J. Dairy Sci. 108:9144–9151
https://doi.org/10.3168/jds.2025-26500

Mid-infrared spectroscopy can be applied to authenticate A2 milk

S. V. Chirife, E. Albanell, X. Such, and C. L. Manuelian*
Group of Ruminant Research (G2R), Department of Animal and Food Sciences, Universitat Autònoma de Barcelona (UAB), 08193, Bellaterra, Spain

ABSTRACT

Due to a genetic variation in β-casein, A2 milk is more easily digestible than regular milk (A1); presence of the amino acid proline instead of histidine in position 67 of the peptide chain prevents the release of β-casomorphin-7 during digestion. This study evaluated the application of mid-infrared (MIR) spectroscopy as a rapid, noninvasive, and routinely large-scale method to authenticate the A2 variant in Holstein cow milk. Spectral, genetic, and milk quality (fat, protein, lactose, and SCC) data from 2,270 milk samples from 2 consecutive routine milk controls were retrieved from 1,356 animals from 6 farms located in the same area that raised both A1 and A2 cows. Genetic information included β-casein, κ-casein, and β-lactoglobulin variants. Milk compositional differences were statistically assessed before the spectral modeling. Then, a preliminary principal component analysis (PCA) on spectra information was conducted, followed by a partial least squares discriminant analysis (PLS-DA) with 30% of the samples as the test set. Results indicated that milk quality was similar across all protein fractions but differed slightly among farms ($P < 0.05$). The preliminary spectral evaluation revealed that the first 2 components of the PCA explained 73.2% of the variance. Still, it could not segregate A1 and A2 milk samples based on β-casein genetic information. The PLS-DA model revealed the lowest balanced accuracy in the training and testing set for the genotype A1A1 (50%). For genotypes A1A2 and A2A2, a better balanced accuracy was recorded in the training than in the testing set and slightly greater for A2A2 than for A1A2. For A1A2, balanced accuracy was 80% for the training set and 81% for the testing set. For A2A2, the balanced accuracy was 81% for the training set and 82% for the testing set. Moreover, balanced accuracy improved when only considering 2 levels, A1 milk (comprising genotypes A1A1 and A1A2) and A2 milk (genotype A2A2), reaching 94% for the training set and 88% for the testing set. In conclusion, MIR spectral information is a promising method to authenticate A2 milk based on a PLS-DA model.
Key words: mid-infrared spectroscopy, β-casein, PCA, PLS-DA

INTRODUCTION

Milk with the amino acid proline instead of histidine at position 67 in the β-CN amino acid chain is called **A2** milk and generates notably lower quantities of the bioactive opioid peptide β-casomorphin-7 (**BCM-7**) than regular (**A1**) milk during its digestion (Kamiński et al., 2007). The BCM-7 peptide has been associated with several adverse effects on human health at gastrointestinal, cardiovascular, and neurological levels (Kamiński et al., 2007; Summer et al., 2020; Fernández-Rico et al., 2022). Those at the gastrointestinal level have the greatest consensus in the scientific community. Therefore, farmers in Western countries are transitioning to A2 milk–producing cows (Alfonso et al., 2019) to offer a functional product for consumers who experience gastrointestinal discomfort after consuming milk despite not being allergic to milk protein or lactose intolerant.

The standard methods for genotypes identification of milk CN are polymerase chain reaction, PCR-restriction fragment length polymorphism (Vafin et al., 2022), and high-performance liquid chromatography (Bonfatti et al., 2008; 2014). Although these tests are highly accurate, they are expensive and time-consuming, and cannot be used for large-scale applications. Thus, a need exists for a more efficient method that meets the dairy industry's demands for speed and scalability. Mid-infrared (**MIR**) spectroscopy offers advantages compared with standard methods, including nondestructive green chemical analysis and faster delivery of results. Consequently, it is considered a promising alternative to traditional methods for measuring food quality (Su and Sun, 2019). This technique is used daily in the dairy industry to evaluate milk gross composition for milk testing and payment systems (Okpara, 2019; Castro et al., 2022; Du et al., 2023). However, very few studies have evaluated the potential

Received February 20, 2025.
Accepted May 27, 2025.
*Corresponding author: carmen.manuelian@uab.cat

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-25. Nonstandard abbreviations are available in the Notes.

of infrared spectroscopy to predict milk protein fraction genotypes.

