Proc. Assoc. Advmt. Anim. Breed. Genet. 19:267-270

PRINCIPAL COMPONENT ANALYSIS IN A POPULATION OF BRAHMAN BULLS GENOTYPED WITH 50K SNP CHIP REVEALED A GENETIC STRUCTURE

M.R.S. Fortes[1,2], S. Bolormaa[3], L.R. Porto Neto[1,4], R.G. Holroyd[5], A. Reverter[1]

Cooperative Research Centre for Beef Genetic Technologies
[1]CSIRO Livestock Industries, Queensland Bioscience Precinct South, St Lucia, QLD 4067
[2]The University of Queensland, School of Veterinary Science, Gatton, QLD 4343
[3]Biosciences Research Division, Department of Primary Industries Victoria,
1 Park Drive, Bundoora 3083
[4]The University of Queensland, School of Animal Studies, Gatton, QLD 4343
[5]Department of Employment, Economic Development and Innovation,
Rockhampton Qld 4701, Australia

SUMMARY

We report on the principal component analysis (PCA) carried out on single nucleotide polymorphism (SNP) genotype data for a population of 1,130 Brahman bulls from the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC). Bulls were born between 2004 and 2008 in 5 different locations (places of birth or origins) and represented 55 sire families. Bulls were genotyped with the 50k Illumina SNP chip. Quality control and genotype imputing resulted in 41,028 SNP with complete genotypes across 1,115 bulls. These genotypes were used in the PCA that revealed the existence of 3 (PC1 vs. PC2) or 5 (PC1 vs. PC3) groups in the population. The results indicate that there is genetic structure in the population, which is partially explained by sire families and bull origin.

INTRODUCTION

Principal component analyses (PCA) have been widely used to detect population structure in animals and humans. Population structure could be the result of geographical migration or reflect isolation. Both events are detected by PCA as groups that appear genetically divergent (Reich *et al.* 2008). Groups that are observed in PCA may also reflect cattle breed differences (Gibbs *et al* 2009; Porto Neto and Barendse 2010) and be influenced by family structure (Patterson *et al.* 2006). Further, knowledge about population structure can be used in correcting for stratification bias in genome wide association studies (GWAS) (Price *et al.* 2006).

In this study, we investigated the genetic structure of a population of Brahman bulls from the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC). These bulls are central to a project focused on measuring reproductive traits (Corbet *et al.* 2009). The project includes genome wide association studies to identify chromosomal regions associated with male cattle reproduction. The pedigree of the bulls under investigation is known and we hypothesise that the presence of 55 sire families will be reflected in the results of principal component analysis.

MATERIALS AND METHODS

Animals. Blood samples for DNA extraction were obtained from 1,130 Brahman bulls, which were the progeny of 55 industry sires mated to the cows from the Beef CRC Lifetime Performance Population previously described (Barwick *et al.* 2009; Johnston *et al.* 2010; Johnston *et al.* 2009). They were born between 2004 and 2008, in 5 properties across Queensland, including the Belmont Research Station (25 Km NW of Rockhampton). The different properties defined 5 origins according to place of birth: BEL, CPC, MDH, TTS and CCK.

Genotypes and edits. The BovineSNP50 bead chip (Matukumalli *et al.* 2009) was used to genotype the samples according to the manufacturer's protocols (Illumina Inc., San Diego, CA). Repeat samples were included in the genotyping for quality assurance and the Bead Studio software (Illumina Inc., San Diego, CA 2006) was used to determine genotypes. Genotype edits were carried out as follows: SNP were discarded if they did not have a call rate greater than 90% and genotypes of animals with genotype calls (GC) < 0.6 were treated as missing genotypes. After this step, SNP not located in chromosome X were discarded if they departed from Hardy-Weinberg equilibrium at P < 0.0001. Finally, SNP were discarded if the proportion of missing genotypes was greater than 20% or if the minor allele frequency (MAF) was less than 0.05. After these edits, missing genotypes were imputed using the BEAGLE 3.2 program (Browning and Browning 2010). Quality control and genotype imputing resulted in 41,028 SNP with complete genotypes for 1,115 bulls.

