



OPEN Genomic prediction and genetic parameter estimation for unsaturated and saturated fatty acids in Canadian dairy cattle

S. O. Peters^{1,5}✉, K. Kizilkaya², E. M. Ibeagha-Awemu³ & X. Zhao⁴✉

The current study aimed to obtain the estimates of heritabilities and genetic correlations and the prediction abilities and accuracy of Bayesian GBLUP and Bayesian alphabet (BayesA, BayesB, BayesC) models for total and individual monounsaturated, polyunsaturated and saturated fatty acids from Canadian Holstein cows by using genome-wide SNP markers from genotyping-by-sequencing method. The heritability estimates were obtained from Bayesian GBLUP and Bayesian alphabet models. They ranged from 0.61 to 0.67 for total monounsaturated, from 0.35 to 0.45 for polyunsaturated and from 0.51 to 0.60 for saturated fatty acids, respectively. For thirty-three individual monounsaturated, polyunsaturated and saturated fatty acids, the heritability estimates ranged from 0.27 to 0.69 for individual monounsaturated, from 0.27 to 0.68 for individual polyunsaturated and from 0.35 to 0.69 for individual saturated fatty acids. These results indicated that total and individual monounsaturated, polyunsaturated and saturated fatty acids were under moderate genetic control and can be improved through genomic selection. The estimates of genetic correlations among total and individual monounsaturated, polyunsaturated and saturated fatty acids showed a moderate to high genetic relationships and pointed out the need for consideration of genetic relationships in successful genomic selection for fatty acids traits. The accuracies of BayesC and BayesA models were similar and better than that of GBLUP and BayesB models which indicated that fatty acids were determined by many genes having non-null effects, which are assumed to follow a univariate or multivariate *Student's t* distribution.

Keywords Fatty acid traits, Bayesian GBLUP, BayesA, BayesB, BayesC, Genomic prediction, Random cluster, Genotyping-by-sequencing

Bovine milk is an important source of nutrients in human nutrition and fat is one of the main components in bovine milk. Milk fat contains a large number of individual fatty acids (FA) which can be identified based on their number of carbons, the saturation of their carbon chain, and the conformation of double bonds. Milk fat consists of around 70% of saturated FA (SFA), 25% monosaturated FA (MUFA), and 5% polyunsaturated FA (PUFA)¹. Milk FA composition is one of the most important aspects of milk nutritional quality and has gained the interest of manufacturers and consumers as it influences nutritional, physical and flavor properties of dairy products². The increasing consumer interest on the nutritional quality of dairy product pushes toward the inclusion of fine milk composition traits among breeding goals of dairy cattle. In particular, milk PUFA like isomers of C18:2n9c11t and C18:2n10t12c (Conjugated linoleic acid), C20:4n6 (Arachidonic acid), C20:5n3 (Eicosapentanoic acid) and C22:5n3 (Docosapentanoic acid) have known positive associations with a range of human health conditions like cardiovascular diseases, anticancer effects, antiadipogenic, antiatherogenic, antidiabetogenic and anti-inflammatory^{3,4}. For these reasons, dairy producers are looking for ways to optimize milk beneficial components. Genetic variability of FA indicates the possibility of using genomic selection to improve milk traits⁵.

¹Department of Animal Science, Berry College, Mount Berry, GA 30149, USA. ²Department of Animal Science, Faculty of Agriculture, Aydin Adnan Menderes University, 09100 Aydin, Turkey. ³Sherbrooke Research and Development Centre, Agriculture and Agri-Food Canada, 2000 Rue College, Sherbrooke, QC J1M 0C8, Canada. ⁴Department of Animal Science, McGill University, 21,111 Lakeshore Road, Ste-Anne-De-Bellevue, Montreal, QC H9S 3V9, Canada. ⁵Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA. ✉email: speters@berry.edu; xin.zhao@mcgill.ca

Studies in dairy production system were carried out looking for efficient strategies to optimize milk beneficial components such as milk fatty acids. They indicated that the milk FA composition can be modified by altering dietary or feeding supplementation in the dairy production system^{6–8}; however, genetic improvement for milk fatty acids through genomic selection can provide better permanent alterations compared with the dietary changes in the milk fatty acid profile. The studies on FA indicated that FA composition in bovine milk could vary among dairy breeds, herds and cows^{9–11} and be affected by environmental factors, especially dairy cows' ration^{12–14} and genetic factors^{6,10,15}. The genetic variability for FA is an important characteristic to modify the FA composition in bovine milk through breeding methods. Heritability is a well-known genetic parameter indicating the proportion of the genetic variation to the total variation. Genomic selection programs require to obtain the estimates of genetic parameters (heritability and genetic correlation) of traits and the accuracy of breeding values of animals from the models used in genomic selection.

Therefore, recent studies have focused on estimating genetic parameters of major milk FA by using pedigree-based BLUP models for Korean Holstein cattle¹⁶, for New Zealand dairy cattle¹⁷, for Italian Brown Swiss cows¹⁸, genome-based BLUP model for Danish Holstein cattle¹⁹ and for North American Holstein cattle²⁰ and have indicated the existence of genetic variability among cows with respect to milk fatty acids.

The accuracy of breeding values is key to the successful application of genomic selection in animal breeding. The aims of this study were (1) to investigate the genetic relationship between FA within MUFA, PUFA and SFA groups by estimating of genetic correlations from the Bayesian bivariate Genomic Best Linear Unbiased Prediction (GBLUP) and (2) to compare the predictive ability and accuracy of the Bayesian univariate GBLUP and Bayesian alphabet models by estimating heritability for milk FA within MUFA, PUFA and SFA groups from Canadian Holstein cows randomly allocated to the training and cross-validation datasets. Results from this endeavor will positively impact the genomic selection approach in livestock and improve accuracies for selecting animals that will produce more nutritious milk for human consumption.

Results

Descriptive statistics of fatty acids

Descriptive statistics of the total and individual FA traits in MUFA, PUFA and SFA (% total FA) are given in (Table 1). Total SFA made up 71.20% (on average 68.568 with the range of 56.215 to 77.283% total FA) of total FA, while total MUFA and PUFA represented 25.30% (on average 24.513 with the range of 16.894 to 33.740% total FA) and 3.50% (on average 3.370 with the range of 2.002 to 5.375% total FA) of total FA in Canadian Holstein cows, respectively.

As seen in (Table 1), the most abundant individual FA (on average) in SFA were C16:0 (34.589), C14:0 (12.643) and C18:0 (9.371% total FA), respectively and they consisted of 36.07% (50.67%), 13.17% (18.49%) and 9.70% (13.63%) of total FA (total SFA). In addition, the means of C12:0, C10:0, C15:0 and C6:0 were 3.599, 2.703, 1.147 and 1.008% total FA, and they represented 3.75% (5.27%), 2.81% (3.95%), 1.20% (1.68%) and 1.05% (1.48%) of the total FA (total SFA). However, C4:0, C8:0, C11:0, C13:0, C17:0, C20:0, C22:0, C23:0 and C24:0 in SFA with the mean of lower than 1.000 g/100 g total FA represented relatively low proportion (< 1% each) of the total FA.

In MUFA, Oleic acid with mean of 19.787% total FA had 20.55% of the total FA which was the most prominent in MUFA with 81.19% of the total MUFA. In addition, C16:1t, C14:1t and C18:1n11t with the means of 1.807, 1.058 and 0.987% total FA made up 1.87% (7.39%), 1.08% (4.28%) and 1.01% (4.00%) of the total FA (total MUFA). However, C14:1, C16:1 and C18:1n9t in MUFA represented less than 1% of the total FA.

In PUFA, C18:2n6cc with the mean of 1.831% total FA was the most abundant individual PUFA which made up 1.91% of the total FA and 54.65% of the total PUFA. However, C18:2n6tt, C18:2n9c11t, C18:2n10t12c, C18:3n3, C18:3tcc, C20:3n6, C20:4n6, C20:5n3 and C22:5n3 in PUFA with the mean of lower than 1.000% total FA represented relatively low proportion (< 1% each) of the total FA.

Estimates of phenotypic and genetic correlations

A total of 630 phenotypic and genetic correlations with standard errors between the total and individual MUFA, PUFA and SFA for Canadian Holstein cows were presented in Table S1 in the supplemental file and those for FA within MUFA, PUFA and SFA groups presented by using blue and red colors indicating correlations from 1 to -1 were given in Figs. 1, 2 and 3, respectively.

As seen in Fig. 1, most of the phenotypic and genetic correlations for MUFA were positive. The very strong positive phenotypic (0.97) and genetic (0.97) correlations were obtained between total MUFA and Oleic acid which was expressed in 81.19% in total MUFA. There were moderate positive phenotypic (0.52) and genetic (0.48) correlations between total MUFA and C16:1t within MUFA. For individual MUFA, the phenotypic and genetic correlations were moderate positive between C14:1 and C16:1 (0.46 and 0.41, respectively), between C16:1t and C18:1n9c (0.46 and 0.41, respectively), C14:1t and C18:1n9t (0.43 and 0.46, respectively) and between C18:1n9t and C18:1n11t (0.59 and 0.56, respectively). However, moderate negative genetic correlations were observed between C16:1 and C18:1n9t (-0.42) and C18:1n11t (-0.43) acids and low negative phenotypic and genetic correlations between C14:1 and C18:1n9t (-0.14, -0.20, respectively), C18:1n11t (-0.25, -0.34, respectively), and between C16:1 and C14:1t (-0.27, -0.26, respectively).

For PUFA, as seen in Fig. 2, the phenotypic and genetic correlations between total and individual PUFA were mostly positive. Total PUFA resulted in positive phenotypic and genetic correlations ranging from 0.12 to 0.92 with individual FA in PUFA.

The very strong positive phenotypic (0.91) and genetic (0.92) correlations were obtained between total PUFA and C18:2n6cc which was represented with high percentage (54.65%) in total PUFA. There were moderate positive correlations (0.40 and 0.70, respectively) between total PUFA and C18:3n3, C22:5n3, C20:3n6 and C18:3tcc within PUFA. For individual PUFA, the phenotypic and genetic correlations were moderate positive between C18:2n6cc and C18:3n3 (0.46, 0.45, respectively), C20:3n6 (0.40, 0.46, respectively), C20:4n6 (0.47, 0.41,

