Journal of Heredity 2011:102(6):697–704
doi:10.1093/jhered/esr101

© The American Genetic Association. 2011. All rights reserved.
For permissions, please email: journals.permissions@oup.com.

Combining US and Brazilian Microsatellite Data for a Meta-Analysis of Sheep (*Ovis aries*) Breed Diversity: Facilitating the FAO Global Plan of Action for Conserving Animal Genetic Resources

Samuel Rezende Paiva, Arthur da Silva Mariante, and Harvey D. Blackburn

From the EMBRAPA Recursos Genéticos e Biotecnologia (Embrapa Genetic Resources and Biotechnology), Laboratório de Genética Animal, Parque Estação Biológica, Brasília, DF 70770-917, Brazil (Paiva and Mariante); and the National Animal Germplasm Program, National Center for Genetic Resources Preservation, Agricultural Research Service, United States Department of Agriculture, Fort Collins, CO (Blackburn).

Address correspondence to Samuel R. Paiva at the address above, or e-mail: samuel@cenargen.embrapa.br.

Abstract

Microsatellites are commonly used to understand genetic diversity among livestock populations. Nevertheless, most studies have involved the processing of samples in one laboratory or with common standards across laboratories. Our objective was to identify an approach to facilitate the merger of microsatellite data for cross-country comparison of genetic resources when samples were not evaluated in a single laboratory. Eleven microsatellites were included in the analysis of 13 US and 9 Brazilian sheep breeds ($N = 706$). A Bayesian approach was selected and evaluated with and without a shared set of samples analyzed by each country. All markers had a posterior probability of greater than 0.5, which was higher than predicted as reasonable by the software used. Sensitivity analysis indicated no difference between results with or without shared samples. Cluster analysis showed breeds to be partitioned by functional groups of hair, meat, or wool types ($K = 7$ and 12 of STRUCTURE). Cross-country comparison of hair breeds indicated substantial genetic distances and within breed variability. The selected approach can facilitate the merger and analysis of microsatellite data for cross-country comparison and extend the utility of previously collected molecular markers. In addition, the result of this type of analysis can be used in new and existing conservation programs.

Key words: *genetic structure, merging data sets, molecular markers*

Originally, all domestic sheep were imported to the Western Hemisphere (WH; Dohner 2001). As a result of these importations, there are issues impacting our understanding of these genetic resources including the similarity of WH breeds with a similar country of origin, and how divergent populations may have become since importation. Also, confounding the evaluations are WH sheep populations that share phenotypic similarities but are known by different names. The proposal for collection and use of a common set of genetic markers for the purpose of understanding genetic diversity was suggested by Barker et al. (1993). As a result, Food and Agriculture Organization (FAO) and the International Society proposed a common set of microsatellite markers for Animal Genetics (Hoffmann et al. 2004). Despite the FAO recommendation, the merger of data sets from different countries has been limited (Baumung et al. 2004; Freeman et al. 2006; SanCristobal et al. 2006; Boitard et al. 2010), and to our knowledge, no sheep data sets have been merged. While these previous efforts made cross-country comparisons, the methodology used was not fully detailed, and there appeared to be opportunities for improvement, such as analysis of samples that were not genotyped in the same laboratory and utilization of allele frequency as a basis for comparison.

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018

Assessment of genetic differences between livestock populations with microsatellites has provided important information on domestication, breed development, and current breed differences across geographic regions (Bruford et al. 2003; Hanotte 2007). In most of these instances, samples were acquired and analyzed in one laboratory or were used with a class of markers that have a genetic characteristic that permit a straightforward combination (e.g., sequence data). In order to improve the conservation of animal genetics resources, the global community has approved and published a Global Plan of Action for Animal Genetic Resources through FAO (2007). The Global Plan of Action articulated strategic priority for characterization of genetics resources in a way to optimize the cross-national comparability of data in order to monitor trends and risks to animal genetic resources at regional and global levels (FAO 2007).

To date, some of the issues surrounding the merger of data sets from different laboratories/teams include the utilization of different analytical platforms, genotyping (binning) methods, and size standards to estimate allele calls. Some laboratories with formal linkages make use of control samples to validate data sets. However, such associations exclude the growing number of laboratories performing molecular genetic analysis across a variety of platforms. In addition, much of the validation of data sets is based on base-pair size matching criterion alone (Amos et al. 2007; Morin et al. 2009). Appropriate tools to accomplish the merging task with sufficient accuracy and precision to enable robust conclusions are needed. There are some approaches that have been proposed to accomplish the task of merging microsatellite data (Freeman et al. 2006; Presson et al. 2006, 2008). The Presson et al. (2006, 2008) approach, which we ultimately used, consists of a Bayesian model and Markov chain Monte Carlo (MCMC) algorithm for sampling the posterior distribution under the conditions of the model. An output and measure of how successfully 2 independent measures of the same locus have been merged is the resultant posterior probability. That is, the posterior probability provides insight as to the confidence in the merged data, which the user can determine if the merged results are acceptable. To date, such an approach has not been fully utilized in livestock genetic diversity studies but is needed given the variety of laboratories that have collected microsatellite data and wish to make full use of this type of data.

