Genetic parameter estimates for buffalo milk yield, milk quality and mozzarella production and Bayesian inference analysis of their relationships

R.R. Aspilcueta-Borquis[1], R. Di Palo[2], F.R. Araujo Neto[1], F. Baldi[1], G.M.F. de Camargo[1], L.G. de Albuquerque[1], L. Zicarelli[2] and H. Tonhati[1]

[1]Departamento de Zootecnia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, São Paulo, SP, Brasil
[2]University of Naples "Federico II", Napoli, Italy

Corresponding author: H. Tonhati
E-mail: tonhati@fcav.unesp.br

ABSTRACT. Buffalo milk has excellent physical and chemical qualities as a consequence of the high percentage of constituents. This milk property is desirable for the dairy industry because it facilitates manufacture of mozzarella cheese. We estimated genetic parameters for milk yield, milk fat and protein and their effects on mozzarella cheese production using Bayesian inference. Using information from 4907 lactation records of buffaloes, genetic and non-genetic parameters were estimated for accumulated 305-day milk yield (MY), milk fat (%F) and protein (%P) percentages and mozzarella production per lactation (MP). The (co)variance components were obtained by Bayesian inference using a multiple trait model, which included as fixed effects contemporary group, milking number and buffalo age at calving as covariables (linear

and quadratic), along with the additive genetic, permanent environmental and residual random effects. Mean *a posteriori* heritability distributions for MY, %F, %P, and MP were 0.25, 0.30, 0.38, and 0.23, respectively. The genetic correlation estimates between MY with %P and %F were negative and moderate. Positive genetic correlation estimates varying from 0.19 (%P/MP) to 0.95 (MY/MP) were obtained among the traits. Milk yield, milk components, and mozzarella production in Murrah buffaloes have enough genetic variation for selection purposes. We conclude that selection to increase milk yield would be effective in improving mozzarella production.

Key words: *Bubalus bubalis*; Milk composition; Gibbs analyses

INTRODUCTION

Buffalo milk accounts for over 50% of drinking milk in certain developing countries, such as India, Pakistan, Egypt, and Nepal, while in Italy and in Brazil, buffalo milk is almost exclusively used for mozzarella cheese production, an expensive semi-soft fresh cheese. During the process of mozzarella cheese making, fat and casein are the major milk solids incorporated into the final product, these milk components being routinely used in many countries as criteria to determine the milk price.

Compared with cow's milk, buffalo milk has a higher percentage of all components, such as protein and fat. The mean protein and fat percentages reported for buffalo milk varies from 4.13 to 4.55% (Macedo et al., 2001; Rosati and Van Vleck, 2002) and from 6.87 to 8.59% (Tonhati et al., 2000; Rosati and Van Vleck, 2002), respectively. In spite of its higher fat percentage, milk and mozzarella cholesterol content is lower for buffalo than for cow's milk (275 *vs* 330 mg and 1562 *vs* 2287 mg, respectively) (Zicarelli, 2004). This is of major interest, together with some studies that report a larger number of small fat globules in buffalo milk as compared to bovine and sheep milk. It is well known that small fat globules are rich in polyunsaturated fatty acids (Martini et al., 2003).

Improvement through the selection of traits associated with milk quality and milk yield for dairy buffaloes is dependent on the availability of reliable genetic parameter estimates for these traits. The estimates of genetic parameters are helpful in determining the method of selection to predict direct and correlated response to selection, choosing a breeding system to be adopted for future improvement as well as in the estimation of genetic response. Moreover, the accuracy of genetic parameter estimates is determined by many factors, such as the quantity and quality of information (records and pedigree), the statistical model applied and the method of covariance estimation.

