META-ANALYSIS OF MICROSATELLITE DATA FROM US AND BRAZIL SHEEP BREEDS

S.R. Paiva [†], H. Blackburn [*], A.da S. Mariante [†]

Introduction

Originally, all domestic sheep were imported to the western hemisphere (WH). As a result of these importations there are issues impacting our understanding of these genetic resources including the similarity of breeds in WH with a similar country of origin, and how divergent populations may have become since importation. Also confounding the evaluations are WH populations that share phenotypic similarities but are known by different names. It has been proposed that these issues can be addressed by using a common set of microsatellite markers as proposed by FAO and the International Society for Animal Genetics (Hoffman et al., 2004: Baumung et al., 2004). Despite the recommendation by FAO, the merger of data sets has been limited during the past 15 years (e.g. San Cristobal 2006, Lenstra et al., 2008). Therefore this study adapted a methodology for merging independent microsatellite data sets and exploring the differences between Brazilian and US hair breeds of sheep.

Material and methods

Microsatellite Datasets. US samples were derived from 28 breeds (N=674) and genotyped with 28 FAO loci and Brazil (BZ) data had 10 breeds (N=383) genotyped by 22 Loci (11 from the FAO list). All the country datasets were obtained by different platforms with exception of 23 samples from the Hampshire were genotyped in by both countries.

Merging Procedure and Genetic of Population Analyses. Eleven markers were in common for the US-BZ dataset and were merged for further analysis. Two software packages (using maximum likelihood or a Bayesian approach) were evaluated. Based upon the results we selected the Bayesian approach with MicroMerge v.2 (Presson et al., 2008) which matches allele frequencies in the two datasets rather than allele size. The analysis was accomplished with a burn-in of 5,000 iterations and 5,000,000 iterations during the actual merging of data. Using this approach we were able to obtain convergence of the data with acceptable levels of posterior probabilities that ranged from 0.6 to 1.0 for the eleven markers. It should be noted that higher posterior probabilities were obtained when the common samples were included in the analysis vs performing the analysis without the Hampshire. After merging the consolidated dataset was analyzed by Structure (Pritchard et al., 2000) to test the level of genetic structure between breeds from both countries. Within breed diversity was estimated by three free software GenAlex, Fstat and Molkin.

[*] National Animal Germplasm Program, NCGRP, Agricultural Research Service United States Department of Agriculture, Fort Collins, CO 80521-4500, US
[†] Embrapa Genetic Resources and Biotechnology, Laboratory of Animal Genetics/PCG, Brasília, DF, 02372, Brazil; CNPq Productivity Scholarship Holder

Results and Discussion

The Structure analysis corroborated the merging process; as some BZ sampled breeds (Ile de France and Hampshire) were more close related with US breeds when K=2 (Figure 1). When K=7 we found that breeds were classified by their commercial function and not just country of origin. The K values of 10 and 12 were selected by method proposed by Evanno et al (2005) and they show a more refined substructure of the pattern already identified with K=7. Of particular interest are the placement of the US and BZ Hampshires and the Dorpers. As expected the Hampshires were placed in the same cluster however the Dorper were not. This situation with the Dorper was also reported by Kijas et al. (2009) using SNPs. While sampling could explain the difference the Hampshire results would seem to question this as a general explanation. An alternative hypothesis is that the Dorper, a more recently formed composite (Dorset X Blackhead Persian), still has not stabilized sufficiently to breed true. We also note the US Dorper sample had a higher F_{is} value which could also explain the separation between the BZ and US populations (Table 1).

