Genome Wide Scan for Age at First Calving in Gyr Dairy Cattle

A. T. H. Utsunomiya,[1,2] **S. A. Boison,[2] D. J. A. dos Santos,[1,2] Y. T. Utsunomiya,[1]**
M. A. Machado,[4] R. S. Verneque,[4] J. Sölkner,[2] J. F. Garcia,[3] R. da Fonseca,[5] M. V. G. B. da Silva[4]
[1]UNESP – Univ Estadual Paulista, Jaboticabal-SP, Brazil, [2]BOKU – University of Natural Resources and Life Sciences, Vienna, Austria, [3]UNESP – Univ Estadual Paulista, Araçatuba-SP, Brazil, [4]Brazilian Agricultural Research Corporation, Juiz de Fora-MG, Brazil, [5]UNESP – Univ Estadual Paulista, Dracena-SP, Brazil

ABSTRACT: Deregressed estimated breeding values for age at first calving of 1,606 Gyr cows were used in a genome-wide scan using 490,009 SNPs. The 20 largest effects, estimated via the Heteroscedastic Effects Model, were used to prospect positional candidate genes, and *CALCRL*, *PRDM8* , *ANTXR2* and *FGF5* were investigated in depth as they flank the two SNPs explaining the largest proportions of additive genetic variance across all markers. Most of these genes have important roles on the development of primordial germ cells and reproductive organs. This study contributes to the identification of loci affecting age at first calving in Gyr cattle, and the positional and functional candidate genes reported represent a starting point to the dissection of the molecular biology underlying the trait.

Keywords: GWAS; age at first calving; *Bos indicus.*

INTRODUCTION

Milk yield is one of the most important traits in dairy cattle. Along with intensive selection for increased milk yield, reproductive performance has declined in many countries, which is partially explained by an unfavorable genetic relationship between these traits (Berlung, 2008). However, reproductive traits, such as age at first calving (AFC), have a huge impact on profitability in the dairy industry. AFC is easy and inexpensive to measure, and it can be used as an indicator of precocity and fertility. As indicine cows (*Bos indicus*) generally take longer to reach puberty than taurines (*Bos taurus*) (Martin et al., 1992), characterizing genomic regions explaining differences in AFC may contribute to the identification of precocious animals. Olsen et al. (2011) also pointed out that the detection of DNA markers associated with improved reproductive performance through genome wide association studies could lead to genetic gain that is more balanced between fertility and production. The aim of this study was to identify candidate genomic regions explaining differences in AFC in Gyr cattle via genome-wide mapping.

MATERIALS AND METHODS

Sample and Phenotype Data

Animals used in this study are part of the national program for genetic evaluation of Gyr dairy cattle in Brazil. Estimated breeding values (EBVs) were available from routine genetic evaluations which comprised 28,133 individuals. EBVs were deregressed following Garrick et al. (2009), and considered as pseudo-phenotypes in the genome-wide mapping analysis.

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 drastic drop in the number of SNPs was mainly due to 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 and 490,009 markers remained.

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 animals' genotype is AA, AB or BB, respectively, and p_i is the B allele frequency at marker i) and

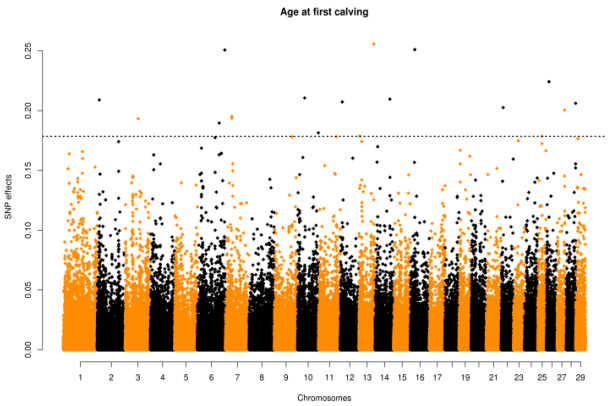

Figure 1. Manhattan plot of the SNP effects estimated from HEM.

$k = 2\sum_{i-1}^{N} p_i\left(1-p_i\right)$. 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

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

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.

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 $2p_i\left(1-p_i\right)\hat{\beta}_1^2$, where p_i is the allele frequency and $\hat{\beta}_i$ is the allele substitution effect estimated by HEM.

