammatory properties (Williams, 2000; Parodi, 2004; Haug et al., 2007; Pepe et al., 2013; Lordan and Zabetakis, 2017). Therefore, increasing the proportion of favorable FA in milk can contribute to improved human health and disease resistance. In this context, selection for improved FA composition will enable the dairy industry to label products with a more favorable milk FA profile. This could have a positive effect on consumer preferences and, consequently, on industry profitability.

In summary, milk fat composition can be altered in various ways, including dietary changes and genetic selection (Palmquist et al., 1993; Kęsek et al., 2014; Narayana et al., 2017). Changes due to diet are faster to observe than those due to genetic selection; however, the latter results in permanent and cumulative changes at the population level. Thus, to select for milk FA composition, mid-infrared spectroscopy (**MIR**) technology has become an attractive method to accurately predict FA composition in milk (Soyeurt et al., 2006; Narayana et al., 2017). Thereafter, milk FA predicted based on MIR can be used to evaluate animal performance and to predict EBV of the animals along the lactation curve using random regression models (**RRM**; e.g., Narayana et al., 2017; Freitas et al., 2020).

The calculation of EBV for each lactation day enables modeling of the shape of the lactation curve (Oliveira et al., 2019a; Schaeffer et al., 2000), in addition to more precisely identifying external factors that may be affecting the animal production in different stages of lactation (Jensen, 2002). With the advancement of genomics and the availability of high-density SNP panels, it became possible to predict more accurate genomic EBV (**GBEV**) for young animals compared with the traditional parent average (Meuwissen et al., 2001; Hayes et al., 2009). Therefore, especially for sex-limited (e.g., milk production-related traits) and expensive-to-measure traits (e.g., direct individual FA), predicting more accurate GEBV at a young age using relatives' phenotypes predicted based on MIR is a very promising approach. In this context, we aimed to (1) estimate genetic parameters for 5 milk FA groups (i.e., short-chain, medium-chain, long-chain, SFA, and UFA) predicted based on milk spectra of first-parity Holstein cows; (2) perform genomic predictions over time for these 5 traits; and (3) conduct single-step GWAS to identify genomic regions and candidate genes associated with milk FA, and consequently, understand the

underlying biology of these traits based on candidates genes and metabolic pathways of FA predicted by MIR.

MATERIALS AND METHODS

No animal care committee approval was necessary for the purposes of this study, as all information required was obtained from pre-existing databases.

Phenotypic, Genotypic, and Pedigree Data

The pedigree, phenotypic, and genomic data sets were provided by the Canadian Dairy Network (Guelph, ON, Canada). Milk test-day records used in this study were collected between January 2013 and July 2018, from Holstein cows in first lactation. Milk FA composition was predicted for specific days (i.e., test-days in which milk was recorded) from MIR spectra obtained from the routine milk recording systems, using the calibration equations developed by Fleming et al. (2017) as well as the same instrument for spectra acquisition. The coefficients of determination of cross-validation (R_{cv}^2) for the prediction equations were 0.93, 0.83, 0.73, 0.89, and 0.81 for SFA, UFA, SCFA, MCFA, and LCFA groups, respectively (Fleming et al., 2017). In brief, Fleming et al. (2017) analyzed the FA profile of individual milk samples using gas chromatography methodologies, and through a partial least squares regression method the authors developed equations to predict FA concentrations from MIR milk spectra. The FA were classified into 5 groups based on the length of the carbon chain and degree of saturation: (1) SCFA (4 to 10 carbons); (2) MCFA (12 to 16 carbons); (3) LCFA (17 to 22 carbons); (4) SFA (no double bonds); and (5) UFA (one or more double bonds).

Only data from first-parity cows with records from 5 to 305 DIM and calving between 19 and 43 mo of age were kept in the data set for further analyses. The minimum number of records per cow ranged from 2 to 9. Milk FA predictions for samples with MIR spectral data above or below 3 standard deviations from the mean were considered outliers and removed from further analyses. In addition, test-day records assumed as an outlier in at least one FA group were removed from all other FA groups, because it indicates that the FA groups for that specific sample could have been poorly predicted. Herd-test-days with <4 cows, and records above or below 3 standard deviations from the mean, within each herd-test-day, were removed from the analyses. Animals were either genotyped with the Illumina BovineSNP50K BeadChip (Illumina Inc., San Diego, CA) or with a lower-density panel and accurately imputed to 50K using the FImpute software (Sargolzaei

et al., 2014). Genotypic quality control was performed using the preGSf90 software (Aguilar et al., 2014; Misztal et al., 2014), and SNP with Mendelian conflicts, duplicated or unknown position, call rate <0.95, and minor allele frequency <0.05 were removed. Individual genotypes with call rate <0.95 were also filtered out. After the phenotypic and genotypic quality controls, the data sets included 629,769 records from 201,465 animals in 6,105 herds, and 8,865 animals genotyped with 44,368 SNP. The pedigree file included 10 generations back for all phenotyped animals and contained 7,587,436 individuals.

Variance Component Estimation

Variance components for all traits were estimated through Bayesian inference of single-trait RRM under a Markov chain Monte Carlo framework using the GIBBS2F90 software (Misztal et al., 2002). The following RRM was used:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_h\mathbf{h} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_p\mathbf{p} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of longitudinal observations (FA content; g/100 g of milk); $\boldsymbol{\beta}$ is the vector of systematic effects, which included herd-test-day and the systematic regression coefficients for age-season of calving effect; $\mathbf{h}$ is the vector of random regression polynomial coefficients for herd-year of calving effect; $\mathbf{a}$ is the vector of random regression coefficients for animal additive genetic effect; $\mathbf{p}$ is the vector of regression coefficients for permanent environmental effect; and $\mathbf{e}$ is the residual vector; $\mathbf{X}$, $\mathbf{Z}_h$, $\mathbf{Z}_a$, and $\mathbf{Z}_p$ are the corresponding incidence matrices for the mentioned effects. Both systematic and random effects were modeled through third-order Legendre orthogonal polynomials (Kirkpatrick et al., 1990), which were suggested by Narayana et al. (2017) using a subset of the same population.

The data distribution was given by

$$\mathbf{y} | \boldsymbol{\beta}, \mathbf{h}, \mathbf{a}, \mathbf{p}, \mathbf{D}, \mathbf{T}, \mathbf{P}, \sigma_e^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_h\mathbf{h} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_p\mathbf{p}, \mathbf{I}\sigma_e^2),$$

where $\mathbf{D}$, $\mathbf{T}$, and $\mathbf{P}$ are the covariance matrices associated with random regression polynomial coefficients for herd-year of calving, for animal additive genetic, and for permanent environmental effects; $\mathbf{I}$ is the identity matrix, and σ_e^2 is the residual variance, assumed as a scaled inverted chi-squared distribution $\sigma_e^2 | \nu_e, S_e \sim \chi^{-2}(\nu_e, S_e)$, where ν is the number of degrees of freedom and S is the scale parameter. The prior distribution for the vector of systematic effects was $\boldsymbol{\beta} \sim N(0, \boldsymbol{\Sigma}_\beta \otimes \mathbf{I})$, $\boldsymbol{\Sigma}_\beta$ being a known diagonal ma-

trix with values $1e+10$ (large variances) to represent vague prior knowledge. For the random effects, the following prior distributions were assumed: $\mathbf{h} \sim N(0, \mathbf{D} \otimes \mathbf{I})$, $\mathbf{a} \sim N(0, \mathbf{T} \otimes \mathbf{A})$, and $\mathbf{p} \sim N(0, \mathbf{P} \otimes \mathbf{I})$, $\mathbf{I}$ being the identity matrix and $\mathbf{A}$ the additive pedigree-based relationship matrix. Furthermore, we assumed that $\mathbf{D}$, $\mathbf{T}$, and $\mathbf{P}$ follow the inverted Wishart (IW) distributions, $IW(\nu_h, \mathbf{V}_h)$, $IW(\nu_a, \mathbf{V}_a)$, and $IW(\nu_p, \mathbf{V}_p)$, respectively, where ν is the number of degrees of freedom and $\mathbf{V}$ is the scale matrix.

