



OPEN ACCESS

EDITED BY

Isidore Houaga,
University of Edinburgh, United Kingdom

REVIEWED BY

Getinet Mekuriaw Tarekegn,
Scotland's Rural College, United Kingdom
Luc Hippolyte Dossa,
University of Abomey-Calavi, Benin

*CORRESPONDENCE

Carlos I. Arbizu,
✉ carbizu@inia.gob.pe

SPECIALTY SECTION

This article was submitted
to Livestock Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 19 October 2022

ACCEPTED 27 February 2023

PUBLISHED 10 March 2023

CITATION

Corredor F-A, Figueroa D, Estrada R,
Salazar W, Quilcate C, Vásquez HV,
Gonzales J, Maicelo JL, Medina P and
Arbizu CI (2023), Genetic diversity and
population structure of a Peruvian cattle
herd using SNP data.
Front. Genet. 14:1073843.
doi: 10.3389/fgene.2023.1073843

COPYRIGHT

© 2023 Corredor, Figueroa, Estrada,
Salazar, Quilcate, Vásquez, Gonzales,
Maicelo, Medina and Arbizu. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Genetic diversity and population structure of a Peruvian cattle herd using SNP data

Flor-Anita Corredor¹, Deyanira Figueroa¹, Richard Estrada¹,
Wilian Salazar¹, Carlos Quilcate¹, Héctor V. Vásquez²,
Jhony Gonzales³, Jorge L. Maicelo², Percy Medina¹ and
Carlos I. Arbizu^{1*}

¹Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Lima, Peru, ²Facultad de Ingeniería Zootecnista, Agronegocios y Biotecnología, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Amazonas, Peru, ³Laboratorio de Biología Molecular, Universidad Nacional de Frontera, Piura, Peru

New-generation sequencing technologies, among them SNP chips for massive genotyping, are useful for the effective management of genetic resources. To date, molecular studies in Peruvian cattle are still scarce. For the first time, the genetic diversity and population structure of a reproductive nucleus cattle herd of four commercial breeds from a Peruvian institution were determined. This nucleus comprises Brahman ($N = 9$), Braunvieh ($N = 9$), Gyr ($N = 5$), and Simmental ($N = 15$) breeds. Additionally, samples from a locally adapted creole cattle, the Arequipa Fighting Bull (AFB, $N = 9$), were incorporated. Female individuals were genotyped with the GGFBovine100K and males with the BovineHD. Quality control, and the proportion of polymorphic SNPs, minor allele frequency, expected heterozygosity, observed heterozygosity, and inbreeding coefficient were estimated for the five breeds. Admixture, principal component analysis (PCA), and discriminant analysis of principal components (DAPC) were performed. Also, a dendrogram was constructed using the Neighbor-Joining clustering algorithm. The genetic diversity indices in all breeds showed a high proportion of polymorphic SNPs, varying from 51.42% in Gyr to 97.58% in AFB. Also, AFB showed the highest expected heterozygosity estimate (0.41 ± 0.01), while Brahman the lowest (0.33 ± 0.01). Besides, Braunvieh possessed the highest observed heterozygosity (0.43 ± 0.01), while Brahman the lowest (0.37 ± 0.02), indicating that Brahman was less diverse. According to the molecular variance analysis, 75.71% of the variance occurs within individuals, whereas 24.29% occurs among populations. The pairwise genetic differentiation estimates (F_{ST}) between breeds showed values that ranged from 0.08 (Braunvieh vs. AFB) to 0.37 (Brahman vs. Braunvieh). Similarly, pairwise Reynold's distance ranged from 0.09 (Braunvieh vs. AFB) to 0.46 (Brahman vs. Braunvieh). The dendrogram, similar to the PCA, identified two groups, showing a clear separation between *Bos indicus* (Brahman and Gyr) and *B. taurus* breeds (Braunvieh, Simmental, and AFB). Simmental and Braunvieh grouped closely with the AFB cattle. Similar results were obtained for the population structure analysis with $K = 2$. The results from this study would contribute to the appropriate management, avoiding loss of genetic variability in these breeds and for future improvements in this nucleus. Additional work is needed to speed up the breeding process in the Peruvian cattle system.

KEYWORDS

cattle breeds, genotypes, diversity, genomics, NGS

1 Introduction

Livestock production around the world is a large sector with an important contribution of 40% and 20% to agriculture production in developed and developing countries, respectively (Herrero et al., 2013; Baltenweck et al., 2020; FAO, 2022). A large part of Peru's livestock economy revolves around cattle production (León-Velarde and Quiroz, 2004). According to the latest Peruvian National Agricultural Census (Instituto Nacional de Estadística e Informática, 2012), Peruvian creole cattle (PCC) is the most predominant cattle population (64.03%). PCC is prevalent in the Andean sector of the country, where it has been adapted to the highlands climate conditions (Quispe, 2016; Delgado et al., 2019). However, in comparison with exotic breeds, PCC achieves smaller body weights and milk production records (Espinoza and Urviola, 2005; Dipas Vargas, 2015; Ruiz et al., 2021).

Due to the low productivity, small farmers breeding strategy is to crossbred PCC with other specialized breeds in order to take advantage of the heterosis effect (Seré et al., 1996; W. et al., 2019). Nowadays, there are a high availability of bovine breeds that can be used to improve milk, meat or double purpose production (Thibier and Wagner, 2002; Mebratu et al., 2020). However, Peruvian initiatives are lacking the understanding of the genetics behind. Genetic diversity knowledge is essential for the effective management of genetic resources (Groeneveld et al., 2010; Hoban et al., 2013). In recent years the availability of genotyping technology has become affordable in livestock allowing to increase genetic studies (Mukhopadhyay et al., 2020). As a result, SNP markers are becoming increasingly common for diversity analysis and population structure studies (Morin et al., 2009; Haas and Payseur, 2011). SNP markers have the advantage of being abundant in the genome, as well as the ability to be automated through high-through genotyping panels (Beuzen et al., 2000; Vignal et al., 2002).

In developing countries, nucleus breeding systems represent a good strategy for animal genetic improvement for ruminants. Concentrating nucleus cattle in one or a few herds to disseminate genetic material to other populations is helpful (Kiwuwa, 1992; Schrooten and van Arendonk, 1992). In 1993, a Peruvian government herd composed of Brahman, Braunvieh, Gyr and Simmental breeds was established with the aim to develop reproductive technology research, such as artificial insemination and embryo transfer. Currently, the herd is distributing semen straws and embryos to producers' associations in order to disseminate specialized cattle breed genetics. This herd is been called a genetic nucleus herd, however, there is scarcity of data available in pedigrees and production records. Therefore, this study aims to provide understanding of the genetic diversity among the breeds on this herd, and its population structure, including a PCC group on the study. We expect to genomic characterize the nucleus using SNP markers, by obtaining genetic diversity and population structure parameters.

