



## Study of the Genetic Structure and Diversity of Holstein cattle in the small holder system in Mexico



Felipe de Jesús Ruiz-López <sup>a</sup>

José G. Cortés-Hernández <sup>b</sup>

José Luis Romano-Muñoz <sup>a</sup>

Fernando Villaseñor-González <sup>c</sup>

Adriana García-Ruiz <sup>a\*</sup>

<sup>a</sup> Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP). Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal. Ajuchitlán Colón, 76280 Querétaro, México.

<sup>b</sup> Universidad Nacional Autónoma de México. Facultad de Medicina Veterinaria y Zootecnia. Ciudad Universitaria, Ciudad de México. México.

<sup>c</sup> INIFAP. Campo Experimental Centro-Altos de Jalisco. México.

\*Corresponding author: [garcia.adriana@inifap.gob.mx](mailto:garcia.adriana@inifap.gob.mx)

### Abstract:

The objective was to know the population structure of Holstein animals in the family dairy system, to identify possible origins of the genetic material, to know the degree of inbreeding and to identify possible traces of selection in the genome, which allow glimpses of the traits that have been improved over the years. The study included 270 animals genotyped with the GGP-50K® chip. After genotype quality control, 43,548 autosomal SNPs were included. To know the population structure, analyses of mixtures and principal components (PCs) were performed. To know genomic inbreeding and detect traces of selection, information on runs of homozygosity (ROH) was used. Mixture analysis was performed with the Admixture software, and PC, ROH and inbreeding analyses were performed with SVS-v7.6.8. Mixture

analysis showed evidence of six components, all linked to Holstein bulls families with different country of origin. The PCs did not show stratification of the population by herd. The mean inbreeding coefficient was  $0.59 \pm 0.53$  %. In the regions of the genome with ROHs most frequent in the population ( $\geq 20$  animals), numerous associations, QTLs and genes related to milk production and composition, fertility parameters, susceptibility to diseases, body conformation, feed efficiency and some characteristics of carcass composition have been reported. The results reflect the existence of a wide genetic diversity in this population and the possibility of carrying out genetic improvement work through selection without affecting inbreeding levels.

**Keywords:** Genetic diversity, Traces of selection, Inbreeding, Family dairy system.

Received: 17/12/2022

Accepted: 09/05/2023

## Introduction

The cattle dairy industry in Mexico produced around 11.489 billion liters of milk nationwide in 2020 (SIAP, 2020)<sup>(1)</sup>, of which more than 30 % of the volume was produced in the family dairy system (FDS), which includes approximately 78 % of the farms<sup>(2)</sup>. In the FDS farms, Holstein animals predominate, although Brown Swiss animals and their crosses can be found<sup>(3)</sup>. Currently, in this system there is little information on production records by animal and on rare occasions genealogical information can be collected, which makes it unfeasible to carry out genetic evaluations of the animals in this system. The genetic improvement of these animals has been carried out by the selection performed by the cattle farmer within their herd, or by the introduction of genetic material, but there is no evidence of directed mating for a specific genetic purpose.

The use of genomic information has made it possible to describe the structure of populations that do not have genealogical information or records. The study of these populations or animals has been carried out from the study of single nucleotide polymorphism (SNP) markers or their clustering pattern, such as for example, runs of homozygosity (ROHs), which are homozygous segments in the genome, identical by descent, which can be used to study population structure, demographic history and to decipher the genetic structure of complex diseases<sup>(4)</sup>. ROHs are the result of crossbreeding between related individuals<sup>(5)</sup> from populations with a high level of selection intensity, influenced by the availability of

replacements and the adoption of technological and reproductive tools<sup>(6)</sup>, or low rates of recombination<sup>(7)</sup>. Their distribution and length depend on the intensity of selection, being more frequent and more extensive when it is higher<sup>(8)</sup>, when mating between close relatives is frequent or when the size of populations is small<sup>(4)</sup>.

The potential of ROHs to help the genetic improvement of production animals is big due to the fact that they contain a large number of genes that encode traits of interest<sup>(9)</sup>. In addition, the identification of ROHs can help to visualize and recognize haplotype patterns characteristic of breeds or species<sup>(7)</sup>, allowing the identification of genomic regions with possible traces of selection for the breed<sup>(10)</sup> and the calculation of individual inbreeding levels. The latter is done by evaluating the portion of the genome covered by ROH segments, especially since there is a high probability of detecting genomic information from ancient relationships<sup>(11)</sup>. This is a useful tool for populations that do not have genealogical information<sup>(12)</sup>.

Traces of selection are regions of the genome that have been conserved for generations in populations due to natural or artificial selection. These sequences of genetic material are related to functionally important traits<sup>(13)</sup> and their detection helps to identify candidate genes that have been favored in the selection processes to which populations have been exposed, and to identify beneficial mutations. In addition, they help to understand the molecular pathways related to phenotypic traits<sup>(14,15)</sup>.

With SNP marker analyses, it is also possible to know the population structure through mixture analysis and to know the most influential origins in a population. In addition, through information-reductive methods, such as principal component analyses, it is possible to determine patterns of population structure, important information for establishing the basis for a genetic improvement program.

The objective of this study was to know the population structure, identify possible origins of the genetic material, know the degree of inbreeding, and identify possible traces of selection in the genome that allow us to glimpse the characteristics that have been improved over the years by the decisions of cattle farmers in family production systems in Mexico.

