Use of single-step GWAS for prospecting genomic regions related to milk production and milk quality of buffalo cows

Camila da Costa Barros[1*], Daniel Jordan de Abreu Santos[1], Rusbel Raul Aspilcueta-Borquis[2], Gregório Miguel Ferreira de Camargo[3], Francisco Ribeiro de Araújo Neto[4], Humberto Tonhati[1]

São Paulo State University (FCAV/UNESP), Jaboticabal, São Paulo Brazil

[2] Universidade Federal da Grande Dourados, Dourados, Mato Grosso do Sul, Brazil

[3]Universidade Federal da Bahia, Salvador, Bahia, Brazil

[4]Instituto Federal Goiano, Rio Verde, Goias, Brazil

[*]Correspondence Camila da Costa Barros

Department of Animal Science

São Paulo State University (FCAV/UNESP)

Jaboticabal, São Paulo 14884-900, Brazil

E-mail: mila_costabarros@hotmail.com

Supplementary File

Phenotypic data

Test-day records obtained after the 5th day of production were considered and only lactations with a duration of more than 90 days were maintained. The %F and %P were obtained as the mean of monthly records per lactation. The contemporary groups were formed by the concatenation of herd and year and season of calving (October-March and April-September). The restrictions applied were that each contemporary group should contain a minimum of four animals and the record of the traits should be ± 3 standard deviations of the mean of the group. Table 1 shows the descriptive statistics of the data.

A repeatability animal model was used for all traits. The variance components (Table 2) were estimated by the restricted maximum likelihood method in single-trait analysis using the REMLF90 program (Misztal, 2005). The model included the contemporary groups as fixed effects, age of cow at calving as covariate (linear and quadratic), and additive genetic, permanent environmental and residual effects as random effects.

Genome-wide association analysis using information from genotyped and non-genotyped animals (ssGWAS)

The general model used for estimate breeding values (GEBVs) for the traits can be written in matrix form as:

$$y = X\beta + Za + Wp + e$$

where y is the vector of observations for each trait; X is the incidence matrix for fixed effects; β is the vector of fixed effects; Z is the incidence matrix for random additive genetic effects; a is the vector of random additive genetic effects assuming $a \sim N(0, H\sigma_a^2)$, where H is the matrix that combines the pedigree-based relationship matrix and genomic relationship matrix and σ_a^2 is the additive genetic variance; W is the incidence matrix for random permanent environmental effects; p is the vector of random permanent environmental effects, and e is the vector of random residual effects.

The combination of the traditional relationship matrix (A) and the genomic relationship matrix (G) resulted in a new matrix, called H^{-1} (Aguilar et al., 2010):

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

where A^{-1}, G^{-1} and A^{-1}_{22} are, respectively, the inverse of the traditional relationship matrix, the inverse of the genomic matrix and the inverse of the relationship matrix between genotyped animals based only on pedigree.

The genomic matrix was obtained according to Vanraden (2008):

$$\mathrm{G} = \frac{\mathrm{ZZ}'}{2\sum_{\mathrm{j}=1}^{\mathrm{m}} \mathrm{p_j}(1-\mathrm{p})}$$

where Z is the subtraction of $M - P$, where M is the matrix of genotypes, with the columns representing the markers and the rows the animals, and P is the frequency matrix of the second allele p_j, expressed as $2p_j$.

The SNP effects were obtained from the GEBVs of genotyped animals in an iterative manner using the postGSf90 program of the BLUPF90 family (Misztal et al., 2002). Followed by the preGSf90 and BLUPF90 programs, postGSf90 calculates the effect of SNPs as described by Wang et al. (2012) using different weights in the genomic relationship matrix, thus permitting the application of different weighting factors for the SNPs. The equation used to calculate the SNP effects can be written in matrix form as:

$$\hat{\mathrm{u}} = \mathrm{DZ}'[\mathrm{ZDZ}']^{-1}\hat{\mathrm{a}}_{\mathrm{g}}$$

where $\hat{u}$ is the vector of the effect of each SNP; D is the diagonal matrix containing weighting factors for the SNP effect; Z is the matrix of genotypes, and $\hat{a}_g$ is the vector of predicted breeding values for genotyped animals.

In the present study, two iterations were performed to estimate the SNP effects. In the first iteration, weighting factors equal to one were assumed, which were calculated as a function of the squared effects of the SNPs and allele frequencies and used in the second iteration. According to Wang et al. (2014), weighting is important to identify regions of larger effect on a trait, i.e., increasing the weights attributed to large-effect SNPs and reducing the weights attributed to small-effect SNPs.

The variance of each SNP was calculated by multiplying the squared effect of SNP i ($\hat{u}_i^2$) by $2p_iq_i$, where p_i is the frequency of the second allele of SNP i and q_i is ($1-p_i$) (Zhang et al., 2010). The percentage of genetic variance explained by each SNP was calculated as described by Wang et al. (2014):

$$\frac{Var(a_i)}{\sigma_a^2} \text{ x } 100\%$$

where a_i is the breeding value of each region with only one SNP and σ_a^2 is the total genetic variance. The 10 SNPs explaining the highest percentage of additive genetic variance were selected for the investigation of genes.

Table 1. Descriptive statistics of milk yield (MY) and fat (%F) and protein (%P) percentage of Murrah buffalo cows.

Trait	N	Mean	Standard deviation	CG
MY (kg)	10,507	2,012.80	697.54	169
%F	4,545	6.63	1.01	49
%P	4,542	4.29	0.29	49

N = number of animals, CG = contemporary group.

Table 2. Additive genetic (σ_a^2), permanent environmental (σ_{pe}^2) and residual variance (σ_e^2), heritability (h^2) and repeatability (r) obtained for milk yield (MY) and fat (%F) and protein (%P) percentage of Murrah buffalo cows.

Trait	σ_a^2	σ_{pe}^2	σ_e^2	h^2	r
MY	75,670	87,550	184,900	0.22	0.47
%F	0.1666	0.1817	0.4447	0.21	0.44
%P	0.0153	0.0177	0.0389	0.21	0.46