Received: 2 November 2021 | Revised: 28 February 2022 | Accepted: 13 April 2022

DOI: 10.1111/jbg.12681

ORIGINAL ARTICLE

Journal of Animal Breeding and Genetics WILEY

Breeding for improved protein fractions and free amino acids concentration in bovine milk

Giulio Visentin[1] | **Donagh P. Berry**[2] | **Angela Costa**[1,3] | **Audrey McDermott**[2,3] | **Massimo De Marchi**[3] | **Sinead McParland**[2]

[1]Department of Veterinary Medical Sciences, Alma Mater Studiorum - University of Bologna, Ozzano dell'Emilia, Italy

[2]Teagasc Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

[3]Department of Agronomy, Food, Natural resources, Animals and Environment, University of Padova, Legnaro, Italy

Correspondence
Angela Costa, Department of Veterinary Medical Sciences, Alma Mater Studiorum – University of Bologna, Via Tolara di Sopra 50, 40064 Ozzano dell'Emilia (BO), Italy and Department of Agronomy, Food, Natural resources, Animals and Environment, University of Padova, Legnaro, Italy.
Emails: angela.costa@unipd.it; angela.costa2@unibo.it

Abstract

Considerable resources are required to routinely measure detailed milk compositional traits. Hence, an insufficient volume of phenotypic data can hinder genetic progress in these traits within dairy cow breeding programmes. The objective of the present study was to quantify the opportunities for breeding for improved milk protein and free amino acid (FAA) composition by exploiting mid-infrared spectroscopy (MIRS) predictions routinely recorded from milk samples. Genetic parameters for protein fractions and FAA composition were estimated using 134,546 test-day records from 16,166 lactations on 9,572 cows using linear mixed models. Heritability of MIRS-predicted protein fractions ranged from 0.19 (α-lactalbumin) to 0.55 (β-lactoglobulin A), while heritability of MIRS-predicted FAA ranged from 0.08 for glycine to 0.29 for glutamic acid. Genetic correlations among the MIRS-predicted FAA were moderate to strong ranging from −0.44 (aspartic acid and lysine) to 0.97 (glutamic acid and total FAA). Adjustment of the genetic correlations for the genetic merit of 24-h milk yield did not greatly affect the correlations. Results from the current study highlight the presence of exploitable genetic variation for both protein fractions and FAA in dairy cow milk. Besides, the direction of genetic correlations reveals that breeding programmes directly selecting for greater milk protein concentration carry with them favourable improvement in casein and whey fractions.

KEYWORDS

bovine milk, free amino acid, genetic correlation, heritability, protein fraction

1 | INTRODUCTION

Although most dairy cow breeding objectives include milk protein and fat concentration and yield at the macro-level (Cole & VanRaden, 2018; Miglior et al., 2005), few consider the detailed milk composition. However, detailed milk composition affects the quality of a variety of dairy products that can be manufactured (Bonfatti, Di Martino, et al. 2011; Wedholm et al., 2006). For example, the concentration of casein (CN) in milk has a favourable effect on the quantity of protein transferred from milk into cheese curd. High concentrations of α_{s1}-casein (α_{s1}-CN), β-casein (β-CN), κ-casein (κ-CN) and β-lactoglobulin B (β-LG B) are known to increase cheese yield (Wedholm et al., 2006).

Milk rich in FAA tends to present unfavourable cheese-making characteristics mainly due to protein hydrolysis

(Auldist et al., 1995; Urech et al., 1999). Nevertheless, presence of FAA is important for some products like infant formulas even though the milk from mammals differs in FAA composition and concentration. In fact, cow milk presents lower quantity of FAA than human milk (McDermott et al., 2016; van Sadelhoff et al., 2020). Some FAA are known to have a key role in the development of the immune system, particularly in infants (van Sadelhoff et al., 2020). However, bovine milk FAA have generally been overlooked thus far, even if some of them are essential or conditionally essential, such as leucine, valine and glycine. For nutritional reasons, FAA-enriched infant formulas and amino acid-based products may be valuable when breastfeeding is not possible (van Sadelhoff et al., 2020). Despite the high cost, amino acid-based supplements provide proteins in non-allergenic forms and are thus suitable for people with bovine milk protein allergies (Verduci et al., 2019). Given these considerations, it is worth investigating whether and how bovine milk FAA profile could be manipulated through breeding.

Interbreed differences in the concentration of protein fractions and FAA have been documented (Auldist et al., 2004; Lopez-Villalobos, 2012; McDermott et al., 2017) and individual protein fractions, including α-lactalbumin (α-LA) and γ-casein (γ-CN), are known to be heritable (Graml & Pirchner, 2003). For example, Schopen et al. (2009) reported a heritability of between 0.25 (β-CN) and 0.80 (β-LG) for protein fractions quantified using capillary zone electrophoresis. Similarly, Huang et al. (2012) reported that the proportion of total phenotypic variance attributable to additive genetic variation varied between 0.33 (α_{s1}-CN, β-CN and α-LA) and 0.68 (β-LG). Differences in the gold standard methods used to quantify milk protein fractions, as well as the characteristics of the studied population such as the breeds investigated, parities and stages of lactation represented, could have contributed to the difference in estimates (Schopen et al., 2009). Heritability estimates calculated by Bonfatti, Cecchinato, et al. (2011) in a population of Italian Simmental dairy cows, for the relative proportions of protein fractions expressed as a percentage of total CN (0.18 for γ-CN to 0.69 for β-CN) were greater than those for protein fraction contents expressed as g/L of milk (0.11 g/L for α-LA to 0.53 g/L for κ-CN). At present, little is known about the genetic parameters of FAA.

The usefulness of mid-infrared spectroscopy (MIRS) analysis of milk to predict milk composition and animal health traits on a large scale is now well established (De Marchi et al., 2014; Niero et al., 2016; Rutten et al., 2011; Soyeurt et al., 2011). Despite this, genetic parameters of MIRS-predicted protein fractions have been reported in only a limited number of studies (Bonfatti et al., 2017; Sanchez et al., 2017), while no estimates are available in the literature for FAA.

Therefore, the objective of the present study was to quantify the potential of breeding for milk protein fractions and FAA in dairy cows, but doing so by exploiting the MIRS predictions routinely available for individual cows.

2 | MATERIALS AND METHODS

2.1 | Milk samples and datasets

2.1.1 | Gold standard dataset and development of MIRS prediction models

Seven hundred and fifteen milk samples were collected from seven research dairy farms operated by the Teagasc Animal and Grassland Research and Innovation Centre (Moorepark, Fermoy, Co. Cork) between August 2013 and August 2014, inclusive. Individual milk proteins (α_{s1}-CN, α_{s2}-CN, β-CN, κ-CN, α-LA, and β-LG variant A and variant B) were determined in 557 of the 715 milk samples using reverse-phase high-performance liquid chromatography (HPLC). The FAA-related traits, namely total FAA, glutamic acid (or glutamate, Glu), glycine (Gly), lysine (Lys), arginine (Arg), aspartic acid (Asp), serine (Ser) and valine (Val), were quantified in all 715 milk samples using cation-exchange HPLC. Further details regarding gold standard analyses for both protein fractions and FAA are described in detail in McDermott et al. (2016). Each milk sample was also analysed using a mid-infrared spectrometer MilkoScan FT6000 located at the Teagasc Animal and Grassland Research and Innovation Centre (Moorepark, Fermoy Co. Cork, Ireland) for milk chemical composition quantification and spectra determination. Gold standard data were matched to their respective milk spectra, which were subsequently used to develop equations to predict individual and groups of milk proteins and FAA using partial least squares regression analysis (PROC PLS) in the SAS software v. 9.4 (SAS Institute Inc., Cary, NC). More details on the methodology used to develop the MIRS prediction models are outlined in McDermott et al. (2016). Overall, the accuracy of prediction, defined as the correlation coefficient between the reference and predicted values using (leave-one-out) cross-validation of protein fractions, ranged from 0.43 (β-LG variant A) to 0.76 (total whey and total LG). Cross-validation prediction accuracy for FAA varied from 0.51 (Ser) to 0.75 (Gly).