Monounsaturated fatty acids (MUFA)	Min	Max	Mean	Sd	% in Total	% in MUFA
C14:1 (myristoleic acid)	0.121	0.678	0.311	0.125	0.33	1.29
C14:1t (myristelaidic acid)	0.271	1.825	1.058	0.281	1.08	4.28
C16:1 (palmitoleic acid)	0.198	0.468	0.304	0.051	0.32	1.25
C16:1t (palmitelaidic acid)	0.784	2.961	1.807	0.365	1.87	7.39
C18:1n9c (oleic acid)	11.014	29.833	19.787	3.000	20.55	81.19
C18:1n9t (elaidic acid)	0.014	0.473	0.234	0.103	0.24	0.95
C18:1n11t (trans vaccenic acid)	0.150	1.929	0.987	0.369	1.01	4.00
MUFA (total)	16.894	33.740	24.513	3.176	25.30	100.00
Polyunsaturated fatty acids (PUFA)	Min	Max	Mean	Sd	% in Total	% in PUFA
C18:2n6cc (linoleic acid)	0.974	2.988	1.831	0.414	1.91	54.65
C18:2n6tt (trans-linoleic acid)	0.061	0.316	0.176	0.049	0.18	5.18
C18:2n9c11t (cis-9 trans-11 CLA)	0.118	0.764	0.396	0.110	0.41	11.73
C18:2n10t12c (trans-10 cis-12 CLA)	0.001	0.032	0.016	0.006	0.02	0.48
C18:3n3 (alpha linolenic acid)	0.035	0.883	0.416	0.148	0.43	12.40
C18:3tcc (gamma linolenic acid)	0.049	0.172	0.111	0.022	0.12	3.30
C20:3n6 (dihomogamma linolenic acid)	0.023	0.176	0.095	0.027	0.10	2.84
C20:4n6 (arachidonic acid)	0.063	0.222	0.124	0.028	0.13	3.58
C20:5n3 (eicosapentanoic acid)	0.001	0.071	0.031	0.012	0.03	0.92
C22:5n3 (docosapentaenoic acid)	0.007	1.108	0.172	0.175	0.18	5.16
PUFA (total)	2.002	5.375	3.370	0.646	3.50	100.00
Saturated fatty acid (SFA)	Min	Max	Mean	Sd	% in Total	% in SFA
C4:0 (butyric acid)	0.375	1.238	0.796	0.130	0.83	1.16
C6:0 (caproic acid)	0.480	1.450	1.008	0.143	1.05	1.48
C8:0 (caprylic acid)	0.478	1.214	0.904	0.125	0.94	1.32
C10:0 (capric acid)	1.331	3.892	2.703	0.436	2.81	3.95
C11:0 (undecanoic acid)	0.103	0.366	0.239	0.045	0.25	0.35
C12:0 (lauric acid)	1.587	5.506	3.599	0.608	3.75	5.27
C13:0 (tridecylic acid)	0.201	0.727	0.448	0.077	0.47	0.66
C14:0 (myristic acid)	7.513	16.048	12.643	1.222	13.17	18.49
C15:0 (pentadecylic acid)	0.395	2.163	1.147	0.244	1.20	1.68
C16:0 (palmitic acid)	24.923	42.896	34.589	3.693	36.07	50.67
C17:0 (margaric acid)	0.498	0.873	0.649	0.067	0.68	0.95
C18:0 (stearic acid)	5.258	15.290	9.371	1.880	9.70	13.63
C20:0 (arachidic acid)	0.010	0.225	0.118	0.033	0.12	0.17
C22:0 (behenic acid)	0.021	0.088	0.054	0.013	0.06	0.08
C23:0 (tricosanoic acid)	0.001	0.043	0.021	0.010	0.02	0.03
C24:0 (lignoceric acid)	0.001	0.082	0.037	0.015	0.04	0.05
SFA (total)	56.215	77.283	68.568	3.731	71.20	100.00

Table 1. Descriptive statistics of minimum (Min), maximum (Max), mean, standard deviation (Sd) and percentages of fatty acids (% total fatty acids).

respectively), C22:5n3 (0.38, 0.46, respectively); between C18:2n9c11t and C18:3tcc (0.54, 0.55, respectively), between C18:2n10t12c and C20:5n3 (0.38, 0.42, respectively); between C18:3n3 and C20:5n3 (0.38, 0.64, respectively) and C22:5n3 (0.45, 0.45, respectively); between C20:3n6 and C20:4n6 (0.55, 0.62, respectively).

For SFA, there were negative correlations between total SFA and individual FA in SFA and between individual FA in SFA. The strong positive phenotypic (0.75) and genetic (0.72) correlations were obtained between total SFA and C16:0 which was represented with high percentage (50.67%) in total SFA. There were moderate positive correlations (0.40 and 0.70, respectively) between total SFA and C6:0 (0.39, 0.43, respectively), C8:0 (0.52, 0.54, respectively), C10:0 (0.57, 0.66, respectively), C12:0 (0.53, 0.60, respectively) and C14:0 (0.51, 0.56, respectively); however, negative low correlations with C17:0 (−0.01, −0.16, respectively), C18:0 (−0.25, −0.22, respectively) and C20:0 (−0.11, −0.05, respectively) within SFA.

For individual FA in SFA, in contrast to the phenotypic and genetic correlations between individual FA in MUFA and PUFA, strong/moderate positive phenotypic and genetic correlations were obtained for seven even-numbered SFA, 0.93 and 0.92 for C10:0 and C12:0, respectively, 0.88 and 0.90 for C6:0 and C8:0, respectively, 0.87 and 0.89 for C8:0 and C10:0, respectively, 0.66 and 0.83 for C20:0 and C22:0, respectively, 0.77 and 0.68 for C12:0 and C14:0 (Myristic), respectively, 0.66 and 0.70 for C8:0 and C12:0, respectively, 0.69 and 0.69 for C4:0 and C6:0, respectively and for two odd-numbered SFA, 0.63 and 0.59 for C15:0 and C17:0, respectively, and 0.64 and

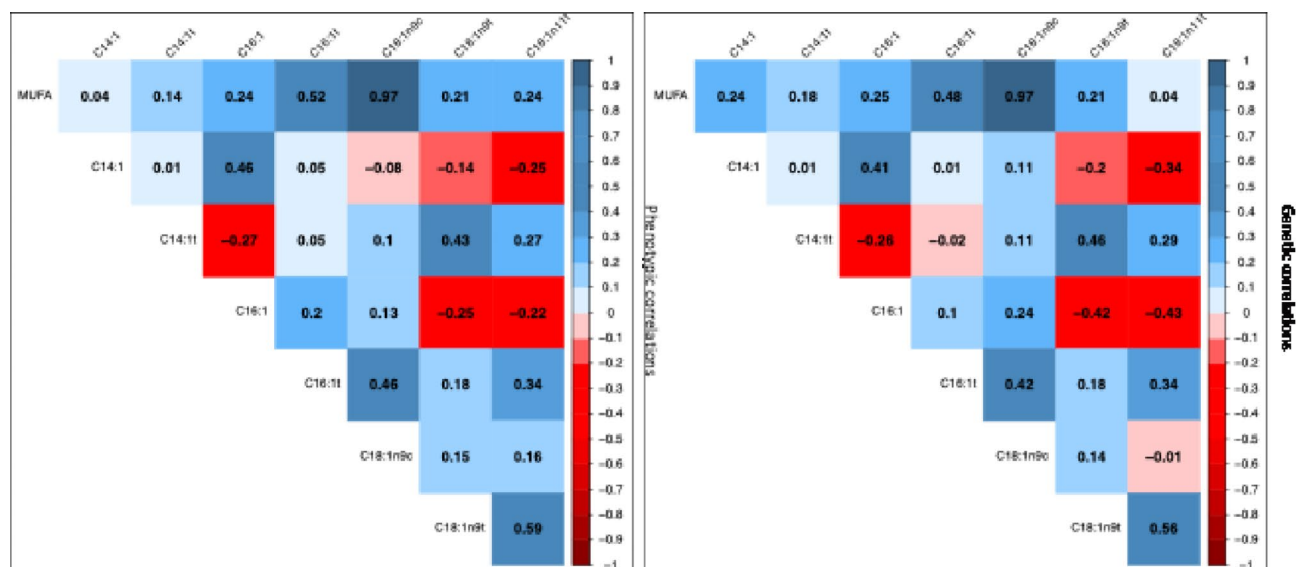


Fig. 1. Phenotypic and genetic correlations between monounsaturated fatty acids (MUFA).

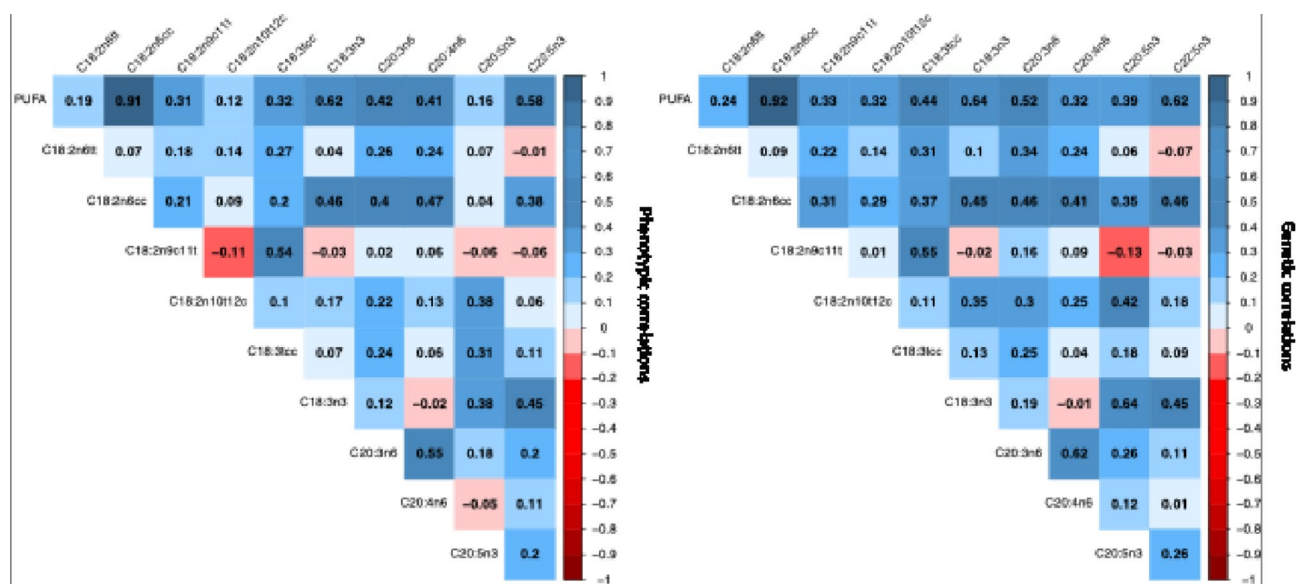


Fig. 2. Phenotypic and genetic correlations between polyunsaturated fatty acids (PUFA).

0.55 for C13:0 and C15:0 acids, respectively. However, moderate negative phenotypic and genetic correlations were obtained between C4:0 and C13:0 (-0.42, -0.29, respectively) and C15:0 (-0.43, -0.40, respectively) acids, between C18:0 and C11:0 (-0.48, -0.54, respectively), and C15:0 (-0.45, -0.39, respectively), and C14:0 (-0.35, -0.40, respectively), and C16:0 (-0.59, -0.58, respectively) acids, C20:0 and C11:0 (-0.32, -0.48, respectively) acids. Phenotypic and genetic correlations between other individual FA in SFA were found generally low positive or negative.

