PAB
Pesquisa Agropecuária Brasileira

Animal Science/ Scientific Notes

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Validation of a customized subset of SNPs for sheep breed assignment in Brazil

Tiago do Prado Paim[1 ✉], Concepta McManus[2], Fábio Danilo Vieira[3], Stanley Robson de Medeiros Oliveira[3], Olivardo Facó[4], Hymerson Costa Azevedo[5], Adriana Mello de Araújo[6], José Carlos Ferrugem Moraes[7], Michel Eduardo Beleza Yamagishi[3], Paulo Luiz Souza Carneiro[8], Alexandre Rodrigues Caetano[9] and Samuel Rezende Paiva[9]

[1] Instituto Federal de Educação, Ciência e Tecnologia Goiano, Campus Iporá, Avenida Oeste, nº 350, Parque União, CEP 76200-000 Iporá, GO, Brazil. E-mail: tiago.paim@ifgoiano.edu.br

[2] Universidade de Brasília, Instituto de Biologia, Campus Darcy Ribeiro, Asa Norte, CEP 70910-900 Brasília, DF, Brazil. E-mail: concepta@unb.br

[3] Embrapa Informática Agropecuária, Avenida André Tosello, nº 209, Campus da Unicamp, Barão Geraldo, CEP 13083-970 Campinas, SP, Brazil. E-mail: fabio.vieira@embrapa.br, stanley.oliveira@embrapa.br, michel.yamagishi@embrapa.br

[4] Embrapa Caprinos e Ovinos, Fazenda Três Lagoas, Estrada Sobral-Groairas, Km 4, Caixa Postal 71, CEP 62010-970 Sobral, CE, Brazil. E-mail: olivardo.faco@embrapa.br

[5] Embrapa Tabuleiros Costeiros, Avenida Beira Mar, nº 3.250, Jardins, CEP 49025-040 Aracaju, SE, Brazil. E-mail: hymerson.azevedo@embrapa.br

[6] Embrapa Meio-Norte, Avenida Duque de Caxias, nº 5.650, Buenos Aires, Caixa Postal 001, CEP 64008-780 Teresina, PI, Brazil. E-mail: adriana.araujo@embrapa.br

[7] Embrapa Pecuária Sul, Rodovia BR-153, Km 632,9, Vila Industrial, Zona Rural, Caixa Postal 242, CEP 96401-970 Bagé, RS, Brazil. E-mail: jose.ferrugem-moraes@embrapa.br

[8] Universidade Estadual do Sudoeste da Bahia, Departamento de Ciências Biológicas, Avenida José Moreira Sobrinho, Jequiezinho, CEP 45205-490 Jequié, BA, Brazil. E-mail: plscarneiro@gmail.com

Abstract – The objective of this work was to evaluate the usefulness of a subset of 18 single nucleotide polymorphisms (SNPs) for breed identification of Brazilian Crioula, Morada Nova (MN), and Santa Inês (SI) sheep. Data of 588 animals were analyzed with the Structure software. Assignments higher than 90% confidence were observed in 82% of the studied samples. Most of the low-value assignments were observed in MN and SI breeds. Therefore, although there is a high reliability in this subset of 18 SNPs, it is not enough for an unequivocal assignment of the studied breeds, mainly of hair breeds. A more precise panel still needs to be developed for the widespread use in breed assignment.

Index terms: *Ovis aries*, animal genetic resources, certification of origin, genomics, traceability.

Validação de um subconjunto de SNPs específicos para certificação racial de ovinos no Brasil

Resumo – O objetivo deste trabalho foi avaliar a utilidade de um subconjunto de 18 polimorfismos de nucleotídeo único (SNPs) para a certificação das raças de ovinos Crioula Brasileira, Morada Nova (MN) e Santa Inês (SI). Dados de 588 animais foram analisados com o programa Structure. Em 82% dos casos, observou-se designação racial correta com confiança acima de 90%. A maioria dos casos de designação incorreta de raça foi observada em MN e SI. Portanto, apesar de o subconjunto de 18 SNPs ter confiabilidade elevada, ele não é suficiente para a inequívoca certificação das raças estudadas, principalmente das deslanadas. É necessário o desenvolvimento de um painel mais preciso para uso amplo em certificação racial.