The national animal genetics resources conservation programs in Brazil and in the United States wanted to determine the similarities and differences in a number of their sheep populations. Both countries have collected substantial numbers of animals and breeds for independent analysis of within country genetic diversity (Paiva et al. 2006; Blackburn et al. 2011). It is of benefit for both countries to merge this data and evaluate genetic differences or similarities that may exist. Therefore, the objectives were to test the utility of merging microsatellite data sets for a meta-analysis of genetic diversity found across the 2 countries.

Materials and Methods

Microsatellite Data Sets

US samples were derived from 28 breeds ($N = 674$) and genotyped with 28 FAO loci (Blackburn et al. 2011), and Brazil (BZ) data had 10 breeds ($N = 383$) and genotyped by 22 Loci (11 from the FAO panel) (Paiva et al. 2006), the subset of breeds used in this study are provided in Supplementary Material 1. Both countries data sets were obtained independently by different genotyping platforms with exception of 23 samples from the Brazilian Hampshire that were genotyped by both countries. For the US samples, a commercial company (GeneSeek) constructed the multiplex system, amplified the DNA, and made the allele calls (Blackburn et al. 2011). Paiva et al. (2006) detailed the collection and genotyping of the Brazilian samples, briefly, genotypes were obtained by capillary electrophoresis on an ABI310 Genetic analyzer (Applied Biosystems, Carlsbad, CA), and genotypes and allele calling were obtained using Gene Scan and Genotyper softwares (Applied Biosystems). It is important to note that Brazil had imported Hampshire in the 1990s from Canada and United States. Eleven markers were common for the US-BZ data set (*HUJ*616, *ILSTS*11, *ILSTS*5, *INRA*63, *MAF*214, *MAF*65, *OarAE*129, *OarFC*B20, *OarFCB*304, *OarJMP*29, and *SRCRSP*5) and were used for merging and genetic evaluation.

Merging Procedure

Two software packages were evaluated. The first approach COMBI.PL (Täubert and Bradley 2008) attempts to merge data sets by assigning allele sizes across studies using maximum likelihood estimation. The second, MicroMerge v.2 (Presson et al. 2008), is based on a Bayesian approach and aligns data sets marker by marker, matching each marker's allele frequencies while preserving size order. Based on the assumptions of both approaches and the structure of the data set, we selected the Bayesian approach contained in MicroMerge v.2. MicroMerge matches allele frequencies between data sets rather than utilizing allele size. This software was run using the following options: the one-to-one alignment format, remerging markers with low posterior probabilities, adjusting the prior on the theoretical allele number, and the genotype error were set from default value (0.02) for the value of 0.06. In addition, alleles with only 1 or 2 reads within a country's data set were deleted. The analysis was accomplished with a burn-in of 5000 iterations and 1 000 000–5 000 000 iterations during the actual merging of data (methods of MCMC). The criterion for acceptance of convergence for the posterior probabilities was considered at 0.6. This value was higher than the MicroMerge default of 0.45.

Population Genetics Analyses

The merged data set consisted of 38 breeds and 1057 animals with genotypes for all 11 selected markers. However, for the purpose of comparison among the countries, a smaller data set for the genetic analyses was

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018

selected and comprised of 21 breeds, 12 from United States, and 9 from Brazil ($N = 706$). The data set was reduced to eliminate many of the US breeds that are composite populations. The first hypothesis tested with the consolidated data set was to verify the level of genetic structure from both countries analyzed by Structure software (Pritchard et al. 2000). With this software, the individuals were probabilistically assigned using Bayesian inference to determined populations or grouped to one or more populations. For these estimates, the methods of MCMC were used, calculating for each individual the probability of one specific genotype "X" to be part of one given population "K" as the natural logarithm: [ln Pr $(X|K)$]. Those probabilities for K varied from 1 to 21 and were estimated by averaging the results of 3 replicate runs each having 650 000 iterations (MCMC) with an initial burn-in value of 200 000. The repetitions for each K were used to estimate an ad hoc indicator of population number based in the averaged likelihood at each K, called ΔK (Evanno et al. 2005). Cluster assignments and their graphical representation were performed using DISTRUCT (Rosenberg 2004).

