Italian Journal of Animal Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjas20

Genomic investigation of milk production in Italian buffalo

Alberto Cesarani, Stefano Biffani, Andre Garcia, Daniela Lourenco, Giacomo Bertolini, Gianluca Neglia, Ignacy Misztal & Nicolo Pietro Paolo Macciotta

To cite this article: Alberto Cesarani, Stefano Biffani, Andre Garcia, Daniela Lourenco, Giacomo Bertolini, Gianluca Neglia, Ignacy Misztal & Nicolo Pietro Paolo Macciotta (2021) Genomic investigation of milk production in Italian buffalo, Italian Journal of Animal Science, 20:1, 539-547, DOI: 10.1080/1828051X.2021.1902404

To link to this article: https://doi.org/10.1080/1828051X.2021.1902404

ITALIAN JOURNAL OF ANIMAL SCIENCE
2021, VOL. 20, NO. 1, 539–547
https://doi.org/10.1080/1828051X.2021.1902404

Taylor & Francis
Taylor & Francis Group

PAPER

Genomic investigation of milk production in Italian buffalo

Alberto Cesarani[a], Stefano Biffani[b], Andre Garcia[a], Daniela Lourenco[a], Giacomo Bertolini[c], Gianluca Neglia[d], Ignacy Misztal[a] and Nicolo Pietro Paolo Macciotta[e]

[a]Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA; [b]Consiglio Nazionale delle Ricerche (CNR), Istituto di biologia e biotecnologia agraria (IBBA), Milano, Italy; [c]Associazione Nazionale Allevatori Specie Bufalina (ANASB), Caserta, Italy; [d]Dipartimento di Medicina Veterinaria e Produzioni Animali, II University of Naples, Napoli, Italy; [e]Dipartimento di Agraria, University of Sassari, Sassari, Italy

ABSTRACT

The aim of this study was to test the feasibility of genomic selection in the Italian Mediterranean water buffalo, which is farmed mainly in the south Italy for milk, and mozzarella, production. 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 4127 cows, born between 1975 and 2009, were analysed in a three-trait model. Cows born in 2008 and 2009 with phenotypes and genotypes were selected as validation animals ($n = 50$). Variance components (VC) were estimated with BLUP and ssGBLUP. Heritabilities for BLUP were 0.25 ± 0.02 (MY), 0.16 ± 0.01 (FY) and 0.25 ± 0.01 (PY). 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: correlation, accuracy, dispersion, and bias statistics were calculated. Average correlations were 0.71 ± 0.02 and 0.82 ± 0.01 for BLUP and ssGBLUP-E, respectively. Accuracies were also higher in ssGBLUP-E (0.75 ± 0.03) compared to BLUP (0.57 ± 0.03). The best dispersions (i.e. closer to 1) were found for ssGBLUP-C. The use of genotypes only for the 35 bulls did not change the validation values compared to BLUP. Results of the present study, even if based on small number of animals, showed that the inclusion of genotypes of females can improve breeding values accuracy in the Italian Buffalo.

HIGHLIGHTS

- The genotypes of males did not improve the predictions.
- Genotypes of females improve breeding values accuracy.
- Slight increase in prediction accuracy with weighted ssGBLUP.

ARTICLE HISTORY
Received 27 January 2021
Revised 4 March 2021
Accepted 8 March 2021

KEYWORDS
Validation; single-step; LR method

Introduction

The domestic water buffalo is the second dairy species of the world, accounting for about 15% of the total milk production in 2018 (FAOSTAT 2020). Largest buffalo milk producers are India, Pakistan, and China. The world buffalo stock exhibited a relevant increase in the last six decades, from 88 million in 1961 to 206 million in 2018 (FAOSTAT 2020), respectively. Reasons for such an increase can be found in the high market value of buffalo milk and of its dairy products. Moreover, it should be remembered the great relevance of this species in many rural areas of the world as a source of meat and of traction power.

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 (di Berardino and Iannuzzi 1981), 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 river type consists mostly of distinct breeds as a result of selection whereas swamp consists of local unselected populations that have evolved by adaptation to a specific environment (Zhang et al. 2016).

CONTACT Alberto Cesarani alberto.cesarani@uga.edu Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA
Supplemental data for this article can be accessed here.

Table 1. Basic statistics and phenotypic correlations.

	MY (kg)	FY (kg)	PY (kg)
Statistics			
Mean	8.74	0.76	0.41
SD	3.27	0.27	11
Max	29.8	3.35	1.65
Phenotypic correlation			
MY		0.83	0.97
FY			0.82
PY			

MY: Milk yield; FY: Fat yield; PY: Protein yield.

River buffalo breeds have been selected for milk production traits, although organised breeding plans are currently running only in few countries (Zhang et al. 2020). Genetic selection in buffalo is hampered by the lack of reliable pedigree information (Rosati and Van Vleck 2002; Abdel-Shafty et al. 2020; Ghoreishifar et al. 2020), the difficult implementation of performance recording due to the farming structure that in several countries is based on small holders (Mokhber et al. 2018; Abdel-Shafty et al. 2020), and to the poor reproductive performances. As a consequence, the potential of buffalo has not been completely exploited (Mokhber et al. 2018).

