Pitfalls and opportunities of genetic and genomic evaluation in the Buffalo species: experiences from Italy

S. Biffani[1], M. Gómez[2] and A. Cesarani[3]

[1] *Institute of Agricultural Biology and Biotechnology, National Research Council, Milan, Italy*

[2] *Italian National Association of Buffalo Breeders, Caserta, Italy*

[3] *Dipartimento di Agraria, University of Sassari, Sassari, Italy*

Abstract

Thanks to the iconic Mozzarella di Bufala Campana cheese, the economic interest in Buffalo breeding has steadily increased making the Italian Mediterranean Buffalo (IMB), a river buffalo breed *(Bubalus bubalis bubalis)*, one of the most important dairy species in the country. In 2018 a new selection index, namely the IBMI, was introduced. New breeding objectives were identified, including milk contents, udder morphology and feet and legs. However, in IMB the use of artificial insemination is still moderate (around 30-40%) and it poses additional problems in developing an accurate BLUP evaluation. We do expect a more extended use of DNA testing and the implementation of genomic selection approaches. However, missing information in the pedigree will still be present and even genomic selection will be faced with the same problem. The use of genetic groups in the BLUP evaluation to account for genetic differences among unknown parents has been recently introduced. This enhancement was based on the results obtained in a study conducted using 15 linear type traits from 7,714 buffalo cows. Moreover, a first test on the feasibility of genomic selection using a ssGBLUP in IMB was developed. Results of this study, even if based on a small number of animals, showed that the inclusion of genotypes of females can improve breeding values accuracy in the IMB. New genotypes have been recently added and additional tests are in progress. Future data recording will include additional traits, related to health and milk quality.

Key words: Buffalo, selection index, genetics groups, genomic, single-step

Introduction

The Water Buffalo (*Bubalus bubalis*) is a large bovid mainly distributed in the Asian continent, where the 97% of its world population is concentrated (FAO, 2020). The name 'water buffalo' is due to its adaptation to flooded or swampy areas, where it partially submerges and walks on the bottom mud without difficulty. Two types of water buffalo exist, the river (*Bubalus bubalis bubalis*) and the swamp (*Bubalus bubalis carabanensis*) buffalo. The two types have different number of chromosomes, 50 for the river and 48 for the swamp, respectively, geographical distribution, and they are characterised by phenotypic and genetic differences. River buffaloes are farmed mainly in the west, from India to Europe, whereas the swamp type can be found mostly in eastern Asian Countries (Iamartino et al. 2017). The Mediterranean area, where the rest of the world population is raised (3%), has been historically characterized by these optimal rearing conditions. In the European continent, only the 0.2% of its world population can be found and about 93% of these animals are raised in south-central Italy (Neglia et al., 2020). Total census in Italy increased considerably, making IMB breed one of the most important dairy species in the country. The main zootechnical interest is to produce iconic traditional dairy products, like

the Mozzarella di Bufala Campana (Boselli et al., 2020). Thanks to the physical-chemical properties of its milk - high concentration in protein and fat (FC ~ 8 %) - and favourable coagulation (Costa et al., 2020b), IMB has a great economic impact in the Italian food industry (ISMEA, 2020). Costa et al., (2020a, 2020b) refers to the impressive increase in heads that the IMB has experienced in the last 15 years, coupled with an increase in terms of kilos of cheese produced, a larger herd size, a constant increase in registered herds, as well as the rise in milk price. Therefore, the economic interest in this specie makes it necessary to develop new innovative tools to improve breeding process. In 1997 a first national selection scheme was implemented. It was based on a traditional progeny testing and a BLUP genetic evaluation which included only milk yields. A first selection index, namely the PKM, was also developed. The breeding objective of the PKM was the mozzarella cheese yield using as selection criteria milk yield, fat and protein yields (Rosati and Van Vleck, 2002). Although a positive selection on milk yields was observed, the PKM had an unfavourable effect on the genetic trend for fat and protein. Moreover, no emphasis was given to health related or functional traits. Attending to mozzarella cheese manufacturers and farmers' community, who both asked for a more balanced breeding objective which included also health related traits, a more comprehensive aggregate selection index, namely the IBMI, was developed and introduced at the end of 2018. The IBMI breeding objectives included milk yield and components, feet and legs, and udder traits. The prediction of breeding values (EBVs) constitutes an integral part of most breeding programs which are based on two fundamental pillars: phenotypic data and genealogical information. However, if animals with unknown parents are present in the pedigree, bias in the prediction of EBV is expected. However, natural mating is still a common reproductive strategy that can increase the

proportion of missing pedigree information. The inclusion of genetic groups in the EBV estimation is a solution. In response to this need, in 2020, the use of genetic groups was implemented in the official routine of the genetic evaluation of the IMB. However, the use of new selective tools such as genomic selection is on the rise in other species and the IMB has not wanted to delay the use of these tools, which will accelerate genetic progress and improve the accuracy of EBVs in Italian buffalo. For all this, the objective of this review is to describe the consolidated advances and future perspectives in the Italian Mediterranean Buffalo breeding.

