

## **Genotype by environment interaction in Brazilian dairy Gir cattle**

*J.C.C. Panetto<sup>1</sup>, L.L. Verardo<sup>1, 2</sup>, M.A. Machado<sup>1</sup>, M.F. Martins<sup>1</sup>, D.R.L.R. Faza<sup>1</sup> & M.V.G.B. Silva<sup>1</sup>*

*<sup>1</sup>Embrapa Gado de Leite, 36038-330, Juiz de Fora, Minas Gerais, Brazil*

*joao.panetto@embrapa.br (Corresponding Author)*

*<sup>2</sup>Departamento de Ciências Exatas, Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, São Paulo, Brazil*

### **Summary**

With the objective of exploring genotype by environment interaction in the Brazilian National Dairy Gir Breeding Program, a total of 97,476 lactation records were separated into three data sets, according to the average milk yield within each management group. They were designated as low input, medium input, or high input management groups. Breeding values were predicted for 305-d milk yield using records from those three data sets as three different traits: Low input 305-d milk yield; Medium input 305-d milk yield; and High input 305-d milk yield. Genetic correlations ranged from 0.75, between the low input and high input traits, to 0.97, between the medium input and high input traits. Reordering of ranking of sires has been observed, especially in the comparison between the low input and high input EBVs. SNP genotypes were obtained for a sample of 2,681 animals, including animals with lactation records (own or from their progeny) within the three data sets. The EBVs of the genotyped animals were later deregressed and used in three Genome Wide Association Studies (GWAS). Different regions were associated with 305-d milk yield while considering the three production scenarios as observed in chromosomes 11 and 21 that were more strongly associated with milk yield in the low input data set than in the other two.

*Keywords: genetic correlation, GWAS, heritability, management groups*

### **Introduction**

Genotype by environment interaction (G x E) is known to influence animals performance in dairy production systems, thus affecting the estimation of individuals breeding values and their suitability in different production scenarios of herds in the context of breeding programs. Heat stress, parasites infection and infestation, poor management or lack of food, are factors that can affect animal's performance in dairy production systems. Genetic parameters of production traits might also be affected, influencing the prediction of breeding values. Bignardi et al. (2015) observed that estimated breeding values were reordered due to G x E while evaluating Holstein cows under tropical conditions. Also, Santana et al. (2015) observed a slight increase of milk yield heritability under higher temperature/humidity indices, and a negative genetic correlation between milk yield and tolerance to heat stress in dairy Gir cows. On the contrary, Fuerst-Waltl et al. (2013) found that genetic correlations between production and functional traits of dairy cattle in Austria, measured in low or high productivity systems, were mostly close to unity, suggesting no need of distinct breeding programs according to the management system.

This study aimed to clarify whether G x E should be considered in the genetic evaluation of

dairy Gir cattle in Brazil, and to identify similarities and dissimilarities in genomic regions associated with milk yield in management groups of different levels of productivity.

## Materials and methods

### Phenotype and genotype data

Production and pedigree information were obtained from the Brazilian breeding program for the dairy Gir (Panetto et al., 2017), and included 97,476 lactation records from 59,271 cows, in 6,803 management groups, and also a related genealogy file containing 116,365 animals over up to 9 generations.

Management groups, formed by herd-year of calving, were separated in three clusters, according to their average 305-d milk yield (305MY), in a way that the three resulting data sets comprised approximately the same number of animals, with the following criteria: Low input (LI) < 2,600 kg ≤ Medium input (MI) < 4,000 kg ≤ High input (HI). Environmental values were defined according to the management group averages, because it was readily available from milk recording data and it was a general measure of a complex of environmental factors, as stated by Kolmodin et al. (2002). The numbers of females with phenotypes within each set used on the estimation of genetic parameters and breeding values are illustrated in Figure 1a. Sampled animals were genotyped, representing each of the three levels of productivity, including females with lactation records and males with at least four offspring with lactation records. The same animals could be evaluated within different clusters, including sires with more than one daughter or females with lactations in different management groups. The numbers of genotyped animals used on the GWAS are illustrated in Figure 1b. Both figures describe each data set and the numbers of animals in common among them.