The availability of a medium density (90 K) SNP platform for the buffalo (Iamartino et al. 2017) has opened the possibility to implement Genomic Selection (GS) programs also for this species. In particular the possibility of calculating the true realised relationships, even when pedigree information is poor or totally lacking (Hayes et al. 2009), could help in addressing a main issue of buffalo breeding. Moreover, the use of the single step GS approach (Aguilar et al. 2010) that combines phenotypic, pedigree and genomic information in the genomic BLUP framework could further help to improve breeding values accuracy in buffalo also when only a limited portion of the population is genotyped.

The genetic improvement of buffalo is a topic of great interest for the Italian livestock industry. Italy is the sixth Country in the world for Buffalo milk production and the sixteenth for number of farmed animals. The milk is mainly processed into mozzarella Campana cheese, that is recognised by Protected Designation of Origin (PDO) certification, and that has a high market value. Together with Brazil, Italy has experienced the largest relative increase (>200%) of the buffalo stock in the last sixty years. Buffaloes farmed in Italy are of the Mediterranean river breed, a group that includes also the breed from Mozambique and Romania (Colli et al. 2018). The Italian buffalo stock consists of 403,093 animals (www.vetinfo.it), and the dairy performance recording plan carried out by the Italian National Association of Buffalo Breeders (ANASB) involved 34,490 buffaloes in 2019. In the last 15 years, the average milk yield per lactation showed an improvement of about 8%, from kg 2184 in 2014 to kg 2356 in 2019, whereas fat (8.06 vs. 8.01) and protein (4.68 vs. 4.63) percentages remained quite constant (www.anasb.it).

In this work, the usefulness of genomic models for the estimation of variance components and the prediction of breeding values for milk production traits in the Italian Buffalo breed is evaluated. In particular, the traditional pedigree based BLUP was compared with the single-step genomic BLUP.

Materials and methods

Data used in this study were obtained from pre-existing databases and therefore the animal care approval was not needed.

Data

The whole data set (Table 1) consisted of 80,417 test day (TD) records of milk (MY, kg), fat (FY, kg) and protein (PY, kg) yields from 4127 buffalo cows born from 1975 to 2009 and farmed in 8 commercial herds.

Phenotypes were provided by ANASB. A sample of 463 buffalo cows of the whole data set and 35 bulls were genotyped with the Axiom Buffalo Genotyping Array 90 K (Iamartino et al. 2017). Quality control was applied both on animals and SNPs with the following criteria: call rate of animals and markers $\geq$95% (27,167 SNPs discarded); minor allele frequency $\geq$5% (8322 SNPs discarded), and *P*-value for the Hardy Weinberg equilibrium test $<10^{-6}$ (167 SNPs discarded). Moreover, markers mapped on sexual chromosomes or without known position were removed. After quality control, all animals and 49,164 markers were retained for the analysis.

A pedigree with 7730 animals was provided by ANASB from which 3 generations were traced back from animals with records or genotypes, resulting in a total of 5404 animals which were included in the analyses.

Analysis

Data were analysed with the following 3-trait repeatability animal model:

$$y_t = Xb_t + W_1 g_t + W_2 u_t + W_3 p_t + e_t$$

y_t is the vector of TD records for trait t (MY, FY, or PY)
b_t is the vector of fixed effects of calving year (23

levels), parity (5 levels), birth year (27 levels), milking events per day (2 levels), calving month (12 levels) and days in milk (linear covariate); g_t, u_t and p_t are the vectors of herd-test-day, additive genetic and permanent environmental random effects, respectively; and e_t is the vector of random residuals. The X, W_1, W_2 and W_3 are the incidence matrices associating phenotypic records to the effects in b_t, g_t, u_t and p_t, respectively. (Co)variance structures of random effects were $I\sigma^2_{HTD}$, $REL\sigma^2_a$, and $I\sigma^2_{pe}$, where I is an identity matrix, REL is the relationship matrix, σ^2_{HTD}, σ^2_a, and σ^2_{pe}, are variance components associated to random effects of herd-test-day, additive genetic, and permanent environment, respectively.

According to the relationship matrix used, two models were used: (i) the pedigree-based (BLUP) with the numerator relationship matrix (A); (ii) the single step genomic BLUP (ssGBLUP) where A and the genomic relationship matrix (G) are blended into H, which inverse (H^{-1}) was built according to Aguilar et al. (2010):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

where A^{-1} and G^{-1} are the inverses of the pedigree and genomic relationship matrices, respectively; A_{22}^{-1} is the inverse of the pedigree relationship matrix for genotyped animals only. The G matrix was created according to VanRaden (2008), with the following equation:

$$G = \frac{ZDZ'}{k}$$

where Z is the matrix of centred gene contents, D is a diagonal matrix of SNP weights, equal to I in ssGBLUP, and k is the scaling parameter defined as $2\sum p_i(1-p_i)$ where p_i is the allele frequency of the i-th SNP. To avoid singularity, G was blended with 5% of A_{22} (VanRaden, 2008). Therefore, in the ssGBLUP model, SNP information is used to construct the G matrix among all the genotyped individuals.

