Genome Regions Associated to Milk Production Traits and Somatic Cell Score in the Mexican Holstein Population

*A. García-Ruiz**, **F.J. Ruiz*†**, **C.P. Van Tassell‡**, **Hugo H. Montaldo***
*National Autonomous University of Mexico, México, D.F., México, †National Institute of Forestry, Agricultural and Livestock Research, Ajuchitlán, Querétaro, México, ‡Bovine Functional Genomics Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland, USA

ABSTRACT: The aim of this study was to identify the genomic markers associated to milk production traits and somatic cell scores of Mexican Holstein cattle. A total of 1,975, 1,602, and 1,595 animals were included in a genome wide association studies (GWAS) for milk yield, milk components, and somatic cell score, respectively. The genotypes included 45,195 SNP. The analyses were performed with a mixed model and regression analysis implemented in GenABEL (R-software). For all traits, a total of 251 markers had statistically significant association ($p < 0.001$) and 69.72% of the SNP were located in genomic regions earlier reported to traits studied in other populations. BTA13 contains 40 SNP with statistically significant association that have not been reported in other studies. These GWAS results confirm the evidence on the well-known genes and identify a suite of new associations in the Mexican Holstein population.
Keywords: GWAS; Somatic cell scores
Milk production

Introduction

A better understanding of genetic improvement of livestock production, has led to the necessity of detecting genomic regions that explain the genetic variation of economical important traits in livestock animals (Pryce et al. (2010)). Conventional quantitative trait locus (QTL) mapping has been used to detect QTL related to important traits; nevertheless this tool was limited because of the low precision of the QTL position (Mai et al. (2010)). The availability of the genome wide scan through the single nucleotide polymorphism (SNP) markers and the reducing cost of genome sequencing had allowed the detection of numerous genetic regions associated to important traits and the identification of the best animals according to their genomic information (Daetwyler et al. (2010)). The genome wide association study (GWAS) is a recent technique for the identification of causal genes and surveys common genetic variation of traits of interest (Zhang et al. (2012); Wang et al. (2005)). These studies have triggered the opportunity of performing QTL mapping more accurate (Calus (2009); Wang et al. (2005)), including causal genes with modest effects (Hirschhorn et al. (2005)).

In dairy cattle, the majority of GWAS used a bull reference population (Bolormaa et al. (2010); Mai et al. (2010); Pryce et al. (2010)) due to the high reliability of their phenotypes, compared with that of cows. But in some populations the availability of biological samples of bulls for extracting DNA is limited and the inclusion of female's phenotypes is necessary. Previous studies show a greater number of detected regions in the genome when males and females are included in the reference population (Calus et al. (2013)). For farm animals, including dairy cattle, numerous genomic regions are reported frequently explaining variance of economically important traits. GWAS studies have included traits for different breeds within countries, but many inconsistencies have been reported. Differences among the same trait and breed in different countries could be attributed to reference population size, panel densities or population structure (Zhang et al. (2012)). The aim of this study was to identify genomic markers associated to milk production traits and somatic cell scores in the Mexican Holstein population and compare them with those reported in other populations.

Materials and Methods

Phenotypes. Breeding values of Holstein males and females estimated by INIFAP in Mexico (Ruiz et al. (2013)) were deregressed according to the methodology described by Garrick et al. (2009) and were used as phenotypes in GWAS.

For milk production, bulls, sires of bulls, dams of bulls, cows, sires of cows and dams of cows with low reliability (<60, <60, <50, <50, <60, <50, respectively) were excluded from the analysis. Reliabilities of milk components and somatic cell scores were slightly lower than milk production. For fat and protein production (Kilograms and percentage), the limits for reliability was decreased 5 units and 10 units for somatic cell score compared to milk production limits.

A total of 1,975 animals (453 sires and 1,522 cows) were included in the GWAS for milk production, 1,602 (387 sires and 1,215 cows) for milk components and 1,595 (411 sires and 1,184 cows) for somatic cell scores. Table 1 contains general statistics of phenotypes for all traits.

Genotypes. Genomic information of 2,018 Mexican animals and 886 ancestors from North America were used for imputing a total of 45,195 single nucleotide polymorphism (SNP), using the Find Hap software Version 2 (VanRaden et al. (2011)). After the imputation process and deletion of low accuracy phenotypes, the quality control

analysis was performed excluding markers with minor allele frequency less than 2.5%, call rate less than 90%, or failure of Hardy Weinberg equilibrium (P <0.15). Individuals with a call rate across loci less than 90% were also excluded. A total of 34,856; 34,776 and 34,505 SNP markers were used in the GWAS for milk production, milk component traits and somatic cell score, respectively.