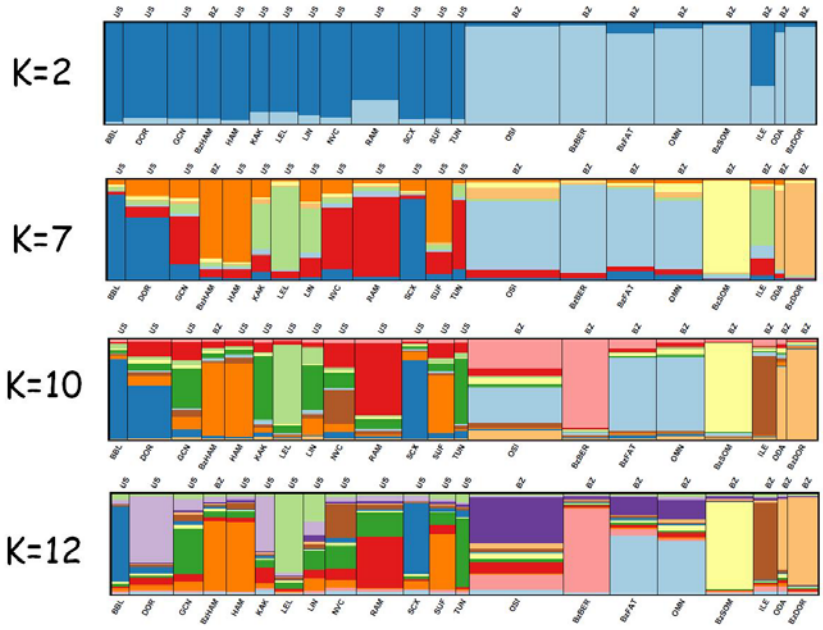

Figure 1: Genetic Structure of US and Brazilian (BZ) sheep breeds by Bayesian analysis with 11 microsatellite loci. K=number of clusters (for breed codes Table 2).

Both countries had similar ranges for mean number of alleles and heterozygosity (Table 1). However, the hair sheep breeds tended to have higher levels genetic diversity than wooled breeds, except for Rambouillet and Navajo Churro. Inbreeding levels tended to be higher across the US population, with the exception of the BZ Hampshire. The AMOVA showed 12.84% ($P<0.01$) of observed variation, that was explained by differences between breeds which is similar to other literature values.

We expected the hair breeds for both countries to be similar. However, neither the Structure nor principal component analysis (Figure 2) supported this assumption. As the principal component analysis also indicates, there is a wide variability among the BZ populations. Of particular interest is the divergence between the BZ Somali, BZ Dorper, and US Dorper; In this analysis, the Somali was found to be an outlier. Given this breed's use in developing the Dorper and the similar phenotypes across eastern and southern Africa a closer association was anticipated.

Table 1: Genetic diversity measures by breed (N=Number of samples; A_M=Mean Number of Alleles; A_E=Effective number of Alleles; He= Expected heterozygosis; Ho= Observed heterozygosis; MolC= Mean molecular coancestry index; F_{IS}= inbreeding coefficient).

Breed	Code	N	A_M	A_E	He	Ho	MolC	F_{IS}
Barbados Blackbelly	BBL	18	4.45	3.084	0.6536	0.4958	0.3722718	0.248*
Columbia	COL	21	4.91	2.933	0.6496	0.5584	0.332646	0.145
Dorper	DOR	44	6.27	3.291	0.6463	0.4814	0.3929878	0.258*
Gulf Coast Native	GCN	30	6.09	3.415	0.6858	0.5543	0.3630535	0.196*
Hampshire-BZ	HABz	23	5.73	3.113	0.6118	0.5217	0.3265669	0.151
Hampshire	HAM	29	5.64	3.056	0.5851	0.5235	0.3173815	0.107
Karakul	KAK	19	3.64	2.603	0.5766	0.3930	0.3709707	0.326*
Leicester Longwool	LEL	29	4.36	2.363	0.5282	0.3976	0.2946019	0.254*
Lincoln	LIN	22	4.55	2.848	0.5967	0.4493	0.3484472	0.252*
Navajo Churro	NVC	31	6.00	3.315	0.6744	0.4982	0.4003195	0.265*
Rambouillet	RAM	47	6.18	3.315	0.6871	0.5488	0.4045425	0.203*
Romney	RME	20	5.00	3.024	0.6634	0.5136	0.3696766	0.234*
St. Croix	SCX	26	5.36	3.345	0.6780	0.5863	0.3637051	0.138
Suffolk	SUF	26	5.36	3.092	0.6348	0.5038	0.3531751	0.210*
Texel	TEX	20	5.00	2.528	0.5810	0.5217	0.2661255	0.107
Tunis	TUN	14	4.91	3.025	0.6320	0.5100	0.3448808	0.200
St. Ines	OSI	94	6.27	3.917	0.7292	0.6882	0.3740563	0.056
Brazilian Bergamasca	OB	46	5.36	3.083	0.6531	0.6842	0.309192	-0.048
Brazilian Fat Tail	ORL	48	5.45	3.085	0.6600	0.7014	0.293827	-0.063
Morada Nova	OMN	48	5.09	3.365	0.6798	0.6225	0.3584191	0.085
Brazilian Somali	OS	48	4.55	2.601	0.5504	0.6570	0.2157592	-0.196
Ile de France	OIF	24	4.64	3.060	0.6060	0.5534	0.3143654	0.089
Corriedale	OC	11	3.18	2.449	0.5434	0.6226	0.2176578	-0.155
Damara	ODA	10	3.64	2.749	0.6689	0.7264	0.2451706	-0.107
Dorper-BZ	ODO	30	4.64	2.919	0.6300	0.6393	0.3019935	-0.015

