Proc. Assoc. Advmt. Anim. Breed. Genet. 20:147-150

POPULATION STRATIFICATION AND BREED COMPOSITION OF AUSTRALIAN TROPICALLY ADAPTED CATTLE

L.R. Porto-Neto[1], S.A. Lehnert[1], M.R.S Fortes[2], M. Kelly[2] and A. Reverter[1]

[1] CSIRO Food Futures Flagship and Animal, Health and Food Sciences,
306 Carmody Road, St Lucia, QLD 4067
[2] The University of Queensland, QAAFI Centre for Animal Science, St Lucia QLD 4072

SUMMARY

Population stratification and differences in individuals' ancestry can potentially bias genome-wide genetic analyses when they are not detected and included in the genetic model. This is especially important in situation where little is known about the extent and sources of stratification. Here a large sample of tropically adapted cattle, Brahman (BB) and Tropical Composite (TC), genotyped for more than 700K SNP loci were evaluated for population stratification using principal components and supervised hierarchical clustering analyses. The BB cattle were more homogeneous than the TC cattle in both analyses, reflecting the TC's more recent and complex origin. Nevertheless, within both breeds there were degrees of variability. The effect of farm of origin was also noticeable, particularly in TC. These analyses indicated that a simple breed designation, BB or TC, encompasses large variation in ancestry within breed. This opens the question whether ancestry composition should be included in downstream analyses. Appropriate use of information on ancestry composition could aid genome-wide association studies and genomic selection.

INTRODUCTION

The detection of population stratification and estimation of ancestry composition are *per se* a field of study that is fits within population genetics and dynamics. There are several factors that might create stratification of a population, some with real biological meaning and others due to experimental artifact. It has been shown that population stratification can cause spurious associations in genome-wide studies (Price *et al.* 2006; Ma *et al.* 2012). Therefore, information on both stratification and ancestry, are very relevant either as a final result or to be taken into account in genome-wide genetic analyses.

In Australia, most beef production operations are located in Northern regions, where the climate is warm, the environment is tropical and infested with parasites. Under these conditions, tropical adaptation is imperative for cattle to thrive. *Bos primigenius indicus* or Zebu cattle (e.g. Nelore) and *Bos primigenius taurus* or Taurine (e.g. Angus) evolved under different environmental pressures, and these natural adaptations are exploited by cattle breeders to improve herd productivity. A good example in Australia is the expansion of Brahman (BB), a Zebu breed that was graded up using Taurine cattle, and the Tropical Composite (TC), which involves crosses of Zebu, Taurine and, in some cases, African cattle. Given their formation, it is expected that both breeds would have a range of ancestry compositions.

In this study a large sample of tropically adapted cattle, BB and TC was evaluated for its potential population stratification and individual ancestry composition were estimated. Furthermore, the estimated ancestry composition was compared to farm origin of the animals.

MATERIAL AND METHODS

Animals. The tropically adapted breeds of BB (n=3,502) and TC (n=2,550), representing 21 and 7 different farms of origin were included in this study. These cattle were from the CRC for Beef Genetic Technologies, Beef CRC, which was described previously (Barwick *et al.* 2009;

Burns *et al.* 2013). To anchor the breed composition estimation a sample of Zebu cattle represented by the Nelore (n=29) and Taurine represented by the Angus (n=81) were included in the analysis.

Genotypes. Cattle were genotyped using either the BovineSNP50 or BovineHD array (Illumina Inc., San Diego, CA 2006). Animals that were genotyped using the smaller array were imputed to a higher density using Beagle 3.2 (Browning and Browning 2009). To reduce the potential bias in the analyses due to a large number of markers with high linkage disequilibrium (LD) and to reduce the computational time, the full dataset was pruned by LD using PLINK v.1.07 software (Purcell *et al.* 2007) to exclude one SNP of a pair that had $r^2>0.5$ calculated in a sliding window of 50 SNP. After pruning, the combined BB and TC dataset included 229,235 SNP genotypes.

Population structure and breed composition estimation. The structure of the population was explored by principal components analysis of the genetic relationship matrix based on the SNP genotypes, both calculated using GCTA (Yang *et al.* 2011). The breed composition estimation was performed using a supervised hierarchical clustering implemented in Admixture (Alexander *et al.* 2009) set at K=2 subpopulations, using the Nelore and Angus breeds as the two reference clusters.

RESULTS AND DISCUSSION

The genetic relationship matrix that considers the four different cattle breeds in a principal components analysis is expected to result in clusters that agree with breed designations. The main split in domestic cattle correspond to Zebu vs Taurine cattle, which can be argued to have occurred 330,000 years ago (Achilli *et al.* 2008). This split, here represented by the Nelore vs Angus distance, accounts for most of the variation resulting in extreme positions in the first principal component (Figure 1), with the main TC cluster positioned approximately half way between Nelore and Angus. This is in agreement with results previously described for other composite breeds (Harrison *et al.* 2009). Comparing the tightness of the Nelore and Angus clusters to TC and BB it is clear that there is more variation within the latter two breeds. This large variation within breed is particularly evident in the TC cluster. However, variation is also seen within BB, where a number of individuals are positioned closer to the TC and Angus clusters.