A second hypothesis analyzed by structure was verifying the power of microsatellite markers to identify common origins among hair sheep breeds presented in both countries. To corroborate Structure results, principal coordinate analysis was performed with Nei's Genetic distance (Nei 1972), and estimates between populations were computed with GENALEX 6 software (Peakall and Smouse 2006). Analysis of molecular variance (AMOVA) was performed with Arlequin software (Excoffier and Lischer 2010) using the codominant allelic distance matrix with 1000 permutations, as an additional test to identify traces of genetic structure in a series of contrasts between all populations studied. Fixation index (F_{IS}) and basic within genetic diversity coefficients were calculated using FSTAT (Goudet 2002) and Molkin (Gutiérrez et al. 2005).

Results

Based on the preset convergence criteria, the Bayesian methodology for merging the 2 data sets was successful and appeared robust. For these data sets, posterior probabilities of each locus ranged from 0.6 to 1.0 and from 0.5 to 0.98 with and without the shared samples (Table 1). The result of comparing with and without common markers was anticipated; however, the level of agreement was unexpected. The majority of articles for livestock diversity studies there are no common samples. For this reason and to test the robustness of MicroMerge, the 2 markers with posterior probabilities of less than 0.6 (*HUJ*616 and *OarAE*129) were discarded, and intra- and intergenetic diversity indices were performed. Within diversity indices between the 2 panels (11 and 9 markers) were very similar (data not shown), and the Mantel test for Nei genetic distances matrices between breeds showed a correlation coefficient of 0.92 ($P < 0.000001$) for the 2 panels, suggesting that merging could proceed with a relatively small number of microsatellite markers within the predefined convergence criteria.

Table 1 Merging results of 11 common microsatellite obtained from 2 independent sheep data sets genotyped by United States and Brazil

	Alleles			Posterior probabilities	
Loci	US	BZ	Final	With shared samples	Without shared samples
*HUJ*616	23	9	10	0.867	0.503
*ILSTS*11	9	8	10	0.730	0.681
*ILSTS*5	12	9	11	0.970	0.943
*INRA*63	21	12	14	0.810	0.679
*MAF*214	10	9	8	0.924	0.983
*MAF*65	11	10	10	0.827	0.925
*OarAE*129	9	8	5	1.000	0.517
*OarFCB*20	15	14	13	0.790	0.606
*OarFCB*304	22	16	15	0.614	0.684
*OarJMP*29	20	17	15	0.784	0.645
*SRCRSP*5	6	5	6	1.000	0.977

Posterior probability provides insight as to the confidence in the merged data, which the user can determine if the merged results are acceptable.

The number of alleles observed in US breeds was always higher than the Brazilian breeds probably because of the number of breeds and larger sample sizes. After the merging procedure, the number of alleles for the unified matrix of genotypes, as expected, had the tendency to be reduced, except for the markers *ILSTS*11 and *SRCRSP*5.

The within genetic diversity among breeds showed that both countries had similar ranges for mean number of alleles and H_e (expected heterozygosis); however, Brazilian breeds had high levels of H_o (observed heterozygosis) (Table 2). Inbreeding (F_{IS}) was lowest for the collection of Brazilian breeds, whereas values for some US populations were significantly larger (Table 2). The hair and hair-coarse wooled sheep breeds tended to have higher levels of genetic diversity than wooled breeds, except for Rambouillet and Navajo Churro.

The ΔK peaked at $K = 2$ and had a minor peak at $K = 12$ suggesting structural differences for these populations (Evanno et al. 2005). Based on the ΔKs, knowledge of breed histories, and inspection of the range of Ks tested, the K values of 2, 7, and 12 were evaluated. When $K = 2$, the Brazilian Hampshire and Ile de France were placed in the same cluster as all US breeds (Figure 1). This placement suggests that the merging process was successful because both breeds genetic backgrounds/formation involve the same breeds as those found in the US breeds. When $K = 7$, it was found that breeds were classified by their commercial function or major phenotype class (e.g., long wool, hair, black face, and meta production) and not just by their country of origin. The K value of 12 showed a more refined substructure of the pattern identified with $K = 7$. Of particular interest was the placement of the US and BZ Hampshires and the Dorpers. As expected Hampshires were placed in the same cluster, but unexpectedly, the

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018

Table 2 Genetic diversity measures by breed (N, number of samples; A_M, mean number of alleles; A_E, effective number of alleles; H_e, expected heterozygosis; H_o, observed heterozygosis; MolC, mean molecular coancestry index; and F_{IS}, inbreeding coefficient)