Bonfatti et al. (2011) employed MIR spectroscopy to predict α_{S1}-, α_{S2}-, β-, and κ-CN, and α-LA and β-LG content in milk samples from Simmental cows applying a partial least square model, obtaining the best prediction models for β-LG A and B contents, with coefficients of determination in cross-validation of 0.60 and 0.44, respectively. Rutten et al. (2011) achieved 76% correct classification for the AA β-LG genotype, 80% for the AB β-LG genotype, and 66% for the BB β-LG genotype using MIR spectroscopy in Dutch Holstein-Friesian milk samples. For β-CN, Xiao et al. (2022) obtained a 96% correct classification of A1 and A2 Holstein milk when applying a partial least squares discriminant analysis (**PLS-DA**) on MIR spectra. Nevertheless, Daniloski et al. (2022) indicated that principal component analysis (**PCA**) on the MIR spectra of Holstein milk could not identify β-CN genotypes. Moreover, Navarro et al. (2024) reported 64% accuracy in discriminating A2 from A1 Holstein milk with PLS-DA on near-infrared spectra.

Therefore, this study evaluated the suitability of MIR spectroscopy as a rapid, economical, and green chemistry method to authenticate A2 milk in Holstein cow milk for implementation in large-scale milk testing routines.

MATERIALS AND METHODS

Milk Sample Spectral, Genetic, and Quality Information

A total of 3,427 spectra of milk samples from 2,536 cows, collected with bronopol (Broad Spectrum Microtabs II, D&F Control Systems, San Ramon, CA) from 2 consecutive official controls between May and July 2023 from 6 farms (F1–F6) located in Catalonia, Spain, were provided by the Interprofessional Dairy Association of Catalonia. Cows' genetic variant information on β-CN (i.e., genotypes A1A1, A1A2, A2A2), κ-CN (i.e., AA, AB, AE, BB, BE, EE), and β-LG (i.e., AA, AB, BB) was accessible through the Frisian Federation of Catalonia (Vic, Spain).

Spectral data were acquired using a MilkoScan 7RM (Foss, Hillerød, Denmark), which operates over the wavelength range from 5,011.54 to 925.92 cm^{-1} and is routinely standardized following the manufacturer's recommendations. The measurements were taken at intervals of 3.85 cm^{-1}, resulting in 1,060 data points per spectrum. Moreover, the MilkoScan 7RM also determined milk composition (fat, protein, and lactose as a percentage), and a Fossomatic (Foss, Hillerød, Denmark) analyzed SCC. The SCC (cells/mL) was transformed into SCS by applying the Wiggans and Shook (1987) equation: SCS = $\log_2$ (SCC/100) + 3.

Data Set Cleaning

A quality control process was conducted on the raw spectral data before the chemometric analyses. During this process, samples with duplicate entries were removed from the database (2.97% of the samples). Additionally, samples from cows without complete genetic variant information (i.e., all 3 fractions) or milk composition and quality parameters were excluded from the analysis (30.7% of the samples). The final data set included 2,270 spectral data from 1,356 cows.

Chemometrics

Chemometric analysis was performed with R version 4.3.3 (R Core Team, 2024). First, spectral data were preprocessed by removing regions 3,690 to 2,990 cm^{-1} and 1,680 to 1,580 cm^{-1}, which correspond to noise induced by water absorption (Grelet et al., 2015). Then, PCA of the centered and scaled variables was performed and visualized with the "FactoMineR" and "factoextra" packages of R, and 2 spectral outliers identified as Mahalanobis distance >3 from the mean were eliminated.

Before PLS-DA modeling, 12 pretreatments were applied to the raw spectrum using the "pretreat_spectra" function from the "waves" package of R before creating the models to investigate which one yielded the best results. The pretreatments were as follows:

- **SNV** = standard normal variate
- SNVD1 = standard normal variate and first derivative
- **SNVD2** = standard normal variate and second derivative
- D1 = first derivative
- D2 = second derivative
- **SG** = Savitzky–Golay filter
- SNVSG = standard normal variate and Savitzky–Golay
- SGD1 = gap-segment derivative
- SGD1W5 = Savitzky–Golay and first derivative (window size = 5)
- SGD1W11 = Savitzky–Golay and first derivative (window size = 11)
- SGD2W5 = Savitzky–Golay and second derivative (window size = 5)
- SGD2W11 = Savitzky–Golay and second derivative (window size = 11)

Then, using the "caret" package of R, the data set was randomly split into a training set (70% of the observations) and a testing set (30% of the observations) with the function "set.seed(108)". Finally, the PLS-DA model of the centered and scaled variables was built using the

"train" function from the "caret" package of R following a 4-fold cross-validation with 3 iterations. The model's performance was assessed with the balanced accuracy, sensitivity, and specificity calculated from the confusion matrix. The balanced accuracy considers the balance between sensitivity (true positive rate) and specificity (true negative rate). Balanced accuracy is particularly useful when dealing with imbalanced data sets, as it gives a more comprehensive view of the model's performance. The adequate number of components to be retained, up to 10, to avoid overfitting of the model, in the final model was automatically selected based on the greatest value for accuracy or the receiver operating characteristic curve. The best model was chosen based on the balanced accuracy in the testing set.