## **Material and methods**

A total of 270 Holstein cow genotypes were used, randomly chosen from the population present in three FDS herds located in the region of Tepatitlán, Jalisco, Mexico. The animals were genotyped with the GeneSeek Genomic Profiler Bovine GGP 50K® chip. Quality

control of genomic information consisted of excluding animals with a call rate  $<0.90$ , excluding SNPs with minor allele frequency (MAF)  $<0.02$ , or with a call rate  $<0.95$ , or with a Hardy Weinberg  $P$ -value  $<0.0001$ <sup>(16,17)</sup>. After quality control, 43,548 autosomal SNPs were included.

In order to know the structure and main population origins, a mixture analysis was carried out through estimation based on likelihood models that define the structure of ancestry in unrelated individuals, a methodology implemented in the Admixture V 1.3.0<sup>(18)</sup> software. On the other hand, Principal Component (PC) analyses were performed to identify possible population groupings by herd. To estimate inbreeding with genomic information and traces of selection in the population, ROHs were searched in the genome. To define ROHs, runs with a minimum length of 500 kb and a minimum number of 25 SNPs, with a minimum density of 1 marker every 50 kb and a maximum gap between contiguous homozygous markers of 500 Kb were included. With the aforementioned parameters, the risk of including very short ROHs was avoided, a common case due to linkage disequilibrium (LD)<sup>(17)</sup>. LD is associated with the presence of linked genes, so when they are inherited from parents to children, they do so jointly, affecting the frequency of recombination (less than 50 %) and the presence of ROHs in the genome<sup>(19)</sup>. In addition, 1 heterozygous SNP and 5 genotypes lost per run were allowed<sup>(20,21)</sup>.

ROH analysis was performed using the bioinformatics platform called SNP & Variation Suite v7.6.8 Win64 (Golden Helix, Bozeman, MT, USA)<sup>(22)</sup>, while the analyses of the data obtained were carried out with SAS Institute 9.3.<sup>(23)</sup> To analyze the distribution of  $L_{ROH}$ , six classes were defined according to their length, which were 0.5 to 4,  $>4$  to 8,  $>8$  to 12,  $>12$  to 16,  $>16$  to 20, and  $>20$  Mb<sup>(24)</sup>.

Traces of selection were detected through the loss of genetic variation, using the ROHs identified in the genome, using the most frequent in the population (in at least 10 % of the animals). According to their physical position in the genome, annotations, or regions previously related to genes, QTLs or traits that have been made in other populations and that are reported in the Animal QTLdb Release 43 database<sup>(25)</sup> were identified. In addition, traces of selection previously reported in the Bovine Genome Variation database (BGVD) were searched in order to find genomic information that helps to know possible characteristics that have been selected in the population of the Family Dairy System (FDS) in Mexico.

For the calculation of the coefficient of inbreeding by runs (FROH), the methodology proposed by Mcquillan *et al*<sup>(27)</sup> was used, who defined it as  $F_{ROH}^i = \Sigma L_{ROH}^i / L_{auto}$ , where  $F_{ROH}^i$  is the endogamy coefficient of individual  $i$  calculated by ROH,  $\Sigma L_{ROH}^i$  is the total sum of the ROH segments of an individual  $i$  above a specified minimum length, in this case  $>500$  kb, and  $L_{auto}$  is the length of the autosomal genome covered by SNPs including centromeres.

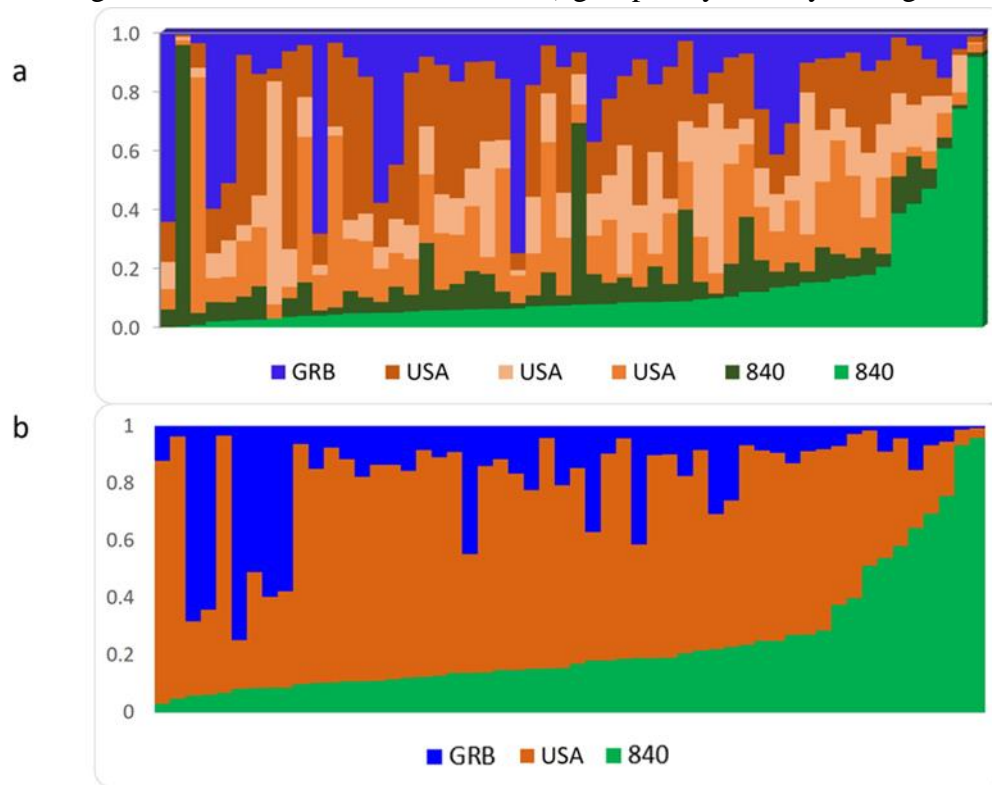
As the year of birth of the animals is unknown, the inbreeding trend by lactation number of the animals at the time of sampling was calculated.

