JOURNAL OF ADVANCED VETERINARY AND ANIMAL RESEARCH
ISSN 2311-7710 (Electronic)
http://doi.org/10.5455/javar.2020.g415
A periodical of the Network for the Veterinarians of Bangladesh (BDvetNET)

June 2020
VOL 7, NO. 2, PAGES 234–241

ORIGINAL ARTICLE

Inspection of real and imputed genotypes reveled 76 SNPs associated to rear udder height in Holstein cattle

Mirvana Gonzalez[1], Rafael Villa[1], Carlos Villa[1], Victor Gonzalez[2], Martin Montano[2], Gerardo Medina[2], Pad Mahadevan[3]

[1]Laboratory of Bioinformatics and Biofotonics, Engineering Institute, Autonomous University of Baja California, Mexicali, B.C, Mexico.
[2]Institute for Research in Veterinary Science, Autonomous University of Baja California, Mexicali, B.C, Mexico.
[3]Department of Biology, University of Tampa, Tampa, FL, USA.

ABSTRACT

Objective: This paper presents the obtained result of a study that realizes to associate a set of real and imputed single nucleotide polymorphisms (SNP) genotypes to the rear udder height in Holstein cows.

Materials and methods: Forty-six Holstein cows from an arid zone of Mexico were phenotyped and genotyped for this study. Blood samples were used for DNA extraction, genotyping was performed with the Illumina BovineLD Bead chip which interrogates 6,912 SNPs genome-wide, and imputation was performed using the Findhap software. After QC filters, a total of 22,251 high quality and informative SNPs were inspected.

Results: The results showed the detection of 76 significant SNPs throughout the complete genome. Significant SNPs fall inside 111 Quantitative Loci Traits related to protein percentage, milk yield, and fat, among others, in chromosomes 1, 2, 3, 5, 6, 9, 10, 11, 12, 13, 19, 20, 21, 23, 26, 27, and 29. Similarly, results confirm that a genotype imputation is a convenient option for genome-wide covering when selecting economic traits with low-density real SNP panels.

Conclusion: This study contributes to establishing a low-cost and profitable strategy for applying genomic selection in developing countries.

ARTICLE HISTORY

Received December 27, 2019
Revised March 24, 2020
Accepted March 30, 2020
Published April 13, 2020

KEYWORDS

Genotypes; QTLs; association; SNPs

Introduction

For centuries, humans have applied a strong selection pressure on cattle favoring health and economically important traits. In response, animals have been shaping their physiological and physical characteristics to adapt. In the last years, a genome-wide association of single nucleotide polymorphisms (SNPs) to genotypes and phenotypes have become a very powerful approach to identify genetic variants associated with the production, conformation, health, and reproduction characteristics [1–4].

Especially, in the northwest of Mexico, production and health animal in arid and semi-arid zones are closely related to several conformation and physiological traits. Previous studies in Holstein cows from arid zones in Mexico have demonstrated that implementing cooling management systems improves physiological status and lactation performance during summer heat [5,6]. Recent studies in Holstein cows from the same arid zones have been focused on associating genes and genetic variations to health and production traits [7,8]. Salomon-Torres et al. [7,9] reported 34 Copy Number Variants overlapping with quantitative loci traits (QTLs) associated with an extensive group of traits, including disease susceptibilities, such as clinical mastitis and dystocia, milk fat, body length, meat color, milk protein, and height.

In a subsequent study where 648,315 SNPs were analyzed, 1,338 genes were published that distinguish cows with ovarian pathology from healthy cows [8].

According to Saowaphak, in dairy cattle rearing, the conformation traits are extremely important since they are related to the high efficiency of the production systems [10].

Correspondence Mirvana Gonzalez ✉ gonzalez.mirvana@uabc.edu.mx Laboratory of Bioinformatics and Biofotonics, Engineering Institute, Autonomous University of Baja California, Mexicali, B.C, Mexico.

How to cite: Gonzalez M, Villa R, Villa C, Gonzalez V, Montano M, Medina G, Mahadevan P. Inspection of real and imputed genotypes revealed 76 SNPs associated to rear udder height in Holstein cattle. J Adv Vet Anim Res 2020; 7(2):234–241.

In this research, we realized a SNP association study to the height of the posterior udder in a Holstein population from the same arid region in northwestern Mexico. In this region, the height of the posterior udder is a good indicator of the cow's ability to produce the milk. This characteristic is important since it can be genotypically and phenotypically related to cow longevity and milk production ability. It should be noted that most of the udder and teats characteristics are hereditary. We sought to identify genes, chromosome segments and QTLs associated with this characteristic. The results show 76 significant SNPs throughout the genome. Besides, an analysis of QTLs covering significant SNPs found 111 QTLs related to milk yield, fat, and protein percentage, among others.