Several studies have reported genetic parameter estimates for buffalo milk yield (Khan et al., 2000; Bajwa et al., 2002; Rosati and Van Vleck, 2002; Hurtado-Lugo et al., 2006, Tonhati et al., 2008) and milk components such as protein and fat percentages (Tonhati et al., 2000; Rosati and Van Vleck, 2002). However, to date, there are few studies reporting genetic parameter estimates for mozzarella production and their relationships with milk yield and milk components (Rosati and Van Vleck, 2002). Although selection to increase milk yield has

received primary emphasis in the selection goals of dairy buffalo herds in Brazil, substantial emphasis has been placed on other traits, such as milk components and mozzarella production. There is concern among breeders about how selection to increase milk yield would affect milk components and mozzarella production, and thus, precise and accurate genetic correlation estimates could support breeders in selection decisions. Moreover, genetic correlation estimates are necessary to develop selection indexes for dairy buffaloes.

Compared to REML, Bayesian methods have the advantage of allowing the inclusion of prior knowledge about unknown parameters in the analysis. Additionally, the Bayes theorem provides a solution for the finite sample size problem, since an exact *a posteriori* distribution exists for each large or small data set from which inferences can be drawn (Jara and Barria, 1999; Misztal, 2008). When a large data set is analyzed, *a priori* information tends to be overwhelmed by the likelihood function in the establishment of the *a posteriori* distribution. In this case, parameter estimates are close to those obtained by frequentist methods based on likelihood functions. However, this may not be true when the sample size is limited because the maximum likelihood procedure only possesses well-defined properties when the sample size is large enough (Gianola and Fernando, 1986).

The objective of this study was to determine estimates of genetic parameters for milk yield, and fat and protein percentages and their relationship with mozzarella cheese production using Bayesian inference.

MATERIAL AND METHODS

A total of 4907 lactation records from 1986 Murrah buffaloes (*Bubalus bubalis*), belonging to 12 herds in São Paulo State, Brazil, and calving from 1985 to 2008, were analyzed. The data belong to the Dairy Bubaline Test Program developed by the Animal Science Department of São Paulo State University, FCAV, Jaboticabal, SP, Brazil. The traits studied were: accumulated 305-day milk yield (MY), milk fat (%F) and protein (%P) percentages, and accumulated 305-day mozzarella production (MP).

For all traits, the lactation records were unadjusted for days in milk and records with a length above 305 days were truncated at this point, as suggested by Tonhati et al. (2008). The %F and %P values were obtained averaging the monthly test-day records per lactation (Tonhati et al., 2000). As proposed by Altiero et al. (1989), MP was based on MY, %F and %P, using the formula adopted in genetic evaluation programs in Italy. In the present study, MP is accumulated at 305 days, because it is based on MY at 305 days.

$$MP \ (kg) = MY * \{[(3.5*\%P) + (1.23*\%F) - 0.88] / 100\}$$

Lactation records shorter than 90 days of lactation were deleted. It was considered that the first test-day milk record was measured from the 5th to the 75th day after calving. Cow's age varied from 2 to 8 years. The contemporary groups (CG) were defined as animals that calved in the same herd, year and season (season = 1 from April to September and season = 2 from October to March). The CG with less than 5 lactations (animals), and animals with records for each trait above or below 3.0 standard deviations from CG average were deleted from the analyses. There were 11,632 animals in the relationship matrix with 140 sires, and all generations back were utilized. The structure of the data is presented in Table 1.

Table 1. Summary of data structure and descriptive statistics for milk yield (MY), mozzarella production (MP) and milk fat (%F) and protein (%P) percentages.

Description	MY	%F	%P	MP
Number of records	4907	1216	1216	1216
Sires	140	103	103	103
Cows	1986	572	572	572
Mean	1614	6.84	4.27	382.30
SD	595.43	0.99	0.30	130.27
CV(%)	38.68	14.53	7.04	34.07
No. GC	197	49	49	49

SD = standard deviation; CV = coefficient of variation; CG = contemporary groups.

The (co)variance components were estimated by Bayesian inference in multi-trait analyses, with a repeatability animal model, using the GIBBS2F90 software (Misztal, 2007). The analysis model included fixed effects of CG, number of milkings (two levels) and the covariable buffalo cow's age at calving (linear and quadratic effects). The additive genetic and permanent environmental effects were included as random effects. The model adopted, represented in the matrix notation, was:

$$y = X\beta + Za + Wc + e \qquad \text{(Model 1)}$$

where, y is a vector of observed traits; X is the incidence matrix of fixed effects; β is a vector of fixed effects; Z is the incidence matrix of additive genetic random effects; a is a vector of additive genetic random effects; W is the incidence matrix of permanent environmental random effect; c is a vector of permanent environmental random effects, and e is a vector of random error effects.