Additionally, the calculation of the inbreeding coefficient was carried out through observed and expected homozygous markers (FHOE) for all animals, which has been reported to have a correlation of  $0.96 \pm 0.001$  with the genomic relationship matrix in Holstein cattle<sup>(28)</sup>, the values of FHOE can range from  $-1$  to  $+1$ . The negative numbers refer to the exogamy present in mating between individuals from different populations and the positive values indicate the level of endogamy of individuals from the same population; the calculation was performed through the method referenced by Ferenčaković *et al*<sup>(11)</sup> with the program called SNP & Variation Suite v7.6.8 Win64 (Golden Helix, Bozeman, MT, USA)<sup>(22)</sup>.

## Results and discussion

In the analysis of mixtures, the value that best defined the number of ancestral populations (K) was six and according to the information collected from some cow parents, six large families were identified, defined mainly by the country of origin of the bulls. Figure 1-a shows the population structure linked to the six main groups by country of origin of the bulls, although some of these families share the same country of origin. Therefore, the groups that shared the same origin were combined, leaving only three large groups represented, two representing the United States of America and one representing the United Kingdom (USA, 840 and GBR, respectively) (Figure 1-b). The origins USA and 840 correspond to the United States of America, only that the 840 is assigned to animals that use radio frequency identifications (RFID), devices issued by the International Committee for Animal Recording (ICAR), while animals registered with the USA country of origin do not carry an RFID and the use of genetic material is locally or more limited than those of the 840<sup>(29)</sup>. The results of this study show the genetic dependence of the family system in Mexico on foreign material, mainly from the United States, since more than 80 % of the origins were linked to families with origins from this country, either from local trade or from those registered internationally.

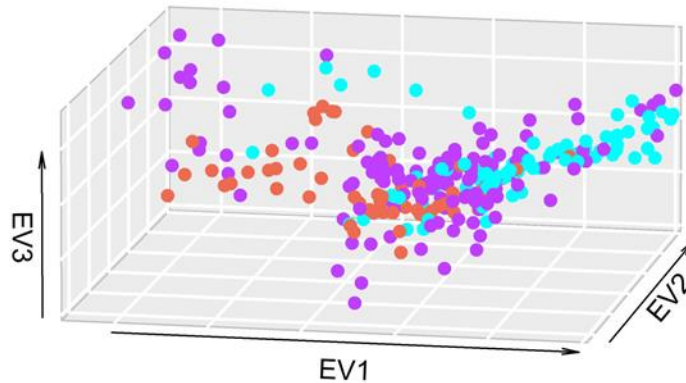
**Figure 1:** Population structure of Holstein cattle in the family system a) including the six origins attributed to bull families and b) grouped by country of origin



Although another study<sup>(29)</sup> had reported the influence of other breeds of dairy cattle on the family system, the present study found no evidence of the use or crossbreeding with other breeds. These results could suggest that cattle farmers have followed a more directed mating system, and that they limit themselves to using animals of the same breed in services.

In the PC analyses (Figure 2), no stratification was found by country of origin of the bull, and when the herd of origin was evaluated, a homogeneous herd (purple) was observed in the population, and a difference was observed between the animals of the other herds (red and blue). The percentage of variability associated with each component was 2.9, 2.0, and 1.8 for components or eigenvalues (EV) 1, 2, and 3, respectively.

**Figure 2:** Principal component analysis of the family system population with Holstein phenotype, defined by herd of origin



The total number of ROHs ( $N_{ROH}$ ) found in the studied population was 15,695, with an average length ( $L_{ROH}$ ) of 4.79 Mb, a minimum and maximum length of 0.5 and 91.49 Mb, respectively. The average length of the genome covered by ROH was 278.76 Mb, with a minimum and maximum of 13.28 and 535.83 Mb, respectively. According to the frequency of ROHs in the population, they were identified as unique (in a single animal) or repeated, the latter with the same length (identical) or of variable length. Thirty-five point eight six (35.86) percent of the ROHs were unique (Table 1), while 64.14 % (10,067) were repeated.