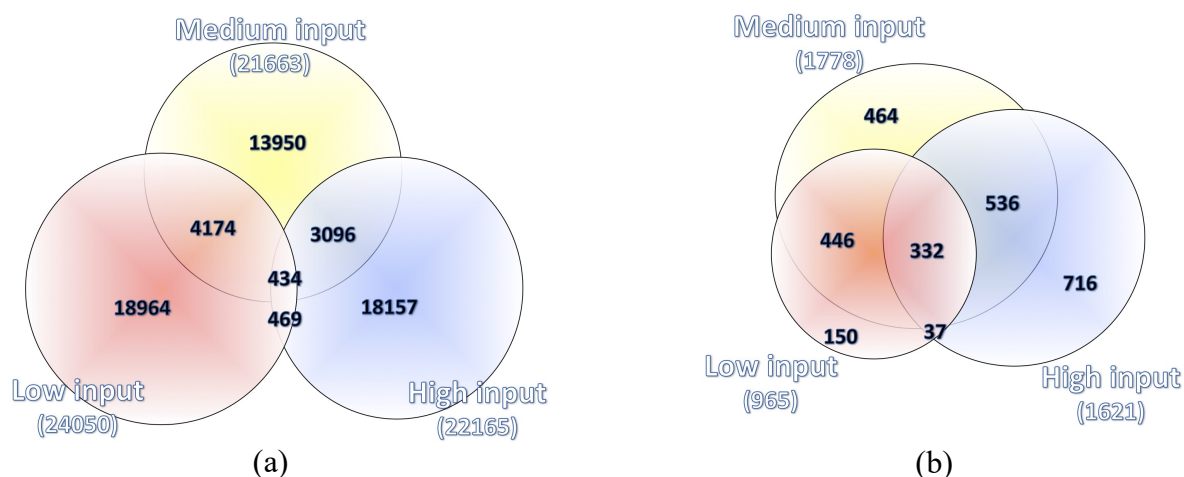


Figure 1. Venn diagrams illustrating the data structure of the three 305-d milk yield traits analysed, including low input, medium input or high input management groups: (a) Phenotyped animals, used on the estimation of genetic parameters and breeding values; (b) genotyped animals, used on the GWAS.

The low input 305MY group included 620 genotyped females and 345 genotyped males. The medium input 305MY group included 1,412 genotyped females and 366 genotyped males. The high input 305MY group included 1,355 genotyped females and 266 genotyped males. All males were

genotyped with the Illumina BovineHD (HD; Illumina, san Diego, CA), while all females were genotyped with the Illumina BovineSNP50 (50K) BeadChip v2. Female 50K genotypes were then imputed to HD, as described in Boison et al. (2017). Thus, all genotypes used in the present study were HD or imputed to HD.

## Prediction of breeding values

Three single-trait animal models were fitted, using respectively the low input, medium input, and high input management group data sets, to obtain 305MY estimated breeding values (EBV) for all animals in the pedigree. The EBVs of the genotyped animals were later deregressed and used in three Genome Wide Association Studies (GWAS).

One multi-trait animal model was fitted to analyze 305MY at the three different levels of productivity as three different traits. EBVs of high accuracy bulls were compared among levels of management groups. The heritabilities of each trait and the genetic correlations among them were also obtained from this analysis.

Models to analyze 305-d milk yield included the effects of management groups (herd-year of calving), genetic composition of cow (about 74% purebred Gir and 26% crossbreds Gir, containing ¼ to ¾ Holstein), the linear and quadratic terms of the regression of 305MY on age at calving, and the 305MY Legendre polynomial regressions of order 3 on the day of calving within year (1 to 365), accounting for the seasonal effect. Animal and permanent environment were both included as random effects. The Wombat software (Meyer, 2007) was used to estimate (co)variances, heritabilities, genetic correlations and breeding values (EBV), with the REML approach.

## GWAS

Estimated breeding values were deregressed following the procedure of Garrick et al. (2009), and applied to association analyses between each 305MY trait (LI, MI or HI) and the imputed HD genotype distributions, which were conducted using logistic regression models implemented with the “mlreg” function of the GenABEL package in R software (Aulchenko et al., 2007). Within this package, the genome-wide degree of inflation ( $\lambda$ ) was calculated to test for any hidden substructure. Thus, the final P-value was corrected by inflation factor and used to plot the SNP effects along the genome.

## Results and discussion

### Prediction of breeding values

The heritabilities estimated for the three traits ranged from 0.23 to 0.26, and the genetic correlations ranged from 0.75 to 0.97 (Table 1).