Breed	Code	N	A_M	A_E	H_e	H_o	MolC	F_{IS}
Barbados Blackbelly	BBL	18	4.45	3.084	0.6536	0.4958	0.3723	0.248*
Dorper	DOR	44	6.27	3.291	0.6463	0.4814	0.3929	0.258*
Gulf Coast Native	GCN	30	6.09	3.415	0.6858	0.5543	0.3631	0.196*
Hampshire-BZ	BzHAM	23	5.73	3.113	0.6118	0.5217	0.3266	0.151
Hampshire	HAM	29	5.64	3.056	0.5851	0.5235	0.3174	0.107
Karakul	KAK	19	3.64	2.603	0.5766	0.3930	0.3709	0.326*
Leicester Longwool	LEL	29	4.36	2.363	0.5282	0.3976	0.2946	0.254*
Lincoln	LIN	22	4.55	2.848	0.5967	0.4493	0.3484	0.252*
Navajo Churro	NVC	31	6.00	3.315	0.6744	0.4982	0.4004	0.265*
Rambouillet	RAM	47	6.18	3.315	0.6871	0.5488	0.4045	0.203*
Saint Croix	SCX	26	5.36	3.345	0.6780	0.5863	0.3637	0.138
Suffolk	SUF	26	5.36	3.092	0.6348	0.5038	0.3532	0.210*
Tunis	TUN	14	4.91	3.025	0.6320	0.5100	0.3449	0.200
Saint Ines	OSI	94	6.27	3.917	0.7292	0.6882	0.3741	0.056
Brazilian Bergamasca	BzBER	46	5.36	3.083	0.6531	0.6842	0.3092	−0.048
Brazilian Fat-Tail	BzFAT	48	5.45	3.085	0.6600	0.7014	0.2938	−0.063
Morada Nova	OMN	48	5.09	3.365	0.6798	0.6225	0.3584	0.085
Brazilian Somali	BzSOM	48	4.55	2.601	0.5504	0.6570	0.2158	−0.196
Ile de France	ILE	24	4.64	3.060	0.6060	0.5534	0.3144	0.089
Damara	ODA	10	3.64	2.749	0.6689	0.7264	0.2452	−0.107
Dorper-BZ	BzDOR	30	4.64	2.919	0.6300	0.6393	0.3019	−0.015

*Significative homozygous excess after Bonferroni correction $P < 0.00018$.

Dorpers were not. Although sampling could explain the difference, the Hampshire results would indicate that this is not an explanation. An alternative hypothesis is that the Dorper, a composite (Dorset × Blackhead Persian) formed during the 1940s (Porter and Mason 2002), still has not stabilized sufficiently to breed true. Alternatively, the US Dorper sample had a higher F_{IS} value, which could contribute to the separation between the Brazilian and US populations (Table 2).

A series of AMOVA were performed using 4 different contrasts (Table 3). In general, the principal contrast observed (last one) showed 12.84% ($P < 0.01$) of observed variation, which was explained by differences between breeds that are similar to other literature values (Handley et al. 2007). The other contrasts tested were not significant.

Principal coordinate analysis was performed, and the first 3 components accounted for 65% of the observed variation (Figure 2). It was expected that the hair breeds for both countries would be similar. However, neither the Structure nor principal component analysis supported this assumption. The principal component analysis also indicated that there was substantial variability among the Brazilian hair sheep populations. Of particular interests are: 1) the divergence between the BZ Somali, BZ Dorper, and US Dorper and 2) the proximity of Barbados Blackbelly to the Brazilian hair sheep breeds by the 2 first principal components provides some insight as to the Barbados Blackbelly relationship to other breeds. The principal component analysis also showed the Somali and Brazilian Dorper to be outliers. Given the Somali's use in developing the Dorper and the similar phenotypes across eastern and southern Africa, a closer association was anticipated.

Discussion

The merging procedure made possible an evaluation of US and Brazilian sheep breeds. The close association and similar measures of genetic variability for the Hampshire populations suggest that the merging process was successful and that the differences in genetic variation and genetic similarity were a function of allele frequencies observed and not the merging process.

The higher levels of inbreeding (F_{IS}) observed in some US breeds could possibly be due to smaller population sizes or a higher level of selection pressure. Across all analysis of AMOVA performed the within populations was the largest and most significant source of variation. These results are similar to a wide range of studies (Handley et al. 2007; Peter et al. 2007; Blackburn et al. 2011). Among populations, within groups were also a significant source of variation, although substantially smaller than the within population component. This result suggests that breeds within countries (e.g., hair vs. wool breeds and even rare breeds within country) exhibited distinct genetic variation useful in utilizing these genetic resources as breeders strive to increase productivity and the breeds' competitive advantage.

