Understanding the structure of the Brazilian Red Sindhi population using genomic information

***J.C.C. Panetto*[*], R.M.H. Leite[†], G.G. Santos[*], F.A.T. Bruneli[*], R.B. Teixeira[‡§], L.G. Castro[*#], D.R.L. Reis[*], M.A. Machado[*], M.G.C. Peixoto[*], R.S. Verneque[*]**

[*]Embrapa Dairy Cattle, Juiz de Fora, MG, Brazil; [†]Paraiba Agricultural State Research Company, João Pessoa, PB, Brazil; [‡]Visiting Scholar grant - CNPq; [§]Minas Gerais Federal Institute, Bambuí, MG, Brazil; [#]Undergraduate grant – FAPEMIG

ABSTRACT: The Red Sindhi is a zebu cattle breed original from Pakistan. A small number of animals were imported to Brazil about the middle of the 20th century. Nowadays, herds are concentrated in Northeast and Southeast regions, with a growing demand from beef and dairy producers. The objective of this study was to determine the structure of subpopulations for the Red Sindhi population in Brazil using genomic information. Genotypes of 218 animals from 15 herds, comprising a set of 20,532 SNP markers, were analyzed to estimate proportions of ancestry through unsupervised hierarchical clustering of individuals, using the maximum likelihood method. The number of inferred subpopulations was 11. Some of those subpopulations were restricted to only one present herd, leading to the conclusion that some optimum contribution selection scheme should be adopted to avoid possible future shrinking on the genetic diversity within this breed.
Keywords: admixture; dairy cattle; genetic diversity; SNP

Introduction

The Red Sindhi is a zebu (*Bos primigenius indicus*) breed named after its region of origin, which is the North part of the Sindh province of Pakistan. Climate in that region is very hot and very dry, with maximum temperatures frequently rising above 46 ºC, and average rainfall around 150 to 180 mm per year. Purebred herds of this breed can also be found in India and many other countries that imported animals from Pakistan.

Brazil also imported Red Sindhi animals. In this country, those animals have been demonstrating good adaptation to a variety of environmental conditions. Figure 1 is a photograph showing Red Sindhi animals in a semi-arid environment in Brazil.

In the year 1930, some red zebu animals were selected among many other cattle being imported at that moment. They were kept breeding in the State of São Paulo, for about two decades without any breed classification. Later, those animals were identified as belonging to the Red Sindhi breed. In a second moment, there was an official importation of 28 females and 3 males selected from purebred Red Sindhi herds in Pakistan, which arrived to the Brazilian island of Fernando de Noronha in the year 1952 (Leite et al. (2001)). The establishment of the Brazilian Red Sindhi herds was based on these two small founder populations only.

Today, most herds are located in the Northeast and Southeast regions of Brazil. This breed is generally considered suitable to be used in dual purpose production systems, dairy and beef. However, selection in the Southeast herds was more emphasized on beef traits, in contrast to the Northeast herds that selected mostly for dairy traits. In the Northeast of Brazil there is a predominance of harsh environmental conditions, with high average temperatures and very low precipitation. The Red Sindhi has been well adapted to those conditions, and most breeders claim that cows can keep good body condition scores and good fertility even in these very harsh environments.

Figure 1 – Red Sindhi cow and calf in a Brazilian semi-arid region.

With an elevated demand for animals capable of producing under harsh environmental conditions, and the small number of Red Sindhi seedstock herds in Brazil, genetic diversity within this breed needs to be monitored. As the number of founders is known to be small, one way to avoid loss of genetic diversity would be the use of optimum contribution selection schemes, as pointed by Sonesson and Meuwissen (2000), and many other authors. Selection schemes with optimum contribution approaches would primarily demand knowledge about genetic divergences and subpopulations.

Population structure, genetic divergences and ancestry have been accessed with the use of increasing numbers of SNP markers, in several breeds and admixed popu-

lations (McKay et al. (2008); Gautier and Naves (2011); Porto-Neto et al. (2013)). The objective of this study was to determine the structure of subpopulations for the Red Sindhi population in Brazil using genomic information.

