GMR

Red Sindhi cattle in Brazil: population structure and distribution

J.C.C. Panetto[1], M.V.G.B. Silva[1], R.M.H. Leite[2], M.A. Machado[1], F.A.T. Bruneli[1], D.R.L. Reis[1], M.G.C.D. Peixoto[1] and R.S. Verneque[1,3]

[1]Embrapa Gado de Leite, Juiz de Fora, MG, Brasil
[2]EMEPA, João Pessoa, PB, Brasil
[3]EPAMIG, Belo Horizonte, MG, Brasil

Corresponding author: J.C.C. Panetto
E-mail: joao.panetto@embrapa.br

ABSTRACT. The Red Sindhi cattle breed was imported to Brazil in small numbers. Nowadays, the herds of this breed are distributed in the Northeast, Southeast and Midwest regions of the country. In this study, DNA samples of animals originating from 15 herds in the Northeast and Southeast regions have been analyzed to obtain the ancestry proportions, and to gain a better understanding of the current population structure of this breed in Brazil. Samples were genotyped using three different single nucleotide polymorphism (SNP) marker panels. Those markers have been used with the approach of unsupervised hierarchical clustering of individuals, and consequently, the ancestry of the population was divided into six different subpopulations. Three of those ancestry subpopulations were identified to be present in various different herds, while the other three were restricted to only one or two herds each. One of those herds has been kept isolated for more than 30 years, and it was identified to contain two almost exclusive subpopulations. To avoid important losses in the genetic diversity within the Red Sindhi breed in Brazil, we recommend the identification

of superior sires from every subpopulation in the establishment of a breeding program for this breed.

INTRODUCTION

The Red Sindhi is a Zebu (*Bos primigenius indicus*) breed named after its region of origin, which is located in the northern part of the Sindh Province of Pakistan, including Tando Jam, Maleer, and Dadu districts. The climate in this region is extremely hot and dry, with the maximum temperature frequently rising above 46°C, and an average annual rainfall of approximately 150 to 180 mm. This breed is especially valued because of its milk production under grazing systems and high tolerance to heat stress and parasites. Purebred herds of this breed can also be found in India and other countries, such as Brazil, which have imported animals from Pakistan.

According to Leite et al. (2001), some red Zebu animals were selected among many other cattle being imported from India in 1930. They were bred in the state of São Paulo for about two decades without any breed classification. However, later on, those animals were identified as belonging to the Red Sindhi breed. On a second occasion, 28 female and three male cattle were officially selected from purebred Red Sindhi herds in Pakistan and imported to Brazil, arriving to the island of Fernando de Noronha in 1952. Thus, the Brazilian Red Sindhi herds are based exclusively on these two small founder populations.

Today, most of the Red Sindhi herds are located in the Northeast and Southeast regions of Brazil. In the Northeast region, harsh environmental conditions predominate, with high average temperatures and very low precipitation. The Red Sindhi cattle have been well adapted to such conditions, and most breeders claim that the cows can maintain good body condition scores and good fertility even in such harsh environments. This breed is generally considered suitable for the dual-purpose production systems, which produce both dairy and beef. However, selection in the Southeastern herds has been more focused on beef traits, in contrast to the Northeastern herds, in which the selection was performed mostly for dairy traits.

The increasing demand for the Red Sindhi cattle in low-input systems in Brazil can represent a risk for the maintenance of the genetic diversity within this breed, because of the small number of operating seedstock herds in the country. As the number of founders is also known to be small, one way to avoid the loss of genetic diversity would be the use of optimum contribution selection schemes, as pointed out by many authors (Sonesson and Meuwissen, 2000; Koenig and Simianer, 2006; Schierenbeck et al., 2011). Selection schemes with optimum contribution approaches would primarily demand knowledge about the genetic divergences and subpopulations. Nowadays, those parameters can be more accurately assessed with the use of genomic tools.

Population structure, genetic divergence, and ancestry have been assessed with the use of single nucleotide polymorphism (SNP) markers in several breeds and admixed populations (McKay et al., 2008; Gautier and Naves, 2011; Porto-Neto et al., 2013).