Statistical Analysis

To determine the minimum number of spectra needed for reliable results, a sample size calculation was performed based on the *F*-test of ANCOVA for β-CN using G*Power software version 3.1.9.6 (Faul et al., 2007, 2009; Heinrich Heine Universität Düsseldorf, Germany) and considering 2 numerator degrees of freedom, 3 groups, 0.25 Cohen's medium effect size, 0.95 power analysis, and 0.05 α-error probability. The number of samples for which we had complete information was above the sample size estimated.

Statistical analysis was performed with R version 4.3.3 (R Core Team, 2024). Data quality control was conducted to identify and exclude outliers by removing records with values outside the range of mean ± 3 SD. Descriptive statistics were obtained with the "dplyr" package of R (Wickham et al., 2023). A multigene approach was followed, as suggested by previous studies (Bovenhuis et al., 1992; Navarro et al., 2024), to explore the factors of variation through a linear mixed model using the "lme4" package of R (Bates et al., 2015). The statistical model included fixed effects farm, β-CN, κ-CN, and β-LG. In a preliminary model, the first-level interaction with the farm was included but removed from the final model because it was not significant. The sample was treated as a random factor. Least squares means were calculated using the "emmeans" package of R, multiple comparisons were performed applying Tukey's honestly significant differences test, and significance was declared at $P < 0.05$.

RESULTS AND DISCUSSION

Milk Quality

Average milk quality traits are shown in Table 1. The CV was calculated for each trait based on SD/mean to evaluate data variability. The greatest CV was reached by SCS (73%), followed by fat (22%) and protein contents (10%). The lowest CV was observed for lactose content (3%). These results are consistent with the ranges identified in previous studies (Franzoi et al., 2020; Bisutti et al., 2022; Navarro et al., 2024).

Table 1. Milk quality descriptive statistics after removing records with values outside mean ± 3 SD

Trait	n	Mean	SD	Minimum	Maximum
Fat, %	2,201	3.41	0.75	0.95	5.87
Protein, %	2,201	3.34	0.35	2.43	4.68
Lactose, %	2,201	4.94	0.16	4.32	5.48
SCS[1]	2,201	2.61	1.91	−0.32	8.62

[1]SCS = $\log_2$ (SCC/100) + 3.

For β-CN genotypes, A2A2 was more frequent than A1A2 (61.5%) compared with A1A2 (35.9%), whereas A1A1 represented only 2.6% of the samples (Table 2). The lower proportion of A1A1 than A1A2 and A2A2 agreed with the findings of Bisutti et al. (2022). However, they reported a higher proportion of A1A2 than A2A2. This discrepancy could be attributed to farm-specific selection practices, as our chosen farms are transitioning to A2 milk production, aligning with observations reported by Kamiński et al. (2023). Those authors observed an increase of the A2 allele from 61% in 2003 to 69% in 2019. For κ-CN, genotypes AB, AA, and BB were the most frequent (41.4%, 23.4%, and 20.5%, respectively), and EE, AE, and BE were the least frequent (0.2%, 6.7%, and 7.8%, respectively; Table 2). For β-LG, AB was the most frequent (45.8%), and BB was the least frequent (20.6%; Table 2). The representativeness of the different alleles for κ-CN and β-LG aligns with results reported by Bisutti et al. (2022) and Kamiński et al. (2023).

Milk quality was similar among the different genotypes of β-CN, κ-CN, and β-LG (Table 2). These results align with those reported by Nguyen et al. (2018), who observed similar fat, protein, and lactose contents between A1 and A2 bulk milk from Kiwi Cross breeds. Moreover, Bisutti et al. (2022) reported similar fat, protein, and lactose contents among β-CN, κ-CN, and β-LG genotypes in individual cow milk from Italian Holsteins. Navarro et al. (2024) also reported similar fat, protein, and lactose contents among β-CN and β-LG genotypes in individual cow milk from Spanish Holsteins. However, they observed a greater protein concentration for κ-CN genotypes AB than BE. The significant difference reported by Navarro et al. (2024) could be related to their study's small sample size, as they included only 168 samples.

The only significant effect in the models was the farm where the cows were raised (Table 2). Differences in milk composition observed among farms could be attributed to factors such as parity, lactation stage, and diet.