The fatty acids in dairy milk are generally classified based on their chain lengths between 4 and 24 carbons. The even-numbered FA being 4–16 carbons in chain length arise from de novo synthesis in the mammary gland and the other FA, which include approximately half of the 16 carbon and all those 18 carbons or greater in length, are taken up preformed from the blood²¹. The phenotypic and genetic correlations between MUFA, PUFA and SFA given in Table S1 in the supplementary file could be evaluated by similarities in the origin of FA. Positive phenotypic and genetic correlations were generally obtained between C6:0, C8:0, C10:0, C12:0, C14:0 and C16:0 acids in SFA which were synthesized de novo in the mammary gland and between C18:1n9c, C18:1n9t and C18:1n11t in MUFA and between C18:2n6cc, C18:2n6t, C18:2n10t12c, C18:3n3 and C18:3tcc acids in PUFA which were extracted from blood. However, as seen in Table S1 in the supplementary file, negative moderate/low phenotypic and genetic correlations were obtained for C6:0, C8:0, C10:0, C11:0, C12:0, C13:0, C14:0, C15:0 and

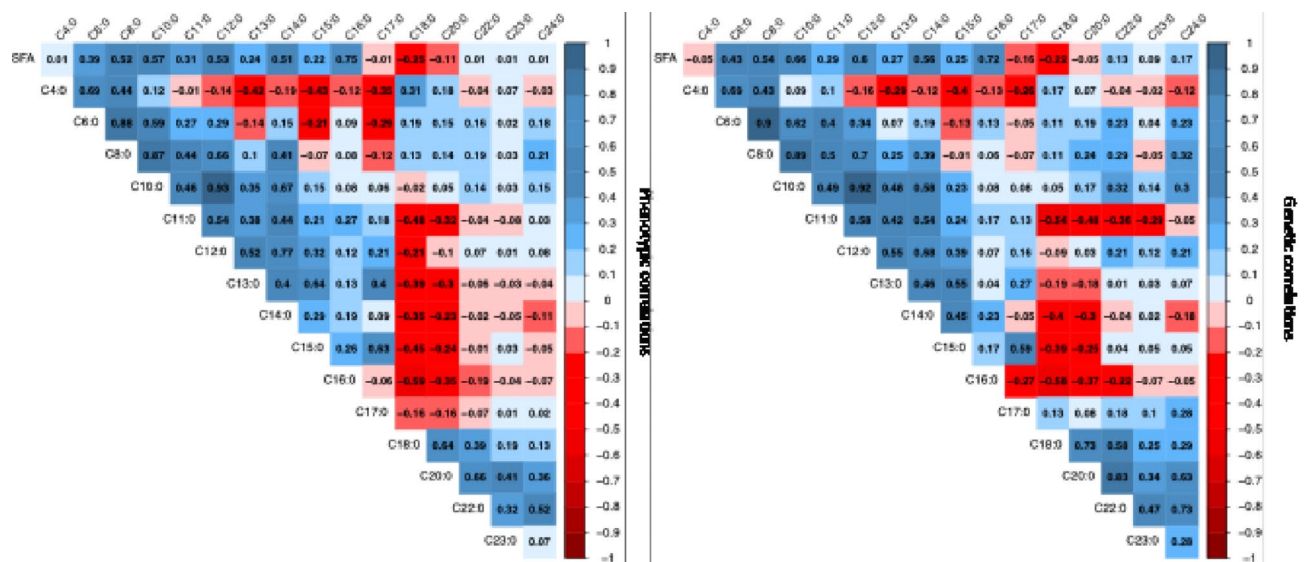


Fig. 3. Phenotypic and genetic correlations between saturated fatty acids (SFA).

C16:0 acids in SFA (de novo synthesis in the mammary gland) with C18:1n9t, C18:1n9c and C18:1n11t acids in MUFA and with C18:2n6cc, C18:2n6t and C18:3n3 acids in PUFA (the mammary uptake of preformed FA).

The estimates of genetic correlations (Table S1 in the supplementary file) indicated the existence of the strong positive and negative genetic relationships between total and individual FA within MUFA, PUFA and SFA groups in bovine milk and the importance of using genetic relationships in the selection programs for fatty acids to improve the nutritional quality of bovine milk and milk products.

Heritability estimates from Bayesian GBLUP and Bayesian alphabet models in random training data sets

Heritability estimates of total and individual FA within MUFA, PUFA and SFA groups from 10-fold random cluster training data sets were obtained from Bayesian GBLUP and Bayesian alphabet models and presented in Fig. 4. As seen in Fig. 4, heritability estimates from Bayesian GBLUP model were higher than those from Bayesian alphabet models for total and individual FA within MUFA, PUFA and SFA groups. In Bayesian alphabet models, BayesB model produced generally higher heritability estimates than BayesA and BayesC models and BayesC model resulted in the lowest heritability estimates for total and individual FA within MUFA, PUFA and SFA groups.

For total MUFA, PUFA and SFA, as seen in Fig. 4A, heritability estimates with standard error of 0.01 indicated that Bayesian GBLUP (0.67, 0.49, 0.60, respectively) yielded higher heritability estimates than BayesA (0.63, 0.41, 0.55, respectively), BayesB (0.63, 0.45, 0.56, respectively) and BayesC (0.61, 0.35, 0.51, respectively) models and heritability estimates from BayesA and BayesB models for total MUFA, PUFA and SFA were found similar and higher than those from BayesC model.

For individual FA in MUFA, as seen in Fig. 4B, Bayesian GBLUP model produced higher heritability estimates than Bayesian alphabet models. The lowest heritability estimates were obtained for C14:1 ranged from 0.27 of BayesC to 0.45 of GBLUP and the highest heritability estimates for C14:1t ranged from 0.63 of BayesC to 0.69 of GBLUP. Heritability estimates from Bayesian GBLUP, BayesA and BayesB models were generally equal to or higher than 0.50 for C14:1t, C16:1, C16:1t, C18:1n9c, C18:1n9t and C18:1n11t.

For individual FA in PUFA, as seen in Fig. 4C, the lowest heritability estimates were obtained for C18:3tcc ranged from 0.27 of BayesC to 0.45 of GBLUP and C18:2n6t ranged from 0.27 of BayesC to 0.44 of GBLUP, and the highest heritability estimates for C18:2n6cc ranged from 0.63 of BayesA and BayesC to 0.68 of GBLUP. Although Bayesian GBLUP model yielded higher heritability estimates for individual PUFA, Bayesian GBLUP and BayesB models resulted in similar heritability estimates for C20:4n6 (0.57 and 0.56) and C22:5n3 (0.47 and 0.46) acids. Heritability estimates from Bayesian GBLUP and Bayesian alphabet models were higher than 0.50 for C18:2n6cc, C18:2n10t12c, C18:3n3 and C20:3n6.

For individual FA in SFA, as seen in Fig. 4D, Bayesian GBLUP model resulted in considerably higher heritability estimates than Bayesian alphabet models. Heritability estimates for individual SFA ranged from 0.35 of BayesC to 0.69 of GBLUP. The lowest heritability estimates ranged from 0.35 to 0.49 were obtained from Bayesian GBLUP and Bayesian alphabet models for C24:0 and the highest heritability estimates ranged from 0.64 to 0.69 for C18:0. Heritability estimates from Bayesian GBLUP and Bayesian alphabet models were higher than 0.50 for C4:0, C6:0, C8:0, C10:0, C11:0, C12:0, C13:0, C14:0, C15:0, C16:0, C17:0, and C18:0.

Predictive ability and accuracy of Bayesian GBLUP and Bayesian alphabet models

Predictive abilities of Bayesian GBLUP and Bayesian alphabet models were explored by using the Pearson's correlation coefficient ($r_{y, GEBV}$) between the observed phenotypic values (y) and GEBV from 10-fold random

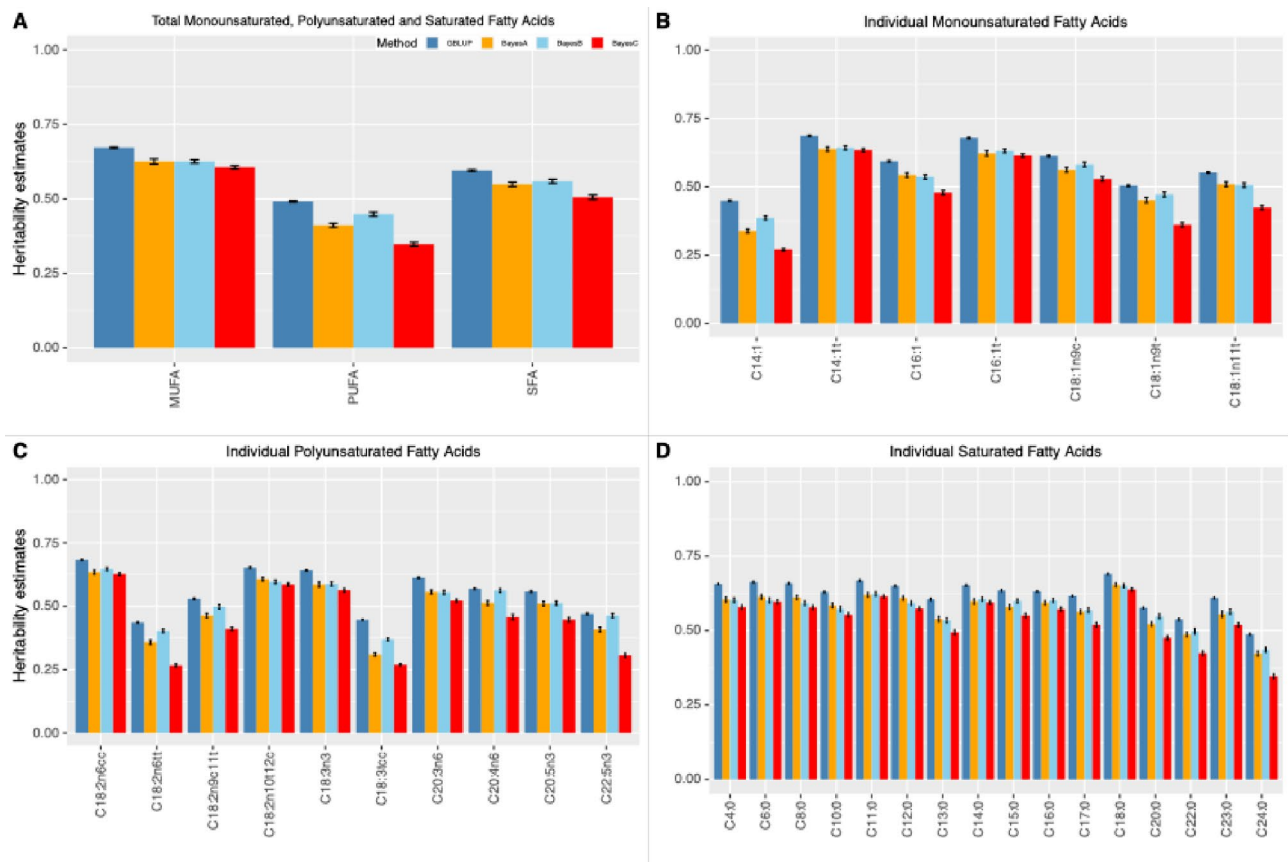


Fig. 4. Heritability estimates for total and individual MUFA, PUFA and SFA from 10-fold cross-validation data sets using Bayesian GBLUP, and Bayesian alphabet (BayesA, BayesB, BayesC) models.

cluster cross-validation data sets and mean values and standard errors of correlations for predictive ability of models were given in Fig. 5A for total MUFA, PUFA and SFA, in Fig. 5B–D for individual MUFA, PUFA and SFA, respectively. The mean values and standard errors of correlations from 10-fold cross-validation data sets for total MUFA, PUFA and SFA in Fig. 5A indicated that the predictive abilities of Bayesian GBLUP and Bayesian alphabet models ranged from 0.46 ± 0.01 to 0.47 ± 0.01 for total MUFA, from 0.37 ± 0.01 to 0.40 ± 0.01 for PUFA and from 0.43 ± 0.01 to 0.45 ± 0.01 for SFA. As seen in Fig. 5A and from the range of predictive abilities of models, Bayesian GBLUP and BayesA models resulted in similar predictive ability for total MUFA, PUFA and SFA and BayesB model yielded the lowest predictive ability for total MUFA, PUFA and SFA.

The predictive ability of Bayesian GBLUP and Bayesian alphabet models for individual FA within MUFA, PUFA and SFA groups indicated that Bayesian GBLUP and BayesA models resulted in similar predictive ability and BayesB model generally yielded the lowest predictive ability in each individual FA within MUFA, PUFA and SFA groups.