Uniform *a priori* distribution was defined for fixed effects (β). Gaussian and inverted Wishart distributions were defined as *a priori* distributions for random effects and (co)variance components (Van Tassel and Van Vleck, 1996), respectively.

$$\beta \propto cons\tan t \qquad\qquad a\,|\,G \sim MVN[0,(G \otimes A)]$$

$$p\,|\,P \sim MVN[0,(P \otimes I_n)] \qquad G\,|\,S_g,v_g \sim IW[S_g v_g, v_g] \qquad \text{(Model 2)}$$

$$P\,|\,S_p,v_p \sim IW[S_p v_p, v_p] \qquad R\,|\,S_r,v_r \sim IW[S_r v_r, v_r]$$

where A, G, P, R, and I_n are the matrices of relationship, (co)variances of additive genetic effects, permanent environmental effects, residual, and identity, respectively; $\otimes$ is the Kronecker product and S_g and v_g, S_p and v_p, S_r and v_r are the *a priori* values and degrees of freedom for additive genetic, permanent environmental and residual (co)variances, respectively.

To verify the impact of selection to MY, %F and %P on MP, the genetic gain for MP was calculated applying MY, %F and %P as selection criteria. Different scenarios were established, as follows: the selection sire intensity (1, 5 and 10% of the sires were selected) and also the number of daughters/sires used (20, 50, 100, and 200 daughters/sire). It was assumed that the selection intensity for females was equal to zero.

A total of 2.0 million samples were generated in the analyses and, a burn-in period of 100,000 samples, with samples taken every 50 cycles. The convergence was verified through graphical inspection (trace plots) and also using the Gibanal program (Van Kaam, 1997). The descriptive statistics (mean, median, mode, and standard deviation) of *a posteriori* distribution for each parameter were obtained from effective samples. The highest posterior density (HPD) region or confidence interval provides the interval that includes 95% of samples and is a measure of reliability. Also, the HPD can be applied to non-symmetric distributions (Hyndman, 1996).

RESULTS AND DISCUSSION

For all parameters estimated, the convergence was confirmed by the inspection of trace plots. The auto-correlation among samples showed different values for each parameter estimated, since the numbers of effective samples varied from 63 to 9999. In general, the estimates of correlation between traits showed the lowest numbers of effective samples and the widest HPD, since there are few animals with records at least for two traits at the same time. Therefore, caution is needed in order to draw conclusions about correlation estimates between traits.

The mean *a posteriori* distribution for heritability estimates was moderate for all traits, varying from 0.23 to 0.38 (Table 2). These results pointed out that a great part of total phenotypic variation is due to the additive genetic action of the genes. Therefore, considerable genetic gain is expected if selection is applied for any trait.

Table 2. *A posteriori* means and highest *a posteriori* density region of variance components and genetic parameter estimates for milk yield (MY), milk fat percent (%F), milk protein percent (%P), and mozzarella production (MP).

Parameter	MY	%F	%P	MP
σ_a^2	46,669.5	0.23	0.03	2,094.9
	[28,800-63,400]	[0.15-0.35]	[0.02-0.04]	[1,172-3,005]
σ_{ap}^2	33,926.64	0.24	0.02	1,637.5
	[21,210-48,050]	[0.15-0.32]	[0.01-0.03]	[880.6-2,476.0]
σ_{at}^2	103,258.4	0.30	0.03	5,483.71
	[95,210.0-111,300]	[0.26-0.33]	[0.03-0.03]	[5,026.0-5,945.0]
σ_f^2	183,572.8	0.77	0.08	9,230.6
	[170,260-197,070]	[0.69-0.86]	[0.07-0.09]	[8,505-9,944]
h^2	0.25	0.30	0.38	0.23
	[0.17-0.34]	[0.18-0.42]	[0.23-0.50]	[0.14-0.32]
c^2	0.18	0.32	0.26	0.18
	[0.11-0.26]	[0.21-0.42]	[0.14-0.38]	[0.09-0.26]

σ_a^2 = additive genetic variance; σ_{ap}^2 = permanent environmental variance; σ_{at}^2 = temporary environmental variance; σ_f^2 = phenotypic variance; h^2 = heritability; c^2 = permanent environmental estimates as proportion of phenotypic variance.

