Technical note

QTL analysis associated to single nucleotide polymorphisms (SNP) involved in the dairy phenotype of Holstein cattle

Jose Manuel Valdez-Torres a
Juan Alberto Grado Ahuir a
Beatriz Elena Castro-Valenzuela a
M. Eduviges Burrola-Barraza a*


*Corresponding author: mburrola1@uach.mx

Abstract:

The aim was to identify QTLs associated with single nucleotide polymorphisms (SNPs) whose action contributes to the productive, reproductive and health phenotypic development of Holstein dairy cattle. 341 QTLs located in 120 genes of the Bos taurus UMD_3.1.1 genome and associated with 189 SNPs with effects on productive (FY, NM, MY, MTCAS, MBLF and PL), reproductive (CONCRATE, DPR, EMBSUR, DAYOPEN and CONCEPT) and health traits (SCC, BTBS and RESRATE) were identified. SNPs were verified in the dbSNP-NCBI database, according to which 42% were located in introns. The Jvenn platform revealed that the SNPs rs135744058, rs110828053 and rs109503725 were common in all three traits. The network of correlations between traits and genes generated by MetScape (Cytoscape 3.4), showed a positive correlation between PL, DPR, DAYOPEN, CONCRATE and CONCEPT, and a negative correlation of FY with PL, NM, DPR and CONCEPT. The functionality of each gene was validated in the Gene-NCBI and UniProt databases, and ClueGo (Cytoscape 3.4) was used to select functional pathways with a significance value less than 0.05, which rendered an intertwining between the development of the mammary gland, the activation of the immune system and the response to steroid hormones evident, the GH
gene being the one that directs this functionality. Although the genetic panorama shows that there is an antagonism between productive and reproductive traits, the functional genetic activity due to the 189 SNPs analyzed exhibits an interwoven action in ontological pathways that influence the production processes, as well as in reproductive and health pathways.

**Key words: QTL, SNP, Holstein**

Received: 19/03/2019

Accepted: 17/09/2019

In Holstein cattle, several studies have identified quantitative trait loci (QTL) associated with productive, reproductive and health traits(1-3). Thanks to the improvement of statistical methods and the development of molecular tools, it has been possible to carry out whole genome association studies (GWAS)(4,5) to identify QTLs associated with a locus, whose influence on the phenotype can vary between individuals of the same species through the change of a single base in the genome; this is what is known as single nucleotide polymorphism (SNP). This type of QTL which correlates with a single SNP has an effect on the functionality of a specific gene and, therefore, an immediate action in the development of a phenotypic trait of interest(6). Thus, identification of the biological pathways and genes that are associated with significant SNPs may provide a deeper biological understanding of the expression mechanism of a particular phenotypic trait(7).

The objective of this study was to detect, with the aid of bioinformatics tools, those QTLs associated with a SNP with a potential effect on the phenotypic traits of production, reproduction and health of Holstein dairy cattle. 15 phenotypical characters were used for the QTL search: duration of productive life (PL), net merit (NM), milk yield (MY), milk protein yield (PY), milk fat yield (FY), casein content (MTCAS), β-lactoglobulin content (MBLG), susceptibility to bovine tuberculosis (BTBS), respiratory rate (RESRATE), somatic cell content (SCC), conception rate (CONCRATE), daughters’ pregnancy rate (DPR), early embryo survival (EMBSUR), parturition-conception interval (DAYOPEN) and services by conception (CONCEPT). The selection was made Out of the 114,685 QTL reported in the QTLdb database (Animal QTL Database)(2), those QTL associated with one of the productive and health traits and whose peak was related to a single SNP with a significance value of less than 0.05 were selected, while the QTL selected for the reproductive SNPs were those reported by Cochran et al(8) that did not have a negative effect on any productive trait. Based on the *Bos taurus* UMD_3.1.1 genome, each SNP was verified in the dbSNP database of the NCBI (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/snp/), where its chromosomal location was verified, and the affected gene was classified according to its
location, as: intron (if it was in an intronic region), change of direction (if it caused a change in the amino acid sequence), synonym (if the change of base did not imply a change in the amino acid sequence), promoter (if it was located in the promoter of the affected gene), deletion / insertion (if the SNP resulted in a deletion or a base insertion), or UTR3 'and UTR5' (if the SNP was located in the 5'or 3'untranslated region of mRNA).