Table 2. Milk quality traits (LSM ± SE) for the fixed effects included in the model

Fixed effect	n	Fat, %	Protein, %	Lactose, %	SCS[1]
β-Casein genotype					
A1A1	59	3.44 ± 0.47	3.41 ± 0.28	4.96 ± 0.13	3.06 ± 1.19
A1A2	815	3.47 ± 0.29	3.32 ± 0.18	4.91 ± 0.07	2.94 ± 0.74
A2A2	1,396	3.48 ± 0.29	3.28 ± 0.18	4.93 ± 0.08	2.94 ± 0.76
κ-Casein genotype					
AA	532	3.46 ± 0.22	3.34 ± 0.14	4.95 ± 0.06	2.79 ± 0.59
AB	939	3.45 ± 0.19	3.41 ± 0.13	4.94 ± 0.05	2.68 ± 0.53
AE	151	3.63 ± 0.31	3.37 ± 0.14	4.94 ± 0.09	3.10 ± 0.85
BB	466	3.39 ± 0.23	3.39 ± 0.19	4.97 ± 0.06	2.57 ± 0.63
BE	177	3.42 ± 0.29	3.28 ± 0.19	4.93 ± 0.08	2.89 ± 0.8
EE	5	3.42 ± 1.49	3.23 ± 0.9	4.89 ± 0.4	3.86 ± 3.85
β-Lactoglobulin genotype					
AA	766	3.42 ± 0.28	3.31 ± 0.17	4.92 ± 0.08	3.04 ± 0.72
AB	1,037	3.51 ± 0.28	3.34 ± 0.16	4.93 ± 0.07	2.93 ± 0.72
BB	467	3.46 ± 0.3	3.37 ± 0.19	4.95 ± 0.08	2.98 ± 0.79
Farm					
F1	347	3.18 ± 0.31^{a}	3.17 ± 0.2^{a}	4.97 ± 0.09^{b}	3.01 ± 0.83
F2	301	3.81 ± 0.31^{c}	3.35 ± 0.2^{ab}	4.94 ± 0.09^{ab}	3.08 ± 0.83
F3	332	3.74 ± 0.32^{c}	3.40 ± 0.2^{c}	4.92 ± 0.09^{a}	2.70 ± 0.84
F4	671	3.37 ± 0.29^{b}	3.51 ± 0.17^{d}	4.91 ± 0.08^{a}	3.12 ± 0.76
F5	341	3.34 ± 0.31^{ab}	3.34 ± 0.18^{bc}	4.92 ± 0.09^{a}	2.77 ± 0.8
F6	278	3.33 ± 0.33^{ab}	3.25 ± 0.21^{bc}	4.95 ± 0.09^{ab}	3.22 ± 0.88

$^{a–d}$Within each trait and fixed effect, different letters indicate significant difference at $P < 0.05$.

[1]SCS = $\log_2$ (SCC/100) + 3.

Mikóné Jónás et al. (2016) observed that higher parity and advanced stage of lactation led to lower fat, protein, and lactose contents and higher SCC. In addition, Veena et al. (2021) have associated the greater activity of acetyl CoA carboxylase enzyme at the end than in earlier stages of lactation with greater fat content.

Spectral Information

Figure 1 shows the raw spectra after removing 2 outliers. According to Grelet et al. (2015), protein peaks observed around 1,550 cm^{-1} are associated with the stretching vibrations of carbon-nitrogen (C–N) and nitrogen-nitrogen (N–N) bonds. Fatty acids exhibit peaks at approximately 1,390 cm^{-1} and 1,454 cm^{-1}, which are related to carbon-hydrogen (C–H) bonds in methyl ($-CH_3$) and methylene ($-CH_2$) groups. Additionally, fatty acids show peaks around 2,862 cm^{-1} and 2,927 cm^{-1}, attributed to C–H bonds. Noisy areas induced by water absorption are between 1,600 cm^{-1} and 1,689 cm^{-1} and between 3,008 cm^{-1} and 5,010 cm^{-1}. Additionally, the region between 3,008 cm^{-1} and 5,010 cm^{-1} is considered noninformative, often referred to as the short-wavelength infrared region, with minimal spectral information (Karoui et al., 2010; El Jabri et al., 2019). Karoui et al. (2003) have also described that water presents strong bands centered at 1,640 cm^{-1} related to H–O–H bending vibration, at 2,310 cm^{-1} (water association band), and 3,360 cm^{-1} linked to H–O stretching band.

PCA to Discriminate β-CN Variants Based on Spectral Information

The first 2 principal components (PC1 and PC2) of the PCA explained 73.2% of the total variance. However, they did not cluster A1 and A2 milk (Figure 2). Other components (up to the 10 first principal components) were also tested for clustering without success in discriminating A1 and A2 milk. Similar results were reported by Daniloski et al. (2022), where no apparent clustering was observed among β-CN variants in 114 Australian Holstein cow milk samples unless performing the PCA on a smaller tailored subset of samples instead of using the complete data set. Moreover, Navarro et al. (2024) could not cluster β-CN, κ-CN, and β-LG genotypes when applying the PCA in near-infrared spectra of milk samples from 168 Spanish Holstein cows. This lack of clustering may be influenced by genetic variability across different farms, as suggested by Daniloski et al. (2022).