2.1.2 | Large dataset with MIRS-predicted traits

Between the years 2013 and 2015, inclusive, 171,279 milk samples from 17,353 lactations on 10,162 cows were

collected from 7 research herds and 69 commercial dairy herds (Visentin et al., 2017). Animals in the research herds were subjected to different experimental treatments, based mainly on different feeding and management strategies. All animals were milked twice daily. The commercial dairy farms ($n = 69$) were located in the south-west of Ireland (Munster Region) and represented typical Irish dairy farms.

Individual milk samples were sporadically taken and delivered to the laboratory of the Teagasc Animal and Grassland Research and Innovation Centre. All milk samples collected in both research and commercial herds were analysed using the same Fourier transform infrared spectrometer MilkoScan FT6000 (Foss Electronic A/S, Hillerød, Denmark) based at the Animal and Grassland Research Centre, Teagasc. Somatic cell count (SCC) was determined using a Fossomatic (Foss Electronic A/S, Hillerød, Denmark). Prediction equations developed by McDermott et al. (2016) based on the gold standard dataset were applied to these milk spectra in order to predict (a) groups of proteins (total CN, total whey, total LG), (b) individual proteins (α_{s1}-CN, α_{s2}-CN, β-CN, κ-CN, α-LA, β-LG A and β-LG B) and (c) FAA traits (total FAA, Glu, Gly, Lys, Arg, Asp, Ser and Val). Spectral variables and information on 24-h milk yield along with cow calving date and parity number were available for all milk samples. The average number of milk samples available per cow was equal to 72 for cows in the experimental farms (median = 57) and 5.15 for those in the commercial herds (median = 4).

In the final dataset, spectra with a Mahalanobis distance relative to the centroid of the gold standard dataset greater than three were discarded prior the application of MIRS prediction models (Williams, 2007). Furthermore, MIRS-predicted values of proteins and FAA, as well as 24-h milk yield and composition, greater than three standard deviations from the mean were set as missing. Moreover, the skewed-distributed values of Glu, Gly, Lys, Arg, Asp, Ser and Val were $\log_e$-transformed. Breed composition was available for each cow and only data from Holstein, Friesian, Jersey, Norwegian Red, Montbeliarde cows as well as their crosses sampled between 5 and 305 days in milk (DIM) and from parities ≤10 were retained. Parities greater than 6 were grouped together for analysis and only contemporary groups with a minimum of ten records were used. In particular, as both research and commercial farms were present, contemporary groups were generated either by combining the experimental treatment and the test date or by combining the commercial herd ID and the test date. Somatic cell count (SCC, n/μl) was converted to somatic cell score (SCS), as conventionally: $SCC = 3 + \log_2(SCC/100)$. Following all edits, 134,546 records from 16,166 lactations on 9,572 cows were available; of these, 40,260 were from commercial herds (i.e. 13,337 lactations and 8,260 cows).

2.2 | Variance components

Pedigree information was retrieved from the Irish Cattle Breeding Federation database and each animal was traced back (where available) up to at least four generations, resulting in a pedigree file of 33,949 and 8,210 individuals for the large dataset and gold standard dataset, respectively. Coefficients of general heterosis and recombination loss were quantified based on the equations presented by both VanRaden and Sanders (2003) and Visentin et al. (2017). Genetic, permanent environmental and residual (co)variances for protein fractions and FAA composition were quantified using linear mixed models in ASReml v. 3.0 (Gilmour et al., 2011).

2.2.1 | Large dataset

Fixed effects included in the linear mixed models were the contemporary group, the interaction between parity (1, 2, 3, 4, 5 and ≥6) and stage of lactation (10 classes with length of 30 d, from 5 to 305 DIM), milking time (AM or PM), proportion of cow breed (Friesian, Jersey, Norwegian Red, Montbeliarde and other; Holstein was not included to avoid linear dependencies), and general heterosis and recombination loss coefficients of the cow. The latter two effects were fitted as linear covariates. Random effects included in the models were the direct additive genetic effect of the animal and both a within- and an across-lactation cow permanent environmental effect.

Genetic and phenotypic (co)variances among the MIRS-predicted protein fractions were estimated using a series of trivariate (3 × 3) analyses that included two MIRS-predicted protein fractions as dependant variables, along with the total protein concentration (%). Covariances between MIRS-predicted FAA were estimated in a similar manner. Thus, two MIRS-predicted FAA together with the 24-h milk yield were the dependent variables. In matrix notation, the model was

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \\ \mathbf{b_3} \end{bmatrix} + \begin{bmatrix} \mathbf{Z_1} & 0 & 0 \\ 0 & \mathbf{Z_2} & 0 \\ 0 & 0 & \mathbf{Z_3} \end{bmatrix} \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \mathbf{a_3} \end{bmatrix} + \begin{bmatrix} \mathbf{W_1} & 0 & 0 \\ 0 & \mathbf{W_2} & 0 \\ 0 & 0 & \mathbf{W_3} \end{bmatrix} \begin{bmatrix} \mathbf{p_1} \\ \mathbf{p_2} \\ \mathbf{p_3} \end{bmatrix} + \begin{bmatrix} \mathbf{S_1} & 0 & 0 \\ 0 & \mathbf{S_2} & 0 \\ 0 & 0 & \mathbf{S_3} \end{bmatrix} \begin{bmatrix} \mathbf{c_1} \\ \mathbf{c_2} \\ \mathbf{c_3} \end{bmatrix} + \begin{bmatrix} \mathbf{e_1} \\ \mathbf{e_2} \\ \mathbf{e_3} \end{bmatrix},$$

where, for traits 1, 2 and 3, $\boldsymbol{y}$ is the vector of phenotypic observations of the dependent variable; $\boldsymbol{b}$ is the vector of fixed effects; $\boldsymbol{a}$ is the vector of random additive genetic effect; $\boldsymbol{p}$ is the vector of permanent environmental effect of the cow within a given lactation; $\boldsymbol{c}$ is the vector of permanent environmental effect of the cow across lactations; $\boldsymbol{e}$ is the vector of random residuals. Incidence matrices relating the corresponding effects to the dependent variable were **X**, **Z**, **W** and **S**, respectively. The following structure was assumed for the (co)variance between the random effects:

$$\text{var}\begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \\ \boldsymbol{p_1} \\ \boldsymbol{p_2} \\ \boldsymbol{p_3} \\ \boldsymbol{c_1} \\ \boldsymbol{c_2} \\ \boldsymbol{c_3} \\ \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \boldsymbol{e_3} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma^2_{a_1} & \mathbf{A}\sigma_{12} & \mathbf{A}\sigma_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{A}\sigma_{12} & \mathbf{A}\sigma^2_{a_2} & \mathbf{A}\sigma_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{A}\sigma_{13} & \mathbf{A}\sigma_{23} & \mathbf{A}\sigma^2_{a_3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\sigma^2_{p_1} & \mathbf{I}\sigma_{12} & \mathbf{I}\sigma_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\sigma_{12} & \mathbf{I}\sigma^2_{p_2} & \mathbf{I}\sigma_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\sigma_{13} & \mathbf{I}\sigma_{23} & \mathbf{I}\sigma^2_{p_3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma^2_{c_1} & \mathbf{I}\sigma_{12} & \mathbf{I}\sigma_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma_{12} & \mathbf{I}\sigma^2_{c_2} & \mathbf{I}\sigma_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma_{13} & \mathbf{I}\sigma_{23} & \mathbf{I}\sigma^2_{c_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma^2_{e_1} & \mathbf{I}\sigma_{12} & \mathbf{I}\sigma_{13} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma_{12} & \mathbf{I}\sigma^2_{e_2} & \mathbf{I}\sigma_{23} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma_{13} & \mathbf{I}\sigma_{23} & \mathbf{I}\sigma^2_{e_3} \end{bmatrix}$$

where for the traits 1, 2 and 3: σ_{a12}, σ_{a13} and σ_{a23} are the additive genetic covariances, σ_{p12}, σ_{p13} and σ_{p23} are the within-lactation permanent environmental covariances; σ_{c12}, σ_{c13} and σ_{c23} are the across-lactation permanent environmental covariances; σ_{e12}, σ_{e13} and σ_{e23} are the residual covariances; σ^2_a, σ^2_p, σ^2_c and σ^2_e indicate the additive genetic variance, the within- and across-lactation permanent environmental variance and the residual variance; **A** was the pedigree-based relationship matrix of size equal to number of individuals ($n = 33{,}949$), while the **I** denoted identity matrixes of appropriate order, that is equal to number of lactations, number of cows and number of records for σ^2_p, σ^2_c and σ^2_e, respectively.

