

Merging Molecular Data for Evaluating Cross Country Genetic Diversity of Pigs

C. O. Ibeka¹, H. D. Blackburn², S. R. Paiva³ and A. S. Mariante^{3*}

¹Department of Animal Sciences, Colorado State University, Fort Collins USA, ²National Animal Germplasm Program ARS-USDA, Ft. Collins, USA, ³Embrapa Headquarters – Secretariat of International Affairs, Brasilia, Brazil,

⁴Embrapa Genetic Resources and Biotechnology, Brasilia, Brazil

ABSTRACT: Integration of molecular data generated by microsatellite panels recommended by FAO around the world should be initiated in order to accomplish objectives stated in the Global Plan of Action for Animal Genetic Resources. To that end microsatellite datasets from US (n=179, including imported Chinese breeds) and Brazil (n=204) pig breeds were merged and the genetic diversity compared using 14 microsatellite markers. MICROMERGE software successfully combined 12 of 14 markers with posterior probabilities exceeding minimum criteria. Global F_{st} of 21% was observed, suggesting a high degree of population segregation. STRUCTURE analysis with K=4 indicated geographic separation of breed types into BR, US and Chinese groups. Cluster number K=9 provided insight into the within and among country genetic diversity. The ease and effectiveness with which the data was combined suggest two countries can share and co-analyze preexisting data to maximizing previous investments.

Keywords: Conservation of animal genetic resources; Bayesian clustering; swine

Introduction

Numerous genetic diversity (GD) studies have been conducted within country using microsatellite markers. What is often lacking in such studies is the ability to extend results to a regional or global level. Using Bayesian approaches (Presson et al., 2008; Paiva, et al. 2011) these investments can be extended to provide greater insight into the status of GD among countries without additional sampling and genotyping. In an effort to better understand GD among western hemisphere sheep breeds Paiva, et al. (2011) showed how such an approach was useful in comparing Brazilian and US sheep breeds. Brazil (BR) and United States (US) pig breeds originated from different European countries or China. However, once imported breeding decisions were made to mold the populations to meet the needs within each country and in some instances to develop new breeds (e.g., Duroc). In this study we merged BR and US microsatellite data (Sollero et al., 2009 and Blackburn et al., 2014) for pigs using the method developed by Presson et al. (2008). Once the microsatellites are merged we explore GD parameters within and among breeds for the two countries

Materials and Methods

Data. Tissue or semen samples for BR and US populations were withdrawn from the national genetic resource repositories for this study. The ten BR populations (including 5 commercial breeds) in the study were: Large White (n=3), Brazilian Duroc (n=4), Landrace (n=31), Pietrain (n=4), MS60 (n=48, an industry composite based on Duroc, Pietrain and Large White and four locally adapted breeds: Moura (n=35), Nilo (n=5), Piau (n=31) and Monteiro (n=37); as well as one Mixed group (n=6), consisting of small number of samples from three locally adapted pig breeds (Caruncho, Mulefoot and Pirapetinga). The six US populations were: Meishan (n=64), Fengjing (n=22), Minzhu (n=20), Yorkshire (n=21), Berkshire (n=26) and Duroc (n=26). The Meishan were subdivided into a group originally imported from China in the 1980's (Meishan-China, n=22) and samples taken during the last decade from randomly mated populations (Meishan-US, n=42). Previously, microsatellites from BR (n=30) and US (n=35) had been used to genotype the populations in their respective country (Sollero et al. 2009; Blackburn et al., 2014). There were no common samples analyzed by either laboratory.

Merging of Datasets. Merging the datasets was accomplished by using MICROMERGE Version 2.0 (Presson et al., 2008). Our group had used this approach with sheep and validated MICROMERGE results by genotyping samples from both countries in the same laboratory (Paiva et al., 2011). In that study MICROMERGE was run with and without the common samples and good agreement between the results was achieved. Due to the previous success in using MICROMERGE and the validation performed in this study, the need for a set of samples to serve as controls was not necessary. MICROMERGE uses allele frequencies as the basis for merging the data in a Bayesian approach to establish posterior probabilities as a means for determining if a locus can be merged.

There were 14 loci that both countries had in common and were used as a basis for merging the datasets. The programs merger criteria were set to a burn-in of 10,000 and MCMC 5,000,000 iterations, with rare alleles being rejected at less than 0.05 and an assumed genotyping error of 0.05. Criteria for accepting the merger of common microsatellites consisted of having a posterior probability > 0.55. There were 12 microsatellites that met the criteria (Table 1).

