Looking at genetic structure and selection signatures of the Mexican chicken population using single nucleotide polymorphism markers

M. G. Strillacci,[*] V. E. Vega-Murillo,[†] S. I. Román-Ponce,[‡] F. J. Ruiz López,[‡] M. C. Cozzi,[*] E. Gorla,[*] S. Cerolini,[*] F. Bertolini,[§,#] L. Fontanesi,[§] and A. Bagnato[*,1]

[*]*Department of Veterinary Medicine, Universitá degli Studi di Milano, via Celoria 10, 20133 Milano, Italy;* [†]*Campo Experimental La Posta, INIFAP, km 22.5 Carretera Federal Veracruz-Córdoba, Paso del Toro, Municipio de Medellín, 94277, Veracruz, México;* [‡]*Centro Nacional de Investigación en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP), Km. 1 Carretera a Colón, Auchitlán, 76280, Querétaro, México;* [§]*Department of Agricultural and Food Sciences - Division of Animal Sciences, University of Bologna, Viale Fanin 46, Bologna, Italy; and* [#]*Department of Animal Science, Iowa State University, 1221 Kildee Hall, Ames, IA 50011*

ABSTRACT Genetic variation enables both adaptive evolutionary changes and artificial selection. Genetic makeup of populations is the result of a long-term process of selection and adaptation to specific environments and ecosystems. The aim of this study was to characterize the genetic variability of México's chicken population to reveal any underlying population structure. A total of 213 chickens were sampled in different rural production units located in 25 states of México. Genotypes were obtained using the Affymetrix Axiom® 600 K Chicken Genotyping Array. The Identity by Descent (IBD) and the principal components analysis (PCA) were performed by SVS software on pruned single nucleotide polymorphisms (SNPs).

ADMIXTURE analyses identified 3 ancestors and the proportion of the genetic contribution of each of them has been determined in each individual. The results of the Neighbor-Joining (NJ) analysis resulted consistent with those obtained by the PCA. All methods utilized in this study did not allow a classification of Mexican chicken in distinct clusters or groups. A total of 3,059 run of homozygosity (ROH) were identified and, being mainly short in length (<4 Mb), these regions are indicative of a low inbreeding level in the population. Finally, findings from the ROH analysis indicated the presence of natural selective pressure in the population of Mexican chicken.

The study indicates that the Mexican chicken clearly appear to be a unique creole chicken population that was not subjected to a specific artificial selection. Results provide a genetic knowledge that can be used as a basis for the genetic management of a unique and very large creole population, especially in the view of using it in production of hybrids to increase the productivity and economic revenue of family farming agriculture, which is widely present in México.

Key words: autochthonous chicken, selection signature, ROH, SNP, biodiversity

INTRODUCTION

The knowledge of the genetic variation within and across populations is essential in the process of identification of local genetic resources (i.e., individuals of local poultry breeds) to be maintained in animal genetics conservation efforts (Cavalchini et al., 2007). Microsatellites markers have been widely used to analyze genetic variability in the chicken population (Strillacci et al., 2009; Al-Qamashoui et al., 2014; Ceccobelli et al., 2015). Recently, the availability of high throughput genomic information, i.e., sequencing data and high-density single nucleotide polymorphism (**SNP**) arrays, has opened the possibility to investigate the genetic structure of populations using a very large number of markers and to highlight genomic regions where events related to selection pressure occur (Fleming et al., 2016; Strillacci et al., 2017). Chickens can be easily utilized for the study of the signatures of selection under artificial breeding conditions, thanks to their relatively fast reproduction time (Brown et al., 2003). Theoretically, functional genes under selection are exposed to a change in allele frequency that can be identified analyzing the characteristic DNA pattern that derives, known as selection signature (Fan et al., 2014). In other words, selection signatures are particular patterns of DNA that can be identified in regions of the genome that include a mutation, that is or has been under selection pressure in the population (Qanbari and

1

Simianer, 2014). Whenever in positive selection for a particular allele, these regions are expected to exhibit larger homozygosity than expected under the Hardy-Weinberg equilibrium. Many measures can be utilized to estimate genetic variability pattern along the genome using marker data; among them, runs of homozygosity (**ROH**) are contiguous lengths of homozygous genotypes that develop as a result of parental transmission of identical haplotypes (Gibson et al., 2006). Long ROH (∼10 Mb) are a consequence of recent inbreeding (up to 5 generations ago), whereas shorter ROH (∼1 Mb) can be related to a more distant ancestral positive selection effect (up to 50 generations ago), because recombination events that break long chromosomes into segments (Mastrangelo et al., 2016) have reduced their size along the reproductive events. Recently, the availability of sequencing and high-throughput SNP datasets has permitted to release chromosome-wide molecular diversity and population structure studies (Nimmakayala et al., 2014). Furthermore, it is possible to disclose traces of positive selection and identify possible candidate genes associated with selection (Fan et al., 2014).

Local chicken populations are considered an important genetic resource, derived after hundreds of years of successful adaptation in areas with peculiar environmental characteristics, with limited veterinary and management support (Hall and Bradley, 1995). Phenotypic traits variability is little known in backyard poultry population, as well as those genes that cause their adaptability to local environments. It is also not clear if the geographical origins of that local chicken population is one of the causes of their genetic differentiation, making them so various (Mahammi et al., 2016). In México, poultry population is not classified in breeds, but there is a diffusion of the creole chicken (*Gallus gallus domesticus*), coming from European chickens brought to México by the Spanish conquerors during the 16th century. They originate from undefined crosses among different breeds for almost 500 years. Because of that, creole chickens include a wide range of variable biotypes, having different morphological features and characterized by high feed conversion, low growth rate, low egg production, and small egg size under semi-intensive or scavenging conditions (Segura-Correa et al., 2004, 2005). The Mexican population is, de facto, under natural adaptive selection for more than 5 centuries, making it a very interesting one to disclose genetic variation related to resilience in harsh environments. The Mexican population is in fact a genetic resource that can express genes lost in the industrial selection process, targeted to increase meat and egg productions. As recently well disclosed by Fleming et al. (2016, 2017) studying genetic variation in African native populations, the existence of proprietary genetic variation in native breeds related to specific environmental conditions (e.g., hot and humid climates or heat waves) is the basic knowledge for its introgression in F1 individuals, crossing for native females populations (natural selection occurring in population) and artificially selected males (artificial selection). To our knowledge, there have been some attempt to characterize phenotype and performance of Mexican creole chickens but up to now, no molecular characterisation studies related to genetic variability and phylogenetic analysis of this population have yet been realized using dense panels of SNPs, except the recent study of Gorla et al. (2017) who used copy number variants to dissect genetic variability in the Mexican population.

The aim of this study was to describe the genetic variability of Mexican chickens to reveal any underlying population structure using a dense SNP panel and to identify selection signatures using ROH, characterizing the inbreeding level of this chicken population and disclosing the genes under positive selection for adaptive variants in an outbred population.

MATERIALS AND METHODS

Sampling and Genotyping

In the present study, a total number of 213 chickens' feathers were sampled in different rural production units located in 25 states of México (Aguascalientes, Baja California Sur, Campeche, Chiapas, Chihuahua, Coahuila, Colima, México City, Durango, Estado de México, Guanajuato, Guerrero, Hidalgo, Jalisco, Morelos, Nayarit, Nuevo León, Oaxaca, Querétaro, Tabasco, Tamaulipas, Tlaxcala, Veracruz, Yucatan, and Zacatecas) by the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP). This sample collection is part of an institutional project entitled, "Identificación de los recursos genéticos pecuarios para su evaluación, conservación y utilización sustentable en México. Aves y cerdos. SIGI NUMBER 10,551,832,012" [Identification of livestock genetic resources for their evaluation, conservation, and sustainable utilization in México. Fowl and pigs.] The project was coordinated with the activities of the Centro Nacional of Recursos Genéticos (**CNRG**) at Tepatitlán, Jalisco (México) engaged in promoting strategic research to solve the most important problems of productivity, competitiveness, equity and sustainability at the forest, agricultural and livestock sectors in México (http://www.inifap.gob.mx/SitePages/centros/cnrg.aspx). The samples are owned by the CNRG, who controls their access and reuse. Original owners of individuals have donated the samples to CNRG who gave consent for reuse for research purposes. The study did not require any ethical approval according to national rules, according to EU regulation, as it does not foresee sampling from live animals. The University of Milan permit for the use of collected samples in existing bio-banks was released with n. OPBA-56–2016.