Genetic correlations among the MIRS-predicted protein fractions were adjusted for their respective genetic correlation with total protein concentration as follows:

$$\boldsymbol{r}_{xy,z} = \frac{\boldsymbol{r}_{xy} - \boldsymbol{r}_{xz}\boldsymbol{r}_{yz}}{\sqrt{\left(\mathbf{1} - \boldsymbol{r}^2_{xz}\right)\left(\mathbf{1} - \boldsymbol{r}^2_{yz}\right)}} \quad (1)$$

where $\boldsymbol{r}_{xy,z}$ is the partial correlation coefficient between traits x and y independent of trait z, $\boldsymbol{r}_{xy}$ represents the correlation coefficient between traits x and y, $\boldsymbol{r}_{xz}$ is the correlation coefficient between traits x and z, and $\boldsymbol{r}_{yz}$ is the correlation coefficient between traits y and z. In a similar manner, genetic correlations among the MIRS-predicted FAA were adjusted for their respective genetic correlation with 24-h milk yield.

The phenotypic and genetic variance adjusted for their respective correlation (phenotypic or genetic) with protein concentration (protein fractions) or 24-h milk yield (FAA) were calculated for each trait afterwards, as:

$$\boldsymbol{\sigma}^2_{\boldsymbol{i}} = \boldsymbol{\sigma}^2_{\boldsymbol{i}_{\boldsymbol{unadjusted}}} * \left(\mathbf{1} - \boldsymbol{r}^2_{\boldsymbol{xy}}\right) \quad (2)$$

where $\boldsymbol{\sigma}^2_{\boldsymbol{i}_{\boldsymbol{unadjusted}}}$ is the variance (phenotypic or genetic) estimated by the REML algorithm for each trait, and $\boldsymbol{r}_{\boldsymbol{xy}}$ is the correlation (phenotypic or genetic) between trait x (protein fraction or FAA) and trait y (protein concentration or 24-h milk yield). Heritability was the ratio of the additive genetic variance (adjusted or unadjusted) to the phenotypic variance calculated by summing up of the additive genetic (adjusted or unadjusted), the within- and the across-lactation permanent environmental and the residual variances.

2.2.2 | Gold standard dataset

The model used for genetic parameter estimation in the gold standard dataset differed from that used in the large dataset. In particular, the contemporary group was replaced by the fixed effect of herd ($n = 7$) and the two random permanent environmental effects were not considered, since only one record per cow (in a single lactation) was present. Heritability was calculated as the ratio between the additive genetic variance to the sum of the additive genetic and residual variance. The genetic correlation between the gold standard trait and its MIRS prediction was estimated using (co)variance components obtained from the bivariate analyses.

TABLE 1 Number of records (n), mean, genetic standard deviation (σ_g), heritability (standard error) and coefficient of genetic variation (CV_g) for milk yield and composition, gold standard protein fractions and free amino acids (FAA) and the genetic correlation (r_g) between gold standard versus MIRS-predicted FFA and protein fractions

Trait	*N*	Mean	σ_g	Heritability	CV_g, %	r_g
24-h Milk yield, kg/d	497	22.08	2.64	0.16 (0.10)	11.96	–
Protein, %	585	3.74	0.22	0.45 (0.12)	5.88	–
Casein, %	585	2.82	0.19	0.51 (0.12)	6.74	–
Fat, %	584	4.71	0.53	0.30 (0.10)	11.25	–
Lactose, %	584	4.61	0.07	0.15 (0.10)	1.52	–
SCS, $\log_2$ units	252	2.48	0.23	0.03 (0.12)	9.27	–
Protein fractions, g/L						
Total casein	441	36.92	2.47	0.20 (0.12)	6.69	0.69 (0.18)
α_{s1}-casein	442	14.18	1.19	0.22 (0.12)	8.39	0.71 (0.15)
α_{s2}-casein	442	3.67	0.51	0.38 (0.14)	13.90	0.82 (0.15)
β-casein	441	12.95	0.47	0.05 (0.09)	3.63	0.63 (0.15)
κ-casein	440	6.01	0.60	0.21 (0.14)	9.98	0.64 (0.24)
Total whey	437	6.18	0.97	0.41 (0.15)	15.70	0.73 (0.15)
α-Lactalbumin	441	1.13	0.27	0.24 (0.16)	23.89	0.45 (0.33)
Total lactoglobulin	438	5.04	0.84	0.38 (0.15)	16.67	0.76 (0.13)
β-Lactoglobulin A	441	2.59	0.44	0.15 (0.12)	16.99	0.68 (0.17)
β-Lactoglobulin B	400	2.74	1.29	0.69 (0.16)	47.08	0.77 (0.11)
FAA, μg/mL[a]						
Total FAA	568	73.28	10.48	0.19 (0.11)	14.30	0.53 (0.12)
Glutamic acid	568	27.41	0.29	0.41 (0.13)	5.46	0.83 (0.11)
Glycine	565	5.64	0.15	0.11 (0.09)	4.02	0.78 (0.32)
Lysine	565	3.41	0.29	0.23 (0.12)	8.98	0.38 (0.35)
Arginine	507	3.15	0.20	0.27 (0.13)	6.36	0.51 (0.26)
Aspartic acid	483	2.12	0.29	0.24 (0.13)	10.54	0.52 (0.28)
Serine	488	1.22	0.30	0.37 (0.15)	13.64	0.38 (0.11)
Valine	524	1.39	0.24	0.15 (0.11)	10.32	0.46 (0.17)

[a]Traits were log-transformed before analysis.

3 | RESULTS

3.1 | Mean and variation

The mean of the gold standard measurement and the mean of the respective MIRS-predicted milk protein fractions were comparable. In fact, the mean value of the gold standard total CN and the mean of MIRS-predicted total CN were 36.92 g/L and 36.16 g/L, respectively (Tables 1 and 2, respectively). Similarly, the mean of the gold standard FAA was similar to the mean of their relative MIRS predictions (Tables 1 and 2). The genetic standard deviation of protein fractions ranged from 0.44 g/L (β-LG A) to 2.47 g/L (total CN; Table 1) for the gold standard data. The MIRS predictions ranged from 0.04 g/L (α-LA) to 1.85 g/L (total CN) for the unadjusted analyses (Table 2). The genetic standard deviation of the gold standard and the respective MIRS-predicted FAA variables were also similar, with the exception of total FAA. In fact, gold standard total FAA had a genetic standard deviation of 10.48 μg/mL, which was approximately twice that for the MIRS-predicted values (5.05 μg/ml). The coefficient of genetic variation differed among traits ranging from 3.63 (β-CN) to 47.08 (β-LG B) for the gold standard protein fractions and from 4.02 (Gly) to 14.30 (total FAA) for the gold standard FAA (Table 1).