Variance components were estimated with both methods using the aiREML algorithm implemented in the airemlf90 software (Misztal et al. 2014).

Breeding values were estimated with BLUP and ssGBLUP models. The genomic model was tested in five scenarios that differ in the number of genotyped animals:

A = genotypes available only for 35 bulls;
B = genotypes available only for the 50 candidates;
C = genotypes available for 50 candidates + 35 bulls;
D = genotypes available for 463 cows;
E = genotypes available for 463 cows + 35 bulls.

In all scenarios, candidate animals were 50 females born in 2008 and 2009 with phenotypes and genotypes (except scenario A) available.

To evaluate prediction accuracy and bias, the LR method was applied (Legarra and Reverter, 2018). Breeding values estimation was run twice: in the first run, candidate animals had their phenotypes available (whole data), whereas in the second they had their phenotypes masked (partial data). The following four statistics were computed to evaluate the models:

Dispersion:

$$b_{w,p} = \frac{\mathrm{cov}(\hat{u}_w, \hat{u}_p)}{\mathrm{var}(\hat{u}_p)}$$

where $\hat{u}_w$ and $\hat{u}_p$ are the vectors of breeding values estimated for the candidates in the whole and partial datasets, respectively.

Prediction accuracy:

$$\widehat{acc} = \sqrt{\frac{\mathrm{cov}(\hat{u}_w, \hat{u}_p)}{(1-\overline{F})\hat{\sigma}^2_u}}$$

where $\overline{F}$ is the average inbreeding of the candidates and $\hat{\sigma}^2_u$ is the additive genetic variance.

Correlation:

$$\rho_{w,p} = \mathrm{cor}(\hat{u}_w, \hat{u}_p)$$

Bias:

$$\mu_{wp} = -\hat{u}_p - -\hat{u}_w$$

Variance components estimated with the BLUP model were used in the validation. Convergence criterion was set to 10^{-15}.

A further development is the weighted single-step (WssGBLUP) where different weights are assigned to SNP in the calculation of the genomic relationship matrix (Wang et al. 2012). In this procedure, the optimum weights are determined iteratively, with the first iteration being equivalent to ssGBLUP. In the first iteration, *p*-values for SNPs effects were computed according to Aguilar et al. (2019); Bonferroni correction, i.e. 0.05 divided by the number of SNPs (49,164), was set as significance threshold. Usually, 3–5 iterations are needed for obtaining the maximum prediction accuracy. Fragomeni et al. (2019) suggested that non-linear weights would be a better option in WssGBLUP compared to linear weights proposed by Wang et al. (2012). In this study, the non-linear A weights were calculated as follows (VanRaden 2008; Legarra et al. 2018):

$$d_i = CT^{\frac{|\hat{a}_i|}{sd(\hat{a})}-2}$$

where d_i is the weight for SNP i; CT is a constant the determines the departure from normality (usually

Table 2. Genetic parameters estimated using two different methods, BLUP and single-step GBLUP.

	BLUP			ssGBLUP		
	MY	FY	PY	MY	FY	PY
Genetic parameters[a]						
HTD	0.24 ± 0.02	0.41 ± 0.02	0.06 ± 0.02	0.23 ± 0.01	0.40 ± 0.02	0.06 ± 0.01
Additive	1.87 ± 0.13	1.01 ± 0.08	0.38 ± 0.03	1.71 ± 0.11	0.91 ± 0.06	0.34 ± 0.02
Permanent	1.05 ± 0.10	0.73 ± 0.06	0.23 ± 0.02	1.11 ± 0.08	0.77 ± 0.04	0.24 ± 0.01
Residual	4.21 ± 0.02	4.08 ± 0.02	0.86 ± 0.02	4.21 ± 0.02	4.08 ± 0.02	0.86 ± 0.01
Heritability	0.25 ± 0.02	0.16 ± 0.01	0.25 ± 0.01	0.23 ± 0.01	0.15 ± 0.01	0.23 ± 0.01
Repeatability	0.40 ± 0.01	0.28 ± 0.01	0.40 ± 0.01	0.39 ± 0.01	0.27 ± 0.01	0.39 ± 0.01
Genetic correlation						
MY		0.92 ± 0.02	0.98 ± 0.04		0.92 ± 0.01	0.97 ± 0.01
FY			0.96 ± 0.04			0.95 ± 0.01
PY						

[a]Values for FY and PY were multiplied by 10.
BLUP: best linear unbiased prediction; ssGBLUP: single-step genomic best linear unbiased prediction; MY: Milk yield; FY: Fat yield; PY: Protein yield; HTD: herd-test-day.

assumed for the SNP effects); $|\hat{a}_i|$ and $sd(\hat{a})$ are the absolute effect for SNP i and the standard deviation of the SNP effects, respectively. A CT value of 1.0 is equivalent to the unweighted ssGBLUP and deviations from it determine the departure from normality. Because this value is empirically determined, we tested three different values (1.025, 1.125, and 1.25) and the same LR validation was applied over 5 iterations of WssGBLUP to evaluate which weight and iteration combination would maximise accuracy of predictions with the least bias.