Several conclusions can be derived when following the STRUCTURE analysis for 2, 7, and 12 clusters. From the perspective of the merging procedure when K equaled 2, the Brazilian Hampshire was placed within the US breeds as was a high proportion of the Ile de France. The similarities of Hampshire from both countries are likely explained by the importation of animals in the 1990s from Canada and United States (ABCOHD 2009) and their common UK origin. For breeds within country, there was agreement with

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018

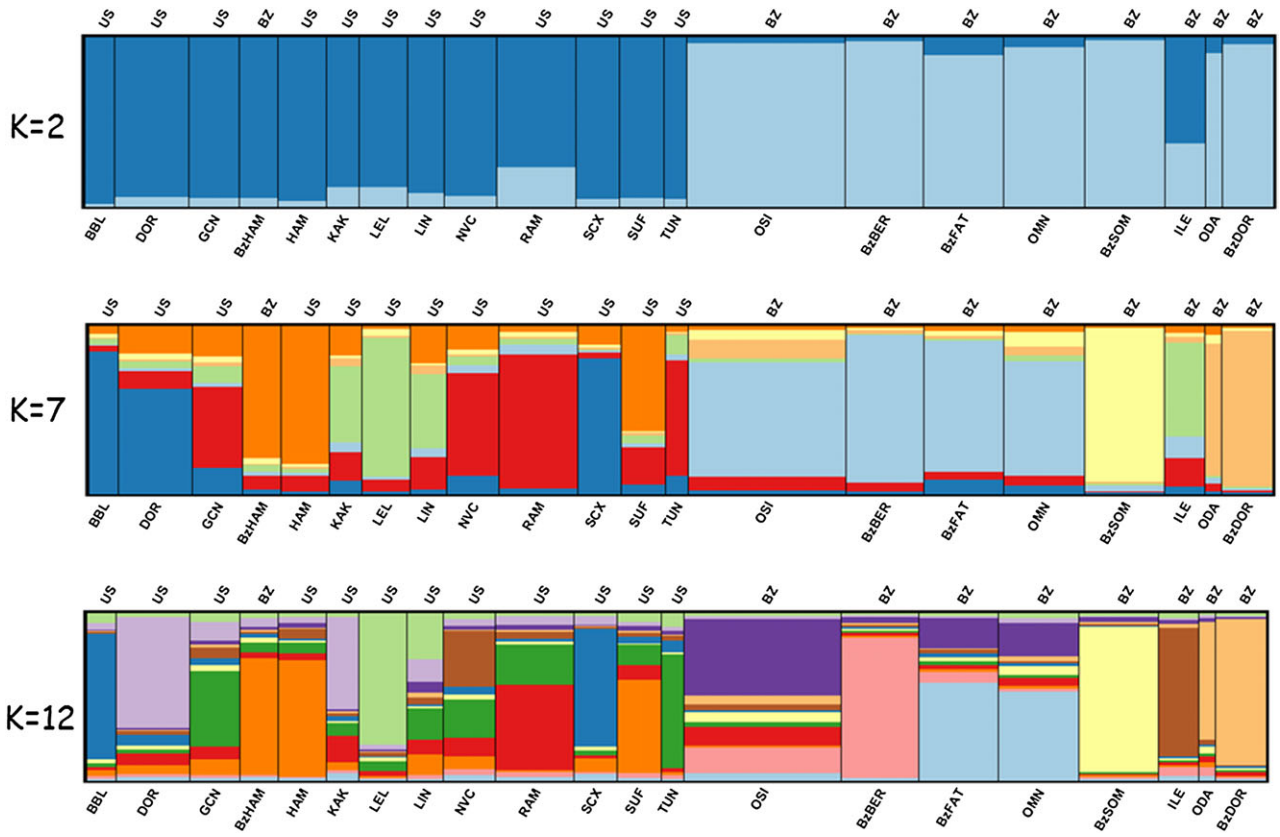

Figure 1. Genetic structure of US and Brazilian (BZ) sheep breeds by Bayesian analysis with 11 microsatellite loci. K = number of clusters (for breed codes, Table 2).

previously published results about the association of breeds within a cluster (Paiva et al. 2006; Blackburn et al. 2011). For example: Barbados Blackbelly—Saint Croix, Damara—Brazilian Dorper, Hampshire—Brazilian Hampshire-Suffolk, and Morada Nova—Brazilian Fat-tail.