PLS-DA Discriminate β-CN Variants Based on Spectral Information

From all 12 pretreatments applied to the raw spectrum, the best balanced accuracy in the testing set of the model was achieved for 3 components with the raw spectra, which indicated that the raw spectrum can be used directly without the need to apply pretreatments. This approach saves time and reduces complexity in data

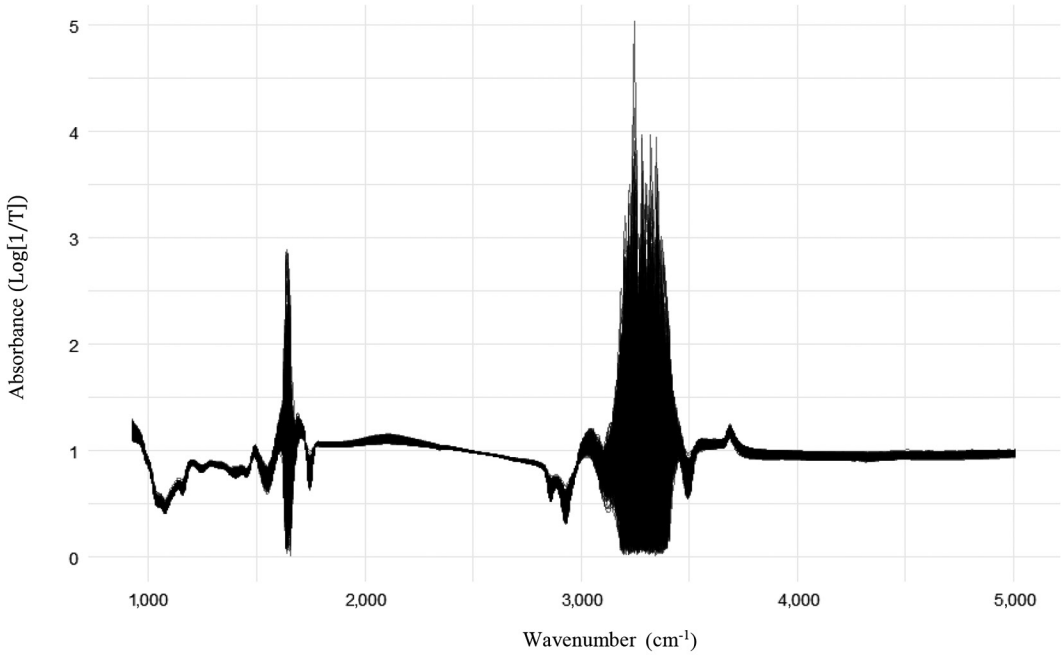

Figure 1. Raw spectra of milk samples after removing 2 samples with Mahalanobis distance >3. T = transmittance.

preprocessing. However, when applying the pretreatment SNVD2, the model improved when only 2 levels of type of milk were considered: A1 milk (comprising genotypes A1A1 and A1A2) and A2 milk (genotype A2A2). The SNV is a scattering correction preprocessing that normalizes individual spectra to remove additive and multiplicative scattering effects (Rinnan et al., 2009). The second derivative corrects for the effect of overlapping peaks and removes the spectral baseline shift and baseline slope (Agelet and Hurburgh, 2010).

The model for the 3 levels of type of milk (A1A1, A1A2, and A2A2 milk) failed to correctly classify the A1A1 genotype (low sensitivity; Table 3), possibly due to the limited number of A1A1 samples. However, it achieved 71% and 92% sensitivity in the training set for A1A2 and A2A2, respectively, meaning it correctly classifies A1A2 and A2A2 samples as A1A2 and A2A2, respectively, minimizing false negatives. Additionally, the model achieved 100%, 89%, and 71% specificity (true negatives) for A1A1, A1A2, and A2A2 in the training set, respectively, meaning it correctly classified samples not belonging to a given class, minimizing false positives. Similar sensitivity and specificity values for each group were obtained in the testing set, with sensitivity for A1A2 and A2A2 of 75% and 90%, respectively, and specificity of 87% and 75%, respectively. Thus, most correct assignments were concentrated in A1A2 and A2A2, with balanced accuracy of 80% and 81%, respectively, for the training test, and 81% and 82%, respectively, for the testing set, which means that, on average, the model correctly identified at least 80% of both positive and negative cases. The model for 2 levels, A1 milk (A1A1

Table 3. Confusion matrix and performance of the model for β-casein genotypes discrimination for 3 components based on the best balanced accuracy in the testing set obtained using the SNVD2 correction

	Reference class					
	Training set (n = 1,589)			Testing set (n = 679)		
Predicted class	A1A1	A1A2	A2A2	A1A1	A1A2	A2A2
A1A1	0	0	0	0	0	0
A1A2	31	403	81	14	182	43
A2A2	11	167	896	3	62	375
Sensitivity, %	0	71	92	0	75	90
Specificity, %	100	89	71	100	87	75
Balanced accuracy, %	50	80	81	50	81	82