The objective of this study was to determine the population structure of the Red Sindhi breed in Brazil using genomic information.

MATERIAL AND METHODS

Population distribution amongst Brazilian regions

The genealogical information of the breed, including 28,134 animals born from 1950 to 2013, was obtained from the files of the Brazilian Association of Zebu Breeders (ABCZ).

The distribution of animals among the regions of Brazil was determined from birth records dating from 2006 to 2013, corresponding to the most recent period of one-generation interval (~8 years) for this population. The proportions of numbers of animals in each region of Brazil, according to the origin of the animals of the recent generation, or according to the origin of the sires that were fathers in the recent generation, are shown in Table 1.

Table 1. Distribution of the Red Sindhi population in Brazil: animals born from 2006 to 2013.

Brazilian Region	Proportion of origin for animals (%)	Proportion of origin for sires (%)
Northeast	56.6	68.5
Southeast	36.6	31.5
Midwest	6.5	0.0
North	0.3	0.0
South	0.0	0.0

In the Northeast, the main Brazilian states were Paraíba and Rio Grande do Norte, representing 35 and 13% of the current national herd size, respectively. In the Southeast, the main states were Minas Gerais and São Paulo, representing 21 and 15% of the current national herd size, respectively.

DNA sampling and genotyping

Blood samples, or semen, were collected from animals originating from 15 herds in the Brazilian states of São Paulo (3) and Minas Gerais (2) in the Southeast region, and Pernambuco (1), Paraíba (7), and Rio Grande do Norte (2) in the Northeast region. The goal was to obtain a comprehensive representation of the Red Sindhi breed in Brazil. This sampling strategy aimed to reach the main bulls that were sires in the current generation and animals from the most traditional breeding herds in the country. In cases where the biological samples from important bulls were not available, a progeny from them was included instead. Initially, a total of 315 animals were sampled to provide DNA for SNP chip genotyping. The herds included for sampling in this study represented approximately 51% of the births and, at the same time, 89% of the bulls that were sires in the most recent generation of the whole Red Sindhi population in Brazil. Figure 1 shows a map with the location of the herds included in this study.

To obtain a sample that represents the majority of the current breeding herds of the Red Sindhi breed in Brazil, 128 animals, out of the 315 initially sampled animals, were selected to be included in the genomic approach for the population structure study. The preferred samples were those from the main reference sires, and the proportion of samples from different seedstock herds was thought to be similar to the proportion of animals from these herds in the Brazilian population.

DNA samples were analyzed using three different SNP marker panels: 25 animals were genotyped with the Illumina BovineHD BeadChip (HD; Illumina, San Diego, CA, USA);

Figure 1. Location of the Brazilian Red Sindhi herds included in the study. Each blue point, identified with a letter, corresponds to the location of a herd of origin of the animals, or the location of a herd from which the samples were collected.

232 with the BovineSNP50 BeadChip (50K; Illumina); and 58 with the GeneSeek SGGP-20Ki; (Geneseek, Lincoln, NE, USA). Only autosomal SNPs present simultaneously in the three different chips were used. The quality control (QC) procedures included the exclusion of any SNP with minor allele frequency (MAF) <0.04 or genotyping call rate <0.98. After QC approval, 4466 SNP markers were selected and used in all samples. Only individuals with genotyping call rate > 0.90 were retained in the analysis. The GenCall score (GC) is a confidence measure assigned to each call that can be used to eliminate any poor quality calls, SNPs, or samples. Genotypes with GC < 0.70 were marked as missing.

Ancestry proportions and genetic diversity among subpopulations

Global ancestry proportions were obtained on the basis of the maximum likelihood approach, implemented with the software Admixture (Alexander et al., 2009) while the genetic diversity estimates were obtained using F-Statistics (Weir and Cockerham, 1984).