Materials and Methods

Ethical approval

The use committee and animal care of the Autonomous University of Baja California (UABC), considered not necessary to obtain ethical clearance for this research, since blood samples used for the DNA extraction were collected under the directives on animal research of the Institute for Research in Veterinary Science of UABC (IICV-UABC) based on the Mexican laws on animal research (NOM-003-ZOO-1994 and NOM-062-ZOO-1999).

Description of animals and trait

Forty-six Holstein-Friesian cows belonging to the Institute for Research in Veterinary Science of the Autonomous University of Baja California, in Mexicali, Baja California, Mexico were sampled. All cows were clinically healthy and free of tuberculosis and brucellosis. All cows were between the first and third lactation. All cows are registered with the Mexican Association of Holstein.

The conformation trait for an association was the Rear Udder Height (RUH), scored on a 1–9 scale (1–3 very low, 4–6 intermediate, and 7–9 High), established by (www.holstein.ca). Descriptive statistics for the RUH measured on the cattle samples show a mean of 5.04 with a standard deviation of ±0.20, a coefficient of variation of 3.9, and a minimum value of 4 and maximum of 6.

Genotyping, imputation and quality control

The sample consisted of 46 Holstein cows. Genotypes were obtained using the SNPchip BovineLD (Illumina, San Diego, CA), which interrogate 6,912 SNPs along the 29 autosomal chromosomes. The position of SNPs within chromosomes was obtained from the annotation UMD3.1 of the bovine genome [11]. The genotyping data was coded as 0, 1, and 2 for AA, AB, and BB genotypes, respectively. For the 39,611 SNPs genotype imputation, we use the Holstein cattle reference panel from the Bovine SNP50 V2 SNP chip (http://bovinegenome.org/?q=hapmap_funding) and the Findhap software [12]. The imputation reached 46,523 SNPs along the 29 autosomal chromosomes.

Data quality control (QC) involved removing SNPs and individuals not meeting either one of the following criteria: SNPs with a genotyping error >0.05%, frequency of minor alleles <0.05%, individuals with a missing values percentage >2% and Hardy–Weinberg equilibrium test (p-value < 0.001). A total of 22,151 SNPs distributed along the 29 autosomal chromosomes and 46 cow samples passed all QC criteria. Table 1 shows the obtained SNPs distribution after data quality control application.

Statistical analysis

The statistical and association analyses were performed with two different softwares. The R software was used to

Table 1. Basic statistics of SNPs after quality control.

Chr	Size (Mb)	No. of SNPs
1	158.34	1,471
2	137.06	1,099
3	121.43	1,050
4	120.83	1,064
5	121.19	915
6	119.46	1,106
7	112.64	988
8	113.39	960
9	105.71	851
10	104.31	936
11	107.31	995
12	91.16	776
13	82.24	865
14	84.65	838
15	85.30	745
16	81.72	684
17	75.16	684
18	66.00	640
19	64.06	584
20	72.04	640
21	71.60	604
22	61.44	527
23	52.53	485
24	62.71	518
25	42.90	471
26	51.68	348
27	45.41	402
28	46.31	416
29	51.51	489
Total	2,510.08	22,151

apply generalized linear models (GLM) for multiple association. GLM can be expressed as follow:

$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + ... + b_pX_p + e_i$

where Y_i correspond to the dependent variable; X_{1i}, X_{2i},..., X_{pi} are independent explanatory variables; b_0 correspond to the intersection or constant term; b_1, b_2, ..., b_p are parameters measuring the influence of explanatory variables on regressors; p is the number of independent parameters; e is the error observed from the absence of controlled variables; and i (1, 2, ..., n) is the number of observed variables [13]. Besides, PLINK software [14] was used to confirm association results. Manhattan plots were generated using R software (http://www.r-project.org/).

Results

The realized SNPs association analysis allowed the imputation from 6,912 to 46,523 SNPs as a low-cost optimization strategy. The genotyping costs are reduced compared with using high-density SNP panels. We did not evaluate imputation accuracy since we had only low-density sampled data (6,912 SNPs); however, our results indicated good association results. Previously reported results of imputation using the same algorithm showed imputation accuracies ranging from 84% to 99% from low to high densities in dairy cattle populations [15].

After applying association algorithms, genome-wide significance was assessed by defining a p-value threshold after a Bonferroni correction (0.05/number of SNPs analyzed). A set of 76 SNPs resulted associated with RUH trait (Fig. 1). Table 2 summarizes the significant SNPs including chromosome, position, the nearest genes and the –log10(p). Besides, 35 significant SNPs were found to fall inside genes, while the rest fall outside genes but within an average distance of 323 kb around of annotated genes.

