Genome Wide Association Study for Calving Interval in Gyr Dairy Cattle

M. A. Machado,[1] A. T. H. Utsunomiya,[2,3] S. A. Boison,[2] D. J. A. dos Santos,[2,3] Y. T. Utsunomiya,[2] R. da Fonseca,[4] J. Sölkner,[2] J. F. Garcia,[5] R. S. Verneque,[1] M. V. G. B. da Silva[1]

[1]Brazilian Agricultural Research Corporation, Juiz de Fora-MG, Brazil, [2]BOKU – University of Natural Resources and Life Sciences, Vienna, Austria, [3]UNESP – Univ Estadual Paulista, Jaboticabal-SP, Brazil, [4]UNESP – Univ Estadual Paulista, Dracena-SP, Brazil, [5]UNESP – Univ Estadual Paulista, Araçatuba-SP, Brazil.

ABSTRACT: We used the Heterocedastic Effects Model (HEM) to perform a genome-wide scan for loci affecting calving interval in Gyr dairy cattle. HEM is an extension of SNP-BLUP that allows for marker-specific shrinkages in order to improve QTL mapping. The 20 SNPs with the largest estimated effects accounted for 1.32% of the total additive genetic variance. In particular, markers *rs41888601* and *rs137830830* are located on the chromosome domains encompassing *SMG9*, *SLC39A11* and *SOX*, which are known to be involved in oestrus cycle and fertility. As HEM relaxes the penalization of large effects SNPs, the low percentage of additive genetic variance explained by these markers brings further evidence that calving interval is highly polygenic and controlled by many QTLs.

Key words: GWAS; calving interval; HEM model; Gyr.

INTRODUCTION

Age at first calving, calving interval, lactation length are some of the traits determining productive life and culling decisions in dairy herds (Hare et al. (2006)). These traits impact the costs of dairy and beef farming due to the need for replacing culled animals. In spite of their economic importance, these traits have low heritability and therefore are difficult to improve.

Mapping loci explaining genetic variance can be of help to improve fertility traits such as calving interval (CI). Minozzi et al. (2013) have recently found genomic regions in Bos taurus dairy cattle that potentially influence calving interval (CI). However, the additive genetic variance reported accounted for only a small fraction of the total variance in CI, and many small effects QTLs may be underlying the genetic differences. Arguably, the extent of overlap between loci explaining variance in CI in *Bos taurus* and *Bos indicus* must be determined in order to evaluate how generalizable these results are across bovine species. Therefore, the objective of this study was to perform a genome-wide scan for calving interval in Gyr cattle (*B. indicus*), and compare our findings with the putative loci affecting CI in *Bos taurus* cattle.

MATERIALS AND METHODS

Sample and Phenotype Data

Estimated breeding values (EBVs) were available from routine genetic evaluations which comprised of 28,133 individuals. EBVs were deregressed following Garrick et al. (2009), and considered as pseudo-phenotypes in the genome-wide association study. The mean, median and standard deviation of the deregressed EBV (dEBV) were 10.95, 12.95 and 25.41, respectively. The reliabilities of dEBV ranged from 0.23 to 0.74. Genome wide association study was thus performed with the dEBVs for the remaining 1606 cows.

Genotype Data, Quality Filtering and Imputation

A total of 457 bulls were genotyped using the Illumina® BovineHD BeadChip assay (HD) and 1,684 cows were genotyped using the Illumina® BovineSNP50 BeadChip assay (50K). Initial filtering of 50K genotypes resulted in the exclusion of ~28,000 markers. This drasctic drop in the number of SNPs was mainly due minor allele frequencies (<0.02), as a result of the well documented ascertainment bias of 50K. Subsequent use of the approximately 26,000 markers in genome-wide mapping was not successful. Therefore, cow genotypes were imputed by using a reference sample of 457 important founder bulls genotyped with the HD panel. Markers and samples were filtered using Plink v1.07 (Purcell et al., 2007), according to the following exclusion criteria: *Markers* - Call rate < 0.95; GenCall score < 0.50; minor allele frequency < 0.02 and Fisher's exact test p-value for Hardy Weinberg Equilibrium < 1e-06; *Samples* - Call rate < 0.90. After filtering, 452 sires remained along 490,009 SNP markers.

The imputaion procedure was performed using FImpute v2 (Sargolzaei et al., 2011) with prior pedigree information to link cows to sires. The maximum, mean top 5, 10 and 20 relationship between the sires and imputed cows were 0.42, 0.27, 0.22 and 0.17, respectively.