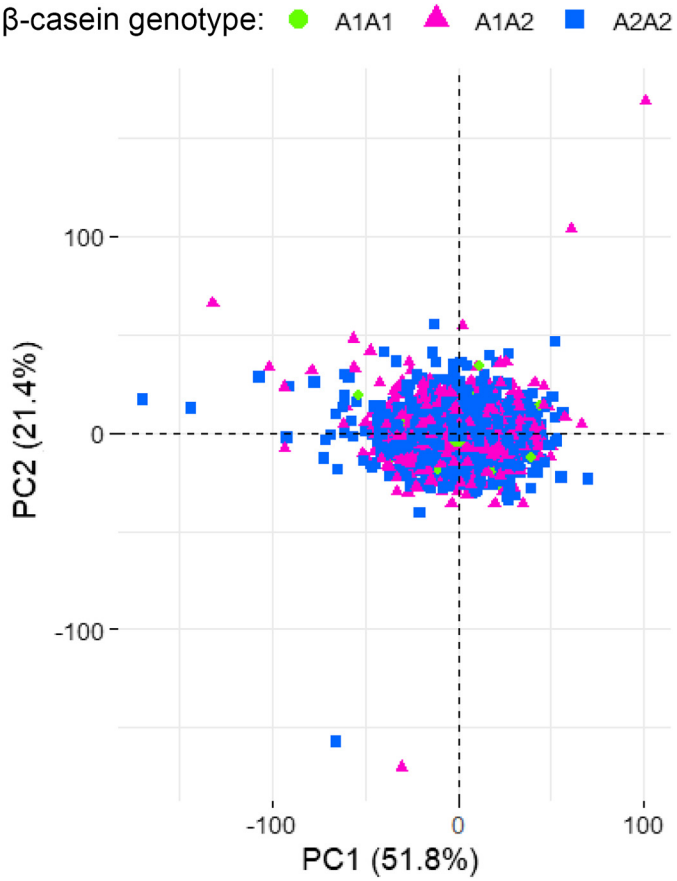

Figure 2. β-Casein principal component (PC) analysis plot for PC1 and PC2.

and A1A2) and A2 milk (A2A2), improved the balanced accuracy for the training set (94%) and in the testing set (88%; Table 4). In that case, the balanced accuracy is considered excellent in the training set and good in the testing set, suggesting their applicability in regular milk controls. Results of both models were better than those reported by Navarro et al. (2024), who obtained 64% balanced accuracy to discriminate A2A2 from A1A2 milk samples using near-infrared spectra. In contrast, Daniloski et al. (2022) reported a complete discrimination of β-CN genotypes when applying PLS-DA on MIR spectra, which they attributed to a tailored selection of samples. Moreover, Xiao et al. (2022) also obtained better classificatory results (96%) in identifying A1 and A2 milk using MIR spectra.

CONCLUSIONS

The study revealed that PCA could not segregate milk samples based on β-CN genetic information, despite explaining 73% of the variance with the first 2 components when milk quality was similar across all protein fractions. The balanced accuracy for both A1A2 and A2A2 in the PLS-DA model could be considered good for the training set (80% and 81%, respectively) and for the testing set (81% and 82%, respectively). The model improved to detect A2 milk when confronted with A1 milk (A1A1 and A1A2), reaching an excellent classification for the training set (94%) and a good classification for the testing set (88%). Thus, MIR spectroscopy is a promising technique for authenticating A2 milk based on a PLS-DA model during routine milk control testing. However, the potential to discriminate A1A1 milk still requires confirmation with a more balanced sample representation.

NOTES

This research was funded by the Ministry of Science, Innovation and Universities of Spain (Madrid, Spain) under the project PID2019-110752RB.I00. Carmen L. Manuelian received funding from the Ramon y Cajal program of the Spanish Ministry of Science, Innovation and Universities (grant ref. RYC2023-042902-I). The authors thank the Interprofessional Dairy Association of Catalonia (Barcelona, Spain) for providing the spectral information and the Frisian Federation of Catalonia (Vic, Spain) for the genetic information. The data presented in this study are available free of charge for any user at the official data repository CORA RDR (https://doi.org/10.34810/data2338). Because no human or animal

Table 4. Confusion matrix and performance of the model for A2 and A1 milk discrimination for 3 components based on the best balanced accuracy in the testing set obtained using raw spectra

	Reference class			
	Training set (n = 1,589)		Testing set (n = 679)	
Predicted class	A1 milk	A2 milk	A1 milk	A2 milk
A1 milk	556	36	215	23
A2 milk	56	941	46	395
Sensitivity, %	91		82	
Specificity, %	96		95	
Balanced accuracy, %	94		88	

subjects were used, this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board. The funders had no role in the study's design, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: A1 = regular milk; A2 = milk with the amino acid proline instead of histidine at position 67 in the β-CN amino acid chain; BCM-7 = β-casomorphin-7; D1 = first derivate; D2 = second derivate; MIR = mid-infrared; PCA = principal component analysis; PLS-DA = partial least squares discriminant analysis; SG = Savitzky–Golay filter; SNV = standard normal variate; W5 = window size 5; W11 = window size 11.