3.2 | Heritability and repeatability estimates

Heritability estimates of milk yield and composition from the gold standard dataset (Table 1), that is the one containing just the reference values of milk protein fractions

and FAA, were similar to those from the large dataset with research and commercial cow data (Table 2). The heritability estimates did, however, differ in terms of standard errors, which was due to the fewer records in the gold standard dataset. Heritability estimates for the gold standard protein fractions, albeit characterized by large standard errors, ranged from 0.05 (β-CN) to 0.69 (β-LG B; Table 1). Heritability estimates for the MIRS-predicted protein fractions (Table 2) were higher than those estimated for the gold standard protein fractions (Table 1), although the range of overall heritability estimates for the MIRS-predicted protein fractions was smaller (0.19 for α-LA to 0.46 for β-LG A; Table 2) than the range for the respective gold standard measures (0.05 for β-CN to 0.69 for β-LG B; Table 1). Similar to protein fractions, heritability estimates for MIRS-predicted FAA had a narrower range (0.08 for Gly to 0.29 for Asp and Glu; Table 2) than the respective gold standard range (0.11 for Gly to 0.41 for Glu; Table 1). Repeatability estimates for MIRS-predicted protein fractions and MIRS-predicted FAA were moderate (0.24 for α-LA to 0.55 for total CN; 0.15 for Gly to 0.40 for Glu; Table 2). The estimated genetic standard deviation for each protein fraction trait, genetically independent of protein concentration, was less than the respective unadjusted measure and this was also reflected in lower heritability estimates; only a small reduction in genetic

TABLE 2 Number of records (*n*), mean, genetic standard deviation (σ_g) with the unadjusted heritability (standard error) and repeatability (standard error) for milk yield and composition, MIRS-predicted protein fractions and MIRS-predicted free amino acids (FAA). Heritability and genetic standard deviation after adjustment[a] are also reported

			Unadjusted			Adjusted	
Trait	***n***	**Mean**	σ_g	**Heritability**	**Repeatability**	σ_g	**Heritability**
24-h Milk yield, kg/day	133,999	20.54	1.53	0.17 (0.02)	0.54 (0.01)	–	–
Protein, %	125,045	3.72	0.18	0.44 (0.02)	0.56 (0.01)	–	–
Casein, %	125,484	2.82	0.15	0.39 (0.02)	0.52 (0.01)	–	–
Fat, %	111,055	4.57	0.36	0.30 (0.01)	0.32 (0.01)	–	–
Lactose, %	110,954	4.77	0.09	0.36 (0.02)	0.49 (0.01)	–	–
SCS, units	63,505	2.50	0.26	0.05 (0.01)	0.47 (0.01)	–	–
Protein, g/L							
Total casein	134,395	36.16	1.85	0.42 (0.02)	0.55 (0.01)	0.27	0.11
α_{s1}-casein	134,409	13.70	0.79	0.44 (0.02)	0.54 (0.01)	0.59	0.36
α_{s2}-casein	134,482	3.66	0.18	0.36 (0.02)	0.44 (0.01)	0.14	0.28
β-casein	134,235	12.96	0.54	0.45 (0.02)	0.48 (0.01)	0.41	0.30
κ-casein	134,463	6.03	0.34	0.36 (0.02)	0.46 (0.01)	0.25	0.27
Total whey	134,383	6.13	0.52	0.37 (0.02)	0.45 (0.01)	0.50	0.36
α-lactalbumin	134,106	1.11	0.04	0.19 (0.01)	0.24 (0.01)	0.03	0.17
Total lactoglobulin	134,236	5.09	0.51	0.39 (0.02)	0.48 (0.01)	0.48	0.37
β-lactoglobulin A	134,489	2.39	0.19	0.46 (0.02)	0.54 (0.01)	0.16	0.42
β-lactoglobulin B	128,391	2.61	0.56	0.42 (0.02)	0.52 (0.01)	0.55	0.42
FAA, μg/ml[b]							
Total FAA	134,546	53.60	5.05	0.24 (0.01)	0.36 (0.01)	5.05	0.24
Glutamic acid	134,426	30.93	4.40	0.29 (0.02)	0.40 (0.01)	4.38	0.29
Glycine	133,650	8.09	0.66	0.08 (0.01)	0.15 (0.01)	0.66	0.08
Lysine	134,105	4.75	0.53	0.17 (0.01)	0.23 (0.01)	0.52	0.16
Arginine	134,425	3.38	0.23	0.15 (0.01)	0.24 (0.01)	0.23	0.15
Aspartic acid	134,433	2.73	0.39	0.29 (0.02)	0.38 (0.01)	0.39	0.28
Serine	112,918	2.74	0.22	0.15 (0.01)	0.20 (0.01)	0.22	0.15
Valine	133,957	1.52	0.16	0.13 (0.01)	0.25 (0.01)	0.16	0.12

[a]Protein factions and FAA were adjusted for protein concentration and 24-h milk yield, respectively. Heritability considered the genetic variance calculated as per Equation 2.

[b]Traits were log-transformed before analysis.

standard deviation and heritability was observed for the whey protein fractions (i.e. α-LA, β-LG A and β-LG B; Table 2). There was also little impact on the heritability estimates for FAA when adjusted for differences in the genetic merit of 24-h milk yield (Table 2).

3.3 | Genetic correlations

Moderate-to-strong genetic correlations existed between the gold standard protein fractions and their respective MIRS-predicted protein fractions, ranging from 0.45 (α-LA) to 0.82 (α_{s1}-CN; Table 1). The genetic correlation between the gold standard Glu and MIRS-predicted Glu was strong (0.83), similar to the correlation between the gold standard and MIRS-predicted Gly (0.78; Table 1). Protein fractions were negatively correlated with 24-h milk yield, but they were positively associated with protein and CN concentration (Table 3). Protein fractions were weakly correlated with both lactose concentration and SCS. With the exception of both β-CN and α-LA, all protein fractions were negatively correlated with lactose concentration. Individual FAA were weakly to moderately genetically correlated with 24-h milk yield, protein, CN, fat concentration, lactose concentration and SCS (Table 4). Correlations between MIRS-predicted FAA and SCS in general did not differ from zero with the exception of the genetic correlation between MIRS-predicted Val and SCS (0.39 ± 0.09; Table 4).

Genetic correlations among the MIRS-predicted protein fractions were weak to strong (Table 5). An almost unity genetic correlation (0.99) existed between protein concentration and total CN, between total CN and α_{s1}-CN, and between total LG and total whey. The genetic correlation between total LG and total whey did not change after adjusting for their respective genetic correlation with protein concentration, albeit most of the correlations among the protein fractions weakened when calculated genetically independent of protein concentration. For example, the strong positive genetic correlation that existed between β-CN and α_{s2}-CN (0.89) became negative (−0.05) once adjusted for genetic merit for protein concentration (Table 5). Genetic correlations among the MIRS-predicted FAA were weak to strong and ranged from −0.44 (Asp and Lys) to 0.97

TABLE 3 Genetic correlations (standard errors) of MIRS-predicted protein fractions with 24-h milk yield, concentration of protein, casein, fat and lactose, as well as somatic cell score (SCS)

Trait	24-h milk yield	Protein	Casein	Fat	Lactose	SCS
Total casein	−0.57 (0.04)	0.99 (0.001)	0.99 (0.002)	0.72 (0.02)	−0.09 (0.04)	0.11 (0.08)
α_{s1}-casein	−0.58 (0.04)	0.99 (0.001)	0.98 (0.002)	0.74 (0.03)	−0.12 (0.04)	0.10 (0.08)
α_{s2}-casein	−0.59 (0.04)	0.95 (0.01)	0.96 (0.01)	0.79 (0.03)	−0.05 (0.04)	0.10 (0.08)
β-casein	−0.57 (0.04)	0.94 (0.01)	0.81 (0.03)	0.69 (0.02)	0.06 (0.04)	0.07 (0.08)
κ-casein	−0.50 (0.04)	0.97 (0.01)	0.94 (0.01)	0.62 (0.03)	−0.20 (0.04)	0.10 (0.08)
Total whey	−0.30 (0.05)	0.49 (0.03)	0.46 (0.03)	0.39 (0.03)	−0.07 (0.04)	0.09 (0.08)
α-lactalbumin	−0.45 (0.05)	0.55 (0.03)	0.59 (0.03)	0.68 (0.03)	0.35 (0.04)	0.11 (0.08)
Total lactoglobulin	−0.29 (0.04)	0.49 (0.03)	0.46 (0.03)	0.38 (0.03)	−0.12 (0.04)	0.08 (0.08)
β-lactoglobulin A	−0.56 (0.04)	0.86 (0.01)	0.87 (0.01)	0.75 (0.02)	−0.09 (0.04)	0.15 (0.08)
β-lactoglobulin B	−0.08 (0.05)	0.18 (0.03)	0.15 (0.04)	0.11 (0.04)	−0.11 (0.04)	0.07 (0.08)