The function of each of the genes involved in this study was validated in both the NCBI Gene database (www.ncbi.nlm.nih.gov/gene/) and UniProt (The Universal Protein Resource, www.uniprot.org). Common SNPs among phenotypic traits were identified through a Venn diagram using the Jvenn platform\(^9\). The Pearson correlation coefficients were calculated using the MetScape algorithm\(^10\), in order to establish associations between phenotypic traits with the presence of a gene. These values were visualized as a colorimetric matrix (heat map) composed of a color spectrum that went from green to red with correlation values from -1 to 1, respectively. These data were also analyzed as a hierarchical grouping and, with them, a network of correlations was generated using the MetScape application of the Cytoscape 3.4 software\(^11\). The ontological functional network of the genes was carried out with the ClueGo application of the Cytoscape 3.4 software, under the following criteria: the ontological data of each gene was taken from the “GO Biological Process-GOA; with a "GO Tree" range from 3 to 8; the selection terms for each pathway included at least 1 gene per cluster with a kappa score of 0.3; in addition, it was tested with a bilateral hypergeometric statistical test with Bonferroni correction, only those pathways with a significance value less than 0.05 were taken into account.

The results published to date on GWAS in dairy cattle have provided information on the influence of SNP on the expression of a QTL\(^4,5,12\). In the genome, SNP are the most abundant forms of DNA and due to their low mutation rate; they are excellent selection markers\(^13,14\), in addition to being easy and inexpensive to perform genotyping\(^15\). Taking the information included in the QTLdb database\(^2\), 341 QTLs associated with a SNP were found, of which 70 % were production QTL, 17 % were involved in health and 13 % in reproductive traits (Figure 1). Within the production phenotype, the most favored trait was PY, followed by MY, then FY, NM, PL, MBLG and lastly MTCAS. In health characters, it was BTBS followed by SCC and RESRATE. In reproductive traits, more QTL were associated with CONCRATE, followed by DPR, CONCEPT, and EMBSUR, which was the least favored. There were QTL that had the same SNP, as well as SNPs that were present in different regions of the same gene. Therefore, in the end, 341 QTL associated with 189 SNP located in 120 genes of the Bos taurus genome were identified. Each of the 189 SNP were verified in the NCBI dbSNP database, along with their location and the gene that they affected.
According to the information collected, 42% of the SNP were intronic; 27% SNP were located in coding regions, of which 17% had an effect on the amino acid change of the protein and 5% were synonymous mutations; 6% were localized at the untranslated 3' end (UTR3'), and 1%, in UTR5'. 6.7% of the SNPs were located within the SAA2 gene, 4.1% were located in the ATF3 gene, 3.8% in the HSD17B7 gene, and 3.0% were located in each of the AP3B1 and CARD15 genes. The BCHE and PDE9A genes had 2.6%, while the PCC8, CDKN1A and CSN2 genes had 2% each. 1.2 to 1.8% of the SNP were located in the regions of the APP, GNAS, NRPL48, SERPINA5, SLC8A1, CACNA1D, COQ9, DGAT1, DSC2, FASN, IGF1R, LEP, PRLR, CSNK1E, BRINP3 and HSD17B13 genes (Figure two). With the exception of the Bta-9 and Bta-12 chromosomes, the 189 SNP were distributed in all autosomes and on the X chromosome of the Bos taurus UMD 3.1.1 genome (Figure 3). However, the distribution was uneven; there were chromosomes that exhibited more SNP than others, and in the same way, there were chromosomes that included few SNP, but these were associated with many QTL.
Figure 2: Percentage distribution of genes with a SNP associated with a QTL
**Figure 3:** Distribution of productive, reproductive and health traits in the *Bos taurus* genome

