R. Bras. Zootec., 53:e20230095, 2024
https://doi.org/10.37496/rbz5320230095

Breeding and genetics
Full-length research article

Revista Brasileira de Zootecnia

Brazilian Journal of Animal Science
e-ISSN 1806-9290
www.rbz.org.br

Genetic structure and diversity of Santa Inês sheep flocks in Central-Northern Brazil

Alzira Regina Silva de Deus[1], Geice Ribeiro da Silva[1], Luciano Silva Sena[1], Fábio Barros Britto[2], Artur Oliveira Rocha[3], Débora Araújo de Carvalho[4], Fabiana Cristina Belchior de Sousa[1]*, Natanael Pereira da Silva Santos[5], José Lindenberg Rocha Sarmento[5]

[1] Universidade Federal do Piauí, Centro de Ciências Agrárias, Teresina, PI, Brasil.
[2] Universidade Federal do Piauí, Departamento de Biologia, Teresina, PI, Brasil.
[3] Purdue University, Department of Animal Sciences, West Lafayette, IN, USA.
[4] Universidade Estadual do Piauí, Teresina, PI, Brasil.
[5] Universidade Federal do Piauí, Departamento de Zootecnia, Teresina, PI, Brasil.

***Corresponding author:**
fabiana.cristina1@hotmail.com

Received: September 14, 2023
Accepted: March 11, 2024

How to cite: Deus, A. R. S.; Silva, G. R.; Sena, L. S.; Britto, F. B.; Rocha, A. O.; Carvalho, D. A.; Sousa, F. C. B.; Santos, N. P. S. and Sarmento, J. L. R. 2024. Genetic structure and diversity of Santa Inês sheep flocks in Central-Northern Brazil. Revista Brasileira de Zootecnia 53:e20230095. https://doi.org/10.37496/rbz5320230095

Editor:
Mateus Pies Gionbelli

ABSTRACT - The objective of this study was to assess the genetic structure and diversity of six Santa Inês sheep flocks from the Central-Northern Brazil. A panel of 20 highly polymorphic and informative microsatellite loci was selected and amplified. The following parameters were obtained: overall mean of number of alleles = 15.4; expected heterozygosity (He) = 0.89; polymorphism information content (PIC) = 0.88; discriminatory capacity = 0.95; combined probability of identity = 1.50×10^{-34}; and probability of exclusion = 1.00. The flocks with the lowest and the highest degrees of genetic variability were Farm 6 (He = 0.70, PIC = 0.653, and allelic richness [Ar] = 3.76) and Farm 1 (He = 0.89, PIC = 0.882, and Ar = 4.39), respectively. Indications of genetic bottleneck were observed in all flocks, as well as moderate genetic differentiation, with F_{ST} = 0.053, R_{ST} = 0.096, and Dest = 0.169. The migration rate in all flocks was high, with a trend towards Farm 1. This finding was not in agreement with the substructure found with the Bayesian admixture analysis and corroborated the array obtained with the principal component analysis and the clustering analysis. The results revealed moderate structuring and high genetic diversity in the flocks. However, management strategies should be reviewed, as evidence of bottleneck and genetic erosion was observed.

Keywords: Brazil, genetic structure, genetic variability, microsatellites, sheep

1. Introduction

Sheep farming is practiced throughout Brazil, especially in the Northeastern and Southern regions of the country, with a focus on meat production. Despite the great potential of sheep meat production in Brazil, it is not enough to meet the domestic demand, hence requiring importation from different countries (Lobo, 2019). The selection of animals with good adaptability and reproductive performance is important to increase productivity. In tropical countries, such as Brazil, Santa Inês is one of the main options for sheep meat production. This breed has higher growth potential, rusticity, and reproductive efficiency, as well as lower susceptibility to parasites, in comparison with other Brazilian sheep breeds (Jucá et al., 2016).

Despite that, the Santa Inês is at risk of losing genetic variability and, consequently, some of its main characteristics. The Food and Agriculture Organization of the United Nations (FAO) and other important

institutions point out that uncontrolled mating without adequate zootechnical control can lead to genetic erosion or dilution, thus compromising the development of sheep flocks (Gaouar et al., 2017). Therefore, research institutions in Brazil have sought to develop strategies for genetic conservation based on population structure, focusing on the preservation of the main characteristics of the Santa Inês.

Genetic markers, such as microsatellites, have been commonly used in conservation and breeding programs, because they are widely distributed throughout the genome, highly polymorphic, multiallelic, highly informative, and potentially heterologous, making their use more accessible (Câmara et al., 2017).

In the present study, we aimed to investigate the genetic structure and diversity of Santa Inês sheep flocks to develop a framework for decision-making in conservation and breeding.

2. Material and Methods

Research on animals was conducted according to the institutional Ethics Committee on the Use of Animals in Research (register # 337/17).

2.1. Biological material and genomic DNA isolation

Peripheral blood samples were collected from 257 animals raised in six Santa Inês flocks in different municipalities of the states of Piauí and Maranhão, in the Central-Northern Brazil, from October 2013 to April 2018 (Table 1).

Prior to DNA extraction, blood samples were stored at −20 °C, in vacuum tubes containing 1% of EDTA. Genomic DNA extraction and purification were performed using a commercial DNeasy® Blood & Tissue kit according to the manufacturer's protocol. After extraction, DNA integrity was checked using 0.8% agarose gel and ethidium bromide staining.

Table 1 - Location of the six herds where peripheral blood samples were collected in different municipalities in the state of Piauí (PI) and Maranhão (MA), from October 2013 to April 2018

Locks	Municipality	Geographic location	Number of animals
Farm 1	José de Freitas – PI	4°39'27.2" S, 42°28'37.8" W, altitude of 140 m	135
Farm 2	Santa Inês – MA	3°42'43.4" S, 45°18'09.3" W, altitude of 24 m	52
Farm 3	José de Freitas – PI	4°39'34.7" S, 42°28'20.1" W, altitude of 140 m	27
Farm 4	Floriano – PI	6°46'19.9" S, 43°03'42.4" W, altitude of 140 m	19
Farm 5	Campo Maior – PI	4°47'58.2" S, 42°08'03.4" W, altitude of 125 m	18
Farm 6	Campo Maior – PI	5°02'50.4" S, 42°01'12.9" W, altitude of 125 m	6

2.2. Amplification of microsatellites

Twenty microsatellite loci were selected from the list of markers recommended by FAO (FAO, 2011). Previously, the reactions were tested and optimized in a volume of 16 µL until the following profile was achieved: at least 2.5 mM dNTPs, 1 × Tris-HCl/KCl buffer, 1.0–2.0 mM $MgCl_2$ (Table 2), 1.25 µM of each primer, one unit of Taq DNA Polymerase, deionized water, and 3 µL of extracted DNA (3 ng/µL).