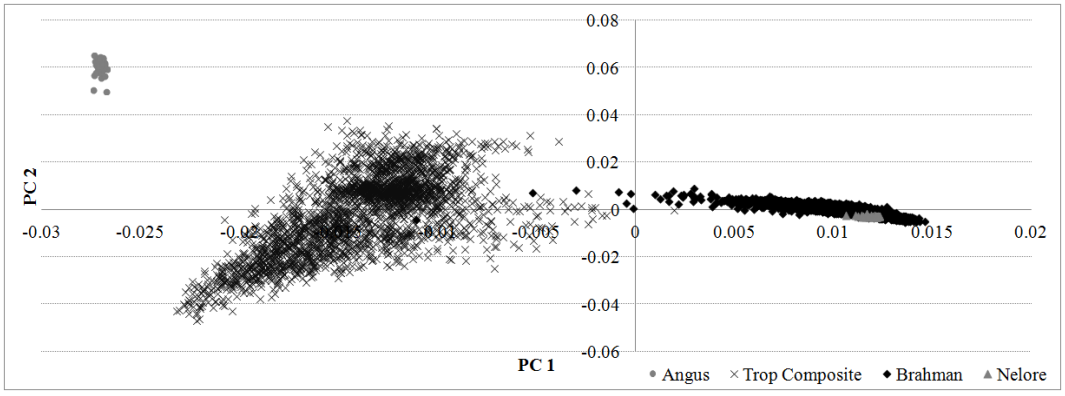

Figure 1. First Cartesian plan of principal components of the genetic relationship matrix between Angus, Tropical Composite, Brahman and Nelore animals.

In hierarchical clustering analyses TC estimated ancestry proportions attributed to Zebu varied from 0.00 to 0.79, and averaged 0.35 (Figure 2). In BB the estimated Zebu content varied from

0.46 to 1.00, and averaged 0.95. Once again, the BB cattle showed more homogeneity than TC. The Angus and Nelore reference breeds were chosen as proxies of original Taurine and Zebu cattle. However, other breeds also contributed to the formation of TC and BB. Whether the inclusion of other reference breeds in the analysis would improve the resolution of those estimates remains to be tested.

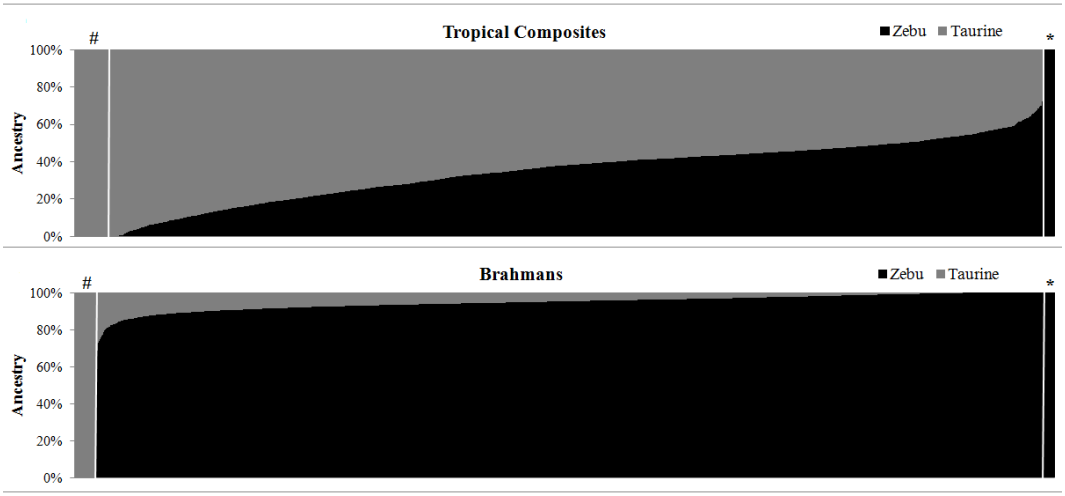

Figure 2. Breed composition (Zebu vs Taurine) estimated by supervised hierarchical clustering. A vertical bar represent each individual along the x-axis, Tropical Composites (n=2,550) and Brahmans (n=3,502) using Nelore (*) and Angus (#) as reference populations.

It is a reasonable assumption that within a breed designation, the farm of origin of an individual would reflect the Zebu vs Taurine proportion of its ancestry. However, as shown in Figure 3, the Zebu content varied between and within farm of origin with large standard deviations, and also within a breed designation.

