Genome-wide association study for milk total unsaturated fatty acids in Brazilian Holstein cows


*University of São Paulo, Department of Animal Science, Piracicaba, Brazil, †Financial support from FAPESP

ABSTRACT: The effects of milk fatty acids in human health have encouraged the research of the genetic control under this trait. In this scenario, the objective in this study was to identify the main single nucleotide polymorphisms (SNP) for total unsaturated fatty acids (UFA) content in a Holstein population reared under tropical conditions. Estimated breeding values (EBV) for UFA content were associated to 6,971 SNP for 757 cows by using Bayesian LASSO methodology. Of these, two SNP in chromosome 14 were responsible for 4.96 and 4.12% of the additive SNP variance, probably because of their proximity with DGAT1 gene. SNP were also identified in chromosomes 2, 11 and 18. The small effects of the majority of SNP confirmed the polygenic nature of this trait. However, the use of genetic markers can improve genetic selection for milk fatty acids.

Keywords: Bayesian Lasso, dairy cattle, milk composition single nucleotide polymorphism

Introduction

Total unsaturated fatty acids (UFA) have been related to several benefits to human health (German and Dillard (2006)). However, in milk, these fatty acids correspond to 30% of lipid fraction (Grummer (1991)) whereas the ideal content would be approximately 70% of the milk fat (Hayes and Khosla (1992)). Moreover, changes in this composition through traditional genetic selection can be difficult, since UFA present low to moderate heritability coefficients, varying between 0.05 and 0.21 (Soyeurt et al. (2007); Stoop et al. (2008)), and negative correlations with important selection criteria in animal breeding programs, such as milk yield and fat percentage, with values ranging from -0.21 to -0.39 (Stoop et al. (2008)), and -0.43 to -0.78, respectively (Soyeurt et al. (2007)).

Nevertheless, the recent identification of polymorphisms linked to genes involved in milk fat biosynthesis represents a helpful tool to accelerate the genetic progress in this trait simultaneously to an enhancement in the understanding of the genetic control under UFA content. Genetic markers within DGAT1 and SCD1 genes (Schennink et al. (2009); Stoop et al. (2009)) and nearby PPARGC1A, AGPAT2, FASN, ACLY and STAT5A genes (Bouwman et al. (2011); (2012)) were associated to genetic variation in fatty acids profile in Holstein cows populations. Also, the use of genomic information for animal selection has increased the reliability of breeding values with a consequent increase in genetic gain.

Despite of these results, there is still a lack of information about marker effects on herds under tropical conditions. Therefore, the aim of this study is to identify major effects single nucleotide polymorphisms (SNP) for unsaturated fatty acids content in a Brazilian Holstein population.

Materials and Methods

Data. 17,165 monthly records of milk unsaturated fatty acids (%) from first through eighth-parity Brazilian Holsteins were used. These records were from 3,404 cows, daughters of 222 sires, with days in milk between zero and 305 (average of 149 days). Unsaturated fatty acids content was measured by Fourier transform infrared spectroscopy and ranged from 0.4 to 1.7%, averaging 1.01%. In addition, 763 of these cows were genotyped using low density panels of SNP (Illumina Bovine LD BeadChip, 6,909 SNP). Samples with call rate lower than 90% (n=4) as well as SNP with minor allele frequency lower than 0.05 (n=36), proportion of missing genotypes higher than 20% (n=2), frequency of heterozygote higher than 0.15 compared with the expected from Hardy-Weinberg equilibrium (n=2), located in Y chromosome (n=9) or monomorphic (n=69) were excluded. At last, 757 individuals and 6,971 SNP remained for the analyses.

Phenotypes: Estimated breeding values (EBV) for the genotyped animals were considered as phenotypes in the association study. The breeding values were predicted by restricted maximum likelihood method under an animal mixed model (REML/BLUP), using VCE and PEST software (Groeneveld et al. (2008); (2009)). The model included the fixed effects of contemporary groups and lactation order, the cubic effect of days in milk as a covariate whereas the additive genetic, the permanent environmental, and the residual effects were considered as random. A total of 7,395 animals from six generations were considered in this analysis (3,404 cows with own records + 3,470 dams + 521 sires).