DNA extraction from feathers and genotyping were performed at GeneSeek (Lincoln, NE) using a commercial kit and the Affymetrix Axiom® 600 K Chicken Genotyping Array, containing 580,954 SNPs, distributed across the genome with an average spacing of 1.7 Kb, respectively. The galGal4 chicken assembly was

used in this study as reference genome. Only markers positioned on chromosome 1 to 28 were used in this study.

A quality control of raw intensity files using the standard protocol in the Affymetrix Power Tools package (www.affimetrix.com) was performed in order to guarantee a high quality of genotyping data. Samples with Dish Quality Control (**DQC**: the closer the value is to 1, the better the signal separates from the background) <0.82 and with quality control (**QC**) call rates <97% were excluded from downstream analysis. The quality verified samples were used for subsequent SNP analyses using dedicated software.

Morphological Chicken Characterization

Morphological characteristics of collected Mexican creole individuals are extremely variable in terms of feathers colors, shapes (i.e., naked neck/breast or not, fighting characteristics), comb, and size. The measurement of morphological characteristics of birds was done according to the FAO Guidelines (2012), which is the recognized standard. Measures were taken at sampling and recorded in a database. The STANDARD procedure of SAS 9.4 (2013) was used to create a dataset with standardized values (mean = zero; standard deviation = 1) for the following 4 quantitative variables: 1) body length—length between the tip of the rostrum maxillare (beak) and that of the cauda (tail, without feathers); the bird's body should be completely drawn throughout its length; 2) wingspan: length in cm between tips of right and left wings after both are stretched out in full; 3) breast circumference: circumference of the chest (taken at the tip of the pectus, hind breast); 4) length of the shank: (length in cm of the shank from the hock joint to the spur of either leg).

Subsequently, a FASTCLUS procedure of SAS 9.4 (2013) was used to perform a disjoint cluster analysis on the basis of distances computed from one or more quantitative variables. The observations were divided into clusters such that every observation belongs to one and only one cluster. Four clusters were generated by the program with a Cubic Clustering Criterion (CCC) of 15.74. Following the FASTCLUS the CANDISC procedure was run on the 4 body measures as variables using the clusters previously created as classes. The CANDISC procedure performs a multivariate one-way analysis of variance and provides 4 multivariate tests under the hypothesis that the class means vectors are equal.

Genetic Characterization

Different approaches and software were used in order to disclose the genetic structure of Mexican chickens:

a) using SVS Golden Helix 8.4 software (SVS) (Golden Helix Inc., Bozeman, MT) the Identity by Descent (**IBD**) estimation and the Principal Components Analysis (**PCA**) were performed. The IBD is a measure of the relatedness of the pair of individuals and indicates how many alleles at any marker in each of 2 individuals came from the same ancestral chromosomes. The estimation of the IBD between all pairs of samples was done after the application of LD pruning option. Relationship-based pruning was performed and one member of each pair of animals with an observed genomic relatedness greater than 0.25 was removed from further analyses. The PCA, of pairwise individual genetic distances, was performed based on allele frequencies of pruned SNPs. To visualize the individual samples relatedness graphically in multi-dimensions the rgl R package (https://CRAN.R-project.org/package = rgl) was used.

b) The ADMIXTURE v. 1.3.0 software was employed to infer the most probable number of ancestral populations based on the SNP genotype data (Alexander et al., 2009). ADMIXTURE was run from K = 2 to K = 6, and the optimal number of clusters (**K-value**) was determined as the one having the lowest cross-validation error. Each inferred chicken population structure was visualized using R script suggested in the ADMIXTURE procedure.

c) Wright's statistics, including observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding estimates (F_{IS}) were calculated with SVS.

d) Neighbor-Joining (**NJ**) tree, constructed based on the allele sharing distances (**DAs**) as the genetic distance between not-related individuals, was created and graphically represented using PEAS (Xu et al., 2010) and FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree) software, respectively.

e) The Arlequin v.3.5.2.2 software (Excoffier and Lisher, 2010) was used to perform an Analysis of MOlecular VAriance (**AMOVA**) a tool to check how the genetic diversity is distributed among individuals within groups, whose structure is quantified by F_{ST}.

f) ROH analysis was performed for each individual (complete SNP dataset = 471,730), using the SVS software. The ROH was defined by: 1) a minimum of 1,500 kb in size and 50 homozygous SNPs; 2) one heterozygous SNP was permitted in ROH, so that the length of the ROH was not disrupted by an occasional heterozygote; 3) 5 missing SNPs were allowed in the ROH; 4) maximum gap between SNPs of 100 Kb was predefined in order to assure that the SNP density did not affect the ROH. According to the nomenclature reported by other authors (Curik et al., 2014), ROH were grouped into 5 classes of length: <2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb and >16 Mb. Number, total length and the average of ROH length were calculated across
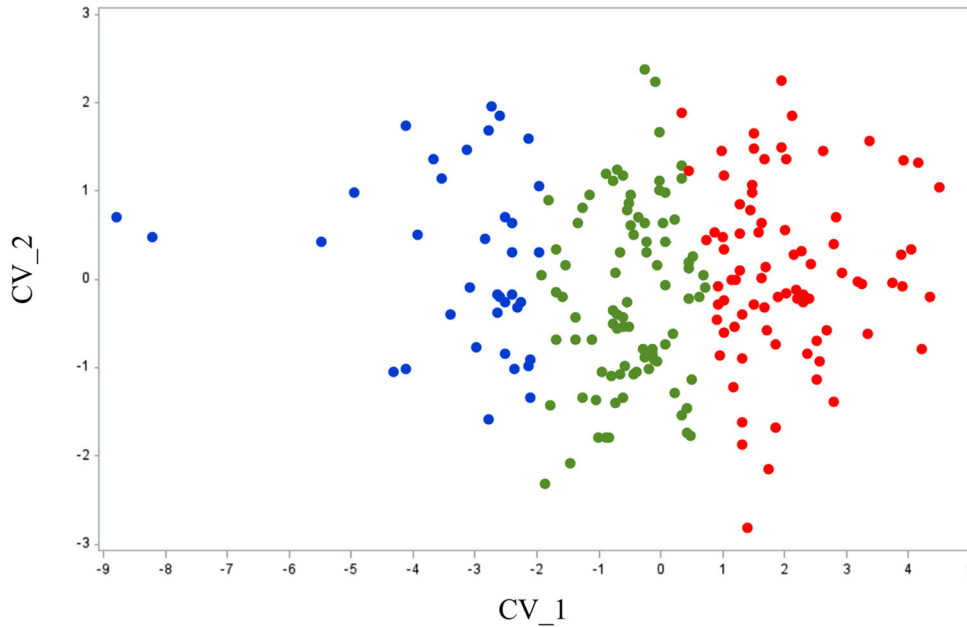
Figure 1. PCA based on morphological features (body length, wingspan, breast circumference, length of the shank): Cl_1: blue, Cl_3: red, and Cl_4: green. Canonical variable 1: CV_1; Canonical variable 2: CV_2

individuals within chicken population. In addition, the percentage of the total genome length affected by ROH was also estimated.

RESULTS

Morphological Chicken Characterization

The analysis of morphological measures (body length, wingspan, breast circumference, length of the shank) to cluster individuals separated the population into 4 different groups. While the clusters 1 (Cl_1), 3 (Cl_3), and 4 (Cl_4) were composed by 36, 74, and 72 animals, respectively, only one individual belonged to cluster 2. This latter individual, as such as cluster 2, were eliminated from subsequent analyses.

In Figure 1 the scatter plot of the canonical variables 1 *vs.* 2 based on morphological measures is shown. The distinction among the 3 clusters was clearly displayed on the canonical variable plotted as x-axis representing 99% of the total variance.