Termos para indexação: *Ovis aries*, recursos genéticos animais, certificação de origem, genômica, rastreabilidade.

Precise breed identification is a key step in genetic and genomic studies as accurate breed assignment can improve accuracy of the genomic breeding value estimation, especially when mixed-breed populations are used for developing or applying prediction equations (Kachman et al., 2013; Vandenplas et al., 2016). Moreover, many examples of protected denomination of origin (PDO) and protected geographical indications (PGI) for animal-derived products are directly associated with specific breeds (Dimauro et al., 2015; Mateus & Russo-Almeida, 2015), and proper certification is therefore dependent on the correct identification of livestock breed. Issuing of PDO and PGI certifications, associated with robust methods to monitor marketed

(9) Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, PqEB, Avenida W5 Norte (Final), Caixa Postal 02372, CEP 70770-917 Brasília, DF, Brazil.
E-mail: alexandre.caetano@embrapa.br, samuel.paiva@embrapa.br

✉ Corresponding author

Received
February 1, 2018

Accepted
June 27, 2018

How to cite
PAIM, T. do P.; MCMANUS, C.; VIEIRA, F.D.; OLIVEIRA, S.R. de M.; FACÓ, O.; AZEVEDO, H.C.; ARAÚJO, A.M. de; MORAES, J.C.F.; YAMAGISHI, M.E.B.; CARNEIRO, P.L.S.; CAETANO, A.R.; PAIVA, S.R. Validation of a customized subset of SNPs for sheep breed assignment in Brazil. **Pesquisa Agropecuária Brasileira**, v.54, e00506, 2019. DOI: https://doi.org/10.1590/S1678-3921.pab2019.v54.00506.

animal products have contributed to prevent breed extinctions, mainly in Europe (Di Stasio et al., 2017).

Most Brazilian sheep breeds are considered local genetic resources which are currently facing the challenges associated with uncontrolled crossbreeding (McManus et al., 2010). Hair sheep breeds (as Morada Nova and Santa Inês) are found mainly in the Northeastern Brazil, that is characterized by high heat-stress challenges and is associated with lower-productivity indices. Wool sheep (as Crioula) are reared mainly in the Southern part of the country (McManus et al., 2014). Both regions have great potential for development of PDO and PGI products and depend on inexpensive and accurate methods for breed certification.

As individual animals have low-overall values, and sheep farming in Brazil is performed by small and low-income farmers, the use of low-density SNP panels for breed-assignment to lower-genotyping costs is highly appealing. Therefore, a key goal is the identification of a subset of SNPs (up to 96) that can be used for accurate breed assignment.

Vieira et al. (2015) used information generated with the Ovine SNP50 BeadChip (Illumina Inc., San Diego, CA, USA) to identify a subset of SNPs to differentiate between Crioula, Morada Nova, and Santa Inês. These authors applied three different prediction methods (least absolute shrinkage and selection operator – Lasso, Random Forest, and boosting prediction methods) to select a minimum number of SNP markers for sheep breed identification. They were able to define a set of 18 SNPs able to distinguish samples between these three breeds. However, Vieira et al. (2015) had used a reduced sampling of genotypes from only 72 animals (23 Crioula, 22 Morada Nova, and 27 Santa Inês), whose validation with an independent dataset remains necessary.

The objective of this work was to verify the usefulness of this subset of SNPs previously reported for breed identification of Crioula (BC), Morada Nova (MN), and Santa Inês (SI) sheep.