The length of the Gibbs sampler chain was 270,000 cycles, considering a burn-in period of 30,000 cycles, and a sampling interval (thin) of 15 cycles for all traits. The convergence criterion was based on graphical analysis and Raftery and Lewis criterion (Raftery and Lewis, 1992) available in the Bayesian Output Analysis (Smith, 2007) package of the R software (R Core Team, 2018).

The posterior marginal distribution samples for heritabilities over DIM were obtained at each Gibbs sampler iteration (q) as follows:

$$h^{2(q)} = \frac{\sigma_g^{2(q)}}{\sigma_a^{2(q)} + \sigma_h^{2(q)} + \sigma_p^{2(q)} + \sigma_e^{2(q)}},$$

where σ_g^2 is the additive genetic variance, σ_h^2 is the herd-year variance, σ_p^2 is the permanent environmental variance, and σ_e^2 is the residual variance. The h^2 estimate for each trait was obtained as the average of daily h^2 estimates.

Genomic Prediction of Breeding Values

Single-Step Genomic BLUP Approach. The same RRM used for the estimation of variance components were used in the single-step genomic BLUP (ssGBLUP) analyses, except that $\mathbf{A}$ was replaced by the $\mathbf{H}$ matrix. Because the direct estimation of $\mathbf{H}$ is computationally demanding, $\mathbf{H}^{-1}$ was calculated as in Aguilar et al. (2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & \tau(0.95\mathbf{G} + 0.05\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{pmatrix},$$

where $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$, $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix [calculated using the first method proposed by VanRaden (2008) and the observed allele frequencies], and $\mathbf{A}_{22}^{-1}$ is the inverse of the portion

of $\mathbf{A}$ related to genotyped animals. As usual, $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ are not in the same scale, so scaling factors (τ and ω) were used to make the matrices more compatible. The values reported by Oliveira et al. (2019b) for milk fat yield, in the same Holstein population, were used in this study and considered, to differentiate from the default values, to be the optimal values (i.e., $\tau = 1.5$ and $\omega = 0.6$). The ssGBLUP analyses were performed using the BLUPF90 software (Miszta et al., 2014). The variance components previously estimated were assumed as the true variance components for the analyzed population. Animals born between 1957 and 2009 ($n = 8,489$ genotyped animals and 158,765 animals with phenotypes) and between 2010 and 2012 ($n = 376$ genotyped animals) were grouped as the training and validation populations, respectively. Thus, the phenotypes of animals born between 2010 and 2012 were excluded from the analyses, to create a reduced data set to be used to predict the GEBV.

Reliability and Bias of Genomic Predictions.

The GEBV of the validation animals were used to evaluate the reliability (r^2) and bias of genomic predictions for each trait. The reliability of genomic predictions was estimated as the squared Pearson correlation coefficient (r) between daily GEBV, predicted using the reduced data set, and daily EBV predicted using the complete data set (i.e., without excluding phenotypes for the validation animals or their descendants). Using the full-data set EBV instead of pseudo-phenotypes to validate the performance of the genomic prediction relies on the fact that animals in the validation population have reliable breeding values (0.84 ± 0.03). Full-data set EBV have also been used to validate the performance of genomic predictions using RRM in other studies (e.g., Oliveira et al., 2019b,c). The genomic prediction bias was calculated by obtaining the regression coefficient (b_1) estimated using a linear regression of EBV (from the complete data sets) on GEBV (from the reduced data sets); that is, $\text{EBV} = b_0 + b_1 \times \text{GEBV} + e$. Only animals from the validation population were used to calculate r^2 and b_1 . To analyze the advantages of including genomic information in the genetic evaluation of milk FA, the parent average (PA) was predicted for the validation animals; PA was used to calculate r^2 and b_1 using daily PA (predicted using the reduced data set) and daily EBV (predicted using the complete data set) for the animals in the validation population.

GWAS and Functional Analyses. Genome-wide association analyses were performed using the postGSf90 software (Aguilar et al., 2014), considering the variance explained by moving windows of 20 adjacent SNP. The postGSf90 software back-solves the additive genomic random regression coefficients (i.e., the GEBV for the

additive random regression coefficients) to SNP effects and it can be described as follows (Wang et al., 2012):

$$\hat{\mathbf{u}}_c = \mathbf{Z}'[\mathbf{ZZ}']^{-1} \hat{\mathbf{a}}_{gc},$$

where $\hat{\mathbf{u}}_c$ is the vector of estimated SNP solutions for the c th random regression coefficient; $\mathbf{Z}$ is the matrix containing the centered genotypes; and $\hat{\mathbf{a}}_{gc}$ is the vector of GEBV for the c th random regression coefficient, estimated in the ssGBLUP analyses, which contains the c th random regression coefficients for all the genotyped animals. Further, to calculate the effect of each SNP along the lactation, the SNP-specific solutions for all random regression coefficients ($c = 1, 2, 3, 4$) were combined into a vector ($\hat{\mathbf{u}}_k = [\hat{\mathbf{u}}_{k1} \hat{\mathbf{u}}_{k2} \hat{\mathbf{u}}_{k3} \hat{\mathbf{u}}_{k4}]$), and used to estimate the SNP effects for all DIM (from 5 to 305), as suggested by Oliveira et al. (2019b):

$$\widehat{\mathbf{SNP}}_k = \mathbf{T} \hat{\mathbf{u}}_k,$$

where $\widehat{\mathbf{SNP}}_k$ is the vector that contains the effect of the k th SNP on the trait estimated for each DIM, $\mathbf{T}$ is a matrix of covariates for each DIM, associated with the third-order Legendre orthogonal polynomials, and $\hat{\mathbf{u}}_k$ is the vector of SNP solutions for all random regression coefficients related to the k th SNP. The percentage of genetic variance explained by the i th region was calculated as described by Wang et al. (2014):

$$\frac{\text{Var}(a_i)}{\sigma_a^2} \times 100 = \frac{\text{Var}\left(\sum_{j=1}^{20} \mathbf{Z}_j \hat{\mathbf{u}}_j\right)}{\sigma_a^2} \times 100,$$

where a_i is the genetic value of the i th region that consisted of 20 consecutive SNP, σ_a^2 is the total genetic variance, $\mathbf{Z}_j$ is the matrix that contains the centered genotypes for the j th SNP, and $\hat{\mathbf{u}}_j$ is the marker effect of the j th SNP within the i th region.