TABLE 4 Genetic correlations (standard errors) between log-transformed MIRS-predicted free amino acids (FAA) with 24-h milk yield, concentration of protein, casein, fat and lactose, as well as somatic cell score (SCS)

Trait	24-h Milk yield	Protein	Casein	Fat	Lactose	SCS
Total FAA	0.01 (0.06)	−0.10 (0.04)	−0.09 (0.04)	−0.07 (0.04)	0.29 (0.04)	0.07 (0.09)
Glutamic acid	0.10 (0.05)	−0.18 (0.04)	−0.17 (0.04)	−0.15 (0.04)	0.33 (0.04)	0.08 (0.09)
Glycine	0.05 (0.06)	−0.19 (0.04)	−0.17 (0.04)	−0.24 (0.04)	0.28 (0.05)	−0.15 (0.10)
Lysine	−0.33 (0.05)	0.52 (0.03)	0.51 (0.03)	0.51 (0.03)	−0.16 (0.04)	0.07 (0.09)
Arginine	−0.18 (0.06)	0.21 (0.04)	0.17 (0.04)	0.40 (0.04)	−0.33 (0.04)	0.13 (0.12)
Aspartic acid	0.14 (0.05)	−0.19 (0.04)	−0.20 (0.04)	−0.16 (0.04)	0.25 (0.04)	−0.01 (0.09)
Serine	−0.13 (0.06)	−0.15 (0.04)	−0.11 (0.04)	0.24 (0.04)	0.26 (0.04)	0.18 (0.09)
Valine	−0.21 (0.06)	0.32 (0.04)	0.33 (0.04)	0.25 (0.04)	−0.16 (0.04)	0.39 (0.09)

TABLE 5 Genetic correlations (standard errors) among the concentration of MIRS-predicted protein fractions[a] (above diagonal); estimates adjusted for milk protein concentration are reported below diagonal[b]

Trait	Total CN	α_{s1}-CN	α_{s2}-CN	β-CN	κ-CN	Total-W	α-LA	Total LG	β-LG A	β-Lg B
Total CN	–	0.99 (0.00)	0.92 (0.01)	0.97 (0.002)	0.96 (0.002)	0.47 (0.03)	0.60 (0.03)	0.47 (0.03)	0.85 (0.01)	0.17 (0.03)
α_{s1}-CN	0.91	–	0.92 (0.01)	0.96 (0.002)	0.95 (0.002)	0.51 (0.03)	0.61 (0.03)	0.50 (0.03)	0.81 (0.01)	0.18 (0.03)
α_{s2}-CN	−0.24	−0.24	–	0.89 (0.01)	0.90 (0.01)	0.42 (0.03)	0.56 (0.03)	0.41 (0.03)	0.88 (0.01)	0.10 (0.03)
β-CN	0.79	0.63	−0.05	–	0.80 (0.02)	0.48 (0.03)	0.70 (0.02)	0.46 (0.03)	0.86 (0.01)	0.13 (0.04)
κ-CN	0.07	−0.09	−0.12	−0.22	–	0.47 (0.03)	0.38 (0.04)	0.47 (0.03)	0.72 (0.02)	0.22 (0.03)
Total-W	0.09	0.16	−0.17	0.05	−0.01	–	0.47 (0.03)	0.99 (0.00)	0.35 (0.03)	0.91 (0.01)
α-LA	0.53	0.56	0.14	0.63	−0.37	0.25	–	0.44 (0.03)	0.72 (0.02)	0.17 (0.04)
Total LG	0.04	0.12	−0.17	−0.02	0.01	0.99	0.24	–	0.34 (0.03)	0.92 (0.01)
β-LG A	0.11	0.25	0.44	0.29	−0.79	−0.01	0.58	−0.18	–	−0.03 (0.03)
β-LG B	−0.05	−0.01	−0.20	−0.13	0.18	0.96	0.09	0.90	−0.36	–

[a]Total casein (Total CN), α_{s1}-casein (α_{s1}-CN), α_{s2}-casein (α_{s2}-CN), β-casein (β-Casein), κ-casein (κ-CN), total whey (Total-W), α-lactalbumin (α-LA), total lactoglobulin (Total LG) and β-lactoglobulin A (β-LG A) and β-lactoglobulin B (β-LG B).

[b]Adjusted genetic correlations calculated as per Equation 1.

(Glu and Total FAA; Table 6). Adjusting the correlations for the genetic merit of 24-h milk yield did not greatly affect the correlation estimates (Table 6). For example, the unadjusted correlation between Arg and Val was 0.63 and changed to 0.62 when adjusted for the genetic merit of 24-h milk yield.

4 | DISCUSSION

The present study aimed to quantify the extent of genetic variability in detailed milk protein and FAA composition predicted by MIRS in individual cow milk samples. Nowadays, global milk recording programmes rely on infrared tools for determination of fat and protein content of samples intended to herd testing and genetic evaluations (De Marchi et al., 2014; Franzoi et al., 2021). Currently, protein fractions and FAA are not either determined by MIRS on a routine basis or used for within-country genetic evaluations. For both protein fractions and FAA, the genetic correlations estimated in the present study between both the gold standard and MIRS prediction were moderate to strong, demonstrating that the MIRS predictions are genetically very similar to their corresponding gold standard measures. All traits were heritable and exhibited considerable and exploitable genetic variation; therefore, MIRS is a viable method to collect a large quantity of accurate data for use in genetic evaluations with the goal of improving the added value of granular milk.

4.1 | Response to genetic selection

Response to genetic selection is a function of the extent of genetic variability, the accuracy of selection, selection intensity and generation interval (Rendel & Robertson, 1950). Therefore, to achieve genetic gain, the trait must exhibit genetic variation and sufficient phenotypic or genomic information on the trait must be available to ensure a high accuracy of selection (Costa et al., 2020; Simm et al., 2020). Considerable genetic variation clearly exists for all milk quality traits examined in the present study with the traits being low to moderately heritable. Therefore, this suggests that there is potential to alter both the protein composition and FAA composition in bovine milk using selective breeding. The greater the heritability, the fewer progeny records required to achieve a high accuracy of selection. In other words, the greater the heritability, the smaller the reference population required to develop genomic predictions (Daetwyler et al., 2008).

Heritability estimates of all the MIRS-predicted protein fractions in the present study (0.19 to 0.46; Table 2) were higher than the heritability estimate for 24-h milk yield (0.17; Table 2). Moreover, the heritability of 24-h

TABLE 6 Genetic correlations (standard errors) among the concentration of MIRS-predicted free amino acids (FAA; above diagonal); estimates adjusted for 24-h milk yield are reported below diagonal[a]

Trait	Total FAA	Glutamic acid	Glycine	Lysine	Arginine	Aspartic acid	Serine	Valine
Total FAA	–	0.97 (0.002)	0.56 (0.04)	−0.20 (0.04)	0.31 (0.04)	0.88 (0.01)	0.60 (0.03)	0.49 (0.04)
Glutamic acid	0.99	–	0.53 (0.04)	−0.35 (0.04)	0.19 (0.04)	0.91 (0.01)	0.57 (0.03)	0.42 (0.04)
Glycine	0.56	0.53	–	−0.36 (0.04)	−0.27 (0.05)	0.57 (0.03)	0.37 (0.04)	−0.09 (0.05)
Lysine	−0.20	−0.31	−0.36	–	0.56 (0.03)	−0.44 (0.04)	0.06 (0.04)	0.49 (0.04)
Arginine	0.24	0.21	−0.22	0.53	–	0.13 (0.04)	0.32 (0.04)	0.63 (0.03)
Aspartic acid	0.90	0.90	0.57	−0.42	0.16	–	0.32 (0.04)	0.18 (0.04)
Serine	0.38	0.59	0.39	0.02	0.30	0.30	–	0.49 (0.04)
Valine	0.49	0.45	−0.09	0.45	0.62	0.19	0.27	–