Materials and Methods

Aggregate Selection Index (IBMI)

Different selection indices and breeding objectives were constructed using selection index theory (Hazel, 1943).

The selection indices investigated in this study considered the following traits: milk yield (MY), fat content (F%), protein content (P%), mozzarella cheese production (MCY) estimated as $116.615 + 2.015 * (P\% * F\%) + 2.929 \times (P\%)^2$, and two composite traits feet and legs (FL) and mammary system (MS). Three alternative breeding objectives were formulated according to relative weights given to MY, MCY, FL and MS. The different breeding objectives and relative weights considered are shown in Table 1.

Table 1. – Relative emphasis (%) on traits included in alternative breeding scenario for the IMB

Breeding Scenario	Traits in breeding objective[a]			
	MY	MCY	FL	MS
S1	5	45	30	20
S2	30	35	15	20
S3	45	45	5	5

[a] MY = Milk Yield (kg/270d), MCY = Mozzarella Cheese Production, FL = Feet & Legs, MS = Mammary System

The three scenarios were formulated in order to give an increasing relative emphasis on milk

yield, but keeping the emphasis to MCY above 30%. The dataset used included phenotypic information from 7,199 buffalo cows and a pedigree with 19,574 animals. A multi-trait animal model was fitted using a Bayesian implementation via Gibbs sampling. Genetic response to selection for each trait considered in the selection indices was calculated with a deterministic procedure (Cameron, 1997).

Genetic groups

To evaluate the use of genetic groups (GG), two genetic grouping strategies were used: 1) the year of birth (Y); 2) genetic clustering (GC). The effect of the GC implementation was evaluated on variance components (VC) and EBV for 5 composite and 10 linear traits. The data was composed of 7,714 buffalo cows plus a pedigree file that includes 18,831 individuals. The single-trait animal model with groups was used to estimate VC (AIREMLF90) and EBV (BLUPF90) (Misztal et al., 2002).

As a first step, VC and EBV were estimated using the official DNA verified pedigree. Successively, the EBVs were re-estimated using modified pedigrees with two different proportions of missing pedigrees (30 or 60% buffalo with records) and using the 2 grouping strategies (Y30/Y60/GC30/GC60).

Results from different scenarios were compared based on descriptive statistics of VC, Pearson's correlations between EBVs grouped by animal status (i.e., bulls with at least 10 daughters, buffalo cows with or without progeny), re-rankings of first 10 bulls, efficiency of selection and genetic trends, estimated by the linear regression of EBV on year of birth (Gómez et al., 2021).

Genomic selection

In order to investigate the feasibility of genomic selection in the IMB, a total of 498 animals were genotyped at 49,164 loci. Test day records (80,417) of milk (MY), fat (FY) and protein (PY) yields, from 4,127 cows born between 1975 and 2009, were analysed in a three-trait animal model. Cows born in 2008 and 2009 with phenotypes and genotypes were selected as validation animals (n=50). Breeding values were computed using BLUP and ssGBLUP, using VC estimated from BLUP. ssGBLUP was applied in five scenarios, each with a different number of genotypes available: (A) bulls (35); (B) validation cows (50); (C) bulls and validation cows (85); (D) all genotyped cows (463); (E) all genotypes (498). Model validation was performed using the LR method (Legarra and Reverter, 2018): correlation, accuracy, dispersion, and bias statistics were calculated (Cesarani et al., 2021).

Results & Discussion

Aggregate Selection Index (IBMI)

Mozzarella cheese production had an unfavourable genetic correlation with MY (-0.54), a null correlation with MS (-0.01), and a favourable correlation with FL (0.25), F% (0.87) and P% (0.96). Those values had an impact on the selection response for all traits considered in the three scenarios which are shown in Figure 1.

