Implementation of A Routine Genetic Evaluation of Milk Coagulation Properties in Italian Holstein Using a Mixed Reference Population of Bulls and Cows

F. Galluzzo[1,2], G. Visentin[2], M.P.G. de Rezende[1], J.B.C.H.M. van Kaam[1], R. Finocchiaro[1], M. Marusi[1] and M. Cassandro[1,3]

[1] Associazione Nazionale Allevatori della razza Frisona, Bruna e Jersey Italiana (ANAFIBJ), Cremona, Italia

[2] Dipartimento di Scienze Mediche Veterinarie – Università di Bologna Alma Mater Studiorum, Bologna, Italia

[3] Dipartimento di Agronomia, Alimenti, Risorse Naturali, Animali e Ambiente – Università degli Studi di Padova, Legnaro, Italia

Abstract

Cheese production is one of the most important segments of the agrifood sector in Italy and milk coagulation properties (MCP) are a key factor for an efficient cheesemaking process. Milk coagulation properties, referred as rennet coagulation time (RCT), curd firmness (a30) and curd-firming time (k20), are available from mid-infrared spectroscopy (MIRS) prediction models implemented within the official national milk recording system (LEO project, PSRN mis 16.2, AIA, 2023). Aim of this study was to assess the possibility to genetically improve MCP in Italian Holstein population and to develop a routine genetic evaluation for such traits. A multiple-trait repeatability linear animal model was employed, with RCT, a30, k20 and casein percentage (CAS) as outcome variables. Fixed effects were the interaction between year and season of recording, between parity (1,2+), year and age at calving class (7 classes) and between parity, year and days in milk class (10 classes). Random effects were contemporary groups, animal permanent environment, animal additive genetic and residuals. The models for RCT, a_{30} and k_{20} accounted also for somatic cell score as covariate. A total of 64,720 records from 150 herds, randomly sampled from the full dataset of 4,001,769 observations after edits, were used for variance components estimation using THRGIBBS1F90. The pedigree was traced back to 4 generations and was composed of 59,124 individuals. Convergence was assessed using R package BOA. The posterior mean (PM) for heritability was 0.33 for CAS, 0.11 for RCT, 0.16 for a30, and 0.15 for k20. The genetic correlation between RCT and a30 was -0.87, highlighting their antagonistic relationship; the same conclusion can be drawn from the correlation between k20 and a30 (-0.98). RCT and K20 were positively correlated (0.77). CAS was negatively genetically correlated to both RCT and k20 (-0.04 and -0.76, respectively), and positively to a30 (0.51). A SNPBLUP model was employed for estimating genomic breeding values (GEBV) using two distinct training populations: solely bulls and both bulls and cows (mixed reference population). The validation of GEBVs, conducted with complete and partial datasets (with a three-year back cutoff date for phenotypes), consistently demonstrated that employing a mixed training population results in reduced dispersion and heightened reliability for these traits. These results showed the feasibility of selecting for MCP improvement within the Italian Holstein population. Furthermore, they establish the foundation for implementing a routine genetic evaluation aimed at enhancing cheese production, utilizing a mixed reference population for SNP effects estimation.

Key words: cheesemaking, dairy cattle, genomic selection, mixed reference population, mid-infrared spectroscopy, genetic parameters

Introduction

In Italy, 77% of the milk is used for cheese production, the 55% of which is specifically utilized for crafting the 56 geographical indications and traditional specialties officially recognized by the European Union (ISMEA, 2022). In this scenario, the significance of dairy production in the Italian agrifood sector is readily evident. The importance of milk coagulation properties (MCP), referred as rennet coagulation time (RCT, minutes), curd firming time (k20, minutes) and curd firmness (a30, millimeters), for an efficient cheesemaking process has been widely discussed (Riddel-Laurence et al, 1989; Wedholm et al, 2006; Pretto et al, 2012). Moreover, MCP were demonstrated to be moderately heritable, indicating the potential for improvement through genetic selection (Cassandro et al, 2008; Visentin et al, 2017). Milk coagulation properties are now available at the National Breeders Association of Italian Holstein, Brown Swiss and Jersey (ANAFIBJ) from mid-infrared spectroscopy (MIRS) prediction models implemented within the official national milk recording system (LEO project, PSRN mis 16.2, AIA, 2023). Considering the aforementioned, the aim of this study was the implementation of a routine genetic and genomic evaluation of MCP for the Italian Holstein breed. Furthermore, given the substantial availability of genotypes of Italian Holstein cows, the feasibility of incorporating them into the training population for the estimation of the SNP effects was explored.