A threshold of 1% of the total genetic variance explained by each genomic window was used to define the important genomic regions associated with the traits included in this study. Positional candidate genes were mapped using the Biomart tool (Kinsella et al., 2011) embedded in the Ensembl Genes database version 96 (<http://useast.ensembl.org/index.html>). Based on the start and end chromosomal positions, important genomic regions were further investigated to understand the biological processes related to the studied traits and to define the most likely functional candidate genes. Complete gene functions were obtained from the Na-

tional Center for Biotechnology Information database (www.ncbi.nlm.nih.gov/gene/) based on the *Bos taurus* reference genome ARS-UCD1.2 (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1/). The biological functions and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa, 2000; Kanehisa et al., 2017, 2019) in which these genes are involved were assessed using the Database for Annotation, Visualization and Integrated Discovery v6.8. (DAVID, <https://david.ncifcrf.gov/home.jsp>; Huang et al., 2009a,b).

RESULTS

Descriptive Statistics

Table 1 shows the descriptive statistics for the 5 groups of milk FA. Higher proportions of MCFA, followed by LCFA and SCFA (lowest) were observed. On average, the proportions of LCFA, MCFA, SCFA, SFA, and UFA were 33.85, 55.5, 10.60, 73.2, and 26.8%, respectively.

Genetic Parameters

The average daily h^2 estimates for each FA are shown in Table 2, and the complete pattern of daily h^2 across the lactation is shown in Figure 1. The average daily h^2 ranged from 0.24 (UFA) to 0.47 (SFA and MCFA). In general, h^2 observed at the beginning of the lactation was similar for all 5 FA groups (~ 0.20). For LCFA, the h^2 estimates increased until the first third of the lactation, stabilized until the end of the lactation, and finally slightly decreased. For MCFA and SCFA, the h^2 pattern was similar, presenting an increase from the middle of the lactation to the end. Following the same pattern of MCFA and SCFA, the h^2 curve for SFA increased at the beginning of lactation and continued to increase until reaching the maximum value at the end of lactation. For UFA and LCFA, daily h^2 values had a slight increase at the beginning of the lactation and

Table 2. Mean and SD of daily heritability estimates for 5 groups of milk fatty acids in the Holstein breed

Fatty acid group ¹	Mean	SD
LCFA	0.25	0.07
MCFA	0.47	0.10
SCFA	0.31	0.09
SFA	0.47	0.09
UFA	0.24	0.07

¹Fatty acid groups were defined as long-chain fatty acids (LCFA; 17 to 22 carbons), medium-chain fatty acids (MCFA; 12 to 16 carbons), short-chain fatty acids (SCFA; 4 to 10 carbons), SFA (no double bond), and UFA (one or more double bonds).

after reaching its peak, decreased until the end of the curve.

Reliability and Bias of Genomic Predictions

Average validation r^2 and their standard deviations considering both analyses, with and without (i.e., considering the default values) optimal τ and ω , are shown in Table 3. When considering τ and ω equal to 1 (default), the average r^2 ranged from 0.564 (LCFA) to 0.737 (SCFA); when using the optimal τ and ω values, the GEV r^2 ranged from 0.583 (LCFA) to 0.732 (SCFA). The inclusion of optimal values of τ and ω slightly changed (in both directions) the average r^2 , as observed for LCFA and UFA where there was an increase in r^2 ; however, a slight decrease was observed for the other 3 traits. For all traits, the inclusion of genomic information increased the average r^2 in both scenarios (with and without the optimal values for τ and ω).

The pattern of GEV r^2 curves over time using default and optimal values for τ and ω are shown in Figure 2. The GEV r^2 were approximately constant over DIM for all groups of FA. Slight differences in the pattern of the GEV r^2 curves were observed when including, or not, τ and ω in the analyses. The average b_1 (indicator of bias) and its respective standard deviations are shown in Table 4. The average b_1 for PA ranged from 0.690 (UFA) to 1.062 (SCFA). When no optimal values of τ and ω were used (i.e., the default values were applied), the b_1 coefficients ranged from 0.756 (LCFA) to 0.905 (SCFA). When the optimal values of τ and ω were used, b_1 ranged from 1.075 (LCFA) to 1.172 (SCFA). In summary, an improvement in the bias (i.e., b_1 closer to 1) was observed for all traits when optimal τ and ω were used. The pattern of bias over time using the default and the optimal values for τ and ω are shown in Figure 3. Similar pattern of bias over time were found when using, or not, the optimal values of τ and ω . In general, GEV predicted over time without the optimal values were deflated, and GEV predicted over time using the optimal values were inflated.

Table 1. Descriptive statistics for the different groups of milk fatty acids in the Holstein breed

Fatty acid group ¹ (g/100 g of milk)	Mean	SD	Minimum	Maximum
LCFA	1.331	0.39	0.143	4.480
MCFA	2.183	0.44	0.513	4.497
SCFA	0.417	0.09	0.063	0.848
SFA	2.896	0.53	0.702	5.689
UFA	1.065	0.26	0.171	3.596

¹Fatty acid groups were defined as long-chain fatty acids (LCFA; 17 to 22 carbons), medium-chain fatty acids (MCFA; 12 to 16 carbons), short-chain fatty acids (SCFA; 4 to 10 carbons), SFA (no double bond), and UFA (one or more double bonds).

Figure 1. Daily heritability estimates of long-chain, medium-chain, short-chain, saturated, and unsaturated milk fatty acid groups from 5 to 305 DIM of first-parity Holstein cows.

GWAS and Functional Analyses

A total of 39 genomic windows were associated with the 5 FA groups. These genomic regions were located in chromosomes BTA5, BTA13, and BTA14, and the maximum proportion of the additive genetic variance explained was 15.12% (LCFA) on BTA14. The largest variances explained were found in BTA14, indicating that this chromosome contains important genes associated with milk FA groups. All common candidate genes mapped through Ensembl considering 20-SNP windows for the SNP selected in each trait, using default and optimal τ and ω values, are shown in Table 5. Table 6 shows the different candidate genes discovered between the 2 scenarios (default and optimal) for τ and ω . The biological pathways in which these genes are involved are shown in Table 7. The main biological pathways

are carbon metabolism, lipid metabolism, cellular lipid metabolic process, and lipid catabolic process. The Manhattan plots of the proportion of additive genetic

Table 3. Mean validation reliability (and SD) estimated for parent average (r^2_{PA}) and genomic breeding values predicted using ($r^2_{GEBV\tau\omega}$) or not (r^2_{GEBV}) the optimal values¹ for the scaling factors τ and ω , for 5 groups of milk fatty acids

Trait ²	r^2_{PA}	r^2_{GEBV}	$r^2_{GEBV\tau\omega}$
SCFA	0.59 (0.02)	0.74 (0.02)	0.73 (0.01)
SFA	0.55 (0.02)	0.70 (0.02)	0.68 (0.01)
MCFA	0.55 (0.02)	0.70 (0.02)	0.68 (0.01)
UFA	0.44 (0.05)	0.59 (0.04)	0.62 (0.03)
LCFA	0.41 (0.03)	0.56 (0.03)	0.58 (0.03)

¹The optimal values assumed were $\tau = 1.5$ and $\omega = 0.6$.

²SCFA = short-chain fatty acids; MCFA = medium-chain fatty acids; LCFA = long-chain fatty acids.

variance explained for each third-order Legendre coefficients are shown in Supplemental Figures S1 to S10 (<https://doi.org/10.3168/jds.2019-17628>). The Manhattan plots show that the peaks of the markers were similar when scaling factors with optimal values or default values (equal to 1) of τ and ω were used.

