



An attempt to identify milk protein fraction genotypes using unsupervised and supervised near-infrared spectroscopy methods

Norma S. Navarro, Elena Albanell, Massimo De Marchi & Carmen L. Manuelian

To cite this article: Norma S. Navarro, Elena Albanell, Massimo De Marchi & Carmen L. Manuelian (2024) An attempt to identify milk protein fraction genotypes using unsupervised and supervised near-infrared spectroscopy methods, Italian Journal of Animal Science, 23:1, 313-319, DOI: [10.1080/1828051X.2024.2314157](https://doi.org/10.1080/1828051X.2024.2314157)

To link to this article: <https://doi.org/10.1080/1828051X.2024.2314157>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 11 Feb 2024.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)

An attempt to identify milk protein fraction genotypes using unsupervised and supervised near-infrared spectroscopy methods

Norma S. Navarro^a , Elena Albanell^a , Massimo De Marchi^b  and Carmen L. Manuelian^a 

^aDepartament de Ciència Animal i dels Aliments, Grup de Recerca en Remugats (G2R), Universitat Autònoma de Barcelona, Bellaterra, Spain; ^bDipartimento di Agronomia, Animali, Alimenti, Risorse naturali e Ambiente, Università di Padova, Legnaro, Italy

ABSTRACT

The aim was to evaluate near-infrared spectroscopy (NIRS) potential to discriminate among β -casein (CN), κ -CN and β -lactoglobulin (LG) genotypes to be used as an authentication method. A total of 168 milk samples with known genetic information for β -CN, κ -CN and β -LG were collected at the same farm and paired with the NIRS spectrum. Spectra were evaluated with an unsupervised method (principal component analysis, PCA) and a supervised method (partial least squares-discriminant analysis, PLS-DA). For the PLS-DA, data were split into a train (75%) and a test set (25%), and the variable in projection >1 criterion was applied to select informative wavelengths. Results obtained confirmed that milk quality was similar among genetic variants. For the PCA, the observed variance explained by the first two principal components was 94%, but samples were not clustered by their genotypes of β -CN (i.e. A1A2, A2A2), κ -CN (i.e. AA, AB, AE, BB, BE) and β -LG (i.e. AA, AB, BB). The best accuracy for the PLS-DA models was reached by β -CN (train and test set, 64%), followed by β -LG (train set, 56%; test set, 52%) and κ -CN (train set, 41%; test set, 36%). In conclusion, the PCA on milk spectra was not able to cluster β -CN, κ -CN and β -LG genotypes, but the PLS-DA models revealed promising results for β -CN and β -LG. It could be interesting to increase the number of samples to equilibrate genetic variants and to apply a sampling selection method before discarding the applicability of NIRS as an authentication method.

HIGHLIGHTS

- Near-infrared spectroscopy discriminates β -casein (CN) and β -lactoglobulin (LG) more accurately than κ -CN genotypes.
- Scarce cluster ability of β -CN, κ -CN and β -LG genotypes with principal component analysis (PCA).
- Partial least squares-discriminant analysis (PLS-DA) moderately discriminates β -CN and β -LG genotypes.
- PLS-DA lowly discriminates κ -CN genotypes.

ARTICLE HISTORY

Received 1 June 2023
Revised 14 November 2023
Accepted 30 January 2024

KEYWORDS

A2 milk; genetic polymorphism; NIRS; protein variants

Introduction

Nowadays, there is an interest in producing A2 milk to fulfil consumers' demands because it is considered a healthier alternative to conventional A1 milk (Bodnár et al. 2018). The genetic variant of β -casein (CN) for milk A2 codes for the amino acid proline instead of histidine at position 67. This change prevents β -CN hydrolysis at this position and avoids peptide β -casomorphin-7 (β -CM-7) release after digestion (Thiruvengadam et al. 2021). The β -CM-7 has been associated with worsening gastrointestinal symptoms, cognitive traits and increasing intestinal transit time

(Jianqin et al. 2016). Moreover, β -CN genotype variants impact milk production, protein yield and fat percentage (Bovenhuis et al. 1992). Genotypes of κ -CN affect protein yield and content (Bovenhuis et al. 1992; Tsiaras et al. 2005). Genetic variants of β -lactoglobulin (LG) have been associated with changes in whey composition and properties (Tsiaras et al. 2005; Rutten et al. 2011), milk production, protein yield and fat percentage (Bovenhuis et al. 1992). Thus, all these genetic variations impact milk quality and technological properties of milk (Čítek et al. 2021).