The Table S1 shows that the individuals in the 3 clusters exhibit different sizes with CL_1 being in general the smaller individuals, CL_4 the intermediate, CL_3 the cluster with birds of larger dimension. This appears also from Figure 1 where CV_1 clearly differentiate the 3 clusters with CL_4, the green one, and intermediate respect the other 2.

Genetic Characterization

SNPs with Minor Allele Frequency (**MAF**) value ≤0.01, HWE value 0.00001, SNPs not on autosomal chromosomes from 1 to 28, and SNPs having a call rate

<97% were excluded, reducing to 471,730 markers the number of SNPs used for the statistical analysis. SNPs passing the QC were pruned for LD using a threshold of $r^2 = 0.5$. LD trimming resulted in another 207,245 SNPs pruned from the dataset, ensuing in a final set of 264,485 SNPs used in the downstream analysis. Of the 213 animals sampled, 31 showed an IBD value greater than 0.25 with at least one other individual of the population, and then were subsequently removed leaving 182 animals for the population structure analyses. The remaining population is thus holding individuals with IBD less then 0.25 IBD value as maximum value. Out of the 16,471 IBD values only 337 were comprised in the interval $0.125 \leq$ IBD <0.25, 561 in the interval $0.0625 \leq$ IBD < 0.125, 646 between $0.0625 \leq$ IBD < 0.03125, while the remaining all less than 0.03125. According to this distribution we considered all individuals unrelated. The Heat map created using the IBD estimates values is showed in Figure 1S in Additional File 1.

The program ADMIXTURE was run for K values from 1 to 6 (Figure 2A). The lowest cross validation error was found at K = 3, that represent the number of ancestors in the Mexican populations (Figure 2B). A number of K greater than 3 does not produce a larger number of ancestors' contribution in the living population, as it is visible in Figure 2A.

The Figure 2C is a graphical representation of the 182 individuals grouped according to the proportion of the 3 ancestors' contribution. One individual results to derive entirely from ancestor 1, while 7 derives entirely from ancestor 2 and 11 from ancestor 3. A total of 25 individuals showed to derive from 2 ancestors while the largest proportion of the sample, 138 individuals, showed a genetic composition that derived from all the
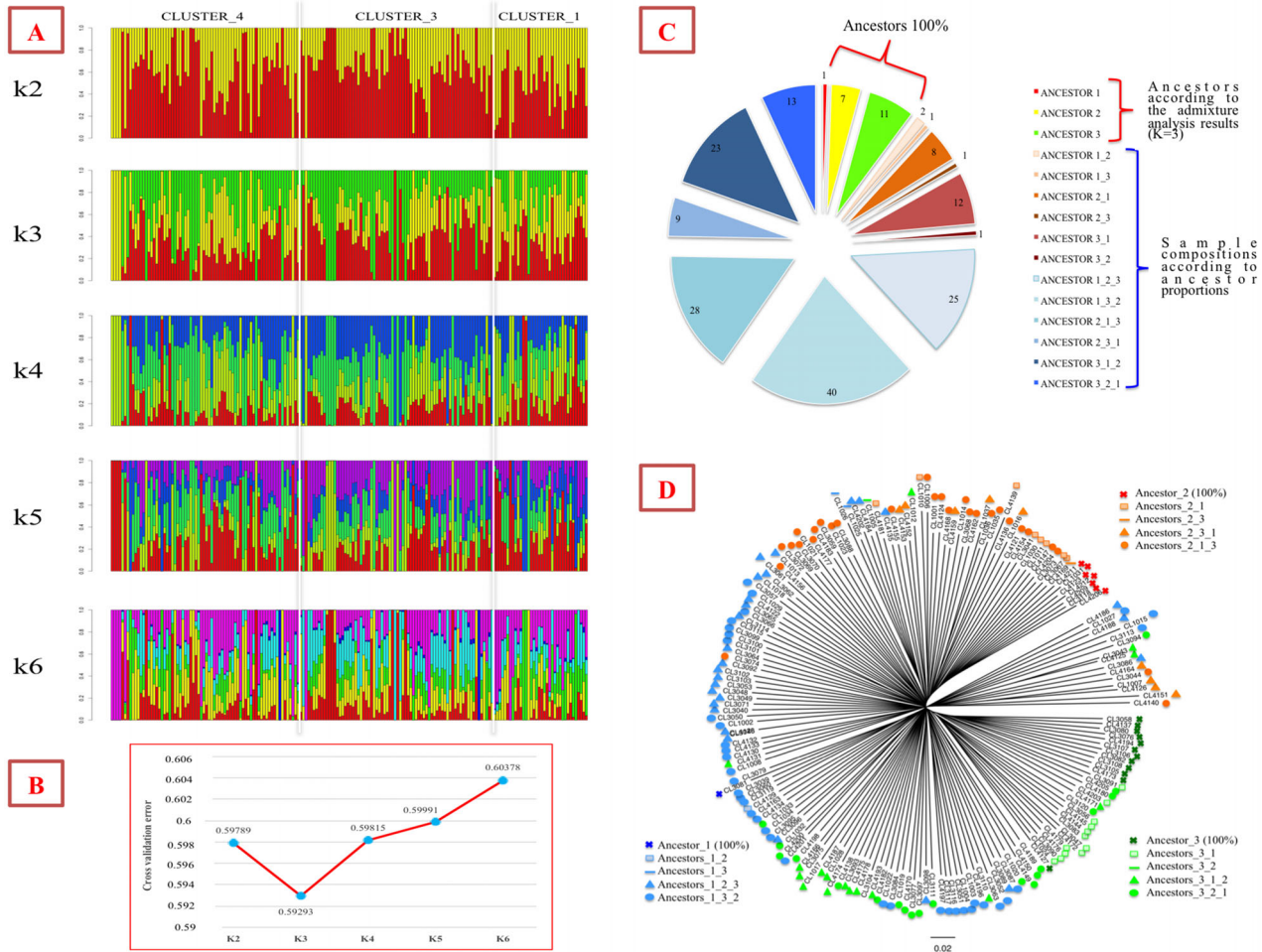
Figure 2. Graphical representation of Mexican chicken population genetic structure. A) ADMIXTURE k = 2-K = 6 barplots; B) Optimal number of clusters according to cross-validation error; C) Count of individuals based on 3 ancestors' composition; D) NJ tree: classification of individuals according to allele sharing distances. Individuals were labeled according to the morphological cluster (i.e., from 1 to 4) they belong and their individual (e.g., CL1012 = morphological cluster 1, individual 012) and the ancestors' composition from ADMIXTURE.

3 identified ancestors. Apparently there is no clear relationship between the morphological clustering and the ancestor's composition. Table S2 shows the bird count according to the ancestor's composition classes and the morphological clusterization. The largest part of individuals pertaining to ancestor 1 class (i.e., 57%) showed morphological characteristics of birds classified as cluster 3, while individuals pertaining to ancestor 2 (i.e., 50%) and 3 (i.e., 49%) showed characteristics of animals classified as cluster 4.

The results of the NJ analysis depicted in Figure 2D are consistent with those obtained by the PCA.

It is possible to note that the major part of samples are grouped according to the ancestor's composition, but individual differences based on DAs did not allow a clear division of birds in well separated clusters (Figure 2D).

The results of the PCA agreed well with the findings outlined above, as showed on Figure 3A e 3B. In both PCA analyses of Figure 3 there is no clustering of individuals neither for morphological cluster than for the ancestor classification, as points are mixed in all distributions depicted.