The thermocycler (Applied Amplifications by Life Technologies) was programmed with the following parameters: 94 °C for 5 min, followed by 30 to 35 cycles of denaturation (94 °C,45 s); annealing (50 to 62 °C, 45 s; see Table 2 for details); and extension (72 °C, 50 s). After the cycles, a final extension (72 °C, 7 min) was performed. The amplification products were visualized after electrophoresis in 6% non-denaturing polyacrylamide gel stained with silver nitrate solution (0.0015 g/mL) with a 0.4% (v/v) formaldehyde solution, followed by NaOH (1.5 mg/mL) with 0.4% (v/v) formaldehyde, using a modification of the method proposed by Benbouza et al. (2006). The results were coded according to the allele sizes and recorded in a spreadsheet to obtain a database of allele frequencies.

Table 2 - Experimental parameters used for the amplification of 20 microsatellite loci and allele size range (ASR) detected in PCR

Locus	Primer sequence	Ta (°C)	$MgCl_2$ (mM)	NC	ASR (bp)
MAF065	F: AAAGGCCAGAGTATGCAATTAGGAG R: CCACTCCTCCTGAGAATATAACATG	60	1.3	30	100-158
MAF209	F: GATCACAAAAAGTTGGATACAACCGTG R: TCATGCACTTAAGTATGTAGGATGCTG	62	1.3	30	100-154
BM6526	F: CATGCCAAACAATATCCAGC R: TGAAGGTAGAGAGCAAGCAGC	55	1.3	30	140-180
OarFCB304	F: CCCTAGGAGCTTTCAATAAAGAATCGG R: CGCTGCTGTCAACTGGGTCAGGG	55	1.0	30	160-198
CSRM60	F: AAGATGTGATCCAAGAGAGAGGCA R:AGGACCAGATCGTGAAAGGCATAG	55	1.5	30	60-92
CSSM66	F: ACACAAATCCTTTCTGCCAGCTGA R: AATTTAATGCACTGAGGAGCTTGG	55	1.5	30	12-208
ILSTS11	F: GCT TGC TAC ATG GAA AGT GC R: CTA AAA TGC AGA GCC CTA CC	55	1.5	30	254-294
INRA23	F: GAGTAGAGCTACAAGATAAACTTC R: TAACTACAGGGTGTTAGATGAACTC	58	1.0	30	190-296
ETH10	F: GTTCAGGACTGGCCCTGCTAACA R: CCTCCAGCCCACTTTCTCTTCTC	57	1.5	30	200-242
BM8125	F: CTCTATCTGTGGAAAAGGTGGG R: GGGGGTTAGACTTCAACATACG	57	1.3	30	108-132
MM12	F: CAAGACAGGTGTTTCAATCT R: ATCGACTCTGGGGATGATGT	54	1.3	30	100-144
BM1329	F: TTGTTTAGGCAAGTCCAAAGTC R: AACACCGCAGCTTCATCC	53	1.5	30	146-180
BM1818	F: AGCTGGGAATATAACCAAAGG R: AGTGCTTTCAAGGTCCATGC	55	1.5	30	248-298
INRA063	F: ATTTGCACAAGCTAAATCTAACC R: AAACCACAGAAATGCTTGGAAG	58	1.5	30	150-184
ILSTS87	F: AGCAGACATGATGACTCAGC R: CTGCCTCTTTTCTTGAGAG	55	1.5	30	140-182
INRABERN172	F: CCACTTCCCTGTATCCTCCT R: GGTGCTCCCATTGTGTAGAC	56	2.0	35	132-192
CSRD247	F: GGACTTGCCAGAACTCTGCAAT R: CACTGTGGTTTGTATTAGTCAGG	50	1.5	35	200-264
TGLA122	F: CCCTCCTCCAGGTAAATCAGC R: AATCACATGGCAAATAAGTACATAC	52	1.5	30	136-182
ETH225	F: GATCACCTTGCCACTATTTCCT R: ACATGACAGCCAGCTGCTACT	50	2.0	35	130-162
OarFCB48	F: GAGTTAGTACAAGGATGACAAGAGGCAC R: GACTCTAGAGGATCGCAAAGAACCAG	58	1.5	30	138-170

PCR - polymerase chain reaction; Ta - annealing temperature; F - forward; R - reverse; NC - number of cycles; bp - base pairs.

2.3. Quality control and data analysis

The Micro-Checker software v.2.2.3 (Van Oosterhout et al., 2004) was used to detect possible null alleles, allele dropout, and scoring errors due to stuttering. FreeNA (Chapuis and Estoup, 2007) was used to assess the effect of null alleles on genetic differentiation estimates. This software calculates the genetic distance and the F_{FST} estimates with and without the Excluding Null Alleles (ENA) correction.

The probability of loci being under selection was calculated using Bayescan software v.2.1 (Foll and Gaggiotti, 2008), with Bayesian inference. Several runs were performed previously to achieve acceptance rates between 0.25 and 0.45. The following parameters were used: prior odds of 10:1 for the neutral model; 20 pilot runs each consisting of 5,000 iterations; and 100,000 iterations with burn-in of 50,000. According to Jeffrey's scale of evidence in the manual of Bayescan, loci with a posterior probability of ≥ 0.76 were considered as being under "substantial selection" odds.

The R package diveRsity (Keenan et al., 2013) was used to assess genetic variability by calculating the following parameters: observed heterozygosity (Ho), expected heterozygosity (He), and number of alleles (A). Allelic richness (Ar) and private allelic richness (PvAr) were obtained using the Hp-Rare software (Kalinowski et al., 2007). The polymorphic information content (PIC) was estimated using the Cervus software (Kalinowski et al., 2007).

The probability of exclusion (PE), the probability of identity (PI), and the combined probability of identity (CPI) were obtained using Genalex software (Peakall and Smouse, 2006) to investigate the power of markers to assess the kinship among flocks. The discriminatory capacity was estimated using the FORSTAT software (Ristow and D'Amato, 2017). Deviations from the Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium were tested using the R package Genepop (Rousset, 2008). Subsequently, the Bonferroni correction was carried out to prevent false significant estimates ($P<0.05$).