[a]Adjusted genetic correlations calculated as per Equation 1.

milk yield and protein concentration (0.44; Table 2) in the present study were in the range of the heritability estimates for the gold standard protein fractions (0.05 to 0.69; Table 1), but only the heritability of 24-h milk yield was in the range of the heritability of gold standard FAA (0.11 to 0.41; Table 1) and of the MIRS-predicted FAA (0.08 to 0.29; Table 2). Previous heritability estimates of protein fractions in bovine milk have been documented to be moderate to high ranging from 0.25 for β-CN to 0.80 for total LG (Huang et al., 2012; Schopen et al., 2009), although they differed across studies. Recent heritability estimates for gold standard protein fractions in milk from Simmental cows ranged from 0.18 (κ-CN) to 0.68 (α_{s1}-CN; Bonfatti, Cecchinato, et al. (2011) and in milk from both Holstein and Holstein-Jersey crosses the range was between 0.33 (α_{s1}-CN, β-CN and α-LA) and 0.68 (β-LG; Huang et al., 2012).

In the present study, all protein fractions were positively genetically correlated with protein concentration, contradicting previous findings by Schopen et al. (2009) who estimated genetic parameters for the same traits but determined with capillary zone electrophoresis in milk from 1,940 primiparous cows. By way of example, α_{s1}-CN, β-CN, α-LA and total whey were negatively genetically correlated with the protein concentration (Schopen et al., 2009), but those estimates were characterized by quite large standard errors and sometimes did not differ from zero. Genetic correlations among the protein fractions obtained in the present study were also different from those of Schopen et al. (2009) and Bonfatti,Cecchinato, et al. (2011) in dairy cows. Again, different methods were used to determine milk protein fractions across studies, as well as the characteristics of the studied population such as breeds, parities and seasons represented all of which could have contributed to the difference in correlations. Indeed, Bonfatti, Cecchinato, et al. (2011) used data from protein fractions determined by high-performance liquid chromatography from 2,167 Simmental cows. In the present study, protein fractions were predicted by MIRS on up to 134,100 milk samples from Holstein, Friesian, Jersey, Norwegian Red and Montbeliarde cows as well as their crosses across a wide range of parities and lactation stages.

Many current national breeding programmes indirectly select for protein concentration through a negative weighting on milk yield concurrent with a positive weighting on protein yield (Cole & VanRaden, 2018). The coefficients of genetic variation for milk yield and protein concentration in the present study were 7.45% and 4.84%, respectively. Such values were similar or even smaller than coefficient of genetic variation for the gold standard total casein and total whey (6.69% and 15.70%, respectively), as well as lower than the coefficient of genetic variation for FAA (14.30%). Favourable trends in genetic gain for milk yield in dairy cows are well documented (Berry, 2008; Berry et al., 2014; Norman & Powell, 1999), with lactation milk yield per cow doubling over the past 40 years (Oltenacu & Broom, 2010). Given the coefficient of variation for milk yield, protein fractions and FAA, a similar genetic gain may be reached for protein fractions and FAA if these traits were included in a selection index with a high accuracy of selection. Satisfactory selection accuracy could be achieved by generating these phenotypes on a representative pool of cows for evaluation purpose as could be the case for a reference population for genomic selection.

However, a relevant question is the expected response to selection when including these traits, detailed proteins or FAA, in the official selection index and among the "true" breeding objectives. The official selection index in Ireland already includes a variety of features such as milk yield and protein concentration. One approach to estimate

such a benefit in response is to derive the coefficient of genetic variation of the traits after adjustment for the genetic merit of (1) protein concentration in the case of protein fractions (2) or milk yield in the case of FAA.

In the present study, the near unity correlation between protein concentration and total CN (0.99) and the resulting small coefficient of genetic variation for total CN after adjustment for the genetic merit of protein concentration (0.75%) suggest that directly including CN in the selection index may result in little additional benefit. Similarly, the genetic correlations of protein concentration with α_{s1}-CN, β-CN and κ-CN were strong (0.99, 0.94 and 0.97, respectively) and the coefficient of genetic variation for α_{s1}-CN, β-CN and κ-CN after adjustment was just 4.30%, 3.16% and 4.14%, respectively. Hence, a very high selection pressure would be required for α_{s1}-CN, β-CN and κ-CN, that means reducing both the selection pressure on the other traits present in the index and their genetic gain. The correlation between protein concentration and α-LA was only 0.55 and the corresponding coefficient of genetic variation after adjustment for the genetic merit was 2.70%. Therefore, even if genetic gain would be slow, it could be advantageous to include α-LA per se in the selection index. High concentrations of α-LA are indeed desirable for infant formula and dairy powders manufacturing (Lien, 2003). In some circumstances, the addition of valuable proteins like α-LA would make sense in dairy cattle selection index, for example in those countries with a wide dairy product portfolio.

The weak genetic correlations between FAA and 24-h milk yield as well as between FAA and protein concentration signify that existing selection indexes that consider milk yield and protein concentration are not fully exploiting the potential to genetically improve the desirable protein fractions and FAA. The genetic correlation close to 1 between total FAA and Glu (0.99) suggests that one should expect little benefit when including both of them in the selection index. The lack of very strong genetic correlations among the other FAA (i.e. Gly, Lys, Arg, Asp, Ser and Val) plus the existence of a moderate to large coefficient of genetic variation for these traits (Table 1) suggests that direct selection can be pursued simultaneously on all these traits together with traditional traits, like milk yield and protein concentration.

As proposed by Henchion et al. (2016), a potential approach would be to define the emphasis to be placed to a milk quality sub-index, for example to the milk α-LA in this study, that finds the consensus of different stakeholders (i.e. breed associations and companies, milk producers and processors, farm advisors and researchers). In Henchion et al. (2016), all the actors of the dairy chain were questioned to gauge their opinion about detailed milk quality traits within an overall main breeding goal. Based on their findings, the relative importance on a milk quality sub-index was suggested to range from 6% to 10%. Nevertheless, the biggest concern raised by the interviewed stakeholders was their limited knowledge on the fine quality characteristics, followed by the scarce consumer demand for enhanced-quality milk and dairy foodstuffs. In addition, some categories of consumers are not familiar with nutritional concepts and health benefits associated with consumption of animal products, including milk and dairy. Opinion and perception of different actors must be taken into account to properly quantify the selection index economic weights going in favour of desirable milk traits.

However, given a scenario where certain specific milk traits (e.g. protein fractions) are considered in the payment system, sub-indexes can be useful to farmers and facilitate more bespoke breeding decisions based on the payment system. A good example is the Trentingrana cheese Consortium (Italy) that incentivizes farmers to produce milk with more favourable cheese-making characteristics by applying specific rewards (Penasa et al., 2016).

4.2 | Practical implications

Routine access to vast quantities of phenotypic data for protein fractions and FAA is imperative to achieve a high accuracy of selection and thus genetic gain. In addition, the actual response achievable is a function of how closely the selection index trait (i.e. MIR-predicted phenotype) reflects the goal trait (i.e the true measure of the trait). This is a function of the genetic correlation between the MIR-predicted value and its true value. In cattle, the genetic correlation between the reference trait and its MIRS prediction is often less than unity. Therefore, while selection using proxies may still achieve progress in a given direction, the actual response is likely to be less than optimum relative to if the gold standard was actually measured (assuming no difference in heritability).

Using the deterministic equation provided by Goddard (2009) to evaluate the accuracy of genomic evaluations with different sized reference populations, the number of genotyped individuals in the reference population necessary to achieve an animal EBV accuracy of 0.70 is approximately equal to 6,000 for predicted α_{s1}-CN. On the contrary, a larger reference population size is needed for predicted α-LA (≈14,000), Glu (≈9,000) and Gly (≈33,000) to achieve the same accuracy of genomic evaluations. For this calculation, the effective population size was set at 75 as reported by McParland et al. (2007) for Irish Holsteins. Mid-infrared spectroscopy is an efficient method commonly used by milk recording organizations worldwide to measure milk gross composiiton.