2 Materials and methods

2.1 Animal sampling and DNA extraction

A total of 63 blood samples were collected from four commercial breeds of taurus (Braunvieh and Simmental) and indicus cattle

(Brahman and Gyr). According to their pedigree, up to grandfathers, genetic origins for Brahman and Gyr were predominantly from Brazil; for Braunvieh, Switzerland and Colombia; while for Simmental was Germany (Supplementary Table S1). Blood sampling was performed from a government herd, the Donoso Agricultural Research Station (EEA Donoso in Spanish) located in Huaral, Lima (128 masl; 11°31'18" S and 77°14'06" W). Pedigree was checked to avoid sampling from related individuals, animals were not siblings or had a parental relationship. Blood samples were collected from related individuals, animals were not siblings or have a parental relationship from the epidural vein using a vacutainer containing EDTA as an anticoagulant and were immediately transferred to the laboratory for DNA extraction. Additionally, we got access to 12 hair samples that were collected from the tail of individuals that were considered as "Arequipa fighting bull" (AFB), which are bovines from Arequipa region (2,335 masl; 15°29'58" S and 72°21'36" W). Most of these individuals were selected as they possessed most of the morphological characteristics of a PCC as identified by their owners, where its body is unbalanced with the topline being higher on the front and becoming smaller toward the rear. For the PCC, the hooks to pin are lower-levelhipped when compared to other breeds of cattle, dairy or beef. The length of the body is shorter, as is the topline. Colors of hair have multiple variations. The diversity of colors ranges from a total color cover to mixed ones and spotting ones.

We extracted genomic DNA from whole blood and hair samples with the Wizard Genomic DNA Purification Kit (Fitchburg, WI, United States) following the manufacturer's instructions. The quality and quantity of genomic DNA were assessed using agarose gel electrophoresis and a Nanodrop spectrophotometer (Model ND 2000, Thermo Fisher Scientific, Wilmington, DE, United States) prior to genotyping. In addition, 40 genotypes from reference breeds were included in the analyses. The reference breeds were sourced from Decker et al. (2014) and the world reference dataset in Web-Interfaced Next-Generation Database (WIDDE) database (Sempéré et al., 2015). Breeds included were Brahman, Braunvieh, Gyr, and Simmental.

2.2 SNP genotyping and quality control

DNA samples were genotyped using Illumina Bovine HD Genotyping BeadChip and Illumina GGP Bovine 100K BeadChip with the help of the commercial genotyping service provider (Neogen, Geneseek, NL, United States). Female individuals were genotyped with the GGPBovine100K and males with the BovineHD. The Bovine HD and 100K chips possess 777,962 and 95,256 SNPs, respectively, uniformly spanning over the entire bovine genome. A total of 87,669 common markers between both SNP panels were used for the following analysis. From the total of 71 animals sampled, we discarded the ones with a genotype call rate minor to 85% (Purfield et al., 2016). A total of 18 samples were discarded before starting the SNP quality control. We started the SNP quality control with 53 animals and 47 remained for the following analysis after quality control.

SNPs quality control was performed using the PLINK v1.9 program (Purcell et al., 2007). SNPs assigned to sex

TABLE 1 Genetic diversity, showing the name of the breed, sample size (N), the proportion of polymorphic SNPs (Pn), minor allele frequency (MAF), expected heterozygosity (He), observed heterozygosity (Ho), and inbreeding coefficient (F_{IS}).

Population	N	Pn (%)	MAF	He	Ho	F_{IS}
Brahman	9	62.18	0.15 ± 0.16	0.33 ± 0.01	0.37 ± 0.02	-0.07
Braunvieh	9	92.06	0.27 ± 0.15	0.38 ± 0.01	0.43 ± 0.01	-0.07
Simmental	15	95.37	0.28 ± 0.14	0.38 ± 0.01	0.41 ± 0.02	-0.05
Gyr	5	51.42	0.13 ± 0.16	0.35 ± 0.01	0.41 ± 0.02	-0.05
AFB	9	97.58	0.32 ± 0.13	0.41 ± 0.01	0.42 ± 0.04	0.03

chromosomes and those lacking genomic locations were excluded from the analysis. SNPs with missing genotypes in more than 10% of individuals, missing rate per SNP of 10%, and minor allele frequency (MAF) lower than 0.05 were excluded. However, SNP filtering based on the Hardy–Weinberg equilibrium was not performed since we expected Hardy–Weinberg deviations in the studied populations due to their small and possibly sub-structured population and genetic drift (Chen et al., 2017). We used 80,178 autosomal SNPs that remained after applying filtering criteria to assess genetic diversity. Additionally, linkage disequilibrium pruning, using the parameter indep (50 5 2), was performed before the population structure analysis. A total of 16,345 SNPs were obtained after pruning for LD.

2.3 Genetic diversity

To assess the genetic diversity within the studied population we used different genetic diversity parameters. The proportion of polymorphic SNPs (Pn), MAF, expected heterozygosity (He), observed heterozygosity (Ho), and inbreeding coefficient (F_{IS}) were estimated using R package *dartR* (Gruber et al., 2018). The distribution of MAF was grouped into five different categories based on the frequency of rare alleles ($0 < \text{MAF} \leq 0.1$), intermediate alleles ($0.1 < \text{MAF} \leq 0.2$, $0.2 < \text{MAF} \leq 0.3$, and $0.3 < \text{MAF} \leq 0.4$), and common alleles ($0.4 < \text{MAF} \leq 0.5$).

2.4 Population structure

Different approaches were employed to investigate the genetic structure among the cattle populations of the EEA Donoso herd, and assess their relationships with the AFB cattle. First, an analysis of molecular variance (AMOVA) was performed with ARLEQUIN v.3.5.2 software (Excoffier and Lischer, 2010), with the locus by locus option and 1,000 permutations. PGDSpider v.2.1.1.5 software (Lischer and Excoffier, 2012) was used to convert files between PLINK and Arlequin formats. We used ARLEQUIN to assess the divergence among breeds. Genetic differentiation among breed (F_{ST}) fixation indices were calculated using 20,000 permutations and a significance level of 0.05. Also, Reynold's distance was performed. Second, a principal component analysis (PCA), and a discriminant analysis of principal components (DAPC) were performed with PLINK. The *factorextra* (Kassambara and Mundt, 2017) and *adeget* (Jombart and Collins, 2017) R packages were used to generate eigenvectors and eigenvalues, and the outputs were visualized using the package *ggplot2* (Gómez-Rubio, 2017). For PCA and DAPC, animals from the referenced population were included in the analyses. Third, an assessment of population genetic structure was performed using the default settings of ADMIXTURE v.1.3 software (Alexander et al., 2009). The most appropriate K value was selected after considering 10-fold cross-validations whereby the best K exhibits low cross validation error compared to other K values (Alexander and Lange, 2011). Finally, a Neighbor-Joining tree was constructed using *vcfR* (Knaus and Grünwald, 2017), *pegas* (Paradis, 2010), and *ape* (Paradis and Schliep, 2019) packages in R. Additionally, 1,000 bootstrap replicates were conducted.