Table 1. Merging Results of 10 Brazilian Swine Breeds and 7 US-English Swine Breeds Genotyped at 12 Common Microsatellite Loci

Loci	Allele			Posterior Probabilities
	US	BR	Final	
SW24	12	10	15	0.554
SW240	13	10	14	0.764
S0227	9	10	12	0.870
SW951	5	10	12	0.813
SW2406	11	10	14	0.863
SW830	9	9	13	0.966
SW2008	8	9	12	0.572
S0026	9	7	11	0.811
SW2410	13	7	17	0.744
SW72	8	5	9	0.843
SW857	9	8	11	0.924
S0225	11	6	12	0.906

Genetic Population Analyses. With the merged dataset genetic diversity analysis was performed within and among populations using GENALEX version 6.501 (Peakall and Smouse 2012) and FSTAT (Goudet, 2002). Population structure was evaluated using STRUCTURE version 2.3 (Pritchard et al., 2010) with burn-in length of 50,000; MCMC replication of 100,000 and 5 replications. Delta K peaks were observed at K=2, 4, 13 and 15 with the highest delta at K=2.

Results and Discussion

The range of number of alleles (2.4 to 5.8) per population was relatively small, with Monteiro ranking highest and Large White the lowest. Breeds—with observed heterozygosity of less than 0.50 and their respective expected heterozygosity were BR Duroc (0.451, 0.472), Monteiro (0.475,0.58), Mixed (0.475,0.668) and Berkshire (0.412,0.414). Among the US breeds, measures of diversity (e.g., observed and expected heterozygosity) tended to be of similar magnitude (Figure 1) to the previous study that generated this dataset (Blackburn et al., 2014), however the previous estimates for Duroc and Minzhu were noticeably larger. These results suggest that while the number of loci was reduced (data not shown) in the merging process, sufficient loci remained in the analysis to explore genetic diversity among the breeds. A global F_{st} was computed to be 0.208 suggesting that a relatively high degree of population segregation has occurred. AMOVA indicated that the source of variation among populations accounted for 21% of the total variance tending to be higher than other reports (Paiva et al., 2011; Blackburn et al. 2014).

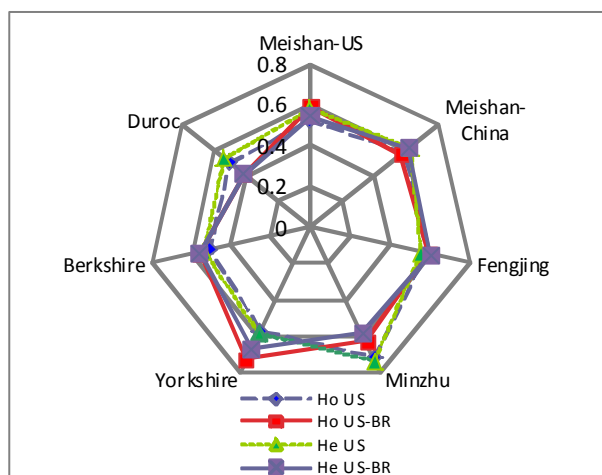


Figure 1. Observed heterozygosity (Ho) and Expected heterozygosity (He) comparing this study (US-BR) and a previous study (US) by breed.

Principal coordinate analysis was performed to evaluate breed differences (Figure 2). The breeds around the parameter of the plot indicate the most divergent populations (Meishan, Fengjing, Monteiro, Berkshire and Duroc). Interestingly the BR Duroc was not closely placed with the US counterpart most likely due to the small BR sample size. The Monteiro, a locally adapted BR breed was the most distinct while the other non-commercial breeds were closely grouped likely due to a lack of selection. The relative proximity of the MS60, Piau, Moura, Nilo is interesting given the MS60 is a composite population based upon industrial breeds. The same trend of non-selected populations, like Piau, Moura, and Nilo, appearing amongst commercially important breeds has been observed when sheep breeds from Kazakhstan were compared to US breeds (Blackburn et al., 2011).