The Table 1 reports the results for the AMOVA analysis. The analysis account of individual classification was based on morphological clustering (Cl_1; Cl_3 and Cl_4). We considered 3 hypotheses: Hypothesis 1) Cl_1 + Cl_3 *vs* Cl_4; Hypothesis 2) Cl_1 + Cl_4 *vs* Cl_3; Hypothesis 3) Cl_3 + Cl_4 *vs* Cl_1. All the hypotheses indicate that the most part of variability is observed within clusters, 99.72% (Hypothesis 1), 99.56% (Hypothesis 2) and 99.49% (Hypothesis 3), with a much smaller amount of the variance component occurring among groups 0% (Hypothesis 1), 0.18% (Hypothesis 2) and 0.22% (Hypothesis 3) (Table 1). The AMOVA confirmed the results obtained with the PCA. In other words the genetic variation of the Mexican population appears to be mostly related to the individual genetic variability rather than to the genetic diversity expressed by the clustering classification obtained on the basis of the morphological characteristics.
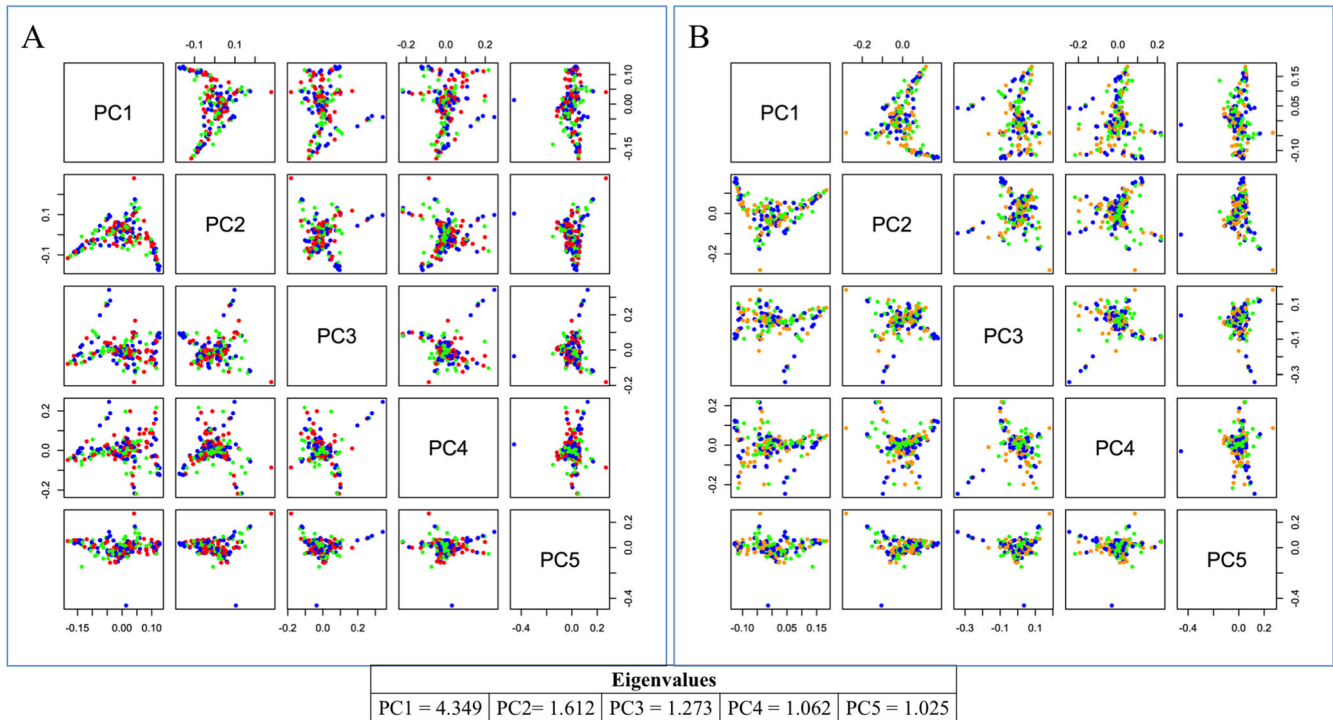
Eigenvalues				
PC1 = 4.349	PC2= 1.612	PC3 = 1.273	PC4 = 1.062	PC5 = 1.025

Figure 3. A) PCA based on allele frequencies of SNPs (individuals were colored according to the 3 morphological clusters: Cl_1: blue, Cl_3: red, and Cl_4: green); B) PCA based on allele frequencies of SNPs (individuals were colored according to the individual ancestor's composition: ancestor_1: blue, ancestor_2: orange, and ancestor_3: green).

Table 1. Hierarchical AMOVA analysis among the clusters obtained based on allele frequencies of pruned SNPs.

	Variance component (%)			Fixation indexes[a]					
Hypotheses	Among groups	Among clusters within groups	Within clusters	ΦCT	P-value[b]	ΦSC	P-value[b]	ΦST	P-value[b]
Cl_1+Cl_3 vs Cl_4	− 0.47[c]	0.75	99.72	− 0.005	1.000 ± 0.000	0.007	0.000 ± 0.000***	0.003	0.000 ± 0.000***
Cl_1+Cl_4 vs Cl_3	0.18	0.26	99.56	0.002	0.658 ± 0.011	0.003	0.043 ± 0.001*	0.004	0.001 ± 0.001*
Cl_3+Cl_4 vs Cl_1	0.22	0.29	99.49	0.002	0.332 ± 0.016	0.003	0.004 ± 0.002*	0.005	0.000 ± 0.000***

[a]ΦCT = variation among groups divided by total variation, ΦSC = variation among sub-groups divided by the sum of variation among sub-groups within groups and variation within sub-groups, ΦST = the sum of variation groups divided by total variation.
[b]ns = $P > 0.05$, * = $P < 0.05$, *** = $P < 0.001$.
[c]Negative values are presented, but we can consider this value effectively equal to zero.

Run of Homozygosity (ROH) Analysis

The SVS software identified a total of 3,059 runs across Mexican chicken population (Supplementary Table S3). Six individuals did not show any ROH in any of the 28 chromosomes. Likewise, the chromosomes 16 and 25 showed no evidence of ROH in all genotyped individuals. Results revealed that there were marked differences in terms of number and length of ROH across individuals.

The ROH have been defined with 305 and 6,629 SNPs as minimum and maximum number of SNPs. The average number of SNPs falling into a ROH was consistent among ROH length category, ranging from 824 (ROH < 2 Mb) to 3,977 (ROH > 8–16 Mb) SNPs.

The identified ROH are mainly short in length; in fact, the ROH of 2–4 Mb and <2 Mb are the most frequent classes of length identified (i.e., 84%). Instead, no ROH were found within the >16 Mb length class (Table 2).

The number of ROH per individual ranged from 1 to 115, with a mean number of ROH for sample of 17.38 (Figure 3). The Figure 3 also shows the relationship between number and averaged total length of ROH for each individual (mainly ranged from 1.7 to 2.8 Mb). Only 2 samples showed a very high number of ROH (i.e., 110 and 115 ROH). The average size of ROH of these 2 individuals is nevertheless similar to other subject. ROH larger than 3 Mb were found in 38 individuals representing 21% of the total sample and showing a count range of ROH from 1 to 65. The amount of the genome covered by ROH per individual ranged (as mean values) from 1,563,036 bp to 4,387,646 bp (Figure 4).

The relative frequencies of ROH (calculated as number of ROH per class on total number of ROH) within each chromosome and by length classes (Table 2) were also calculated. The ROH of 8 to 16 Mb size were found in longer chromosomes, the total number of ROH appeared to be proportional to their lengths

Table 2. Numbers of ROH per chromosome according to ROH classes of length.