**Table 1:** Number and percentage of unique runs of homozygosity (ROHs) and repeated ROHs with the same start and end position, as well as variable positions

Type of ROH	Length	Number of ROHs	Percentage
Unique	-	5,628	35.86
	Identical	5,663	36.08
Repeated	Variable	4,404	28.06

The  $N_{ROH}$  was lower compared to that reported in animals that come from specialized production systems, which could be attributed to a lower intensity of selection since the loss of genetic variation or the formation of ROHs in the genome is influenced, among other factors, by the level of intensity of selection in the populations, which in turn is determined

by the availability of replacements and adoption of technological and reproductive tools, such as artificial insemination (AI) and embryo transfer (ET). In dairy cattle, the intensity of selection is very high and the selection of genetic material is influenced by a limited number of parent families, so the mating of related individuals may be common<sup>(30)</sup>. In Holstein cattle from the specialized production system in Mexico, the  $N_{ROH}$  was 88,529, with a larger population size (~4,500 animals) and  $L_{ROH}$  was greater than 8.95 Mb<sup>(24)</sup>. In other studies in Holstein cattle from the specialized system, the  $L_{ROH}$  reported are even higher; for example, in the US it is 299.6 Mb<sup>(31)</sup> and in Italy it is 297 Mb<sup>(32)</sup>.

The average number of ROHs per animal was  $58.13 \pm 11.89$ , with a maximum and minimum of 92 and 10, which is a high value compared to the results of Holstein cattle from the specialized system in Mexico<sup>(24)</sup>, reported on average at 20.07 ROHs per animal, with a maximum of 283 and a minimum of 1. Studies in other Holstein populations of intensive production have reported around  $82.3 \pm 9.83$  ROHs per animal in Holstein cattle from the US<sup>(31)</sup> and  $81.7 \pm 9.7$  runs per animal in Holstein cattle from Italy<sup>(32)</sup>. These differences may be due to the high degree of selection in the populations of specialized production systems in both the US and Italy, as well as the availability of genetic material from highly selected bulls compared to the FDS that was analyzed, where the selection objectives are not so defined.

According to the classification of  $L_{ROH}$ , the most frequent runs were the shortest (0.5 to 4 Mb) with 64.42 %, followed by those from 4 to 8 Mb with 20.45 %, and the least frequent were the longest runs (>16 to 20 and >20; Table 2). The lengths of the ROHs provide information on the number of generations in which the common ancestor is shared, with the longest being those formed in recent generations<sup>(21)</sup>, so the length of the ROHs found in this population reflects recent and low inbreeding.

**Table 2:** Frequency of runs of homozygosity (ROH) at different lengths

Length (Mb)	Number	Percentage
0.5 to 4	10,110	64.42
>4 to 8	3,209	20.45
>8 to 12	1,131	7.21
>12 to 16	539	3.43
>16 to 20	296	1.89
>20	410	2.61

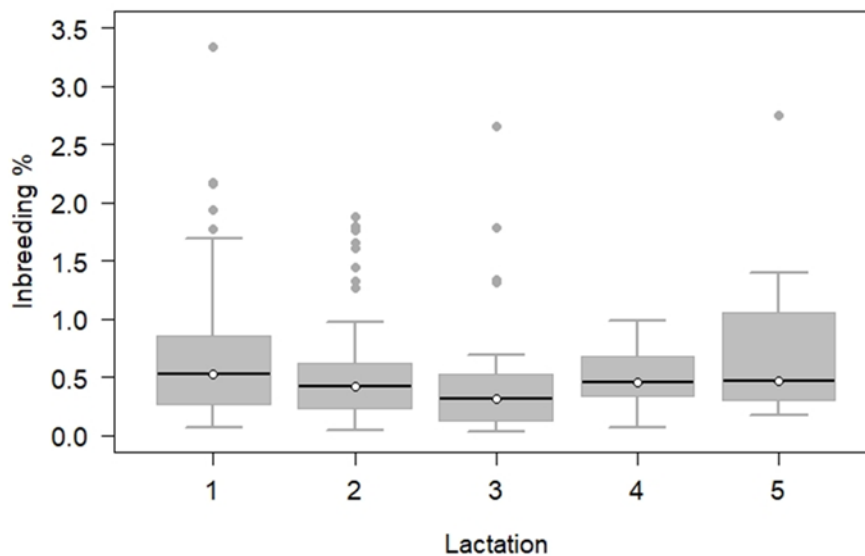
Mb= Megabases.

The average inbreeding coefficient (FROH) in the population was  $0.59 \pm 0.53$  %, with a maximum of 3.35 % and a minimum of 0.034 %. The results are consistent with the small