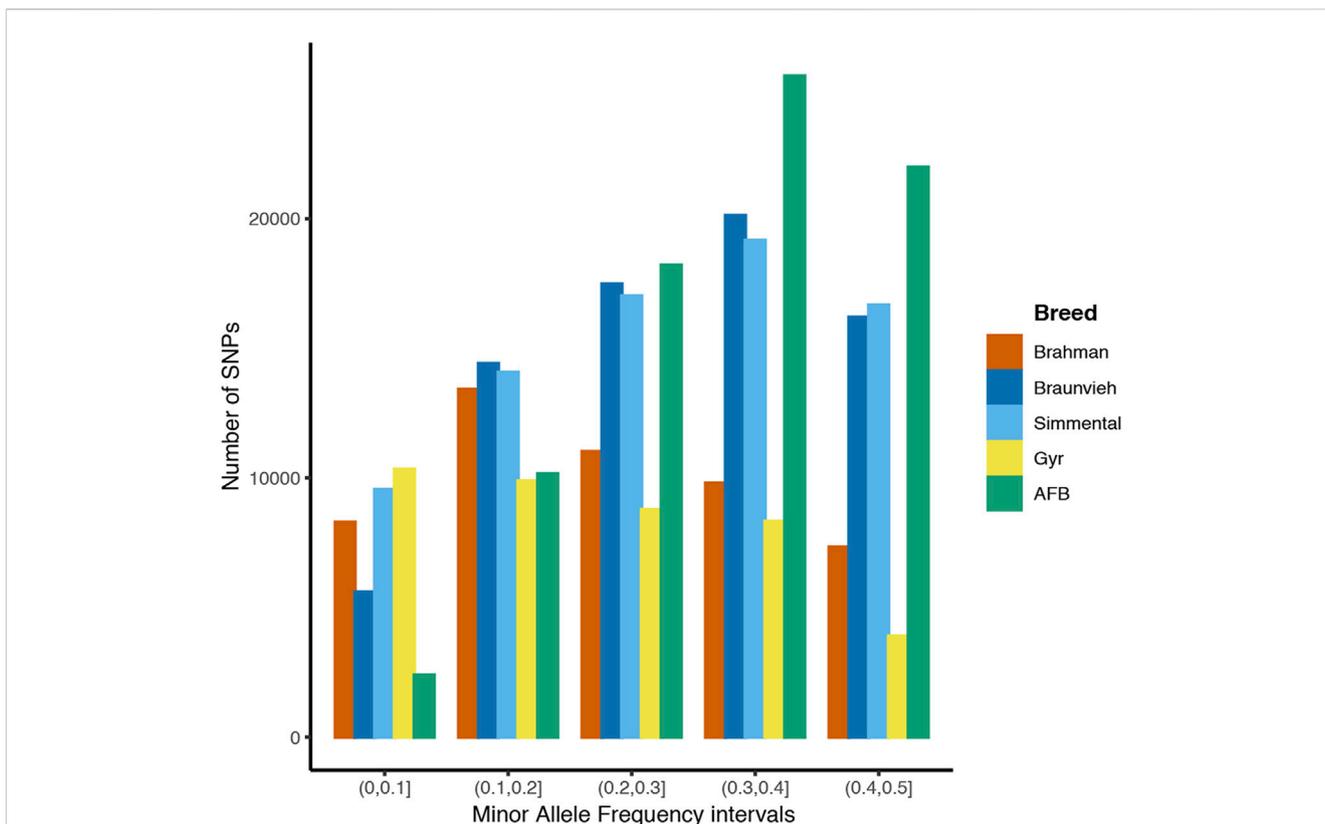


FIGURE 1
Distribution of minor allele frequency for each cattle breed.

TABLE 2 Analysis of molecular variance among five cattle breeds.

Source of variation	Degree of freedom	Sums of squares	Variance component	% of variations
Among population	4	365574.76	4,301.64	24.29
Within individuals	89	1184932.66	13411.11	75.71
Total	93	1550507.41	17712.74	

3 Results

3.1 Genetic diversity analysis

The results of the genetic diversity parameters calculated for the different cattle breed groups genotyped are summarized in Table 1. Most of the breeds show a high Pn, varying from 51.42% in Gyr to 97.58% in AFB. The highest mean MAF value was observed in AFB (0.32 ± 0.13), and the lowest value was observed in Gyr (0.13 ± 0.16) with a mean value of 0.23 across populations. The He ranged from 0.33 (Brahman) to 0.41 (AFB). The highest observed heterozygosity was observed in Braunvieh (0.43 ± 0.01), while the lowest was in Brahman (0.37 ± 0.02). The Ho was greater than the He and the inbreeding coefficient was negative for the breeds, except for AFB.

Minor allele frequency distribution for different categories is shown in Figure 1. Among the five cattle breeds, AFB (25,504) and Gyr (3,903) showed the highest and the lowest count of SNPs when MAF greater than or equal to 0.3.

TABLE 3 Estimates of the pairwise genetic differentiation statistic (F_{ST} statistics; below the diagonal) and the Reynold’s genetic distance (above the diagonal) among five cattle breeds.

Breed	Brahman	Braunvieh	Gyr	AFB	Simmental
Brahman		0.46	0.14	0.38	0.43
Braunvieh	0.37		0.44	0.09	0.13
Gyr	0.13	0.36		0.36	0.42
AFB	0.31	0.08	0.30		0.09
Simmental	0.35	0.12	0.34	0.09	

The Gyr breed had a higher count of SNPs in the lowest MAF interval (MAF≤0.1) compared to the counts of SNPs in the higher MAF intervals. The count of SNPs for the Braunvieh, AFB, and Simmental cattle breeds was shown to be higher as the MAF interval