The MY heritability estimate was close to those obtained by Tonhati et al. (2000, 2008), who also worked with Murrah buffaloes but utilized REML, which resulted in values of 0.24 and 0.28, respectively. However, lower MY heritability estimates (0.17 and 0.14) than that obtained in the present study, were reported by Thevamanoharan et al. (2000) and Rosati and Van Vleck (2002), with Nili-Ravi buffaloes in Pakistan and Mediterranean buffaloes in Italy, respectively. Differences in genetic variability of breed, in methodology applied to estimate the (co)variance components, and also in the production system (management and environment), probably explain the divergence between the results obtained in the present study and those reported in the literature.

The %F and %P heritability estimates were higher than those reported in previous studies. Tonhati et al. (2000) and Rosati and Van Vleck (2002) working with dairy buffaloes and applying REML, reported lower heritability estimates for %F and %P than those obtained in this study, the values being 0.21 and 0.26, 0.14 and 0.17, respectively. For all traits, the permanent environmental estimates as a proportion of phenotypic variance (c^2 estimates) were moderate (Table 2). With the exception of %P, all c^2 estimates were similar to those reported by Rosati and Van Vleck (2002), being 0.24, 0.29, 0.12, and 0.22 for MY, %F, %P, and MP, respectively.

The genetic correlation estimates between MY and %F and %P were low and negative (Table 3). These results agree with those reported by Tonhati et al. (2000) and Rosati and Van Vleck (2002), who reported genetic correlation estimates between MY and %F and %P, which were -0.18/-0.23 and -0.08/-0.12, respectively. These results showed that the genes that affect MY have an antagonistic effect on %F and %P traits. It suggests that selection to increase MY, would in the long term probably cause a reduction in milk constituents (%F and %P). In dairy cattle, De Paula et al. (2008) reported the same trend for the genetic correlations between milk yield and milk components. The genetic correlation estimates between %F and %P were high and positive. Similar results were obtained by Tonhati et al. (2000) (0.50) and Rosati and Van Vleck (2002) (0.31). Therefore, selection for higher %F in milk would also increase %P.

Table 3. *A posteriori* means and highest *a posteriori* density region of genetic (above the diagonal) and phenotypic correlation estimates (below the diagonal) among MY, %F, %P, and MP.

	MY	%F	%P	MP
MY	-	-0.19 [(-0.50)-(0.13)]	-0.20 [(-0.49)-(0.13)]	0.89 [(0.83)-(0.94)]
%F	-0.09 [(-0.16)-(-0.02)]	-	0.48 [(0.14)-(0.84)]	0.22 [(-0.13)-(0.51)]
%P	-0.14 [(-0.22)-(-0.08)]	0.32 [(0.24)-(0.38)]	-	0.19 [(-0.10)-(0.49)]
MP	0.95 [(0.94)-(0.5)]	0.15 [(0.09)-(0.22)]	0.11 [(0.04)-(0.17)]	-

MY = milk yield; %F = percent of milk fat; %P = percent of milk protein; MP = mozzarella production.

The genetic correlation estimates between MY and MP were high and positive, and close to those reported by Rosati and Van Vleck (2002) (0.95). In agreement with Rosati and Van Vleck (2002), high genetic correlation estimate between MP with MY was obtained since MY is the main component of the formula applied to calculate MP. However, the genetic correlations between MP and %F and %P were positive and moderate. Rosati and Van Vleck (2002) obtained higher genetic correlations between MP and %F (0.62) and %P (0.87) than those reported in the present study.