Chr	< 2 Mb (*)	2–4 Mb (*)	4–8 Mb (*)	8–16 Mb (*)	> 16 Mb (*)	Total
			Classes of ROH			
1	247 (0.34)	350 (0.48)	123 (0.17)	11 (0.02)	0 (0)	731
2	190 (0.35)	255 (0.47)	83 (0.15)	11 (0.02)	0 (0)	539
3	147 (0.38)	182 (0.47)	56 (0.14)	5 (0.01)	0 (0)	390
4	110 (0.34)	162 (0.5)	45 (0.14)	8 (0.02)	0 (0)	325
5	97 (0.39)	104 (0.42)	46 (0.18)	3 (0.01)	0 (0)	250
6	51 (0.42)	46 (0.38)	23 (0.19)	1 (0.01)	0 (0)	121
7	41 (0.32)	71 (0.56)	14 (0.11)	1 (0.01)	0 (0)	127
8	39 (0.41)	50 (0.53)	6 (0.06)	0 (0)	0 (0)	95
9	23 (0.34)	42 (0.62)	3 (0.04)	0 (0)	0 (0)	68
10	27 (0.37)	36 (0.49)	10 (0.14)	0 (0)	0 (0)	73
11	27 (0.44)	30 (0.48)	5 (0.08)	0 (0)	0 (0)	62
12	20 (0.48)	20 (0.48)	2 (0.05)	0 (0)	0 (0)	42
13	16 (0.43)	19 (0.51)	2 (0.05)	0 (0)	0 (0)	37
14	15 (0.52)	14 (0.48)	0 (0)	0 (0)	0 (0)	29
15	6 (0.22)	18 (0.67)	3 (0.11)	0 (0)	0 (0)	27
16	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0
17	10 (0.36)	18 (0.64)	0 (0)	0 (0)	0 (0)	28
18	9 (0.53)	8 (0.47)	0 (0)	0 (0)	0 (0)	17
19	12 (0.52)	10 (0.43)	1 (0.04)	0 (0)	0 (0)	23
20	21 (0.49)	19 (0.44)	3 (0.07)	0 (0)	0 (0)	43
21	3 (0.38)	5 (0.63)	0 (0)	0 (0)	0 (0)	8
22	1 (0.33)	2 (0.67)	0 (0)	0 (0)	0 (0)	3
23	3 (0.6)	2 (0.4)	0 (0)	0 (0)	0 (0)	5
24	6 (0.67)	3 (0.33)	0 (0)	0 (0)	0 (0)	9
25	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0
26	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	1
27	1 (0.25)	3 (0.75)	0 (0)	0 (0)	0 (0)	4
28	2 (1)	0 (0)	0 (0)	0 (0)	0 (0)	2
Total	1124 (0.36)*	1470 (0.48)*	425 (0.14)*	40 (0.02)*	0 (0)*	3059

(*) Proportion calculated as number of ROH per class over the total number of ROH.
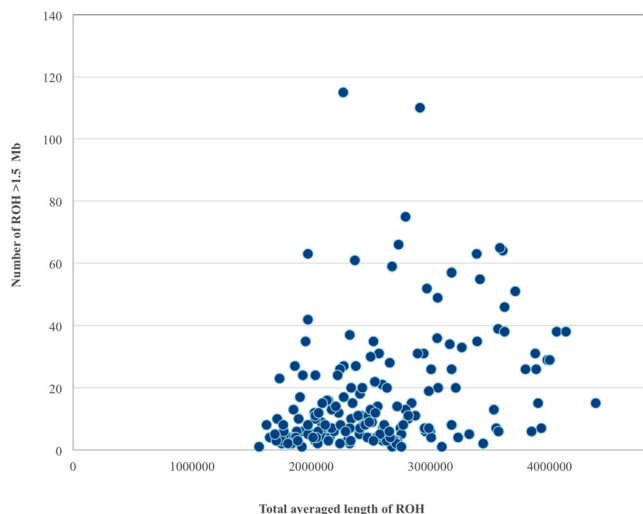

Figure 4. Relationship between number and averaged length of ROH in each individual.

and were distribution appeared homogeneous across them.

The genomic regions most commonly associated with ROH have been identified by selecting the top 1% of the SNPs most frequently observed in the ROH (Top 1% ROH). Figure 5 shows the incidence of ROH segments across the genome and as appear, the genomic distribution of ROH segments was clearly non-uniform across chromosomes. A total of 11 regions were identified with frequencies of ROH segments exceeding 1%

of the whole population (Top 1% ROH) in the first 8 chromosomes, excluding chromosome 6.

After downloading the list of chicken autosome galGal4 genes (GCA_0,00002315.2) from Ensembl database (http://www.ensembl.org), the annotation of gene mapping within the Top 1% ROH is reported in Table 3.

DISCUSSION

Patterns of high-density SNPs variation were used in this study to detect genetic variability in a chicken population collected in several states of México. All findings provided in this research, using several statistical approaches, confirmed and highlighted a non-structural classification of individuals in well-differentiated subpopulations, even if ADMIXTURE statistic identified 3 possible ancestors to define the predominant genetic background in the population.

The effective number of polymorphic SNPs (considered as the number of SNP in which at least one heterozygous individual was identified) represents the 99.9% of the total loci. The moderately high values of H_O (0.319) and H_E (0.348) reflect the high percentage of polymorphic SNP; the low F_{is} value (0.084) are indicative of a low level of inbreeding in the population and of the relatively high number of birds in a heterozygous state. In other native populations where the heterozygosity and F_{is} was recently calculated using a
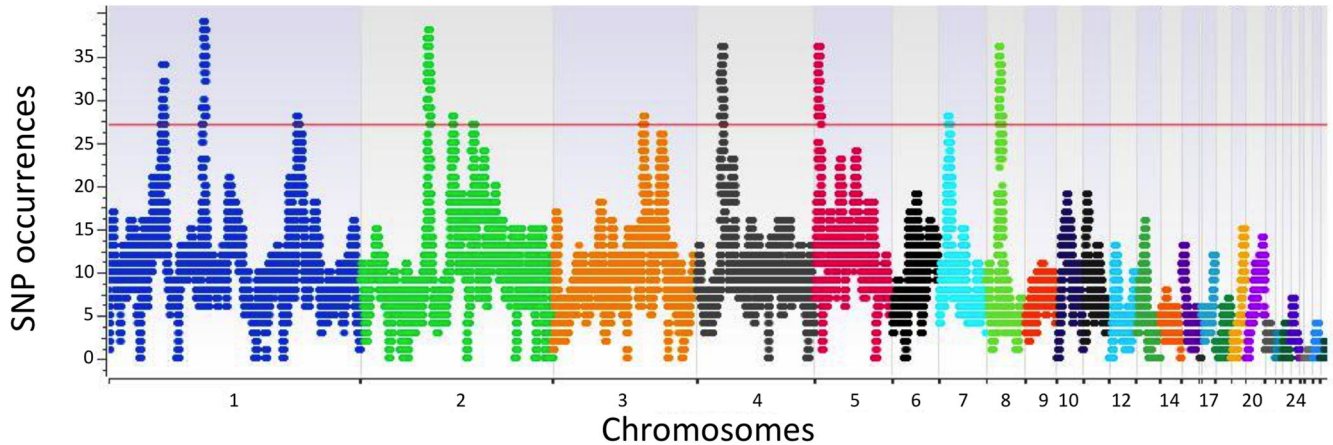
Figure 5. SNPs incidence in ROH identified by SVS. Red line indicates the adopted threshold: Top 1% of the observations.

high-density SNP chip (Strillacci et al., 2017), the H_O varies from 0.21 to 0.34, the H_E from 0.17 to 0.32 and the F_{is} from −0.19 to 0.094. These populations nevertheless are very well characterized in different breeds, thus showing more homogeneity within the same group of individuals. Using a 60 K SNP chip (Johansson and Nelson, 2015) found an F_{is} value of −0.09 and 0.17 in 2 local chicken populations indicating that farmers do not increase inbreeding excessively. Our results thus show that in outbred creole populations as the Mexican one, the genetic variability appear larger respect to local populations defined in breeds.

As expected, the results of AMOVA showed that the most of the genetic variation occurred within populations in all the 3 hypotheses here considered and confirm the absence of a genetic structure in the Mexican chicken population. The slightly negative value for the variance in fact, as obtained in hypothesis 1 (i.e., −0.47), can occur in absence of genetic structure, and is a quite common occurrence in AMOVA, as the real parameter value has to be considered zero. The negative or slightly positive values of among groups variance and ΦCT for all hypothesis (Excoffier, 2007), thus confirm the absence of a hierarchical genetic structure in the Mexican poultry population. These findings also confirm the results from Gorla et al. (2017) who, using a different approach and a different class of genetic markers, did not disclosed a genomic structure in the Mexican chicken population. We did not consider the analysis by ancestor as the classes are extremely numerous and unbalanced among them (see Table 1).

It is to be recalled that the Mexican poultry population is a creole unique genetic pool that have not been selected for target traits for more than 500 years. As consequences of its adaptation to the environmental conditions and production, some genomic region may be fixed in individuals as a result of positive selection.