For individual FA in MUFA, as seen in Fig. 5B, the lowest predictive ability from Bayesian GBLUP and Bayesian alphabet models was obtained for C14:1 ranging from 0.36 ± 0.02 to 0.38 ± 0.02 and the highest predictive ability for C14:1t ranging from 0.47 ± 0.01 to 0.49 ± 0.01 . Bayesian GBLUP and Bayesian alphabet models also showed similar predictive ability (0.42 ± 0.01 to 0.47 ± 0.01) for C18:1n9c, C16:1, C16:1t and C18:1n11t acids. For individual FA in PUFA, as seen in Fig. 5C, the lowest predictive ability from Bayesian GBLUP and Bayesian alphabet models was obtained for C18:3tcc ranging from 0.25 ± 0.01 to 0.27 ± 0.01 and C22:5n3 ranging from 0.29 ± 0.02 to 0.31 ± 0.02 . Besides, the highest predictive ability from Bayesian GBLUP and Bayesian alphabet models was obtained for C18:2n6cc, C18:3n3 and C20:3n6 ranging from 0.43 ± 0.01 to 0.47 ± 0.01 . For individual FA in SFA, as seen in Fig. 5D, Bayesian GBLUP and Bayesian alphabet models resulted in similar predictive ability ranging from 0.38 ± 0.02 to 0.48 ± 0.01 . The lowest predictive ability from Bayesian GBLUP and Bayesian alphabet models was obtained for C24:0 ranging from 0.38 ± 0.02 to 0.41 ± 0.02 and the highest predictive ability for C11:0 ranging from 0.46 ± 0.01 to 0.48 ± 0.01 . Fatty acids in total and individual MUFA, PUFA and SFA having higher heritability resulted in higher predictive ability regardless of Bayesian GBLUP and Bayesian alphabet models.

After removing the effect of heritability on the predictive ability of Bayesian GBLUP and Bayesian alphabet models (Eq. 9), mean values and standard errors of accuracies of Bayesian GBLUP and Bayesian alphabet models for total and individual MUFA, PUFA and SFA were calculated and presented in Fig. 6 for total (Fig. 6A) and individual MUFA, PUFA and SFA (Fig. 6B–D). As seen in Fig. 6, BayesC and BayesA models resulted in slightly higher accuracy than Bayesian GBLUP and BayesB models. As seen in Fig. 6A, the accuracy of Bayesian GBLUP

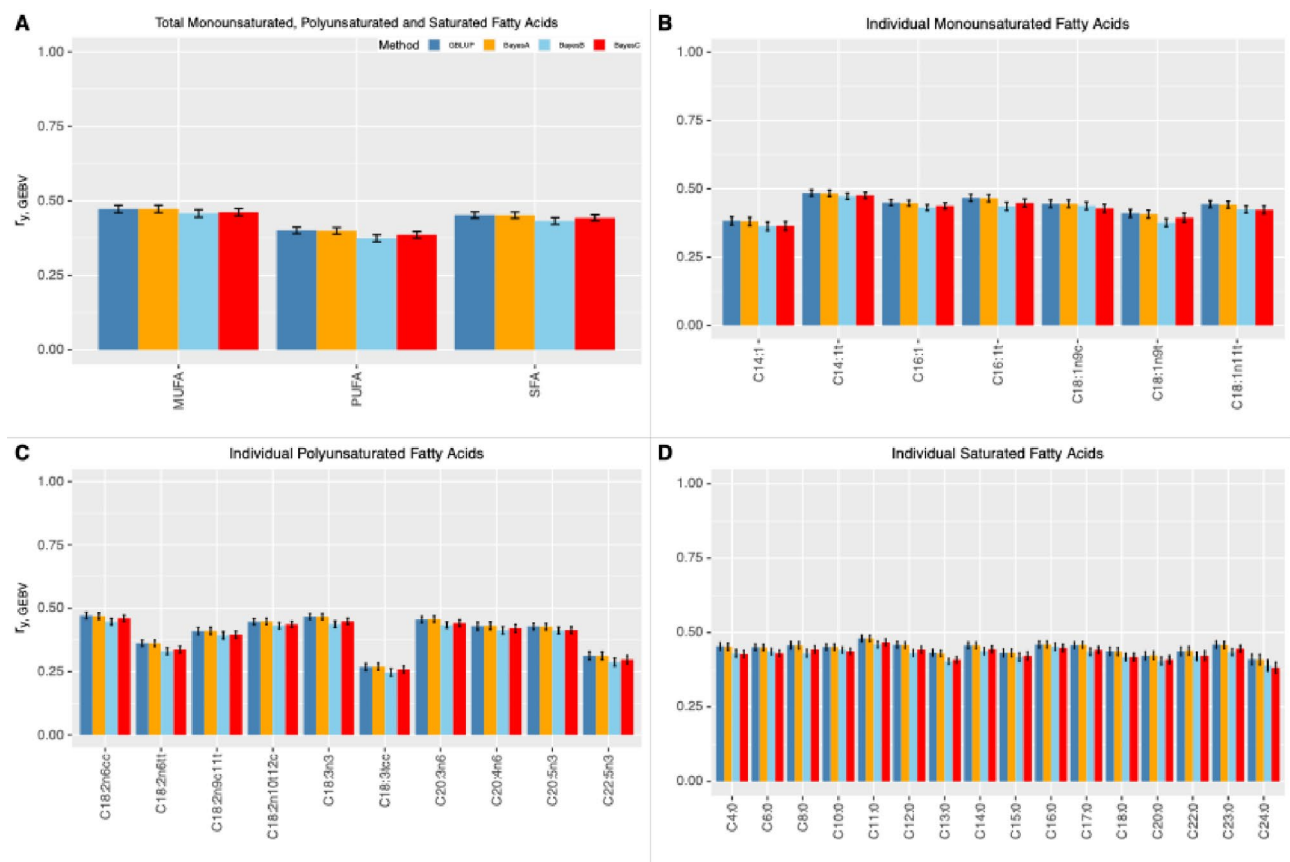


Fig. 5. Mean of Pearson's correlation coefficient ($r_{y, GEBV}$) between the observed phenotypic values (y) and GEBV for individual and total MUFA, PUFA and SFA from 10-fold cross-validation data sets using Bayesian GBLUP, and Bayesian alphabet (BayesA, BayesB, BayesC) models.

and Bayesian alphabet models for total MUFA, PUFA and SFA ranged from 0.56 ± 0.02 to 0.65 ± 0.02 and BayesC and BayesA models resulted in higher accuracy than Bayesian GBLUP and BayesB models. Accuracies in Fig. 6B, C indicated that accuracies ranged from 0.55 ± 0.02 to 0.70 ± 0.03 for individual MUFA and from 0.41 ± 0.02 to 0.65 ± 0.03 for individual PUFA. Figure 6B, C also showed that BayesC model generally produced higher accuracy than Bayesian GBLUP, BayesA and BayesB models for individual MUFA and PUFA. Higher accuracies from BayesC model were obtained for C14:1 (0.70 ± 0.03), C16:1 (0.63 ± 0.02), C18:1n9t (0.66 ± 0.03) and C18:1n11t (0.65 ± 0.02) in individual MUFA and for C18:2n6tt (0.65 ± 0.03), C18:2n9c11t (0.62 ± 0.02), C20:4n6 (0.62 ± 0.02), C20:5n3 (0.62 ± 0.02), C22:5n3 (0.54 ± 0.03) and C18:3tcc (0.50 ± 0.03) in individual PUFA. As seen in Fig. 6D, the accuracies from Bayesian GBLUP and Bayesian alphabet models ranged from 0.52 ± 0.02 to 0.65 ± 0.03 and BayesA model resulted in higher accuracies for individual SFA. The lowest accuracy was obtained for C18:0 ranging from 0.52 ± 0.02 to 0.54 ± 0.02 and the highest accuracies for C24:0 and C22:0 ranging from 0.60 ± 0.02 to 0.65 ± 0.03 . Accuracies for total and individual MUFA, PUFA and SFA in Fig. 6 indicated that Bayesian alphabet models yielded more reliable prediction than Bayesian GBLUP model for FA traits.

In addition to accuracy of Bayesian GBLUP and Bayesian alphabet models for reliable prediction, MSE was calculated to determine the better fit of Bayesian GBLUP and Bayesian alphabet models for total and individual MUFA, PUFA and SFA and presented in Table 2. As seen in Table 2, although Bayesian GBLUP and Bayesian alphabet models yielded same or similar MSE for C14:1 (0.11), C16:1 (0.09 and 0.10, respectively) and C18:1n9t (0.06) in MUFA, C18:1n11t (0.03), C18:2n9c11t (0.16 and 0.17, respectively), C18:2n10t12c (0.0002 and 0.0003, respectively), C18:3n3 (0.18 and 0.19, respectively), C18:3tcc (0.01), C20:3n6 (0.01), C20:4n6 (0.02), C20:5n3 (0.0011) and C22:5n3 (0.06 and 0.07, respectively) acids in PUFA, C11:0 (0.05 and 0.06, respectively), C13:0 (0.20 and 0.21, respectively), C20:0 (0.01 and 0.02, respectively) and C23:0 (0.0004 and 0.0005, respectively) in SFA, MSE values indicated that Bayesian alphabet models generally outperformed Bayesian GBLUP model and in Bayesian alphabet models, BayesA model outperformed BayesB and BayesC models for total and individual MUFA, PUFA and SFA.

Discussion

Descriptive statistics of fatty acids

The percentages of total SFA (71.20%), MUFA (25.30%) and PUFA (3.50%) found in this study were similar to the percentages of SFA (71.09%), UFA (28.95%) and PUFA (4.10%) from New Zealand dairy (Holstein-Friesian

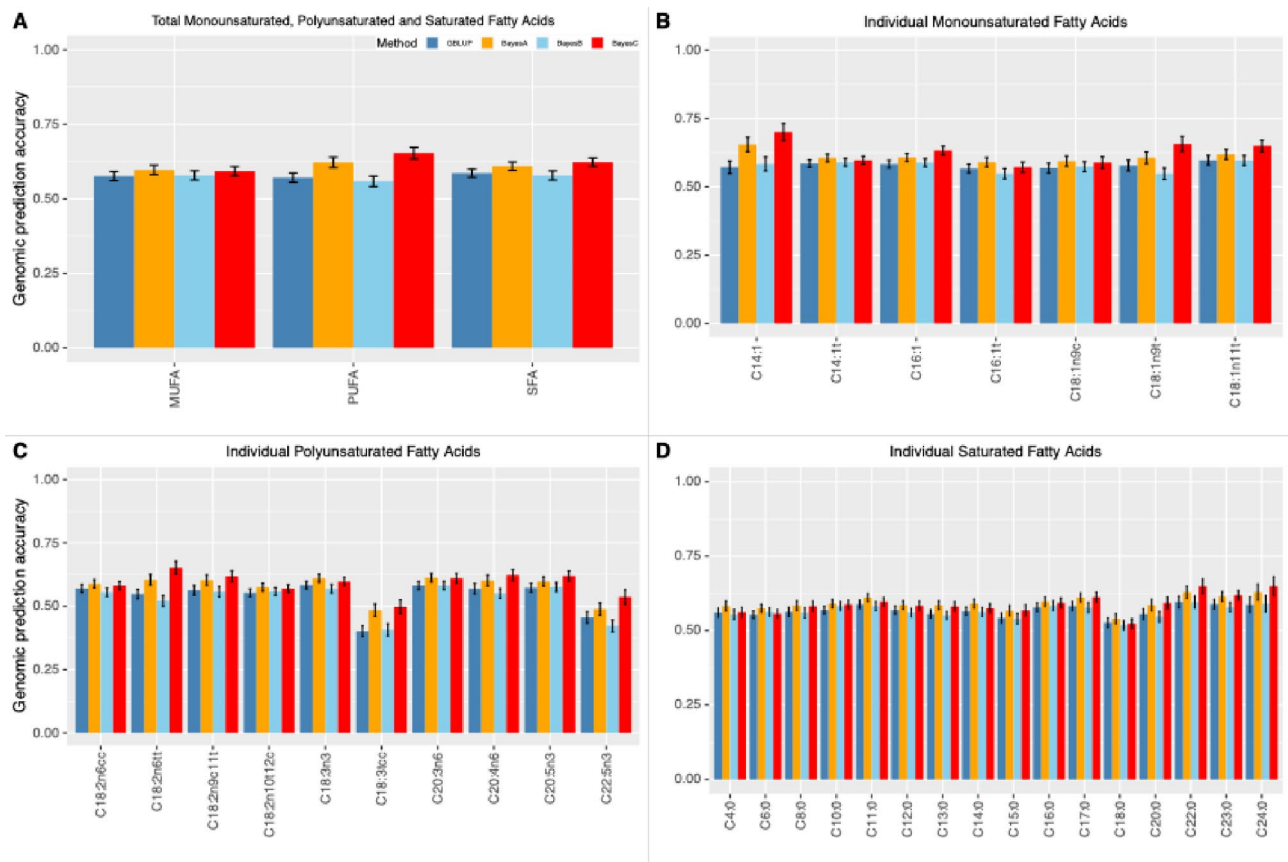


Fig. 6. Accuracies from Bayesian GBLUP, and Bayesian alphabet (BayesA, BayesB, BayesC) models for total and individual MUFA, PUFA and SFA.