*Table 1. Heritabilities (diagonal) of 305-d milk yield measured within three different levels of average productivity within management groups (LI: low input; MI: medium input; HI: high input), and the genetic correlations amongst those three traits above the diagonal (standard errors of all estimates are between parenthesis)*

	LI_305MY	MI_305MY	HI_305MY
LI_305MY	0.23 (0.01)	0.88 (0.03)	0.75 (0.05)
MI_305MY		0.25 (0.01)	0.97 (0.02)
HI_305MY			0.26 (0.01)

The genetic correlation between MI\_305MY and HI\_305MY was close to unity, indicating that breeding values for 305-d milk yield would be almost the same for both production environments, with minor rearrangements of sire ranking in regard to this trait. On the other hand, the genetic correlations between LI\_305MY and the other two levels of productivity were different from the unity, especially between the LI and HI data sets, indicating possible rearrangements on sire ranking. Robertson (1959) stated that the quantitative expression of genotype-environment interactions in terms on genetic correlation between performances in two or more environments was of value in giving a measure of the practical, rather than the statistical, significance of the results, but he suggested a genetic correlation should fall to a figure around 0.8 before the genotype-environment interaction has biological or agricultural importance. A sample of 305-d milk yield estimated breeding values, for the three levels of productivity, are illustrated in Figure 2, including the 12 sires with the highest accuracies.

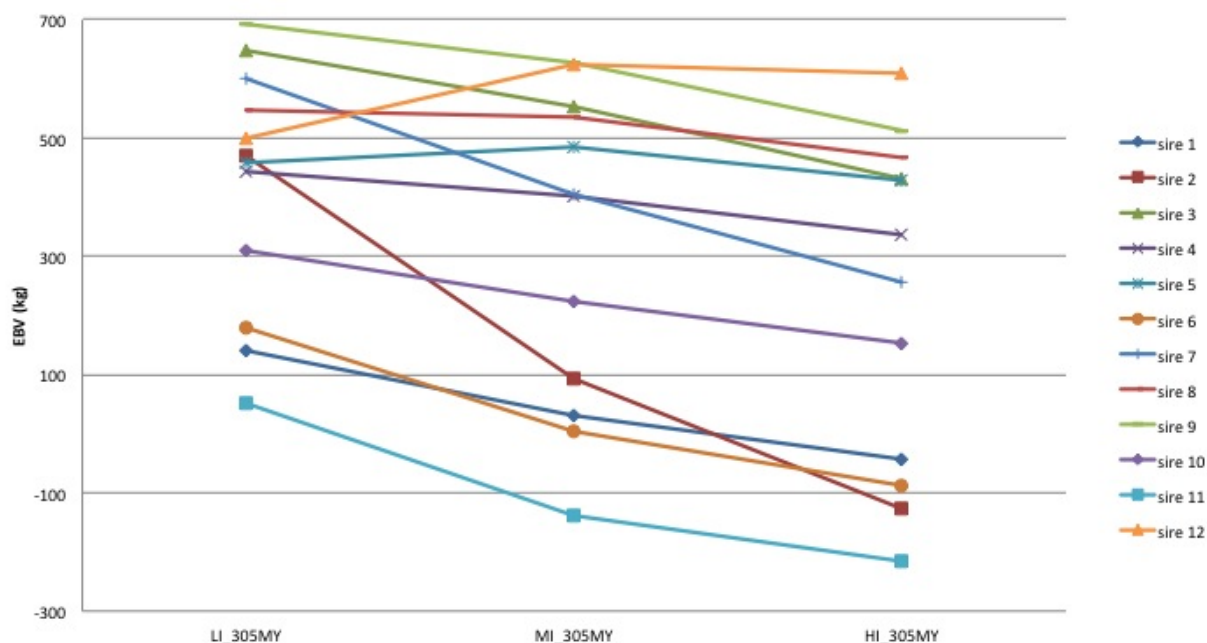


Figure 2. Top 12 accuracy sire's 305-d milk yield (305MY) estimated breeding values (EBV) for three different levels of productivity: Low Input (LI\_305MY), Medium Input (MI\_305MY), or High Input management groups (HI\_305MY)

The importance of the genotype-environment interaction, for 305MY, can be observed through the comparison between some of those higher accuracy animals. Most sires presented their best EBVs for the worst environment, distinctly sires 7 and 2, while sire 12 presented better EBVs for improved environments (MI or HI).

## GWAS

Different patterns were found across the three regression analyses, for the three data sets, as illustrated in the Manhattan plots of Figure 3.