Genetic variants are usually identified through genetic tests (Caroli et al. 2009) which are highly accurate

CONTACT Carmen L. Manuelian  carmen.manuelian@uab.cat

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

but expensive, labour-intensive and not feasible for large-scale screening (Xiao et al. 2022). Near-infrared spectroscopy (NIRS) is considered as a quick method, objective, non-destructive with the sample, free of chemical reagents, low cost and environmentally friendly analysis. It is routinely used in the dairy industry to predict milk quality and represents an interesting approach to detecting food adulteration (dos Santos Pereira et al. 2020). Few studies have attempted to apply principal component analysis (PCA; unsupervised method) or/and partial least squares-discriminant analysis (PLS-DA; supervised method) to infrared spectra to identify protein and whey milk fraction genotypes. Those studies used mid-infrared (MIR) spectroscopy instead of NIRS, and focused on β -CN and β -LG genetic variants. Daniloski et al. (2022) reported that the PCA plot failed to discriminate among β -CN genetic variants. On the other hand, Rutten et al. (2011) and Xiao et al. (2022) obtained a 74% for β -LG and 96% for β -CN accuracy, respectively, when applying a PLS-DA.

Therefore, this study aimed to evaluate NIRS potential to discriminate among β -CN, κ -CN and β -LG genetic variants.

Materials and methods

Milk samples analysis

A total of 168 individual milk (2×50 mL) samples from Holstein-Friesian cows were collected in plastic tubes with preservative bronopol (Broad Spectrum Micro-tabs II, D&F Control Systems, San Ramon, CA) in June 2022 from a commercial farm located in the area of Barcelona, Spain ($42^{\circ}00'11.2''N$ $2^{\circ}13'16.8''E$). This farm has been selecting milk A2 and has the specific genotype information for β -CN (i.e. A1A2, A2A2), κ -CN (i.e. AA, AB, AE, BB, BE) and β -LG (i.e. AA, AB, BB) provided by the Frisian Federation of Catalonia (FEFRIC, Barcelona, Spain). One aliquot (50 mL) was analysed by the Interprofessional Dairy Association of Catalonia (ALLIC) for gross composition (fat, protein, lactose and milk urea nitrogen (MUN)) by MilkoScan (FOSS, Hillerød, Denmark) and somatic cell count (SCC) by Fossomatic (FOSS, Hillerød, Denmark). As described by Manuelian et al. (2021), SCC was transformed into SCS by applying the Wiggans and Shook (1987) equation.

The other aliquot was analysed with a NIRSystems 5000 spectrophotometer (FOSS, Hillerød, Denmark) equipped with a scanning monochromator working from 1100 to 2500 nm every 2 nm at the Agriculture and Animal Production laboratory of the Universitat Autònoma de Barcelona (Barcelona, Spain). Before

scanning the samples, they were heated at $40^{\circ}C$ for 10 min in a water bath (Aparatos Científicos, J.P. Selecta SA, Barcelona, Spain) and homogenised by gently rotating the bottles four times. Then, it was transferred into a quartz glass (diameter 4.8 cm and height 3.9 cm) where a gold reflector (0.5 mm path length) was placed. All samples were scanned in duplicate, manually shaking gently between each scan, and the average spectrum of each sample was used. The absorbance was recorded as a log (1/transflectance). Each spectrum was then matched with the corresponding milk protein genotype information.

Chemometric analysis

The chemometric analysis was conducted following Manuelian et al. (2021) procedure with R version 4.2.0 (R Core Team 2022). Briefly, 'stats' and 'ggbiplot' (Vu 2011) packages were used to perform the PCA. The 'DiscriMiner' package (Sanchez 2013) was applied to perform the PLS-DA for two components after removing the low signal-to-noise ratio wavelength, splitting the dataset into a train set (75% of the observations) and a test set (25% of the observations), and selecting the wavelengths with a variable importance in projection (VIP) score >1 . The PLS-DA model performance was assessed by calculating the error rate and accuracy.

Statistical analysis

The sample size was calculated with G*Power software ver. 3.1.9.6 (Faul et al. 2007, 2009; Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany). The normality of the traits was evaluated with PROC UNIVARIATE of SAS ver. 9.4 (SAS Institute Inc., Cary, NC). A multigene approach was applied to evaluate milk quality through a PROC GLM of SAS ver. 9.4 as suggested by Bovenhuis et al. (1992). The model included cow β -CN, κ -CN and β -LG genotypes, stage of lactation (i.e. class 1–4, being the first class from 3 to 100 DIM; class 2 from 101 to 200 DIM; class 3 from 201 to 300 DIM; and class 4 from >300 DIM), parity (i.e. primiparous and multiparous) and stage of lactation \times parity as fixed effects. Results are reported as least squares means (LSM) and multiple comparisons were performed using Tukey's adjustment when necessary. Significance was declared at $p < .05$ unless otherwise stated.