Materials and Methods

Data editing

The input dataset was composed of 6.7 million records from 2017 onwards. Only records from regions that provided a consistent data flow were kept (10 regions out of 20). Records from parity 1 to parity 5 and from 5 to 405 days in milk (DIM) were considered. Regarding MCP traits, accepted range of values for RCT, k20 and a30 where five to 60 minutes, one to 20 minutes and five to 60 millimeters respectively: all records out of these ranges were removed as obvious errors. In order to detect laboratory measurement anomalies, isolation forest algorithm implemented in the python module Scikit-learn was employed (Pedregosa et al, 2011). Briefly, reference values from Visentin et al, 2015 were used for the phenotypic correlations between the three traits: -0.73 for RCT-a30, 0.80 for rct-k20 and -0.79 for k20-a30. Phenotypic correlations within herd-year-test-day (HTD) groups were calculated: all milk samples collected from the same herd in the same day are processed in the same laboratory. All the HTD groups with an anomalous value compared to the reference have been excluded. All test-day observations had to have a record for casein percentage (CAS) too to be included in the analysis. Finally, only herd-year-season of recording groups with at least 20 contemporaries were kept. Pedigree was traced back to four generations.

Statistical model

A multiple trait repeatability linear animal model was used, with CAS, RCT, a30 and k20 as correlated dependent variables.
The model for CAS was the following:

$$CAS_{ijklmnop} = hys_i + S_j{*}Y_k + DIM_l{*}PARC_m{*}Y_k + AGEC\_PAR_n{*}Y_k + a_o + pe_o + e_{ijklmnop}$$

with $CAS_{ijklmnop}$ as the p*th* phenotypic observation of casein percentage. Fixed effects were $S_j{*}Y_k$ as the crossed effect of season j by year k, $DIM_l{*}PARC_m{*}Y_k$ as the l*th* days in milk class (10 classes of 40 days) by parity class m (3 classes: 1, 2, 3+) and year k, $AGEC\_PAR_n{*}Y_k$ as the n*th* age at calving by parity class (9 classes: 3 age at calving classes for every parity class) by year k. Random effects were hys_i as the i*th* contemporary group for herd-year-season of recording, a_o as the additive genetic effect of the o*th* animal,

pe_o as the permanent environmental effect of the o*th* animal and $e_{ijklmnop}$ as the residual of observation p. The same model with the addition of the fixed linear regression of somatic cell score was applied to MCP traits.

Variance components estimation, genetic and genomic evaluation, approximate genetic correlations

Variance components estimation was performed with the software THRGIBBS1F90 (Misztal et al, 2002) on a sample of 64,720 animals (150 herds). Convergence was assessed with R package BOA, Bayesian output analysis (Smith, 2007). Conventional estimated breeding values (EBVs) were estimated with MiX99 software (MiX99 development team, 2012). Genomic evaluation was performed with a SNPBLUP model using GS3 software (Legarra et al, 2007). For estimated deregressed proofs (EDPs), the method from Degano et al (2009) was applied. A conventional quality control was applied to SNP data. For the imputation process, PedImpute software was used (Nicolazzi et al, 2011). Approximate genetic correlations were calculated as Pearson correlation coefficients between genomic estimated breeding values (GEBVs) of 87,569 heifers born after 2016.

Genomic validation

Genomic validation was performed as described in Finocchiaro et al (2012) and Galluzzo et al (2022). Briefly, two datasets were used for EBVs estimation: one full (with records up to 2308 run) and one reduced (with a 3-years back cutoff date). For both sets of EBVs, EDPs were calculated and used as pseudo-phenotypes for SNP effects estimation. Bulls with daughters in the full datasets but without in the reduced one were selected as validation bulls. Finally, a linear regression with EDPs of validation bulls from the full run as dependent variable and their direct genomic values (DGVs) from the reduced run as the independent one was fitted. The validation process was performed either using a training population composed of bulls only and of bulls and cows. Parameters considered for the comparison were the dispersion coefficient and the reliability of the linear regression model.

Results & Discussion

The dataset after edits was composed of 4,001,769 records: phenotypes averaged 2.72%, 25.40 minutes, 20.73 millimeters and 7.29 minutes for CAS, RCT, a30 and k20 respectively. The results of variance components estimation are listed in Table 1. Posterior mean for heritability was moderate to high for CAS and moderate for MCP traits: genetic correlations were high for all combinations of traits except for CAS and RCT.

Table 1. – Results of variance components estimation.

	CAS	RCT	a30	k20
CAS	0.33(0.01)	-0.04	0.51	-0.67
RCT		0.11(0.01)	-0.87	0.77
a30			0.16(0.01)	-0.98
k20				0.15(0.01)

Posterior means of heritability on diagonal with posterior standard deviations in parentheses, genetic correlations above diagonal.