A total of 444 SNP were selected as the most relevant (top 1%) for each trait. When considering τ and ω equal to 1 (default), the proportion of variance explained by the 20-SNP window ranged from 0.72% (UFA) to 15.12% (LCFA); when using the optimal

τ and ω values, the proportion of variance explained ranged from 0.71% (SFA) to 15.10% (LCFA). In relation to SNP effects over DIM, when considering default values for τ and ω , it ranged (in module) from 0.57% (SCFA) to 14.96% (SFA). When using the optimal τ and ω values, the SNP effect over DIM, it ranged from 0.59% (SCFA) to 14.97% (SFA). Both the largest SNP effects and the genomic windows explaining the highest proportion of the additive genetic variance were found for the LCFA group. The trajectory of SNP effects over DIM, estimated for the top 10 SNP of each trait

Figure 2. Daily reliabilities using scaling factors τ and ω equal to 1 (default) and optimal values for long-chain, medium-chain, short-chain, saturated, and unsaturated milk fatty acid groups from 5 to 305 DIM of first-parity Holstein cows

Table 4. Regression coefficients (and SD) estimated for parent average (b_{PA}), and genomic EBV (GEBV) predicted using ($b_{GEBV\tau\omega}$) or not (b_{GEBV}) the optimal values¹ for the scaling factors τ and ω , for 5 groups of milk fatty acids

Trait ²	b_{PA}	b_{GEBV}	$b_{GEBV\tau\omega}$
SCFA	1.062 (0.025)	0.905 (0.046)	1.172 (0.028)
SFA	1.030 (0.018)	0.883 (0.035)	1.140 (0.021)
MCFA	1.054 (0.018)	0.759 (0.032)	1.164 (0.020)
UFA	0.690 (0.052)	0.800 (0.052)	1.120 (0.033)
LCFA	0.965 (0.021)	0.756 (0.045)	1.075 (0.022)

¹The optimal values assumed were $\tau = 1.5$ and $\omega = 0.6$.

²SCFA = short-chain fatty acids; MCFA = medium-chain fatty acids; LCFA = long-chain fatty acids.

considering τ and ω equal to 1 and optimal values, are shown in Supplemental Figure S11 (<https://doi.org/10.3168/jds.2019-17628>) and Figure 4, respectively. The chromosome and position of the top 10 SNP are shown in Supplemental Table S1 (<https://doi.org/10.3168/jds.2019-17628>). The pattern of the SNP effects over lactation for LCFA, SCFA, and UFA tended to remain constant, with small changes in magnitude. Higher deviations in the pattern of SNP effects were observed for MCFA and SFA, where the effects of SNP increased as lactation progressed.

DISCUSSION

Descriptive Statistics and Genetic Parameters

Current recommendations are for consumers to reduce SFA intake and consume more UFA (FAO, 2010). Coupled with an appropriate feeding system, a reduction in the proportion of SFA in milk and an increase in the proportion of FA that promote good health, may therefore be of great interest in dairy cattle (Fleming, 2016). Previous authors have indicated that higher prediction model accuracies were obtained when defining the FA measurement unit on a per-milk basis versus a per-fat basis (Soyeurt et al., 2006; Rutten et al., 2009; Fleming et al., 2017). This might be due to the fact that milk samples contain different amounts of total fat, and samples with the same relative concentrations of FA can contain very different total quantities of the FA (Fleming, 2016). Therefore, the unit used in this study was grams per 100 g of milk. Fleming et al. (2017) reported mean values of FA based on GC analysis of 2.81, 1.10, 0.464, 2.09, and 1.40 g/dL of milk for SFA, UFA, SCFA, MCFA, and LCFA, respectively. These estimates are in a similar range to the averages (based on MIR predictions) observed in this study. Milk FA are highly correlated with milk fat percentage and fat yield (e.g., Fleming et al., 2018). If the breeding goal is to alter a specific FA profile, then expressing the traits on a grams per 100 g of fat basis could be a good ap-

proach. However, we might be interested in increasing the amount of certain FA in the milk (independently of fat yield). Because predictions based on grams per 100 g of fat are less accurate (e.g., Soyeurt et al., 2006; Rutten et al., 2009; Fleming et al., 2017), this could greatly limit their use in breeding programs. Therefore, we used grams per 100 g of milk in this study.

The heritability estimates obtained for all 5 FA groups indicated that they are under moderate genetic control; therefore, it is genetically possible to manipulate milk FA proportions. The higher h^2 estimates observed for de novo synthesized SCFA compared with dietary- and adipose-derived LCFA was expected. Our results were mostly in agreement with the literature (Bastin et al., 2011; Narayana et al., 2017; Hein et al., 2018). The h^2 estimates were similar to those found by Narayana et al. (2017), using data from the same population of first-parity Canadian Holsteins (data subset) and FA MIR prediction equations for LCFA and UFA (0.23 and 0.21, respectively), but were slightly higher for MCFA (0.32), SCFA (0.24) and SFA (0.33). In contrast, the values for these FA groups were similar to those reported by Bastin et al. (2011) in Holstein cattle from the Walloon Region of Belgium, who found h^2 of 0.44 and 0.43 for MCFA and SFA, respectively. These differences may be because the present study had more observations than the aforementioned and the distinct populations used in each study. In addition, the accuracy of the calibration equations can affect the results, as reported by Fleming et al. (2017). In the current data set, a limited number of test-days were available for some animals because MIR spectra was recorded for only a proportion of all milk samples during DHI milk recording. Consequently, there were fewer records per cow and the traits may not have been as well modeled along one lactation compared with Bastin et al. (2011).

Reliability and Bias of Genomic Predictions

The τ and ω parameters are used to make $\mathbf{G}^{-1}$ more compatible with $\mathbf{A}_{22}^{-1}$ and $\mathbf{A}^{-1}$ (Misztal et al., 2017; Oliveira et al., 2019b). In this context, Misztal et al. (2017) showed that τ is used to account for the reduced genetic variance, whereas ω is used to account for different depths of pedigree. For all traits, the GEBV r^2 were higher than the GEBV r^2 calculated for PA, even when no scaling factors were used. These results are in agreement with other studies, such as Oliveira et al. (2019b) and Baba et al. (2017), which showed that the use of genomic information through ssGBLUP increases the r^2 of breeding values. It is also worth noting that the animals used for these analyses have highly accurate EBV for all traits (and consequently highly accurate PA).

Figure 3. Daily bias (b_1) using scaling factors τ and ω equal to 1 (default) and optimal values for long-chain, medium-chain, short-chain, saturated, and unsaturated fatty acid groups from 5 to 305 DIM of first-parity Holstein cows.

With regards to b_1 , the use of optimal τ and ω scaling factors yielded less biased GEBV (i.e., b_1 closer to 1) compared with analyses using the default values (τ and $\omega = 1$). In general, GEBV predicted over time with and without the optimal values were inflated and deflated, respectively. However, GEBV estimated for the majority of traits in this study would be accepted for use in the international comparisons performed by Interbull (Interbull, 2016), if estimated with or without the scaling factors. Compared with PA, GEBV were more biased for the majority of traits. In this context,

Lourenco et al. (2017) showed that the use of a genomic relationship matrix weighting markers according to the analyzed trait can generate more reliable and less biased estimates. Thus, posterior studies using the weighted ssGBLUP are needed.