In terms of using the merged data in a Bayesian and principal component analysis, it was possible to assess genetic variability between the 2 countries. When *K* was set to 7, breeds within country tended to be placed in clusters based on function. Within the United States, these groups consisted of hair, fine wooled, long wooled, and meat breeds. In addition to the previously discussed associations between US and Brazilian breeds, multibreed Brazilian clusters consisted of Santa Ines, Bergamasca, Morada Nova, and Brazilian Fat-tail. The Damara, Brazilian Dorper, and Brazilian Somali were in separate clusters. It was noted that the Ile de France a composite of Leicester Longwool and Merino (Porter and Mason 2002) was admixed with cluster

Table 3 AMOVA obtained among US and Brazil sheep breeds by a set of 11 microsatellite markers

Sample contrast	Source of variation	df	Sum of squares	Percentage variation
United States versus Brazil	Among groups	1	53.172	0.78^{ns}
	Among populations within groups	23	596.790	12.38*
	Within populations	1531	4125.672	86.84*
Wool versus hair breeds	Among groups	1	56.684	0.96^{ns}
	Among populations within groups	23	593.278	12.27*
	Within populations	1531	4125.672	86.76*
US versus Brazil rare breeds	Among groups	1	51.192	1.81^{ns}
	Among populations within groups	9	262.412	10.84*
	Within populations	833	2377.139	87.35*
Between breeds	Among populations	24	649.962	12.84*
	Within populations	1531	4125.672	87.16*

df, deghrees of freedom; ns, non significative.

*Significative $P < 0.01$.

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018

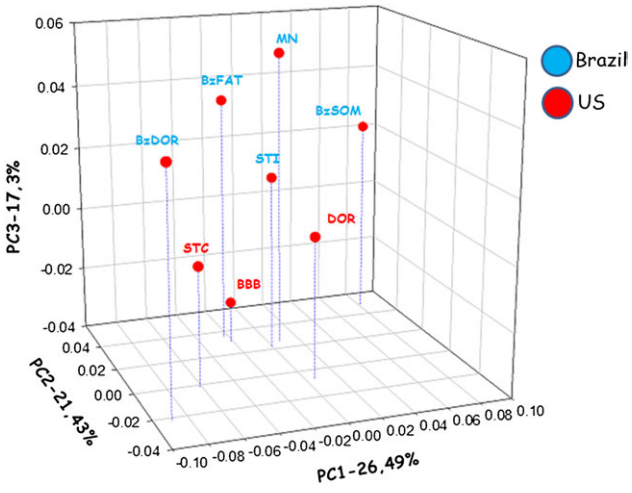

Figure 2. Relative placement of hair sheep breeds using principal coordinate analysis obtained by Nei unbiased genetic distance. BBB, Barbados Blackbelly; STC, Saint Croix; DOR, Dorper; STI, Saint Ines; BzFAT, Brazilian Fat Tail; MN, Morada Nova; BzSOM, Brazilian Somali; and BzDOR, Brazilian Dorper.

assignments corresponding to the Leicester Longwool and the Rambouillet.

The evaluation of breeds when $K = 12$, which the approach of Evanno et al. (2005) indicated was an appropriate number of clusters for this set of breeds and yielded a number of insights about the breeds and their relationship to one another. Among the Brazilian breeds, the Bergamasca, Brazilian Fat-tail—Morada Nova, Brazilian Somali, Ile de France, and Damara—Brazilian Dorper formed unique clusters. Although the Santa Ines was placed in a separate cluster and shown to have a substantial amount of admixture (in part due to its formation with Bergamasca and Brazilian Fat-tail). The Brazilian Fat-tail and Morada Nova were also shown to have a substantial proportion of their genetic composition in the Santa Ines cluster. Across countries, both Hampshire populations were placed in the same cluster corroborating the previous K values and recent importations.

The association among the hair breeds of the 2 countries is of interest due to a growing attention in the production of sheep with little or no wool. Using principal component analysis (Figure 2) showed the hair breeds widely dispersed across the 3 dimensions. The Brazilian Dorper and Brazilian Somali were placed at the extreme corners of the first and second principal components, whereas the Barbados Blackbelly and Morada Nova were placed at the extremes of the third principal component. Breeds placed in closest proximity to one another were the Barbados Blackbelly—Brazilian Fat-tail and the Santa Ines—Morada Nova. The relatively close association between the Barbados Blackbelly and the Brazilian Fat-tail is the first close association seen with the Barbados Blackbelly. However, the origin of the Brazilian Fat-tail is not clearly known. Muigai et al. (2002) reported the Barbados Blackbelly to be more closely associated with Iberian wooled breeds instead of West African hair breeds, thereby rejecting their original hypothesis. The Brazilian and US Dorpers were not closely associated in the STRUCTURE or principal component analyses; in addition, the Brazilian Somali was quite distinct from the 2 Dorper populations. The difference among Dorper populations was also observed by Kijas et al. (2009), suggesting that distinct subpopulations were exported, or gene frequencies have not been sufficiently fixed within South Africa.