Results and discussion

Database description and statistics

Average fat and protein content (Table 1) are in line with Sola-Larrañaga and Navarro-Blasco (2009) who evaluated 348 bulk milk from herds located in northern Spain (Navarra region). Moreover, milk quality variability agreed with Franzoi et al. (2020), with Italian Holstein-Friesian individual records with similar MUN content, lower lactose content (4.76%), and greater fat (4.08%) and protein (3.31%) content, and SCS (2.59).

For β -CN, the greater number of cows A2A2 (62.5%) than A1A2 (37.5%; Table 2) is linked to the selection strategy of the farm. For κ -CN, genotypes AB and AA were the most frequent ones (33.9% and 31.0%, respectively); and AE and BE were the least frequent ones (9.5% and 8.3%, respectively; Table 2). For β -LG, AA was the most frequent (51.2%) and BB was the least frequent (8.3%; Table 2). The unbalanced frequency of the genetic variants for κ -CN and β -LG could be attributed to favouring A2A2 for β -CN as indicated by Comin et al. (2008) and supports the multigene statistical approach (Bovenhuis et al. 1992). In agreement with our results, Comin et al. (2008) also

reported a greater frequency of AA and AB for κ -CN when evaluating 1042 Italian Holstein cows. Some authors also observed that BE and EE genotypes for κ -CN are infrequent (Comin et al. 2008; Gai et al. 2021). In line with our results, the Gai et al. (2021) review indicated that genetic variant A is more frequent than B for β -LG in Holstein-Friesian cows.

In the present study, milk quality parameter estimates (Table 2) did not differ among β -CN and β -LG variants. Whereas some authors have reported higher protein yield and lower fat percentage in A2 than A1 variants for β -CN, others have not identified differences in protein and fat percentage (Gai et al. 2021). Moreover, Cendron et al. (2021) also obtained similar fat, protein, MUN and SCS between A1A2 and A2A2 genotypes of β -CN in Italian Holstein. Nevertheless, Albarella et al. (2020) found significant differences in protein percentage in β -CN genotypes in the autochthonous agerolese cattle breed. In line with our results, Botaro et al. (2008) also described a similar milk composition among the β -LG genetic variants (i.e. AA, AB, BB) in Holstein cows. However, Cendron et al. (2021) reported greater fat percentage, MUN and SCS in genetic β -LG variant BB than AA. The lower protein content for BE than AB variants of κ -CN ($p = .031$) agreed with Cendron et al. (2021) results; however, we cannot discard that the observed differences were related to the lower number of cows with the BE variant. Although Heck et al. (2009) and Cendron et al. (2021) also reported similar milk fat percentage among κ -CN genotypes as we did, Cendron et al. (2021) reported significantly lower protein and MUN, and a

Table 1. Descriptive statistics of milk quality traits.

Trait ^a	N	Mean	SD	Minimum	Maximum	CV, %
Milk composition						
Fat, %	162	3.28	0.65	1.74	5.32	20
Protein, %	164	3.15	0.30	2.38	3.02	10
Lactose, %	161	5.18	0.28	4.31	5.84	5
MUN, mg/L	162	193.34	53.56	48.00	369.00	28
SCS, score	164	1.76	1.63	-0.32	7.69	93

^aMUN: milk urine nitrogen; SCS: somatic cell score.

Table 2. Least squares means (\pm SE) of milk quality¹ traits for the main effects.