The breeding objective S1 showed large and positive genetic response in terms of MCY, F and P%. However, MY was strongly penalized, with a negative genetic response. The breeding objective S2 had the best results for health-related traits, namely FL and MS. In this scenario, 35% of the relative emphasis in the breeding objective was given to health-related traits. The last scenario, S3 showed the best overall results with a positive genetic response for all traits included in the breeding objective ranging from 0.02 (MCY) to 0.21 (P%). The scenario S3 was eventually chosen as the official new selection index for the IMB, namely IBMI.

The new selection index for the Italian Mediterranean buffalo (IBMI) is extremely different from the former selection index (PKM).
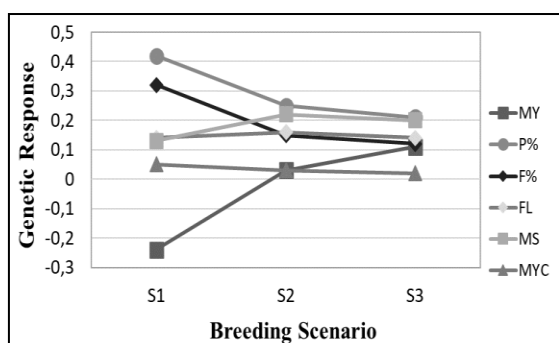
Figure 1. Genetic response (expressed in genetic standard deviations) for the three-breeding objective

Introducing it as a new selection tool has actually been a cornerstone of buffalo breeding in Italy, affecting bulls and dam rankings. Indeed, the best individuals are no longer those with the highest milk EBV, but those that best match the need for a more balanced breeding objective (production, quality, and functionality).

Genetic groups

In the last three years the IMB has experienced an exponential increase in term of registered animals in the Herd Book. Consequently, IMB is facing a situation where phenotypic data are available for many animals, but some animals lack complete genealogical data. Records from individuals without pedigree information has been excluded from the genetic evaluation or assumed to have an unknown sire.

Table 2. – Heritability for the composite traits obtained with the use of GG

Scenario	Trait[a]				
	FS	ST	FL	UT	YP
GOLD	0.22	0.25	0.14	0.15	0.20
Y30	0.21	0.23	0.14	0.14	0.18
Y60	0.21	0.22	0.14	0.14	0.17
GC30	0.20	0.24	0.14	0.17	0.18
GC60	0.19	0.22	0.16	0.22	0.17

[a] final score (FS), structure (ST), feet and legs (FL), yield potential (YP) and udder teat (UT)

With the use of the GG, the estimates of VC and EBV are quite close to the "GOLD" (original pedigree) estimates (Table 2 and 3). These results support the efficiency of the methodology to estimate the true parameters.

Table 3. – Heritability for the linear traits obtained with the use of GG

Scenario	Trait[a]				
	STAT	BD	BL	FA	FUA
GOLD	0.35	0.19	0.23	0.09	0.17
Y30	0.33	0.18	0.23	0.10	0.16
Y60	0.32	0.18	0.23	0.10	0.16
GC30	0.35	0.17	0.23	0.10	0.16
GC60	0.38	0.15	0.26	0.10	0.14

Scenario	RUW	UD	TP	TL	BCS
GOLD	0.21	0.25	0.13	0.26	0.16
Y30	0.21	0.22	0.12	0.26	0.13
Y60	0.19	0.22	0.13	0.26	0.14
GC30	0.18	0.23	0.13	0.25	0.15
GC60	0.18	0.22	0.14	0.25	0.14

[a] stature (STAT), body depth (BD), body length (BL), foot angle (FA), fore udder attachment (FUA), rear udder width (RUW), udder depth (UD), teat placement (TP), teat length (TL), and body condition score (BCS)

As expected, the accuracy of the EBV decreased when an increased proportion of missing pedigree (60%) was observed. However, when the proportion of missing pedigree was 30%, the average percentage point drop in accuracy was negligeable. We can therefore hypothesize that the contemporary use of the available pedigree information and of the most appropriate GG strategy will mitigate the loss in accuracy of the EBV due to missing pedigree information. On the other hand, Pearson's correlations between EBVs were generally high in all clustering scenarios. However, Y30 and GC30 scenarios showed the highest correlations (Table 4). In addition, the Y30 genetic group strategy showed the highest coefficients for buffalo cows, while for bulls GC30 was the most appropriate. This result was somewhat expected because the strategy based on the

hierarchical clustering is strictly related to the relationship matrix, i.e., on the pedigree information.