The QTLs most often identified were mainly productive, which supports the fact that the quantity and quality of milk produced go hand in hand with economic gains\(^\text{(16)}\). In Bta-14, the SNP rs109421300 located in the DGAT1 gene has more influence on the MY, FY and PY traits\(^\text{(14,15,17-19)}\). Bta-5 associates with MY, while FY and PY associate with the SNP rs41591907, rs41256890, rs41592943, rs41592948, rs137408198, and rs133449166\(^\text{(7,17,20)}\), located in the BCATI, MGP, GUCY2C, CDEABAR1, and CDEABAR1 genes. Bta-18 has
been associated with both MY and PY, due to the influence of the SNP rs41581694\(^{(14,17)}\), which is located in the FOXA3 gene. In Bta-20, the SNP rs385640152 causes a change in the GHR protein that affects MY and PY\(^{(7)}\). In Bta-21, the SNP rs416444615 located in the SERPINE5 gene is associated with MY and PY\(^{(19)}\). Finally, the SNP rs110475419 is associated with PY\(^{(14)}\), which is in the ADAM12 gene in Bta-26.

Milk is made up of the proteins \(\alpha\)-lactalbumin, \(\beta\)-lactoglobulin, and the \(\alpha\), \(\beta\) and \(\kappa\) caseins\(^{(21)}\). Bta-11 has been associated with \(\beta\)-lactoglobulin (MBLG) concentrations in milk, due to the effect of the \(PAEP\) gene, which has the SNP rs41255679\(^{(16,21,22)}\), rs110066229\(^{(21)}\), and rs110180463\(^{(21)}\). On the other hand, Bta-6 has been related to MTCAS through the SNP rs109299401\(^{(23)}\), which, when expressed in the CSN2 gene, causes a functional change in \(\beta\)-casein. MTCAS has also been associated with Bta-14 and Bta-19 due to the SNPs rs110757796 and rs41923484, respectively\(^{(23)}\). In dairy cattle, the most common disease and the one that causes the most economic losses is mastitis, a disease with which the somatic cell count (SCC) in milk is associated as a predictor\(^{(12)}\). The chromosomes that have the greatest influence on SCC are Bta-6, Bta-13, Bta-14, Bta-19 and Bta-20\(^{(24)}\). There are 6 SNPs in Bta-6, of which rs43703013, rs43703011and rs109299401 are in the coding region of the CSN2 gene, where each one of them causes an amino acid change in the \(\beta\)-casein protein, while rs110239379 and rs110118210 are located in the same intron of the KIT gene, and rs109757609 is located in the promoter of the CSNIS1 gene. The association with SCC in Bta-13 was related to two SNPs in the SRC gene, rs41703851 and rs41602996. Bta-14 contributed to SCC due to SNPs rs109162116 and rs109234250 in the \(DGAT1\) gene. Other SNPs associated with SCC are rs109149276 and rs109149276, of the \(FASN\) gene of Bta-19 and of the promoter of the \(PRLR\) gene in Bta-20, respectively\(^{(24)}\), as well as rs43315150, located in Bta-2, in an intron of the \(CYP27A1\) gene\(^{(20)}\).