Among the advantages of the method proposed in this study to evaluate FA is the fact that GEBV and PA are predicted for all DIM. Thus, different selection criteria can be defined in the breeding schemes. In this regard, it becomes possible to change the pattern of the genetic (and phenotypic) FA lactation curves. Increase

Table 5. Chromosome information and common candidate gene symbols mapped through Ensembl for SNP considering the 20-SNP windows selected in each trait between default and optimal values of the scaling factors τ and ω ¹

Group ²	Common genes
LCFA	BTA14: <i>LY6D</i> , ³ <i>LYNX1</i> , <i>LYPD2</i> , <i>SLURP1</i> , <i>TSNARE1</i> , <i>SLC45A4</i> , <i>PTK2</i> , ³ <i>AGO2</i> , ³ <i>ARC</i> , <i>JRK</i> , <i>TRAPPC9</i> , ³ <i>FAM135B</i> ³
MCFA	BTA5: <i>EPS8</i> , <i>RERG</i> , <i>ACO2</i> , ³ <i>PLBD1</i>
	BTA13: <i>ITCH</i>
	BTA14: <i>LYNX1</i> , <i>SLURP1</i> , <i>TSNARE1</i> , <i>PTK2</i> , ³ <i>AGO2</i> ³
SCFA	BTA5: <i>ERP27</i> , <i>ART4</i> , <i>PLBD1</i>
	BTA14: <i>LY6D</i> , ³ <i>LYNX1</i> , <i>LYPD2</i> , <i>SLURP1</i> , <i>TSNARE1</i> , <i>SLC45A4</i> , <i>PTK2</i> , ³ <i>AGO2</i> , ³ <i>ARC</i> , <i>JRK</i> , <i>TRAPPC9</i> , ³ <i>FAM135B</i> ³
SFA	BTA5: <i>RERGL</i> , <i>EPS8</i> , <i>RERG</i> , <i>ARHGDIB</i> , <i>GUCY2C</i>
	BTA14: <i>LY6D</i> , ³ <i>LYNX1</i> , <i>LYPD2</i> , <i>SLURP1</i> , <i>TSNARE1</i> , <i>ARC</i> , <i>JRK</i> , <i>SLC45A4</i> , <i>PTK2</i> , ³ <i>AGO2</i> , ³ <i>TRAPPC9</i> , ³ <i>KCNK9</i> ³
UFA	BTA14: <i>LY6D</i> , ³ <i>LYNX1</i> , <i>LYPD2</i> , <i>SLURP1</i> , <i>TSNARE1</i> , <i>SLC45A4</i> , <i>PTK2</i> , ³ <i>AGO2</i> , ³ <i>ARC</i> , <i>JRK</i> , <i>TRAPPC9</i> , ³ <i>KCNK9</i> ³

¹The default values were τ and ω equal to 1, and the optimal values assumed were $\tau = 1.5$ and $\omega = 0.6$.²LCFA = long-chain fatty acids; MCFA = medium-chain fatty acids; SCFA = short-chain fatty acids.³Genes previously reported to be related to milk fat traits.**Table 6.** Chromosome information and different candidate gene symbols mapped through Ensembl for SNP considering the 20-SNP windows selected in each trait between default and optimal values of the scaling factors τ and ω ¹

Group ²	Default τ and ω	Optimal τ and ω
LCFA	BTA14: <i>ARC</i> , <i>JRK</i> , <i>TRAPPC9</i> , ³ <i>FAM135B</i> ³	—
MCFA	BTA5: <i>RERGL</i> , <i>ARHGDIB</i> , <i>ERP27</i> , <i>ART4</i> , <i>SMCO3</i>	BTA5: <i>GUCY2C</i>
	BTA13: <i>ITCH</i>	BTA14: <i>ARC</i> , <i>JRK</i> , <i>GPR20</i>
	BTA14: <i>LY6D</i> , ³ <i>LYPD2</i> , <i>SLC45A4</i> , <i>TRAPPC9</i> , ³ <i>FAM135B</i> ³	
SCFA	BTA5: <i>RERGL</i> , <i>EPS8</i> , <i>RERG</i> , <i>ARHGDIB</i> , <i>ERP27</i> ;	BTA5: <i>ERP27</i> , <i>WBP11</i> , <i>GUCY2C</i>
	BTA13: <i>DIP2C</i> , <i>PRNP</i> , <i>PRND</i>	BTA14: <i>FAM135B</i> ³
SFA	BTA5: <i>RERGL</i> , <i>EP27</i> , <i>ART4</i> , <i>SMCO3</i> , <i>PLBD1</i>	BTA14: <i>ST3GAL1</i> , <i>NDRG1</i> , <i>SLA</i>
UFA	BTA14: <i>FAM135B</i> ³	—

¹The default values were τ and ω equal to 1, and the optimal values assumed were $\tau = 1.5$ and $\omega = 0.6$.²LCFA = long-chain fatty acids; MCFA = medium-chain fatty acids; SCFA = short-chain fatty acids.³Genes previously reported to be related to milk fat traits.**Table 7.** Biological functions and pathways (Kyoto Encyclopedia of Genes and Genomes database) of selected genes from long-chain, medium-chain, short-chain, saturated, and unsaturated milk fatty acids¹

Gene symbol	Gene name	Biological pathway
<i>ACO2</i>	Aconitase 2	Citrate cycle (tricarboxylic acid cycle), Glyoxylate and dicarboxylate metabolism, Carbon metabolism, ¹ 2-Oxocarboxylic acid metabolism, biosynthesis of amino acids
<i>ARC</i>	Activity regulated cytoskeleton associated protein	Cytoskeleton organization, Anterior/posterior pattern specification, Cell migration, Regulation of cell morphogenesis
<i>ARHGDIB</i>	Rho GDP dissociation inhibitor beta	Neurotrophin signaling pathway, Vasopressin-regulated water reabsorption
<i>DIP2C</i>	Disco interacting protein 2 homolog C	Lipid metabolism, ¹ Secondary metabolites biosynthesis, transport, and catabolism
<i>FAM135B</i>	Family with sequence similarity 135 member B	Cellular lipid metabolic process ¹
<i>ITCH</i>	Itchy E3 ubiquitin protein ligase	Ubiquitin mediated proteolysis, Endocytosis, tumor necrosis factor signaling pathway, Non-alcoholic fatty liver disease (NAFLD)
<i>KCNK9</i>	Potassium two pore domain channel subfamily k member 9	Aldosterone synthesis and secretion
<i>PLBD1</i>	Phospholipase B domain containing 1	Lipid catabolic process, ¹ Phospholipase B-like
<i>PRNP</i>	Prion protein	Prion diseases
<i>PTK2</i>	Protein tyrosine kinase 2	ErbB signaling pathway, Chemokine signaling pathway, PI3K-Akt signaling pathway, Axon guidance, Vascular endothelial growth factor (VEGF) signaling pathway, Focal adhesion, Leukocyte transendothelial migration, Regulation of actin cytoskeleton, Bacterial invasion of epithelial cells

¹Biological pathways related to milk fat traits.

in GEBV r^2 can be achieved by increasing the numbers of both genotyped and phenotyped animals and the use of more accurate FA prediction equations (developed using a larger number of animals with gold standard FA measurements). Different τ and ω values may affect the GEBV obtained and, therefore, may likely change the SNP effects. Here, optimal τ and ω values were used that maximized the observed accuracy of the GEBVs for milk fat yield in the same Holstein population.