Fixed effect ²	n	Fat, %	Protein, %	Lactose, %	MUN, mg/L	SCS
β -CN genotypes						
A1A2	63	3.36 \pm 0.10	3.20 \pm 0.04	5.17 \pm 0.04	200.3 \pm 8.2	1.79 \pm 0.23
A2A2	105	3.27 \pm 0.10	3.16 \pm 0.04	5.23 \pm 0.04	196.2 \pm 8.2	1.89 \pm 0.23
κ -CN genotypes						
AA	52	3.41 \pm 0.11	3.16 \pm 0.04 ^{ab}	5.11 \pm 0.05	187.4 \pm 8.9	2.103 \pm 0.246
AB	57	3.30 \pm 0.11	3.30 \pm 0.04 ^a	5.14 \pm 0.05	192.6 \pm 9.1	2.099 \pm 0.252
AE	16	3.36 \pm 0.18	3.15 \pm 0.07 ^{ab}	5.29 \pm 0.08	206.9 \pm 14.5	1.792 \pm 0.404
BB	29	3.43 \pm 0.14	3.27 \pm 0.05 ^{ab}	5.20 \pm 0.06	199.0 \pm 11.1	1.492 \pm 0.307
BE	14	3.10 \pm 0.19	3.03 \pm 0.08 ^b	5.25 \pm 0.08	205.2 \pm 15.8	1.732 \pm 0.439
β -LG genotypes						
AA	86	3.20 \pm 0.08	3.13 \pm 0.03	5.218 \pm 0.031	193.4 \pm 6.2	1.71 \pm 0.17
AB	68	3.39 \pm 0.09	3.12 \pm 0.04	5.223 \pm 0.037	212.4 \pm 7.3	1.65 \pm 0.20
BB	14	3.36 \pm 0.18	3.29 \pm 0.07	5.155 \pm 0.076	188.9 \pm 14.3	2.17 \pm 0.40
Lactation stage						
3–100 d	49	3.33 \pm 0.12 ^{ab}	2.99 \pm 0.05 ^c	5.21 \pm 0.05	175.7 \pm 9.4 ^b	1.64 \pm 0.26
101–200 d	37	3.03 \pm 0.12 ^b	3.12 \pm 0.05 ^{bc}	5.26 \pm 0.05	196.1 \pm 10.0 ^{ab}	1.72 \pm 0.28
201–300 d	51	3.23 \pm 0.10 ^b	3.20 \pm 0.04 ^b	5.16 \pm 0.04	194.7 \pm 8.2 ^{ab}	1.92 \pm 0.23
>300 d	31	3.69 \pm 0.13 ^a	3.42 \pm 0.05 ^a	5.17 \pm 0.05	226. \pm 10.6 ^a	2.08 \pm 0.30
Parity						
Primiparous	87	3.24 \pm 0.09	3.21 \pm 0.04	5.30 \pm 0.04 ^a	200.7 \pm 7.3	1.19 \pm 0.21 ^b
Multiparous	81	3.40 \pm 0.09	3.15 \pm 0.04	5.10 \pm 0.04 ^b	195.8 \pm 7.6	2.50 \pm 0.20 ^a

¹MUN: milk urine nitrogen; SCS: somatic cell score.

² β -CN: β -casein; κ -CN: κ -casein; β -LG: β -lactoglobulin.

^{a,b}Superscript letters within trait and effect are significantly different ($p < .05$).

greater SCS for AA than AB variants. In agreement with our results, Albarella et al. (2020) did not find differences in milk from the agerolese breed for AB and BB variants. Differences observed across lactation and parities are in agreement with the literature, and we are not going to discuss them further.

PCA and PLS-DA

The two first principal components (PC1 = 78.0%; PC2 = 16.3%) of the PCA applied to the NIR spectra explained 94.3% of the observed variance, but were unable to cluster β -CN, κ -CN and β -LG genotypes (Figure 1). Similar results were reported by Daniloski et al. (2022) for β -CN where overlapping clusters were observed when evaluating 114 milk samples from Australian Holstein-Friesian. However, those authors improved the PCA model's performance by selecting only five samples of each variant that were separated enough to avoid overlaps. We are not aware of any other study trying to cluster samples based on protein fraction genetics applying PCA to infrared spectra.

The average spectra of the samples presented a peak around 1200 nm, a slightly broad peak between 1380 nm and 1550 nm, a small peak around 1750 nm, and two broad peaks between 1860 and 2220 nm and between 2300 and 2498 nm (Figure 2). For β -CN, VIP was >1 from 1100 nm to 1116 nm; 1190 nm to 1222 nm; 1692 nm to 1774 nm; 1888 nm to 2002 nm; 2408 nm; 2410 nm; 2414 nm to 2422 nm; and 2430 nm to 2490 nm (Figure 2). For κ -CN, the VIP was >1 from 1100 nm to 1138 nm; 1888 nm to 2018 nm; and 2380 nm to 2498 nm (Figure 2). For β -LG the, VIP was >1 from 1708 nm to 1736 nm; 1880 nm to 2084 nm; and 2382 nm to 2498 nm (Figure 2). Those VIP peaks identified in all three fractions around 1940 nm have been linked to water (Coppa et al. 2010); around 2000 nm and 2400 nm to protein and fat content, respectively (Mohamed et al. 2021); and around 2466 nm to protein (Núñez-Sánchez et al. 2016). Moreover, those VIP peaks in β -CN and κ -CN fractions around 1700 nm have been associated with fat and protein content (Mohamed et al. 2021). Additionally, the VIP peak in β -CN fraction around 1200 nm has been associated with fat content (Núñez-Sánchez et al. 2016).