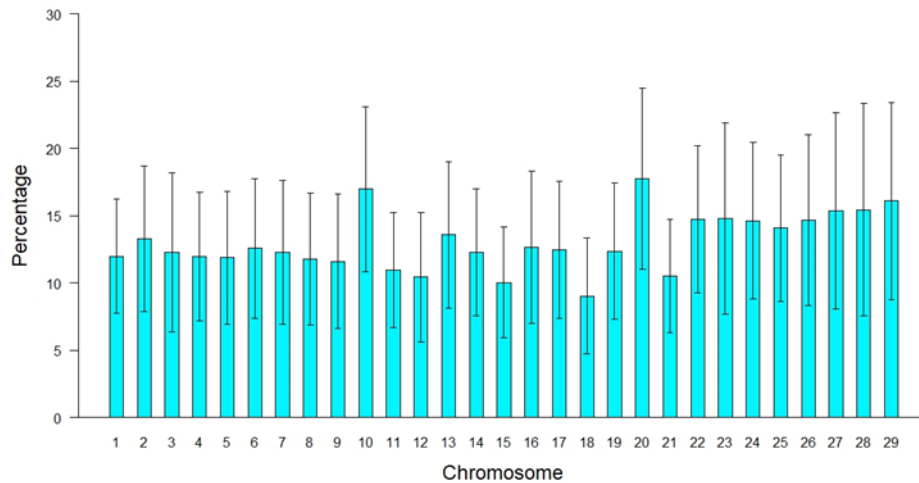


number of ROHs found and the short average length. These values were well below what was reported in other highly selected Holstein cattle populations; for example, 4.2 % in the US<sup>(33)</sup>. Although the inbreeding found in this population was insignificant, a value that is confirmed by the values calculated for FHOE, which were  $-0.02 \pm 0.08$ , when reviewing the averages by number of births, a slight increase in FROH was found in recent generations, which could indicate the beginning of an unfavorable trend for the group studied (Figure 3).

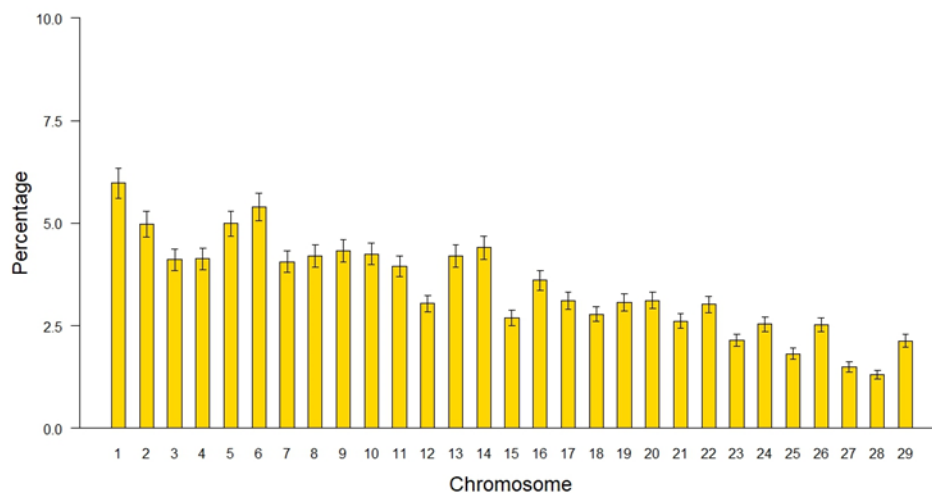
**Figure 3:** Genomic inbreeding (FROH) percentages by lactation number at the time of sampling



To identify traces of selection throughout the genome, it was searched for specific chromosomes and regions in the location and distribution of ROHs. The presence of ROHs occurred to a greater extent in the long chromosomes than in the short ones, although the latter presented a higher proportion of the genome covered by homozygous regions, as was the case of chromosomes 10 and 20, which presented 10 and 20 with 16.98 % and 17.76 % (Figure 4), behavior similar to what was reported by Szmatoła *et al*<sup>(34)</sup> in Holstein cows from Poland, suggesting that these regions have been subject to greater selection, due to association with traits of economic interest. The percentages of homozygosity per chromosome are higher than the average FROH value because the length of the chromosome is taken as one hundred percent and not the total length of the genome covered by the SNPs; this gives a better perception of the length of the chromosome covered by ROHs.

**Figure 4:** Percentage of chromosome covered by runs of homozygosity (ROH)

Throughout the genome, a positive relationship was observed between chromosome size and the number of ROHs detected on that chromosome, but this was not the case for the percentage of chromosome length covered by ROHs since short chromosomes showed a higher proportion covered by ROHs (Figure 5), this is because the average length of ROHs was greater in short chromosomes than in long chromosomes, because long chromosomes have more recombination than short chromosomes<sup>(8)</sup>. Of the total ROHs determined in the population, chromosomes 1 and 6 were the ones with the highest number of ROHs (5.98 and 5.39 %) and the chromosomes with the lowest number were chromosomes 28 and 27 (1.31 and 1.50 %), results that are similar to those of Purfield *et al*<sup>(17)</sup>, who also reported a higher number of ROHs on the long chromosomes than on the short ones.

**Figure 5:** Percentage of runs of homozygosity (ROH) on each chromosome

According to the frequency of repeated ROHs in the population, only 35 were found in 10 or more animals and were distributed on chromosomes 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 21, 22, 23, 26, and 29. The most frequent ROHs were found on chromosomes 2 and 22, in 27 and 23 animals, respectively, with the length in these runs being 1.82 Mb and 1.61 Mb. In the same position as the runs found on chromosome 2 (83.84-85.66 Mb), Cole *et al*<sup>(35)</sup> reported QTL (Quantitative Trait Loci) related to hip width and height in Holstein cattle from the US; Cai *et al*<sup>(36)</sup> reported QTLs associated with milk fat production in Holstein cattle from Nordic countries, which may indicate traces of selection on these chromosomes<sup>(37)</sup>.