Association study. The estimation of SNP effects on milk unsaturated fatty acids content was performed by Bayesian Least Absolute Shrinkage and Selection Operator (BLASSO) method, using GS3 software (Legarra et al. (2012)). This method assumes the double exponential distribution as the prior for SNP effects, producing a larger shrinkage of regression coefficients closer to zero simultaneously to a reduced shrinkage in coefficients with larger absolute values (de los Campos et al. (2009)). Also, equivalent to the original LASSO (Tibshirani (1996)), there are two variances, residual ($\sigma^2_e$) and due to SNP ($\sigma^2_s$), and a regularization parameter $\lambda$ (Legarra et al. (2011)). The
initial value for $\sigma_e^2$ ($1.04 \times 10^{-6}$) was calculated as $\sigma_e^2 / \sum_i^{n SNP} 2p_iq_i$ (VanRaden (2008)), where $\sigma_u^2$ is the additive polygenic variance and $p_i$ and $q_i$ are the SNP’s allelic frequencies. The prior distribution for $\sigma_e^2$ was an inverted chi-squared distribution with 4 degrees of freedom whereas for $\lambda$ was adopted an uniform prior between zero and 1,000,000. Initial values for $\sigma_u^2(0.05)$ and $\sigma_e^2(3.1 \times 10^{-7})$ were obtained in pedigree-based REML/BLUP analysis. Markov Chain Monte Carlo (MCMC) was used to estimate the posterior distributions of the model parameters. A burn-in period of 20,000 cycles was run before saving results every 100 cycles out of 200,000.

**Results and Discussion**

The mean of residual variance and additive SNP variance obtained by Bayesian LASSO in 2,000 cycles was $5.7 \times 10^{-3}$ and $2.68 \times 10^{-6}$ $\sigma^2_e$, respectively, whereas the absolute values for SNP effects ranged from $5.79 \times 10^{-8}$ to $9.4 \times 10^{-3}$ %. The 1% SNP with higher effects (n=70) represented 21% of the additive SNP variance. Of these, two SNP in chromosome 14 were responsible for 4.96 and 4.12% of $\sigma^2_e$ (Figure 1).

![Figure 1. Proportion of additive SNP variance explained (%) by single nucleotide polymorphisms in each chromosome (chromosome X = 30) for unsaturated fatty acids content.](image)

Despite chromosome 14, the majority of top 1% SNP were located in chromosomes 2 (n=8), 11 (n=6) and 18 (n=5). Regions in these chromosomes were connected to variation in lauroleic acid (C12:1), palmitoleic acid (C16:1), vaccenic acid (C18:1 trans 11), linoleic acid (C18:2 cis 9,12), conjugated linoleic acid (CLA cis 9 trans 11) content in previous studies (Schennink et al. (2009); Stoop et al. (2009); Bouwman et al. (2011); (2012)). Nevertheless, these SNP together corresponded to 3.7 % of the SNP variance in this study and, therefore, had a small impact on UFA content.

**Table 1. Description of the four single nucleotide polymorphisms (SNP) with major effects on unsaturated fatty acids content.**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position, bp</th>
<th>Var,%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARS-BFGL-NGS-4939</td>
<td>14</td>
<td>1801116</td>
<td>4.96</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-57820</td>
<td>14</td>
<td>1651311</td>
<td>4.12</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-113105</td>
<td>11</td>
<td>76012286</td>
<td>0.38</td>
</tr>
<tr>
<td>Hapmap30383-BTC-005848</td>
<td>14</td>
<td>1489496</td>
<td>0.30</td>
</tr>
</tbody>
</table>

1Based on UMD 3.1 Assembly.  
2Proportion of the additive SNP variance explained.

No important SNP effects, considering this sample size and population, were founded in chromosome 26, in which is located SCD1 gene mutation – a quantitative trait loci for milk fatty acids (Schennink et al. (2007)). This behavior suggests the presence of a genotype-environment interaction and, consequently, highlights the necessity to evaluate SNP effects in several cattle populations under different production conditions.

**Conclusion**

The majority of SNP were associated with a small fraction of SNP additive variance, confirming the polygenic nature of the trait. Major effects were only observed for two SNP nearby a known quantitative trait loci (DGAT1 gene) with relation already reported for other milk production and
composition traits. Due to the high correlation between unsaturated milk fatty acids content and these traits, it was expected a shared genetic control and possible pleiotropic effects associated with these SNP.

On the other hand, important SNP for fat yield and fat percentage it was not identified in this study, suggesting the necessity to evaluate SNP effects under several production scenarios for a better understanding of genetics under these traits and also a good prediction of breeding genetic values.

**Literature Cited**