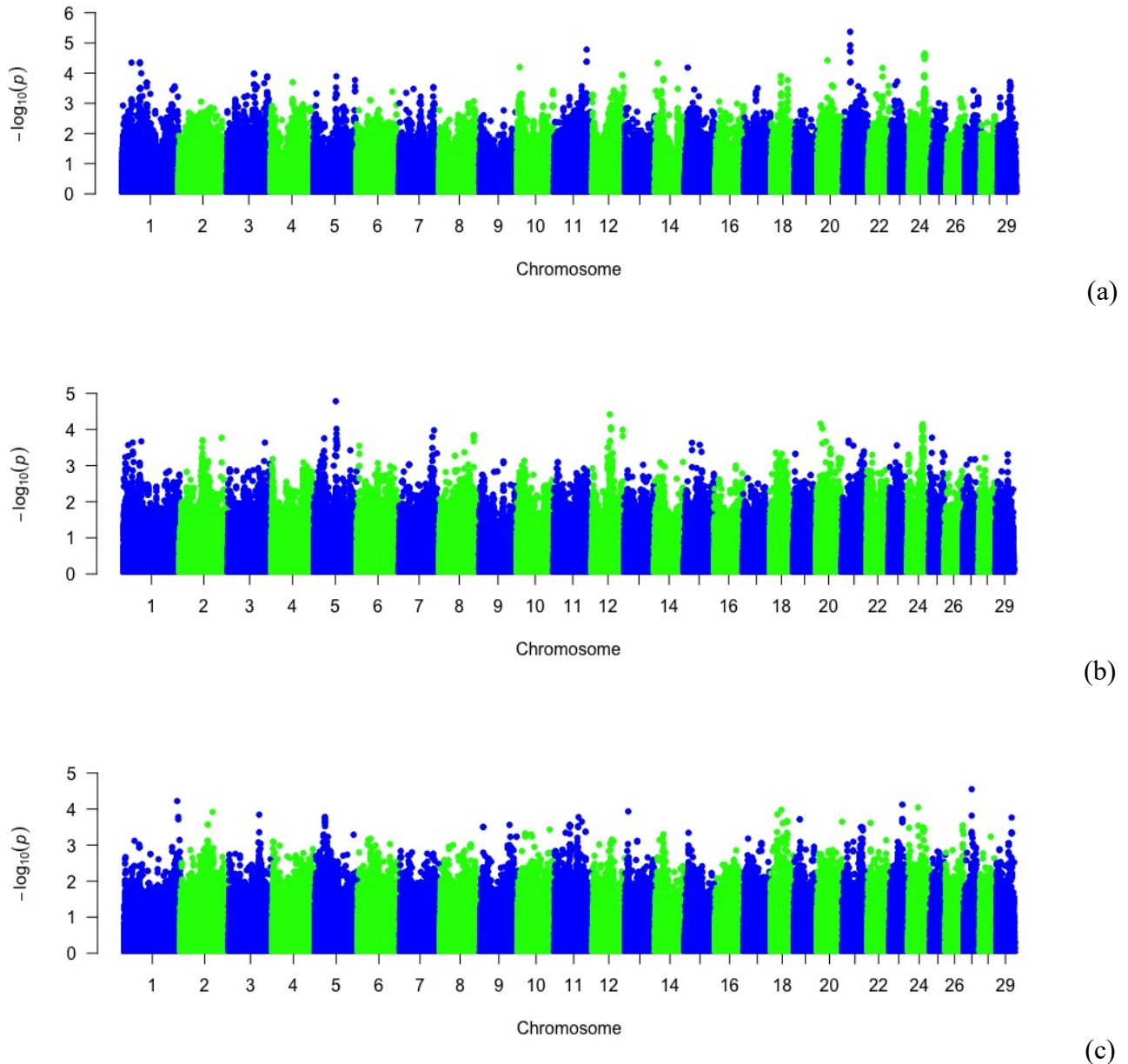


Figure 3. Manhattan plots illustrating SNP effects estimated from GWAS on three different levels of productivity for 305-d milk yield: a) LI: low input; b) MI: medium input; c) HI: high input

Considering the top 10 significant SNPs from each analysis, we observed relevant regions on chromosomes (BTA) 1, 2, 3, 5, 11, 12, 13, 18, 20, 21, 23, 24 and 27. See Table 2 for results. Only one SNP was common to the top 10 of two analyses simultaneously, which was the BovineHD2400012129, on BTA24, between the Low and the Medium input data sets. Several QTL for milk yield traits were also previously reported at the same or near chromosomal regions identified by the top 10 SNPs in this study for BTA 1, 3, 5, 11, 18, 21, 23, 24 and 27 (Michinet et al., 2016; Meredith et al., 2012; Awad et al., 2009; Olsen et al., 2002; Cole et al., 2011; Guo et al., 2012).