For the analysis of genetic differentiation among flocks, the parameters F_{ST}, R_{ST}, and D_{est} were evaluated using the software SPAGeDi (Hardy and Vekemans, 2002) and R package diveRsity. R_{ST} is a population genetic distance that is analogous to F_{ST}, but includes information on allele size. D_{est} is a parameter of real differentiation based on the frequency of unique alleles of each subpopulation (Jost, 2008).

Also using SPAGeDi, allele size permutation test was implemented to assess the level of markers that fit the stepwise mutation model (SMM) ($R_{ST}>pR_{ST}$, in which pR_{ST} represents the mean of R_{ST} after 10,000 permutations).

A Bayesian clustering analysis applying Markov Chain Monte Carlo (MCMC) estimation was performed to assess the relationship among flocks using the admixture model implemented in the software STRUCTURE v2.3.2 (Pritchard et al., 2000). For the purpose of estimation, in each of the five runs performed for each *K* (20 simulations each), the first 5×10^5 iterations were discarded for the burn-in process and the 1×10^6 remaining iterations were used. The ΔK method (Evanno et al., 2005) was used to determine the value of *K* that best fit the data using STRUCTURE HARVESTER v.0.6.1. The R package POPHELPER v.2.2.9 was used to calculate the mean of the five replicates of the best *K* and to generate a final barplot of the array among genotypes.

A dendrogram using Neighbor-Joining algorithm based on Nei's genetic distance was constructed using the R packages Poppr, ADEgenet, Ape, Polysat, and Ggplot2. The principal coordinate analysis (PCoA) based on shared allele distance matrix was performed using the same packages mentioned above to obtain a distribution profile of the genotypes in a 2D plane.

The Wilcoxon sign-rank test was conducted to detect population bottlenecks using the software Bottleneck v.2.1.02 (Piry et al., 1999). The genetic bottleneck effect tends to lead to higher HWE heterozygosity than heterozygosity under mutation-drift equilibrium. Thus, for these analyses, markers were considered following mutational model that aim to determine the expected number of alleles based on the observed heterozygosity. Three models were used: stepwise mutation model (SMM) (Slatkin, 1995), which assumes that a mutation in a locus will result in the gain or loss of a repeat unit; infinite allele model (IAM) (Wright, 1931), in which a new mutation always generates

a new allele; and the two-phase mutation model (TPM) (Di Rienzo et al., 1994), which assumes that mutations may occur gradually or stepwise. The following parameters were adopted for the TPM: 95% single-step mutations and 5% multiple-step mutations (with variance among multiple steps of 12).

Network graphs were constructed using the function divMigrate of the R package diveRsity to assess the bidirectional distribution of gene flow among flocks based on the following estimators: G_{ST}, D'Jost, and Nm.

3. Results

3.1. Performance of microsatellite markers

Possible null alleles were detected for nine of the 20 microsatellite loci used. However, only one locus was not included in the analyses (CSRM60) because its null allele frequency was greater than 0.20, as suggested by Dakin and Avise (2004). No significant differences (P>0.05) were detected between the parameters of global genetic differentiation before (F_{ST} = 0.052) and after (F_{ST} = 0.049) the ENA correction. Thus, further analysis proceeded considering all the 19 loci with null allele frequency < 0.20.

All markers deviated from the Hardy-Weinberg equilibrium and 65% had significant deficit of heterozygotes (P<0.001), whereas no markers had significant heterozygote excess. Sixty-six percent of the pairwise combinations of loci were not significant for linkage disequilibrium. This suggests that most loci are not statistically associated with each other (Table 3).

All 20 loci were polymorphic, totaling 308 alleles, ranging from 9 (ILSTS11) to 36 (INRABERN172), with an average of 15.4 alleles. Half of the loci were neutral, and the remaining showed evidence of selection ranging from strong (0.91 to 0.97) to decisive (0.99 to 1.00) according to Bayescan estimates (Table 3). When all sampled animals were included in the evaluation, markers had high performance, high polymorphism, and high discriminatory capacity among individuals, with mean PIC and DC values of 0.88 ± 0.05 and 0.95 ± 0.02, respectively (Table 3).

When the 257 animals were included to evaluate the capacity of the 20 markers to infer kinship, PI values were between 0.009 (CSRD247) and 0.094 (ILSTS11), whereas PE values were 0.740 (ILSTS11) and 0.969 (CSRD247). The combined values of PI and PE were 1.5×10^{-34} and 1.0, respectively (Table 3).

The level of genetic diversity (He) of markers ranged from 0.760 (ILSTS11, Ho = 0.550) to 0.930 (CSRD247, Ho = 0.900), with a mean of 0.890 ± 0.21 (Ho = 0.780 ± 0.04) (Table 3).

3.2. Genetic structure and diversity of sheep flocks

All flocks deviated from the Hardy-Weinberg equilibrium. Except for Farm 6, all other flocks had significant deficit of heterozygotes, without significant heterozygote excess (P>0.05). Farms 2 and 3 exhibited the greatest number of markers with possible null alleles, when frequencies ≤ 0.20 were considered as acceptable (Table 4).

The highest degree of genetic diversity was observed in Farm 1 (Ar = 4.690, He = 0.890, and PIC = 0.882), and the lowest in Farm 6 (Ar = 3.760, He = 0.700, PIC = 0.653). Private allelic richness was highest in Farms 1 and 2, with 21 and 13 alleles, respectively (Table 4).

The allele size permutation test detected the influence of the stepwise mutation model on the markers used in the current study. R_{ST} (0.096) was significantly higher than pR_{ST} (0.038) (P-value = 0.005), suggesting that mutation is affecting the genetic differentiation among some flocks. Nevertheless, when each locus was considered separately, this was not observed for most of the loci. This suggests that microsatellite markers could also fit multiple-step mutation models, such as IAM, which indicates that gene flow could be also contributing in some extent to differentiation among flocks.

In general, genetic structuring of flocks was moderate according to SPAGeDi results (F_{ST} = 0.053, R_{ST} = 0.096, and D_{est} = 0.169). This might be due to high gene flow among flocks. High relative migration index was observed, especial towards Farm 1 (Figure 1), with bootstrap confidence level higher than 0.82 (Data not shown).