GWAS and Functional Analyses

This study supports the polygenic nature of the studied traits, with many regions across the genome making small contributions to total genetic variation. The coefficients of the Legendre polynomial, despite presenting

differences in the patterns of markers identified, have no biological explanation, because they are used to model the lactation curve. Thus, some coefficients allow identification of different markers at different points in the curve. Important contributions were found on BTA14. This chromosome is well known to have a substantial number of QTL for milk and fat production traits (Viitala et al., 2003; Wang et al., 2013), including *DGAT1*, a gene involved in the last step of the synthesis of triacylglycerol and that has a major effect on milk fat content (Grisart, 2002; Cruz et al., 2019). Despite its importance, *DGAT1* was not found in this study because of the absence of markers located in its proximity. In addition, *DGAT1* has changed position in the new cattle reference genome (ARS-UCD1.2): from BTA14: 1,795,425–1,804,838 Mb (Bos_taurus_UMD_3.1.1) to

Figure 4. Single nucleotide polymorphism effect patterns of the top 10 SNP over DIM for long-chain (LCFA), medium-chain (MCFA), short-chain (SCFA), saturated (SFA), and unsaturated (UFA) milk fatty acids using optimal values of scaling factors τ and ω . SNPSol = effect of the SNP.

BTA14:603,981–614,153 Mb (ARS-UCD1.2). However, the peak observed at the beginning of BTA14 is likely in high linkage disequilibrium with markers located in *DGAT1*. In fact, conducting a GWAS for milk FA in the same population, Cruz et al. (2019) identified *DGAT1* when using the previous reference genome.

Candidate genes were identified and shown to have important biological functions associated with the traits under study, including lipid metabolism and cholesterol synthesis. For instance, *PTK2*, associated with all FA, is related to several signal transduction pathways, such as cell motility and microtubule stability (Harte et al., 1996; Ezratty et al., 2005), and an important trait statistic related to fat percentage in Chinese Holstein cattle (Wang et al., 2013). Similarly to *PTK2*, *TRAPPC9* is associated with all the FA traits. Wang et al. (2015) reported that *TRAPPC9* plays a role in regulation of cell differentiation and is associated with mastitis susceptibility in Chinese Holstein cattle. Moreover, *ACO2* is related to the second step of the tricarboxylic acid cycle (Palombo et al., 2018). This gene encodes a protein that acts in the mitochondrion to catalyze the interconversion of citrate to isocitrate via *cis*-aconitate.

Some genes were found to be related to all FA groups. For instance, *SLURP1* was reported by Zhao et al. (2019), who suggested that this gene may play an important role in the regulation of triglycerides through secreted frizzled-related protein 5, an anti-inflammatory adipocytokine secreted by adipocytes in mice. Kim et al. (2011) reported that *AGO2* has a leading function in microRNA-induced RNA silencing, which is critical for stem cell self-renewal, development, and other functions in humans. Also related to all traits, *TSNARE1* is reported to be associated with inflammatory processes in humans (Lighthart et al., 2016; Arpón et al., 2019).

Biological pathways related to lipid catabolism and metabolism have been identified and present important processes in the animal body. The pathway “Lipid catabolism” has FA as a direct product. The FA are catabolized in the β -oxidation process, sequentially removing 2-carbon acetyl groups from the fatty acid chain ends, reducing nicotinamide adenine dinucleotide (NAD^+) and flavin adenine dinucleotide (FAD) to produce NADH and FADH_2 , respectively, and these electrons can be used in energy production by the oxidative phosphorylation process (Stryer, 1995). Another important biological pathway identified is related to “Lipid metabolism.” Keller et al. (1993) showed that FA, especially polyunsaturated FA, have great potential in activating the peroxisome proliferator-activated receptors, the receptors responsible for regulating β -oxidation of FA, thus preventing them from accumulating at lethal levels in blood.

CONCLUSIONS

All 5 groups of MIR-predicted milk fatty acids (SCFA, MCFA, LCFA, SFA, and UFA) are under moderate genetic control in the North American Holstein breed, with heritability estimates varying throughout lactation. Genetic improvement can be made in all of these traits through selection. Moderately to highly accurate genomic breeding values can be obtained for daily points throughout lactation using ssGBLUP. Inclusion of scaling factors improved the accuracy for MCFA, UFA, and LCFA. Various genomic regions are associated with the traits, but many regions across the genome have small individual effects on the total genetic variation of each trait. Candidate genes and pathways were identified, which helps in our understanding of the genetic background of milk fatty acid composition. These results contribute to a better understanding of the genetic architecture of predicted milk fatty acid in dairy cattle and will be of great value for the implementation of genomic selection for milk fatty acid composition.

ACKNOWLEDGMENTS

The first author acknowledges the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; Brasília, DF, Brazil). Funding for the collection of milk spectra was provided by Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network (Ottawa, ON, Canada), and the Canadian Dairy Commission under the Agri-Science Clusters Initiative (Dairy Farmers of Canada, Agriculture and Agri-Food Canada, the Canadian Dairy Network, and the Canadian Dairy Commission). The authors have not stated any conflicts of interest.

REFERENCES

- Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752. <https://doi.org/10.3168/jds.2009-2730>.
- Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Pages 1–3 in *Proc. 10th World Congr. Genet. Appl. Livest. Prod.*, Vancouver, BC, Canada. American Society of Animal Science, Champaign, IL.
- Arpón, A., F. I. Milagro, O. Ramos-Lopez, M. L. Mansego, J.-I. Riezu-Boj, and J. A. Martínez. 2019. Methylome-wide association study in peripheral white blood cells focusing on central obesity and inflammation. *Genes (Basel)* 10:444. <https://doi.org/10.3390/genes10060444>.
- Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2017. Application of single-step genomic best linear unbiased prediction with a multiple-lactation random