Table 4. – Average correlation across traits from different scenarios

rg	Y30	Y60	GC30	CG60
Buffalo cows[a]	0.91	0.84	0.88	0.79
AI bulls	0.89	0.76	0.92	0.81

[a] Buffalo cows with record

The inclusion of GG in the genetic evaluation could have unpredictable but substantial effects on the estimated genetic trend (Saavedra, 2019). Furthermore, the exclusion of genetic groups or having paternities with "phantom" parents could lead to biased estimates of selection response (Theron et al., 2002). In our study, these expectations are met, observing how the cumulative genetic trends without genetic groups were slightly lower than those estimated with the Y30/Y60 genetic group. Upward trends may indicate that the grouping type "year of birth" may be comparable to those obtained in GOLD. To see this trend, a trait was chosen (UT) as shown in Figure 2.
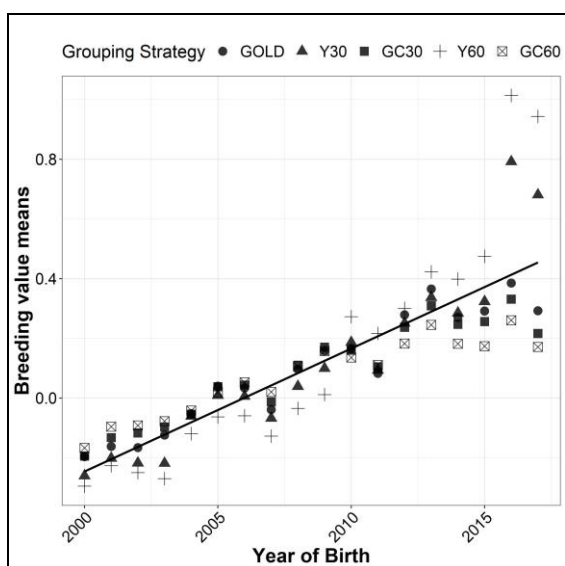

Figure 2. Genetic trend by year of birth for the UT composite trait, according to the different genetic group

The prediction of the genetic value with models that consider the uncertainty in paternity have been shown to have better precision (Cardoso and Tempelman, 2003, Sapp et al., 2007, Shiotsuki et al., 2013, Carneiro et al., 2017, Shiotsuki et al., 2018). Its effectiveness depends on the definition of the grouping strategy (Petrini et al., 2015), which requires prior knowledge of: (a) the selection process of the breed, (b) the sources of genetic variation present in the population, (c) the intensity of selection or the generational interval. Therefore, the use of genetic groups can be considered an effective alternative in the absence of relationship data for VC and EBV estimate.

Genomic selection

Genomic selection in the IMB is possible as in other species of zootechnical interest. It was recently developed the first test on the feasibility of genomic selection using a ssGBLUP in IMB. In this study, the usefulness of genomic models for the estimation of VC and the prediction of breeding values for milk production traits in the IMB is evaluated. In particular, the traditional pedigree based BLUP was compared with the single-step genomic BLUP (ssGBLUP).

Heritability estimates were very similar in both considered models; lower values of genetic parameters were obtained for FY (Table 5). Moreover, ssGBLUP provided smaller estimates and smaller standard errors for genetic correlations (Table 5).

The correlation between breeding values of candidate cows estimated with BLUP and ssGBLUP were 0.96, 0.95, and 0.95 for MY, FY, and PY, respectively. Moreover, the rank correlations between BLUP and ssGBLUP breeding values was high (0.98), suggesting that the methods are comparable.

Table 5. – Heritability (diagonal) and genetic correlations (above diagonal) estimated using two different methods, BLUP and single-step GBLUP

Method	Trait[a]		
	MY	FY	PY
BLUP			
MY	**0.25 ± 0.02**	0.92 ± 0.02	0.98 ± 0.04
FY		**0.16 ± 0.01**	0.96 ± 0.04
PY			**0.25 ± 0.01**
ssGBLUP			
MY	**0.23 ± 0.01**	0.92 ± 0.01	0.97 ± 0.01
FY		**0.15 ± 0.01**	0.95 ± 0.01
PY			**0.23 ± 0.01**

[a] MY=Milk yield; FY= Fat yield; PY=Protein yield

Results from LR validation for ssGBLUP, suggest that correlations and prediction accuracies increased as the number of genotypes included in the analysis increased. The stability of the model is shown by correlation statistics (LR method), which in the best scenario (i.e., with all 498 genotypes) ranged from 0.81 for FY to 0.83 form MY. Interesting results have been obtained for the dispersion statistics, even if a clear pattern across the different scenarios could not be detected. ssGBLUP dispersion statistics were slightly worse than BLUP ones, but they were within a 10% difference from the ideal value (1). Indeed, Tsuruta et al., (2011) reported that values within 15% are acceptable.