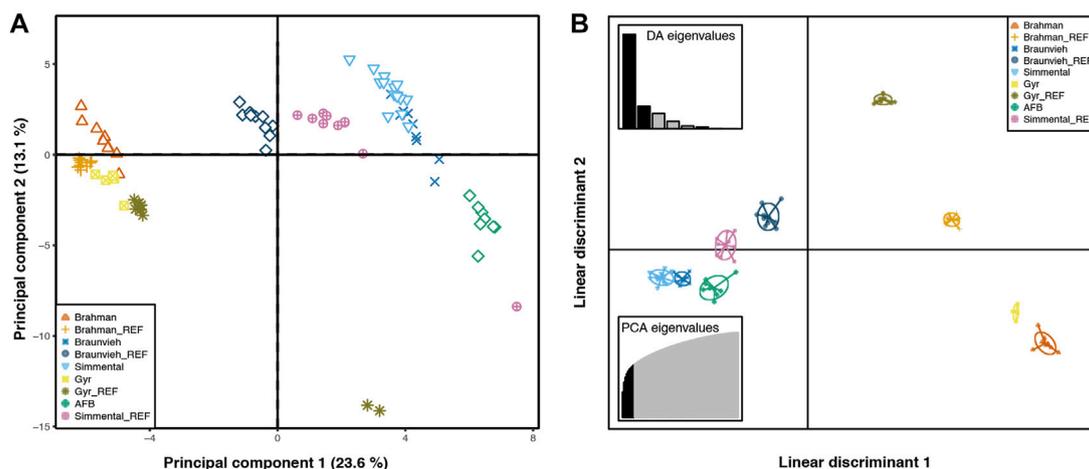


FIGURE 2 Principal component analysis (PCA) and Discriminant analysis of principal components (DAPC) plots. Samples belong to a reproductive cattle herd that comprises with Brahman ($N = 9$), Braunvieh ($N = 9$), Gyr ($N = 5$), and Simmental ($N = 15$) breeds; a locally adapted creole cattle, the Arequipa Fighting Bull (AFB, $N = 9$); and 40 genotype samples from reference breeds included in the analyses (subscript with _REF). Symbols and colors indicate breed affiliation, each symbol represents an individual. **(A)** For PCA plot, the x- and y-axes are indicated by the first and second components, respectively, and the values in parentheses show the percentages of total variance explained. **(B)** For DAPC plot, the scatterplot shows only the first two linear discriminants of the analysis.

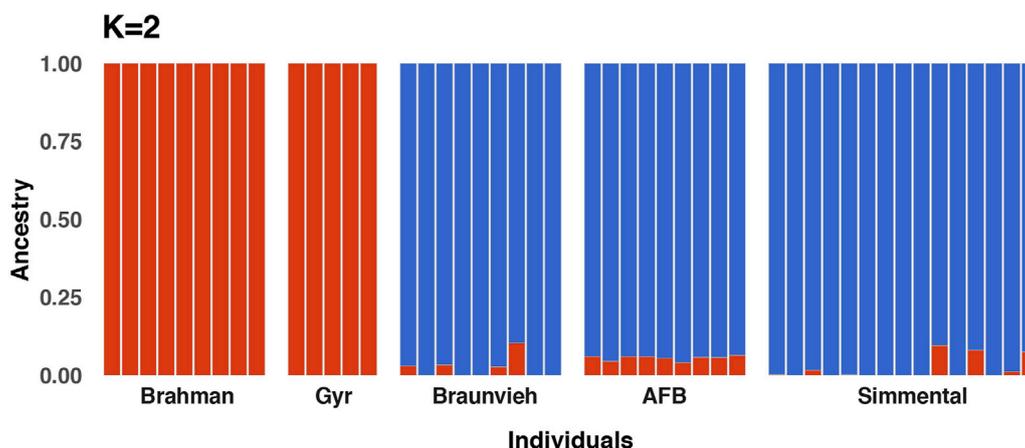


FIGURE 3 Population structure using 16,345 SNPs for five cattle breeds consisting of 47 individuals. Admixture analysis showing the proportions of ancestral populations for $K = 2$, each vertical bar exemplifies an individual.

increased. For the Brahman and Gyr breeds the count of SNPs decreased.

3.2 Population structure

The AMOVA results (Table 2) showed that the most important part of the genetic variation (75.71%) was observed within the cattle breeds and variability among the cattle breeds was 24.29%. Also, pairwise F_{ST} and Reynold’s distance among all populations were estimated (Table 3). The pairwise F_{ST} estimates among breeds ranged from 0.08 (AFB-Braunviehpair) to 0.37

(Braunvieh-Brahman pair). The pairs Braunvieh-Brahman and Braunvieh-Gyr showed high pairwise F_{ST} values, with 0.37 and 0.36, respectively. Furthermore, the pairs Braunvieh-AFB, and AFB-Simmental showed the lowest pairwise F_{ST} values, with 0.08 and 0.09, respectively. The pairwise Reynold’s distance showed a pattern similar to the one obtained with the F_{ST} statistics, with values ranging from 0.09 (AFB-Braunvieh pair) to 0.46 (Braunvieh-Brahman pair). The pairs Brahman-Braunvieh and Braunvieh-Gyr showed high pairwise Reynold’s distance values, with 0.46 and 0.44, respectively. Furthermore, the pairs AFB-Braunvieh, and AFB-Simmental showed the lowest pairwise F_{ST} values, with 0.09, respectively each.