- regression test-day model for Japanese Holsteins. *Anim. Sci. J.* 88:1226–1231. <https://doi.org/10.1111/asj.12760>.
- Baer, R. J. 1991. Alteration of the fatty acid content of milk fat. *J. Food Prot.* 54:383–386.
- Bastin, C., N. Gengler, and H. Soyeurt. 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. *J. Dairy Sci.* 94:4152–4163. <https://doi.org/10.3168/jds.2010-4108>.
- Bergamaschi, M., G. Stocco, C. Valorz, I. Bazzoli, E. Sturaro, and M. Ramanzin. 2016. Cheesemaking in highland pastures: Milk technological properties, cream, cheese and ricotta yields, milk nutrients recovery, and products composition. *J. Dairy Sci.* 99:9631–9646.
- Bilal, G., R. I. Cue, A. F. Mustafa, and J. F. Hayes. 2014. Short communication: Genetic parameters of individual fatty acids in milk of Canadian Holsteins. *J. Dairy Sci.* 97:1150–1156. <https://doi.org/10.3168/jds.2012-6508>.
- Briggs, M. A., K. S. Petersen, and P. M. Kris-Etherton. 2017. Saturated fatty acids and cardiovascular disease: Replacements for saturated fat to reduce cardiovascular risk. *Healthcare (Basel)* 5:29. <https://doi.org/10.3390/healthcare5020029>.
- Cozma, A., D. Miere, L. Filip, R. Banc, and F. Loghin. 2013. A review of the metabolic origins of milk fatty acids. *Not. Sci. Biol.* 5:270–274.
- Cruz, V. A. R., H. R. Oliveira, L. F. Brito, A. Fleming, S. Larmer, F. Miglior, and F. S. Schenkel. 2019. Genome-wide association study for milk fatty acids in Holstein cattle accounting for the DGAT1 gene effect. *Animals (Basel)* 9:997. <https://doi.org/10.3390/ani9110997>.
- Ezratty, E. J., M. A. Partridge, and G. G. Gundersen. 2005. Microtubule-induced focal adhesion disassembly is mediated by dynamin and focal adhesion kinase. *Nat. Cell Biol.* 7:581–590. <https://doi.org/10.1038/ncb1262>.
- Fleming, A. 2016. Phenotypic and genetic variation of milk fat components incorporating mid-infrared technology. PhD Thesis. Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada.
- Fleming, A., F. S. Schenkel, J. Chen, F. Malchiodi, V. Bonfatti, R. A. Ali, B. Mallard, M. Corredig, and F. Miglior. 2017. Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. *J. Dairy Sci.* 100:5073–5081. <https://doi.org/10.3168/jds.2016-12102>.
- Fleming, A., F. S. Schenkel, F. Malchiodi, R. A. Ali, B. Mallard, M. Sargolzaei, J. Jamrozik, J. Johnston, and F. Miglior. 2018. Genetic correlations of mid-infrared-predicted milk fatty acid groups with milk production traits. *J. Dairy Sci.* 101:4295–4306. <https://doi.org/10.3168/jds.2017-14089>.
- FAO (Food and Agriculture Organization of the United Nations). 2010. Fats and fatty acids in human nutrition. Report of an expert consultation. 91. FAO, Rome, Italy.
- Freitas, P. H. F., H. R. Oliveira, F. F. Silva, A. Fleming, F. S. Schenkel, F. Miglior, and L. F. Brito. 2020. Short communication: Time-dependent genetic parameters and single-step genome-wide association analyses for predicted milk fatty acid composition in Ayrshire and Jersey dairy cattle. *J. Dairy Sci.* 103:5263–5269. <https://doi.org/10.3168/jds.2019-17820>.
- Grisart, B. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res.* 12:222–231. <https://doi.org/10.1101/gr.224202>.
- Harte, M. T., J. D. Hildebrand, M. R. Burnham, A. H. Bouton, and J. T. Parsons. 1996. p130Cas, a substrate associated with v-Src and v-Crk, localizes to focal adhesions and binds to focal adhesion kinase. *J. Biol. Chem.* 271:13649–13655. <https://doi.org/10.1074/jbc.271.23.13649>.
- Haug, A., A. T. Høstmark, and O. M. Harstad. 2007. Bovine milk in human nutrition—A review. *Lipids Health Dis.* 6:25. <https://doi.org/10.1186/1476-511X-6-25>.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. (Camb.)* 91:47–60. <https://doi.org/10.1017/S0016672308009981>.
- Hein, L., L. P. Sørensen, M. Kargo, and A. J. Buitenhuis. 2018. Genetic analysis of predicted fatty acid profiles of milk from Danish Holstein and Danish Jersey cattle populations. *J. Dairy Sci.* 101:2148–2157. <https://doi.org/10.3168/jds.2017-13225>.
- Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4:44–57. <https://doi.org/10.1038/nprot.2008.211>.
- Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009b. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37:1–13. <https://doi.org/10.1093/nar/gkn923>.
- Interbull. 2016. Genetic evaluation. <http://interbull.org>.
- Jensen, R. G. 2002. The composition of bovine milk lipids: January 1995 to December 2000. *J. Dairy Sci.* 85:295–350. [https://doi.org/10.3168/jds.S0022-0302\(02\)74079-4](https://doi.org/10.3168/jds.S0022-0302(02)74079-4).
- Kanehisa, M. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28:27–30.
- Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima. 2017. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45(D1):D353–D361. <https://doi.org/10.1093/nar/gkw1092>.
- Kanehisa, M., Y. Sato, M. Furumichi, K. Morishima, and M. Tanabe. 2019. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47(D1):D590–D595. <https://doi.org/10.1093/nar/gky962>.
- Keller, H., C. Dreyer, J. Medin, A. Mahfoudi, K. Ozato, and W. Wahli. 1993. Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. *Proc. Natl. Acad. Sci. USA* 90:2160–2164. <https://doi.org/10.1073/pnas.90.6.2160>.
- Kęsek, M., T. Szulc, and A. Zielak-Steciwo. 2014. Genetic, physiological and nutritive factors affecting the fatty acid profile in cows' milk—A review. *Anim. Sci. Pap. Rep.* 32:95–105.
- Kim, B. S., J. S. Jung, J. H. Jang, K. S. Kang, and S. K. Kang. 2011. Nuclear Argonaute 2 regulates adipose tissue-derived stem cell survival through direct control of miR10b and selenoprotein N1 expression. *Aging Cell* 10:277–291. <https://doi.org/10.1111/j.1474-9726.2011.00670.x>.
- Kinsella, R. J., A. Kahari, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek. 2011. Ensembl BioMart: A hub for data retrieval across taxonomic space. *Database (Oxford)* 2011:bar030. <https://doi.org/10.1093/database/bar030>.
- Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124:979–993.
- Ligthart, S., C. Marzi, S. Aslibekyan, M. M. Mendelson, K. N. Conneely, T. Tanaka, E. Colicino, L. L. Waite, R. Joehanes, W. Guan, J. A. Brody, C. Elks, R. Marioni, M. A. Jhun, G. Agha, J. Bressler, C. K. Ward-Caviness, B. H. Chen, T. Huan, K. Bakulski, E. L. Salfati, G. Fiorito, S. Wahl, K. Schramm, J. Sha, D. G. Hernandez, A. C. Just, J. A. Smith, N. Sotoodehnia, L. C. Pilling, J. S. Pankow, P. S. Tsao, C. Liu, W. Zhao, S. Guarrera, V. J. Michopoulos, A. K. Smith, M. J. Peters, D. Melzer, P. Vokonas, M. Fornage, H. Prokisch, J. C. Bis, A. Y. Chu, C. Herder, H. Grallert, C. Yao, S. Shah, A. F. McRae, H. Lin, S. Horvath, D. Fallin, A. Hofman, N. J. Wareham, K. L. Wiggins, A. P. Feinberg, J. M. Starr, P. M. Visscher, J. M. Murabito, S. L. R. Kardia, D. M. Absher, E. B. Binder, A. B. Singleton, S. Bandinelli, A. Peters, M. Waldenberger, G. Matullo, J. D. Schwartz, E. W. Demerath, A. G. Uitterlinden, J. B. J. van Meurs, O. H. Franco, Y.-D. I. Chen, D. Levy, S. T. Turner, I. J. Deary, K. J. Ressler, J. Dupuis, L. Ferrucci, K. K. Ong, T. L. Assimes, E. Boerwinkle, W. Koenig, D. K. Arnett, A. A. Baccarelli, E. J. Benjamin, and A. Dehghan. 2016. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 17:255. <https://doi.org/10.1186/s13059-016-1119-5>.