Of the total ROHs of variable length, only 37 were found in 10 or more animals, distributed throughout the genome, except on chromosome 8. In the region where the most frequent ROHs in the population were found ( $\geq 20$  animals), numerous associations, QTLs and genes related to milk production and composition, fertility parameters, susceptibility to diseases, body conformation, feed efficiency and some characteristics of carcass composition (Table 3) have been reported. The results show that, although the ROH lengths in this population ( $\sim 4.79$  Mb) suggest crossbreeding of animals related approximately 16 generations ago<sup>(11)</sup>, the conserved regions could indicate that selection in this population is aimed at improving milk production, composition, fertility, and health, as could be expected in milk production systems. On chromosomes 1 and 2, in addition to associations with characteristics of interest in dairy cattle, associations with carcass characteristics are observed, findings that could suggest possible crosses with other breeds.

To search for ROH annotations with variable length, the shortest ROH with respect to the final position was used as a reference (Table 3), to avoid providing information outside the region common to all animals with a specific ROH.

**Table 3:** Genome annotations found in regions where the most frequent runs of homozygosity (ROH) in the family dairy population were detected

BTA	Start position	Length (pb)	NoA	Associations /QTL reported	Genes NCBI	Traces of selection
13	389,736	1,593,318	38	CALEASE, PTAT, FY, MY, NM, PY, UHT, SB, STA, FANG, FTLEG, UA, RLEGR, RLEGS, RTPL, SCS, MRCT, FSC, CONCEPT, MBCASP, MPFRAT, DYF, DYST, TPL, TLGTH, UDPTH, PP, FP, HTINT.	Associated: 287026.	TMX4, PLCB1, MIR2285M-1.
1	761,316	1,116,706	32	PP, MKCASP, CONCRATE, MUGKCASP, BTBS, SCS, FATTH, PY, RFI.	Associated: 506426. Candidate: 282257.	ATP50, ITSN1, CRYZL1, DONSON, SON, GART, DNAJC28, TMEM50B, IFNGR2, IFNAR1, LOC104970778, IL10RB, IFNAR2, LOC526226, OLIG1, OLIG2.
17	67,686	1,209,677	32	FY, PY, CONCRATE, CONCEPT.		TMEM192, KLHL2, MSMO1, CPE, LOC101903170.
2	83,841,602	1,794,112	31	FSC, NRR, CONCRATE, CONCEPT, MUGKCASP, MSPD, BTBS, FY, BD, CALEASE, PTAT, FTPL, UA, NM, PL, RTPL, UHT, RUMWD, SCS, SB, STA, STR, UC, UDPTH, LMY, EY, BW, BV DV.	Candidate: 526800 Candidate: 19122 Gene: 521004.	SLC39A10, DNAH7, STK17B, LOC531691.
7	153,780	811,014	31	CALEASE, SB, FANG, FTLEG, PTAT, FTPL, UA, NM, PL, RLEGR, RLEGS, UHT, SCS, STA, UC, UDPTH, MBCASP.	Gene: 338031.	LOC107131408, LOC100125913, LOC101902704, FLT4, CNOT6, GFPT2, MAPK9, RASGEF1C.

BTA= chromosome, NoA= number of animals.

**Associations /QTLs reported.** CALEASE= calving ease, PTAT= conformation final score, FY= milk fat yield, MY= milk yield, NM= net merit, PY= milk protein yield, UHT=udder attachment height,

SB=stillbirths, STA= stature, FANG= foot angle, FTLEG= foot and leg conformation, UA=udder attachment, RLEGR= rear leg placement - rear view, RLEGS= rear legs- side view, RTPL= rear teat placement, SCS= somatic cell score, MRCT= milk rennet coagulation time, FSC= first service conception, CONCEPT= number of inseminations per conception, MBCASP= milk B-casein percentage, MPFRAT= milk protein-to-fat ratio, DYF= dairy character, DYST= dystocia, TPL= teat placement, TLGTH= teat length, UDPTH=udder depth, PP= milk protein percentage, FP= milk fat percentage, HTINT= heat intensity, MKCASP= milk Kappa casein percentage, CONCRATE= conception rate, MUGKCASP= milk non-glycosylated kappa casein percentage, BTBS= bovine tuberculosis susceptibility, FATTH= fat thickness in the 12th rib, RFI= residual feed intake, NRR= non-return rate, MSPD= milking speed, BD= Body depth, FTPL=front teat placement, PL= productive life span, RUMWD= rump width, STR=milk strength, UC=udder cleft, LMY=lean meat yield, EY= energy of milk yield, BW= birth body weight, BVDV= bovine viral diarrhea susceptibility.

**NCBI genes.** 287026= phospholipase C beta 1, 506426= crystallin zeta protein encoder, 282257= subunit 1 of interferon alpha and beta receptor, 526800= ankyrin repeat domain 44, 19122= prion protein, 521004= solute carrier family 39 member 10, 338031= FMS-related receptor tyrosine kinase 4.