Samples from 19 BC, 308 MN, and 261 SI animals were genotyped with Ovine SNP50 BeadChip (Illumina Inc., San Diego, CA, USA). The full set of genotypes was used to calculate the genomic relationship matrix for each breed, normalized by an individual marker (GCTA method) (Yang et al., 2011). The average relationship between the animals used by Vieira et al. (2015) (reference population) and the animals evaluated in the present study (validation population) was calculated. The results showed a low relationship between animals from the two datasets (Crioula, 0.029±0.132 (mean±standard deviation); Morada Nova, 0.012±0.049; and Santa Inês, 0.008±0.053).

The eighteen SNPs selected by Vieira et al. (2015) were extracted from the dataset, and minor allele frequencies (MAF) were determined for each breed (Table 1). As the minor allele can be different from one breed to another, and can differ between the two datasets, contrasts were performed between breeds and studies. Only one SNP in Santa Inês (s32131) and one in Morada Nova (s69653) differed in minor allele between the reference population (Vieira et al., 2015) and the validation population used in the present study.

The Structure software version 2.3.4 (Pritchard et al., 2000) was used to estimate individual allocation probabilities in each of the three breeds. The definition of clusters was based on the admixture model and assumption that allele frequencies were correlated between breeds. Run parameters were as follows: 588 individuals; 18 loci, without a priori information of populations; length of burn-in period of 10,000; and 200,000 repetitions after burn-in for Markov Chain Monte Carlo (MCMC). The number of clusters (K) was set to 2, 3, 4, and 5, with five runs for each cluster. Following the method of Evanno et al. (2005), the best K was 3, which agrees with breeds in the data, and shows that this extremely small panel is able to identify

this structure in the samples. Thereafter, we used the results for K=3 to evaluate the correct classification rate.

The percentage of individuals classified in each cluster was determined by the estimated proportion of the association of each individual genotype to each of the clusters. Tests of individual allocation were performed with and without a priori information about the source population of individuals, yielding similar outcomes. Therefore, results without a priori information were used, as they represent a real situation of breed assignment analyses more properly, since there is no previous knowledge or information about the sample.

Accurate breed assignments (confidence >90%) were observed in 89, 86, and 75% of BC, MN, and SI animals, respectively. Mean cluster allocation values ranged from 90.9 to 93.7% (Table 2). SI has been previously shown to have been formed by crossbreeding of MN, Bergamasca, and Somalis (McManus et al., 2010). MN and SI animals were observed to have some degree of admixture and estimated fixation index (Fst) of 6.59% (Genome-wide..., 2012). Therefore, some allocation errors between MN and SI were expected. Nonetheless, high levels of correct breed allocation (>90%) were observed.

The results obtained here using 18 SNPs were less accurate than those of previous studies, most likely because of the higher-information content of microsatellite markers compared to SNPs, and the great difference in number of SNPs used. SNPS for parentage... (2014) identified a set of 163 SNPs for accurate parentage testing and traceability, in many of the world's main sheep breeds. Mateus & Russo-Almeida (2015) identified 12 microsatellite markers able to correctly classify animals into their respective breeds, while Di Stasio et al. (2017) used 15 microsatellite markers for breed certification in Italian sheep breeds. Other studies (Bertolini et al., 2015; Dimauro et al., 2015) showed that at minimum of 100 SNPs are required for correct and accurate breed assignment of cattle and sheep breeds.

Table 1. Minor allele frequency estimates for each SNP marker used in the analyses, for each breed (Crioula, Morada Nova, and Santa Inês), and the datasets "Reference", according to Vieira et al. (2015), and "Validation", from present study.