Table 3 - Genetic diversity parameters for 20 polymorphic microsatellites in Santa Inês sheep flocks from the Mid-North sub-region of Brazil

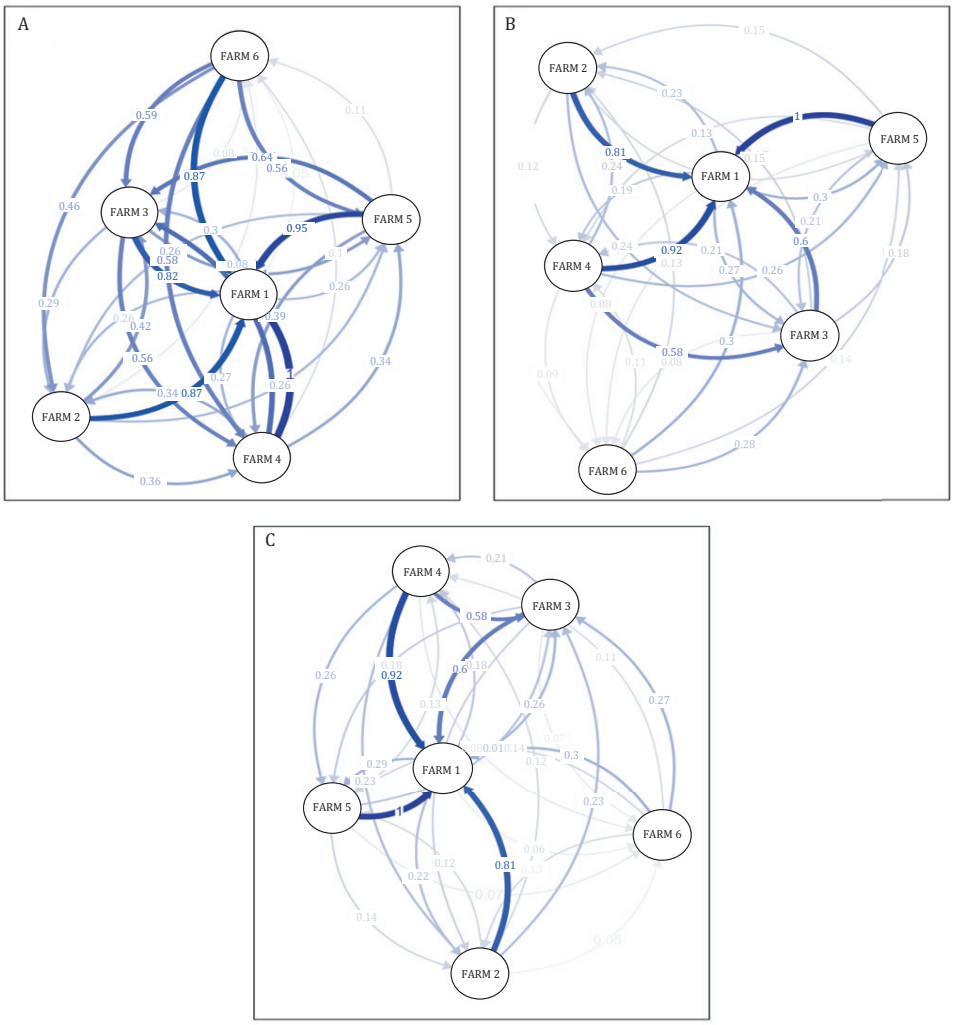
Locus	A	Ho	He	PIC	DC	PI	PE	F_{IS}	pHWE			Null	F_{ST}	D_{est}	R_{ST}	pR_{ST}	P-value (R_{ST}>pR_{ST})	Bayescan prob
									Prob	Def	Exc							
MAF065	19	0.910	0.920	0.918	0.939	0.011	0.959	−0.0522	0.000	0.324	0.684	0.005	0.088	0.640	0.168	0.086	0.090	0.974 ± 0.003
MAF209	15	0.870	0.930	0.921	0.967	0.010	0.962	0.0454	0.000	0.000	1.000	0.029	0.028	0.336	0.063	0.027	0.113	0.045 ± 0.004
BM6526	13	0.830	0.900	0.891	0.971	0.018	0.935	0.0587	0.000	0.000	1.000	0.036	0.031	0.202	0.041	0.032	0.319	0.044 ± 0.003
OarFCB304	18	0.930	0.910	0.907	0.936	0.014	0.951	−0.0711	0.000	0.059	0.941	−0.009	0.073	0.402	0.145	0.070	0.129	0.954 ± 0.003
CSRM60	14	0.000	0.900	0.896	0.908	0.017	0.939	1.0000	0.000	0.000	1.000	0.475	0.213	0.765	0.320	0.203	0.169	1.000 ± 0.000
CSSM66	10	0.660	0.880	0.874	0.966	0.025	0.914	0.2262	0.000	0.000	1.000	0.120	0.058	0.385	0.024	0.056	0.759	0.032 ± 0.003
ILSTS11	9	0.550	0.760	0.721	0.896	0.094	0.740	0.2248	0.000	0.000	1.000	0.120	0.102	0.184	0.143	0.080	0.164	1.000 ± 0.000
INRA023	17	0.920	0.880	0.872	0.940	0.025	0.915	−0.0453	0.000	0.000	1.000	−0.020	0.007	0.119	0.018	0.007	0.177	0.995 ± 0.001
ETH10	13	0.860	0.880	0.871	0.962	0.024	0.921	−0.0094	0.000	0.078	0.922	0.009	0.044	0.350	0.109	0.041	0.024	0.065 ± 0.003
BM8125	12	0.850	0.890	0.880	0.962	0.022	0.923	0.0117	0.000	0.000	1.000	0.019	0.046	0.298	0.059	0.045	0.308	0.029 ± 0.001
MM12	10	0.610	0.800	0.782	0.933	0.057	0.836	0.2275	0.000	0.000	1.000	0.103	0.014	0.050	0.023	0.012	0.262	0.924 ± 0.012
BM1329	10	0.680	0.830	0.815	0.953	0.044	0.866	0.0885	0.000	0.000	1.000	0.082	0.148	0.399	0.230	0.125	0.158	0.044 ± 0.002
BM1818	22	0.900	0.930	0.925	0.981	0.009	0.966	0.0341	0.000	0.256	0.744	0.014	−0.004	−0.103	0.009	−0.004	0.108	1.000 ± 0.000
INRA063	12	0.830	0.890	0.880	0.965	0.022	0.925	0.0305	0.000	0.001	1.000	0.030	0.054	0.294	0.159	0.052	0.055	0.052 ± 0.003
ILSTS087	12	0.900	0.890	0.882	0.970	0.021	0.926	−0.0050	0.000	0.702	0.298	−0.004	−0.002	−0.064	0.006	−0.002	0.213	1.000 ± 0.000
INRABERN172	36	0.880	0.910	0.903	0.976	0.015	0.949	0.0092	0.000	0.009	0.991	0.017	0.044	0.330	0.100	0.038	0.076	0.025 ± 0.002
CSRD247	29	0.900	0.930	0.929	0.983	0.009	0.969	0.0219	0.000	0.018	0.983	0.016	0.020	0.152	0.000	0.015	0.725	0.981 ± 0.005
TGLA122	12	0.860	0.900	0.891	0.958	0.019	0.934	0.0262	0.000	0.001	0.999	0.022	0.034	0.260	0.077	0.032	0.115	0.066 ± 0.004
ETH225	14	0.860	0.900	0.891	0.969	0.018	0.937	0.0245	0.000	0.117	0.883	0.021	0.033	0.263	0.118	0.033	0.026	0.051 ±0.002
OarFCP48	11	0.870	0.890	0.882	0.955	0.022	0.924	0.0211	0.000	0.653	0.347	0.012	0.011	0.118	0.060	0.011	0.033	0.998 ± 0.002
Mean	308	0.780	0.890	0.877	0.954	1.5E−34	1.000	0.0851					0.053	0.169	0.096	0.038	0.005	