The permanent environmental correlation estimates were low between milk constituents and MP and high between MY and MP (Table 4). A similar permanent environmental correlation estimate between %P and MP (-0.03) was observed by Rosati and Van Vleck (2002), but lower permanent environmental correlation estimates than those obtained in the present study were reported between %F and MP (0.04) and between MY and MP (0.54).

Table 4. *A posteriori* means and highest *a posteriori* density region of permanent environmental (above the diagonal) and temporary environmental correlation estimates (below the diagonal) among MY, %F, %P, and MP.

	MY	%F	%P	MP
MY	-	-0.10 [(-0.43)-(0.19)]	-0.27 [(-0.60)-(0.05)]	0.95 [(0.92)-(0.99)]
%F	-0.05 [(-0.13)-(0.03)]	-	0.22 [(-0.02)-(0.62)]	0.16 [(-0.18)-(0.46)]
%P	-0.06 [(-0.14)-(0.02)]	0.21 [(0.13)-(0.29)]	-	-0.04 [(-0.40)-(0.32)]
MP	0.96 [(0.95)-(0.97)]	0.12 [(0.04)-(0.19)]	0.12 [(0.05)-(0.20)]	-

MY = milk yield; %F = percent of milk fat; %P = percent of milk protein; MP = mozzarella production.

The genetic gain estimates for MP applying MY, %F, %P, and MP as selection criteria are presented in Figure 1. Applying MY or MP as selection criteria, the genetic response obtained for mozzarella production was similar. These results were expected, since the heritability estimates for both traits were similar and the genetic correlation estimates was high between MY and MP. However, applying milk constituents (%F and %P) as selection criteria to improve mozzarella production, the genetic gain obtained for this trait was low. Although these traits have a large additive genetic influence, the genetic correlation estimates between MP and %F and %P were low and, thus, the correlated response obtained for MP was also low.
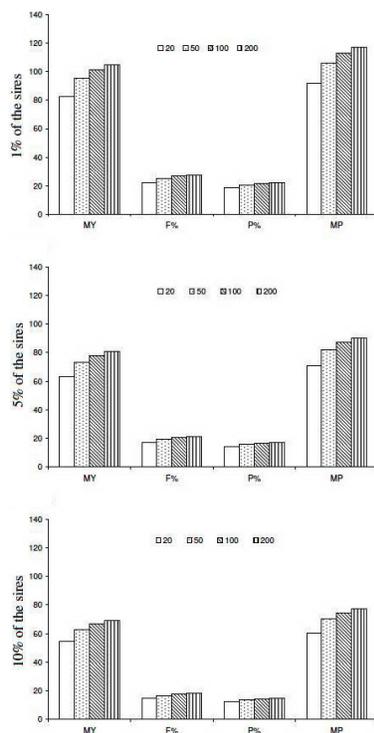

Figure 1. Expected direct response to selection for mozzarella production (MP) and correlated response for MP with direct selection for total milk yield (MY), milk fat (%F) and protein percentage (%P) varying the intensity of sire selection (1, 5 and 10% of the sires were selected) and also the number of daughters/sires used (20, 50, 100, and 200 daughters/sire).

For dairy buffalo herds in which the aim is mozzarella production, MY could be an effective selection criterion to increase the mozzarella production. The milk constituents were not useful criteria to increase mozzarella production. However, Seno et al. (2007) found that the economic values of milk constituents were high, and for a long-term selection, a selection index combining MY, %F and %P was a suitable alternative for the dairy industry.

In the simulated selection intensity scenarios, the number of daughters/sire influenced the genetic response. As expected, for all traits, as selection intensity increased the genetic response was higher. For %F and %P, the genetic response stabilized at 100 daughters/sire. However, for MP and MY, considerable genetic responses were obtained for over 100 daughters/sire, since heritability estimates for these traits were lower than those obtained for milk constituents (%F and %P).

CONCLUSION

Milk yield, milk components and mozzarella production in Murrah buffaloes have enough genetic variation for selection purposes. The results obtained in the present study suggest that selection to increase milk yield would be effective to improve mozzarella production.