These results here obtained for ROH are in concordance with those identified in previous studies (Gibson et al., 2006) where short ROH with high frequencies were identified in outbred individuals, as well as the intermediate sizes runs. ROH greater than 10 Mb, gen-

erally identified in individuals belonging to populations with high levels of background relatedness, have been also identified in 2% to 26% of individuals pertaining to outbred populations (Pemberton et al., 2012), and in a proportion of 14% in our birds. These findings may reflect a recent parental relatedness, or be the result of a recombination lack that allows uncommonly long ancestral genomic segments to persist in the population (Pemberton et al., 2012). Additionally findings from the ROH analysis indicated that natural selection affected allele frequencies in specific regions of the Mexican chicken genome (Figure 5).

Among the annotated genes in the ROH regions, in fact, some are worth mentioning because their functions could play important roles in the historical genetic dynamic occurred to the Mexican chicken population.

On chr1 within the ROH_1 (at 41.38–43.21 Mb) lies the KIT ligant (**KITLG**) gene that has a role in controlling the migration, survival and proliferation of melanocytes; also rare mutations in the mouse homolog of the KITLG gene are known to affect coat colour (Sulem et al., 2007). Additionally Metzger et al. (2015) highlighted the role of this gene in the horse reproduction efficiency, claiming its general effect in all livestock populations. The activation-induced cytidine deaminase (**AICDA**) gene mapping within the ROH_2, encodes for a DNA-editing protein that plays an essential role in some events of immunoglobulin (**Ig**) diversification: somatic hypermutation, class switch recombination and Ig gene conversion (Carãtao et al., 2013). These processes generate the vast diversity of antibodies required to challenge a nearly infinite number of antigens that immune systems encounter (Keim et al., 2013). In the same ROH_2 the von Willebrand factor (**VWF**) gene and the fibrinogen beta chain (**FGB**), the fibrinogen gamma chain (**FGG**) and the fibrinogen alpha chain (FGA) genes located within ROH_8, are 4 of the 8 hemostatic genes resulted down regulated in studies based on RNA-Seq analysis on breast muscle of chickens affected by "wooden breast disease" (Mutryn et al., 2015). The ROH_9 on chr5 (2.60–3.95 Mb) harbors the brain derived neurotrophic factor (**BDNF**)

Table 3. The eleven top 1% ROH identified on Mexican chicken autosomes by SVS.

ROH_ID	Chr	Start	End	Length	Genes*	QTL (http://www.animalgenome.org/cgi-bin/QTL_IDdb/GG/index)
ROH_1	1	41,387,392	43,210,859	1823,467	**NTS, KITLG**, *DUSP6*	Femur bending strength QTL_ID (6758); Yolk weight QTL_ID (24,938, 24,939, 24,940); Breast muscle percentage QTL_ID (95,427); Growth (post-challenge) QTL_ID (65,829)
ROH_2	1	73,476,359	75,663,705	2187,346	*CCND2, NDUFA9,* **NTF3, VWF**, *TEAD4, RHNO1,* **FKBP4**, *FOXM1, NANOG,* **AICDA**, *PHC1*	
ROH_3	1	146,817,860	147,817,564	999,704		
ROH_4	2	51,510,051	54,136,958	2626,907	*PSMA2, STK17A,* **EGFR**, *SEC61G*	Body weight (70 days) QTL_ID (12,390); Body weight (56 days) QTL_ID (12,391); Body weight (70 days) QTL_ID (12,392)
ROH_5	2	70,951,155	71,838,862	887,707	*ENS-3, mir6545*	Body weight (42 days) QTL_ID (6899); Abdominal fat weight QTL_ID (6900); Breast muscle weight QTL_ID (6968)
ROH_6	2	86,255,347	87,498,657	1243,310	*IRX1, IRX2*	Egg shell color QTL_ID (1914); Body weight (42 days) QTL_ID (6901); Breast muscle percentage QTL_ID (12,569)
ROH_7	3	68,723,915	70,012,473	1288,558		
ROH_8	4	18,018,373	20,496,038	2477,665	*IDS, TLR2A, TLR2B, TRIM2, MND1, SFRP2,* **FGB, FGA, FGG**, *NPY2R, CTSO, mir7469*	Abdominal fat weight QTL_ID (19,531, 19,535, 19,538); Body weight (40 days) QTL_ID (6659); Egg shell color QTL_ID (3348); Muscle fiber density QTL_ID (19,534, 19,537); Muscle fiber diameter QTL_ID (19,533,19,536,19,340); Residual feed intake QTL_ID (7057); Subcutaneous fat thickness QTL_ID (19,532, 19,539); Yolk weight QTL_ID (3349)
ROH_9	5	2126,161	4221,327	2095,166	*PRMT3, ANO5, SLC17A6, GAS2, SVIP, ANO3, FIBIN, LIN7C,* **BDNF,** *KIF18A, METTL15, BBOX1, LGR4,* **SLC5A1,** *mir1775, mir1760*	Body weight (28 days) QTL_ID (95,415, 195,416)
ROH_10	7	6661,093	8199,140	1538,047	*COL18A1, SLC19A1,* **COL6A1, COL6A2**, *FTCD, LSS,* **S100B**, *ITGB2, ADARB1,* **GLS, STAT1, STAT4**	Shank weight QTL_ID (9161); Breast muscle weight QTL_ID (6982)
ROH_11	8	9141,018	11,122,757	1981,739	**PLA2G4A, PTGS2**, *C8H1ORF27, AMY1AP, AMY1A, SLC30A7,* **CRK**, *CDC14A, DBT, SASS6, MFSD14A, SLC35A3, HOXA3, PALMD, mir6561, mir1610*	Thigh meat-to-bone ratio QTL_ID (6721); Abdominal fat percentage QTL_ID (2183); Body weight (day of first egg) QTL_ID (14,465); Tibia bone mineral density QTL_ID (24,365)

*Genes in bold are those included in networks.

gene, which is considered important for the heat stress response in chicken (Lamont et al., 2014). Furthermore, previous findings indicating that the BDNF gene prevents the death of cultured chick retinal ganglion cells, and as reported by Herzog et al. (1994) the tightly controlled expression of the BDNF gene might be important in the coordinated development of the visual system in chicks.

The same ROH_9 includes the leucine rich repeat containing G protein-coupled receptor 4 (**LGR4**) gene that in human is associated with low bone mineral density (Styrkarsdottir et al., 2013).

Within the ROH_8 and ROH_10 map genes that are closely linked to immune system (Table 3). More precisely, within the first region map 2 duplicated genes, the toll-like receptor 2 family member A (**TLR2A**) and the toll-like receptor 2 family member B (**TLR2B**), both orthologs of the single TLR2 of mammals. These genes mediate innate immune responses via recognition of pathogen-associated molecular patterns (**PAMPs**) such as dsRNA of some viruses, or lipopolysaccha-

ride of gram-negative bacteria (Downing et al., 2010). Miyagi et al. (2007) demonstrated that regulation of basal levels of particular STATs including STAT1 and STAT4 and their receptor association, contributes to innate production of the IFN-γ of NK cells. Also, the STAT4 gene encodes a transcription factor involved in the signalling pathways of several cytokines, including interleukin-12 and interleukin-23 (**IL-12**) (Martinez et al., 2008).

A recent work by Fleming et al. (2017) has mapped ROH in several indigenous African and European populations. The authors do not report the list of genomic position of the 4167 consensus ROH mapped in their populations, so it is not possible to compare the overlapping with our results. Nevertheless, the number of ROH mapped is comparable with the one found in this study. Finally, among the 3 ROH that Fleming et al. (2017) are reporting in detail, no one is overlapping those found in this study.