- Lordan, R., and I. Zabetakis. 2017. Invited review: The anti-inflammatory properties of dairy lipids. *J. Dairy Sci.* 100:4197–4212. <https://doi.org/10.3168/jds.2016-12224>.
- Lourenco, D. A. L., B. O. Fragomeni, H. L. Bradford, I. R. Menezes, J. B. S. Ferraz, I. Aguilar, S. Tsuruta, and I. Misztal. 2017. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. *J. Anim. Breed. Genet.* 134:463–471. <https://doi.org/10.1111/jbg.12288>.
- MacGibbon, A. K. H., and M. W. Taylor. 2006. Composition and Structure of Bovine Milk Lipids. *Advanced Dairy Chemistry. Volume 2 Lipids*. 1–42. https://doi.org/10.1007/0-387-28813-9_1.
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Misztal, I., H. L. Bradford, D. A. L. Lourenco, S. Tsuruta, Y. Masuda, and A. Legarra. 2017. Studies on inflation of GEBV in single-step GBLUP for type. *Interbull Bull.* XX:38–42.
- Misztal, I., S. Tsuruta, D. A. L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. University of Georgia, Athens.
- Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. Lee. 2002. BLUPF90 and related programs (BGF90). Pages 21–22 in *Proc. 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France*. Editions Quae, Montpellier, France.
- Narayana, S. G., F. S. Schenkel, A. Fleming, A. Koeck, F. Malchiodi, J. Jamrozik, J. Johnston, M. Sargolzaei, and F. Miglior. 2017. Genetic analysis of groups of mid-infrared predicted fatty acids in milk. *J. Dairy Sci.* 100:4731–4744. <https://doi.org/10.3168/jds.2016-12244>.
- Ntambi, J. M. 1995. The regulation of stearoyl-CoA desaturase (SCD). *Prog. Lipid Res.* 34:139–150. [https://doi.org/10.1016/0163-7827\(94\)00010-J](https://doi.org/10.1016/0163-7827(94)00010-J).
- Oliveira, H. R., L. F. Brito, D. A. L. Lourenco, F. F. Silva, J. Jamrozik, L. R. Schaeffer, and F. S. Schenkel. 2019a. Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. *J. Dairy Sci.* 102:7664–7683. <https://doi.org/10.3168/jds.2019-16265>.
- Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, J. P. Cant, and F. S. Schenkel. 2019c. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. *J. Dairy Sci.* 102:9995–10011. <https://doi.org/10.3168/jds.2019-16821>.
- Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, and F. S. Schenkel. 2019b. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. *J. Dairy Sci.* 102:2365–2377. <https://doi.org/10.3168/jds.2018-15466>.
- Palmquist, D. L., A. D. Beaulieu, and D. M. Barbano. 1993. Feed and animal factors influencing milk fat composition. *J. Dairy Sci.* 76:1753–1771. [https://doi.org/10.3168/jds.S0022-0302\(93\)77508-6](https://doi.org/10.3168/jds.S0022-0302(93)77508-6).
- Palombo, V., M. Milanese, S. Sgorlon, S. Capomaccio, M. Mele, E. Nicolazzi, P. Ajmone-Marsan, F. Pilla, B. Stefanon, and M. D'Andrea. 2018. Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. *J. Dairy Sci.* 101:11004–11019. <https://doi.org/10.3168/jds.2018-14413>.
- Parodi, P. 2004. Milk fat in human nutrition. *Aust. J. Dairy Technol.* 59:3–59.
- Pepe, G., G. C. Tenore, R. Mastrocinque, P. Stusio, and P. Campiglia. 2013. Potential anticarcinogenic peptides from bovine milk. *J. Amino Acids* 2013:939804. <https://doi.org/10.1155/2013/939804>.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Raftery, A. L., and S. Lewis. 1992. Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Stat. Sci.* 7:493–497. <https://doi.org/10.1214/ss/1177011143>.
- Rutten, M. J. M., H. Bovenhuis, K. A. Hetingtinga, 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>.
- Samková, E., J. Špička, M. Pešek, T. Pelikánová, and O. Hanuš. 2012. Review: Animal factors affecting fatty acid composition of cow milk fat: A review. *S. Afr. J. Anim. Sci.* 42:83–100.
- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478.
- Sauer, F. D., V. Fellner, R. Kinsman, J. K. G. Kramer, H. A. Jackson, A. J. Lee, and S. Chen. 1998. Methane output and lactation response in Holstein cattle with monensin or unsaturated fat added to the diet. *J. Anim. Sci.* 76:906–914.
- Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and J. Van Doormaal. 2000. Experience with a test-day model. *J. Dairy Sci.* 83:1135–1144. [https://doi.org/10.3168/jds.S0022-0302\(00\)74979-4](https://doi.org/10.3168/jds.S0022-0302(00)74979-4).
- Smith, B. J. 2007. boa: An R Package for MCMC output convergence assessment and posterior inference. *J. Stat. Softw.* 21:1–37.
- Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. Estimating fatty acid content in cow milk using mid-infrared spectrometry. *J. Dairy Sci.* 89:3690–3695. [https://doi.org/10.3168/jds.S0022-0302\(06\)72409-2](https://doi.org/10.3168/jds.S0022-0302(06)72409-2).
- Stryer, L. 1995. Fatty acid metabolism. Pages 603–628 in *Biochemistry*. 4th ed. W. H. Freeman and Company, New York, NY.
- Temme, E. H., R. P. Mensink, and G. Hornstra. 1996. Comparison of the effects of diets enriched in lauric, palmitic, or oleic acids on serum lipids and lipoproteins in healthy women and men. *Am. J. Clin. Nutr.* 63:897–903. <https://doi.org/10.1093/ajcn/63.6.897>.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- Viitala, S. M., N. F. Schulman, D. J. de Koning, K. Elo, R. Kinoshita, A. Virta, J. Virta, A. Mäki-Tanila, and J. H. Vilkkii. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *J. Dairy Sci.* 86:1828–1836. [https://doi.org/10.3168/jds.S0022-0302\(03\)73769-2](https://doi.org/10.3168/jds.S0022-0302(03)73769-2).
- Wang, H., L. Jiang, X. Liu, J. Yang, J. Wei, J. Xu, Q. Zhang, and J.-F. Liu. 2013. A post-GWAS replication study confirming the PTK2 gene associated with milk production traits in Chinese Holstein. *PLoS One* 8:e83625. <https://doi.org/10.1371/journal.pone.0083625>.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 20:134.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb.)* 94:73–83. <https://doi.org/10.1017/S0016672312000274>.
- Wang, X., P. Ma, J. Liu, Q. Zhang, Y. Zhang, X. Ding, L. Jiang, Y. Wang, Y. Zhang, D. Sun, S. Zhang, G. Su, and Y. Yu. 2015. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. *BMC Genet.* 16:111. <https://doi.org/10.1186/s12863-015-0263-3>.
- Williams, C. M. 2000. Dietary fatty acids and human health. *Ann. Zootech.* 49:165–180. <https://doi.org/10.1051/animres:2000116>.
- Zhao, A., F. Jiang, G. Yang, H. Liu, and L. Li. 2019. Sfrp5 interacts with Slurp1 to regulate the accumulation of triglycerides in hepatocyte steatosis model. *Biochem. Biophys. Res. Commun.* 512:256–262. <https://doi.org/10.1016/j.bbrc.2019.03.035>.
- Zong, G., Y. Li, A. J. Wanders, M. Alssema, P. L. Zock, W. C. Willett, F. B. Hu, and Q. Sun. 2016. Intake of individual saturated fatty acids and risk of coronary heart disease in US men and women: Two prospective longitudinal cohort studies. *BMJ* 355:i5796. <https://doi.org/10.1136/bmj.i5796>.