Presently, there is little overlap between Brazilian and US sheep breeds. The distinctness of both countries hair, meat, fine wool, and long wool populations indicate a wide range of genetic variability has been transferred to the WH. These results would suggest that WH sheep breeders have an opportunity to evaluate each other's populations and determine if their respective breeds might benefit from the introgression of genes that would yield significant levels of heterosis for production traits of interest. In addition, these results would suggest that among the WH, these 2 countries have established sheep populations with considerable variability important for conservation.

The present study was the first to apply for livestock the proposed method by Presson et al. (2008), which was developed for humans. The results appear robust enough to support the merging of microsatellite data for other livestock species. The higher posterior probabilities obtained for most of the markers showed the importance to have common animals (breed Hampshire) genotyped in both analyses (Table 1). Because of the complexity of electrophoresis and the variability of analytical platforms, the simple comparison based on base-pair size as a matching criterion alone can increase the probability of errors in the merging process. The Bayesian approach appears to enhance the matches between allele frequencies for these 2 data sets and therefore circumvent the issues created by merging based on allele size. It was found that alleles with low frequencies were more difficult to estimate a posterior probability greater than the cutoff level of 0.6. However, this problem will likely remain an issue regardless of the approach used to merge data sets, thus requiring careful evaluation of potential reasons for the alleles to have a low frequency (e.g., genotyping errors). Furthermore, potential exclusion of alleles with very low frequencies may not have a large impact on cross-country evaluations given the results from STRUCTURE when $K = 2$.

The robustness of the adapted method was observed when the results obtained with 11 markers (common samples) and 9 markers (without common samples) did not vary significantly. In addition, Presson et al. (2008) suggested a default threshold of 0.425 to discard a marker; however, in this study, the threshold was increased to 0.6 for data sets with common samples, in order to obtain a more conservative evaluation. Freeman et al. (2006) proposed a regression method that a simple presence of a common breed is enough to support necessary information to merge the results between 2 or more data sets. The results of this study were in agreement because the presence of Hampshire

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018

breed in both data sets helps the merging process and we had better results (higher posterior probabilities for each locus) when the common samples were used, illustrating the desirability of having common samples whenever possible. However, as the posterior probabilities illustrate, the MicroMerge approach does offer researchers an opportunity to explore combining data sets with no genotyping of common animals/breeds.

In conclusion, it is suggested that the method evaluated can be used to successfully merge livestock data sets. Application of the approach permitted an evaluation of genetic diversity among US and Brazilian sheep breeds. As a result, both national programs are in a better position to assess their national sheep populations and plan in situ and ex situ conservation activities. Based on these results, it was also suggested that the tested approach can be further utilized to make cross-country or region comparisons for livestock populations in general. Such evaluations are timely as many countries and laboratories within countries have expended considerable resources in collecting and genotyping livestock populations via microsatellites markers and have not yet made full use of their investment. Furthermore, as genotyping platforms are converting from microsatellites to SNPs and whole-genome analysis, such evaluations may bring microsatellite analysis to a logical conclusion.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

ARS/USDA, CNPq and Embrapa.

Acknowledgments

The authors thank the owners of the numerous flocks that provided tissue samples for this project. Mention of a trade name or proprietary product does not constitute a guaranty or warranty by the United States Department of Agriculture (USDA) and Embrapa and does not imply approval to the exclusion of other products that may be suitable. In addition, the authors would like to thank the 2 anonymous reviewers. USDA, Agricultural Research Service, and Northern Plains Area are an equal opportunity/affirmative action employer. All agency services are available without discrimination.

References

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS. 2007. Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes. 7:10–14.

Associação Brasileira de Criadores de Ovinos Hampshire Down (ABCOHD). 2009. Hampshire no Brasil. Brazilian Association of Hampshire Down Breeders; [cited 2011 Apr 25]. Available form: http://hampshire.com.br/?pg=brasil.

Barker JSF, Bradley DG, Fries R, Hill WG, Nei M, Wayne RK. 1993. An integrated global programme to establish the genetic relationships among the breeds of each domestic animal species. Rome (Italy): FAO. Report of a working group for the Animal Production and Health Division.

Baumung R, Simianer H, Hoffmann I. 2004. Genetic diversity studies in farm animals—a survey. J Anim Breed Genet. 121:361–373.

Blackburn HD, Paiva SR, Wildeus S, Getz W, Waldron D, Stobart R, Bixby D, Purdy PH, Welsh C, Spiller S. 2011. Genetic structure and diversity among US sheep breeds: identification of the major gene pools. J Anim Sci. 89:2336–2348.