RESULTS AND DISCUSSION

The top20 were distributed along chromosomes 2, 3, 6, 7, 10, 13, 14, 16, 22, 25, 26, 27 and 28 (Figure 1). These markers explained 28.48% of the additive genetic variance of AFC together. Forty-three genes were found surrounding these SNPs. Four genes flanking the two SNPs (*rs133983601* and *rs135323951*) explaining the largest proportion of additive genetic variance (5.11%) were investigated in depth (Table 1).

Marker *rs133983601* was found in an intronic region of *CALCRL*. This gene encodes a G protein-coupled receptor related to the calcitonin receptor. Combination of his receptor with the activity-modifying protein transmembrane 2 (*RAMP2*) produces an adrenomedullin receptor (*AM1*). The expression of adrenomedullin receptor inhibits the effects of gonadotropins, affecting the secretion of estradiol and progesterone and consequently the activation and growth of primordial follicles in rats (Li et al., 2008). This phenomenon can be related with the onset of reproductive life in cows and directly affect AFC, since the bovine *CALCRL* is orthologous to the rat.

Table 1. Description of the SNPs *rs133983601* and *rs135323951*

SNP name	Chr	Gene	Description
rs133983601	2	*CALCRL*	Calcitonin receptor-like
rs135323951	6	*ANTXR2*	Anthrax toxin receptor 2
rs135323951	6	*FGF5*	Fibroblast growth factor 5
rs135323951	6	*PRDM8*	PR domain containing 8

Marker *rs135323951* was found in an intronic region of *ANTXR2*, and *PRDM8* and *FGF5* were found 139.42kb and 225.48kb upstream of *rs135323951*, respectively.

ANTXR2 is a positive regulator of metalloproteinases transmembrane activity, a family of proteolytic enzymes that degrade several extracellular matrix components. Reeves et al. (2012) performed a histological analysis of the *ANTXR2*-knockout mouse uterus and cervix and found aberrant deposition of extra cellular matrix in these tissues They concluded that *ANTXR2* expression is required for murine parturition in young pregnant mice and for preserving fertility in aged female mice. In humans, the *ANTXR2* is associated with a rare disorder named Infantile Systemic Hyalinosis (ISH), which may retard intrauterine growth.

FGF5 belongs to the family of fibroblast growth factors (FGF). This family of genes is closely linked to the development of the primordial and primary follicles. Nilsson et al. (2001) found that high levels of FGF in oocytes promote growth of granulosa, stromal and theca cells during early follicular development.

Although *PRDM8* has been identified as a positional candidate in the present study, available functional data is scarce to speculate about its biological relevance in AFC. However, the other genes (*CALCRL*, *ANTXR2* and *FGF5*) have important roles for the development of primordial germ cells as well as reproductive organs.

CONCLUSION

The present genome-wide scan for loci affecting age at first calving in Gyr Dairy cattle identified positional candidate genes with important roles in the development of primordial germ cells, as well as reproductive organs. Combining these results with agnostic hypothesis-driven studies may allow for a better understanding of the relationship between these genes and puberty in indicine cattle.

LITERATURE CITED

Berglund, B. 2008. *Reprod. Domest. Anim.* 43 Suppl 2:89–95.

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

Kinsella, R. J., A. Kähäri, S. Haider, et al. 2011. *Database (Oxford)*. 2011:bar030.

Li, Y., L. Li, I. Hwang, et al. 2008. *Biol. Reprod.* 79:200–208.

Martin, L. C., J.S. Brinks, R. M. Bourdon, et al. 1992. *J. Anim. Sci.* 70:4006–17.

Nilsson, E., J. A. Parrott, and M. K. Skinner. 2001. *Mol. Cell. Endocrinol.* 175:123–130.

Olsen, H. G., B. J. Hayes, M. P. Kent, et al. 2011. *Anim. Genet.* 42:466–74.

Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. *Am. J. Hum. Genet.* 81:559–75.

Reeves, C. V., X. Wang, P. C. Charles-Horvath, et al. 2012. *PLoS One*. 7:e34862.

Sargolzaei, M., J. P. Chesnais, and 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. *Genetics*. 193:1255–68.

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