Table 1. Mean and standard deviation of phenotypes for Milk production, milk components and somatic cell score included in the analysis.

Trait Breeding Values	Mean	Standard deviation
Milk production (kg)	1178.21	886.88
Fat production (kg)	20.40	27.69
Protein Production (kg)	23.28	24.47
Fat test (%)	-0.107	0.191
Protein test (%)	-0.041	0.100
Somatic cell score	-0.189	0.388

Model. GWAS were performed using a mixed model and regression analysis (GRAMMAR). This method, obtains the residuals adjusted for family effects first (constructing the relationship matrix with SNP information) and subsequently analyzes the association between the residuals and genetic polymorphisms using rapid least-squares methods (Aulchenko et al. (2007)). To account for relatedness, the covariance matrix was estimated from the genomic kinship matrix. This methodology was implemented in the GenABEL package of R (Aulchenko et al. (2007b)). The same package was used for the quality control analysis.

Results and Discussion

To reference the location of SNP with estimated effects statistically different than zero, these were searched by base pair position in the cattle QTL data base (http://www.animalgenome.org/QTLdb/cattle.html, January, 2013) and compared with our results. Previous studies reported QTL, genes or significant SNP in the same positions that our findings and others had not been identified. SNP markers with a p <0.001 were assumed as significant. Genomic regions which included more than 3 significant SNP closely located and were not previous referenced in Holstein cattle will be discussed.

A total of 251 SNP markers had statistically significant association in this study. Twenty eight for milk yield, 30 and 83 for milk fat yield and percentage, respectively, 35 and 44 for milk protein yield and percentage and 31 for somatic cell score. From all the statistically significant association SNP, 70% were previously reported in the same phenotype in this study. A total of 40 SNP with significant association were newly identified on chromosome (BTA) 13 in this study.

Milk yield. Four SNP on BTA2 had statistically significant association (BTB-00088621, BTB-00969752, ARS-BFGL-BAC-30703 and ARS-BFGL-NGS-25349) were related to milk production in our study, and those markers were not previously associated with this trait in Holsteins, but QTL related to milk yield in Ayrshire cattle (Viitala et al. (2003)) were reported in the same genomic region.

Milk fat yield. Four SNP on BTA15 (ARS-BFGL-NGS-117777, ARS-BFGL-NGS-32123, Hapmap52074-ss46527004, BTB-00605195) had statistically significant association for milk fat yield in this study that had not previously found associated with fat yield in dairy cattle.

Milk protein yield. This trait presented the largest number of genomic regions across 12 BTA with low to moderate association level. Nine regions were previously identified. Novel SNP associations were identified in protein yield for Mexican Holsteins. Those associations were located in BTA4 (ARS-BFGL-NGS-102687, ARS-BFGL-NGS-113152), BTA13 (Hapmap42348-BTA-88290, Hapmap51209-BTA-32563, BTA-117294-no-rs, ARS-BFGL-NGS-31560, ARS-BFGL-NGS-38064, ARS-BFGL-NGS-21302, ARS-BFGL-NGS-32055) and BTA15 (Hapmap52074-ss46527004, ARS-BFGL-NGS-31679, ARS-BFGL-NGS-25439). In the same region of BTA13, were detected QTL (Plante et al. (2001)) affecting fat yield, but no associations were identified with milk traits on BTA15.

Milk Fat percentage. The GWAS results for this trait showed fewer genomic regions associated with fat percentage than with milk production, but with moderate to large effect. SNP on BTA9 (ARS-BFGL-NGS-75693 and Hapmap45687-BTA-83338) had not been associated with fat percentage previously, although they had been linked with fat yield (Wiener et al. (2000)), and some SNP on BTA14 (Hapmap30381-BTC-005750, Hapmap30383-BTC-005848, BTA-34956-no-rs and ARS-BFGL-NGS-57820) have not previously been associated with fat percentage, although they are in the neighborhood of DGAT1 (Kaupe et al. (2007)). This result might indicate a wider DGAT1 region or linkage disequilibrium between these markers and that mutation.