Tuberculosis is another disease that causes economic losses in the livestock industry; bovines are not only one of the animal species that most coexists with humans but have also become one of the main sources of spread of this disease worldwide\(^{(25)}\). Richardson et al\(^{(1)}\) found BTBS associated to Bta-1 with 9 SNP (rs42294486, rs29020933, rs29020933, rs42294431, rs42294441, rs110098599, rs132953892, rs41665131 and rs43741780 gene, located in the \(BCHE\) region), and with 2 SNPs (rs109186526 and rs110679397) located in the same intron of the \(KALRN\) gene. In the chromosomes Bta-3, Bta-8, Bta-10 and Bta-23, the same research identified the SNP rs110622046, rs135916795, rs109277058 and rs41642913, located in intronic regions of the MAEL, PTPRD, AP3B1 and \(FKBP5\) genes, respectively. On the other hand, Finlay et al\(^{(26)}\) reported that in Bta-22 the SNP rs42286978, rs42287005 and rs42724727, which are located in the promoter region of the \(SLC6A6\) gene, are also associated with susceptibility to tuberculosis. Wang et al\(^{(27)}\) reported that in Bta-18 the \(CARD15\) gene had 3 SNPs within the coding region of exon 4, as well as one SNP located in the intron prior to exon 4.
As for reproductive traits, these were located very close to productive traits in chromosomes Bta-1, Bta-3, Bta-5, Bta-10, Bta-15, Bta-16, Bta-18, Bta-22, Bta-23 and Bta-24 (Figure 3). Notably, although these SNPs represent only a small portion of the genes involved in the reproductive process, they were selected from the works of Cochran et al(8) and Ortega et al(28), who proved that these SNP have a positive effect on the reproductive traits associated with fertility and, at the same time, they exert no negative influence on productive traits. Of the total SNPs only rs135744058, rs110828053 and rs109503725 were common to productive, health and reproductive traits alike. The SNP rs135744058 is associated with the MY(8), CONCRATE(8,28), DPR(8), and RESRATE characters(29). In the genetic language, it is composed of the A/G variants(8) and is located in the exonic region of the CACNAID gene where it brings about the change in the α1D subunit of the calcium voltage channel(30). In cattle, its presence has been detected at the level of the hypothalamus during the development of the central nervous system(31). The SNP rs110828053 has been reported to be associated with NM(8), DPR(28), CONCRATE(28), CONCEPT(28) and RESTATE(29) and it generates an A/G swap in the HSD17B7 gene that causes a change that substitutes alanine for threonine at position 308 of the protein. The HSD17B7 gene encodes for dehydrogenase 7 hydrosteroid 17-β, which participates in the biosynthesis of sex steroids and cholesterol(28); in bovines, it has been located in the oviduct epithelial cells(32), being essential for ovarian function and regulation of fertility(33). The SNP rs109503725, composed of the variables T/C in the DSC2 gene, is associated with the PL(8), DPR(8,28), CONCRATE(28) and SCC(8) characters. This SNP generates a change in amino acid 535, substituting arginine for tryptophan in the protein Desmochochine 2, which is a protein involved in cell-to-cell junctions, forming desmosomes in epithelial cells(34).

During the last two decades in dairy cattle, selection strategies have focused on developing highly dairy producing animals, which has resulted in a decline in reproductive capacity. This has generated a negative genetic association, i.e. an antagonist, between the productive and reproductive traits(20,35-37). Now, in this study, reproductive SNP were previously reported as non-antagonists with productive traits(8,28); in fact, these SNP are located in the same regions as SNPs associated with productive traits (Figure 3). In order to locate those traits that are positively related, a correlation analysis of phenotypical characters was performed, based on the genes that they shared (Figure 4A). This analysis managed to classify the characters in two groups, grouped into two clades. Clade 1 comprised all the reproductive traits together with the PL and NM characters of the productive traits. The rest of the productive traits (MY, PY, FY, MTCAS and MBLG) were grouped together with the health traits in clade 2 (Figure 4B).
**Figure 4:** Correlation of phenotypic traits based on a SNP associated with a QTL

(A) Pearson's correlation matrix (Heat map), the color spectrum goes from green to red, with values from -1 to 1, respectively.

(B) Hierarchical grouping of A. DPR daughters’ pregnancy rates; CONCRATE = conception rate; PL = duration of reproductive life; NM = Net merit; DAYOPEN = birth-conception interval; CONCEPT = conception inseminations; RESRATE = respiratory rate; EMSUR = early embryo survival; MY = milk yield; PY = protein yield in milk; FY = fat yield in milk; SCC = somatic cell count in milk; MTCAS = casein content in milk; MBLG = content of β-lactoglobulin in milk; BTBS = susceptibility to bovine tuberculosis.