Results and discussion

Genetic parameters

Compared to values in Table 1, slightly lower average daily milk yield (7.33 ± 3.03 kg) was reported for the Nili-Ravi buffalo breed (Abdel-Shafy et al. 2020).

Heritability and repeatability estimates were very similar in both considered models (Table 2); lower values of genetic parameters were obtained for FY. As far as the kind of model is concerned, the ssGBLUP provided smaller estimates (Table 2) and smaller standard errors for genetic correlations. These differences were basically due to the larger additive genetic variance and the lower permanent environmental variance estimated with BLUP. On the contrary, residual variances were the same across methods, and variance associated to HTD were very similar (Table 2).

Genetic parameters of milk production traits have been estimated in buffalo with traditional pedigree-based BLUP methods, both on cumulated yields and on TD records. Lower h^2 values for lactational traits (0.14 for MY, 0.11 for FY, and 0.14 for PY) were reported for Italian Buffalo, whereas genetic correlations were similar to those of the present study (Rosati and Van Vleck 2002). Heritability values for lactational milk traits similar to those found in the present study with BLUP and sGBLUP models were estimated in Brazilian Murrah and Philippine buffaloes (Tonhati et al. 2008; Aspilcueta-Borquis et al. 2010; Flores et al. 2013). Lower genetic correlation values among yield traits were reported for Murrah buffaloes (Aspilcueta-Borquis et al. 2010). The modelling of TD records for estimating variance components resulted in lower heritability estimates. For example, Flores et al. (2013) reported values of 0.15, 0.09, and 0.03 for MY, FY and PY, respectively. Genetic correlations reported by these authors were 0.99 between MY and FY and 0.97 between MY and PY, respectively.

Genomic information has already been used to estimate variance components for milk traits in buffaloes. Liu et al. (2018) reported h^2 of 0.33, 0.35 and 0.27 for lactational MY, FY, and PY, respectively. Abdel-Shafy et al. (2020) estimated a slightly lower h^2 (0.20) for daily milk yield in Egyptian buffalo using a ssGBLUP approach. Some comparisons between BLUP and GBLUP were performed. Aspilcueta-Borquis et al. (2015) found similar heritabilities using the two methods, whereas Iamartino et al. (2017), reported h^2 for milk yield of 0.38 using the pedigree relationship matrix and 0.45 using a genomic relationship matrix.

In the present study, genetic parameters estimated with BLUP and ssGBLUP were quite similar. Unbiased variance components can be estimated when the whole data about the studied population is used in the computation. The unbiased estimations with complete data can be realised also in population under selection if the data used to make the selection decision are included in the analysis (Cantet et al. 2000). However, sometimes the use of complete dataset is not possible, either because the data is too big which makes variance components computationally expensive or because of incomplete and or incorrect pedigrees in the historical dataset. The ssGBLUP seems to be better than BLUP with subsets of data in simulated

Table 3. LR validation results with BLUP and single-step genomic BLUP (ssGBLUP).

		ssGBLUP[a]				
	BLUP	A	B	C	D	E
N genotypes	–	35	50	85	463	498
Correlation						
Milk	0.72	0.72	0.75	0.77	0.82	0.83
Fat	0.71	0.70	0.75	0.76	0.81	0.81
Protein	0.69	0.69	0.73	0.75	0.82	0.82
Accuracy						
Milk	0.60	0.60	0.67	0.68	0.77	0.77
Fat	0.55	0.55	0.62	0.63	0.71	0.72
Protein	0.57	0.57	0.65	0.66	0.76	0.76
Dispersion						
Milk	0.93	0.93	0.92	0.95	0.89	0.90
Fat	0.98	0.98	0.99	1.01	0.92	0.93
Protein	0.92	0.92	0.93	0.96	0.92	0.92
Bias						
Milk	−0.107	−0.106	−0.103	−0.099	−0.014	−0.019
Fat	−0.013	−0.013	−0.013	−0.015	−0.005	−0.006
Protein	−0.005	−0.005	−0.005	−0.005	−0.001	−0.001

[a]A: genotypes only for 35 bulls; B: genotypes only for the 50 candidates; C: genotypes for 50 candidates + 35 bulls; D: genotypes for 463 cows; E: genotypes for 463 cows + 35 bulls.
BLUP: best linear unbiased prediction; ssGBLUP: single-step genomic best linear unbiased prediction.