REFERENCES

Agelet, L. E., and C. R. Hurburgh Jr. 2010. A tutorial on near infrared spectroscopy and its calibration. Crit. Rev. Anal. Chem. 40:246–260. https://doi.org/10.1080/10408347.2010.515468.

Alfonso, L., O. Urrutia, and J. A. Mendizabal. 2019. Conversión de las explotaciones de vacuno de leche a la producción de leche A2 ante una posible demanda del mercado: Posibilidades e implicaciones. Inf. Téc. Econ. Agrar. 115:231–251. https://doi.org/10.12706/itea.2019.001.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67:1–48. https://doi.org/10.18637/jss.v067.i01.

Bisutti, V., S. Pegolo, D. Giannuzzi, L. F. M. Mota, A. Vanzin, A. Toscano, E. Trevisi, P. Ajmone Marsan, M. Brasca, and A. Cecchinato. 2022. The β-casein (CSN2) A2 allelic variant alters the milk protein profile and slightly worsens coagulation properties in Holstein cows. J. Dairy Sci. 105:3794–3809. https://doi.org/10.3168/jds.2021-21537.

Bonfatti, V., G. Chiarot, and P. Carnier. 2014. Glycosylation of κ-casein: Genetic and nongenetic variation and effects on rennet coagulation properties of milk. J. Dairy Sci. 97:1961–1969. https://doi.org/10.3168/jds.2013-7418.

Bonfatti, V., G. di Martino, and P. Carnier. 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. J. Dairy Sci. 94:5776–5785. https://doi.org/10.3168/jds.2011-4401.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier. 2008. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. J. Chromatogr. A 1195:101–106. https://doi.org/10.1016/j.chroma.2008.04.075.

Bovenhuis, H., J. A. M. van Arendonk, and S. Korver. 1992. Associations between milk protein polymorphisms and milk production traits. J. Dairy Sci. 75:2549–2559. https://doi.org/10.3168/jds.S0022-0302(92)78017-5.

Castro, M. M. D., R. D. Matson, D. E. Santschi, M. I. Marcondes, and T. J. DeVries. 2022. Association of housing and management practices with milk yield, milk composition, and fatty acid profile, predicted using Fourier transform mid-infrared spectroscopy, in farms with automated milking systems. J. Dairy Sci. 105:5097–5108. https://doi.org/10.3168/jds.2021-21150.

Daniloski, D., N. A. McCarthy, T. F. O'Callaghan, and T. Vasiljevic. 2022. Authentication of β-casein milk phenotypes using FTIR spectroscopy. Int. Dairy J. 129:105350. https://doi.org/10.1016/j.idairyj.2022.105350.

Du, C., X. Ren, C. Chu, L. Ding, L. Nan, A. Sabek, G. Hua, L. Yan, Z. Zhang, and S. Zhang. 2023. Assessing the relationship between somatic cell count and the milk mid-infrared spectrum in Chinese Holstein cows. Vet. Rec. 193:e3560. https://doi.org/10.1002/vetr.3560.

El Jabri, M., M. P. Sanchez, P. Trossat, C. Laithier, V. Wolf, P. Grosperrin, E. Beuvier, O. Rolet-Répécaud, S. Gavoye, Y. Gaüzère, O. Belysheva, E. Notz, D. Boichard, and A. Delacroix-Buchet. 2019. Comparison of Bayesian and partial least squares regression methods for mid-infrared prediction of cheese-making properties in Montbéliarde cows. J. Dairy Sci. 102:6943–6958. https://doi.org/10.3168/jds.2019-16320.

Faul, F., E. Erdfelder, A. Buchner, and A.-G. Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav. Res. Methods 41:1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39:175–191. https://doi.org/10.3758/BF03193146.

Fernández-Rico, S., A. C. Mondragón, A. López-Santamarina, A. Cardelle-Cobas, P. Regal, A. Lamas, I. S. Ibarra, A. Cepeda, and J. M. Miranda. 2022. A2 milk: New perspectives for food technology and human health. Foods 11:2387. https://doi.org/10.3390/foods11162387.

Franzoi, M., C. L. Manuelian, M. Penasa, and M. De Marchi. 2020. Effects of somatic cell score on milk yield and mid-infrared predicted composition and technological traits of Brown Swiss, Holstein Friesian, and Simmental cattle breeds. J. Dairy Sci. 103:791–804. https://doi.org/10.3168/jds.2019-16916.

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci. 98:2150–2160. https://doi.org/10.3168/jds.2014-8764.