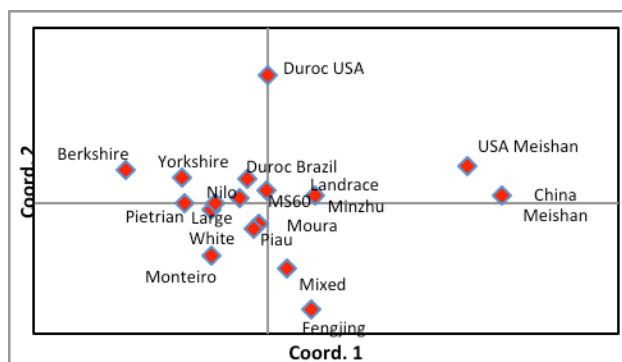
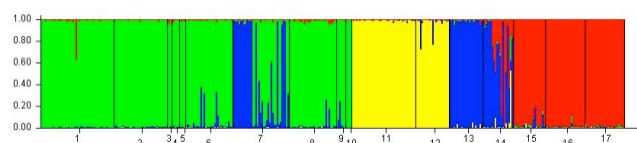


Figure 2. Breed placement derived from principal coordinate analysis, where coordinate 1 and 2 explain 43 and 21 % of the variation of the dataset.

STRUCTURE analysis when K=4 indicated a strong geographic separation of breed types into BR, US and Chinese groups (Figure 3a). We noted with interest that the Chinese breeds were split into two groups the Meishan

and Fengjing plus Minzhu suggesting very genetically divergent populations from that country. The other US breeds were placed together in a group, while the BR breeds were placed in their own group except the Monteiro, which showed admixture with the Fengjing and Minzhu breeds. Increasing the cluster number to K=9 provided further insight into the within country and among country genetic diversity (Figure 3b). Among the BR breeds, the composite population MS60 was clustered with the Pietrain and Large White. The BR breeds Nilo, Duroc, Mixed, Piau and Monteiro were clustered together and all had varying amounts of admixture. The Moura and Landrace were placed in their own unique clusters. Among the US breeds, the Yorkshire and the Berkshire were clustered together despite presenting very different phenotypes and industry functions. It was also note that the maternal breeds Landrace, Large White and Yorkshire did not cluster together; but they were in relative close proximity to one another in the principal coordinate analysis (Figure 2).

3a.



3b.

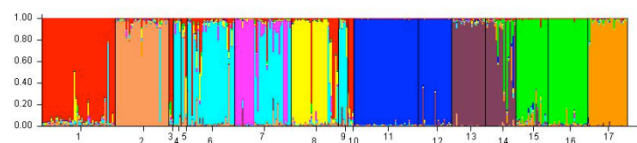


Figure 3. Genetic structure of Brazilian and US breeds when cluster number equaled 4 an 9.

1-MS60, 2-Moura, 3-Large White, 4-Nilo, 5-BR Duroc, 6-Piau, 7-Monteiro, 8-Landrace, 9-Mixed, 10-Pietrain, 11-Meishan-US, 12-Meishan-China, 13-Fengjing, 14-Minzhu, 15-Yorkshire, 16-Berkshire, 17-US Duroc

Conclusion

By merging preexisting data we were able to explore genetic diversity differences for a number of breeds across three continents. Placing the locally adapted BR breeds in the context of a wider collection of breeds showed they were not as genetically unique as was previously assumed (Figure 2). In comparing levels of heterozygosity among US breeds the merging process did result in lower levels when compared to a previous study which was likely due to the reduced number of loci in the analysis. Such a result suggests that if merged marker panels are reduced, a country can perform a separate within country analysis in conjunction with the merged analysis and compare genetic diversity parameters. Despite this finding, and the increased use of SNP data, merging country microsatellite datasets provides an effective and least cost approach to better understand genetic diversity across international boundaries.

Literature Cited

- Blackburn, H.D., Plante, Y., Rohrer, G. et al. 2014. *J. Anim. Sci.* doi:10.2527/jas2014-7274.
- Blackburn, H.D., Toishibekov, Y., Toishibekov, M. et al. 2011. *Genetica*, 139:1169-1178.
- Goudet, J. 2002. <http://www2.unil.ch/popgen/softwares/fstat.htm>
- Paiva, S.R., Mariante, A. and Blackburn, H.D. 2011. *J. Heredity*, 102:697-704.
- Peakall, R., and P.E. Smouse. 2012. *Bioinformatics*, 28:2537-2539.
- Presson, A.P., Sobel, E.M., Pajukanta, P. et al. 2008. *BMC Bioinformatics*, 9:317
- Pritchard, J.P., Stephens, M. and Donnelly, P. 2010. *Genetics*, 155:945-959.
- Silva, E.C., Dutra Junior, E.M., Ianella, P. et al. 2011. *R. Bras. Zootec.*, 8:1691-1699.
- Sollero, B.P., PAIVA, S.R., FARIA, D.A. et al. 2009. *Livestock Science*, 123:8-15.