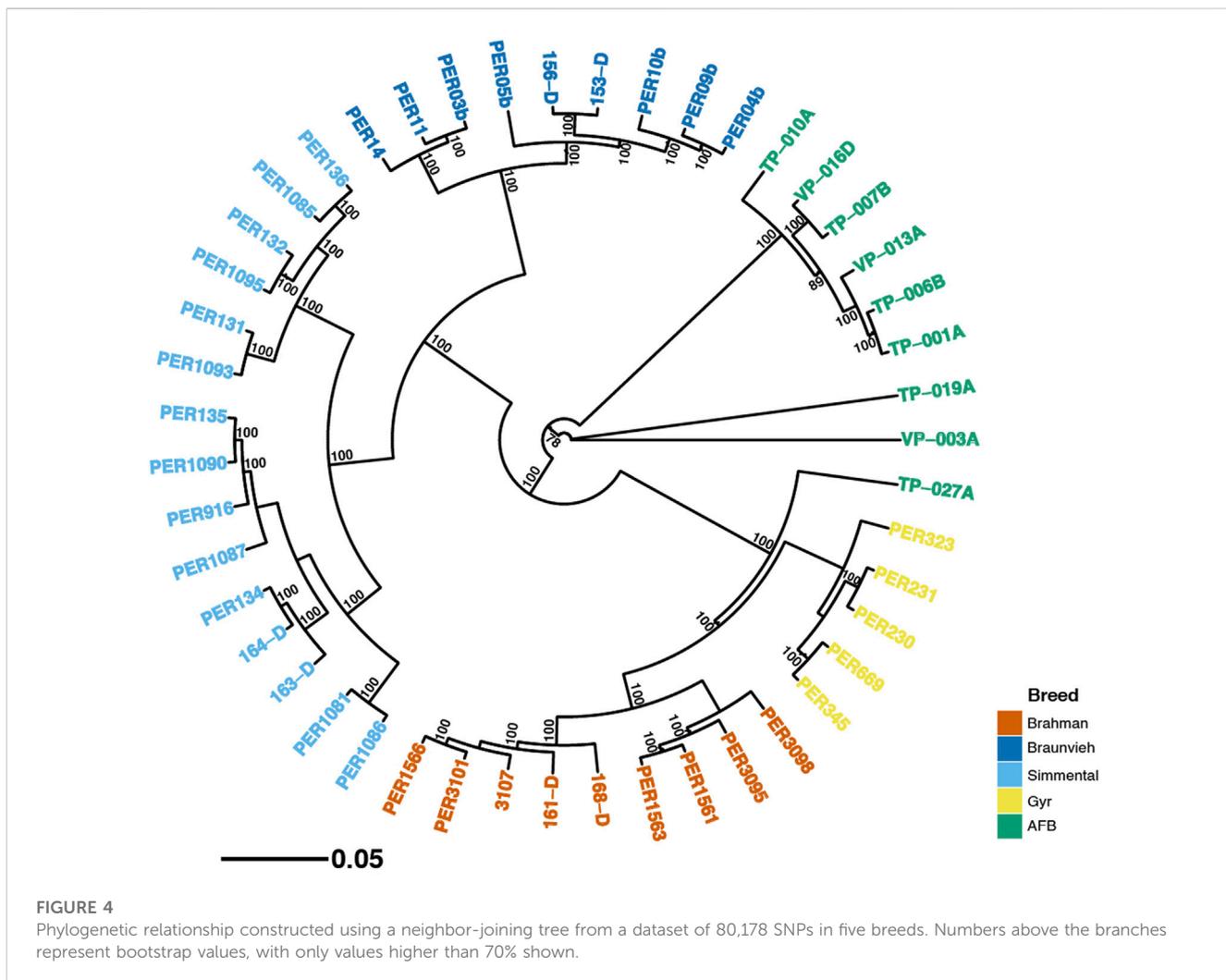


Figure 2 presents the result of PCA and DAPC analysis performed to visualize individual relationships among populations. Individuals were grouped according to their breed origins. The first and second component accounted for a total of 23.60% and 13.10%, respectively. PCA and DAPC showed a low differentiation among the AFB, Braunvieh, and Simmental populations, while the Brahman and Gyr herds are clearly separated from the other three populations. In the PCA a substructure was observed corresponding to samples from Gyr, from the reference populations. The populations included in this study come from different selection environments. Brahman and Gyr individuals have been selected for tropical climates. The Simmental and Braunvieh groups have been selected in temperate environments, while AFB individuals have been mainly selected under artificial selection pressure.

A graphic representation of cluster structure analysis is depicted in Figure 3. Based on the ΔK value, $K = 2$ was the most optimal number for the inferred genetic structure of the populations (Supplementary Figure S1). At $K = 2$ a considerable source of variation among cattle breeds was perceptible. Cluster 1 comprised of the Brahman and Gyr breed groups ($N = 14$ genotypes), whereas cluster 2 consisted of the Braunvieh Simmental, and AFB cattle groups ($N = 33$ genotypes). The

Brahman and Gyr populations displayed a separated cluster, whereas the Braunvieh, and AFB, and Simmental populations presented similar genetic construction.

A neighbor-joining tree was constructed from SNPs (Figure 4), displaying bootstrap support greater than 70%. The first group is composed of fifteen Simmental individuals with 100% bootstrap support. The second group is composed of nine Braunvieh individuals with 100% bootstrap support. Nine individuals AFB composed the third group. In concordance with the principal coordinate analysis, these groups are together. The fourth group is composed of nine Braunvieh individuals with 100% bootstrap support. Five individuals Gyr composed the fifth group. These groups are together with 100% bootstrap support, also in agreement with principal coordinate analysis. However, an individual AFB (TP-027A) was integrated into this group of Brahman and Gyr with 100% bootstrap support.

4 Discussion

Investigating genetic diversity parameters of populations is critical for developing future breeding objectives (Notter, 1999).

Here for the first time, we examined population genetic structure of a group of animals that are been used as a genetic nucleus in Peru, as well as a group of PCC (AFB). Our results of Pn, MAF, He, Ho showed that the populations have a moderate genomic diversity. In this study, all the average values of MAF recorded in taurine cattle (i.e., Braunvieh, Simmental, and AFB) were higher than those recorded in zebu cattle (i.e., Brahman and Gyr). This result might be due to the low representation of zebuine cattle breeds in the SNP genotyping array used (Chagunda et al., 2018). Most of the bovine SNP panels available have been developed of the sequences of individuals belonging to European bovine breeds (European Cattle Genetic Diversity Consortium et al., 2006; The Bovine Hapmap Consortium et al., 2009). This might explain why the observed polymorphism of our SNP data set was higher in the Braunvieh, Simmental, and AFB breeds.

Previous studies have shown that breeding practices have a great effect on reducing genetic diversity, leading to a lower level of genetic diversity in selected germplasm compared with wild varieties (Tisdell, 2003; Zenger et al., 2007). Interestingly, our genetic diversity analysis with the four specialized breeds from the nucleus herd seems to agree. We observed a significantly higher genetic diversity level in the AFB group breed than in the specialized breeds, which is similar from those reported in previous studies (Giovambattista et al., 2001; Egito et al., 2007; Edea et al., 2015). In the case of AFB, it showed the highest levels of He and one of the highest for Ho. For these cattle population, there is a marked effect due to mating control by breeders, which can certainly play an important role (Hidalgo et al., 2015; Delgado et al., 2019). Creole breeds are primarily used in the Peruvian livestock systems to establish crosses with other species of *B. taurus*, particularly Brown Swiss and Simmental in high Andean areas (Primo, 1992; Quispe, 2016). Considering AFB, the mean value of Ho (0.42) obtained in this study is lower than that (0.77) reported by Martínez et al. (2015) in Costa Rica, (0.75) Lirón et al. (2006) in Argentina and Bolivian Creole Breeds, (0.68) Egito et al. (2007), in Brazil, (0.67) Ginja et al. (2010) in Portuguese Native Cattle, (0.70) Acosta et al. (2012) in Cuban cattle breeds. However, most of these studies are also in creole cattle from Latin America where the values greatly differ from ours. One explanation for these differences is that our study was based on SNP markers, whereas the other studies used microsatellite markers. As population genetic statistics can easily be applied to SNPs because they are often bi-allelic, however, a greater number of polymorphic loci may be required to match the power of multi-allelic SSR loci (Guichoux et al., 2011; Laoun et al., 2020). Also, the reduced Ho of the AFB may be explained on the fact that these individuals, compared to other local breeds, go through a process of strict artificial selection as growers always look for fighting traits. AFB are always part of the traditional bullfight activity of Arequipa. It should be noted that the Brahman and Gyr breed presented the lowest levels of He and Ho. This lower level of heterozygotes is generally interpreted as a deviation from random mating (Zeng et al., 2013; Lamkey and Edwards, 2015).