A - number of alleles; Ho - observed heterozygosity; He - expected heterozygosity; PIC - polymorphism information content; DC - discriminatory capacity; PI - probability of identity; PE - probability of exclusion with both parents unknown; F_{IS} - inbreeding coefficient; pHWE - P-value of the test for Hardy-Weinberg equilibrium; Prob - probability; Def - deficit of heterozygotes; Exc - excess of heterozygotes; Null - frequency of null alleles; F_{ST} - fixation index; R_{ST} - population genetic differentiation based on allele size; Dest - parameter of real differentiation based on the proportion of unique alleles of each subpopulation; pR_{ST} - permuted R_{ST} ;R_{ST}>pR_{ST} - P-value of the test for allele size; Bayescan prob - posterior probability for the test of selection.

Table 4 - Analysis of the profile of genetic variability of Santa Inês sheep flocks from different municipalities of the Mid-North sub-region of Brazil

Parameter		Farm 1	Farm 2	Farm 3	Farm 4	Farm 5	Farm 6
Ar		4.690	3.990	4.190	4.190	4.170	3.760
Ho		0.810	0.730	0.790	0.750	0.760	0.750
He		0.890	0.790	0.810	0.810	0.800	0.700
PIC		0.882	0.832	0.791	0.785	0.777	0.653
F_{IS}		0.088	0.066	0.029	0.078	0.052	−0.072
	Prob	0.000	0.000	0.000	0.000	0.000	0.000
HWE	Def	0.000	0.000	0.000	0.000	0.000	0.103
	Exc	1.000	1.000	1.000	1.000	1.000	0.897
Null		6	6	4	7	4	Nd
PvAr		20.924	13.474	12.512	10.495	11.087	9.538

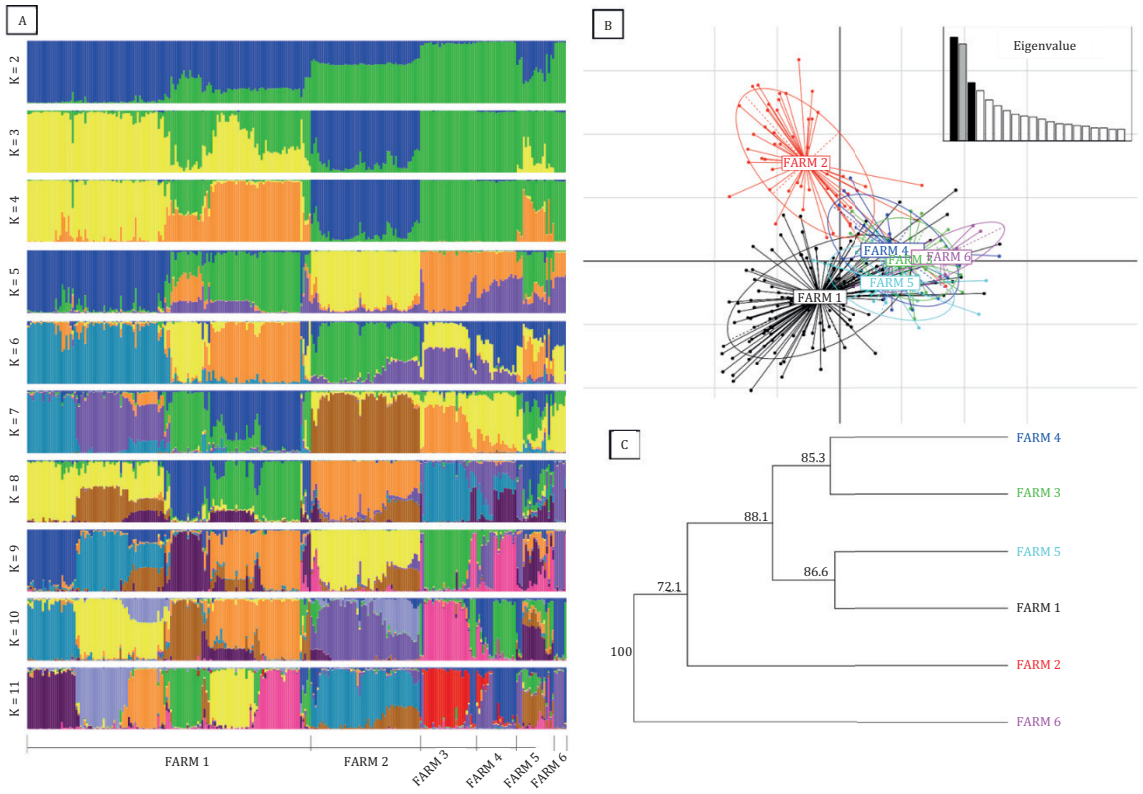
Farm 1 - José de Freitas – PI; Farm 2 - Santa Inês – MA; Farm 3 - José de Freitas – PI; Farm 4 - Floriano – PI; Farm 5 - Campo Maior – PI; Farm 6 - Campo Maior – PI; Ar - allelic richness; Ho - observed heterozygosity; He - expected heterozygosity; PIC - polymorphic information content; F_{IS} - coefficient of inbreeding; Null - frequency of null alleles; PvAr - private allelic richness; HWE - Hardy-Weinberg equilibrium; Prob - probability; Def - deficit of heterozygotes; Exc - excess of heterozygotes; Nd - not done.



Farm 1 - José de Freitas – PI; Farm 2 - Santa Inês – MA; Farm 3 - José de Freitas – PI; Farm 4 - Floriano – PI; Farm 5 - Campo Maior – PI; Farm 6 - Campo Maior – PI.
Wider and darker arrows indicate higher relative proportion.

