Proceedings, 10th World Congress of Genetics Applied to Livestock Production.

Genome-wide Association Study of First Service Conception Rate in Brangus Heifers using Probit, Robit and Logit Models

S. Peters[1], K. Kizilkaya[2], D. Garrick[3], R. Fernando[3], J. Reecy[3] I. Imumorin[4] and M. Thomas[5]
[1]Berry College, [2]Adnan Menderes University, Turkey,
[3]Iowa State University, [4]Cornell University, [5]Colorado State University

ABSTRACT: First Service Conception (FSC) is an important trait in cattle reproduction. Genome-wide association study (GWAS) was conducted using BayesC probit, robit and logit model and marker genotypes from the Bovine SNP50. Yearling heifers were estrus synchronized, A.I and pregnancy verified by ultrasound. Model treated SNP effects as random with an assumed fraction pi=0.9995 having no effect on phenotype. Fixed effects were year, location of birth, calving season, age of dam and contemporary group. Results showed there was concordance among the three models for 9 out of the 15 top SNP windows associated with variation in the phenotype. The concordant regions were found on chromosomes 1, 6, 8, 16, 23 and 26 and they explained 10.9% of the variance in FSC. Seventy positional candidates were identified within 1Mb upstream and downstream of the flanking SNP of the associated windows but only 34 were expressed in the pre and post pubertal hypothalamic transcriptome using RNA-Seq.
Keywords: GWAS; bovine; fertility; heifer.

INTRODUCTION

First Service Conception (FSC) is one of the most important complex fertility traits in both beef and dairy cattle with low heritability (Peters et al., 2013). Genetic evaluation for FSC trait for beef cattle has been reported to be challenged by factors such as lack of whole-herd data collection system, low heritability of the trait and the ample time required to collect data for genetic evaluation (Peters et al., 2013). One of the suggested solutions to this problem is the use of information from markers that are in linkage disequilibrium (LD) with genes that contributes to genetic variation of these fertility traits in genome wide association (GWAS) or Genomic Prediction (GP) settings (Kizilkaya et al., 2010; Fortes et al., 2012). Several statistical methodologies have been developed to make use of SNP panels to detect association between single nucleotide polymprphisms (SNP) and traits of interest (Habier et al., 2011) Among the Bayesian methodologies is the probit also known as threshold model commonly used when analyzing binomial or multinomial phenotypes by augmenting the joint posterior density with normally distributed latent variables (Kizilkaya et al., 2003; Peters et al., 2013). Others include the student's t-link (robit) model which provide greater modeling flexibility when compared to probit and then the logit model which has been the preferred model for most statistical applications. We have previously shown the application of Bayesian linear regression model (Bayes C) extended for the probit model for GWAS of FSC and heifer pregnancy (HPG) fertility traits (Peters et al., 2013). The objective of this study was to further extend Bayes C for logit and robit models and then verify the positional candidates underlying the concordant regions in the hypothalamic transcriptome.

MATERIALS AND METHODS

Phenotypes are from Brangus heifers previously described by Peters et al. (2013) and genotypes were successfully obtained from 796 heifers. Genotypes were from a SNP panel including 53695 markers (i.e., Infinium BeadChip, Illumina, San Diego, CA) described by Matukumalli et al. (2009). The SNP loci positions were according to the UMD 3.1 bovine assembly.

Probit, Logit and Robit models for GWAS. In order to provide normal, logistic or Student's-t distributed latent variable in the analysis of ordinal categorical data, the latent variable for animal i can be represented by a scale mixture of normal:

$$l_i \mid \beta, a, R \sim N\left(\mathbf{x}_i'\beta + \sum_{k=1}^{K} z_{ik} a_k \delta_k, R\right)$$

where $R = \sigma_e^2$ is for probit (Albert and Chib, 1993; Sorensen et al., 1995), $R = \sigma_e^2 (2\lambda_i)^2$ with Kolmogororv Smirnov distributed λ_i for logit (Holmes and Held, 2006; Devroye, 1986) and $R = \sigma_e^2 (\lambda_i)^{-1}$ with $\lambda_i \sim Gamma(v/2, v/2)$ for where v is the degrees of freedom of Student's-t distribution for robit models (Kizilkaya et al., 2003), and σ_e^2 is equal to 1 for the models. β is a $p \times 1$ vector of fixed effects that accounted for cohort groups (i.e., animals with the same calving season, ranch location and trait contemporary group) defined as classes, and dam age (in years) as classes, $\mathbf{x}_i'$ is a known incidence vector corresponding to fixed