Regarding the content of Pn, a study in six breeds including Simmental, determined an average proportion of polymorphic SNPs of 79% (Dadi et al., 2012), while in this study was 79.72%. F_{IS} presented an average value of -0.04 , which ranged from 0.03 (AFB) to -0.07 (Brahman and Braunvieh). So, this negative F_{IS} values could indicate that the population was in outbreeding (Caballero and

Toro, 2002). In addition, mating could be occurring between individuals from different populations (Wright, 1965; Chesser, 1991). The F_{IS} value was negative for the studied breeds of the reproductive herd, where Brahman and Braunvieh had the lowest F_{IS} values, suggesting an excess of heterozygotes and a lack of population structure (Tantia et al., 2006). This could be due to the small sample population size.

According to the AMOVA results (Table 2), the proportion of genetic variability attributable to the difference variation among populations, and within individuals was 24.29% and 75.71%, respectively. These results implied lower genetic differentiation among breeds than within breeds maintained at EEA Donoso. Similar studies have reported lower values for variation across populations (Cañón et al., 2001; Lirón et al., 2006; Egito et al., 2007). Lirón et al. (2006) reported that 8.8% of the total genetic variation corresponded to differences between populations (zebu and taurine breeds), while 91.2% was explained by differences between and within individuals. Cañón et al. (2001) indicated that about 7% of the total genetic variation corresponded to differences between racial groups, while the remaining 93% corresponded to differences between and within individuals. On the other hand, Egito et al. (2007) reported a value of 12% for genetic variation attributable to differences between breed groups. The higher value obtained in the present study may be linked to the characteristics of the sampling (Kitada et al., 2021). The AFB breed group is made up of highly heterogeneous animals, which magnifies the within-group (within individual) variance compared to the between-group (among population) variance. Many of these groups are also highly related to each other (i.e., Gyr-Brahman pair, Simmental-Braunvieh pair), which is further confirmed in the population structure analyses. The little degree of variation is consistent with the F_{ST} for AFB-Braunvieh pair (0.08) and AFB-Simmental pair (0.09).

Table 3 showed that the lowest genetic distance (0.08) was observed for the AFB and Braunvieh. Similarly, for the AFB and Simmental breeds the genetic distance was low (0.09). These values close to zero indicate that these breeds shared their genetic material through breeding. Likewise, Figure 2 showed that Simmental, Braunvieh, and AFB grouped closely to each other. In addition, many individuals of AFB possess traits of Braunvieh and Brown Swiss as they are also employed for beef and milk production. Also, concordant to the genetic distance, PCA and DAPC indicated that Brahman and Gyr are closely related. The share of genetic material between Brahman and Gyr can be explained as they belong to the same species, *B. indicus*. On the contrary, higher genetic distances were observed between Brahman with 1) Braunvieh (0.37), 2) Simmental (0.35), 3) and AFB (0.31), indicating some degree of isolation between these breeds, that is, they are not currently breeding with one another. This could be because the region where the AFB samples were collected is located in the southern parts of the country, where climatic conditions are cold. Hence, European breeds such as Brown Swiss, Braunvieh, Simmental, Overo negro, Jersey, etc., are more commonly used in these regions because of the cold climate, whereas Zebuine breeds, such as the Brahman and Gyr, are preferably used in the Amazon region of the country, in the north. Our phylogenetic reconstruction is in concordance with ADMIXTURE analysis and genetic distances. We identified that the AFB breed is closer to

Braunvieh than the Simmental breed and others. Arbizu et al. (2022) examined the relationship between PCC and other *B. taurus* with an analysis of the mitochondrial genome and validated these relationships, probably influenced by the high introgression and crossing over. Also, a strong relationship between Gyr and Brahman breeds was identified.

We analyzed the genetic structure of the AFB cattle by using SNP markers. This information will be valuable to our farmers as well as future studies. The results of this study provide some insight that AFB can become a separate breed in the future. The analysis also provides evidence for two subgroups within the AFB group (Figure 4), with one level higher of genetic differentiation than the other one. Also, this new information of a Peruvian reproductive cattle herd would offer valuable information to establish a genetic nucleus herd and modern breeding programs. In addition, we expect molecular tools become widely employed in favor of the cattle industry in Peru.

5 Conclusion

We here determined for the first time the genetic diversity and population structure of a Peruvian cattle herd using SNP data. Braunvieh breed possessed the highest genetic diversity while Brahman the lowest. Most of the variance occurs within individuals among the five breeds evaluated in this study. A total of two clusters were identified, showing, as expected, a clear separation between *B. indicus* (Brahman and Gyr) and *B. taurus* breeds (Braunvieh, AFB and Simmental). Interestingly, the AFB was placed in a single cluster, providing evidence that this may be considered a breed as farmers from Arequipa breed their animals in favor of fighting traits. Additional work is needed to also characterize other cattle herd of INIA located in San Martin region. We hope this work will pave the way towards developing a modern cattle breeding program in Peru.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://datadryad.org/stash/share/xHP03DhsK4QVn2wh8QZzaOCGEBUYN1Kg5Pnui2xkqIo>.

Ethics statement

The animal study was reviewed and approved by Peruvian National Law No. 30407: “Animal Protection and Welfare”. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

Conceptualization, F-AC, DF, and CA; Methodology, F-AC, DF, CQ, and CA; Formal analysis, F-AC, DF, and RE; Investigation,

F-AC, DF, JG, RE, and CA; Data curation, F-AC, and DF; Writing—original draft preparation, F-AC; Writing—review and editing, F-AC, and CA; Project administration, HV, PM, JG, WS, and CA; Funding acquisition, WS; HV, JM, JG, PM, JG, and CQ. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the following two research projects: “Creación del servicio de agricultura de precisión en los Departamentos de Lambayeque, Huancavelica, Ucayali y San Martín 4 Departamentos” and “Mejoramiento de la disponibilidad de material genético de ganado bovino con alto valor a nivel nacional 7 departamentos” of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government, with grant numbers CUI 2449640 and CUI 2432072, respectively. CA was supported by PP0068 “Reducción de la vulnerabilidad y atención de emergencias por desastres”.