One of the assumptions for this approach, as pointed out by its authors, was that markers should be independent from each other; in other words, the markers should be in linkage disequilibrium. Different limits for the pruning of markers were tested for verifying the robustness of the model regarding the inferred numbers of subpopulations within this population. This was implemented with the command "--indep-pairwise 50 10 0.15" within the PLINK software (Purcell et al., 2007).

Boison et al. (2015), in their study with the zebu Gyr cattle in Brazil, indicated that a 100 kb distance between markers corresponded to an R^2 minimum of 0.15 for linkage disequilibrium. In the present study, the minimum R^2 (0.15) was chosen as the threshold for eliminating linkage disequilibrium in the set of markers used on the determination of ancestry proportions.

RESULTS AND DISCUSSION

Population distribution amongst Brazilian regions

In Brazil, animals from the Red Sindhi breed were mainly raised in the Northeast, followed by the Southeast and Midwest regions. Through comparison of the origins of current-generation animals with the origins of the current-generation sires, it could be observed that the Red Sindhi population was growing more rapidly in the Southeast and Midwest regions in relation to the Northeast region.

Cross-validation error and the number of subpopulations

In the population structure analysis of the Red Sindhi breed in Brazil, the observed cross-validation errors, according to the numbers of assumed subpopulations at different thresholds for the correlation coefficients (R^2) among markers, are shown in Figure 2.
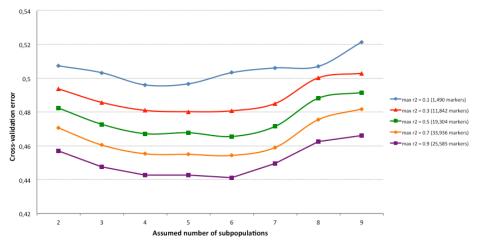

Figure 2. Cross-validation errors for various numbers of assumed subpopulations at different R^2 (correlation coefficient) thresholds for the pruning of markers in the population structure analysis of the Red Sindhi breed in Brazil.

The testing of different correlation limits for the pruning of markers - with R^2 thresholds ranging from 0.10 to 0.90 - led to the detection of local minimum cross-validation errors when six subpopulations were considered as the structure of the population, in four out of the five different values tested, but always within the range of four to six. Thus, in our study, the inferred number of subpopulations showed to be robust regarding different levels of pruning for linkage disequilibrium.

With the application of a R^2 threshold of 0.15 for the maximum correlation between markers, following the example of the study of Boison et al. (2015), the minimum cross-validation error was also detected for the inferred number of subpopulations equal to six. Same results were found with more permissive thresholds and consequently bigger numbers of markers used. Thus, with the observed robustness of the model, we decided to follow the recommendation of thinning the marker set for linkage disequilibrium, from the Admixture software authors (Alexander et al., 2009), and we used this pruned file, which contained 1,775 markers in the panel, for the following admixture analysis.

Porto-Neto et al. (2013) applied an R^2 threshold of 0.1 for the same purpose. McKay et al. (2008) have also tested different numbers of SNP markers, although using a random sampling instead of the pruning for linkage disequilibrium. They concluded that the number of SNP markers should be greater than 100, which was completely fulfilled in the panel defined with the methodology based on linkage disequilibrium used in this study.

Subpopulations

Ancestry proportions for each of the 128 animals sampled were plotted in Figure 3. For this figure, the sampled animals were ordered according to the major subpopulation in their ancestry and, within each subpopulation, according to the herds they belonged to. Subpopulations 1 to 6 were located from left to right in this Figure.
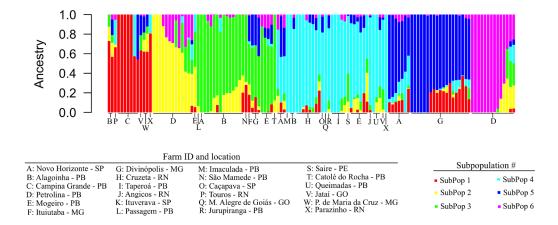