× Jersey) cows reported by Lopez-Villalobos et al.¹⁷ and the percentages of SFA (70.3%), MUFA (26.6%) and PUFA (3.1%) from Danish Holstein cattle reported by Krag et al.¹⁹. However, these percentages were slightly different than the percentages of SFA (68.9%), MUFA (28.2%) and PUFA (2.9%) from Norwegian Red cows reported by Olsen et al.²² and the percentages of SFA (73.6%) and MUFA (26.4%) from the dairy cattle in the Walloon region of Belgium reported by Soyeurt et al.²³. These differences observed in the concentration of FA across studies could be attributed to the effects of diet, stage of lactation, parity, breed and genetic variation between cows^{6,10,17}.

Bovine milk fat includes fatty acyls, triglycerides, glycerophospholipids, sphingolipids and saccharolipids; however, 98% of milk fat is made up of triglycerides which are synthesized from more than 400 different FA. Many of the FA are about trace quantities and only a few FA are at the 1% level or higher²⁴. Lopez-Villalobos et al.¹⁷ indicated that C14:0 10.53%, C16:0 23.92% and C18:0 12.20% of the total FA in SFA and C18:1n9c 16.95% of the total FA in MUFA were the most abundant FA in New Zealand dairy cows, respectively. Krag et al.¹⁹ also reported that while the most abundant FA was C16:0 comprising 28.95% of the fat, the combination of C14:0, C16:0 and C18:0 in SFA, and C18:1n9c in MUFA accounted for 70.56% of the total fat content in Danish Holstein cattle. Olsen et al.²² also studied the FA in Norwegian Red cow population and found that C14:0, C16:0 and C18:0 acids in SFA, and C18:1n9c in MUFA were the most abundant FA by comprising 12.03, 27.11, 12.12 and 22.98% of the total fat, respectively. As seen in Table 1, the most abundant FA were C18:1n9c (20.55%) in MUFA and C14:0 (Myristic) (13.17%), C16:0 (36.07%) and C18:0 (9.70) in SFA and accounted for 79.49% of the total fat content in Canadian Holstein cows which were in a similar range to the percentages reported in the literature.

About 50–60% of the total SFA in bovine milk fat is C14:0, C16:0 and C18:0²⁵. Astrup et al.²⁶ indicated that C14:0, C16:0 and C18:0 are associated with increased levels of cholesterol and an increased risk of cardiovascular diseases. Desirable changes in milk FA in regard to human health are to increase the amounts of MUFA and PUFA, particularly C18:2n9c11t and n-3 FA, and to decrease in the amounts of SFA, particularly C14:0 and C16:0. FAO²⁷ also recommended to consumers to reduce SFA intake and consume more UFA.

Estimates of phenotypic and genetic correlations

Estimates of phenotypic and genetic correlations among 36 total and individual FA in MUFA, PUFA and SFA groups from Canadian Holstein cows are presented in Table S1 (the supplementary file) and correlations only for MUFA, PUFA and SFA are also showed in Figs. 1, 2 and 3. As seen in Figs. 1, 2 and 3 and Table S1, overall, the estimates of phenotypic correlations had the same direction of variation (negative or positive) as the

Fatty acid				
Monounsaturated fatty acids (MUFA)	GBLUP	BayesA	BayesB	BayesC
C14:1 (myristoleic acid)	0.11	0.11	0.11	0.11
C14:1t (myristelaidic acid)	1.19	1.00	0.97	0.97
C16:1 (palmitoleic acid)	0.10	0.09	0.09	0.09
C16:1t (palmitelaidic acid)	3.38	3.15	3.05	3.12
C18:1n9c (oleic acid)	399.70	361.13	360.90	366.45
C18:1n9t (elaidic acid)	0.06	0.06	0.06	0.06
C18:1n11t (trans vaccenic acid)	1.09	0.91	0.93	0.96
MUFA (total)	609.73	566.52	571.37	574.83
Polyunsaturated fatty acids (PUFA)				
C18:2n6cc (linoleic acid)	3.51	3.38	3.41	3.42
C18:2n6tt (trans-linoleic acid)	0.03	0.03	0.03	0.03
C18:2n9c11t (cis-9 trans-11 CLA)	0.17	0.16	0.16	0.16
C18:2n10t12c (trans-10 cis-12 CLA)	0.0003	0.0002	0.0002	0.0002
C18:3n3 (alpha linolenic acid)	0.19	0.18	0.18	0.18
C18:3tcc (gamma linolenic acid)	0.01	0.01	0.01	0.01
C20:3n6 (dihomogamma linolenic acid)	0.01	0.01	0.01	0.01
C20:4n6 (arachidonic acid)	0.02	0.02	0.02	0.02
C20:5n3 (eicosapentanoic acid)	0.0011	0.0011	0.0011	0.0011
C22:5n3 (docosapentaenoic acid)	0.06	0.07	0.06	0.06
PUFA (total)	11.73	10.90	10.85	11.07
Saturated fatty acids (SFA)				
C4:0 (butyric acid)	0.65	0.60	0.59	0.59
C6:0 (caproic acid)	1.03	0.96	0.96	0.98
C8:0 (caprylic acid)	0.83	0.74	0.74	0.75
C10:0 (capric acid)	7.47	6.67	6.59	6.77
C11:0 (undecanoic acid)	0.06	0.05	0.05	0.05
C12:0 (lauric acid)	13.28	12.07	12.27	12.56
C13:0 (tridecylic acid)	0.21	0.20	0.20	0.20
C14:0 (myristic acid)	161.20	148.18	149.89	149.45
C15:0 (pentadecylic acid)	1.37	1.29	1.26	1.28
C16:0 (palmitic acid)	1208.27	1124.80	1105.52	1128.73
C17:0 (margaric acid)	0.43	0.41	0.41	0.41
C18:0 (stearic acid)	91.01	80.62	78.54	78.29
C20:0 (arachidic acid)	0.02	0.01	0.01	0.01
C22:0 (behenic acid)	0.0031	0.0026	0.0026	0.0027
C23:0 (tricosanoic acid)	0.0005	0.0004	0.0004	0.0004
C24:0 (lignoceric acid)	0.0016	0.0014	0.0013	0.0014
SFA (total)	4714.61	4569.82	4616.49	4597.56

Table 2. Mean squared errors from Bayesian GBLUP, and Bayesian alphabet (BayesA, BayesB, BayesC) models for total and individual MUFA, PUFA and SFA.

corresponding estimates of genetic correlations. As seen in Fig. 1, phenotypic (−0.27 to 0.97) and genetic (−0.43 to 0.97) correlations for total and individual MUFA from Canadian Holstein cows were consistent with those reported for Italian Holstein-Friesian cows by Mele et al., for Italian Brown Swiss cows by Pegolo et al.¹⁸ and for New Zealand crossbred Holstein-Friesian×Jersey cows by Lopez-Villalobos et al.¹⁷.

For individual FA in PUFA, higher correlations between C18:3n3 and C18:2n9c11t acids for Danish Holstein cows¹⁹ and strong positive genetic correlations for C18:2n6cc with C18:3n3 (0.92) and C18:2n9c11t (0.63) for Canadian Holstein cattle²⁸ compared with those from this study were reported. However, genetic correlations among C20:3n6, C20:4n6, C20:5n3 and C22:5n3 reported by Bilal et al.²⁸ were found to be similar with those from this study.

As seen in Fig. 1 and Table S1, C14:1, C16:1 and C18:1n9c in MUFA were positively genetically correlated with each other (0.11 to 0.41) and with C18:2n9c11t (0.40) in PUFA which was consistent with those reported for Canadian Holstein cattle by Bilal et al.²⁸. Moreover, Bilal et al.²⁸ indicated that the possible explanation of positive genetic correlations among C14:1, C16:1 and C18:1n9c in MUFA and between MUFA and C18:2n9c11t lies in their common mode of synthesis. C18:1n9c which is expressed in 81.19% in total MUFA and constitutes

about 20.55% of the total milk fat, is predominantly synthesized by mammary uptake of C18:0 and its conversion to C18:1n9c acid by the enzymatic activity of the stearoyl-CoA desaturase enzyme (SCD)²⁹.

Correlation estimates between total and individual MUFA and PUFA ranged from −0.31 to 0.49 for phenotypic correlations and from −0.34 to 0.50 for genetic correlations, which were similar with those reported by Krag et al.¹⁹ for Danish Holstein cows. Correlation estimates between total and individual SFA and MUFA (PUFA) ranged from −0.97 to 0.69 (from −0.51 to 0.74) for phenotypic correlations and from −0.97 to 0.57 (from −0.53 to 0.81) for genetic correlations. There were also very strong phenotypic and genetic correlations for total SFA and MUFA (−0.97 and −0.97, respectively), for total SFA and C18:1n9c (−0.93 and −0.93, respectively) which was expressed in 81.19% in total MUFA. The very strong negative phenotypic and genetic correlations between total SFA and total MUFA and C18:1n9c in MUFA were also reported for Danish Holstein cows by Krag et al.¹⁹; for Danish Holstein and Danish Jersey cattle by Buitenhuis et al.³⁰ and Hein et al.³¹; for Italian Brown Swiss cows by Pegolo et al.¹⁸ and for New Zealand crossbred Holstein-Friesian×Jersey cows by Lopez-Villalobos et al.¹⁷ and indicated that dairy cows with a reduced the stearoyl-CoA desaturase enzyme (SCD) activity in the mammary gland tended to produce dairy milk with lower MUFA content and with greater preformed FA³². Moderate negative phenotypic and genetic correlations among total MUFA (C18:1n9c in MUFA) and the individual SFA were consistent with the negative genetic correlation pattern among the total MUFA and the individual SFA reported by Stoop et al.¹⁰, Mele et al.¹⁵, Lopez-Villalobos et al.¹⁷ and Soyeurt et al.²³.

The fatty acids in dairy milk are generally classified based on their chain lengths between 4 and 24 carbons. The even-numbered fatty acids being 4–16 carbons in chain length arise from de novo synthesis in the mammary gland and the other fatty acids, which include approximately half of the 16 carbon and all those 18 carbons or greater in length, are taken up preformed from the blood²¹. The phenotypic and genetic correlations between MUFA, PUFA and SFA given in Table S1 in the supplementary file could be evaluated by similarities in the origin of fatty acids. Positive phenotypic and genetic correlations were generally obtained between C4:0, C6:0, C8:0, C10:0, C12:0, C14:0 and C16:0 in SFA which were synthesized de novo in the mammary gland and between C18:1n9c, C18:1n9t and C18:1n11t in MUFA and between C18:2n6cc, C18:2n6tt, C18:2n9c11t, C18:2n10t12c, C18:3n3 and C18:3tcc in PUFA which were extracted from blood. However, as seen in Table S1 in the supplementary file, negative phenotypic and genetic correlations were obtained for C4:0, C6:0, C8:0, C10:0, C12:0, C14:0 and C16:0 in SFA (de novo synthesis in the mammary gland) with C18:1n9c, C18:1n9t and C18:1n11t in MUFA and with C18:2n6cc, C18:2n6tt, C18:2n9c11t, C18:2n10t12c, C18:3n3 and C18:3tcc in PUFA (the mammary uptake of preformed fatty acids).