Figure 1 - Estimate of the gene flow among Santa Inês sheep flocks from different municipalities of the Mid-North sub-region of Brazil based on the estimators D'Jost (A), GST (B), and Nm (C).

After several simulations, $K = 11$ was considered the most likely number of genetic groups. This finding contributed to the subdivision of Farm 1 into six different groups, which corroborated the high levels of genetic diversity of this flock (He, PIC, and Ar). When other simulations were performed, Farm 2 began differentiation when $K = 3$. On the other hand, when $K = 4$, Farm 1 and 5 had some genotypes in common and began distancing from the other flocks. Simultaneously, Farms 4 and 3 shared most of their alleles. This corroborates the dendrogram constructed with the unweighted pair group method with arithmetic mean (confidence level of knots > 72%, with cophenetic correlation of 0.84) and PCoA, despite Farm 6 being different from the other flocks (Figure 2).



Farm 1 - José de Freitas – PI; Farm 2 - Santa Inês – MA; Farm 3 - José de Freitas – PI; Farm 4 - Floriano – PI; Farm 5 - Campo Maior – PI; Farm 6 - Campo Maior – PI.

Figure 2 - Bayesian clustering analysis (A), principal coordinates analysis (PCoA) (B), and dendrogram generated from Nei's genetic distance with the unweighted pair group method with arithmetic mean (UPGMA) (cophenetic correlation of 85%) of the flocks analyzed (C).

When the stepwise mutation model (SMM) was carried out, evidence of genetic drift was observed only in Farm 1. Based on the TPM, genetic drift was detected in Farms 1, 4, and 6. However, by assuming that all markers were under the influence of IAM, all flocks showed signals of genetic bottleneck (Table 5).

4. Discussion

According to the performance of the microsatellite markers evaluated, the presence of null alleles in some loci was not enough to change the population structure significantly. A high level of polymorphism was observed, with elevated values of A, PIC, Ho, and He, above those reported by Petroli et al. (2014) (Ho = 0.34, He = 0.63, A = 7.45) and Souza et al. (2012) (Ho = 0.68, He = 0.745, A = 10, PIC = 0.712)

Table 5 - Wilcoxon sign-rank test for heterozygote excess in six Santa Inês sheep flocks from different municipalities of the Mid-North sub-region of Brazil

Pop	L	N ± SE	He ± SE	Wilcoxon sign-rank test											
				IAM				TPM				SMM			
				Heq	Obs LHexc	Exp LHexc	P-value	Heq	Obs LHexc	Exp LHexc	P-value	Heq	Obs LHexc	Exp LHexc	P-value
Farm 1	20	117.250 ± 9.391	0.895 ± 0.026	0.756 ± 0.072	20	12	0.000	0.864 ± 0.042	18	12	0.000	0.876 ± 0.038	17	12	0.001
Farm 2	20	44.050 ± 5.206	0.794 ± 0.166	0.719 ± 0.140	17	12	0.001	0.809 ± 0.119	13	12	0.273	0.819 ± 0.117	12	12	0.594
Farm 3	20	24.300 ± 2.080	0.830 ± 0.100	0.776 ± 0.117	16	12	0.000	0.837 ± 0.099	13	12	0.493	0.844 ± 0.096	11	12	0.663
Farm 4	20	17.250 ± 1.333	0.833 ± 0.104	0.770 ± 0.137	19	12	0.000	0.822 ± 0.120	14	12	0.027	0.827 ± 0.119	12	12	0.115
Farm 5	20	16.300 ± 1.455	0.827 ± 0.122	0.773 ± 0.137	18	12	0.000	0.823 ± 0.120	12	12	0.226	0.828 ± 0.119	11	12	0.406
Farm 6	20	5.250 ± 0.910	0.774 ± 0.164	0.766 ± 0.149	15	11	0.024	0.766 ± 0.149	13	13	0.045	0.769 ± 0.149	13	13	0.062

Farm 1 - José de Freitas – PI; Farm 2 - Santa Inês – MA; Farm 3 - José de Freitas – PI; Farm 4 - Floriano – PI; Farm 5 - Campo Maior – PI; Farm 6 - Campo Maior – PI; L - number of polymorphic loci; N - mean number of individuals sampled by locus; SE - standard error; He - expected heterozygosity in the scenario in which the populations are in Hardy-Weinberg equilibrium; IAM - infinite allele mutation model; SMM - stepwise mutation model; TPM - two-phase mutation model; Heq - heterozygosity in the scenario with mutation-drift equilibrium; LHexc - number of loci with excess of heterozygosity (He >Heq); Obs - observed; Exp - expected.

for Santa Inês sheep. However, it is important to consider that our results are based on makers recommended by FAO, while those authors used different sets of markers. Also, Petroli et al. (2014) aimed to analyze markers linked to the Major Histocompatibility Complex (Chromosome 20), while FAO markers are randomly distributed in different chromosomes.

Considering comparisons with other breeds, the level of polymorphism observed in the current study was lower than that reported in the Turkish sheep breed Savak-Akkaraman (Ozmen et al., 2020). This could be due to the similarity of the latter with pre-historic breeds. Furthermore, the Savak-Akkaraman has been raised by local tribes and nomads for centuries, which might have contributed to the conservation and variability of most genes. On the other hand, the results shown in the present research indicated higher levels of genetic diversity in the Santa Inês breed compared with the Moroccan (Gaouar et al., 2016) and Algerian breeds (Gaouar et al., 2015; Abdelkader et al., 2018).

In a scenario of random mating, the markers evaluated in the current study demonstrated potential for kinship evaluation. The probability that two unrelated individuals had the same genotype in different flocks was almost null, especially when markers were combined (PI = 1.5×10^{-34}). In general, the probability of exclusion was high (> 0.74) and reached 100% when all markers were included. According to some studies, the required minimum combined probability of exclusion (CPE) for the identification of a true parent is 0.9999 (Piry et al., 1999; Ozmen et al., 2020). This indicates that the markers used in the present study are suitable for kinship evaluation.

The relative allele frequency of all evaluated flocks was not constant. Except for Farm 6, all the other flocks were not on Hardy-Weinberg Equilibrium ($P<0.05$), possibly due to the Wahlund effect (Yilmaz, 2016). Despite this, Farm 6 only had a few individuals included in the analysis. Considering the other farms, this result might have been intensified by non-random mating and directed selection, resulting in the levels of inbreeding (F_{IS}) observed for each flock. This could have also led to genetic bottlenecks, which were observed especially when all markers fitted IAM. However, this was not found when TPM and SMM models were used. Despite that, the three models revealed signals of genetic bottleneck in Farm 1. These findings differed from those reported by Selvam and Kathiravan (2019) and by Radha et al. (2011) in the Indian sheep breeds Madras Red, Mecheri, and Kilakharsal. In these studies, the authors did not find genetic bottlenecks in any of the mutation models.