Acknowledgments

The authors thank Jerry Valdeiglesias and Fernando Gomez for the logistic support in Arequipa. We also thank Ivan Ucharima, Maria Angélica Puyo, Cristina Aybar, and Erick Rodriguez for supporting the logistic activities in the laboratory. Also, we would like to thank Cesar Osorio for his comments. We thank the comments of two reviewers that greatly improved our manuscript. Finally, the authors thank the Bioinformatics High-Performance Computing server of Universidad Nacional Agraria la Molina for providing resources for data analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1073843/full#supplementary-material>

References

- Acosta, A. C., Uffo, O., Sanz, A., Ronda, R., Osta, R., Rodellar, C., Martín-Burriel, I., and Zaragoza, P. (2013). Genetic diversity and differentiation of five Cuban cattle breeds using 30 microsatellite loci. *Journal of Animal Breeding and Genetics* 130 (01), 79–86. doi:10.1111/j.1439-0388.2012.00988.x
- Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinforma.* 12, 246. doi:10.1186/1471-2105-12-246
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/GR.094052.109
- Arbizu, C. I., Ferro-Mauricio, R. D., Chávez-Galarza, J. C., Vásquez, H. V., Maicelo, J. L., Poemape, C., et al. (2022). The complete mitochondrial genome of a neglected breed, the Peruvian creole cattle (*Bos taurus*), and its phylogenetic analysis. *Data* 7, 76. doi:10.3390/data7060076
- Baltenweck, I., Cherney, D., Duncan, A., Eldermire, E., Lwoga, E. T., Labarta, R., et al. (2020). A scoping review of feed interventions and livelihoods of small-scale livestock keepers. *Nat. Plants* 6, 1242–1249. doi:10.1038/s41477-020-00786-w
- Beuzen, N. D., Stear, M. J., and Chang, K. C. (2000). Molecular markers and their use in animal breeding. *Vet. J.* 160, 42–52. doi:10.1053/TVJL.2000.0468
- Caballero, A., and Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* 2002 33 (3), 289–299. doi:10.1023/A:1019956205473
- Cañón, J., Alexandrino, P., Bessa, I., Carleos, C., Carretero, Y., Dunner, S., et al. (2001). Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genet. Sel. Evol.* 33, 311–332. doi:10.1186/1297-9686-33-3-311
- Chagunda, M. G. G., Mujibi, F. D. N., Dusingizimana, T., Kamana, O., Cheruiyot, E., and Mwai, O. A. (2018). Use of high density single nucleotide polymorphism (SNP) arrays to assess genetic diversity and population structure of dairy cattle in smallholder dairy systems: The case of Girinka Programme in Rwanda. *Front. Genet.* 9, 438. doi:10.3389/fgene.2018.00438
- Chen, B., Cole, J. W., and Grond-Ginsbach, C. (2017). Departure from Hardy weinberg equilibrium and genotyping error. *Front. Genet.* 8, 167. doi:10.3389/fgene.2017.00167
- Chesser, R. K. (1991). Influence of gene flow and breeding tactics on gene diversity within populations. *Genetics* 129, 573–583. doi:10.1093/GENETICS/129.2.573
- Dadi, H., Kim, J. J., Yoon, D., and Kim, K. S. (2012). Evaluation of single nucleotide polymorphisms (SNPs) genotyped by the Illumina bovine SNP50K in cattle focusing on hanwoo breed. *Asian-Australasian J. Anim. Sci.* 25, 28–32. doi:10.5713/AJAS.2011.11232
- Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J. W., Molina Alcalá, A., Sonstegard, T. S., et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10, e1004254. doi:10.1371/journal.pgen.1004254
- Delgado, A., García, C., Allcahuamán, D., Aguilar, C., Estrada, P., and Vega, H. (2019). Caracterización fenotípica del ganado criollo en el Parque nacional Huascarán – ancash. *Rev. Investig. Vet. del Perú* 30, 1143–1149. doi:10.15381/RIVEP.V30I3.16611
- Dipás Vargas, E. S. (2015). *Zoometría e índices corporales del vacuno criollo en el matadero de Quicapata de la provincia de Huamanga, a 2720 m.s.n.m. Ayacucho - 2014*. Ayacucho: Universidad Nacional de San Cristóbal de Huamanga.
- Edea, Z., Bhuiyan, M. S. A., Dessie, T., Rothschild, M. F., Dadi, H., and Kim, K. S. (2015). Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. *Animal* 9, 218–226. doi:10.1017/S1751731114002560
- Egito, A. A., Paiva, S. R., Albuquerque, M. D. S. M., Mariante, A. S., Almeida, L. D., Castro, S. R., et al. (2007). Microsatellite based genetic diversity and relationships among ten Creole and commercial cattle breeds raised in Brazil. *BMC Genet.* 8, 83–14. doi:10.1186/1471-2156-8-83
- Espinoza, R., and Urviola, G. (2005). Biometría y constantes clínicas del bovino criollo en el centro de investigación y producción Chuquibambilla de Puno (Perú). *Arch. Zootec.* 54, 233–236.
- European Cattle Genetic Diversity Consortium Lenstra, J. A., Nijman, I. J., Eding, H., Engelsma, K. A., Bijma, P., et al. (2006). Marker-assisted conservation of European cattle breeds: An evaluation. *Anim. Genet.* 37, 475–481. doi:10.1111/J.1365-2052.2006.01511.X
- Excoffier, L., and Lischer, H. E. L. (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. doi:10.1111/J.1755-0998.2010.02847.X
- FAO (2022). Producción animal. Available at: <https://www.fao.org/animal-production/es> (Accessed April 22, 2022).
- Ginja, C., Telo Da Gama, L., and Penedo, M. C. T. (2010). Analysis of STR markers reveals high genetic structure in Portuguese native cattle. *J. Hered.* 101, 201–210. doi:10.1093/jhered/esp104
- Giovambattista, G., Ripoli, M. V., Peral-García, P., and Bouzat, J. L. (2001). Indigenous domestic breeds as reservoirs of genetic diversity: The argentinean creole cattle. *Anim. Genet.* 32, 240–247. doi:10.1046/J.1365-2052.2001.00774.X
- Gómez-Rubio, V. (2017). ggplot2-elegant graphics for data analysis. *J. Stat. Softw.* 77, 1–3.
- Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals – A review. *Anim. Genet.* 41, 6–31. doi:10.1111/J.1365-2052.2010.02038.X
- Gruber, B., Unmack, P. J., Berry, O. F., and Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* 18, 691–699. doi:10.1111/1755-0998.12745
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., et al. (2011). Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* 11, 591–611. doi:10.1111/J.1755-0998.2011.03014.X
- Haasl, R. J., and Payseur, B. A. (2011). Multi-locus inference of population structure: A comparison between single nucleotide polymorphisms and microsatellites. *Hered. (Edinb.)* 106, 158–171. doi:10.1038/hdy.2010.21
- Herrero, M., Grace, D., Njuki, J., Johnson, N., Enahoro, D., Silvestri, S., et al. (2013). The roles of livestock in developing countries. *Animal* 7, 3–18. doi:10.1017/S175713112001954
- Hidalgo, C., Tamargo, C., Fernández, A., and Merino, M. (2015). Recursos zogenéticos. Banco de razas domésticas autóctonas en peligro de desaparición. *Tecnol. Agroaliment.* 15, 45–52.
- Hoban, S. M., Hauflé, H. C., Pérez-Espona, S., Arntzen, J. W., Bertorelle, G., Bryja, J., et al. (2013). Bringing genetic diversity to the forefront of conservation policy and management. *Conserv. Genet. Resour.* 2013 52 (5), 593–598. doi:10.1007/S12686-013-9859-Y
- Instituto Nacional de Estadística Informática, e. (2012). IV censo nacional agropecuario 2012. (INEI). Available at: <http://censos.inei.gob.pe/Cenagro/redatam/#> (Accessed March 14, 2022).
- Jombart, T., and Collins, C. (2017). *A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.1.0*. London, United Kingdom: Imp. Coll.
- Kassambara, A., and Mundt, F. (2017). Factoextra: Extract and visualize the results of multivariate data analyses. *R. Packag. version 1*, 337–354.
- Kitada, S., Nakamichi, R., and Kishino, H. (2021). Understanding population structure in an evolutionary context: Population-specific FST and pairwise FST. *G3 (Bethesda)* 11. doi:10.1093/g3journal/jkab316
- Kiwuwa, G. H. (1992). “Breeding strategies for small ruminant productivity in Africa.” in Proceeding of the Biennial Conference of the African Small Ruminant Research Network, 10-14 Dec 1990 (Nairobi (Kenya): African Small Ruminant Research Network, ILCA), 568.
- Knaus, B. J., and Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17, 44–53. doi:10.1111/1755-0998.12549
- Lamkey, K. R., and Edwards, J. W. (2015). “Quantitative genetics of heterosis,” in *Genet. Exploit. Heterosis crop*, 31–48. doi:10.2134/1999.geneticsandexploitation.c4
- Laoun, A., Harkat, S., Lafri, M., Gaouar, S. B. S., Belabdi, I., Ciani, E., et al. (2020). Inference of breed structure in farm animals: Empirical comparison between SNP and microsatellite performance. *Genes* 2020 11, 57. doi:10.3390/GENES11010057
- León-Velarde, C. U., and Quiroz, R. (2004). The development of livestock production systems in the andean region: Implications for smallholder producers. *Anim. Prod. Anim. Sci. Worldw. A Rev. Dev. Res. Livest. Syst.* 1, 233–240. doi:10.3920/978-90-8686-533-8
- Lirón, J. P., Peral-García, P., and Giovambattista, G. (2006). Genetic characterization of Argentine and Bolivian creole cattle breeds assessed through microsatellites. *J. Hered.* 97, 331–339. doi:10.1093/jhered/esl003
- Lischer, H. E. L., and Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28, 298–299. doi:10.1093/bioinformatics/btr642
- Martínez, M., Vargas, B., Cordero, J. M., Chacón, I., León, B., and Rica, C. (2015). Diversidad genética entre subpoblaciones raciales bovinas de Costa Rica. *Agron. Costarric.* 39, 33–45.
- Mebratu, B., Fesseha, H., and Goa, E. (2020). Embryo transfer in cattle production and its principle and applications. *Int. J. Pharm. Biomed. Res.* 7, 40–54. doi:10.18782/2394-3726.1083
- Morin, P. A., Martien, K. K., and Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Mol. Ecol. Resour.* 9, 66–73. doi:10.1111/J.1755-0998.2008.02392.X
- Mukhopadhyay, C. S., Kumar, A., and Deb, R. (2020). Cattle genomics: Genome projects, current status, and future applications. *Genomics Biotechnol. Adv. Vet. Poult. Fish.*, 3–28. doi:10.1016/B978-0-12-816352-8.00001-1
- Notter, D. R. (1999). The importance of genetic diversity in livestock populations of the future. *J. Anim. Sci.* 77, 61–69. doi:10.2527/1999.77161X
- Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26, 419–420. doi:10.1093/bioinformatics/btp696

- Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi:10.1093/bioinformatics/bty633
- Primo, A. (1992). El ganado bovino ibérico en las américas: 500 años después. *Arch. Zootec.* 41, 421–432.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795
- Purfield, D. C., McClure, M., and Berry, D. P. (2016). Justification for setting the individual animal genotype call rate threshold at eighty-five percent. *J. Anim. Sci.* 94, 4558–4569. doi:10.2527/jas.2016-0802
- Quispe, J. E. (2016). El bovino criollo del altiplano peruano: Origen, producción y perspectivas. *Rev. Investig. Altoandinas - J. High. Andean Res.* 18, 257–270. doi:10.18271/RIA.2016.215
- Ruiz, R. E., Saucedo-Uriarte, J. A., Portocarrero-villegas, S. M., Quispe-ccasa, H. A., and Cayo-colca, I. S. (2021). Zoometric characterization of creole cows from the Southern Amazon region of Peru. *Diversity* 13, 510. doi:10.3390/d13110510
- Schrooten, C., and van Arendonk, J. A. M. (1992). Stochastic simulation of dairy cattle breeding schemes: Genetic evaluation of nucleus size and type. *J. Anim. Breed. Genet.* 109, 1–15. doi:10.1111/j.1439-0388.1992.tb00374.x
- Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L. (2015). Widde: A web-interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics* 16, 940. doi:10.1186/s12864-015-2181-1
- Séré, C., Steinfeld, H., and Groenewold, J. (1996). *World livestock production systems*. Rome: Food and Agriculture Organization of the United Nations.
- Tantia, M. S., Vijn, R. K., Mishra, B., Bharani Kumar, S. T., and Arora, R. (2006). Multilocus genotyping to study population structure in three buffalo populations of India. *Asian-Australasian J. Anim. Sci.* 19, 1071–1078. doi:10.5713/AJAS.2006.1071
- The Bovine Hapmap Consortium Gibbs, R. A., Taylor, J. F., Van Tassel, C. P., Barendse, W., Eversole, K. A., et al. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Sci.* (80- 324, 528–532. doi:10.1126/science.1167936
- Thibier, M., and Wagner, H. G. (2002). World statistics for artificial insemination in cattle. *Livest. Prod. Sci.* 74, 203–212. doi:10.1016/S0301-6226(01)00291-3
- Tisdell, C. (2003). Socioeconomic causes of loss of animal genetic diversity: Analysis and assessment. *Ecol. Econ.* 45, 365–376. doi:10.1016/S0921-8009(03)00091-0
- Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* 34, 275–305. doi:10.1186/1297-9686-34-3-275
- Williams, W., G., and Anderson, D. P. (2019). Growth of the Latin American livestock industry. *Choices* 34, 1–12.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evol. (N. Y.)* 19, 395. doi:10.2307/2406450
- Zeng, J., Toosi, A., Fernando, R. L., Dekkers, J. C. M., and Garrick, D. J. (2013). Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genet. Sel. Evol.* 45, 11. doi:10.1186/1297-9686-45-11
- Zenger, K. R., Khatkar, M. S., Cavanagh, J. A. L., Hawken, R. J., and Raadsma, H. W. (2007). Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. *Anim. Genet.* 38, 7–14. doi:10.1111/j.1365-2052.2006.01543.X