The best PLS-DA model based on the accuracy of the test set was observed for β -CN (64%), followed by β -LG (53%) and κ -CN (36%; Table 3). In contrast with our results, Xiao et al. (2022) obtained a greater accuracy in discriminating A1 and A2 milk in the test set (96%) using 754 MIR spectra of Chinese Holstein cows

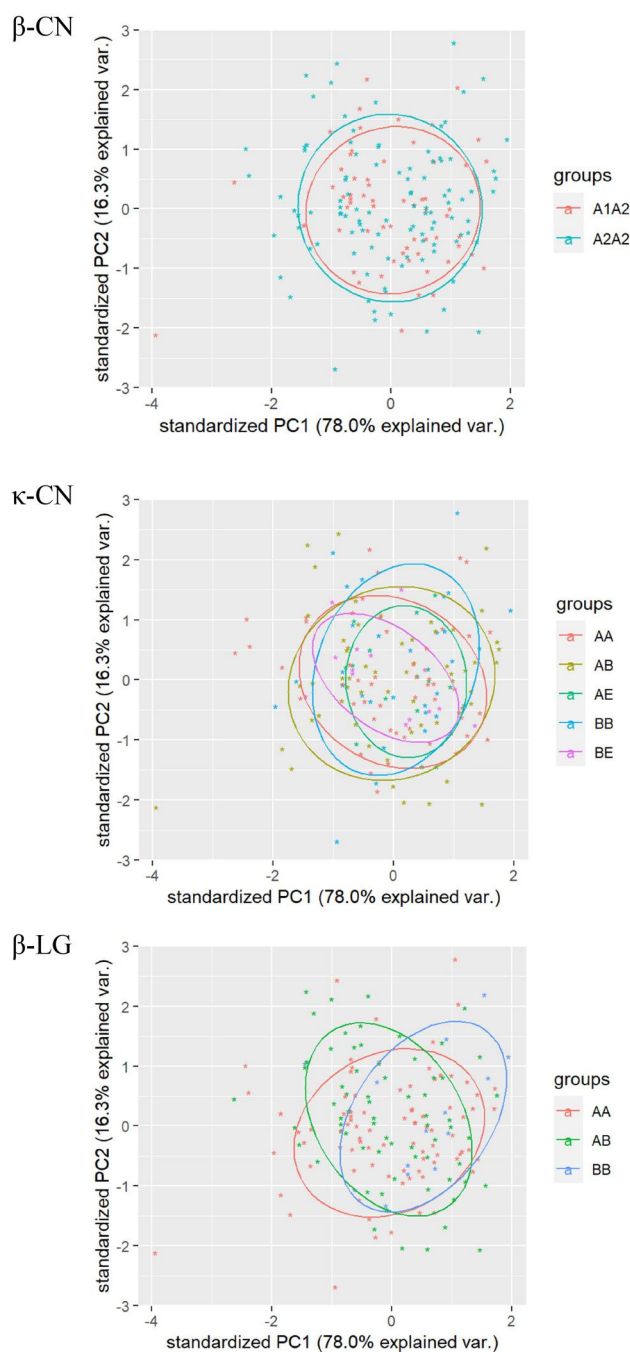


Figure 1. Standardised plot of first (PC1) and second (PC2) principal components for β -casein (β -CN), κ -casein (κ -CN) and β -lactoglobulin (β -LG). The ellipse represents 95% confidence interval.

reared on the same farm. Although Daniloski et al. (2022) observed that the PLS-DA analysis with a subset of samples (30 out of 114) was able to separate milk samples collected in the same farm into different groups based on their β -CN variants, they stressed that this could be purely by chance due to the selected samples included in the final model. Similarly, Rutten et al. (2011) also reported a better accuracy for