* Significative homozygous excess after Bonferroni correction $p<0.00018$

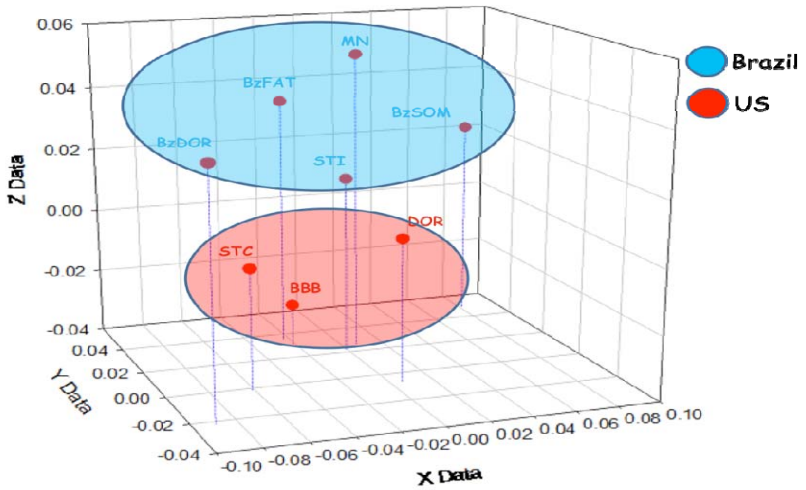

Figure 2. Relative Placement of Hair Sheep Breeds using Principal Component Analysis obtained by Nei Unbiased Genetic Distance. BBB-Barbados Blackbelly; STC-St. Croix; DOR-Dorper; STI-St. Ines; BzFAT-Brazilian Fat Tail; MN-Morada Nova; BzSOM-Brazilian Somali; BzDOR-Brazilian Dorper.

Conclusion

These results suggest the merging process facilitated the combined analysis among independently derived microsatellite datasets. With this approach it was possible to infer at least seven major gene pools among the 21 analyzed breeds. These gene pools were based upon physiological functions rather than country of origin. We also document the breadth of genetic diversity found, especially among hair sheep breeds. Such information may be useful to breeders as they move forward in selecting and developing new commercial populations. The information is also critical to the US and BZ conservation programs as they proceed with development and implementation of in-situ and ex-situ conservation activities. The analysis provided both countries an opportunity to leverage the information generated independently, as a result additional collaborations of this nature are planned.

References

Baumung, R., Simianer, H., Hoffmann, I. (2004). *J. Anim. Breed. Genet.*, 121: 361-373.

Evanno, G.; S. Regnaut, S.; Goudet, J. (2005), *Molecular Ecology*, 14: 2611–2620

Hoffmann, I.; Ajmone Marsan, P.; Barker, J.S.F. *et al*. (2004). *Proceedings of the 29th International Conference on Animal Genetics*, 123 (abstract).

Kijas, L.W.; Townley, D.; Dalrymple, B.P. *et al*. (2009). PLoS ONE, 4(3): e4668.

Lenstra et al. (2008). Proceedings of the 31th International Conference on Animal Genetics.

Presson , A.P.; Sobel, E.M.; Pajukanta, P.; *et al*. (2008). *BMC Bioinformatics*, 9:317

San Cristobal, M.; Chevalet, C.; Haley. C. S. (2006). Animal Genetics, 37: 189–198.