**Traces of selection (protein-coding genes).** TMX4= thioredoxin-related transmembrane protein 4, PLCB1= phospholipase C beta 1, MIR2285M-1= microRNAs involved in post-transcriptional regulation of gene expression, ATP5O= ATP synthase peripheral stalk subunit, OSCP= ATP synthase peripheral stalk subunit, ITS1= intersection 1, CRYZL1= crystallin zeta protein encoder, DONSON= DNA replication fork stabilization factor, SON= DNA and RNA-binding protein, GART=phosphoribosylglycinamide formyltransferase synthetase, Phosphoribosylaminoimidazole synthetase, DNAJC28= heat shock protein family, TMEM50B= transmembrane protein 50B, IFNGR2=interferon gamma receptor 2, IFNAR1= interferon alpha and beta receptor subunit 1, LOC104970778= uncharacterized RNA gene, IL10RB= interleukin receptor subunit beta, IFNAR2= interferon alpha and beta receptor subunit 2, LOC526226= histone H4, OLIG1= oligodendrocyte transcription factor 1, OLIG2= oligodendrocyte transcription factor 2, TMEM192= transmembrane protein 192, KLHL2= kelch of family 2, MSMO1= methylsterol monooxygenase 1, CPE= carboxypeptidase E, LOC101903170= uncharacterized RNA gene, SLC39A10= solute carrier family 39, DNAH7= dynein axonemal heavy chain 7, STK17B=serine/threonine kinase 17b, LOC531691= HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2, LOC107131408= olfactory receptor family 5 subfamily W member 39, LOC100125913= uncharacterized gene, LOC101902704= C-type lectin domain family, 7, A, FLT4= fms-related receptor tyrosine kinase 4, CNOT6= CCR4-NOT transcription complex subunit 6, GFPT2= glutamine-fructose-6-phosphate transaminase 2, MAPK9= mitogen-activated protein kinase 9, RASGEF1C= RasGEF domain family member 1C.

In the study population, ROHs that have been maintained as a result of the selection process of the family system population were identified. In these conserved regions, associations of SNP markers, QTL and gene are found, which are mostly related to characteristics of economic interest in the dairy industry, such as milk production and composition, fertility parameters, susceptibility to diseases, body conformation, feed efficiency and some other characteristics such as carcass composition, which could be taken as traces of selection (Table 3). These results show the traits that have been included in the selection processes in the population, either intentionally because of the selection made by cattle farmers or unintentionally because of the availability of genetic material in the market, since, by using AI, the choice of bulls guides the cattle farmer to modify the genetics of their animals in the way that AI companies do.

## Conclusions and implications

The dairy cattle from the FDS have ancestral origins from countries that are suppliers of genetic material internationally, such as the US and GBR, showing no evidence of recent crossbreeding with other dairy breeds. Within the studied population, it can be observed genetically homogeneous, with a lower number and length of ROHs than animals in specialized production systems, reflecting a wide genetic variation caused by a low intensity of selection. In this work, ROHs that have been maintained as a result of the selection process were identified, which are mostly related to characteristics of economic interest in the dairy industry. The results of this study reflect the existence of a low level of inbreeding in the population and a greater genetic diversity in this population compared to those found in specialized systems, so it is possible to carry out genetic improvement work aimed at the characteristics of interest of the producers through selection, without inbreeding compromising the productivity and health of the population.

## Acknowledgments and funding source

Project funded by INIFAP-CENIDFyMA with the name “Development of a sustainable comprehensive strategy to increase the availability of good quality Holstein replacements in the family milk production system in Mexico” with SIGI Number: 15352034772.

## Conflict of interest

The authors declare that there are no conflicts of interest.

## Literature cited:

1. Servicio de Información Agroalimentaria y Pesquera (SIAP) 2020. <https://www.gob.mx/siap/acciones-y-programas/produccion-pecuaria>.
2. Sainz-Sánchez PA, López-González F, Estrada-Flores JG, Martínez-García CG, Arriaga-Jordán CM. Effect of stocking rate and supplementation on performance of dairy cows grazing native grassland in small-scale systems in the highlands of central Mexico. *Trop Anim Health Prod* 2017;49(1):179-186.
3. Villamar AL, Oliveira CE. Situación actual y perspectiva de la producción de leche de bovino en México. Coordinación General de Ganadería. México, DF. SAGARPA. 2005.

4. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics* 2018; 19(106):1–12. <https://doi.org/10.1186/s12864-018-4489-0>.
5. Hillestad B, Woolliams JA, Boison SA, Grove H, Meuwisse T, Våge DI, *et al.* Detection of runs of homozygosity in Norwegian Red : Density, criteria and genotyping quality control. *Acta Agric Scandinavica, Section A — Anim Sci* 2018;67(3-4):107-116. <https://doi.org/10.1080/09064702.2018.1501088>.
6. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: Windows into population history and trait architecture. *Nat Rev Gen* 2018b;19(4):220–234. <https://doi.org/doi:10.1038/nrg.2017.109>.
7. Rebelato AB, Caetano AR. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. *Pes Agropec Brasileira* 2018;53(9):975–984. <https://doi.org/10.1590/S0100-204X2018000900001>.
8. Kim ES, Sonstegard TS, Van Tassell CP, Wiggans G, Rothschild MF. The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. *PloS One* 2015;10(7):e0129967. <https://doi.org/10.1371/journal.pone.0129967>.
9. Moravčíková N, Kasarda R, Kadlečík O, Trakovická A, Halo M, Candrák, J. Runs of homozygosity as footprints of selection in the norik of muran horse genome. *Acta Univ Agric Silv Mendel Brun* 2019; 67(5): 1165–1170. <https://doi.org/10.11118/actaun201967051165>.
10. Zavarez LB, Utsunomiya YT, Carmo AS, Neves HHR, Carvalheiro R, Ferencakovic M, Garcia JF. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. *Front Genet* 2015;5(1):1–8. <https://doi.org/10.3389/fgene.2015.00005>.
11. Ferencaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, *et al.* Estimates of autozygosity derived from runs of homozygosity : empirical evidence from selected cattle populations. *Anim Breed Genet* 2013;(130):286–293. <https://doi:10.1111/jbg.12012>.
12. Gurgul A, Szmatoła T, Topolski P, Jasielczuk I, Żukowski K, Bugno-Poniewierska M. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *J Appl Genet* 2016;57(4):527-530.
13. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. *Livest Sci* 2014;(166):133-143. <https://doi.org/10.1016/j.livsci.2014.05.003>.