Other than this, the ability of MIRS to predict individual protein fractions and FAA with a reasonable accuracy has been previously documented (McDermott et al., 2016). Thus, MIRS is the most popular, rapid and cost-effective method for generating substantial quantities of phenotypes related to milk detailed composition, health traits and more (Benedet et al., 2020; Costa et al., 2021; Grelet et al., 2018).

Of potential interest to many processors are the similarities, or lack thereof, between human and bovine milk and the potential strategies to make them more similar, ideally at low cost and preferably even within the farm gate. Human and bovine milk differ in their protein profile; human milk has a whey to casein ratio of approximately 60:40, while bovine milk has a whey to casein ratio of approximately 20:80. While human milk does not contain β-LG, it is present in large amount in bovine milk; on the contrary, α-LA is the dominant protein in human milk but is relatively low in concentration in bovine milk (Lien, 2003). Therefore, it may be advantageous for infant formula producers to select bovine milk with a higher concentration of α-LA and a lower concentration of β-LG. The amino acid profile in human and bovine milk also differs (Chuang et al., 2005): human milk contains more Glu, Lys, Arg, Asp, Ser and Val than bovine milk (Ghadimi & Pecora, 1963).

Processors also aim to maximize the efficiency of transformation of the milk they purchase into saleable high-quality and nutritious products. A high concentration of CN to total protein is essential for the transfer of proteins from milk to cheese with high concentrations of α_{s1}-CN, β-CN, κ-CN and β-LG B being associated with greater cheese yield (Wedholm et al., 2006). The heritability estimates of the MIRS-predicted CN fractions ranged from 0.36 (κ-CN and α_{s2}-CN) to 0.45 (β-CN) and the genetic standard deviation calculated for gold standard CN fractions in the present study (0.79 g/L for α_{s1}-CN) were comparable to results obtained by Graml and Pirchner (2003), i.e. 0.94 g/L for α_{s1}-CN). Variability in predicted CN fractions also existed in the present study, as the genetic standard deviation ranged from 0.18 g/L for α_{s2}-CN to 0.79 g/L for α_{s1}-CN. This was true eve after adjusting for differences in protein concentration, indicating an opportunity to improve milk-specific protein fractions through selective breeding.

5 | CONCLUSIONS

Gold standard and MIRS-predicted traits were heritable and exhibited considerable exploitable genetic variation in their own right. Nonetheless, variability in some cases was eroded once expressed as relative to total protein concentration, which is present in most dairy cattle breeding goal worldwide. Traditional and genomic evaluations of detailed protein composition make sense if MIRS prediction equations will be implemented for phenotyping and whenever dairy chain stakeholders will conclude that certain protein fractions deserve inclusion in the breeding objective.

ACKNOWLEDGEMENT

Laboratory support by Bernard Corrigan, Jim Flynn, Owen Kenny and Paula O'Connor is gratefully acknowledged. Open Access Funding provided by Universita degli Studi di Padova within the CRUI-CARE Agreement.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Original data used in this study are available from the corresponding author upon reasonable request.

ORCID

Giulio Visentin https://orcid.org/0000-0003-0869-5516
Donagh P. Berry https://orcid.org/0000-0003-4349-1447
Angela Costa https://orcid.org/0000-0001-5353-8988
Massimo De Marchi https://orcid.org/0000-0001-7814-2525
Sinead McParland https://orcid.org/0000-0003-3320-2330

REFERENCES

Auldist, M. J., Johnston, K. A., White, N. J., Fitzsimons, W. P., & Boland, M. J. (2004). A comparison of the composition, coagulation characteristics and cheesemaking capacity of milk from Friesian and Jersey dairy cows. *Journal of Dairy Research*, *71*, 51–57. https://doi.org/10.1017/S0022029903006575

Auldist, M. P., Coats, S., Rogers, G. L., & McDowell, G. H. (1995). Changes in composition of milk from healthy and mastitic cows during the lactation cycle. *Australian Journal of Experimental Agriculture.*, *35*, 427–436. https://doi.org/10.1071/EA9950427

Benedet, A., Costa, A., De Marchi, M., & Penasa, M. (2020). Heritability estimates of predicted blood *β*-hydroxybutyrate and nonesterified fatty acids and relationships with milk traits in early-lactation Holstein cows. *Journal of Dairy Science*, *103*, 6354–6363. http://dx.doi.org/10.3168/jds.2019-17916

Berry, D. P. (2008). Genetics - a tool to improve productivity and profitability. *International Journal of Dairy Technology*, *61*, 30–35. https://doi.org/10.1111/J.1471-0307.2008.00371.X

Berry, D. P., Wall, E., & Pryce, J. E. (2014). Genetics and genomics of reproductive performances in dairy and beef cattle. *Animal*, *8*, 105–121. https://doi.org/10.1017/S1751731114000743

Bonfatti, V., Di Martino, G., & Carnier, P. (2011). Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. *Journal of Dairy Science*, *94*, 5776–5785. https://doi.org/10.3168/jds.2011-4401

Bonfatti, V., Cecchinato, A., Gallo, L., Blasco, A., & Carnier, P. (2011). Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. *Journal of Dairy Science*, *94*, 5183–5193. https://doi.org/10.3168/jds.2011-4297

Bonfatti, V., Vicario, D., Lugo, A., & Carnier, P. (2017). Genetic parameters of measured and population-wide infrared predictions of 92 traits describing the fine composition and technological properties of milk in Italian Simmental cattle. *Journal of Dairy Science*, *100*, 5526–5540. https://doi.org/10.3168/jds.2016-11667

Chuang, C. K., Lin, S. P., Lee, H. C., Wang, T. J., Shih, Y. S., Huang, K. Y., & Yeung, C. Y. (2005). Free amino acids in full-term and pre-term human milk and infant formula. *Journal of Pediatric Gastroenterology and Nutrution*, *40*, 496–500. https://doi.org/10.1097/01.mpg.0000150407.30058.47

Cole, J. B., & VanRaden, P. M. (2018). Symposium review: Possibilities in an age of genomics: The future of selection indices. *Journal of Dairy Science*, *101*, 3686–3701. https://doi.org/10.3168/jds.2017-13335

Costa, A., Bovenhuis, H., & Penasa, M. (2020). Changes in milk lactose content as indicators for longevity and udder health in Holstein cows. *Journal of Dairy Science*, *103*, 11574–11584. https://doi.org/10.3168/jds.2020-18615

Costa A., Niero G., Franzoi M., Cassandro M., De Marchi M., & Penasa M. (2021). Short communication: Iodine content in bovine milk is lowly heritable and shows limited genetic variation. *Journal of Dairy Science*, *104*, 3292–3297. http://dx.doi.org/10.3168/jds.2020-19486

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*, *3*, e3395. https://doi.org/10.1371/journal.pone.0003395

De Marchi, M., Toffanin, V., Cassandro, M., & Penasa, M. (2014). Invited review: Mid-infrared spectroscopy as a phenotyping tool for milk traits. *Journal of Dairy Science*, *97*, 1171–1186. https://doi.org/10.3168/jds.2013-6799

Franzoi, M., Costa, A., Penasa, M., & De Marchi, M. (2021). Genetic background of calcium and phosphorus phases predicted from milk mid-infrared spectra in Holstein cows. *Italian Journal of Animal Science*, *20*, 777–783. https://doi.org/10.1080/1828051X.2021.1912663

Ghadimi, H., & Pecora, P. (1963). Free amino acids of different kinds of milk. *American Journal of Clinical Nutrition*, *13*, 75–81. https://doi.org/10.1093/ajcn/13.2.75

Gilmour, A. R., Cullis, B. R., Welham, S. J., & Thompson, R. (2011). *ASREML reference manual*. VSN International Ltd.

Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*, *136*, 245–257. https://doi.org/10.1007/s10709-008-9308-0

Graml, R., & Pirchner, F. (2003). Effects of milk protein loci on content of their proteins. *Archives Animal Breeding*, *46*, 331–340. https://doi.org/10.5194/aab-46-331-2003

Grelet, C., Vanlierde, A., Hostens, M., Foldager, L., Salavati, M., Ingvartsen, K.L., Crowe, M., Sorensen, M.T., Froidmont, E., Ferris, C.P., Marchitelli, C., Becker, F., Larsen, T., Carter, F., & Dehareng, F. (2019). Potential of milk mid-IR spectra to predict metabolic status of cows through blood components and an innovative clustering approach. *Animal*, *13*, 649–658. http://dx.doi.org/10.1017/s1751731118001751

Henchion, M., McCarthy, M., Resconi, V. C., Berry, D. P., & McParland, S. (2016). Stakeholder involvement in establishing a milk quality sub-index in dairy cow breeding goals: A Delphi approach. *Animal*, *10*, 878–891. https://doi.org/10.1093/ajcn/77.6.1555S

Huang, W., Penagaricano, F., Ahmad, K. R., Lucey, J. A., Weigel, K. A., & Khatib, H. (2012). Association between milk protein gene variants and protein composition traits in dairy cattle. *Journal of Dairy Science*, *95*, 440–449. https://doi.org/10.3168/jds.2011-4757

Lien, E. L. (2003). Infant formulas with increased concentrations of α-lactalbumin. *American Journal of Clinical Nutrition*, *77*, 1555–1558. https://doi.org/10.1093/ajcn/77.6.1555S

Lopez-Villalobos, N. (2012). Analysing the genetic basis of milk production traits. *CAB Reviews*, *5167*, 1–18. https://doi.org/10.1079/PAVSNNR20127028

McDermott, A., Visentin, G., De Marchi, M., Berry, D. P., Fenelon, M. A., O'Connor, P. M., Kenny, O. A., & McParland, S. (2016). Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. *Journal of Dairy Science*, *99*, 3171–3182. https://doi.org/10.3168/jds.2015-9747

McDermott, A., De Marchi, M., Berry, D. P., Visentin, G., Fenelon, M. A., Lopez-Villalobos, N., & McParland, S. (2017). Cow and environmental factors associated with milk protein fractions and free amino acids predicted by mid-infrared spectroscopy in bovine milk. *Journal of Dairy Science*, *100*, 6272–6284. https://doi.org/10.3168/jds.2016-12410

McParland, S., Kearney, J. F., Rath, M., & Berry, D. P. (2007). Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *Journal of Animal Science*, *85*, 322–331. https://doi.org/10.2527/jas.2006-367

Miglior, F., Muir, B. L., & Van Doormaal, B. J. (2005). Selection indices in Holstein cattle of various countries. *Journal of Dairy Science*, *88*, 1255–1263. https://doi.org/10.3168/jds.S0022-0302(05)72792-2

Niero, G., Penasa, M., Gottardo, P., Cassandro, M., & De Marchi, M. (2016). Short communication: Selecting the most informative mid-infrared spectra wavenumbers to improve the accuracy of prediction models for detailed protein composition. *Journal of Dairy Science*, *99*, 1853–1858. https://doi.org/10.3168/jds.2015-10318

Norman, H. D., & Powell, R. L. (1999). Dairy cows of high genetic merit for yields of milk, fat, and protein - a review. *Journal of Animal Science*, *12*, 1316–1323. https://doi.org/10.5713/ajas.1999.1316

Oltenacu, P. A., & Broom, D. M. (2010). The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Animal Welfare*, *19*, 39–49.

Penasa, M., Toffanin, V., Cologna, N., Cassandro, M., & De Marchi, M. (2016). Effects of dairy factory, milk casein content and titratable acidity on coagulation properties in Trentingrana dairy industry. *Journal of Dairy Research*, *83*, 242–248. https://doi.org/10.1017/S0022029916000212

Rendel, J. M., & Robertson, A. (1950). Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *Journal of Genetics*, *50*, 1–8. https://doi.org/10.1007/BF02986789

Rutten, M. J. M., Bovenhuis, H., Heck, J. M. L., & van Arendonk, J. A. M. (2011). Predicting bovine milk protein composition based

on Fourier transform infrared spectra. *Journal of Dairy Science*, *94*, 5683–5690. https://doi.org/10.3168/jds.2011-4520

Sanchez, M. P., Ferrand, M., Gelé, M., Pourchet, D., Miranda, G., Martin, P., Brochard, M., & Boichard, D. (2017). Short communication: Genetic parameters for milk protein composition predicted using mid-infrared spectroscopy in the French Montébeliarde, Normande, and Holstein dairy cattle breeds. *Journal of Dairy Science*, *100*, 6371–6375. https://doi.org/10.3168/jds.2017-12663

Schopen, G. C. B., Heck, J. M. L., Bovenhuis, H., Visker, M. H. P. W., van Valenberg, H. J. F., & van Arendonk, J. A. M. (2009). Genetic parameters for major milk proteins in Dutch Holstein-Friesians. *Journal of Dairy Science*, *92*, 1182–1191. https://doi.org/10.3168/jds.2008-1281

Simm, G., Pollott, G., Mrode, R. A., Houston, R., & Marshall, K. (2020). *Genetic improvement of farmed animals*. CAB International.

Soyeurt, H., Dehareng, F., Gengler, N., McParland, S., Wall, E., Berry, D. P., Coffey, M., & Dardenne, P. (2011). Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. *Journal of Dairy Science*, *94*, 1657–1667. https://doi.org/10.3168/jds.2010-3408

Urech, E., Puhan, Z., & Schallibaum, M. (1999). Changes in milk protein fraction as affected by subclinical mastitis. *Journal of Dairy Science*, *82*, 2402–2411. https://doi.org/10.3168/jds.S0022-0302(99)75491-3

VanRaden, P. M., & Sanders, A. H. (2003). Economic merit of cross-bred and purebred US dairy cattle. *Journal of Dairy Science*, *86*, 1036–1044. https://doi.org/10.3168/jds.s0022-0302(03)73687-x

van Sadelhoff, J. H. J., Wiertsema, S. P., Garssen, J., & Hogenkamp, A. (2020). Free amino acids in human milk: A potential role for glutamine and glutamate in the protection against neonatal allergies and infections. *Frontiers in Immunology*, *11*, 1007. https://doi.org/10.3389/fimmu.2020.01007

Verduci, E., D'Elios, S., Cerrato, L., Comberiati, P., Calvani, M., Palazzo, S., Martelli, A., Landi, M., Trikamjee, T., & Peroni, D. G. (2019). Cow's milk substitutes for children: Nutritional aspects of milk from different mammalian species, special formula and plant-based beverages. *Nutrients*, *11*, 1739. https://doi.org/10.3390/nu11081739

Visentin, G., McParland, S., De Marchi, M., McDermott, A., Fenelon, M. A., Penasa, M., & Berry, D. P. (2017). Processing characteristics of dairy cow milk are moderately heritable. *Journal of Dairy Science*, *100*, 6343–6355. https://doi.org/10.3168/jds.2017-12642

Wedholm, A., Larsen, L. B., Lindmark-Mansson, H., Karlsson, A. H., & Andren, A. (2006). Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *Journal of Dairy Science*, *89*, 3296–3305. https://doi.org/10.3168/jds.S0022-0302(06)72366-9

Williams, P. (2007). Statistical terms for evaluation of accuracy and precision. In near infrared technology – Getting the best out of light. PDK grain, Nanaimo, BC, and Winnipeg, man., Canada, 5:1–17.

How to cite this article: Visentin, G., Berry, D. P., Costa, A., McDermott, A., De Marchi, M. & McParland, S. (2022). Breeding for improved protein fractions and free amino acids concentration in bovine milk. *Journal of Animal Breeding and Genetics*, *139*, 517–529. https://doi.org/10.1111/jbg.12681