The WssGBLUP is a weighted version of ssGBLUP in which different weights are used for each SNP. Giving more weights to some SNPs allows the model to take into account the presence of major genes or QTL that affect the trait of interest. This is particularly appealing for species like buffalo, where milk protein composition or cheesemaking traits are key features. Several publications (Dagnachew et al., 2011, Teissier et al., 2018, Pizarro Inostroza et al., 2020) concluded that the inclusion of detailed information on major genes (e.g., *DGAT1* for fat content or as1 casein for protein content) with additive, dominance, and epistatic effects in the genetic evaluation could improve both the statistical power of the model and the accuracy of breeding values for milk yield and composition compared to a model without these effects.

Indeed, the slight improvement in accuracy with WssGBLUP in our study may indicate the presence of important markers associated with these traits. However, only one SNP showed a p-value of its effect passing the significance threshold. The low number of genotypes involved in the present study could have been influenced the results; anyway, further studies will be needed to determine the genomic make up of milk production in Italian buffalos.

It should be also pointed out that the use of WssGBLUP resulted in an increase of prediction accuracy but together with an increase of breeding value dispersion. This latter statistic should be as much as closer to 1, which means that breeding values are not under or over estimated. In our results, the dispersion was 0.90, 0.93, and 0.92 for ssGBLUP, compared to 0.86, 0.90, and 0.89 for WssGBLUP for MY, FY, and PY, respectively. In this scenario, one would have to establish a balance between the highest gains in accuracy and the least bias and dispersion of the breeding values when choosing the model for the official evaluations.

Conclusions

The new selection index for the Italian Mediterranean buffalo provides animals that best match the need for a more balanced breeding objective (production, quality, and functionality). This index can be adjusted especially as concerns the phenotypes used to improve the quality of milk for mozzarella cheese production. Several studies have already showed that milk coagulation properties like rennet coagulation time or curd firmness are important traits for enhancing the efficiency of the dairy industry.

Pedigree completeness is a fundamental requirement of any genetic evaluation. We do expect a more extended use of DNA testing which will eventually increase the

implementation of genomic selection approaches in Buffalo species as well. However, missing information in the pedigree will still be present and even genomic selection will be faced with the same problem. In the IMB population the best results were obtained when grouping was based on the year of birth. These findings confirmed the possibility of developing a genetic evaluation in populations with uncertain paternities without the need to exclude data or to use only a select of the available population.

The use and future application of ssGBLUP in Italian buffalo provided interesting information. Variance components estimated using BLUP and ssGBLUP were similar, but the latter model showed lower standard deviations for genetic correlations. To implement a genomic evaluation in the Italian Buffalo, it is important to emphasize that female genotyping could increase the accuracies of the prediction compared to both pedigree-based and genomic evaluations with only male genotypes models.

Acknowledgments

References

Boselli, C., M. De Marchi, A. Costa, and A. Borghese. 2020. Study of Milkability and Its Relation With Milk Yield and Somatic Cell in Mediterranean Italian Water Buffalo. Front Vet Sci 7(432):432.

Cameron, N. D. 1997. Selection indices and prediction of genetic merit in animal breeding. CAB international.

Cardoso, F. F. and R. J. Tempelman. 2003. Bayesian inference on genetic merit under uncertain paternity. Genet Sel Evol 35(5):469-487.

Carneiro, F. F. D., A. M. B. O. Lôbo, L. P. d. Silva, K. d. M. Silva, A. V. Landim, and R. N. B. Lôbo. 2017. Genetic evaluation is possible on community pastoral small ruminant flocks in the presence of multiple sires and uncertain of paternity. Small Ruminant Res. 151:72-81.

Cesarani, A., S. Biffani, A. Garcia, D. Lourenco, G. Bertolini, G. Neglia, I. Misztal, and N. P. P. Macciotta. 2021. Genomic investigation of milk production in Italian buffalo. Ital. J. Anim. Sci. 20(1):539-547.

Costa, A., G. Neglia, G. Campanile, and M. De Marchi. 2020a. Milk somatic cell count and its relationship with milk yield and quality traits in Italian water buffaloes. J. Dairy Sci. 103(6):5485-5494.