populations under selection (Cesarani et al. 2019). Recently, Aldridge et al. (2020) in a pig study (with 10 different traits, including mean litter birth weight, total number born, number still born and litter mortality) found very similar variance components using BLUP and different ssGBLUP models. However, because genomic estimates exhibited lower standard errors, they concluded that H is a more informative than A. Moreover, they also suggested that the genomic approach is to be preferred for more complex models, as those with repeated records where a permanent environment effect is fitted. Similar results for variance components estimated with BLUP and ssGBLUP were recently reported for two functional traits in Italian Simmental (Cesarani, Gaspa, et al. 2020). Hidalgo et al. (2020) investigated the changes in genetic parameters for fitness and growth traits in pigs under GS. These authors reported that the genetic parameters estimated with the traditional BLUP are biased in populations undergoing genomic selection and the differences between BLUP and ssGBLUP could be caused by genomic preselection unaccounted for by BLUP.

Breeding value prediction

Correlation between breeding values estimated for candidate cows with BLUP and ssGBLUP were 0.96, 0.95, and 0.95 for MY, FY, and PY, respectively. Correlations of BLUP and ssGBLUP breeding values estimated for the other animals were 0.99 for all the three considered traits. Also, the rank correlations between BLUP and ssGBLUP breeding values was high (0.98), indicating that the methods are comparable.

Correlations and prediction accuracies increased as the number of genotypes included in the analysis increased (Table 3). MY showed better values compared to the other two traits, probably because its higher h^2. Indeed, prediction accuracy is influenced by the trait heritability (Hayes et al. 2009). Average correlations among traits in the scenario with the largest number of genotypes available (E) were 0.11 larger in ssGBLUP compared to BLUP. The highest correlation between the two sets of BVs (i.e. stability of the model) was found with the largest number of genotypes: values of correlation were 0.83, 0.81, and 0.82 for MY, FY, and PY. Average predictions accuracies increased from 0.57 to 0.75 moving from BLUP to ssGBLUP.

Interesting results have been obtained for the dispersion (Table 3), even if a clear pattern across the different scenarios could not be detected. The lowest dispersion was observed for ssGBLUP in the scenario C (i.e. with genotypes for candidates and 35 bulls). Even if dispersions for ssGBLUP-E were worser than BLUP ones, they were within a 10% difference from the ideal value (1). Tsuruta et al. (2011) reported that values within 15% are acceptable. The lowest biases were found when genotypes of all cows ($N = 463$) were included in the analysis (i.e. ssGBLUP-D).

The differences in correlations and accuracies between the use of genotypes of only females or both females and males (i.e. scenarios B vs C and D vs E) were limited, suggesting that the number of genotypes available and not the source (males vs females) mainly affects genomic prediction performances. The inclusion of only bull genotypes (i.e. ssGBLUP-A) did not change prediction statistics in comparison with BLUP, apart from bias, probably because the sire contribution was already included in the analysis through the phenotypes of the cows.

Higher breeding values accuracies (+8%), based on the prediction error variance (Gonzáles-Récio et al. 2006), were reported for ssGBLUP compared to BLUP in Brazilian buffaloes (Aspilcueta-Borquis et al. (2015), together with lower standard deviations for ssGBLUP.

More accurate predictions were reported by genomic methods (GBLUP and ssGBLUP) compared to BLUP in Philippine dairy buffaloes (Herrera et al. 2019). These authors compared accuracies for genotyped and for non-genotyped animals and they reported the highest increase in accuracy for the ungenotyped animals (+0.07 on average among milk, fat and protein)

Table 4. Weighted ssGBLUP-E using the largest number of genotypes (463 cows + 35 bulls).

CT		1.025				1.125				1.250			
Round	ssGBLUP*	2	3	4	5	2	3	4	5	2	3	4	5
Correlation													
Milk	0.83	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.81	0.81	0.81	0.81
Fat	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.80	0.80	0.80	0.80
Protein	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.81	0.81	0.81	0.80
Accuracy													
Milk	0.77	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.81	0.82	0.82	0.82
Fat	0.72	0.73	0.73	0.73	0.73	0.75	0.75	0.75	0.75	0.76	0.77	0.77	0.77
Protein	0.76	0.77	0.77	0.77	0.77	0.79	0.79	0.79	0.79	0.80	0.81	0.81	0.81
Dispersion													
Milk	0.90	0.89	0.89	0.89	0.89	0.88	0.88	0.88	0.88	0.86	0.85	0.85	0.85
Fat	0.93	0.92	0.92	0.92	0.92	0.91	0.91	0.91	0.91	0.90	0.89	0.88	0.88
Protein	0.92	0.92	0.92	0.92	0.92	0.90	0.90	0.90	0.90	0.89	0.88	0.87	0.87
Bias													
Milk	−0.019	−0.0172	−0.0172	−0.0172	−0.0172	−0.0113	−0.0099	−0.0096	−0.0095	−0.0047	−0.0005	0.0023	0.0028
Fat	−0.006	−0.0060	−0.0060	−0.0060	−0.0060	−0.0057	−0.0056	−0.0055	−0.0055	−0.0052	−0.0049	−0.0048	−0.0048
Protein	−0.001	−0.0012	−0.0012	−0.0012	−0.0012	−0.0010	−0.0009	−0.0009	−0.0009	−0.0007	−0.0005	−0.0004	−0.0004

*Results from the 1st iteration of WssGBLUP are the same as unweighted ssGBLUP.
CT: carcass traits; ssGBLUP: single-step genomic best linear unbiased prediction.

compared to the group of genotyped animals only (+0.03 on average among milk, fat and protein).