Boitard S, Chevalet C, Mercat MJ, Meriaux JC, Sanchez A, Tibau J, Sancristobal M. 2010. Genetic variability, structure and assignment of Spanish and French pig populations based on a large sampling. Anim Genet. 41:608–618.

Bruford MW, Bradley DG, Luikart G. 2003. DNA markers reveal the complexity of livestock domestication. Nat Rev Genet. 4:900–910.

Dohner JV. 2001. The encyclopedia of historic and endangered livestock and poultry breeds. New Haven (CT): Yale University Press.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14:2611–2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 10:564–567.

FAO. 2007. The Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration. Rome (Italy): FAO; [cited 2011 Mar] Available from: ftp://ftp.fao.org/docrep/fao/010/a1404e/a1404e00.pdf.

Freeman AR, Bradley DG, Nagda S, Gibson JP, Hanotte O. 2006. Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. Anim Genet. 37:1–9.

Goudet J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9. 3.2). Lausanne (Switzerland): University of Lausanne, Department of Ecology and Evolution.

Gutiérrez JP, Royo LJ, Alvarez I, Goyache F. 2005. MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. J Hered. 96:718–721.

Handley LJL, Byrne K, Santucci F, Townsend S, Taylor M, Bruford MW, Hewitt GM. 2007. Genetic structure of European sheep breeds. Heredity. 99:620–631.

Hanotte OH. 2007. Origin and history of livestock diversity. In: Richkowsky B, Pilling D, editors. The state of the world's animal genetic resources for food and agriculture. Rome (Italy): FAO. p. 5–19.

Hoffmann I, Marsan PA, Barker JSF, Cothran EG, Hanotte O, Lenstra JA, Milan D, Weigend S, Simianer H. 2004. New MoDAD marker sets to be used in diversity studies for the major farm animal species: recommendations of a joint ISAG/FAO working group. Proceedings of the 29th International Conference on Animal Genetics, 2004 Sep 11–16; Tokyo, Japan 123 [Abstract].

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, et al. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 4:e4668.

Morin PA, Manaster C, Mesnick SL, Holland R. 2009. Normalization and binning of historical and multi-source microsatellite data: overcoming the problems of allele size shift with allelogram. Mol Ecol Resour. 9:1451–1455.

Muigai AWT, Hirbo J, Sharkey S, Rege JEO , Blackburn H, Hanotte O. 2002. Genetic diversity and relationships of hair sheep breeds of the Americas: first results. Proceedings of the 7th World Congress on Genetics Applied to Livetock Production; Montpellier, France p. 573–576.

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018

Nei M. 1972. Genetic distance between populations. Am Nat. 106: 283–292.

Paiva SP, Faria DF, Dergam JD, McManus CM, Guimarães SG, Egito AE, Albuquerque MA, Silvia SC, Mariante AS. 2006. Genetic structure of hair sheep breeds in Brazil by microsatellites markers. Proceedings of the 30th International Conference on Animal Genetics; 2006 Aug 20–25; Porto Seguro, Brazil. Belo Horizonte (Brazil): CBRA, section A [Abstract].

Peakall ROD, Smouse PE. 2006. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes. 6:288–295.

Peter C, Bruford M, Perez T, Dalamitra S, Hewitt G, Erhardt G. 2007. Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. Anim Genet. 38:37–44.

Porter V, Mason IL. 2002. Mason's world dictionary of livestock breeds, types, and varieties. New York: CABI.

Presson AP, Sobel E, Lange K, Papp J. 2006. Merging microsatellite data. J Comput Biol. 13:1131–1147.

Presson AP, Sobel EM, Pajukanta P, Plaisier C, Weeks DE, Aberg K, Papp JC. 2008. Merging microsatellite data: enhanced methodology and software to combine genotype data for linkage and association analysis. BMC Bioinformatics. 9:317.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Rosenberg NAR. 2004. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes. 4:137–138.

SanCristobal M, Chevalet C, Haley CS, Joosten R, Rattink AP, Harlizius B, Groenen MA, Amigues Y, Boscher MY, Russell G, et al. 2006. Genetic diversity within and between European pig breeds using microsatellite markers. Anim Genet. 37:189–198.

Täubert H, Bradley DG. 2008. COMBI. PL: a computer program to combine data sets with inconsistent microsatellite marker allele size information. Mol Ecol Resour. 8:572–574.

Received April 26, 2011; Revised August 23, 2011; Accepted August 30, 2011

Corresponding Editor: Ernest Bailey

Downloaded from https://academic.oup.com/jhered/article-abstract/102/6/697/845950 by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) user on 24 September 2018