The results of the genomic validation are listed in Table 2. Adding the females, the training population increased by 40,478 individuals. The mixed training population performed better than the one composed by only bulls for both the parameters considered, dispersion coefficient and reliability of the model.

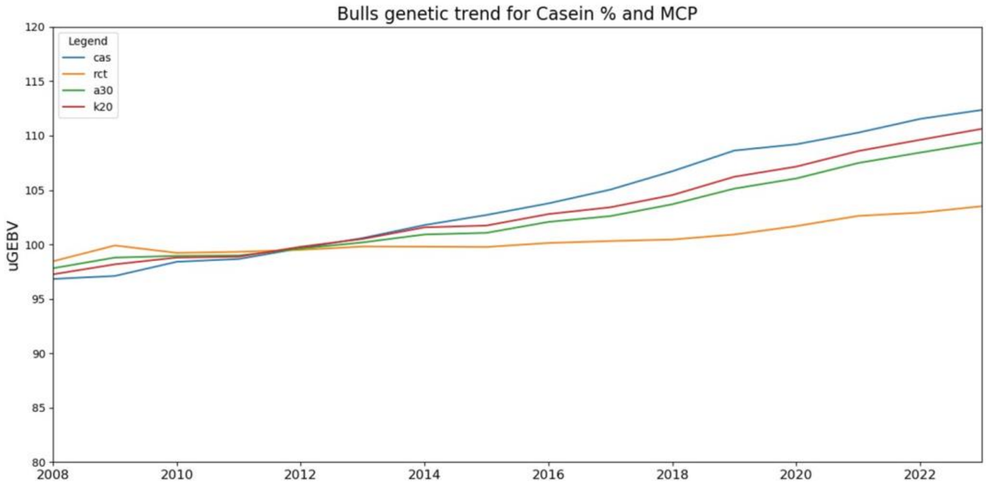

Figure 1. Bulls' genetic trend by birth year. µGEBV=average GEBV.

Regarding dispersion, a mean coefficient of 0.91, compared to a mean of 1.28 for the bulls only, was detected for the mixed reference population; regarding reliability, a mean reliability of 0.48 was detected for the bulls' reference population while a mean of 0.76 resulted for the mixed one. These evidences suggest that the inclusion of females in the reference population would be beneficial for MCP traits.

Based on these results, the mixed reference population was chosen for the subsequent analyses.

Table 2. Results of genomic validation.

	Training	Animals	b	r^2
CAS	B	3,276	1.205	0.452
	M	43,754	0.898	0.790
RCT	B	3,276	1.359	0.421
	M	43,754	0.925	0.737
a30	B	3,276	1.319	0.478
	M	43,754	0.911	0.767
k20	B	3,276	1.246	0.459
	M	43,754	0.895	0.763

B=bulls only; M=mixed; b=dispersion coefficient; r^2=model reliability.

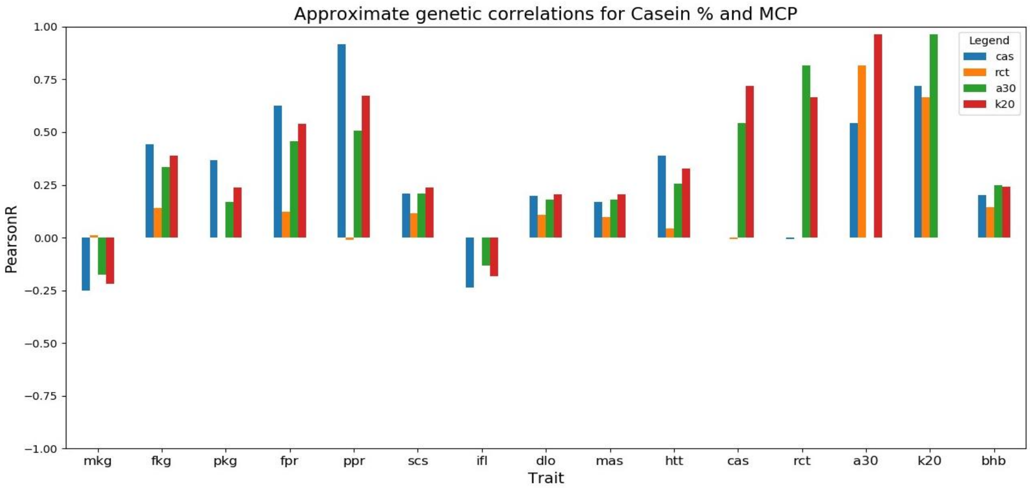

Figure 2. Approximate genetic correlations for MCP traits. Mkg=milk yield, fkg=fat yield,pkg=protein yield, fpr=fat percentage, ppr=protein percentage, scs=somatic cell score, ifl=interval first-last insemination, dlo=direct longevity, mas=mastitis, htt=heat tolerance, cas=casein percentage, rct=rennet coagulation time, a30=curd firmness, k20=curd firming time, bhb=ketosis.