In the present study, the high level of genetic diversity of the flocks was, in part, probably due to miscegenation in their origin and the system of non-random mating adopted by breeders, in which breeding animals acquired from more productive flocks are prioritized. This corroborates the small differentiation observed, even among flocks (F_{ST} = 0.053 and R_{ST} = 0.09), as well as the high flow estimated among populations (the highest migrations were observed towards Farm 1) and the apparent presence of various genetic subgroups.

In general, the presented results highlight the need for conservation policies that take into account technical information for the Santa Inês sheep. However, there is no consensus on the best strategies related to breed conservation (Meuwissen, 2009; Caballero et al., 2010). Ginja et al. (2013) emphasized that conservation programs should consider strategies based on the specific needs of the species, considering the differences in programs that prioritize conservation among breeds or within them. Also, strategies should consider whether programs should focus on short-term or long-term conservation. For short-term conservation programs, it is recommended to consider high levels of heterozygosity, while long-term strategies should take into account allelic richness and differentiation between breeds. In this study, only the Santa Inês breed was evaluated, which showed relatively high estimates of heterozygosity. However, observed heterozygosity values consistently appeared lower than expected estimates in the samples. This pattern was observed as for the entire analyzed sample as for evaluations conducted by farm. Despite the high levels of diversity found in the breed, this result and the absence of Hardy-Weinberg equilibrium can also have negative implications for long-term conservation. However, local breeders usually do not take genomic information into account for management decisions.

We also highlight the role of mutations on the differentiation among populations, as some genes linked to microsatellite markers may be under artificial selection. This could explain the results of marker

neutrality tests, in which part of the loci had some effect on the selection of mutations of potential microsatellites associated with production-related genes.

Despite having the highest degree of genetic diversity, Farm 1 had the greatest number of private alleles, suggesting the beginning of genetic erosion or dilution (Zhang et al., 2018). On the other hand, Farm 6 had the lowest number of private alleles, despite its lowest genetic diversity. This might indicate that allelic loss has already occurred in this flock, but the small sample size could have compromised the results.

Santa Inês sheep farming has been widely spread throughout Brazil, due to its productive capacity and adaptability. Nevertheless, institutions like the Brazilian Agricultural Research Corporation (Embrapa) have pointed out the risk of genetic erosion of this breed, due to the uncontrolled breeding focusing on short-term production increase to meet market demands. The most likely consequences of this practice are elevated inbreeding levels, loss of fertility, and disease resistance, as well as frequent occurrence of recessive genetic diseases.

The level of differentiation among flocks from Central-Northern Brazil evaluated in this study was similar to that reported by Souza et al. (2012), in Santa Inês sheep from farms located in the Central-Western and Northeastern regions of the country (F_{ST} = 0.058). This suggests that sheep farming practices adopted in Brazil may be promoting the formation of moderate structuring among sheep flocks. Thus, the signs of population substructure found within the farms are also crucial for making decisions in conservation programs (Cañón et al., 2011). The differentiation between genotypes within farms observed in the analysis is probably due to mating among individuals with different genotypes. This corroborates the results generated by STRUCTURE, which revealed the participation of various clusters in the structural array of the flocks. This is also the most plausible explanation for most of the genetic variability of Farm 1 and for the greater distancing of Farm 6.

5. Conclusions

The microsatellites used in this research played a crucial role in assessing the high genetic diversity within sheep populations. Their polymorphic nature allowed the identification of subgroups within samples. This information is vital for conservation efforts, helping to identify, preserve, and improve unique genetic resources.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: Deus, A. R. S.; Sena, L. S.; Carvalho, D. A.; Sousa, F. C. B.; Santos, N. P. S. and Sarmento, J. L. R. **Data curation:** Deus, A. R. S. and Sena, L. S. **Formal analysis:** Sena, L. S. and Rocha, A. O. **Funding acquisition:** Sarmento, J. L. R. **Investigation:** Deus, A. R. S. and Sena, L. S. **Methodology:** Deus, A. R. S.; Silva, G. R.; Sena, L. S.; Britto, F. B.; Rocha, A. O.; Carvalho, D. A.; Santos, N. P. S. and Sarmento, J. L. R. **Project administration:** Deus, A. R. S.; Santos, N. P. S. and Sarmento, J. L. R. **Software:** Deus, A. R. S.; Sena, L. S. and Rocha, A. O. **Supervision:** Santos, N. P. S. **Validation:** Santos, N. P. S. and Sarmento, J. L. R. **Writing – original draft:** Deus, A. R. S. **Writing – review & editing:** Silva, G. R.; Sena, L. S.; Britto, F. B.; Rocha, A. O.; Carvalho, D. A.; Sousa, F. C. B. and Sarmento, J. L. R.

Acknowledgments

The authors thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), which granted the scholarship to the first author, and the Instituto Nacional de Ciência e Tecnologia de Ciência Animal (INCT-CA; process 465377/2014-9). We also thank the Coordenação de Aperfeiçoamento

de Pessoal de Nível Superior (CAPES), Universidade Federal do Piauí (UFPI), and the Santa Inês sheep breeders of Piauí and Maranhão for the support.

References

Abdelkader, A. A.; Ata, N.; Benyoucef, M. T.; Djaout, A.; Azzi, N.; Yilmaz, O.; Cemal, I. and Gaouar, S. B. S. 2018. New genetic identification and characterization of 12 Algerian sheep breeds by microsatellite markers. Italian Journal of Animal Science 17:38-48. https://doi.org/10.1080/1828051X.2017.1335182

Benbouza, H.; Jacquemin, J. M.; Baudoin, J. P. and Mergeai, G. 2006. Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. Biotechnology, Agronomy, Society and Environment 10:77-81.