LR method was recently widely adopted to validate breeding values estimation in almost all livestock species, such as cattle (Porto Neto et al. 2019; Cesarani, Hidalgo, et al. 2020; Durbin et al. 2020), sheep (Granado-Tajada et al. 2020; Macedo et al. 2020), pigs (Aliakbari et al. 2020), chicken (Chu et al. 2019; Bermann et al. 2021) and fish (Silva et al. 2019). Results about the comparison between BLUP and ssGBLUP, by using LR method, can change according to the considered species. No clear advantages were found for genomic selection of Latxa dairy sheep breed (Granado-Tajada et al. 2020), whereas other studies reported better performances of ssGBLUP compared to BLUP (Chu et al. 2019; Silva et al. 2019). For example, Chu et al. (2019) reported that the use of ssGBLUP increased population accuracy of BV for both genotyped and non-genotyped birds compared to the use of BLUP. However, bias of BV prediction increased for non-genotyped birds.

The results for the WssGBLUP are presented in Table 4. Overall, there was an increase in prediction accuracy and a reduction in bias which are beneficial, however at the same time there was an increase in the dispersion of the breeding values, which is undesirable. The difference between iterations (2–5) was small, with the second iteration reaching the highest accuracy. On the other hand, there was a considerable increase in accuracy when the CT values were increased. For instance, for MY prediction accuracy went from 0.79 (CT = 1.025) to 0.82 (CT = 1.25) which represent an increase of 2.6 and 6.5% respectively, over ssGBLUP.

The performance of the WssGBLUP compared to ssGBLUP depends on the architecture of the trait and on the sample size. For instance, Lourenco et al. (2017), showed that for highly polygenic traits there were no improvements in prediction accuracy with WssGBLUP compared to ssGBLUP but under oligogenic traits, accuracy can be improved with SNP weighting. Additionally, with larger sample sizes, the benefit of using different weights for markers may be reduced, as pointed out by Karaman et al. (2016) who showed that the prediction accuracy of Bayesian variable selection methods and GBLUP converged to the similar prediction accuracy once the sample size was large enough.

As aforementioned, 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. In this regard, the integration of α_{s1} casein gene in genomic evaluation of dairy goats are effective examples. Dagnachew et al. (2011) analysed 38 SNPs located within the four casein genes in Norwegian goats and showed significant additive effect of a single SNP within two casein variants (i.e. CSN1S1 and CSN3 genes) on fat and protein percentages, milk yield and milk taste. Pizarro Inostroza et al. (2020) analysed the effects of considering 48 casein loci-located SNP in the Murciano-Granadina goats. Teissier et al. (2018) investigated three weighted ssGBLUP methods to integrate information on the α_{s1} *casein* in genomic evaluations of dairy goats. All

authors concluded that the inclusion of detailed information on major genes (e.g. *DGAT1* for fat content or *αs1 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 (Supplementary Figure 1). 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. Liu et al. (2018) reported 4 SNP in 2 genomic regions significantly associated with fat yield and protein percentage in water buffalo and concluded that a more precise map of the buffalo genome as well as a larger sample size is needed in future studies to be able to identify important genes affecting the traits.

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 (scenario E), compared to 0.86, 0.90, and 0.89 for WssGBLUP (CT = 1.25, iteration 2) for MY, FY, and PY, respectively (Table 4). As pointed out by Legarra and Reverter (2017), the bias and the slope of the regression (dispersion) need to be taken in to account specially when proven and young animals are mixed in the population, because they could affect selection decisions, beyond the accuracy of selection. 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

This study, even if based on a small number of genotyped animals, provided interesting insights about the use and future application of ssGBLUP in the Italian Buffalo. Variance components estimated using BLUP and ssGBLUP were similar, but the latter model showed lower standard deviations for genetic correlations. The results showed that the inclusion of genotypes only for bulls (i.e. ssGBLUP-A) did not lead to any improvements compared to BLUP. On the contrary, the inclusion of genotypes for cows led to an increase in almost all validation parameters. Thus, to implement a genomic evaluation in the Italian Buffalo, it is important to emphasise that female genotyping could increase the accuracies of the prediction compared to both pedigree-based and genomic evaluations with only male genotypes models.

Acknowledgements

We gratefully acknowledge the Associazione Nazionale Allevatori Specie Bufalina (ANASB) for providing data.