To investigate if associated genes interact through a gene network and understand co-expression, we used GeneMania [16] to predict a gene network and interrogate interactions. Figure 2 shows the resulting gene network which reveals a dense co-expression network. The network shows the 69 related genes, with 625 interactions through six pathways.

In another analysis, we looked for the association of significant SNPs to QTLs. We searched in the bovine genome database (http://bovinegenome.org/bovineqtl_v2/findQTL!default.html) and found a total of 111 QTLs related to milk fat composition, milk yield, protein percentage, fat thickness, marbling score, slaughter weight, estimated kidney, pelvic and heart fat, meat tenderness, dressing

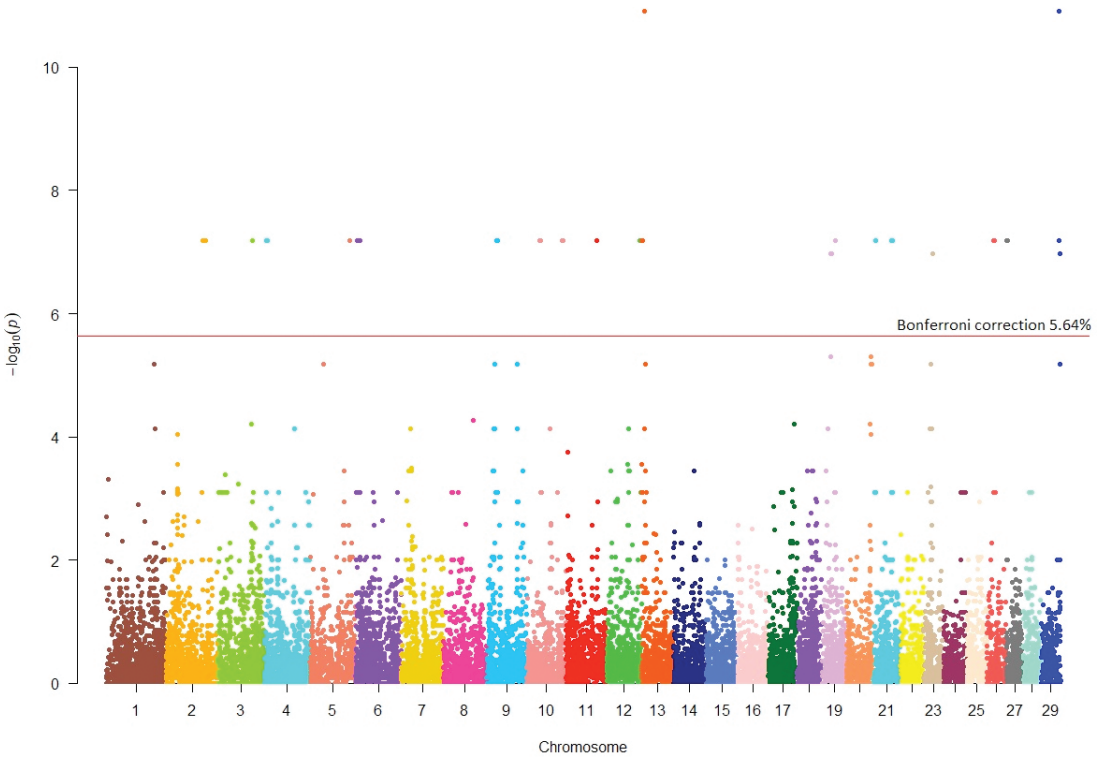

Figure 1. Genome-wide plot $-log_{10}$ (p-values) for associations of SNP with the RUH. The horizontal line represents the Bonferroni-correction significance threshold (p-value = 2.25723E-06).

Table 2. List of significant SNPs for rear udder height trait.