Gene ontology (GO) and pathway analyses for genes included into the Top ROH (Supplementary Table S4)
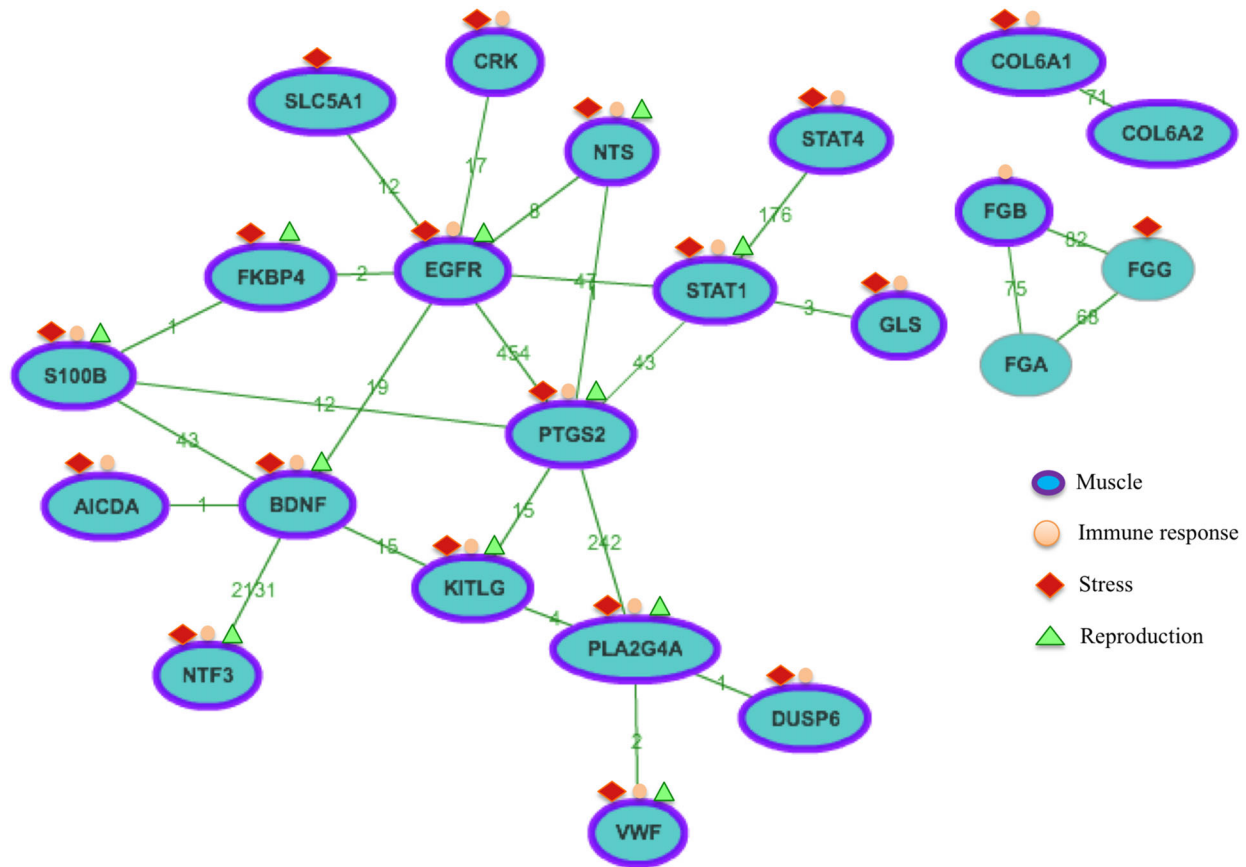
Figure 6. Network of genes included in the Top 1% Mexican chicken ROH.

were performed using GenCLiP2.0, an online server for functional clustering of genes (http://ci.smu.edu.cn/GenCLiP2.0/analysis.php?random=new) accounting for false discovery rate. The GO analysis revealed that they are clustered into a 10 group of genes that were involved in a variety of cellular functions such as sex differentiation, reproductive system development, regulation of response to stress, programmed cell death, tissue and organ development, and so on. KEGG Pathway analysis showed the involvement of several signal pathways, but only 5 were significant after FDR correction (in Supplementary Table S5, as the Q-values).

The Literature Mining Gene Network tool (provided by GenCLiP2.0), that searches for genes linked to keywords based on up-to-date literature profiling, revealed that the 22 genes included within Top 1% ROH have been associated mainly with the keywords "stress", "muscle", "immune response," and "reproduction," as reported in Figure 6. Edges in Network correspond to literature that associate 2 genes with each other, while the relative edge-labels indicate the number of related articles.

To further examine the Top 1% ROH content, quantitative trait loci (**QTL**) that overlapped with these genomic regions were identified by downloading the QTL list from the animal QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/GG/index). We fil-

tered out the QTL that are larger than 5 Mb and only QTL overlapping for at least 50% with the ROH were considered. As reported in Table 3, the most represented QTL are those associated to body conformation and structure (i.e., breast muscle percentage and weight, tibia bone mineral density, body weight, abdominal fat weight, and muscle fiber density and diameter). The same holds for QTLs with a size comprised between 5 and 10 Mb.

The study indicates that the Mexican population is well adapted to the diverse farming conditions that can be found in México. The population clearly appear to be a unique creole chicken population that was not subjected to a specific breeding strategy to improve performance, but shows selection sweeps due to the occurring natural selection for more than 500 years. As the population was maintained mainly as a backyard population, possibly the farmers have reproduced the more productive, more fertile, and more resistant individuals regardless to plumage colour or morphological characteristics. The adaptation of the population to environmental conditions, its resilience to various challenges, makes it very interesting as native genetic resource to be used in family non-intensive farming, in order to raise their income.

In some states of México, where poultry and swine intensive farming is very important, there is no specific financial support for poultry family farming by

the Mexican program "Sin hambre" ("no hunger"—http://sinhambre.gob.mx/). In these states the goal of is to improve sanitary conditions in intensive farming to favor the exportation to the United States and Europe. Nevertheless, the "Sin hambre" project greatly helps local families to farm chickens and increase their revenue, providing a commercial channel for the egg production. This can be easily supported with the local Mexican chicken population adapted to local environmental conditions. A strategy sometime used at present is to cross the local creole population with highly productive breeds, e.g., the Rhode Island Red, or to provide farmers directly with F1 hybrids. This practice nevertheless requires very careful management of the local, well-adapted population to avoid the loss of the genetic variability that guarantee the resilience of the individuals in very harsh environments.

The study provides a genetic knowledge that can be used as a basis for the genetic management of a unique and very large creole population, especially in the view of using it in production of hybrids to increase the productivity and economic revenue of family farming agriculture, which is a large reality in México.

SUPPLEMENTARY DATA

Supplementary data are available at *PSCIEN* online.

Figure S1. Heat map of IBD PI estimates values

Table S1. Mean of normalized Z values of body measures used in the cluster analysis

Table S2. Count of individuals according to classes of ancestor's composition and morphological clusterization

Table S3. Descriptive statistics of identified ROH according of length values on different chromosomes

Table S4. Go analysis results for genes included in the Networks

Table S5. Pathway analysis results for genes included in the Networks.

REFERENCES

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Al-Qamashoui, B., H. Simianer, I. Kadim, and S. Weigend. 2014. Assessment of genetic diversity and conservation priority of Omani local chickens using microsatellite markers. Trop. Anim. Health Prod. 46:747–752.

Brown, W. R. A., S. J. Hubbard, C. Tickle, and S. A. Wilson. 2003. The chicken as a model for large-scale analysis of vertebrate gene function. Nat. Rev. Genet. 4:87–98.

Caratão, N., C. S. Cortesão, P. H. Reis, R. F. Freitas, C. M. Jacob, A. C. Pastorino, M. Carneiro-Sampaio, and V. M. Barreto. 2013. A novel activation-induced cytidine deaminase (AID) mutation in Brazilian patients with hyper-IgM type 2 syndrome. Clin. Immunol. 148:279–286.

Cavalchini, L. G., S. P. Marelli, M. G. Strillacci, M. C. Cozzi, M. Polli, and M. Longeri. 2007. Heterozygosity analysis of Bionda Piemontese and Bianca di Saluzzo chicken breeds by microsatellites markers: a preliminary study. Ital. J. Anim. Sci. 6:63–65.

Ceccobelli, S., P. Di Lorenzo, H. Lancioni, L. V. Monteagudo Ibáñez, M. Tejedor, C. Castellini, V. Landi, A. Martínez Martínez, J. D. Delgado Bermejo, J. L. Vega Pla, J. M. Leon Jurado, M. García, G. Attard, A. Grimal, S. Stojanovic, K. Kume, F. Panella, S. Weigend, and E. Lasagna. 2015. Genetic diversity and phylogeographic structure of sixteen Mediterranean chicken breeds assessed with microsatellites and mitochondrial DNA. Livest. Sci. 175:27–36.