Table 2. Top 10 significant SNPs from each cluster (Low, Medium or High productivity levels), SNP chromosome (Chr) location and position in base pair (bp), P-value obtained from GenABEL, the

*closest genes and the distance in bp between SNPs and genes*

SNP	Chr	Position (bp)	P-value	Pop	Closest genes	Distance (bp)
BovineHD0100043227	1	149203674	3.17E-31	High	RUNX1	254621
BovineHD0200025824	2	90797534	8.36E-29	High	CDK15	Inside
BovineHD0300024417	3	85442743	3.27E-28	High	LOC104971724	Inside
BovineHD0500017331	5	61846208	8.21E-41	Medium	LOC104976959	169369
BovineHD1100026442	11	91160241	5.67E-23	Low	SOX11	230526
BovineHD1200013992	12	50895100	1.95E-37	Medium	TBC1D4	6196
BovineHD1300002478	13	9297527	6.17E-29	High	MACROD2	Inside
BovineHD1800006018	18	19364170	2.75E-28	High	CYLD	65655
BovineHD1800019617	18	30075661	2.85E-29	High	CDH8	212708
BovineHD2000004248	20	13238986	4.38E-35	Medium	SREK1	152739
BovineHD2100005846	21	20196212	1.03E-22	Low		1586
BovineHD2100005847	21	20197144	1.03E-22	Low		2518
BovineHD2100005849	21	20199454	1.07E-23	Low	LOC101905669	4828
BovineHD2100005852	21	20204462	5.85E-26	Low		9836
BovineHD2300010551	23	36379600	1.91E-30	High	SOX4	85303
BovineHD2400007752	24	28614217	8.34E-30	High	CDH2	378449
BovineHD2400011961	24	43161465	2.31E-34	Medium	GNAL	Inside
BovineHD2400011963	24	43166599	2.31E-34	Medium		
BovineHD2400011965	24	43177465	2.31E-34	Medium	MPPE1	Inside
BovineHD2400011967	24	43185216	2.31E-34	Medium		4457
BovineHD2400011968	24	43192358	2.31E-34	Medium	IMPA2	Inside
BovineHD2400011969	24	43195767	2.31E-34	Medium		Inside
BovineHD2400012122	24	43893570	6.73E-22	Low		
BovineHD2400012123	24	43896489	6.73E-22	Low		
BovineHD2400012124	24	43900901	6.73E-22	Low	LDLRAD4	Inside
BovineHD2400012129	24	43919225	3,69E-22/6,05E-35	Low/Medium		
BovineHD2400012130	24	43921662	2.65E-22	Low		
BovineHD2700006349	27	22483556	6.98E-34	High	C27H8orf48	15734
BovineHD2700013577	27	22552624	5.22E-28	High	ENSBTAG00000024526	17619

In general, our results illustrate that different regions (genes) of the genome are the most relevant for the trait (milk yield) across distinct clusters, when it comes to the level of input. It seems that different genes play their biological roles with distinct importance in the expression of the phenotype, resulting in different sets of most relevant SNPs within different production circumstances. However, it is important to note that, in this preliminary observation, we were very restrictive (top 10 SNPs from each production level). When considering the top 0.1% significant SNPs (top 400 SNPs), for example, we might observe other common SNPs across the different productive levels clusters, and that kind of approach might be considered in further studies. Moreover, reordering on the ranking of sires and different SNPs associated with milk yield, according to different levels of productivity of the management groups from which daughters were measured, suggest that genotype by environment interaction is an important phenomena to be considered in genetic evaluation of the dairy Gir cattle in Brazil.

## Acknowledgements

This study and the authors received financial support from EMBRAPA (Brazilian Agricultural Research Corporation), CNPq: INCT-CA (National Council for Scientific and Technological Development: Animal Science National Institute); FAPEMIG (Minas Gerais Research Foundation); and CAPES (Coordination for the Improvement of Higher Education Personnel). The authors would

like to thank the ABCZ (Brazilian Association of Zebu Breeders) for providing the purebred production records and genealogy data files and the ABCGIL (Brazilian Association of the Dairy Gir Breeders) for the crossbred production records and their important contribution to the contact with breeders for the DNA sampling.