Materials and Methods

Genealogical information was obtained at the Brazilian Association of Zebu Breeders (ABCZ), which keeps records of most purebred animals from the Red Sindhi and other zebu breeds in Brazil. Those genealogical information were used in the definition of herds and animals to be included in this study, with the objective of obtaining a sample with a comprehensive representation of this breed in Brazil. One exception was made for one herd, in the State of Pernambuco, in which the animals were not registered at ABCZ, but they were considered purebred since they were originated from the official Red Sindhi importation in 1952. This herd was maintained closed since then.

Blood samples, or semen, were collected from animals originating from 15 herds in the Brazilian States of São Paulo (3), Minas Gerais (2), Pernambuco (1), Paraíba (7) and Rio Grande do Norte (2). The sampling strategy aimed to sample animals from most traditional seed stock herds in the country and sample the main sires of the current generation as well. In cases were biological samples from important sires were not available, progeny from these sires were collected instead. A total of 219 animals were sampled to provide DNA for molecular markers genotyping.

Most samples (n=194) were genotyped using the BovineSNP50K BeadChip (Illumina, Inc. San Diego, USA). A total of 25 samples were genotyped using the BovineHD BeadChip that includes ~777 K SNPs (Illumina, Inc. San Diego, USA). These samples were genotyped with this HD chip because these same animals were included in another study of our group. Only autosomal SNPs that were common to the HD and the 50k chips were used in the present study. Quality control procedures included the exclusion of any SNP with MAF < 0.05 or genotyping call rate < 0.98. Only individuals with genotyping call rate > 0.90 were kept. Genotypes with GCScore < 0.70 were marked as missing.

Proportions of ancestry from each contributing subpopulation were estimated using genomic information, with SNP markers selected as previously indicated, under a model-based approach. This was implemented through unsupervised hierarchical clustering of individuals, using the maximum likelihood method implemented in the program Admixture (Alexander et al. (2009)). Default input parameters were considered (i.e.quasi-Newton convergence acceleration method and termination criterion of $<10^{-4}$ for the log-likelihood increase between successive iterations). Linkage Disequilibrium (LD) was used as an additional criterion for the selection of the SNP markers, since the model applied with the program Admixture does not explicitly take LD into consideration. Different levels of pruning have been tested for the dataset, based on LD between markers, using the PLINK (Purcell et al. (2007)) command "--indep-pairwise", with a slide window of 50 SNP, a step of 10 SNP and maximum r^2 values ranging from 0.50 to 0.98. The number of inferred subpopulations (K) was obtained with the cross-validation procedure of the program Admixture.

Results and Discussion

Out of 54,609 SNP, from the 50K bovine chip, 52,886 (autosomal) were kept in the study. After merging files from animals genotyped with 50K or HD chips, there were 21,498 SNP common to both chips after the restrictions of MAF > 0.05 and Genotyping rate > 0.98. The numbers of SNP markers remaining after exclusion according to maximum r^2 values for sample correlation coefficients, indicating Linkage Disequilibrium (LD), were 15,540 (r^2<0.50), 19,113 (r^2<0.80), 20,532 (r^2<0.95) or 20,856 (r^2<0.98).

Cross validation error plots, according to the number of subpopulations, at different levels of restriction for LD are shown in Figure 2. The minimum cross validation error was obtained with the SNP set obtained for r^2<0.95 for the LD restriction. Thus, this set of markers has been chosen for the proportions of ancestry analysis. This same minimum cross-validation error was reached when the number of inferred populations was K=11. Thus, we could assume that the Brazilian Red Sindhi sampled population was structured into 11 subpopulations.

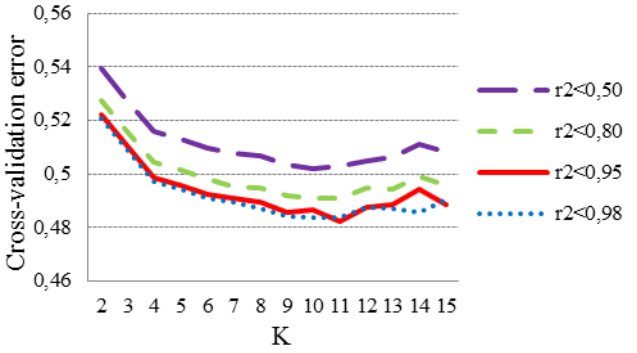

Figure 2 – Cross-validation errors according to the number of inferred subpopulations (K), at different levels of restriction for Linkage Disequilibrium among markers.