rs–id[1]	SNP name	Chr	Position (bp)	Nearest gene	Distance (bp)	*p*_value Bonferroni
rs42530614	BTB-01405008	1	117,345,968	PLOD2	5,976,499	5.17
rs110305210	ARS-BFGL-NGS-20754	2	97,390,598	LOC616092	110,894	7.18
rs110564068	ARS-BFGL-NGS-99030	2	98,160,191	UNC80	Within	7.18
rs109523794	ARS-BFGL-NGS-102802	2	10,282,5867	VWC2L	98,213	7.18
rs41628800	BTA-27081	2	102,947,494	VWC2L	Within	7.18
rs29019825	BTA-05667-rs29019825	2	103,592,247	ABCA12	Within	7.18
rs43193272	BTB-02093517	3	91,068,865	LOC782072	470,179	7.18
rs43075694	BTB-01966650	4	7,051,825	ABCA13	43,281	7.18
rs42591564	ARS-BFGL-NGS-88802	4	7,567,292	C4H7orf57	9,568	7.18
rs41624504	Hapmap49928-BTA-24568	4	7,624,594	SUN3	Within	7.18
rs41642656	UA-IFASA-8535	5	104,659,347	VWF	Within	7.18
rs42655314	Hapmap27853-BTA-101914	5	105,045,500	ANO2	Within	7.18
rs109812267	ARS-USMARC-686	5	36,917,783	PUS7L	Within	5.17
rs42404150	BTB-01280976	6	4,193,024	TRNAY-AUA	140,681	7.18
rs41595536	Hapmap42010-BTA-93975	6	11,570,944	NDST4	112,866	7.18
rs110314239	ARS-BFGL-NGS-105440	9	26,678,862	NKAIN2	9,401	7.18
rs42930065	BTB-01819195	9	27,965,979	LOC781754	25,049	7.18
rs109907884	ARS-BFGL-NGS-21973	9	27,991,255	LOC781754	230	7.18
rs42970704	BTB-01861211	9	28,018,172	MIR2478	28,108	7.18
rs108958089	ARS-BFGL-NGS-45002	9	28,257,531	LOC101904186	372,315	7.18
rs110755049	ARS-BFGL-NGS-112504	9	28,640,808	CLVS2	Within	7.18
rs41609177	BTA-83186	9	30,175,954	TBC1D32	18,574	7.18
rs43100186	BTB-01990582	9	20,937,396	LOC104972971	553,618	5.17
rs43732219	BTB-01182680	9	81,368,713	HIVEP2	Within	5.17
rs41601192	BTA-67177	10	35,048,321	THBS1	265,704	7.18
rs41588497	Hapmap48845-BTA-67174	10	35,104,530	THBS1	209,495	7.18
rs29013243	Hapmap55209-rs29013243	10	36,131,687	IVD	8,901	7.18
rs110502992	ARS-BFGL-NGS-40292	10	36,465,813	ZFYVE19	Within	7.18
rs108977212	ARS-BFGL-NGS-55657	10	36,835,619	CHP1,EXD1	Within	7.18
rs41655610	Hapmap38437-BTA-81372	10	94,258,071	LOC107132890	75,048	7.18
rs41655605	Hapmap39357-BTA-81356	10	95,629,821	LOC788293	78,526	7.18
rs109081781	ARS-BFGL-NGS-85148	10	97,043,822	MIR2293	413,790	7.18
rs41604491	BTA-107309	11	80,283,508	LOC788214	294,318	7.18
rs41617764	BTA-107661	11	81,484,369	FAM49A	449,694	7.18
rs42359906	BTB-01198500	12	88,162,862	MYO16	Within	7.18
rs109266645	ARS-BFGL-NGS-31217	12	88,245,784	MYO16	Within	7.18
rs41681411	ARS-BFGL-NGS-3885	12	90,101,327	SPACA7	79,043	7.18
rs110016361	ARS-BFGL-NGS-180	13	8,737,253	MACROD2	940,238	10.90
rs109707704	ARS-BFGL-BAC-11281	13	2,217,706	PLCB4	Within	7.18
rs29019327	Hapmap35931-SCAFFOLD200024_14429	13	3,110,731	ANKEF1	31,922	7.18
rs109983818	ARS-BFGL-BAC-15070	13	3,246,468	SNAP25	112,687	7.18
rs110099559	ARS-BFGL-NGS-62490	13	3,978,354	LOC1019001915	45,223	7.18
rs43708448	Hapmap34939-BES1_Contig527_922	13	10,302,310	KIF16B	Within	5.17
rs109928500	ARS-BFGL-NGS-23315	13	10,567,492	SNRPB2	166,828	5.17
rs41918067	BTB-00748932	19	36,264,424	SPAG9	9,354	7.18
rs109225314	ARS-BFGL-NGS-18213	19	24,307,384	RAP1GAP2	Within	6.96
rs110564143	ARS-BFGL-BAC-31839	19	24,812,617	SPATA22	Within	6.96
rs110513971	ARS-BFGL-NGS-80289	19	23,397,071	WDR81	Within	5.29
rs41625312	Hapmap48665-BTA-16217	20	65,220,743	MTRR	207,030	5.29