## References

- Awad, A, I. Russ, R. Emmerling, M. Förster & I. Medugorac 2009. Confirmation and refinement of a QTL on BTA5 affecting milk production traits in the Fleckvieh dual purpose cattle breed. *Anim. Genet.* 41(1):1-11
- Aulchenko, Y.S., S. Ripke, A. Isaacs & C.M. van Duijn 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 10:1294–6
- Bignardi, A.B., L. El Faro, R.J. Pereira, D.R. Ayres, P.F. Machado, L.G. Albuquerque & M.L. Santana Jr. 2015. Reaction norm model to describe environmental sensitivity across first lactation in dairy cattle under tropical conditions. *Trop. Anim. Health Prod.*, 47:7, 1405–1410
- Boison, S.A., A.T.H. Utsunomiya, D.J.A. Santos, H.H.R. Neves, R. Carvalheiro, G. Mészáros, Y.T. Utsunomiya, A.S. do Carmo, R.S. Verneque, M.A. Machado, J.C.C. Panetto, J.F. Garcia, J. Sölkner & M.V.G.B. da Silva 2017. Accuracy of genomic predictions in Gyr (*bos indicus*) dairy cattle. *J. Dairy Sci.* 100:1-12
- Cole J.B., G.R. Wiggans, Li Ma, T.S. Sonstegard, T.J. Lawlor Jr, B.A. Crooker, C.P. Van Tassell, J. Yang, S. Wang, L.K. Matukumalli & Y. Da 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics*, 12:408
- Fuerst-Waltl, B., H. Schwarzenbacher & C. Fuerst 2013. Genetic relationships under different management systems and their consequences for dairy cattle breeding. *Agriculturae Conspectus Scientificus*, 78:3, 183-186
- Garrick, D.J., J.F. Taylor & R.L. Fernando 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Gen. Sel. Evol.*, 41:55
- Guo J., H. Jorjani & Ö. Carlborg 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BMC Genet.*, 13:82
- Kolmodin R., E. Strandberg, P. Madsen, J. Jensen & H. Jorjani 2002. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. *Acta Agriculturae Scandinavica, Section A - Animal Science*, 52:1, 11-24
- Meredith B.K., F.J. Kearney, E.K. Finlay, D.G. Bradley, A.G. Fahey, D.P. Berry & D.J. Lynn 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genet.* 13:21
- Meyer K, 2007 WOMBAT – A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8:815-821
- Michenet A., R. Saintilan, E. Venot & F. Phocas 2016. Insights into the genetic variation of maternal behavior and suckling performance of continental beef cows. *Genet Sel Evol.* 48:45
- Olsen H.G., L. Gomez-Raya, D.I. Våge, I. Olsaker, H. Klungland, M. Svendsen, T. Ådnøy, A. Sabry, G. Klemetsdal, N. Schulman, W. Krämer, G. Thaller, K. Rønningen & S. Lien 2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. *J. Dairy Sci.* 85(11):3124-30
- Panetto, J.C.C., R.S. Verneque, M.V.G.B. Silva, M.A. Machado, M.F. Martins, F.A.T. Bruneli, M.G.C.D. Peixoto, G.G. Santos, W.A. Arbex, D.R.L.R. Faza, C.C. Geraldo, C.H.C. Machado, H.T. Ventura, M.A. Pereira, A.E. Vercesi Filho, R.S. Maciel & A.R. Fernandes 2017. Programa

Nacional de Melhoramento do Gir leiteiro: Sumário de Touros: Resultado do Teste de Progênie – 8ª Prova de Pré-seleção de Touros. Juiz de Fora: Embrapa Gado de Leite, 2017. 96 p. (Embrapa Gado de Leite. Documentos, 202).

Robertson, A. 1959. The sampling variance of the genetic correlation coefficient. *Biometrics*. 15:3, 469-485.

Santana Jr., M.L., R.J. Pereira, A.B. Bignardi, A.E. Vercesi Filho, A. Menéndez-Buxadera & L. El Faro 2015. Detrimental effect of selection for milk yield on genetic tolerance to heat stress in purebred Zebu cattle: Genetic parameters and trends. *J. Dairy Sci.* 98:12, 9035-9043