The estimates of genetic correlations (Table S1 in the supplementary file) indicated the existence of the strong positive and negative genetic relationships between total and individual FA in MUFA, PUFA and SFA in bovine milk and the importance of using genetic relationships in the selection programs for FA to improve the nutritional quality of bovine milk and milk products.

Heritability estimates from Bayesian GBLUP and Bayesian alphabet models in random training data sets

In the literature, heritability estimates for SFA, MUFA and PUFA in bovine milk differ across studies based on the combination of individual FA, sample size, breed, statistical method and FA analytical methods. Heritability estimates for SFA, MUFA and PUFA in this study were generally higher than those reported for American Holstein cows by Bobe et al.³³, for Italian Holstein cows by Mele et al.¹⁵, for mixed-breed population of the Walloon region of Belgium by Soyeurt et al.²³, for Danish Holstein cows by Krag et al.¹⁹, for Italian Brown Swiss cows by Pegolo et al.¹⁸, for Norwegian Red cattle by Olsen et al.²² and for New Zealand crossbred Holstein-Friesian × Jersey cows by Lopez-Villalobos et al.¹⁷, but agree well with the heritability estimates of PUFA reported for Korean Holstein cows by Park et al.¹⁶. Olsen et al.²² also indicated that heritability estimates for FA obtained from Fourier transform infrared spectroscopy analyses in Norwegian Red cattle were usually lower than those from FA studies based on mid-infrared spectroscopy, but in the same range as in the FA studies based on gas chromatography. However, our heritability estimates for FA were found higher than those from Fourier transform infrared spectroscopy, mid-infrared spectroscopy or gas chromatography analyses reported in the literature^{15–19,22,23,33}. Lopez-Villalobos et al.¹⁷ indicated that comparisons of heritability estimates among the various studies were difficult because of differences in the analytical methods (GC or spectroscopy), the method of expressing FA contents (g/100 g of milk, g of FA/100 g of fat, or g of FA/100 g of total FA), sample size, breed, parity number, stage of lactation, and the statistical methods (e.g., sire, animal model and random regression).

As seen in Fig. 4, heritability estimates changed across the total and individual MUFA, PUFA and SFA and for individual SFA heritability estimates were also found higher than those for individual MUFA and PUFA in Canadian Holstein cows, which agreed with the results from Stoop et al.¹⁰, Lopez-Villalobos et al.¹⁷, Pegolo et al.¹⁸, Krag et al.¹⁹ and Bilal et al.²⁸. Bovine milk FA were synthesized de novo in the mammary gland of cows or derived from the diet of cows which leads to the possible physiological explanation of the heritability estimates of individual FA¹³. As seen in Fig. 4, the heritability estimates for individual SFA, MUFA and PUFA were generally moderate, and indicated that de novo synthesized FA (such as $h^2_{C4:0} = [0.58, 0.66]$, $h^2_{C8:0} = [0.58, 0.66]$, $h^2_{C10:0} = [0.55, 0.63]$, $h^2_{C12:0} = [0.57, 0.65]$, $h^2_{C14:0} = [0.59, 0.65]$, $h^2_{C16:0} = [0.57, 0.63]$ in SFA) were under genetic control and genetic selection could be effectively applied to alter the FA composition of bovine milk for human and animal health^{16,17}.

Predictive ability and accuracy of Bayesian GBLUP and Bayesian alphabet models

Predictive abilities of Bayesian GBLUP and Bayesian alphabet models were presented in Fig. 5 for total and individual MUFA, PUFA and SFA. As seen in Fig. 5, predictive ability of Bayesian GBLUP was higher than or similar with those of Bayesian alphabet models for total and individual MUFA, PUFA and SFA. The differences and similarities among the predictive abilities of Bayesian GBLUP and Bayesian alphabet models for total and

individual MUFA, PUFA and SFA could result from the model assumptions, heritability and genetic architecture of total and individual MUFA, PUFA and SFA traits.

Bayesian GBLUP model is considered as a direct estimation approach for genetic effects and variances. The estimation of the genetic effects in GBLUP model is based on the genomic relationship matrix (**G**) computed from SNP markers³⁴. The relatively few causative SNP markers hardly affect the genomic relationship matrix because of the equally contribution of each SNP marker to the genomic relationship^{34,35}. However, Bayesian alphabet models are considered as a SNP-based estimation approach for genetic effects and variances. The estimation of the genetic effects in Bayesian alphabet models is based on the π proportion of SNP markers having non-null effects, which are assumed to follow a univariate *Student's t*, multivariate *Student's t* distribution depending on the Bayesian models^{36,37}.

Heritability estimates and predictive ability results for total and individual MUFA, PUFA and SFA traits were based on the direct additive gene effect and variances; however, could depend on the number of loci affecting the FA traits, distribution of their effects, underlying relationship between FA traits and dominance and epistatic gene effects^{38–40}.

Freitas et al.²⁰ studied the estimation of genetic parameters and determination of genomic regions for the short-chain, medium-chain, long-chain, saturated, and unsaturated FA from North American Holstein cattle and indicated that various genomic regions are associated with the short-chain, medium-chain, long-chain, saturated, and unsaturated FA, but many regions across the genome have small individual effects on the total genetic variation of each trait. Wang et al.⁴¹ compared the genomic selection models (RR-BLUP, GBLUP, BayesA, BayesB, BayesCp and Bayesian LASSO,) for breeding value prediction in a simulation study and found that the predictive ability of BayesB model decreased markedly for the trait determined by many genes and were much better than GBLUP for the trait determined by 20 QTL. However, as the number of QTL determined the trait increased, the predictive ability of BayesB decreased gradually to a plateau even lower than that of GBLUP.

Mean accuracy values with standard errors of Bayesian GBLUP and Bayesian alphabet models were given for total and individual MUFA, PUFA and SFA from Canadian Holstein cows in Fig. 6 after removing the effect of heritability of FA traits on the predictive ability of the models. As seen in Fig. 6, error-bar intervals of mean accuracy values of Bayesian GBLUP and Bayesian alphabet models indicated that mean accuracy values of Bayesian GBLUP and Bayesian alphabet models were similar for large number of individual MUFA, PUFA and SFA. However, BayesC model resulted in slightly higher accuracy than Bayesian GBLUP for C14:1, C18:2n6tt, C18:3tcc, C20:4n6, C20:5n3, C22:5n3, C24:0. When the number of SNP markers increases the computational time for Bayesian alphabet models take longer time to obtain the estimates of parameters. Therefore, Bayesian GBLUP is an effective method in providing the best compromise between computational time and prediction ability if there are no major gene effects for FA traits.

Conclusions

The heritability estimates ranged between 0.61 ± 0.01 and 0.67 ± 0.01 for total MUFA, between 0.35 ± 0.01 and 0.49 ± 0.01 for PUFA and between 0.51 ± 0.01 and 0.60 ± 0.01 for SFA, respectively and for individual FA, ranged between 0.27 ± 0.01 and 0.69 ± 0.01 for MUFA, between 0.27 ± 0.01 and 0.68 ± 0.01 for PUFA and between 0.35 ± 0.01 and 0.69 ± 0.01 for SFA. The heritability estimates indicated that total and individual MUFA, PUFA and SFA traits were under moderate genetic control in Canadian Holstein cattle and can be improved through genomic selection. Moderate and high genetic correlations among total and individual MUFA, PUFA and SFA traits were also determined and these underlying genetic linear relationships among individual MUFA, PUFA and SFA traits need to be taken into consideration for successful genomic selection for FA traits. Comparison of mean accuracy values of Bayesian GBLUP and Bayesian alphabet models based on their standard errors indicated that Bayesian GBLUP and Bayesian alphabet models resulted in similar accuracies for large number of individual MUFA, PUFA and SFA. However, BayesC model yielded better accuracies than Bayesian GBLUP for C14:1, C18:2n6tt, C18:3tcc, C20:4n6, C20:5n3, C22:5n3, C24:0, indicating that these individual FA traits might be determined by many genes having non-null effects, which are assumed to follow a univariate or multivariate *Student's t* distribution. Overall comparison and evaluation of Bayesian GBLUP and Bayesian alphabet models based on heritability estimates of FA traits, predictive ability and accuracies of models indicated that when the number of SNP markers increases the computational time for Bayesian alphabet models take longer time to obtain the estimates of parameters. Therefore, Bayesian GBLUP is an effective method to estimate heritabilities of FA traits and to provide the best compromise between computational time and prediction ability and accuracy if there are no major gene effects for FA traits⁴¹.

Methods

Ethical statement

Animal use procedures and protocols were according to the national codes of practice for the care and handling of farm animals (<http://www.nfacc.ca/codes-of-practice>) and the milk samples were collected by the technicians of the Valacta laboratory.

Animals and milk sampling

Milk samples from 1246 Canadian Holstein cows from 16 herds in the province of Quebec, Canada, were collected during the morning milking by Valacta (www.valacta.com) (now Lactanet, <https://lactanet.ca/>) as part of its routine monthly milk collection for Dairy Herd Improvement services. The dairy farms in Quebec are managed following standard procedures instituted by the Dairy Farmers of Quebec and Canada (<https://lait.org/>; <https://dairyfarmersofcanada.ca/>). Milk samples were delivered to our laboratory on the same day of collection

and processed immediately. Detailed procedures for milk processing, fatty acid analysis, DNA isolation and SNP genotyping by the method of genotyping-by-sequencing (GBS) have been reported in our previous work⁴².

Phenotypes

Out of the 1246 cows, the phenotypic data of individual FA within MUFA, PUFA and SFA groups from 695 individuals in mid-lactation and with lactation number ranging from one to five were further analyzed. Data manipulation for the phenotypic data of individual FA traits was carried out to identify outliers and provide normally distributed data before the genomic analysis. Then, the number of observations ranging from 636 to 695 were used for the genomic analysis of the FA traits.

Forty mL of milk sample from each cow was used to extract the milk fat followed by FA analysis as reported previously by Ibeagha-Awemu et al.⁴². Briefly, FA in the extracted milk fat were re-esterified to their methyl esters. Then, FA methyl esters were determined by gas chromatography (GC) method using Varian CP-3900 GC based on O'Fallon et al.⁴³. After determining the individual FA methyl ester peaks, FA proportion is specified by counting peak area proportion to the total peak area of all determined fatty acids and weight% fatty acid data (% total FA) were calculated from the area data. A total of 33 individual FA, C14:1 (Myristoleic), C14:1t (Myristelaic), C16:1 (Palmitoleic), C16:1t (Palmitelaic), C18:1n9c (Oleic), C18:1n9t (Elaidic), C18:1n11t (Trans vaccenic) acids in MUFA; C18:2n6cc (Linoleic), C18:2n6tt (Trans-linoleic), C18:2n9c11t (Cis-9 trans-11 CLA), C18:2n10t12c (Trans-10 cis-12 CLA), C18:3n3 (Alpha linolenic), C18:3tcc (Gamma linolenic), C20:3n6 (Dihomogamma linolenic), C20:4n6 (Arachidonic), C20:5n3 (Eicosapentaenoic), C22:5n3 (Docosapentaenoic) acids in PUFA; and C4:0 (Butyric), C6:0 (Caproic), C8:0 (Caprylic), C10:0 (Capric), C11:0 (Undecanoic), C12:0 (Lauric), C13:0 (Tridecylc), C14:0 (Myristic), C15:0 (Pentadecylc), C16:0 (Palmitic), C17:0 (Margaric), C18:0 (Stearic), C20:0 (Arachidic), C22:0 (Behenic), C23:0 (Tricosanoic), C24:0 (Lignoceric) acids in SFA were identified by the GC method. Total MUFA, PUFA and SFA values were calculated by summing the individual FA values within MUFA, PUFA and SFA groups, respectively.