Marker	Chromosome	Crioula				Morada Nova				Santa Inês			
		Reference		Validation		Reference		Validation		Reference		Validation	
		Minor	MAF	Minor	MAF	Minor	MAF	Minor	MAF	Minor	MAF	Minor	MAF
s03528.1	1	A	0.435	A	0.425	A	0.227	A	0.460	G	0.074	G	0.188
OAR1_194627962.1	1	?	0.000	G	0.025	A	0.273	A	0.387	G	0.043	G	0.106
OAR2_55853730.1	2	C	0.152	C	0.353	?	0.000	A	0.047	A	0.106	A	0.111
s20468.1	2	A	0.152	A	0.225	?	0.000	A	0.032	G	0.277	G	0.194
OAR3_164788310.1	3	G	0.217	G	0.275	G	0.182	G	0.268	A	0.149	A	0.278
s16949.1	3	G	0.152	G	0.200	G	0.182	G	0.252	A	0.160	A	0.295
s69653.1	3	G	0.087	G	0.150	G	0.364	A	0.311	A	0.106	A	0.175
OAR3_165050963.1	3	A	0.022	A	0.100	A	0.068	A	0.055	G	0.202	G	0.322
s32131.1	4	A	0.326	A	0.316	G	0.023	G	0.269	G	0.500	A	0.381
s06182.1	5	A	0.152	A	0.150	G	0.068	G	0.211	A	0.415	A	0.423
OAR15_45152619.1	15	A	0.239	A	0.300	G	0.023	G	0.029	G	0.053	G	0.056
s30024.1	25	A	0.087	A	0.100	C	0.023	C	0.144	C	0.277	C	0.257
s61697.1	X	C	0.065	C	0.100	C	0.045	C	0.097	A	0.319	A	0.222
OARX_29830880.1	X	G	0.196	G	0.200	?	0.000	A	0.026	A	0.074	A	0.157
OARX_53305527.1	X	?	0.000	A	0.079	A	0.091	A	0.021	G	0.277	G	0.226
s56924.1	X	G	0.022	G	0.075	A	0.136	A	0.197	A	0.160	A	0.103
OARX_78903642.1	X	G	0.043	G	0.200	A	0.068	A	0.013	A	0.096	A	0.098
OARX_121724022.1	X	A	0.022	A	0.150	C	0.023	C	0.008	C	0.085	C	0.151

Minor: minor allele in each breed and dataset. MAF: minor allele frequency. The two changes in minor alleles observed between the two datasets is highlighted in gray.

Table 2. Mean cluster allocation of Crioula (BC), Morada Nova (MN), and Santa Inês (SI) sheep obtained with the Structure analysis of data from 18 SNP markers.

Population	Inferred cluster			Number of individuals
	1	2	3	
BC	0.929	0.012	0.059	19
MN	0.021	0.937	0.041	308
SI	0.034	0.056	0.909	261

The 18 SNP panel tested showed 90% correct assignment of the studied breeds. Incorrect assignments ranged between 6 to 9% of the animals (Table 2). Ideally, a system for breed certification requires a correct allocation close to 100% with minimal incorrect assignment. The SNP panel tested showed high levels of correct assignment; however, the obtained results are not enough for its widespread use for breed certification.

The construction and validation of a larger panel with additional SNPs could provide higher correct assignment rates (close to 100%) for other major sheep breeds reared in Brazil, which may contribute to breed identification and certification procedures. Thereupon, this tool could be incorporated in routine inspection services and ongoing genetic improvement and conservation activities.

Acknowledgments

To Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), for financial support; to Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano) and to International Sheep Genomics Consortium, for technical and logistics support; and to Embrapa multiuser bioinformatics laboratory (Laboratório Multiusuário de Bioinformática da Embrapa) for permitting the use of its high-performance computational infrastructure.

References

BERTOLINI, F.; GALIMBERTI, G.; CALÒ, D.G.; SCHIAVO, G.; MATASSINO, D.; FONTANESI, L. Combined use of principal component analysis and random forests identify population-informative single nucleotide polymorphisms: application in cattle breeds. **Journal of Animal Breeding and Genetics**, v.132, p.346-356, 2015. DOI: https://doi.org/10.1111/jbg.12155.