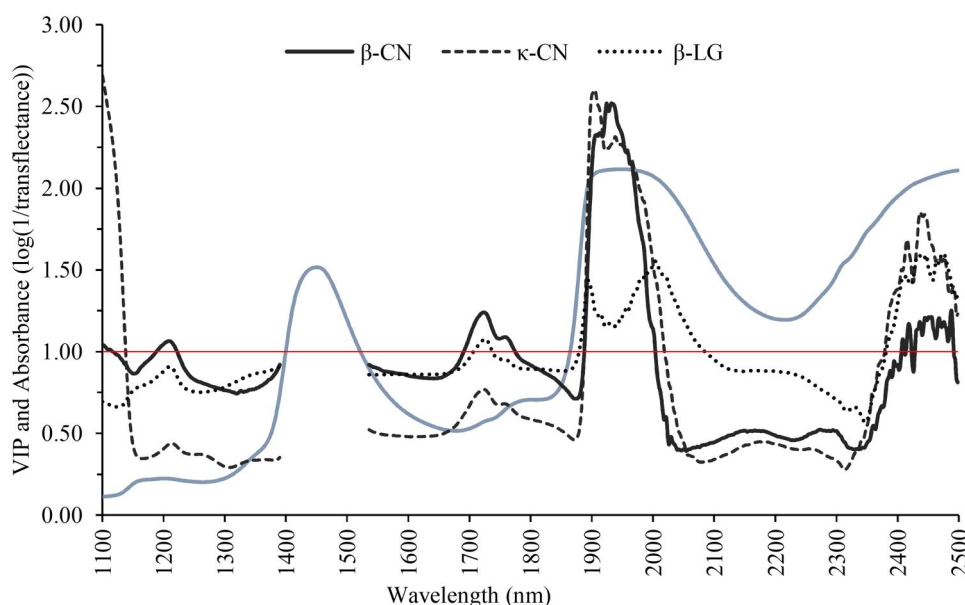


Figure 2. The absorbance of milk samples (blue), and the importance of the variable for projection (VIP) for β -casein, κ -casein and β -lactoglobulin for two components after removing signal-to-noise ratio wavelengths. The red straight line indicates the threshold $VIP = 1$.

Table 3. Partial least squares-discriminant analysis performance.

Fractions	Train set ($n = 126$)	Test set ($n = 42$)
β -CN genotypes		
Error rate	0.36	0.36
Accuracy	0.64	0.64
κ -CN genotypes		
Error rate	0.59	0.64
Accuracy	0.41	0.36
β -LG genotypes		
Error rate	0.44	0.48
Accuracy	0.56	0.52

β -CN: β -casein; κ -CN: κ -casein; β -LG: β -lactoglobulin.

β -LG variants in the test set (74%) using MIR in 3826 milk samples from Dutch Holstein-Friesian cows collected in different farms. To the best of our knowledge, there are no studies regarding κ -CN genotype discrimination analysis.

Conclusions

Our results demonstrated the difficulties of using NIRS to cluster milk samples with an unsupervised method such as the PCA based on their genotypes for β -CN, κ -CN and β -LG when milk quality is similar. The supervised method PLS-DA led to slightly better results depending on the fractions evaluated, revealing promising results for β -CN and β -LG. The best accuracy was achieved for β -CN, reaching up to 64% when dealing with A1A2 and A2A2 milk. The accuracy decreased for β -LG, but it still was above 50% considering the three variants. The worst results were obtained for κ -CN, which included five variants. We cannot discard that

the decrease in the accuracy of the models could be related to the unbalanced number of samples for the different variants of each fraction. Further research trying to equilibrate variants' representativeness within each fraction could be advisable before discarding NIRS as an authentication method. Moreover, it could be interesting to apply sample selection techniques to identify samples with different genetic variants that are farther in the PCA to also improve the models.

Acknowledgements

The authors thank the Interprofessional Dairy Association of Catalonia (ALLIC) (Barcelona, Spain) for providing the samples.

Ethics statement

Procedures adopted in the present study do not fall into the scope of an animal ethics evaluation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the Ministry of Science and Innovation of Spain with the Project PID2019-110752RB.I00. C.L. Manuelian is currently a postdoctoral researcher funded with a María Zambrano Grant from the Spanish Ministry of

Universities (funded by European Union-Next Generation EU; MZ2021-86).

ORCID

Norma S. Navarro  <http://orcid.org/0009-0009-1326-0470>
 Elena Albanell  <http://orcid.org/0000-0002-6158-7736>
 Massimo De Marchi  <http://orcid.org/0000-0001-7814-2525>
 Carmen L. Manuelian  <http://orcid.org/0000-0002-0090-0362>

Data availability statement

The data presented in this study are available free of charge for any user upon reasonable request from the corresponding authors.