Kamiński, S., A. Cieślińska, and E. Kostyra. 2007. Polymorphism of bovine beta-casein and its potential effect on human health. J. Appl. Genet. 48:189–198. https://doi.org/10.1007/BF03195213.

Kamiński, S., T. Zabolewicz, K. Oleński, and A. Babuchowski. 2023. Long-term changes in the frequency of beta-casein, kappa-casein and beta-lactoglobulin alleles in Polish Holstein-Friesian dairy cattle. J. Anim. Feed Sci. 32:205–210. https://doi.org/10.22358/jafs/157531/2023.

Karoui, R., G. Downey, and C. Blecker. 2010. Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships—A review. Chem. Rev. 110:6144–6168. https://doi.org/10.1021/cr100090k.

Karoui, R., G. Mazerolles, and É. Dufour. 2003. Spectroscopic techniques coupled with chemometric tools for structure and texture determinations in dairy products. Int. Dairy J. 13:607–620. https://doi.org/10.1016/S0958-6946(03)00076-1.

Mikóné Jónás, E., S. Atasever, M. Gráff, and H. Erdem. 2016. Non-genetic factors affecting milk yield, composition and somatic cell count in Hungarian Holstein cows. Kafkas Univ. Vet. Fak. Derg. 22:361–366. https://doi.org/10.9775/kvfd.2015.14672.

Navarro, N. S., E. Albanell, M. de Marchi, and C. L. Manuelian. 2024. An attempt to identify milk protein fraction genotypes using unsupervised and supervised near-infrared spectroscopy methods. Ital. J. Anim. Sci. 23:313–319. https://doi.org/10.1080/1828051X.2024.2314157.

Nguyen, H. T. H., H. Schwendel, D. Harland, and L. Day. 2018. Differences in the yoghurt gel microstructure and physicochemical properties of bovine milk containing A1A1 and A2A2 β-casein phenotypes. Food Res. Int. 112:217–224. https://doi.org/10.1016/j.foodres.2018.06.043.

Okpara, M. O. 2019. Milk fatty acids estimation by mid-infrared spectroscopy as proxy for prediction of methane emission in dairy cows. Russ. Agric. Sci. 45:386–392. https://doi.org/10.3103/S1068367419040116.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed Mar. 1, 2024. https://www.R-project.org/.

Rinnan, Å., F. van den Berg, and S. B. Engelsen. 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends Analyt. Chem. 28:1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.

Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2011. Prediction of β-lactoglobulin genotypes based on milk Fourier transform infrared spectra. J. Dairy Sci. 94:4183–4188. https://doi.org/10.3168/jds.2011-4149.

Su, W.-H., and D.-W. Sun. 2019. Mid-infrared (MIR) spectroscopy for quality analysis of liquid foods. Food Eng. Rev. 11:142–158. https://doi.org/10.1007/s12393-019-09191-2.

Summer, A., F. di Frangia, P. Ajmone Marsan, I. de Noni, and M. Malacarne. 2020. Occurrence, biological properties and potential effects on human health of β-casomorphin 7: Current knowledge and concerns. Crit. Rev. Food Sci. Nutr. 60:3705–3723. https://doi.org/10.1080/10408398.2019.1707157.

Vafin, R. R., A. G. Galstyan, S. V. Tyulkin, Kh. Kh. Gilmanov, E. A. Yurova, V. K. Semipyatniy, and A. V. Bigaeva. 2022. Species identification of ruminant milk by genotyping of the κ-casein gene. J. Dairy Sci. 105:1004–1013. https://doi.org/10.3168/jds.2020-19931.

Veena, N., J. Hundal, M. Wadhwa, and A. Puniya. 2021. Factors affecting the milk yield, milk composition and physico-chemical parameters of ghee in lactating crossbred cows. Indian J. Dairy Sci. 74:68–73. https://doi.org/10.33785/IJDS.2021.v74i01.009.

Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughan. 2023. dplyr: A Grammar of Data Manipulation. R package version 1.1.4. Accessed Mar. 1, 2024. https://CRAN.R-project.org/package=dplyr.

Wiggans, G. R., and G. E. Shook. 1987. A lactation measure of somatic cell count. J. Dairy Sci. 70:2666–2672. https://doi.org/10.3168/jds.S0022-0302(87)80337-5.

Xiao, S., Q. Wang, C. Li, W. Liu, J. Zhang, Y. Fan, J. Su, H. Wang, X. Luo, and S. Zhang. 2022. Rapid identification of A1 and A2 milk based on the combination of mid-infrared spectroscopy and chemometrics. Food Control 134:108659. https://doi.org/10.1016/j.foodcont.2021.108659.

ORCIDS

S. V. Chirife, https://orcid.org/0009-0002-9759-9488
E. Albanell, https://orcid.org/0000-0002-6158-7736
X. Such, https://orcid.org/0000-0002-9712-4477
C. L. Manuelian https://orcid.org/0000-0002-0090-0362