Genome Wide Association

The SNP effects were simultaneously estimated using the Heteroscedastic Effects Model (HEM), a generalized form of ridge regression BLUP that allows for marker-specific shrinkage (Shen et al., 2013). The following variance-components model was fitted:

$$Y = 1\mu + a + e$$

where y is the vector of deregressed EBVs, 1 is a vector of 1's, μ is an intercept, a is the vector of unobserved random polygenic effects, and e is a residual term. Polygenic effects were assumed $a \sim N\left(0, G\sigma_a^2\right)$, where G represents the realized genomic relationship matrix (VanRaden, 2008) and σ_a^2 is the additive genetic variance. Residuals were assumed $e \sim N\left(0, I\sigma_e^2\right)$, where σ_e^2 is the residual variance and I is an identity matrix. Variance components were estimated from the data using a penalized quasi-likelihood approach. Random SNP effects (b_j^0) were back-solved as $k^{-1}Z'G^{-1}\hat{a}$, where Z is the matrix of genotype scores (coded as $0\text{-}2p_i$, $1\text{-}2p_i$, $2\text{-}2p_i$, depending if the

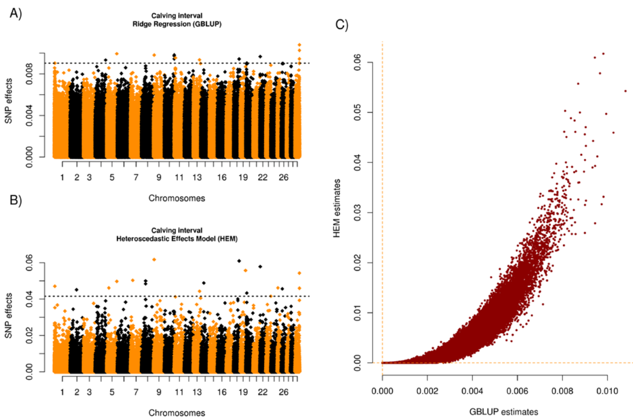

Figure 1. SNP-BLUP effect estimates (A), HEM effect estimates (B), relationship between HEM and SNP-BLUP models (C).

animals' genotype is AA, AB or BB, respectively, and p_i is the B allele frequency at marker i) and $k = 2\sum_{i-1}^{N} p_i\left(1-p_i\right)$. This procedure is known as SNP-BLUP. SNP specific shrinkages (w_{jj}) are calculated from the SNP-BLUP results as:

$$w_{jj} = \hat{\sigma}_{b_j}{}^{2} = \frac{b_j^{z2}}{1-h_{jj}}$$

where h_{jj}, known as the *hat value*, is the (n + j)-*th* diagonal element of the *hat matrix* $H = T(T'T)^{-1}T$, where

$$T = \begin{bmatrix} X & Z \\ 0 & diag\left(\lambda\right) \end{bmatrix}$$

The inverse of matrix $G^* = ZWZ'$ (where W is a diagonal matrix containing SNP-specific shrinkages) is used to re-estimate polygenic effects as described before, and SNP effects are again back-solved. This procedure is referred as HEM, and it takes into account the problem of high dimensionality of the data, where the number of parameters to be estimated is much larger than the number of observations.

Table 1. Description of the SNPs rs41888601 and rs137830830 associated with calving interval (CI).

SNP name	Chr	Gene name	Description
rs41888601	18	SMG9	smg-9 homolog, nonsens mediated mRNA decay Factor (C. elegans).
rs137830830	19	SLC39A11	Solute carrier family 39 (metal ion transporter), Member 11.
rs137830830	19	SOX9	SRY (sex determining region Y) – Box 9.

*Chromosome

Functional Studies

The gene content of windows of ±250kb around the 20 largest SNP effects (top20) was annotated. Gene coordinates in the UMD v3.1 assembly were obtained from the Ensembl genes 73 database using the BioMart tool (Kinsella et al., (2011)). The proportion of the additive genetic variance explained by top20 was also assessed by $2p_i\,(1-p_i)\,\hat{\beta}_i^{\,2}$, where p_i is the allele frequency and $\hat{\beta}_i$ is the allele substitution effect estimated by HEM.

RESULTS AND DISCUSSION

The SNP-BLUP estimates of allele substitution effects did not provide a clear indication of markers with distinctive effects on CI (Figure 1A). As discussed by Shen et al. (2013), this was due to the assumption of equal variance across markers. By applying a different shrinkage to each marker, the HEM method could pinpoint loci contributing more to the additive genetic variance then the genome background (Figure 1B). Still, in spite of the benefit of discriminating better the markers with the largest effects on CI, the proportion of genetic variance explained by these loci is small, being consistent with a lowly heritable ($h^2 = 0.07$) and highly polygenic trait.

The estimated percentage of additive genetic variance explained by the 20 largest effects SNPs was 0.03% and 1.32% for SNP-BLUP and HEM, respectively. Figure 1C shows that, in comparison to SNP-BLUP, the different shrinkages in HEM allows for relaxing the penalization of moderate and large effects SNPs, therefore