References

- Albarella S, Selvaggi M, D'Anza E, Cosenza G, Caira S, Scaloni A, Fontana A, Peretti V, Ciotola F. 2020. Influence of the casein composite genotype on milk quality and coagulation properties in the endangered agerolese cattle breed. *Animals*. 10(5):892. doi: [10.3390/ani10050892](https://doi.org/10.3390/ani10050892).
- Bodnár Á, Hajzsér A, Egerszegi I, Póti P, Kuchtík J, Pajor F. 2018. A2 milk and its importance in dairy production and global market. *Anim Welf*. 14:1–7. doi: [10.17205/SZIE.AWETH.2018.1.001](https://doi.org/10.17205/SZIE.AWETH.2018.1.001).
- Bovenhuis H, van Arendonk JAM, Korver S. 1992. Associations between milk protein polymorphisms and milk production traits. *J Dairy Sci*. 75(9):2549–2559. doi: [10.3168/jds.S0022-0302\(92\)78017-5](https://doi.org/10.3168/jds.S0022-0302(92)78017-5).
- Botaro BG, Lima YVR, Aquino AA, Fernandes RHR, Garcia JF, Santos MV. 2008. Effect of beta-lactoglobulin polymorphism and seasonality on bovine milk composition. *J Dairy Res*. 75(2):176–181. doi: [10.1017/S0022029908003269](https://doi.org/10.1017/S0022029908003269).
- Caroli AM, Chessa S, Erhardt GJ. 2009. Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J Dairy Sci*. 92(11):5335–5352. doi: [10.3168/jds.2009-2461](https://doi.org/10.3168/jds.2009-2461).
- Cendron F, Franzoi M, Penasa M, De Marchi M, Cassandro M. 2021. Effects of β - and κ -casein, and β -lactoglobulin single and composite genotypes on milk composition and milk coagulation properties of Italian Holsteins assessed by FT-MIR. *Ital J Anim Sci*. 20(1):2243–2253. doi: [10.1080/1828051X.2021.2011442](https://doi.org/10.1080/1828051X.2021.2011442).
- Čítek J, Brzáková M, Hanusová L, Hanuš O, Večerek L, Samková E, Křížová Z, Hoštičková I, Kávová T, Straková K, et al. 2021. Gene polymorphisms influencing yield, composition and technological properties of milk from Czech Simmental and Holstein cows. *Anim Biosci*. 34(1):2–11. doi: [10.5713/ajas.19.0520](https://doi.org/10.5713/ajas.19.0520).
- Comin A, Cassandro M, Chessa S, Ojala M, Dal Zotto R, De Marchi M, Carnier P, Gallo L, Pagnacco G, Bittante G. 2008. Effects of composite β - and κ -casein genotypes on milk coagulation, quality, and yield traits in Italian Holstein cows. *J Dairy Sci*. 91(10):4022–4027. doi: [10.3168/jds.2007-0546](https://doi.org/10.3168/jds.2007-0546).
- Coppa M, Ferlay A, Leroux C, Jestin M, Chilliard Y, Martin B, Andueza D. 2010. Prediction of milk fatty acid composition by near infrared reflectance spectroscopy. *Int Dairy J*. 20(3):182–189. doi: [10.1016/j.idairyj.2009.11.003](https://doi.org/10.1016/j.idairyj.2009.11.003).
- Daniloski D, McCarthy NA, O'Callaghan TF, Vasiljevic T. 2022. Authentication of β -casein milk phenotypes using FTIR spectroscopy. *Int Dairy J*. 129:105350. doi: [10.1016/j.idairyj.2022.105350](https://doi.org/10.1016/j.idairyj.2022.105350).
- dos Santos Pereira EV, de Sousa Fernandes DD, de Araújo MCU, Diniz PHGD, Maciel MIS. 2020. Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms. *LWT*. 127:109427. doi: [10.1016/j.lwt.2020.109427](https://doi.org/10.1016/j.lwt.2020.109427).
- Faul F, Erdfelder E, Lang A-G, Buchner A. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 39(2):175–191. doi: [10.3758/bf03193146](https://doi.org/10.3758/bf03193146).
- Faul F, Erdfelder E, Buchner A, Lang A-G. 2009. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 41(4):1149–1160. doi: [10.3758/BRM.41.4.1149](https://doi.org/10.3758/BRM.41.4.1149).
- Franzoi M, Manuelian CL, Penasa M, De Marchi M. 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(1):791–804. doi: [10.3168/jds.2019-16916](https://doi.org/10.3168/jds.2019-16916).
- Gai N, Uniacke-Lowe T, O'Regan J, Faulkner H, Kelly AL. 2021. Effect of protein genotypes on physicochemical properties and protein functionality of bovine milk: a review. *Foods*. 10(10):2409. doi: [10.3390/foods10102409](https://doi.org/10.3390/foods10102409).
- Heck JML, Schennink A, van Valenberg HJF, Bovenhuis H, Visker MHPW, van Arendonk JAM, van Hooijdonk ACM. 2009. Effects of milk protein variants on the protein composition of bovine milk. *J Dairy Sci*. 92(3):1192–1202. doi: [10.3168/jds.2008-1208](https://doi.org/10.3168/jds.2008-1208).
- Jianqin S, Leiming X, Lu X, Yelland GW, Ni J, Clarke AJ. 2016. Effects of milk containing only A2 beta casein versus milk containing both A1 and A2 beta casein proteins on gastrointestinal physiology, symptoms of discomfort, and cognitive behavior of people with self-reported intolerance to traditional cows' milk. *Nutr J*. 15(1):35. doi: [10.1186/s12937-016-0147-z](https://doi.org/10.1186/s12937-016-0147-z).
- Manuelian CL, Vigolo V, Righi F, Simoni M, Burbi S, De Marchi M. 2021. MIR and Vis/NIR spectroscopy cannot authenticate organic bulk milk. *Italian J Anim Sci*. 20(1):1810–1816. doi: [10.1080/1828051X.2021.1954559](https://doi.org/10.1080/1828051X.2021.1954559).
- Mohamed H, Nagy P, Agbaba J, Kamal-Eldin A. 2021. Use of near and mid infra-red spectroscopy for analysis of protein, fat, lactose and total solids in raw cow and camel milk. *Food Chem*. 334:127436. doi: [10.1016/j.foodchem.2020.127436](https://doi.org/10.1016/j.foodchem.2020.127436).
- Núñez-Sánchez N, Martínez-Marín AL, Polvillo O, Fernández-Cabanás VM, Carrizosa J, Urrutia B, Serradilla JM. 2016. Near Infrared Spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. *Food Chem*. 190:244–252.
- R Core Team. 2022. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rutten MJM, Bovenhuis H, Heck JML, van Arendonk JAM. 2011. Prediction of β -lactoglobulin genotypes based on