Continued

rs–id[1]	SNP name	Chr	Position (bp)	Nearest gene	Distance (bp)	*p*_value Bonferroni
rs110728621	ARS-BFGL-NGS-38061	20	65,637,410	ADCY2	Within	5.17
rs108988948	ARS-BFGL-NGS-33801	20	68,065,663	ADAMTS16	Within	5.17
rs41964341	BTB-00798343	20	68,430,871	IRX1	1,041,382	5.17
rs109423241	ARS-BFGL-NGS-44523	21	6,035,370	LOC100301305	66,460	7.18
rs110126411	ARS-BFGL-NGS-34864	21	6,129,342	ASB7	21,119	7.18
rs41606152	BTA-93299	21	48,062,756	MIPOL1	Within	7.18
rs42603510	BTB-01479918	21	48,099,076	MIPOL1	Within	7.18
rs42721237	BTB-01605223	21	48,778,636	SSTR1	19,275	7.18
rs41983805	BTB-00822497	21	49,316,928	LOC101904185	150,666	7.18
rs41983483	ARS-BFGL-NGS-112131	21	49,336,022	LOC101904185	131,572	7.18
rs110665802	ARS-BFGL-NGS-69490	23	23,746,294	PKHD1	49,269	6.96
rs109825181	ARS-BFGL-BAC-5865	23	24,117,682	PKHD1	Within	6.96
rs109579148	ARS-BFGL-NGS-32373	23	20,320,655	LOC100296156	263,263	5.17
rs42084477	BTB-00926954	26	17,163,979	C26H10orf131	Within	7.18
rs109138979	ARS-BFGL-NGS-113660	26	17,246,984	CC2D2B	Within	7.18
rs109442564	ARS-BFGL-NGS-116902	26	18,967,997	CRTAC1	Within	7.18
rs109465094	ARS-BFGL-NGS-25126	26	18,994,785	CRTAC1	Within	7.18
rs109257773	ARS-BFGL-NGS-113339	26	19,237,604	PYROXD2	96,404	7.18
rs41615797	BTA-103632	27	1,250,585	CSMD1	Within	7.18
rs41597984	ARS-BFGL-NGS-117314	27	2,248,130	LOC104970060	Within	7.18
rs109592892	ARS-BFGL-NGS-100576	27	3,076,246	LOC104976024	59,134	7.18
rs42193189	Hapmap24835-BTA-140780	29	48,123,622	PPFIA1	Within	10.90
rs109697888	ARS-BFGL-NGS-111347	29	48,285,322	SHANK2	Within	7.18
rs110507656	ARS-BFGL-NGS-118384	29	48,948,337	DHCR7	Within	6.96
rs109595049	ARS-BFGL-NGS-53937	29	50,202,589	LOC101902793	Within	5.17
rs109039119	ARS-BFGL-NGS-37430	29	50,240,781	LSP1, PRR33	Within	5.17
rs109762888	ARS-BFGL-NGS-3284	29	50,569,383	TSPAN4	Within	5.17

rs-id = reference SNP ID; http://www.ncbi.nlm.nih.gov/projects/SNP/ Chr = Chromosome; bp = Base pair(s), *p* = *p*-value.

percentage, rib fat, birth weight, ovulation rate, yield grade, longissimus muscle area, fat yield, protein yield, twinning rate, hot carcass weight, Back fat, fat percentage, follicle-stimulating hormone concentration, somatic cell score, yearling weight, canonical conformation trait 2, live weight, rump angle, canonical conformation trait 6, dystocia, fecundity, non-return rate of 90, stillbirth, udder cleft, Average Daily Gain on feed, retail product yield, foot angle, pre-weaning average daily gain, productive life, and ribeye muscle area.

The associated QTLs are located within the regions on BTA1, 2, 3, 4, 5, 6, 9, 10, 19, 20, 21, 23, 26, and 29 which have been previously reported by [17,18].

Discussion

As shown in Figure 2, gene network analysis revealed a dense network of co-expression. The network included 69 genes with 625 interactions. The general network shows six internal functional gene networks and several genes presenting at least 10 connections. From the 76 significant SNPs, 35 were found within genes.

The SNP ARS-BFGL-NGS-99030 in chromosome BTA2 which has been previously reported as being associated with fatty acids [18]. It was found within the gene UNC80 which encodes the activator of the NALCN protein. Furthermore, in chromosome BTA2 was also found the SNP BTA-27081 within the gene VWC2L which corresponds to a Von Willebrand factor C domain which contains two proteins [19]. Besides, the SNP BTA-05667-rs29019825 was found within the gene ABCA12 that corresponds to a gene group known as "ATP-binding cassette family" and codes for proteins that transport molecules through the membrane of the cell [19,20]. The SNP Hapmap49928-BTA-2456 was found in chromosome BTA4 within the gene SUN3 [19]. In chromosome BTA5, the SNP UA-IFASA-8535 was found within the VWF gene which is a protein-coding gene von Willebrand factor [21]. The SNP Hapmap27852-BTA-101914 was found within the gene ANO2 which is a protein-coding anoctamin 2 [19], and the SNP ARS-USMARC-686 of was found within the gene PUS7L which is a protein-coding pseudouridylate synthase 7 [22]. All these genes were located within 31 QTLs associated with fat percentage, fat yield, milk production, protein yield, fat