ACKNOWLEDGMENTS

REFERENCES

Altiero V, Moio L and Addeo F (1989). Previsione della resa in "mozzarella" sulla base del contenuto in graso e proteine del latte di bufala. *Sci. Tec. Lattiero-Casearia* 40: 425-433.

Bajwa IR, Khan MS, Ahmad Z and Gondal KZ (2002). Genetic parameters of lactation milk yield as affected by lactation length adjustment procedures. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, August 19-23, 2002, Montpellier, France.

De Paula MC, Martins EN, Silva LOC, Oliveira L, et al. (2008). Estimates of genetic parameters for yield and composition of milk of Holstein cows in Paraná State. *Braz. J. Anim. Sci.* 37: 824-828.

Gianola D and Fernando RL (1986). Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63: 217-244.

Hurtado-Lugo N, Cerón-Muñoz M and Gutierrez-Valencia A (2006). Estimacion de Parametros Geneticos para la Produccion de Leche en el Dia del Control en Bufalos de la Costa Atlantica Colombiana. Livestock Research for Rural Development, 18. Available at [http://www.lrrd.org/lrrd18/3/hurt18039.htm]. Accessed March 12, 2009.

Hyndman RJ (1996). Computing and graphing highest density regions. *Am. Stat.* 50: 120-126.

Jara A and Barria N (1999). Bayesian methods in animal breeding. *Adv. Anim. Prod.* 24: 3-19.

Khan MS, Chaudhry HZ and Raza SH (2000). Impact of lactation length adjustment procedures on accuracy and heritability of adjusted milk yield in buffaloes. *J. Dairy Sci* 83 (Suppl 1): 223.

Macedo MP, Wechsler FS, Ramos AA, Amaral JB, et al. (2001). Composição físico-química e produção do leite de búfalas da raça Mediterrâneo no oeste do Estado de São Paulo. *Rev. Bras. Zoot.* (Suppl 1) 30: 1084-1088.

Martini M, Spinelli S, Scolozzi C and Cecchi F (2003). Studio sulle caratteristiche lipidiche del latte di bufale allevate in Toscana: nota II. Atti II Congresso Nazionale Sull'allevamento del Bufalo, Roma, 28-30 Agosto, 147-151.

Misztal I (2007). BLUPF90 family of programs. Available at [http://nce.ads.uga.edu/~ignacy/newprograms.html]. Accessed March 12, 2009.

Misztal I (2008). Reliable computing in estimation of variance components. *J. Anim. Breed. Genet.* 125: 363-370.

Rosati A and Van Vleck LD (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: 185-190.

Seno LO, Cardoso VL and Tonhati H (2007). Economic values for milk production traits of dairy buffaloes in São Paulo State. *Braz. J. Anim. Sci.* 36: 2016-2022.

Thevamanoharan K, Vandepitte W, Mohiuddin G and Shafique M (2000). Genetic, phenotypic and residual correlation between various performance traits of Nili-Ravi buffaloes. *Buffalo Bull.* 19: 80-86.

Tonhati H, Cerón-Muñoz MF, Oliveira JA, Duarte JMC, et al. (2000). Genetic parameters of milk production, fat and protein contents in buffalo milk. *Braz. J. Anim. Sci.* (Suppl 1) 29: 2051-2056.

Tonhati H, Cerón-Muñoz MF, Oliveira JA, El Faro L, et al. (2008). Test-day milk yield as a selection criterion for dairy buffaloes (*Bubalus bubalis* Artiodactyla, Bovidae). *Genet. Mol. Biol.* 31: 674-679.

Van Kaam JBCHM (1997). Gibanal: Analyzing program for Markov Chain Monte Carlo sequences, version 2.4. Dept. Anim. Sci., Wageningen Agric. Univ., Wageningen.

Van Tassel CP and Van Vleck LD (1996). Multiple-trait Gibbs sampler for animal models; flexible programs for Bayesian and likelihood-based (co)variance components inference. *J. Anim. Sci.* 74: 2586-2597.

Zicarelli L (2004). Buffalo milk: its properties, dairy yield and mozzarella production. *Vet. Res. Commun.* (Suppl 1) 28: 127-135.