- milk Fourier transform infrared spectra. *J Dairy Sci.* 94(8): 4183–4188. doi: [10.3168/jds.2011-4149](https://doi.org/10.3168/jds.2011-4149).
- Sanchez G. 2013. *Discriminer*: tools of the trade for discriminant analysis. R package version 0.1-29; [accessed 2023 May 5]. <https://cran.microsoft.com/snapshot/2022-03-28/web/packages/Discriminer/index.html>.
- Sola-Larrañaga C, Navarro-Blasco I. 2009. Chemometric analysis of minerals and trace elements in raw cow milk from the community of Navarra, Spain. *Food Chem.* 112(1):189–196. doi: [10.1016/j.foodchem.2008.05.062](https://doi.org/10.1016/j.foodchem.2008.05.062).
- Thiruvengadam M, Venkidasamy B, Thirupathi P, Chung IM, Subramanian U. 2021. β -Casomorphin: a complete health perspective. *Food Chem.* 337:127765. doi: [10.1016/j.foodchem.2020.127765](https://doi.org/10.1016/j.foodchem.2020.127765).
- Tsiaras AM, Bargouli GG, Banos G, Boscos CM. 2005. Effect of kappa-casein and beta-lactoglobulin loci on milk production traits and reproductive performance of Holstein cows. *J Dairy Sci.* 88(1):327–334. doi: [10.3168/jds.S0022-0302\(05\)72692-8](https://doi.org/10.3168/jds.S0022-0302(05)72692-8).
- Vu VQ. 2011. *ggbiplot*: a ggplot2 based biplot. R package version 0.55; [accessed 2023 May 5]. <http://github.com/vqv/ggbiplot>.
- Wiggans GR, Shook GE. 1987. A lactation measure of somatic cell count. *J Dairy Sci.* 70(12):2666–2672. doi: [10.3168/jds.S0022-0302\(87\)80337-5](https://doi.org/10.3168/jds.S0022-0302(87)80337-5).
- Xiao S, Wang Q, Li C, Liu W, Zhang J, Fan Y, Su J, Wang H, Luo X, Zhang S. 2022. Rapid identification of A1 and A2 milk based on the combination of mid-infrared spectroscopy and chemometrics. *Food Control.* 134:108659. doi: [10.1016/j.foodcont.2021.108659](https://doi.org/10.1016/j.foodcont.2021.108659).