effects in β, K is the number of SNP loci in the genotype file, z_{ik} is the covariate (0, 1 or 2) at locus k for animal i, a_k is the random substitution effect for locus k, which conditional on σ_a^2 is assumed normally distributed $N(0,\sigma_a^2)$ when δ_k=1 but a_k=0 when δ_k=0, δ_k is a random 0/1 variable indicating the absence with fixed probability π or presence with 1-π of locus k in the model where π=0.9995 (Peters et al., 2013).

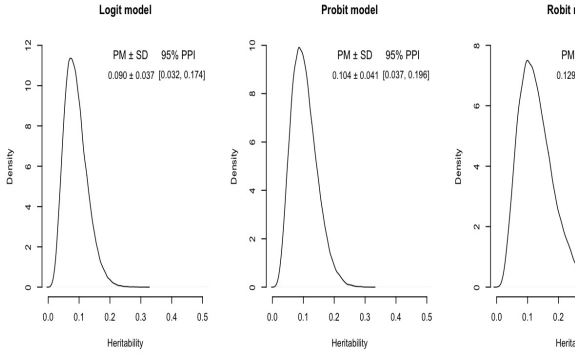
Genetic variance and heritability. Jensen et al. (2012) indicated that the genomic information accounts for between 92% and 98% of the total additive genetic variance depending on the trait in question. In this study, the posterior means of the genetic variance and of the heritability for the latent variable were estimated from the Markov Chain Monte Carlo (MCMC) samples for probit, logit and robit models. The vector of breeding values ($\mathbf{g}$) of all animals in MCMC step (t) was sampled as: $\mathbf{g}^{(t)} = \sum_{k=1}^{K} z_k a_k^{(t)}$, where $a_k^{(t)}$ is the sampled values of a_k in step t. The genetic variance $\sigma_g^{2(t)}$ as $\sigma_g^{2(t)} = \left(\sum_{i=1}^{n}(g_i^{(t)} - \bar{g}^{(t)})^2\right)/n$, where $\bar{g}^{(t)} = \left(\sum_{i=1}^{n} g_i^{(t)}\right)/n$ (Kizilkaya et al., 2013). The heritability was sampled as: $h^{2(t)} = \sigma_g^{2(t)}/(\sigma_g^{2(t)} + \sigma_E^2)$, where $\sigma_E^2 = 1$ for probit, $\sigma_E^2 = 3.29$ for logit and $\sigma_E^2 = v_e^{(t)}/(v_e^{(t)} - 2)$ for robit model, $v_e^{(t)}$ is the sampled values of degrees of freedom in step t. The arithmetic means of these samples for probit, logit or robit model were used to estimate their posterior means.

Hypothalamic transcriptome. RNA was isolated from the entire hypothalamus of both pre and pubertal heifer, cDNA libraries constructed and then sequenced (RNA-seq) to generate gene expression data. The presence or absence of the positional candidate genes alongside with their expression levels were queried in this resource.

RESULTS AND DISCUSSION

The posterior inference on heritability for FSC and posterior probability interval (PPI) for logit, probit and robit models are presented in Figure 1. The posterior mean values presented here are lower than values presented in Peters et al. (2013). This may be due to the fact that values reported in this study are from marker effects while the values reported in Peters et al. (2013) were from additive

Figure 1. Posterior inference on heritability of FSC in Brangus heifers using Logit, Probit and Robit models. PM: Posterior mean, SD: Posterior standard deviation, PPI: 95% posterior probability interval.



relationship from pedigree information. GWAS results are presented in Table 1 and they revealed that there was concordance among the three models for 9 out of the 15 top SNP windows associated with variation in the phenotype. The 9 concordant regions were found on chromosomes 1, 6, 8, 16, 23 and 26 (Table 1, Figure 2) and they explained 10.9% of the variance in FSC. These regions overlaps with previously reported regions associated with fertility by Fortes et al. (2013). Seventy positional candidates were identified within 1Mb upstream and downstream of the flanking SNP of the associated windows but only 34 were expressed in the pre and post pubertal hypothalamic transcriptome. This result underscores the utility of robit and logit models in

Figure 2. Plot of the proportion of window variance accounted for by genome locations for FSC trait in Brangus heifer. Each spot on the plot indicates the proportion of genetic variance contributed by a SNP window defined based on Bos taurus genome assembly (UMD 3.1). The colors represent SNP windows from chromosome 1(left) to 29 and the X and unmapped markers (right).