Statistical Analysis. Principal component analyses (PCA) was conducted using *smartpca* from EIGENSOFT 3.0 (Patterson *et al.* 2006), using default parameters. The resulting eigenvalues for PC1 were plotted against those for PC2 and PC3 for visualizing groups, or structure, in the population. A hypergeometric distribution test (Mood *et al.* 1974) was used to examine if groups of bulls were significantly overlayed by their sire family. The Chi-square test of independence (Mood *et al.* 1974) was used test the overall independency between principal component grouping and the origin of bulls.

RESULTS AND DISCUSSION

The PCA revealed the existence of three main groups when PC1 was plotted against PC2. The plotting of PC1 versus PC3 divided the population in 5 main groups. Both of these observations indicate the presence of a genetic structure in the Beef CRC population (Figure 1, A and B).

When the three groups separated by PC1 versus PC2 were overlayed with sire information (Figure 1 C), one sire was significantly related ($P < 0.0001$) to the distinct group in the lower half of Figure 1 C (bulls with lower PC2 values). Further, out of the remaining 54 sires, only 14 were represented in the top left group of Figure 1 C. Three bulls out of these 14 were exclusive to this group. The probability of being sired by these three bulls and simultaneously belonging to the distinct top left group was high ($P < 0.0004$). Thus, for four of the sires, PCA and sire grouping were completely confounded. These results could indicate that those four Brahman sires are genetically different from the remaining families. It is also possible to speculate that they are carriers of chromosome segments from other breeds (*Bos taurus* crossbred ancestry) or from a distinguishable population of *Bos indicus*. Further, in this population not all sires contributed to a similar proportion of offspring distributed across origins. Unequal contributions of sires can affect PCA results, as PC1 favours correlated data points.

When the 5 groups revealed by PC1 versus PC3 were overlayed with origin information (Figure 1 D) they did not exactly overlap. Nevertheless, PC1 results were not independent from origin grouping, according to Chi-squared test ($P < 6.02E-16$, Table 1). For example, bulls from TTS and CCK were observed to group together and were distant from others by presenting higher PC1 values (Figure 1 D). Therefore, at least some of the variance explained by PC1 and PC2 could be attributed to origin of bulls, as well as sire families and the effect of unequal sire contributions to this population.