Farm ID and location				Subpopulation #	
A: Novo Horizonte - SP	G: Divinópolis - MG	M: Imaculada - PB	S: Saire - PE	SubPop 1	SubPop 4
B: Alagoinha - PB	H: Cruzeta - RN	N: São Mamede - PB	T: Catolé do Rocha - PB		
C: Campina Grande - PB	I: Taperoá - PB	O: Caçapava - SP	U: Queimadas - PB	SubPop 2	SubPop 5
D: Petrolina - PB	J: Angicos - RN	P: Touros - RN	V: Jataí - GO		
E: Mogeiro - PB	K: Ituverava - SP	Q: M. Alegre de Goiás - GO	W: P. de Maria da Cruz - MG	SubPop 3	SubPop 6
F: Ituiutaba - MG	L: Passagem - PB	R: Jurupiranga - PB	X: Parazinho - RN		

Figure 3. Ancestry proportions of 128 animals sampled from the Red Sindhi population in Brazil indicated according to their order of subpopulation and herd. Different colors represent different subpopulations. Each letter represents a different herd.

In Figure 3, it can be observed that the subpopulations 1, 3, and 4 were fairly widespread among various herds, originating from the three different regions included in the study. However, the subpopulations 2, 5, and 6 were relatively more concentrated in a smaller number of herds. Furthermore, animals with their major ancestry in subpopulations 1, 3, or 4 could be found in six, eight, and fourteen different herds, respectively. On the other hand, animals with their major ancestry in subpopulations 2, or 5 could be found in two herds each, while those with major ancestry in subpopulation 6 could be found in only one herd. It is important to note that the subpopulations 2 and 6 were almost exclusively found in the same herd (D), in the Northeast region, whereas subpopulation 5 was concentrated in two herds of the Southeast region.

Composition of herds

The ancestry proportions of the animals ordered according to their herds of origin are illustrated in Figure 4. Animals from the herds A, D, B, H, F, G, I, E, J, K, and C were located from left to right in this order. Subpopulation 1 was mainly associated with herd C, with a significant relation to herd B, both originating from the Northeast region, but also with some extension to other herds. Subpopulation 2 was mainly associated with herd D, with some relation to herd B and two animals from herd E, all in the Northeast region. Subpopulation 3 was mainly associated with the herds B and E, both in the Northeast region and with some animals in herd G, in the Southeast region. Subpopulation 4 was mainly associated with the herds I, H, and J, all of them in the Northeast region, but with some extension to other herds. Subpopulation 5 was mainly associated with the herds G, K, and A, all of them in the Southeast region, but with some extension to other herds. Subpopulation 6 was found to be exclusively originated from herd D, in the Northeast region.
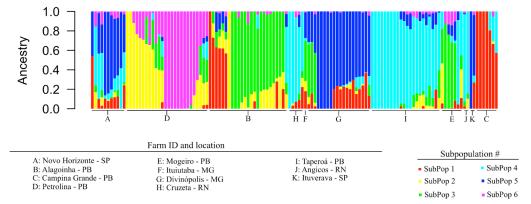

Figure 4. Ancestry proportions of 128 animals sampled from the Red Sindhi population in Brazil. Animals from herds A, D, B, H, F, G, I, E, J, K, and C were located from left to right in this order. Different colors represent different subpopulations.

Population structure

F_{ST} statistics presented in Table 2 show the genetic distances between each pair of subpopulations.

Table 2. Pairwise F_{ST} between subpopulations (below the diagonal) and average pairwise F_{ST} (in the diagonal).

	Subpop. 1	Subpop.2	Subpop. 3	Subpop. 4	Subpop. 5	Subpop. 6
Subpop. 1	0.177					
Subpop. 2	0.215	0.196				
Subpop. 3	0.160	0.167	0.155			
Subpop. 4	0.151	0.180	0.131	0.153		
Subpop. 5	0.144	0.189	0.146	0.119	0.158	
Subpop. 6	0.216	0.228	0.171	0.182	0.190	0.197