Disclosure statement

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

ORCID

Alberto Cesarani http://orcid.org/0000-0003-4637-8669
Stefano Biffani http://orcid.org/0000-0001-5559-3630
Andre Garcia http://orcid.org/0000-0001-9778-7978
Daniela Lourenco http://orcid.org/0000-0003-3140-1002
Gianluca Neglia http://orcid.org/0000-0002-0989-6072
Ignacy Misztal http://orcid.org/0000-0002-0382-1897
Nicolo Pietro Paolo Macciotta http://orcid.org/0000-0001-5504-9459

References

Abdel-Shafy H, Awad MA, El-Regalaty H, Ismael A, El-Assal SED, Abou-Bakr S. 2020. A single-step genomic evaluation for milk production in Egyptian buffalo. Livest Sci. 234: 103977.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci. 93(2): 743–752.

Aguilar LA, Cardoso F, Masuda Y, Lourenco D, Misztal I. 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. Genet Sel E. 51:1–8.

Aldridge MN, Vandenplas J, Bergsma R, Calus MP. 2020. Variance estimates are similar using pedigree or genomic relationships with or without the use of metafounders or the algorithm for proven and young animals. J Anim Sci. 98:skaa019.

Aliakbari A, Delpuech E, Labrune Y, Riquet J, Gilbert H. 2020. The impact of training on data from genetically-related lines on the accuracy of genomic predictions for feed efficiency traits in pigs. Genet Sel E. 52:1–15.

Aspilcueta-Borquis RR, Araujo Neto FR, Santos DJA, Hurtado-Lugo NA, Silva JAV, Tonhati H. 2015. Multiple-trait

genomic evaluation for milk yield and milk quality traits using genomic and phenotypic data in buffalo in Brazil. Genet Mol Res. 14(4):18009–18017.

Aspilcueta-Borquis RR, Neto FA, Baldi F, Bignardi AB, Albuquerque LGD, Tonhati H. 2010. Genetic parameters for buffalo milk yield and milk quality traits using Bayesian inference. J Dairy Sci. 93(5):2195–2201.

Bermann M, Legarra A, Hollifield MK, Masuda Y, Lourenco D, Misztal I. 2021. Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: an application in chicken mortality. J Anim Breed Genet. 138(1):4–13.

Cantet RJ, Birchmeier AN, Santos-Cristal MG, De Avila VS. 2000. Comparison of restricted maximum likelihood and Method R for estimating heritability and predicting breeding value under selection. J Anim Sci. 78(10):2554–2560.

Cesarani A, Gaspa G, Masuda Y, Degano L, Vicario D, Lourenco D, Macciotta NPP. 2020. Variance components using genomic information for 2 functional traits in Italian Simmental cattle: calving interval and lactation persistency. J Dairy Sci. 103(6):5227–5233.

Cesarani A, Hidalgo J, Garcia A, Degano L, Vicario A, Masuda Y, Misztal I, Lourenco D. 2020. Beef trait genetic parameters based on old and recent data and its implications for genomic predictions in Italian Simmental cattle. J Anim Sci. 98:skaa242.

Cesarani A, Pocrnic I, Macciotta NPP, Fragomeni BO, Misztal I, Lourenco D. 2019. Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. J Anim Breed Genet. 136(1):40–50.

Chu TT, Bastiaansen JW, Berg P, Romé H, Marois D, Henshall J, Jensen J. 2019. Use of genomic information to exploit genotype-by-environment interactions for body weight of broiler chicken in bio-secure and production environments. Genet Sel E. 51:1–13.

Colli L, Milanesi M, Vajana E, Iamartino D, Bomba L, Puglisi F, Del Corvo M, Nicolazzi EL, Ahmed SSE, Verrera JRV, et al. 2018. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. Front Genet. 9:53

Dagnachew BS, Thaller G, Lien S, Ådnøy T. 2011. Casein SNP in Norwegian goats: additive and dominance effects on milk composition and quality. Genet Sel E. 43:1–12.

di Berardino D, Iannuzzi L. 1981. Chromosome banding homologies in Swamp and Murrah buffalo. J Hered. 72(3): 183–188.

Durbin HJ, Lu D, Yampara-Iquise H, Miller SP, Decker JE. 2020. Development of a genetic evaluation for hair shedding in American Angus cattle to improve thermotolerance. Genet Sel E. 52:1–14.

FAOSTAT, 2020. http://www.fao.org/faostat/en/#home. Accessed 10 Apr 2020.

Flores EB, Kinghorn BP, Van Der Werf J. 2013. Genetic parameters for milk yield and milk component traits estimated from test day and 305D lactation records of Philippine dairy buffaloes. Buffalo Bulletin, 32 (Special Issue 2): 768-772.

Fragomeni BO, Lourenco DAL, Legarra A, VanRaden PM, Misztal I. 2019. Alternative SNP weighting for single-step genomic best linear unbiased predictor evaluation of stature in US Holsteins in the presence of selected sequence variants. J Dairy Sci. 102(11):10012–10019.

Ghoreishifar SM, Moradi-Shahrbabak H, Fallahi MH, Moradi-Shahrbabak M, Abdollahi-Arpanahi R, Khansefid M. 2020. Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo. BMC Genet. 21:16.