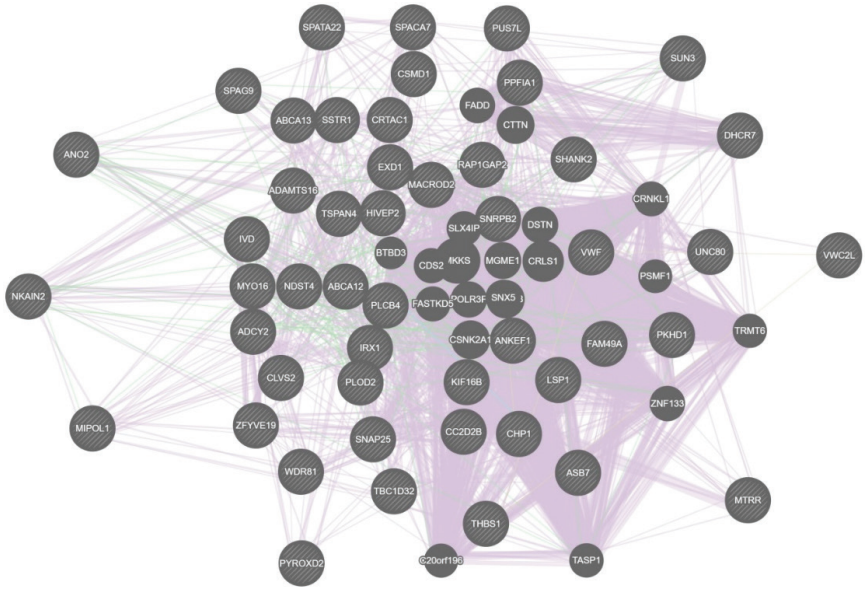

Figure 2. Gene network produced using GeneMania.

Rib, somatic cell score, Backfat EBV, birth weight, dressing percentage, follicle-stimulating hormone concentration, hot carcass weight, Longissimus muscle area, ovulation rate, twinning rate and 1-year weight [23]. In chromosome BTA9, the SNP ARS-BFGL-NGS-112504 was found within the gene CLVS2, which is involved in the regulation of endosomes/lysosomes morphology and associated with defense against infections and recycling of cellular components. Besides, in chromosome BTA9, the SNP BTB-01182680 was found within the gene HIVEP2 that corresponds to a human immunodeficiency virus type I enhancer-binding protein 2 [19]. In chromosome BTA10, the SNP ARS-BFGL-NGS-40292 was found within the gene ZFYVE19 which codes for zinc finger FYVE-type containing protein 19, and the SNP ARS-BFGL-NGS-55657 was found within the gene EXD1 and EXD1 which corresponds to an exonuclease 3′–5′ domain containing 1 and codes for the protein Calcineurin-like EF-hand protein 1, respectively [19].

In chromosome BTA12, the SNPs BTB-01198500 and ARS-BFGL-NGS-31217 were found within the gene MYO16 which codes for the protein myosin XVI. The SNP ARS-BFGL-BAC-11281 was found in chromosome BTA13, within the gene PLCB4 which is sphingomyelin and its metabolic products are now known to have second messenger functions in a variety of cellular signaling pathways. Also, in chromosome BTA13, the SNP Hapmap34939-BES1_Contig527_922 was found the gene KIF16B, supporting the involvement in resistance to bTB [24]. Similarly, in chromosome BTA19, the SNP ARS-BFGL-NGS-18213 was found inside the gene RAP1GAP2 which codes for the protein RAP1 GTPase activating protein 2, the SNP ARS-BFGL-BAC-31839 was found in the gene SPATA22 which is spermatogenesis associated protein 22 [19], and the SNP ARS-BFGL-NGS-80289 was found in the gene WDR81 which is a WD repeat-containing protein [25]. In chromosome BTA20 two significant SNPs were found, the SNP ARS-BFGL-NGS-38061 within the gene ADCY2 which codes for the protein adenylate cyclase 2, and the SNP ARS-BFGL-NGS-33801 within the gene ADAMTS16 which corresponds a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif [19].

In chromosome BTA21 within the gene MIPOL1, SNPs BTA-93299 and BTB-01479918 were found. The MIPOL1 gene corresponds to mirror-image polydactyly 1 protein. In chromosome BTA23 within the gene PKHD1, the SNP ARS-BFGL-BAC-5865 was found and PKHD1 codes for fibrocystin/polyductin protein [19]. Four SNPs were found within chromosome BTA26. The SNP BTB-00926954 was found within the gene C26H10orf131 which codes for chromosome 26 open reading frame protein, human C10orf131 [10], the SNP ARS-BFGL-NGS-113660 was found within the gene CC2D2B which is a coding coiled-coil and C2 domain-containing protein 2B, and the SNPs ARS-BFGL-NGS-116902 and ARS-BFGL-NGS-25126 were found within the gene CRTAC1 which codes for the protein cartilage acidic protein 1 [19,26]. In chromosome BTA27, the SNP BTA-103632 was found within the gene CSMD1 which codes for protein-coding CUB and Sushi multiple domains 1, and the SNP ARS-BFGL-NGS-117314 was found within the gene LOC104970060 which corresponds to a CUB and sushi domain-containing protein 1-like [19].