The genetic trend of bulls' GEBVs by birth year is represented in Figure 1: an increasing trend is evident for CAS, a30 and k20. The trend for RCT is increasing too, but in a milder way compared to the other traits.

The approximate genetic correlations are represented in Figure 2. For all the four traits analyzed, this study revealed a null or favorable approximate genetic correlation with all the traits considered. The only exception is milk yield: for this trait, a negative and unfavorable correlation was detected with CAS, a30 and k20. The strong favorable genetic correlations of CAS, a30 and k20 with protein yield and percentage may explain their increasing genetic trend. In contrast, the milder correlations calculated for RCT can be the motivation of its less pronounced trend.

Conclusions

In conclusion, this study increased the knowledge about the genetic aspects of MCP in the Italian Holstein population, revealed the possibility to genetically improve the breed for these traits and highlighted the benefits of including females in the reference population for SNP effects estimation. A routine genetic evaluation of CAS and MCP traits will be soon implemented in Italy for the Italian Holstein breed.

Acknowledgments

This study was supported by “Latteco2 project, sottomisura 10.2 of the PSRN-Biodiversity 2020–2023” (MIPAAF. D.M. no. 465907 del 24/09/2021, project unique code J12C21004080005).

References

Cassandro, M., Comin, A., Ojala, M., Dal Zotto, R., De Marchi, M., Gallo, L., Carnier, P., and Bittante, G. 2008. Genetic parameters of milk coagulation properties and their relationships with milk yield and quality traits in Italian Holstein cows. *J. Dairy Sci.* 91, 371-376. https://doi.org/10.3168/jds.2007-0308.

Degano, L., Jansen, G., Finocchiaro, R., Rossoni, A., and Vicario, D. 2016. Hybrid One-Step Genomic Evaluation System for the Italian Simmental Breed. *Interbull Bulletin.* 50, 24-28. https://journal.interbull.org/index.php/ib/article/view/1624/1626.

Finocchiaro, R., van Kaam, J.B.C.H.M. and Biffani, S. 2012. The Genomic Selection System in Italian Holstein. 63rd Annual Meeting of the European Federation of Animal Science, Bratislava, Slovensko.

Galluzzo, F., van Kaam, J.B.C.H.M., Finocchiaro, R., Marusi, M., Tsuruta, S., and Cassandro, M. 2022. Estimation of milkability breeding values and variance components for Italian Holstein. *J. Dairy Sci.* Comm. 3, 180–184.https://doi.org/10.3168/jdsc.2021-0167.

ISMEA. 2022. Rapporto ISMEA – qualivita. https://doi.org/10.1017/S0022029912000453.

Legarra, A., Ricardi, A., and Filangi, O. 2011. *GS3*: Genomic Selection, Gibbs Sampling, Gauss-Seidel (and BayesCp).

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D.H. 2002. BLUPF90 and related programs (BGF90). 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=28-07.pdf.

MiX99 Development Team. 2015. MiX99: A software package for solving large mixed model equations.

Nicolazzi, E.L., Biffani, S. and Jansen, G. 2013. Short communication: Imputing genotypes using PedImpute fast algorithm combining pedigree and population information. 2011. *J. Dairy Sci*, 96, 2649–2653. https://doi.org/10.3168/jds.2012-6062.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res*. 12, 2825-2830.

Pretto, D., De Marchi, M., Penasa, M., and Cassandro, M. 2013. Effect of milk composition and coagulation traits on Grana Padano cheese yield under field conditions. *J. Dairy Res*. 80, 1-5.

Riddel-Laurence, S., and Hicks, C. 1989. Effect of Curd Firmness on Stirred Curd Cheese Yield. *J. Dairy Sci.* 72, 313-321. https://doi.org/10.3168/jds.S0022-0302(89)79112-8.

Smith, B.J. 2007. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *J. Stat. Softw*. 21, 1-37. https://doi.org/10.18637/jss.v021.i11.

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

Visentin, G., McDermott, A., McParland, S., Berry, D.P., Kenny, O.A., Brodkorb, A., Fenelon, M.A., and De Marchi, M. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. *J. Dairy Sci*. 98, 6620-6629. https://doi.org/10.3168/jds.2015-9323.

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