Costa, A., R. Negrini, M. De Marchi, G. Campanile, and G. Neglia. 2020b. Phenotypic Characterization of Milk Yield and Quality Traits in a Large Population of Water Buffaloes. Animals 10(2):327.

Dagnachew, B. S., G. Thaller, S. Lien, and T. Ådnøy. 2011. Casein SNP in Norwegian goats: additive and dominance effects on milk composition and quality. Genet Sel Evol 43(1):31.

FAO. 2020. Food and Agriculture Organization (FAO). Live Animals. Available online at: http://www.fao.org/faostat/en/#data/QA [Accessed September 28, 2020).

Gómez, M., D. Rossi, R. Cimmino, G. Zullo, Y. Gombia, D. Altieri, R. Di Palo, and S. Biffani. 2021. Accounting for Genetic Differences Among Unknown Parents in Bubalus bubalis: A Case Study From the Italian Mediterranean Buffalo. Front. Genet. 12(82).

Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics 28(6):476-490.

Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltes JE, Biffani S, Sonstegard TS, Schroeder SG, Ajmone-Marsan P, et al. 2017. Design and validation of a 90K SNP genotyping assay for the water buffalo (Bubalus bubalis). PLOS One. 12(10):e0185220

ISMEA. 2020. Istituto di servizi per il mercato agricolo alimentare è un ente pubblico economico nazionale sottoposto alla vigilanza del Ministero delle politiche agricole alimentari e forestali. http://www.ismeamercati.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/5186 [Accessed Sept 15, 2020].

Legarra, A. and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet Sel Evol 50(1):53.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. Lee. 2002. BLUPF90 and related programs (BGF90). Pages 743-744 in Proc. Proceedings of the 7th world congress on genetics applied to livestock production.

Neglia, G., D. de Nicola, L. Esposito, A. Salzano, M. J. D'Occhio, and G. Fatone. 2020. Reproductive management in buffalo by artificial insemination. Theriogenology 150:166-172.

Petrini, J., S. F. N. Pertile, J. P. Eler, J. B. S. Ferraz, E. C. Mattos, L. G. G. Figueiredo, and G. B. Mourão. 2015. Genetic grouping strategies in selection efficiency of composite beef cattle (Bos taurus × Bos indicus)1. J. Anim. Sci. 93(2):541-552.

Pizarro Inostroza, M. G., V. Landi, F. J. Navas González, J. M. León Jurado, J. V. Delgado Bermejo, J. Fernández Álvarez, and M. D. A. Martínez Martínez. 2020. Integrating Casein Complex SNPs Additive, Dominance and Epistatic Effects on Genetic Parameters and Breeding Values Estimation for Murciano-Granadina Goat Milk Yield and Components. Genes 11(3):309.

Rosati, A. and L. D. Van Vleck. 2002. Estimation of genetic parameters for milk, fat, protein and mozzarella cheese production for the Italian river buffalo Bubalus bubalis population. Livest. Prod. Sci. 74(2):185-190.

Saavedra, L. 2019. Variabilidad genética de ácidos grasos y fracciones nitrogenadas en leche de bovinos. [Phd]. [Chapingo (MX)]: Universidad Autónoma Chapingo.

Sapp, R. L., W. Zhang, J. K. Bertrand, and R. Rekaya. 2007. Genetic evaluation in the presence of uncertain additive relationships. I. Use of phenotypic information to ascertain paternity. J. Anim. Sci. 85(10):2391-2400.

Shiotsuki, L., F. F. Cardoso, and L. G. Albuquerque. 2018. Method for the estimation of genetic merit of animals with uncertain paternity under Bayesian inference. J. Anim. Breed. Genet. 135(2):116-123.

Shiotsuki, L., F. F. Cardoso, J. A. Silva, 2nd, and L. G. Albuquerque. 2013. Comparison of a genetic group and unknown paternity models for growth traits in Nellore cattle. J. Anim. Sci. 91(11):5135-5143.

Teissier, M., H. Larroque, and C. Robert-Granié. 2018. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. Genet Sel Evol 50(1):31.

Theron, H., F. Kanfer, and L. Rautenbach. 2002. The effect of phantom parent groups on genetic trend estimation. S. Afr. J. Anim. Sci. 32:130-135.

Tsuruta, S., I. Misztal, I. Aguilar, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94(8):4198-4204.