As shown in Table 2, six SNPs were found to be significant in chromosome BTA29. The SNP Hapmap24835-BTA-140780 was found inside the gene PPFIA1 that codes for protein PTPRF and is an interacting protein alpha 1, the SNP ARS-BFGL-NGS-111347 that was found within the gene SHANK2 which codes for the SH3 and multiple ankyrin repeat domains-containing protein 2, the SNP ARS-BFGL-NGS-118384 which was found within the gene DHCR7 that codes for the protein 7-dehydrocholesterol reductase, the SNP ARS-BFGL-NGS-53937 which was found within the gene LOC101902793 that is a type of ncRNA. Besides, the SNP ARS-BFGL-NGS-37430 was found within the genes LSP1 and PRR33 that are lymphocyte-specific protein 1 and protein-coding proline-rich 33, respectively, and finally, the SNP ARS-BFGL-NGS-3284 was found within the gene TSPAN4 that codes for the protein tetraspanin 4 [19].

Although this study aimed to identify the characteristic of the rear udder height in Holstein cows that are associated with genes, SNP and QTLs using SNPs imputation as a low-cost optimization strategy, an important limitation of this study is the sample size (a small number of animals).

Conclusion

The R, PLINK, and Findhap software were used to analyze the association of a set of real and imputed SNP genotypes to the rear udder height in a group of cows from arid and semi-arid regions in northwestern of México. The imputation of 6,912 to 46,523 SNPs was performed. 76 SNPs throughout the whole genome were found significant (p-value ≤ 2.25723E-06 after a Bonferroni correction). The significant SNPs fall inside 111 QTLs. Several SNPs were found located within genes that are related to cow production (milk yield, fat, and protein percentage), health, reproduction, and conformation traits. The use of GeneMania revealed a dense network of co-expression that includes 69 genes with 625 interactions. Besides, the general network shows six internal functional gene networks and several genes presenting at least 10 connections. Finally, results confirm that a genotype imputation is a convenient option for genome-wide covering when selecting economic traits with low-density real SNP panels. Besides, this study contributes to establishing a low-cost and profitable strategy for applying genomic selection in the developing countries.

Acknowledgment

The authors are grateful to the National Council of Science and Technology of México for supporting a doctoral studies scholarship for Mirvana Elizabeth González-Macías

Conflict of interests

The authors declare that they have no conflict of interest.

Authors' contribution

MG designed the study, involved in the collection of data, interpreted the data, drafted the manuscript, and decided to publish. RV, CV was involved in the collection of data and also contributed to manuscript preparation and the decision to publish. VG, MM, and GM took part in the supervision and decision to publish. PM took part in preparing and critical checking of this manuscript and decision to publish.

References

[1] Wang WYS, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005; 6(2):109–18; https://doi.org/10.1038/nrg1522

[2] Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da Y. A large-scale genome-wide association study in U.S. Holstein cattle. Front Genet 2019; 10:1–21; https://doi.org/10.3389/fgene.2019.00412

[3] Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. BMC Genomics 2011; 12:408; https://doi.org/10.1186/1471-2164-12-408

[4] Lee S, Dang C, Choy Y, Do C, Cho K, Kim J, et al. Comparison of genome-wide association and genomic prediction methods for milk production traits in Korean Holstein cattle. *Asian-Australas J Anim Sci* 2019; 32(7):913–21; https://doi.org/10.5713/ajas.18.0847

[5] Avendaño-Reyes L, Alvarez-Valenzuela FD, Correa-Calderón A, Saucedo-Quintero JS, Robinson PH, Fadel JG. Effect of cooling Holstein cows during the dry period on postpartum performance under heat stress conditions. Livest Sci 2006; 105(1–3):198–206; https://doi.org/10.1016/j.livsci.2006.06.009

[6] Avendaño-Reyes L, Álvarez-Valenzuela FD, Correa-Calderón A, Algándar-Sandoval A, Rodríguez-González E, Pérez-Velázquez R, et al. Comparison of three cooling management systems to reduce heat stress in lactating Holstein cows during hot and dry ambient conditions. Livest Sci 2010; 132(1–3):48–52; https://doi.org/10.1016/j.livsci.2010.04.020

[7] Salomón-Torres R, Villa-Angulo R, Villa-Angulo C. Analysis of copy number variations in Mexican Holstein cattle using axiom genome-wide Bos 1 array. Genom Data 2015; 7:97–100; https://doi.org/10.1016/j.gdata.2015.12.007