Likewise, when analyzing the correlation network of genes based on the trait, two groups of genetic networks could be observed (Figure 5): Group A, corresponding to those genes present in the characters FY, MY and PY (productive traits), and Group B, with genes present in BTBS, SCC, CONCEPT, CONCRATE, DAYOPEN, DPR, EMSUR, NM, PL, RESRATE, encompassing the health, reproduction and production phenotypes. These results indicate that at the genetic level there is an antagonistic scenario between the production and reproduction QTL.
**Figure 5:** Pearson's correlation network of genes associated with a phenotypic trait

The nodes represent the gene, and the black lines represent the correlations between the nodes. The analysis was carried out with the MetScape application of the Cytoscape software, considering the positive ones from 0.5 to +1. The thicker the line, the higher the correlation. The nodes in yellow, purple, blue, green and pink represent nodes with high correlation generating groups, and the nodes in red represent the genes with the most SNP involved in the panel. The legends within the black ovals represent the phenotypical character associated with the genes in each group of nodes.

In order to encompass the functional information of the genes involved in this study, an ontological network was designed in which those genes whose functions were related to each other were located. Figure 6 shows that, starting from the development of the mammary gland where the genes CSN2, FASN, GH, GPAT4, SRC and TPH1 participate, the GH gene is linked to the pathways that correspond to lipid development, where the DGAT1, GPAT4 and THRSP genes are found. Hence, GH and GPAT4 are linked to cell growth pathways such as the...
development of adipose tissue, in which the APP, ARID5B, FTO, GNAS, GPAT4, LEP, PPARGC1A genes participate; of these, ARID5B is the one that leads to the differentiation of fat cells, while FTO, LEP and PPARGC1A participate in the differentiation into brown fat cells. Likewise, GH, LEP and PPARGC1A are linked, together with CSN1S1, with the steroid hormone response pathways and are related to the regulation of the immune system.

**Figure 6:** Genetic ontology (GO) of the genes that make up the Panel, depicted within a network of biological functions generated with the ClueGo application of the Cytoscape 3.4 software.

Each GO is represented as a circular node, whose size is related to the significance \( P < 0.05 \) and the number of genes associated with a particular biological process; they are grouped by color, each color representing one of these processes. The genes involved in each GO are indicated in bold letters.

The genes involved in the response to steroid hormones are CSN1S1, CYP7B1, GH, HSD17B12, HSD17B7, and LEP; thus, again, GH and CSN1S1 lead to the hormone signaling pathways in which CNOT1, NRH13, SRC, UBR, CTBP2, CYPB7B1, NOD2 and UBR5 participate. GH is also linked to the regulation of the immune system, specifically to the production of the tumor necrosis factor involving the genes APP, BAIAP2, GH, IGF1R, NOD2, SRC, STAT1 and TLR4. Hence, NO2 and TLR4 are the bridge to other pathways of the immune system, such as the activation of cytosines and interleukins and that of the innate
immune system. Interestingly, at this point we find LEP once more, which leads back to the lipid metabolic process. When it is looked at this functional ontological pathway from a global perspective, it seems that GH acts as a master gene that directs, like an orchestra conductor, the functional activity of the other genes involved in the development of dairy cattle. The GH gene is located in Bta-19 and codes for the protein somatotropin. This gene is associated with PY and MTCAS due to the effect of the SNP rs41923484, which causes a change in amino acid 153 by substituting valine for leucine in somatotropin(23). Somatotropin, also known as growth hormone, participates in multiple activities, which range from its effect on cell growth to the differentiation of various tissues, such as the development of follicles(38) and of the mammary gland(39); therefore, it is not surprising that its action is crucial for the entire body of Holstein cattle to function properly. The functional activity of the 120 genes associated with 189 SNPs within 341 QTL is intertwined within the ontological pathways that influence both the production processes, such as the development of the udders, and the reproductive and health pathways. Awareness of this fact opens the door to improving selection in order to generate animals that will gradually acquire both productive and reproductive traits.

**Acknowledgements**

The authors thank CONACYT for the support granted through project No. 216179.

**Literature cited:**