DI STASIO, L.; PIATTI, P.; FONTANELLA, E.; COSTA, S.; BIGI, D.; LASAGNA, E.; PAUCIULLO, A. Lamb meat traceability: The case of Sambucana sheep. **Small Ruminant Research**, v.149, p.85-90, 2017. DOI: https://doi.org/10.1016/j.smallrumres.2017.01.013.

DIMAURO, C.; NICOLOSO, L.; CELLESI, M.; MACCIOTTA, N.P.P.; CIANI, E.; MOIOLI, B.; PILLA, F.; CREPALDI, P. Selection of discriminant SNP markers for breed and geographic assignment of Italian sheep. **Small Ruminant Research**, v.128, p.27-33, 2015. DOI: https://doi.org/10.1016/j.smallrumres.2015.05.001.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology**, v.14, p.2611-2620, 2005. DOI: https://doi.org/10.1111/j.1365-294X.2005.02553.x.

GENOME-WIDE analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. **PLoS Biology**, v.10, e1001258, 2012. DOI: https://doi.org/10.1371/journal.pbio.1001258.

KACHMAN, S.D.; SPANGLER, M.L.; BENNETT, G.L.; HANFORD, K.J.; KUEHN, L.A.; SNELLING, W.M.; THALLMAN, R.M.; SAATCHI, M.; GARRICK, D.J.; SCHNABEL, R.D.; TAYLOR, J.F.; POLLAK, E.J. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. **Genetics Selection Evolution**, v.45, p.1-9, 2013. DOI: https://doi.org/10.1186/1297-9686-45-30.

MATEUS, J.C.; RUSSO-ALMEIDA, P.A. Traceability of 9 Portuguese cattle breeds with PDO products in the market using microsatellites. **Food Control**, v.47, p.487-492, 2015. DOI: https://doi.org/10.1016/j.foodcont.2014.07.038.

MCMANUS, C.; HERMUCHE, P.; PAIVA, S.R.; MORAES, J.C.F.; DE MELO, C.B.; MENDES, C. Geographical distribution of sheep breeds in Brazil and their relationship with climatic and environmental factors as risk classification for conservation. **Brazilian Journal of Science and Technology**, v.1, p.1-15, 2014. DOI: https://doi.org/10.1186/2196-288X-1-3.

MCMANUS, C.; PAIVA, S.R.; ARAÚJO, R.O. de. Genetics and breeding of sheep in Brazil. **Revista Brasileira de Zootecnia**, v.39, p.236-246, 2010. Suplemento especial. DOI: https://doi.org/10.1590/S1516-35982010001300026.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v.155, p.945-959, 2000.

SNPs FOR PARENTAGE testing and traceability in globally diverse breeds of sheep. **PLoS ONE**, v.9, e94851, 2014. DOI: https://doi.org/10.1371/journal.pone.0094851.

VANDENPLAS, J.; CALUS, M.P.L.; SEVILLANO, C.A.; WINDIG, J.J.; BASTIAANSEN, J.W.M. Assigning breed origin to alleles in crossbred animals. **Genetics Selection Evolution**, v.48, p.1-22, 2016. DOI: https://doi.org/10.1186/s12711-016-0240-y.

VIEIRA, F.D.; OLIVEIRA, S.R. de M.; PAIVA, S.R. Metodologia baseada em técnicas de mineração de dados para suporte à certificação de raças de ovinos. **Engenharia Agrícola**, v.35, p.1172-1186, 2015. DOI: https://doi.org/10.1590/1809-4430-Eng.Agric.v35n6p1172-1186/2015.

YANG, J.; LEE, S.H.; GODDARD, M.E.; VISSCHER, P.M. GCTA: a tool for genome-wide complex trait analysis. **American Journal of Human Genetics**, v.88, p.76-82, 2011. DOI: https://doi.org/10.1016/j.ajhg.2010.11.011.