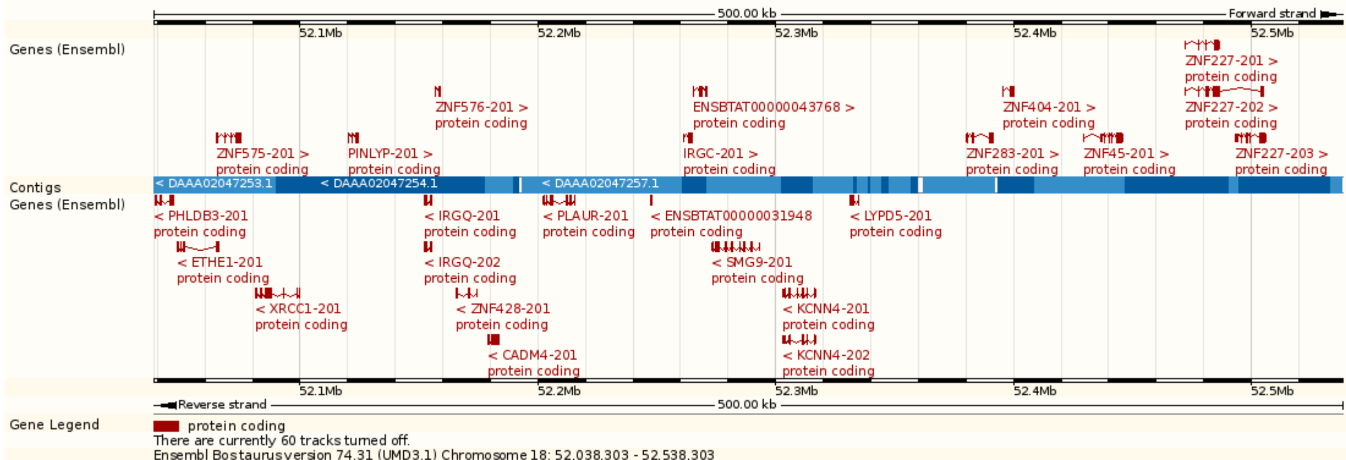

Figure 2. Ensemble region of UMD v3.1 surrounding rs41888601 250kb up and downstream.

improving QTL mapping.

SNP coordinates related with the top20 from HEM model were distributed along chromosomes 1, 2, 5, 7, 8, 9, 13, 14, 18, 19, 20, 22, 23, 25, 26 and 29. Seventy genes were found surrounding these SNPs. However, analyzing the proportion of additive genetic variance explained by markers, 2 of them were investigated in depth (on chromosomes 18 and 19) (Table 1). These two markers explained 0.46% of the additive genetic variance.

Several genes were found surrounding the *rs41888601* on chromosome 18 (Figure 2). Most of them are related with metal ion transporter gene families. One of the most important gene flanking *rs41888601* is localized in an intronic region of *SMG9.*

SMG9 is involved in nonsense-mediated mRNA decay (*NMD*). *NMD* selectively degrades premature termination codon (PTC)-containing mRNAs (Izumi et al. (2012)). The PTC-mRNAs can arise in physiological processes as T and B-cell maturation. According to Trevisol et al. (2013), the luteolysis process causes an increase in the immune cell (including T-cell) levels. Thus the retarded effect on immune cell maturation can affect the luteolysis process, interfering in the oestrus cycle and calving interval.

Associated with *rs137830830* are *SLC39A11* and *SOX9*. *SLC39A11* is either part of the metal ion transporter family (mainly Selenium). The effect of selenium deficiency on reproductive performance is well documented, and selenium is known to have a positive effect on fertility (Mehdi et al. (2013)). The *SOX9* is also well known and it is related with sex-determining and gonad-formation (Harkey et al. (2013)).

As stated earlier, calving interval seems to be highly polygenic. This was also observed by Minozzi et al. (2013) while studying calving interval and other fertility traits in Italian Holstein. In this study we do not find the same regions associated with calving interval.

CONCLUSION

The genome wide association study of calving interval in Gyr cows unravelled interesting genomic regions with genes controlling fertility and oestrus cycle. The genomic regions reported in this study represent interesting positional candidate loci affecting calving interval in Gyr cattle, as they harbor genes known to control fertility, oestrus cycle and gonad differentiation. The lack of evidence supporting chromosome segments previously reported in *Bos taurus* cattle should not be interpreted as absence of common QTLs underlying the trait in both species, as the detection of small effects QTLs remains a challenging caveat.

Literature Cited

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. *Gen. Sel. Evol.*, *41*(1), 55.

Harley, V. R., M. J. Clarkson, and A. Argentaro. 2003. *Endocrine Reviews*, *24*(4), 466–87.

Izumi, N., A. Yamashita, and S. Ohno. 2012. *Nucl*, *3*(1), 29–43.

Kinsella, R. J., A. Kähäri, S. Haider, et al. 2011. *Database : The Journal of Biological Databases and Curation*, *2011*, bar030.

Mehdi, Y., J.-L. Hornick, L. Istasse, et al. 2013. *Mol.*, *18*, 3292–3311.

Minozzi, G., E. L. Nicolazzi, A. Stella, et al. 2013. *PloS One*, *8*(11), e80219.

Sargolzaei, M., J. P. Chesnais, F. S. Schenkel. 2011. *J. Anim. Sci*. 89, E-Suppl. 1 / J. Dairy Sci. 94, E-Suppl. 1: 421 (333).

Shen, X., M. Alam, F. Fikse, et al. 2013. *Gen.*, *193*(4), 1255–68.

Trevisol, E., J. C. P. Ferreira, C. L. Ackermann, et al. 2013. *Rev. Bras. Rep. Anim.* p29-36 (RB384).

VanRaden, P. M. 2008. *J. Dai. Sci.*, *91*(11), 4414–23.