Curik, I., M. Ferenčaković, and J. Sölkner. 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 166:26–34.

Downing, T., A. T. Lloyd, C. O'Farrelly, and D. G. Bradley. 2010. The differential evolutionary dynamics of avian cytokine and TLR gene classes. J. Immunol. 184:6993–7000.

Excoffier, L. 2007. Analysis of Population Subdivision. In: David J. Balding, Martin Bishop, and Chris Cannings (eds.) Handbook of Statistical Genetics 980–1020. Chichester: John Wiley & Sons.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Fan, H., Y. Wu, X. Qi, J. Zhang, J. Li, X. Gao, L. Zhang, J Li, and H. Gao H. 2014. Genome-wide detection of selective signatures in Simmental cattle. J. Appl. Genet. 55:343–351.

FAO, 2012. Phenotypic characterization of animal genetic resources. FAO Animal Production and Health Guidelines No. 11. Rome (http://www.fao.org/docrep/015/i2686e/i2686e00.pdf) (accessed 01.02.2017).

Fleming, D. S., J. E. Koltes, A. D. Markey, C. J Schmidt, C. Ashwell, M. F. Rothschild, M. E. Persia, J. M. Reecy, and S. J. Lamont. 2016. Genomic analysis of Ugandan and Rwandan chicken ecotypes using a 600 k genotyping array. BMC Genomics. 17:407.

Fleming, D. S., S. Weigend, H. Simianer, A. Weigend A, M. F. Rothschild, C. J. Schmidt, C. Ashwell, M. E. Persia, J. M. Reecy, and S. J. Lamont. 2017. Genomic comparison of indigenous African and northern European chickens reveals putative mechanisms of stress tolerance related to environmental selection pressure. G3 (Bethesda). 7(5):1525–1537.

Gibson, J., N. Morton, and A. Collins. 2006. Extended tracts of homozygosity in outbred human populations. Hum. Mol. Genet. 15:789–795.

Gorla, E., M. C. Cozzi, S. I. Roman-Ponce, F. J. Ruiz-Lopez, V. E. Vega-Murillo, S. Cerolini, A. Bagnato, and M. G. Strillacci. 2017. Genomic variability in Mexican chicken population using Copy Number Variants. BMC Genetics. 18:61.

Hall, S. J., and D. G. Bradley. 1995. Conserving livestock breed biodiversity. Trends. Ecol. Evol. 10:267–270

Herzog, K. H., K. Bailey, and Y. A. Barde. 1994. Expression of the BDNF gene in the developing visual system of the chick. Development. 120:1643–1649.

Johansson, A. M., and R. M. Nelson. 2015. Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds. Front. Genet. 6:44.

Keim, C., D. Kazadi, G. Rothschild, and U. Basu. 2013. Regulation of AID, the B-cell genome mutator. Genes Dev. 27:1–17.

Lamont, S. J., D. J. Coble, A. Bjorkquist, M. F. Rothschild, M. Persia, C. Ashwell, and C. Schmidt. 2014. Genomics of heat stress in chickens. Proceedings, 10th World Congress of Genetics Applied to Livestock Production. Vancouver, BC, Canada, August 17–22, 2014.

Mahammi, F. Z., S. B. Gaouar, D. Laloë, R. Faugeras, N. Tabet-Aoul, X. Rognon, M. Tixier-Boichard, and N. Saidi-Mehtar. 2016. A molecular analysis of the patterns of genetic diversity in local chickens from western Algeria in comparison with commercial lines and wild jungle fowls. J. Anim. Breed. Genet. 133:59–70.

Martínez, A., J. Varadé, A. Márquez, M. C. Cénit, L. Espino, N. Perdigones, J. L. Santiago, M. Fernández-Arquero, H. de la Calle, R. Arroyo, J. L. Mendoza, B. Fernández-Gutiérrez, E. G. de la Concha, and E. Urcelay. 2008. Association of the STAT4 gene with increased susceptibility for some immune-mediated diseases. Arthritis Rheum. 58:2598–2602.

Mastrangelo, S., M. Tolone, R. Di Gerlando, L. Fontanesi, M. T. Sardina, and B. Portolano. 2016. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. Animal. 10:746–754.

Metzger, J., M. Karwath, R. Tonda, S. Beltran, L. Águeda, M. Gut, I. G. Gut, and O. Distl. 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics. 16:764.

Miyagi, T., M. P. Gil, X. Wang, J. Louten, W. M. Chu, and C. A. Biron. 2007. High basal STAT4 balanced by STAT1 induction to control type 1 interferon effects in natural killer cells. J. Exp. Med. 204:2383–2396.

Mutryn, M. F., E. M. Brannick, W. Fu, W. R. Lee, and B. Abasht. 2015. Characterization of a novel chicken muscle disorder through differential gene expression and pathway analysis using RNA-sequencing. BMC Genomics. 16:399.

Nimmakayala, P., A. Levi, L. Abburi, V. L. Abburi, Y. R. Tomason, T. Saminathan, V. G. Vajja, S. Malkaram, R. Reddy, T. C. Wehner, S. E. Mitchell, and U. K. Reddy. 2014. Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. BMC Genomics. 15:767.

Pemberton, T. J., D. Absher, M. W. Feldman, R. M. Myers, N. A. Rosenberg, and J. Z. Li. 2012. Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. 91:275–292.

Qanbari, S., and H. Simianer. 2014. Mapping signatures of positive selection in the genome of livestock. Livest. Sci. 166:133–143.

Segura-Correa, J. C., L. Sarmiento-Franco, J. G. Magaña-Monforte, and R. Santos-Ricalde. 2004. Productive performance of Creole chickens and their crosses raised under semi- intensive management conditions in Yucatan, Mexico, Br. Poult. Sci. 45:342–345.

Segura-Correa, J. C., A. Juarez-Caratachea, L. Sarmiento-Franco, and R. Santos-Ricalde. 2005. Growth of Creole chickens raised under tropical conditions of Mexico. Trop. Anim. Health. Prod. 37:327–332.

SAS Institute. 2013. SAS 9.4 language reference concepts. Cary, NC: SAS Institute.

Styrkarsdottir, U., G. Thorleifsson, P. Sulem, D. F. Gudbjartsson, A. Sigurdsson, A. Jonasdottir, A. Oddsson, A. Helgason, O. T. Magnusson, G. Bragi Walters, M. L. Frigge, H. T. Helgadottir, H. Johannsdottir, K. Bergsteinsdottir, M. H. Ogmundsdottir, J. R. Center, T. V. Nguyen, J. A. Eisman, C. Christiansen, E. Steingrimsson, J. G. Jonasson, L. Tryggvadottir, G. I. Eyjolfsson, A. Theodors, T. Jonsson, T. Ingvarsson, I. Olafsson, T. Rafnar, A. Kong, G. Sigurdsson, G. Masson, U. Thorsteinsdottir, and K. Stefansson. 2013. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. Nature. 497:517–520.

Strillacci, M. G., S. P. Marelli, M. C. Cozzi, E. Colombo, M. Polli, M. Gualtieri, A. Cristalli, P. Pignattelli, M. Longeri, and L. Guidobono Cavalchini. 2009. Italian autochthonous chicken breeds conservation: evaluation of biodiversity in Valdarnese Bianca breed (Gallus gallus domesticus). Avian Biol. Res. 2:229–233.

Strillacci, M. G., M. C. Cozzi, F. Schiavini, E. Gorla, S. Cerolini, S. I. Román-Ponce, F. J. Ruiz López, and A. Bagnato. 2017. Genomic and genetic variability of six Italian chicken populations using SNP and CNV as markers. Animal. 11:737–745.

Sulem, P., D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar, K. P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Pálsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K. R. Benediktsdottir, K. K. Aben, L. A. Kiemeney, J. H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, and K. Stefansson. 2007. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nature Genet. 39:1443–1452.

Xu, S., S. Guputa, and L. Jin. 2010. PEAS V1.0: A package for elementary analysis of SNP data. Mol. Ecol. Resour. 10:1085–1088.