14. Elferink MG, Megens HJ, Vereijken A, Hu X, Crooijmans RP, Groenen MA. Signatures of selection in the genomes of commercial and non-commercial chicken breeds. *PLoS One* 2012;7(2):e32720. <https://doi.org/10.1371/journal.pone.0032720>.
15. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livest Sci* 2020;(241): e104257. <https://doi.org/10.1016/j.livsci.2020.104257>.
16. Purfield DC, Berry DP, Mcparland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genetics* 2012;13(70):2–11. <https://doi.org/10.1186/1471-2156-13-70>.
17. Peripolli E, Munari DP, Silva MVB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: Current knowledge and applications in livestock. *Anim Gen* 2017;48(3):255-271. <https://doi.org/10.1111/age.12526>.
18. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;(19):1655–1664.
19. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 2003;165(12):2213–2233. <https://doi.org/10.1534/genetics.104.030692>.
20. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 2010;5(11):e13996. <https://doi.org/10.1371/journal.pone.0013996>.
21. Ferenčaković M, Sölkner J, Curik I. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genet Sel Evol* 2013; 45(1):1–9. <https://doi.org/10.1186/1297-9686-45-42>.
22. Golden Helix I, Bozeman M. SNP & Variation Suite TM Version 8 [Software]. Golden Helix, Inc. <http://www.goldenhelix.com>. Retrieved from [https://www.goldenhelix.com/products/SNP\\_Variation/index.html](https://www.goldenhelix.com/products/SNP_Variation/index.html).
23. SAS Institute© 2019 SAS Institute Inc. Cary, NC, USA [https://www.sas.com/es\\_mx/industry/life-sciences/solution/real-world-evidence.html](https://www.sas.com/es_mx/industry/life-sciences/solution/real-world-evidence.html).
24. Cortes-Hernández JG, Ruiz-López FJ, Vásquez-Peláez CG, García-Ruiz A. Runs of homocigosity and its association with productive traits in Mexican Holstein cattle. *PLoS One* 2022;17(9):e0274743. <https://doi.org/10.1371/journal.pone.0274743>.
25. Hu ZL, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Ac Res* 2019;47(1): D701-D710.



26. Chen N, Fu W, Zhao J, Shen J, Chen Q, Zheng Z, *et al.* BGVD: An integrated database for bovine sequencing variations and selective signatures. *Genom Proteom Bioinform* 2020;18(2):186-193.
27. Mcquillan R, Leutenegger A, Abdel-rahman R, Franklin CS, Pericic M, Barac-lauc, *et al.* Runs of homozygosity in European populations. *Am J Hum Genet* 2008;(83)e359–372. <https://doi.org/10.1016/j.ajhg.2008.08.007>.
28. Cortes-Hernández J, García-Ruiz A, Vásquez-Peláez CG, Ruiz-Lopez FdJ. Correlation of genomic and pedigree inbreeding coefficients in small cattle populations. *Animals* 2021;11(11):3234. <https://doi.org/10.3390/ani11113234>.
29. García-Ruiz A, Ruiz-López FD, Van Tassell CP, Montaldo HH, Huson HJ. Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and US Holsteins. *Front Genet* 2015;(6)7. <https://doi.org/10.3389/fgene.2015.00007>.
30. Bjelland DW, Weigel KA, Vukasinovic N, Nkrumah JD. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J Dairy Sci* 2013;96(7):4697-4706.
31. Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics* 2018;19(1):1–12. <https://doi.org/10.1186/s12864-018-4453-z>.
32. Marras G, Gaspa, G, Sorbolini S, Dimauro C, Ajmone-marsan P, Valentini A, *et al.* Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet* 2014;46(2):1–12. <https://doi.org/10.1111/age.12259>.
33. Mastrangelo S, Tolone M, Gerlando RD, Fontanesi L, Sardina MT, Portolano B. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Anim Consor* 2016;10(5):746–754. <https://doi.org/10.1017/S1751731115002943>.
34. Szmatoła T, Gurgul A, Ropka-Molik K, Jasielczuk I, Zabek T, Bugno-Poniewierska M. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. *Livest Sci* 2016;(188):72–80. <https://doi.org/10.1016/j.livsci.2016.04.006>.
35. 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 US Holstein cows. *BMC genomics* 2011;12(1):1-7.

36. Cai Z, Guldbrandtsen B, Lund MS, Sahana G. Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle. *BMC Genetics* 2019;20(15):1–12.
37. Dixit SP, Singh S, Ganguly, Bhatia AK, Sharma A, Kumar NA, *et al.* Genome-wide runs of homozygosity revealed selection signatures in *Bos indicus*. *Front Genet* 2020;11(2):1–11. doi:<https://10.3389/fgene.2020.00092>.