Among the six subpopulations of the Red Sindhi breed in Brazil, subpopulations 2 and 6 were found to be the most detached ones, in relation to their average pairwise distances. The largest pairwise distances were observed between the subpopulations 2 and 6 (F_{ST} = 0.228), followed by 1 and 6 (F_{ST} = 0.216), and 1 and 2 (F_{ST} = 0.215). The smallest pairwise distances were observed between the subpopulations 4 and 5 (F_{ST} = 0.119), followed by subpopulations 3 and 4 (F_{ST} = 0.131). Some examples of pairwise FST estimated from SNP loci are available in the literature. McKay et al. (2008) found values ranging from 0.015 to 0.373 among eight different breeds of cattle. In the work of Gautier and Naves (2011), the pairwise F_{ST} values ranged from 0.028 to 0.175 among some populations of European and African taurine and Zebu cattle breeds.

Genetic diversity in the future

One possible strategy for the maintenance of genetic diversity in the Red Sindhi population, focusing also on the establishment of its breeding program, is to rely on the systematic use of reference sires from every subpopulation. A progeny test could be run with this approach. Any practical application of optimum contribution schemes, as proposed by Sonesson and Meuwissen (2000), could be facilitated with the knowledge of the population structure, as the superior sires from different subpopulations could be identified as mating options for the females in specific herds. In this regard for the Brazilian Red Sindhi population, special attention should be given to the subpopulations 2 and 6, which were both found almost exclusively in only one herd (D). Furthermore, subpopulations 2 and 6 were the most genetically detached from the other subpopulations, and the use of sires from those subpopulations would contribute to the maintenance of the genetic diversity in future generations.

CONCLUSION

The ancestry of the Brazilian Red Sindhi population has been identified as belonging to six different subpopulations, originating from the Northeast or the Southeast regions of the country. Three of these subpopulations were identified to be currently associated with various different herds, while the other three were restricted to only one or two herds. One of those herds was identified to contain two almost exclusive subpopulations. To avoid important losses in the genetic diversity, we recommend the identification of superior sires from every subpopulation in the establishment of a breeding program for this breed in Brazil.

Conflicts of interest

The authors declare no conflicts of interest.

_____

ACKNOWLEDGMENTS

REFERENCES

Alexander DH, Novembre J and Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19: 1655-1664. http://dx.doi.org/10.1101/gr.094052.109

Boison SA, Santos DJA, Utsunomiya AHT, Carvalheiro R, et al. (2015). Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (*Bos indicus*) dairy cattle: Comparison of commercially available SNP chips. *J. Dairy Sci.* 98: 4969-4989. http://dx.doi.org/10.3168/jds.2014-9213

Gautier M and Naves M (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20: 3128-3143. http://dx.doi.org/10.1111/j.1365-294X.2011.05163.x

Koenig S and Simianer H (2006). Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. *Livest. Sci.* 103: 40-53. http://dx.doi.org/10.1016/j.livsci.2005.12.009

Leite PRM, Santiago AA, Navarro Filho HR, Albuquerque RPF, et al. (2001). Sindi: gado vermelho para o semi-árido. Banco do Nordeste, João Pessoa, EMEPA-PB, Brazil.

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, et al. (2008). An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genet.* 9: 37-45. http://dx.doi.org/10.1186/1471-2156-9-37

Porto-Neto LR, Sonstegard TS, Liu GE, Bickhart DM, et al. (2013). Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14: 876-887. http://dx.doi.org/10.1186/1471-2164-14-876

Purcell S, Neale B, Todd-Brown K, Thomas L, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81: 559-575. http://dx.doi.org/10.1086/519795

Schierenbeck S, Pimentel ECG, Tietze M, Körte J, et al. (2011). Controlling inbreeding and maximizing genetic gain using semi-definite programming with pedigree-based and genomic relationships. *J. Dairy Sci.* 94: 6143-6152. http://dx.doi.org/10.3168/jds.2011-4574

Sonesson AK and Meuwissen THE (2000). Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genet. Sel. Evol.* 32: 231-248. http://dx.doi.org/10.1186/1297-9686-32-3-231

Weir BS and Cockerham CC (1984). Estimating F-Statistics for the analysis of population structure. *Evolution* 38: 1358-1370. http://dx.doi.org/10.2307/2408641