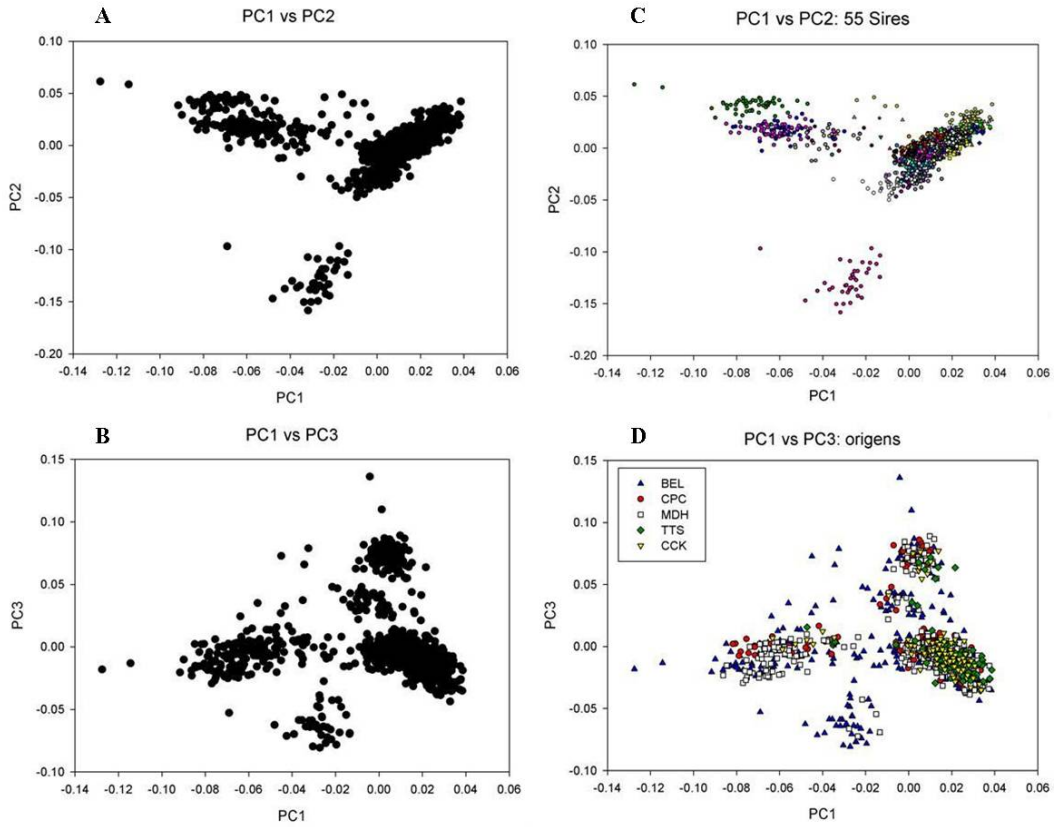

Figure 1. Principal Component Analysis: **A. PC1 vs. PC2 separated the population into 3 main groups. B. PC1 vs. PC3 separated the population into 5 main groups. C. PC1 vs. PC2 colour coded to represent 55 sire families. D. PC1 vs. PC3 colour coded to represent 5 origins.**

Table 1. Number of bulls from each origin corresponding to the groups separated by PC1

PC1 groups	Origin*					
	BEL	CCK	CPC	MDH	TTS	Total
PC1 > 0.02	215	96	160	260	77	808
PC1 ≤ 0.02	153	8	52	89	5	307
Total	368	104	212	349	82	1115

* $P < 6.02E\text{-}16$

Table 2 presents the proportion of variance explained by the first three principal components along with their corresponding eigenvalues. Previous studies performed PCA to compare multiple cattle breeds and reported that PC1 explained between 16 and 19% of the variance (Porto Neto and Barendse 2010). By comparison, the proportion of variance explained by PC1 in the present study seems small. Our results may reflect a degree of homogeneity in the population, a consequence of studying a single breed. This can also be a consequence of groups that are overlayed by family

structure, where individual bulls are more related that they would be in random sets of animals from different populations or breeds.

Table 2. Proportion of the variance explained

Principal component	Proportion of the variance explained (Percent)	Eigenvalue
PC1	1.67%	18.619
PC2	1.40%	15.609
PC3	1.22%	13.601

CONCLUSION

Population structure was detected within the 1,115 Brahman bulls of the Beef CRC, using PCA. Partially, this population structure could be attributed to different origins of bulls and sire families. Further research is needed to elucidate other sources of population structure since not all the groupings we detected with PCA could be explained by origin and sire family. This structure is an important consideration for future genome wide association studies planned for this population, as it may influence SNP association results.

ACKNOWLEDGEMENTS

Northern Pastoral Group, Department of Employment, Economic Development and Innovation (DEEDI) and CSIRO Livestock Industries collaborated to collect phenotypes. Meat and Livestock Australia, Australian Centre for International Agricultural Research, Cooperative Research Centre for Beef Genetic Technologies, Northern Pastoral Group, CSIRO - LI, DEEDI and The University of Queensland gave financial support.

REFERENCES

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) *Animal Production Science* **49:** 367.

Browning S.R. and Browning B.L. (2010) *Am J Hum Genet* **86:** 526.

Corbet N.J., Burns B.M., Corbet D.H., Johnston D.J., Crisp J.M., McGowan M., Prayaga K.C., Venus B.K. and Holroyd R.G. (2009) *Proc. Assoc. Adv. An. Breed. Genet.* **18:** 121.

Gibbs R.A., Taylor J.F. *et al.* Bovine HapMap Consortium (2009) *Science* **324:**528.

Johnston D.J., Barwick S.A., Corbet N.J., Fordyce G., Holroyd R.G., Williams P.J. and Burrow H.M. (2009) *Animal Production Science* **49:** 399.

Johnston D.J., Barwick S., Fordyce G. and Holroyd R.G. (2010) *9th World Congress on Genetics Applied to Livestock Production*, Leipzig, Germany.

Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F., Heaton M.P., O'Connell J., Moore S.S., Smith T.P., Sonstegard T.S. and Van Tassell C.P. (2009) *PLoS One* **4:** e5350.

Mood A.M., Graybill A.G. and Boes D.C. (1974) 'Introduction to the theory of statistics' 3 ed. McGraw-Hill Publishing Company, New York.

Patterson N., Price A.L. and Reich D. (2006) *PLoS Genet* **2:** e190.

Porto Neto L.R. and Barendse W. (2010) *Animal Production Science* **50:** 792.

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

Reich D., Price A.L. and Patterson N. (2008) *Nature Genetics* **40:** 491.