Pairwise F_{ST} statistics ranged from 0.163 (subpop.2 *x* subpop.9 or subpop.7 *x* subpop.10) to 0.344 (subpop.1 *x* subpop.5), indicating a range of degrees of divergence between inferred subpopulations.

Individual ancestry proportions, with animals sorted according to the location of their herds of origin, including State and City, are illustrated in **Figure 3**. A general observation of this figure indicates that one herd in the State

of Minas Gerais, one herd in Paraíba, and the herd sampled in Pernambuco were apparently less admixed than other herds in the study. These observations were consistent with the known histories of those herds, and point to the efficacy of the genomic approach to understand the structure of populations.

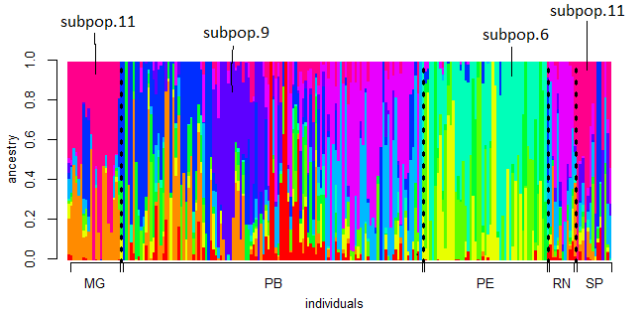

Figure 3 – Proportions of 11 inferred subpopulations for 218 individuals in the Brazilian States of MG (Minas Gerais), PB (Paraíba), PE (Pernambuco), RN (Rio Grande do Norte), and SP (São Paulo).

The observation of animals with major proportions in only one subpopulation, in Figure 3, added to the knowledge of the position of each herd within this figure (known by the authors, but not shown in the figure to avoid any possible commercial implication to individual breeders) provide an indication that some of the subpopulations are almost exclusively represented in one herd only. Examples are subpop.6 (light green in the State of PE) and subpop.9 (dark purple or blue, in the State of PB). Subpop.11 (dark pink) is concentrated only in the Southeast region of Brazil, in the States of MG and SP. These findings reinforce the possibility of introducing optimum contribution selection schemes within the Red Sindhi breed. The recommendation would be that sires and dams from those ancestry lines present in small numbers of existent herds, should be necessarily included in genetic evaluation schemes for the breed. Moreover, the use of superior animals (beef or dairy traits) identified within those lines should be encouraged in such a way that genetic diversity of this breed would not decrease in future generations.

Conclusion

The genomic approach, with unsupervised hierarchical clustering of individuals from SNP markers, allowed the identification of subpopulations for the Red Sindhi in Brazil. Results indicate the existence of 11 subpopulations, with some of them restricted to one or two herds. Thus, optimum contribution selection schemes should be adopted in breeding programs to this breed, to avoid future losses in the genetic diversity.

Acknowledgements

The authors would like to thank the genealogical information from ABCZ, the support of ABCSindi for the contact with breeders and sample collection, and the financial support from CNPq, CAPES and FAPEMIG.

References

Alexander, D.H., Novembre, J., Lange, K. (2009). *Genome Research*, 19:1655–1664

Gautier, M., Naves, M. (2011). *Mol. Ecology*, 20: 3128-3143.

Leite, P.R.M., Santiago, A.A., Navarro Filho, H.R. et al. (2001). Sindi: gado vermelho para o semi-árido. Banco do Nordeste, João Pessoa, PB, Brazil. 174p.

McKay, S.D., Schnabel, R.D., Murdoch, B.M. et al. (2008). *BMC Genetics*, 9:37.

Porto-Neto, L.R., Sonstegard, T.S., Liu, G.E. et al. (2013). *BMC Genomics*, 14:876.

Purcell, S., Neale, B., Todd-Brown, K., et al. (2007). Am. J. *Hum. Genet.*, 81(3):559–575.

Sonesson, A.K., Meuwissen, T.H.E. (2000). *Genet. Sel. Evol.*, 32:231-248.