González-Recio O, Chang YM, Gianola D, Weigel KA. 2006. Comparison of models using different censoring scenarios for days open in Spanish Holstein cows. Anim Sci. 82(2): 233–239.

Granado-Tajada I, Legarra A, Ugarte E. 2020. Exploring the inclusion of genomic information and metafounders in Latxa dairy sheep genetic evaluations. J Dairy Sci. 103(7): 6346–6353.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci. 92(2):433–443.

Herrera JRV, Flores EB, Duijvesteij N, Moghaddar N, van der Werf J. 2019. Accuracy of genomic predictions for milk production traits in Philippine dairy buffaloes. In Proceedings of the 23rd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Armidale, New South Wales, Australia; 27th October–1st November 2019. p. 448–451).

Hidalgo J, Tsuruta S, Lourenco D, Masuda Y, Huang Y, Gray KA, Misztal I. 2020. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. J Anim Sci. 98:skaa032.

Karaman E, Cheng H, Firat MZ, Garrick DJ, Fernando RL. 2016. An upper bound for accuracy of prediction using GBLUP. PloS One. 11(8):e0161054.

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.

Legarra A, Lourenco DAL, Vitezica Z. 2018. Bases for genomic prediction. http://genoweb.toulouse.inra.fr/~alegarra/GSIP.pdf.

Legarra A, Reverter A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet Sel E. 50:1–18.

Legarra A, Reverter A. 2017. Can we Frame and Understand Cross-Validation Results in Animal Breeding? In Proceedings of the 22nd Conference Association for the Advancement of Animal Breeding and Genetics (AAABG), Townsville, Queensland, Australia; 2nd July-5th July 2017. p 2-5.

Liu JJ, Liang AX, Campanile G, Plastow G, Zhang C, Wang Z, Salzano A, Gasparrini B, Cassandro M, Yang LG. 2018. Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. J Dairy Sci. 101(1):433–444.

Lourenco DAL, Fragomeni BO, Bradford HL, Menezes IR, Ferraz JBS, Aguilar I, Tsuruta S, Misztal I. 2017. Implications of SNP weighting on single-step genomic

predictions for different reference population sizes. J Anim Breed Genet. 134(6):463–471.
Macedo FL, Christensen OF, Astruc JM, Aguilar I, Masuda Y, Legarra A. 2020. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. Genet Sel E. 52:1–10.
Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. 2014. Manual for BLUPF90 family of programs. [Accessed Jun 2020]. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.
Mokhber M, Moradi-Shahrbabak M, Sadeghi M, Moradi-Shahrbabak H, Stella A, Nicolzzi EL, Rahmaninia J, Williams JL. 2018. A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. BMC Genomics. 19(1):449.
Pizarro Inostroza MG, Landi V, Navas González FJ, León Jurado JM, Delgado Bermejo JV, Fernández Álvarez J, Martínez Martínez MDA. 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.
Porto Neto LR, Fortes MRS, Reverter A. 2019. GBLUP analysis predicted fertility phenotypes of crossbred bulls using data from Brahman and tropical composite. In Proceedings of the 23rd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Armidale, New South Wales, Australia; 27th October–1st November 2019. p. 87–90.
Rosati A, Van Vleck LD. 2002. Estimation of genetic parameters for milk, fat, protein and mozzarella cheese production for the Italian river buffalo *Bubalus bubalis* population. Liv Prod Sci. 74(2):185–190.
Silva RM, Evenhuis JP, Vallejo RL, Gao G, Martin KE, Leeds TD, Palti Y, Lourenco DA. 2019. Whole-genome mapping of quantitative trait loci and accuracy of genomic predictions for resistance to columnaris disease in two rainbow trout breeding populations. Genet Sel E. 51(1):1–13.
Teissier M, Larroque H, Robert-Granié C. 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):31–42.
Tonhati H, Cerón-Muñoz MF, Oliveira JAD, El Faro L, Lima ALF, Albuquerque LGD. 2008. Test-day milk yield as a selection criterion for dairy buffaloes (*Bubalus bubalis* Artiodactyla, Bovidae. Genet Mol Biol. 31(3):674–679.
Tsuruta S, Misztal I, Aguilar I, Lawlor TJ. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J Dairy Sci. 94(8): 4198–4204.
VanRaden PM. 2008. Efficient methods to compute genomic predictions. J Dairy Sci. 91(11):4414–4423.
Wang H, Misztal I, Aguilar I, Legarra A, Miur WM. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. 94:73–83.
Zhang Y, Colli L, Barker JSF. 2020. Asian water buffalo: domestication, history and genetics. Anim Genet. 51(2): 177–191.
Zhang Y, Lu Y, Yindee M, Li K-Y, Kuo H-Y, Ju Y-T, Ye S, Faruque MO, Li Q, Wang Y, et al. 2016. Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the China/Indochina border region. Mol Ecol. 25(7):1530–1550.