Genotyping-by-sequencing SNP marker genotypes

The genomic DNA for genotyping-by-sequencing was isolated from milk somatic cells obtained from the 40 mL of milk sample taken from each cow followed by GBS analysis as reported previously by Ibeagha-Awemu et al.⁴². Briefly, after preparing GBS libraries, they were multiplexed and subjected to single end 100 bp sequencing on an Illumina HiSeq 2000 system (Illumina Inc., San Diego, CA, USA). The Tassel software package was utilized to process raw Illumina DNA sequence data and to call SNP markers. Further details on the GBS analysis and bioinformatics for GBS SNP marker calling have been given by Ibeagha-Awemu et al.⁴². In this study, a total of 76,299 SNPs out of 515,787 with call rates > 85%, accuracy of imputation score > 50% and MAF ≥ 1.5% were used. After quality control of GBS SNP marker data, considerable amounts of GBS SNP markers were removed from the genotype data set since GBS SNP markers had the capacity to deliver a large number of markers, but with considerable amounts of missing values and unknown marker positions⁴⁴. The findings from Elbasyoni et al.⁴⁴, Gouesnard et al.⁴⁵ and Negro et al.⁴⁶ also indicated that GBS SNP markers had low coverage leading to a high (65%) missing data rate and had low call rate compared to the findings from array SNP markers.

Estimates of phenotypic and genetic correlations between fatty acids

Genetic correlations among 36 total and individual MUFA, PUFA and SFA were calculated based on the (co) variance components obtained from a total of 630 Bayesian bivariate GBLUP models.

A Bayesian bivariate GBLUP model for any two fatty acids is given in Eq. (1) as follows:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (1)$$

where y_1 and y_2 were the vector of observations from any two fatty acids among 36 total and individual MUFA, PUFA and SFA; X_1 , X_2 , Z_1 and Z_2 were the design matrices relating the fixed (overall mean, lactation number, herd) and random (direct additive genetic) effects for two fatty acids; b_1 and b_2 were the vectors of fixed effects (overall mean, lactation number and herd) for two fatty acids; and g_1 and g_2 were the vectors of random direct additive genetic effects following a bivariate *Normal* distribution $\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \sim N(0, G \otimes \Sigma)$ with the mean

zero and (co)variance $G \otimes \Sigma$ where G is the genomic relationship matrix and $\Sigma = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_1 g_2} \\ \sigma_{g_1 g_2} & \sigma_{g_2}^2 \end{bmatrix}$

is the direct additive genetic (co)variance matrix with the direct additive genetic (co)variance components of $\sigma_{g_1}^2$, $\sigma_{g_2}^2$ and $\sigma_{g_1 g_2}$ for two fatty acids; and e_1 and e_2 were the vectors of random residuals following a bivariate *Normal* distribution $\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N(0, I \otimes R)$ with the mean zero and the residual (co)variance

matrix $R = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 \end{bmatrix}$ where $\sigma_{e_1}^2$, $\sigma_{e_2}^2$ and $\sigma_{e_1 e_2}$ were the residual (co)variance components for two fatty acids. The vectors of fixed effects are assigned flat prior (a Gaussian prior with null mean and very large variance) and the covariance matrices of direct additive genetic and residual random effects are assigned Inverse Wishart priors.

The genomic relationship matrix (G) was computed as in Eq. (2)

$$\mathbf{G} = \frac{(\mathbf{M} - \mathbf{P})(\mathbf{M} - \mathbf{P})^T}{2 \sum_{i=1}^k p_i(1 - p_i)} \quad (2)$$

where the matrix of \mathbf{M} included the values of 0, 1 or 2 for SNP markers; the matrix of \mathbf{P} included the allele frequencies of SNP markers multiplied by 2; p_i was the allele frequency of SNP marker i ; and the k was the number of SNP markers^{34,47}.

The Bayesian bivariate GBLUP models were applied using the Multitrait function in the statistical package BGLR (<https://cran.r-project.org/web/packages/BGLR/index.html> (accessed on 15 May 2023)) in the R program^{37,48}.

A total of 630 genetic (r_G) and phenotypic (r_P) correlations among 36 total and individual MUFA, PUFA and SFA were calculated as in Eq. (3)

$$r_G = \frac{\sigma_{g_1 g_2}}{\sqrt{\sigma_{g_1}^2 \sigma_{g_2}^2}} \text{ and } r_P = \frac{\sigma_{p_1 p_2}}{\sqrt{\sigma_{p_1}^2 \sigma_{p_2}^2}} \quad (3)$$

where $\sigma_{p_1}^2$, $\sigma_{p_2}^2$ and $\sigma_{p_1 p_2}$ were the phenotypic (co)variance components for two fatty acid traits. The standard error (r_{se}) of r_G or r_P was calculated as $r_{se} = \sqrt{(1 - r^2)/(n - 2)}$ where r was the r_G or r_P and n was the sample size⁴⁹.

Genomic selection

Analyses of 36 total and individual FA traits in MUFA, PUFA and SFA groups for genomic selection were carried out by applying the univariate Bayesian (GBLUP) and the Bayesian alphabet models^{36,37}.

The phenotypes from a FA were modeled by using the statistical model in Eq. (4):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{g} + \mathbf{e} \quad (4)$$

where \mathbf{y} was the vector of observations from a FA trait in MUFA, PUFA or SFA; \mathbf{X} was the design matrix allocating a FA trait in MUFA, PUFA or SFA to the fixed effects of overall mean, lactation number and herd. \mathbf{b} was the vector of fixed effects (overall mean, lactation number and herd) and flat (un-informative) priors were assigned. \mathbf{e} was the vector of random residuals and assigned a priori multivariate *Normal* distribution $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ with the mean zero and variance $\mathbf{I}\sigma_e^2$ where \mathbf{I} is the identity matrix. σ_e^2 is the residual variance with prior density $\chi^{-2}(\nu_e, S_e)$ with degrees of freedom $\nu_e = 5$ and scale S_e parameters which is derived using the following equation $0.5 \times \text{var}(\mathbf{y}) \times (\nu_e + 2)$ ³⁷.

The term of $\mathbf{Z}\mathbf{g}$ in Eq. (4) was explained based on the Bayesian GBLUP and the Bayesian alphabet models.

Bayesian GBLUP model

For GBLUP, \mathbf{Z} was the design matrix allocating a FA trait in MUFA, PUFA or SFA to additive genomic breeding values of animals and \mathbf{g} was the vector of additive genomic breeding values for animals following a priori multivariate *Normal* distribution $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$ with mean zero and variance $\mathbf{G}\sigma_g^2$ where \mathbf{G} is the genomic relationship matrix and σ_g^2 is the direct additive genetic variance with prior density $\chi^{-2}(\nu_g, S_g)$ with degrees of freedom $\nu_g = 5$ and scale S_g parameters.

The Bayesian alphabet models

The phenotypes from a FA were modeled as a function of the SNP marker effects in the Bayesian alphabet models. Thus, for Bayesian alphabet models, \mathbf{Z} was the matrix of SNP marker values (0, 1 or 2) from animals and \mathbf{g} was the vector of SNP marker effects. The prior distribution for individual SNP marker effect (g_i) depends on the SNP marker-specific variance $\sigma_{g_i}^2$ and the parameter $0 \leq \pi \leq 1$ which represents the prior proportion of non-zero SNP marker effects. In BayesA, every SNP marker is assumed to have an effect on the phenotypes of FA. Hence, $\pi = 1$ and the prior distribution of g_i is assumed to be *Normal* $g_i | \sigma_{g_i}^2 \sim N(0, \sigma_{g_i}^2)$ with prior density $\chi^{-2}(\nu, S)$ for $\sigma_{g_i}^2$. Thus, the marginal distribution of g_i will be the univariate scaled-*Student's t* distribution $t(0, \nu, S)$ with parameters ν and S . In BayesB, only a small proportion (π) of SNP markers have non-zero effect and the majority $(1 - \pi)$ of the SNP markers have exactly zero effect on the phenotypes of FA. The *Gaussian Mixture* prior distribution of g_i is defined as in Eq. (5)

$$g_i | \pi, \sigma_{g_i}^2 \begin{cases} = 0 & \text{with } (1 - \pi) \\ \sim N(0, \sigma_{g_i}^2) & \text{with } \pi \end{cases} \quad (5)$$

with prior density $\chi^{-2}(\nu, S)$ for $\sigma_{g_i}^2$. Thus, the marginal distribution of g_i with the π will be the univariate scaled-*Student's t* distribution $t(0, \nu, S)$ with parameters ν and S . In BayesC, the prior distribution for g_i depends on the SNP marker-common variance σ_g^2 and a small proportion (π) of SNP markers have non-zero effect. Hence, The *Gaussian Mixture* prior distribution of g_i is defined as in Eq. (6)

$$g_i | \pi, \sigma_g^2 \begin{cases} = 0 & \text{with } (1 - \pi) \\ \sim N(0, \sigma_g^2) & \text{with } \pi \end{cases} \quad (6)$$

with SNP marker-common variance σ_g^2 which is distributed $\chi^{-2}(\nu, S)$. Thus, the marginal distribution of g_i with the π will be the mixture of multivariate Student's t distributions $t(\mathbf{0}, \nu, IS)$ with parameters ν and S ^[36,37]. In BayesB and BayesC, π is treated as known with the value of 0.05.

The GEBV of animal j for a FA in Bayesian alphabet models was calculated as in Eq. (7)

$$GEBV_j = \sum_{i=1}^k m_{ji} \hat{g}_i \quad (7)$$

where m_{ji} is the SNP marker score (0, 1 or 2) for animal j at SNP i and \hat{g}_i is the estimated effect of i^{th} SNP marker³⁶. Then, the estimates of heritability for a FA from Bayesian alphabet (BayesA, BayesB, BayesC) models

was computed as $\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2}$ where $\hat{\sigma}_g^2$ and $\hat{\sigma}_e^2$ are the estimates of direct additive genetic and residual variances. The BGLR package (<https://cran.r-project.org/web/packages/BGLR/index.html> (accessed on 15 May 2023)) in the R program^{37,48} was utilized to get the estimates (\hat{b} , \hat{g} , $\hat{\sigma}_g^2$ and $\hat{\sigma}_e^2$) of b , g , σ_g^2 and σ_e^2 . Fixed and random direct additive genetic effects and variance components in Bayesian GBLUP and Bayesian alphabet models were estimated using the Markov chain Monte Carlo sampling algorithm. Markov chains were run for 25,000 cycles of Gibbs sampling. The first 5,000 cycles were discarded as burn-in and then every fifth sample in the cycle was saved for estimating model parameters.

10-fold cross-validation and predictive ability of the models

The cross-validation (CV) is a robust and nonparametric technique to test the predictive ability of the model by splitting the data set into the training and the validation data sets to train and evaluate the model. In this study, after the FA data set based on the CV strategy was randomly split into 10-fold CV data sets, all the observations from the nine CV data sets were used to train the Bayesian GBLUP and the Bayesian alphabet models. Then, all the observations from the leave-one-out CV data set were used to validate the models. This procedure based on the 10-fold CV was replicated five times.