Milk protein percentage. SNP with estimated effects that were statistically different than zero for protein percentage in this study were located in 9 different BTA and 5 of these BTA were earlier reported in other Holstein populations. Nineteen SNP on BTA13 (ARS-BFGL-NGS-23363, ARS-BFGL-NGS-65199, ARS-BFGL-NGS-115482, ARS-BFGL-NGS-40188, ARS-BFGL-NGS-13061, UA-IFASA-4272, ARS-BFGL-NGS-41237, ARS-BFGL-NGS-80072, ARS-BFGL-BAC-15734, BTB-01414766, BTA-92697-no-rs, ARS-BFGL-NGS-14463, ARS-BFGL-NGS-75174, ARS-BFGL-NGS-25461, ARS-BFGL-NGS-31462, ARS-BFGL-NGS-116624, BTB-00529466, ARS-BFGL-NGS-104967 and ARS-BFGL-NGS-14592) with large impact on protein percentage in this

study had not been reported previously. SNP affecting milk fat yield with estimated effects that were statistically different than zero in this study have been reported within the genomic region with statistically significant association reported in other studies (Plante et al. (2001)).

Somatic cell score. SNP markers had a low to moderate effect for this trait. Genome regions without previous references were located in BTA13 (Hapmap52431-ss46526792, BTB-01721718, ARS-BFGL-BAC-11276, ARS-BFGL-NGS-39930 and ARS-BFGL-NGS-103355), and BTA26 (ARS-BFGL-NGS-12828, ARS-USMARC-Parent-EF034086-no-rs and ARS-BFGL-NGS-114564). While QTL associated with different traits have been reported in the BTA26 region reported here, no previous references were found for the SNP in the region where the statistically significant association markers of BTA13 are located. For example, studies identified milk yield QTL (Bennewitz et al. (2003)), milk fat yield QTL (Plante et al. (2001)) and milk fat percentage QTL in Ayrshire cattle (Viitala et al. (2003)).

Conclusion

The majority of the SNP with statistically significant association of this study were located in genomic regions earlier related to the studied traits. Others were near reported QTL or genes or were located in regions previously associated with other production traits. Two groups of SNP with estimated effects that were statistically different than zero were located on BTA13 and had moderate to large effects on milk protein yield and percentage. The region includes 7 and 19 SNP for each trait, respectively. These GWAS results identify a suite of novel associations that are not reported previously as well as confirm the evidence on the well-known genes.

Literature Cited

Aulchenko, Y. S., Koning, D. De, and Haley, C. (2007). Genetics, 585: 577–585.

Aulchenko, Y. S., Ripke, S., Isaacs, A., and Duijn, C. M. Van. (2007b). Bioinformatics 23: 1294–1296.

Bennewitz, J.,Reinsch, N., Grohs, C. et al. (2003). Gen. Sel. Evol., 35, 319–338.

Bolormaa S, Pryce J. E., Hayes B.J. et al.. (2010). J. Dairy Sci. , 93:3818–3833.

Calus, M. P. L. (2009). Animal : Int J. Anim Biosci, 4:157–164.

Calus, M. P. L., de Haas, Y., and Veerkamp, R. F. (2013). J. Dairy Sci, 96: 6703–15.

Daetwyler, H.D., R. Pong-Wong, B. Villanueva, et al. (2010). Genetics 185:1021–1031.

Garrick, D. J., Taylor, J. F. and & Fernando, R. L. (2009 Gen. Sel. Evol.,41:55

Hirschhorn, J. N. and Daly M.J. (2006). Nat Rev. 6: 95-108.

Kaupe, B., Brandt, H., Prinzenberg, E. M. et al. (2007). J. Anim. Sci., 85(1), 11–21.

Mai, M. D., Sahana, G., Christiansen, F. B., et al. (2010). J. Anim. Sci., 88: 3522–3528.

Plante, Y., Gibson, J. P., Nadesalingam, J. et al. (2001). J. Dairy Sci, 84(6):1516-24.

Pryce, J. E., Bolormaa, S., Chamberlain, J. et al. (2010). J. Dairy Sci. , 93(7), 3331–45.

Ruiz, F. L., Toledo, H., Duran, A.M., (2013). ¿Qué toro?, INIFAP, 44.

VanRaden, P. M., O'Connell, J. R., Wiggans, G. R. et al. (2011). Gen. Sel. Evol., 43, 10.

Viitala, S. M., Schulman, N. F., de Koning, D. J. et al. (2003). J. Dairy Sci., 86(5):1828-36.

Wang, W. Y., Barratt, B. J., Clayton, D. G. et al. (2005). Nat. Rev. Genetics, 6: 109–18.

Zhang, H., Wang, Z., Wang, S. et al. (2012). J. Anim. Sci. and Biotechnology 3:26