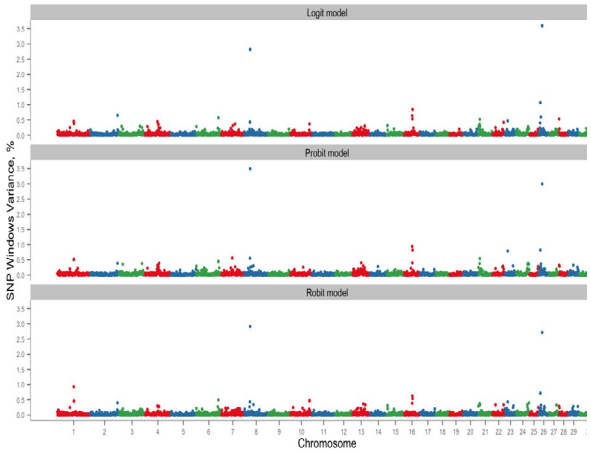


Table 1. Information on the top nine concordant SNP window between Probit, Robit and Logit model in GWAS of FSC trait in Brangus heifer

Map Position	Map position	# SNP in window	% variance	P>0
26_17047506	26_17994478	25	3.5	0.168
8_28983594	8_28027736	20	2.82	0.134
26_8015281	26_8924055	17	1.07	0.056
16_41077367	16_41940324	19	0.85	0.058
16_39033834	16_39967882	17	0.64	0.044
6_108130637	6_108928946	18	0.57	0.042
16_40050652	16_40954414	13	0.53	0.032
23_12069470	23_12978259	20	0.47	0.040
1_80026845	1_80995759	28	0.45	0.046

GWAS. It also shows as expected that multiple loci across the genome are involved in the variation observed in the complex trait of FSC and the positional candidate genes expressed in the hypothalamic transcriptome only explains a fraction of the molecular mechanisms underlying the trait. The hypothalamus controls the reproductive-endocrine axis and therefore plays a big role in fertility. Among the candidate genes in the top concordant regions with differential expression includes *TCTN3, BLNK, ADAMTSL1, PLIN2, RRAGA, SORCS1, BNC2, HAUS6, PSIP1, TTC39B*. Ontology of these genes includes cell signaling and neuron function which may affect reproductive processes.

CONCLUSION

This is believed to be the first study applying robit and logit models to GWAS of FSC and this contributes important data to the dissection of this complex trait in bovine. Fine mapping of the concordant region with whole genome sequence data may shed more light on positional candidate genes, pathways and gene networks underlying this trait.

LITERATURE CITED

Albert J., and Chib, S. (1993). J. Am. Stat. Assoc. 88:669–679.

Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, NY.

Fortes, M. R. S., Lehnert, S. A., Bolormaa S., et al. (2012). Anim. Prod. Sci. 52:143-150.

Fortes, M.R.S., DeAtley, K. L., Lehnert, S.A., et al. (2013). Anim Repro Sci. 141: 1-19.

Habier, D., Fernando, R.L., Kizilkaya, K., et al. (2011). BMC Bioinformatics doi:10.1186/1471-2105-12-186.

Holmes, C., and Held, L. (2006). Bayesian Analysis 1:145–168.

Jensen, J., Su, G., and Madsen, P. (2012). BMC Genetics. doi:10.1186/1471-2156-13-44.

Kizilkaya, K., Carnier, P., Albera, A., et al. (2003). Genet. Sel. Evol. 35:489–512.

Kizilkaya, K., Fernando, R. L., and Garrick, D. J., (2010). J. Anim. Sci. 88:544-551.

Matukumalli, L. K., Schnabel, R. D., Lawley, C. T., et al. (2009). Plant and Animal Genome XVI. San Diego, CA.

Peters, S. O., Kizilkaya, K., Garrick, D. J., et al. (2013). J. Anim. Sci. 91:doi:10.2527/jas.2012-5580.

Sorensen, D. A., Andersen, S., Gianola, D., et al. (1995). Genet. Sel. Evol. 27: 229–249.