Caballero, A.; Rodríguez-Ramilo, S. T.; Ávila, V. and Fernández, J. 2010. Management of genetic diversity of subdivided populations in conservation programmes. Conservation Genetics 11:409-419. https://doi.org/10.1007/s10592-009-0020-0

Cañón, J.; García, D.; Delgado, J. V.; Dunner, S.; Telo da Gama, L.; Landi, V.; Martín-Burriel, I.; Martínez, A.; Penedo, C.; Rodellar, C.; Zaragoza, P. and Ginja, C. 2011. Relative breed contributions to neutral genetic diversity of a comprehensive representation of Iberian native cattle. Animal 5:1323-1334. https://doi.org/10.1017/S1751731111000267

Câmara, T. S.; Nunes, J. F.; Diniz, F. M.; Silva, G. R. and de Araújo, A. M. 2017. Genetic diversity and relatedness between Canindé and British Alpine goat breeds in Northeastern Brazil accessed by microsatellite markers. Genetics and Molecular Research 16:gmr16019569. https://doi.org/10.4238/gmr16019569

Chapuis, M. P. and Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24:621-631. https://doi.org/10.1093/molbev/msl191

Dakin, E. E. and Avise, J. C. 2004. Microsatellite null alleles in parentage analysis. Heredity 93:504-509. https://doi.org/10.1038/sj.hdy.6800545

Di Rienzo, A.; Peterson, A. C.; Garza, J. C.; Valdes, A. M.; Slatkin, M. and Freimer, N. B. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences 91:3166-3170. https://doi.org/10.1073/pnas.91.8.3166

Evanno, G.; Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Foll, M. and Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977-993. https://doi.org/10.1534/genetics.108.092221

FAO - Food and Agriculture Organization of the United Nations. 2011. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. FAO, Rome.

Gaouar, S. B. S.; Silva, A.; Ciani, E.; Kdidi, S.; Aouissatk, M.; Dhimi, L.; Lafri, M.; Maftah, A. and Mehtar, N. 2015. Admixture and local breed marginalization threaten Algerian sheep diversity. PLoS ONE 10:e0122667. https://doi.org/10.1371/journal.pone.0122667

Gaouar, S. B. S.; Kdidi, S. and Ouragh, L. 2016. Estimating population structure and genetic diversity of five Moroccan sheep breeds by microsatellite markers. Small Ruminant Research 144:23-27. https://doi.org/10.1016/j.smallrumres.2016.07.021

Gaouar, S. B. S.; Lafri, M.; Djaout, A.; El-Bouyahiaoui, R.; Bouri, A.; Bouchatal, A.; Maftah, A.; Ciane, E. and da Silva, A. B. 2017. Genome-wide analysis highlights genetic dilution in Algerian sheep. Heredity 118:293-301. https://doi.org/10.1038/hdy.2016.86

Ginja, C.; Gama, L. T.; Cortes, O.; Delgado, J. V.; Dunner, S.; García, D.; Landi, V.; Martín-Burriel, I.; Martínez-Martínez, A.; Penedo, M. C. T.; Rodellar, C.; Zaragoza, P.; Cañon, J. and Consortium, B. B. 2013. Analysis of conservation priorities of Iberoamerican cattle based on autosomal microsatellite markers. Genetics Selection 45:35. https://doi.org/10.1186/1297-9686-45-35

Hardy, O. J. and Vekemans, X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2:618-620. https://doi.org/10.1046/j.1471-8286.2002.00305.x

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. Molecular Ecology 17:4015-4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

Jucá, A. F.; Faveri, J. C.; Melo Filho, G. M.; Ribeiro Filho, A. L.; Azevedo, H. C.; Muniz, E. N.; Pedrosa, V. B. and Pinto, L. F. B. 2016. Effects of birth type and family on the variation of carcass and meat traits in Santa Ines sheep. Tropical Animal Health and Production 48:435-443. https://doi.org/10.1007/s11250-015-0971-8

Kalinowski, S. T.; Taper, M. L. and Marshall, T. C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Keenan, K.; McGinnity, P.; Cross, T. F.; Crozier, W. W. and Prodöhl, P. A. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution 4:782-788. https://doi.org/10.1111/2041-210X.12067

Lobo, R. N. B. 2019. Opportunities for investment into small ruminant breeding programmes in Brazil. Journal of Animal Breeding and Genetics 136:313-318. https://doi.org/10.1111/jbg.12396

Meuwissen, T. 2009. Towards consensus on how to measure neutral genetic diversity? Journal of Animal Breeding and Genetics 126:333-334. https://doi.org/10.1111/j.1439-0388.2009.00839.x

Ozmen, O.; Kul, S. and Gok, T. 2020. Determination of genetic diversity of the Akkaraman sheep breed from Turkey. Small Ruminant Research 182:37-45. https://doi.org/10.1016/j.smallrumres.2019.10.009

Peakall, R. and Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Petroli, C. D.; Paiva, S. R.; Paim, T. P. and McManus, C. M. 2014. Association of microsatellite markers with production traits in Santa Inês and crossbred sheep. Archives of Veterinary Science 19:7-16.

Piry, S.; Luikart, G. and Cornuet, J. M. 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. The Journal of Heredity 90:502-503. https://doi.org/10.1093/jhered/90.4.502

Pritchard, J. K.; Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959. https://doi.org/10.1093/genetics/155.2.945

Radha, P.; Sivaselvam, S. N.; Kumarasamy, P. and Kumanan, K. 2011. Genetic diversity and bottleneck analysis of Kilakarsal sheep by microsatellite markers. Indian Journal of Biotechnology 10:52-55.

Ristow, P. G. and D'Amato, M. E. 2017. Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics. Forensic Science International: Genetics Supplement Series. 6:e52-e54. https://doi.org/10.1016/j.fsigss.2017.09.006

Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8:103-106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Selvam, R. and Kathiravan, P. 2019. Genetic diversity and bottleneck analysis of sheep based on microsatellite markers. Indian Journal of Small Ruminants 25:13-18. https://doi.org/10.5958/0973-9718.2019.00012.6

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462. https://doi.org/10.1093/genetics/139.1.457

Souza, C. A.; Paiva, S. R.; McManus, C. M.; Azevedo, H. C.; Mariante, A. S. and Grattapaglia, D. 2012. Genetic diversity and assessment of 23 microsatellite markers for parentage testing of Santa Inês hair sheep in Brazil. Genetics and Molecular Research 11:1217-1229. https://doi.org/10.4238/2012.May.8.4

Van Oosterhout, C.; Hutchinson, W. F.; Wills, D. P. M. and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159. https://doi.org/10.1093/genetics/16.2.97

Yilmaz, O. 2016. Power of different microsatellite panels for paternity analysis in sheep. Animal Science Papers and Reports 34:155-164.

Zhang, M.; Peng, W. F.; Hu, X. J.; Zhao, Y. X.; Lv, F. H. and Yang, J. 2018. Global genomic diversity and conservation priorities for domestic animals are associated with the economies of their regions of origin. Scientific Reports 8:11677. https://doi.org/10.1038/s41598-018-30061-0