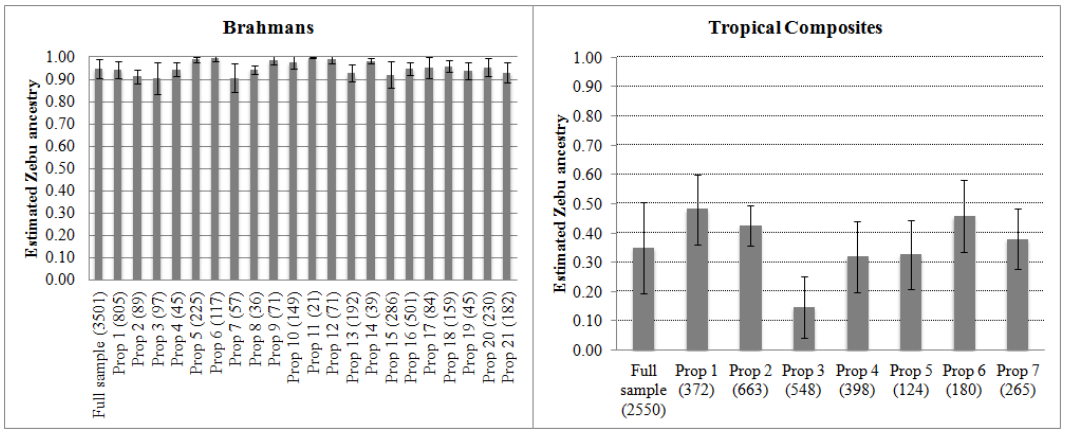

Figure 3. Estimated Zebu ancestry of Brahmans and Tropical Composites, averaged per farm of origin and its standard deviations. X axis: farm and number of animals sampled.

Both analyses demonstrated that there is genetic variation within TC and BB, although this was more evident within TC than in BB. The large spread within breed seen for TC and BB in the principal components analysis is strongly suggestive of differences in breed composition or stratification, and it could be partially attributed to differences in estimated Zebu to Taurine ancestry ratio of each individual. Population stratification on BB animals of the CRC was expected given previous results (Fortes *et al*. 2011). Importantly, the origin of the animal did not fully explain the population stratification; as large variation was also seen within each origin. Hence, using farm of origin or breed designation as factors in genetic analysis of these populations does not correct for the differences seen within. Further analyses are required to better explore and understand potential stratification of this population, to correlate the principal components and hierarchical clustering results, and to evaluate whether including estimated breed ancestry in genome-wide analyses improves the reliability of such analyses.

The large variation observed within and between BB origins and the relative high proportion of Taurine content estimated for some BB farms are interesting findings. Is the animal selection within those farms selecting "Taurine alleles" or "chromosomal segments" that were introduced long ago during the grading up of BB in Australia? On the other hand, is the TC variation in Zebu ancestry due different breeding strategies or due to selection decisions made in response to finding that animals with more or less Zebu content thrive in a particular environment? These open questions should be targeted in future research.

ACKNOWLEDGEMNTS

The authors thank the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC) and its Legacy database for access to the data. We thank W. Barendse, R. McCulloch, B. Harrison and R. Bunch for generating genotypes, and T.S. Sonstegard for access to Nelore genotypes. Financial support for genotyping TC bulls was provided separately by Meat and Livestock Australia (project code B.NBP.0723).

REFERENCES

Achilli A, Olivieri A, Pellecchia M, Uboldi C, Colli L, Al-Zahery N, Accetturo M, Pala M, Kashani BH, Perego UA, Battaglia V, Fornarino S, Kalamati J, Houshmand M, Negrini R, Semino O, Richards M, Macaulay V, Ferretti L, Bandelt HJ, Ajmone-Marsan P and Torroni A (2008) *Curr. Biol.* **18**:R157.

Alexander D.H., Novembre J. and Lange K. (2009) *Gen. Res.* **19**:1655.

Barwick S. A., Johnston D. J., Burrow H. M., Holroyd R. G., Fordyce G., Wolcott M. L., Sim W. D. and Sullivan M. T. (2009) *Anim. Prod. Sci.* **49**:367.

Browning B.L. and Browning S.R. (2009) *Am. J. Hum. Gen.* **84**:210.

Burns B. M., Corbet N. J., Corbet D. H., Crisp J. M., Venus B. K., Johnston D. J., Li Y., McGowan M. R. and Holroyd R. G. (2013) *Anim. Prod. Sci.* **53**:87.

Fortes M.R.S., Bolormaa S., Porto Neto L.R., Holroyd R.G. and Reverter A. (2011) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **19**:267.

Harrison B.E., Bunch R.J., McCulloch R., Williams P., Sim W., Corbet N.J. and Barendse W. (2012) *Anim. Prod. Sci.* **52**:890

Ma L., Wiggans G.R., Wang S.W., Sonstegard T.S., Yang J., Crooker B.A., Cole J.B., Van Tassell C.P., Lawlor T.J. and Da Y. (2012) *BMC Geno.* **13**:536.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. and Reich D. (2006) *Nat. Gen.* **38**: 904.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J. and Sham P.C. (2007) *Am. J. Hum. Gen.* **81**:559.

Yang J.A., Lee S.H., Goddard M.E. and Visscher P.M. (2011) *Am. J. Hum. Gen.* **88**:76.