[8] Salomón-Torres R, Montaño-Gómez MF, Villa-Angulo R, González-Vizcarra VM, Villa-Angulo C, Medina-Basulto GE, et al. Genome-wide SNP signal intensity scanning revealed genes differentiating cows with ovarian pathologies from healthy cows. Sensors 2017; 17(8):1920; https://doi.org/10.3390/s17081920

[9] Salomón-Torres R, González-Vizcarra VM, Medina-Basulto GE, Montaño-Gómez MF, Mahadevan P, Yaurima-Basaldúa VH, et al. Genome-wide identification of copy number variations in Holstein cattle from Baja California, Mexico, using high-density SNP genotyping arrays. Genet Mol Res 2015; 14(4):11848–59; https://doi.org/10.4238/2015.October.2.18

[10] Saowaphak P, Duangjinda M, Plaengkaeo S, Suwannasing R, Boonkum W. Genetic correlation and genome-wide association study (GWAS) of the length of productive life, days open, and 305-days milk yield in crossbred holstein dairy cattle. Genet Mol Res 2017; 16(2):1–11; https://doi.org/10.4238/gmr16029091

[11] Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol 2009; 10(4):R42; https://doi.org/10.1186/gb-2009-10-4-r42
[12] VanRader P, O'Connell J, Wiggans G, Weigel KA. Genomic evaluations with many more genotypes. Genet Sel Evol 2010; 43(10):1–11; https://doi.org/10.1186/1297-9686-43-10
[13] Rosner B. Fundamentals of biostatistics. In: Brooks/Cole CL (ed.). 7th edition, 2011. Available via http://books.google.com/books?hl=en&lr=&id=-CQtWiJJL0cC&oi=fnd&pg=PR7&dq=Fundamentals+of+Biostatistics&ots=W1K4reTwct&sig=eVAhrOjm1iAJhMLlBJH9B_LvqSw
[14] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81(3):559–75; https://doi.org/10.1086/519795
[15] Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics 2011; 12(1):408; https://doi.org/10.1186/1471-2164-12-408
[16] Mostafavi S, Ray D, Warde-farley D, Grouios C, Morris Q. Open Access GeneMANIA : a real-time multiple association network integration algorithm for predicting gene function. Genome Biol 2008; 9(Suppl 1):1–15; https://doi.org/10.1186/gb-2008-9-s1-s4
[17] Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, et al. Genome wide association studies for milk production traits in Chinese Holstein population. Plos One 2010; 5(10):e13661; https://doi.org/10.1371/journal.pone.0013661
[18] Li C, Sun D, Zhang S, Wang S, Wu X, Zhang Q, et al. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. PLoS One 2014; 9(5):e96186; https://doi.org/10.1371/journal.pone.0096186
[19] Lee Y-S, Shin D. Estimation of the genetic substitution rate of hanwoo and holstein cattle using whole genome sequencing data. Genomics Inform 2018; 16(1):14–20; https://doi.org/10.5808/GI.2018.16.1.14
[20] Woolley SA, Eager KLM, Hafliger IM, Bauer A, Drogemuller C, Leeb T, et al. An ABCA12 missense variant in a Shorthorn calf with ichthyosis fetalis, Anim Genet 2019; 50(6):749–52; https://doi.org/10.1111/age.12856
[21] Berber E. The molecular genetics of von Willebrand disease. Turk J Hematol 2012; 29(4):313–24; https://doi.org/10.5505/tjh.2012.39205
[22] Lee PR, Cohen JE, Iacobas DA, Iacobas S, Douglas Fields R. Gene networks activated by specific patterns of action potentials in dorsal root ganglia neurons. Sci Rep 2017; 7:1–14; https://doi.org/10.1038/srep43765
[23] Liu S, Zhou Z, Lu J, Sun F, Wang S, Liu H, et al. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. BMC Genomics 2011; 12(1):53; https://doi.org/10.1186/1471-2164-12-53
[24] González-Ruiz S, Strillacci MG, Durán-Aguilar M, Cantó-Alarcón GJ, Herrera-Rodríguez SE, Bagnato A, et al. Genome-wide association study in mexican holstein cattle reveals novel quantitative trait loci regions and confirms mapped loci for resistance to bovine tuberculosis. Animals 2019; 9(9):1–15; https://doi.org/10.3390/ani9090636
[25] Traka M, Millen KJ, Collins D, Elbaz B, Kidd GJ, Gomez CM, et al. WDR81 is necessary for Purkinje and photoreceptor cell survival. J Neurosci 2013; 33(16):6834–44; https://doi.org/10.1523/JNEUROSCI.2394-12.2013
[26] Pimentel ECG, Edel C, Emmerling R, Götz K-U. How imputation errors bias genomic predictions. J Dairy Sci 2015; 98(6):4131–8; https://doi.org/10.3168/jds.2014-9170