The predictive abilities of the Bayesian GBLUP and the Bayesian alphabet models were determined in the 10-fold CV data sets using Pearson's correlation coefficient ($r_{y, GEBV}$) between the observed phenotypic values (y) and $GEBV$ for each FA trait in MUFA, PUFA or SFA (Eq. 8):

$$r_{y, GEBV} = \frac{Cov(y, GEBV)}{\sqrt{Var(y) Var(GEBV)}} \quad (8)$$

where $Cov(y, GEBV)$ was the covariance between the observed phenotypic values (y) and $GEBV$ for the FA trait in MUFA, PUFA or SFA; $Var(y)$ and $Var(GEBV)$ were the variances for the observed phenotypic values (y) and $GEBV$ for the FA trait in MUFA, PUFA or SFA, respectively.

Accuracy of the Bayesian GBLUP and the Bayesian alphabet models was defined as the correlation of $r(g, \hat{g})$ between genomic breeding values (g) and genomic estimated breeding values (\hat{g}) for animals and approximately estimated from the predictive ability of the models as shown by Legarra et al.⁵⁰ (Eq. (9)):

$$r_{g, \hat{g}} = \frac{r(y, GEBV)}{\sqrt{\hat{h}^2}} \quad (9)$$

where \hat{h}^2 was the heritability estimate from the Bayesian GBLUP or the Bayesian alphabet models for FA.

The mean squared error (MSE) of prediction in the CV data sets was used to measure the overall fit of model, and the computation equation (Eq. 10) was

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - GEBV_i)^2 \quad (10)$$

where N is the number of individuals. A large estimated value of accuracy is indicative of reliable prediction, and a low MSE value means a better overall fit⁵¹.

Data availability

The SNP marker data used in this study have been deposited in the figshare repository and can be accessed via the following link: <https://doi.org/10.6084/m9.figshare.28560845>.

Received: 12 October 2024; Accepted: 1 April 2025

Published online: 22 April 2025

References

1. Van Maurice, M. H. T. et al. Differences in milk fat composition predicted by mid-infrared spectrometry among dairy cattle breeds in the Netherlands. *J. Dairy. Sci.* **96**, 2570–2582 (2013).
2. Bobe, G. et al. Short communication. Composition of milk protein and milk fatty acids is stable for cows differing in genetic merit for milk production. *J. Dairy. Sci.* **90**, 3955–3960 (2007).
3. Yang, T. N-3 PUFAs have antiproliferative and apoptotic effects on human colorectal cancer stem-like cells in vitro. *J. Nutr. Biochem.* **24**, 744–753. <https://doi.org/10.1016/j.jnutbio.2012.03.023> (2013).
4. Burgos, S. A. et al. The role of dairy fat on cardiometabolic health: what is the current state of knowledge? *Can. J. Anim. Sci.* **99**, 429–441 (2019).

5. Bilal, G. et al. Short communication: estimates of heritabilities and genetic correlations among milk FA unsaturation indices in Canadian Holsteins. *J. Dairy. Sci.* **95**, 7367–7371 (2012).
6. Palmquist, D. L. et al. Feed and animal factors influencing milk fat composition. *J. Dairy. Sci.* **76**, 1753–1771 (1993).
7. Jenkins, T. C. et al. Major advances in nutrition: impact on milk composition. *J. Dairy. Sci.* **89**, 1302–1310 (2006).
8. Butler, G. et al. Fat composition of organic and conventional retail milk in Northeast England. *J. Dairy. Sci.* **94**, 24–36 (2011).
9. Beaulieu, A. D. et al. Differential effects of high fat diets on fatty acid composition in milk of Jersey and Holstein cows. *J. Dairy. Sci.* **78**, 1336–1344 (1995).
10. Stoop, W. M. et al. Genetic parameters for major milk fatty acids and milk production traits of Dutch Holstein-Friesians. *J. Dairy. Sci.* **91**, 385–394 (2008).
11. Soyeurt, H. et al. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *J. Dairy. Sci.* **94**, 1657–1667 (2011).
12. Grummer, R. R. Effect of feed on the composition of milk fat. *J. Dairy. Sci.* **74**, 3244–3257 (1991).
13. Bauman, D. E. & Griinari, J. M. Nutritional regulation of milk fat synthesis. *Annu. Rev. Nutr.* **23**, 203–227 (2003).
14. Walker, G. P., Dunshea, F. R. & Doyle, P. T. Effects of nutrition and management on the production and composition of milk fat and protein: a review. *Aust J. Agric. Res.* **55**, 1009–1028 (2004).
15. Mele, M. et al. Genetic parameters of conjugated Linoleic acid, selected milk fatty acids, and milk fatty acid unsaturation of Italian Holstein-Frisian cows. *J. Dairy. Sci.* **92**, 392–400 (2009).
16. Park, C. H. et al. Genetic parameters for milk fatty acid composition of Holstein in Korea Asian-Australas. *J. Anim. Sci.* **33**, 1573–1578 (2020).
17. Lopez-Villalobos, N. et al. Genetic correlations of milk fatty acid contents predicted from milk mid-infrared spectra in new Zealand dairy cattle. *J. Dairy. Sci.* **103**, 7238–7248 (2020).
18. Pegolo, S. et al. Genetic and environmental relationships of detailed milk fatty acids profile determined by gas chromatography in brown swiss cows. *J. Dairy. Sci.* **99**, 1315–1330. <https://doi.org/10.3168/jds.2015-9596> (2016).
19. Krag, K. et al. Genetic parameters for milk fatty acids in Danish Holstein cattle based on SNP markers using a bayesian approach. *BMC Genet.* **14** (79), 1–10 (2013).
20. Freitas, P. H. F. et al. Genomic analyses for predicted milk fatty acid composition along the lactation in North American Holstein cattle. *J. Dairy. Sci.* **103**, 5263–5269 (2020).
21. Bouwman, A. C. et al. Genome-wide association of milk fatty acids in Dutch dairy cattle. *BMC Genet.* **12** (43), 1–12. <https://doi.org/10.1186/1471-2156-12-43> (2011).
22. Olsen, H. G. et al. Genome-wide association mapping for milk fat composition and fine mapping of a QTL for de Novo synthesis of milk fatty acids on bovine chromosome 13. *Genet. Sel. Evol.* **49** (20), 1–13 (2017).
23. Soyeurt, H. et al. Estimation of heritability and genetic correlations for the major fatty acids in bovine milk. *J. Dairy. Sci.* **90**, 4435–4442 (2007).
24. Lindmark Månsson, H. Fatty acids in bovine milk fat. *Food Nutr. Res.* **52** (1821), 1–3 (2008).
25. Mansson, H. L. Fatty acids in bovine milk fat. *Food Nutr. Res.* <https://doi.org/10.3402/fnr.v52i0.1821> (2008).
26. Astrup, A. et al. The role of reducing intakes of saturated fat in the prevention of cardiovascular disease: where does the evidence stand in 2010. *Am. J. Clin. Nutr.* **93**, 684–688 (2011).
27. FAO (Food and Agriculture Organization of the United Nations). *Fats and Fatty Acids in Human Nutrition. Report of an Expert Consultation.* 91 (FAO, 2010).
28. Bilal, G. et al. Short communication: genetic parameters of individual fatty acids in milk of Canadian Holsteins. *J. Dairy. Sci.* **97**, 1150–1156 (2014).
29. Bauman, D. E. et al. Major advances associated with the biosynthesis of milk. *J. Dairy. Sci.* **89**, 1235–1243 (2006).
30. Buitenhuis, B. et al. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. *BMC Genom.* **15** (1112), 1–11. <https://doi.org/10.1186/1471-2164-15-1112> (2014).
31. Hein, L. et al. Genetic analysis of predicted fatty acid profiles of milk from Danish Holstein and Danish Jersey cattle populations. *J. Dairy. Sci.* **101**, 2148–2157 (2018).
32. Kay, J. K. et al. Endogenous synthesis of cis-9, trans-11 conjugated Linoleic acid in dairy cows fed fresh pasture. *J. Dairy. Sci.* **87**, 369–378 (2004).
33. Bobe, G. et al. Short communication: estimates of genetic variation of milk fatty acids in US Holstein cows. *J. Dairy. Sci.* **91**, 1209–1213 (2008).
34. vanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy. Sci.* **91**, 4414–4423 (2008).
35. Meuwissen, T. H. et al. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* **185**, 623–631. <https://doi.org/10.1534/genetics.110.116590> (2010).
36. Habier, D. et al. Extension of the bayesian alphabet for genomic selection. *BMC Bioinform.* **12** (186), 1–12 (2011).
37. Pérez, P. et al. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483–495. <https://doi.org/10.1534/genetics.114.164442> (2014).
38. Daetwyler, H. D. et al. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3** (10), 1–8. (2008).
39. Goddard, M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257. <https://doi.org/10.1007/s10709-008-9308-0> (2009).
40. Meuwissen, T. H. Accuracy of breeding values of unrelated individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* **41** (35), 1–9. <https://doi.org/10.1186/1297-9686-41-35> (2009).
41. Wang, X. et al. A comparison of genomic selection methods for breeding value prediction. *Sci. Bull. (Beijing)* **60**, 925–935. <https://doi.org/10.1007/s11434-015-0791-2> (2015).
42. Ibeagha-Awemu, E. M. et al. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Sci. Rep.* **6** (31109), 1–18. <https://doi.org/10.1038/srep31109> (2016).
43. O'Fallon, J. V. et al. A direct method for fatty acid Methyl ester synthesis: application to wet meat tissues, oils, and feedstuffs. *J. Anim. Sci.* **85**, 1511–1521. <https://doi.org/10.2527/jas.2006-491> (2007).
44. Elbasyoni, I. S. et al. A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant. Sci.* **270**, 123–130 (2018).
45. Gouesnard, B. et al. Genotyping-by-sequencing highlights original diversity patterns within a European collection of 1191 maize Flint lines, as compared to the maize USDA genebank. *Theor. Appl. Genet.* **130** (10), 2165–2189 (2017).
46. Negro, S. S. et al. Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. *BMC Plant. Biol.* **19**, 318. <https://doi.org/10.1186/s12870-019-1926-4> (2019).
47. Habier, D. et al. The impact of genetic relationship information of genome-assisted breeding values. *Genetics* **177**, 2389–2397 (2007).
48. R Core Team. R: A Language and environment for statistical computing; R foundation for statistical computing: Vienna, Austria. (accessed 10 March 2022); <https://www.R-project.org/>
49. Hayter, A. *Probability and Statistics for Engineers and Scientists* (Brooks/Cole, 2012).
50. Legarra, A. et al. Performance of genomic selection in mice. *Genetics* **180**, 611–618. <https://doi.org/10.1534/genetics.108.088575> (2008).

51. Zhu, B. et al. Accuracies of genomic prediction for Twenty economically important traits in Chinese simmental beef cattle. *Anim. Genet.* **50**, 634–643 (2019).

Acknowledgements

We thank Valacta laboratories (www.valacta.com) (now Lactanet, <https://lactanet.ca/>) for sampling and providing the milk samples on the animals. The authors acknowledge the Faculty Development Grant from Berry College to S.O.P. to work on this project in the summer of 2023. The numerical calculations reported in this paper were fully/partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources) in Türkiye.

Author contributions

S.O.P and K.K conceptualized the study. E.I.A and X.Z provided the data. S.O.P and K.K performed all the analyses, E.I.A, X.Z, K.K and S.O.P interpreted the data. S.O.P, K.K, E.I.A and X.Z drafted the manuscript; All authors discussed the results, commented on the manuscript and approved the final version.

Funding

Financial support for generation of GBS genotypes was provided by DairyGen (Dairy Cattle Genetics Research and Development Council of Canadian Dairy Network) and NSERC (The Natural Sciences and Engineering Research Council of Canada) (CRDPJ 453002-13).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-96839-1>.

Correspondence and requests for materials should be addressed to S.O.P. or X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025