

smallholder dairy cattle in Kenya and Uganda. Genotype data resulted from the BovineHD Beadchip array (Illumina Inc). Of the 777,962 SNP on the array, 566,056 were selected after quality control process (Gondro and Gibson, 2012). Since the crossbred dairy cows are studied, Y chromosomes were excluded. Only the SNP on the 29 autosomes were retained for analysis. The genotypes for the three main indigenous breeds Ankole (n=44), Nganda (n=16) and Small East African zebu (SEAZ) (n=58) were also included in the study. Since pedigree data is not available, 5 breeds from the international Bovine Hapmap 770k data set were selected as reference breeds: Nelore (reference *Bos indicus* breed; n=35), Brown Swiss (n=24), Holstein (n=66), Jersey (n=40) and N'Dama (reference African *Bos taurus* breed; n=24) and all the reference animals were used for all the analysis. Although Friesian and Ayrshire have been reportedly widely used in the East Africa region, these breeds were not available in the Hapmap data set.

A principal components analysis (PCA) was performed to help describe the genetic structure of our East African cattle sample. The analyses included the Hapmap animals to provide reference points for interpretation of the East Africa samples. PCA was based on the genomic relationship matrix, obtained using the allele frequency method (VanRaden, 2008). The dairy proportions for crossbred animals were estimated using the ADMIXTURE program (Alexander *et al.* 2009). The program was run setting the number of breed origins from K=2 to K=6. The farmer's prediction of exotic breed proportion was recorded and categorised into 7 groups: 0%, 25%, 50%, 65%, 75%, 85% and 100% dairy proportions. The correlation between farmer prediction and genomic prediction was estimated.

RESULTS AND DISCUSSION

Principal component 1 (PC1) separates *Bos taurus* from *Bos indicus* (Figure 1A) and explains 87% of the variation among animals (Table 1) PC2 separates the African *Bos taurus* N'Dama breed from other breed groups (*Bos indicus* and European *Bos Taurus* ;Figure 1A), and explains 1.84% of the variation (Table 1). Altogether the first five principal components explain 92% of the variation in the dataset.

Table 1. Summary of PCA for the first five principal components for the African dairy cattle and bovine Hapmap animals

Parameter	PC1	PC2	PC3	PC4	PC5
Standard deviation	3.677	0.533	0.474	0.347	0.217
Proportion of variance	0.877	0.018	0.014	0.009	0.003
Cumulative proportion	0.877	0.895	0.909	0.918	0.921

Ankole is one of the Sanga type breeds which are believed to include some African *Bos taurus* and *Bos indicus* ancestry. This is confirmed by their position in Figure 1A on the axis between the Nelore and N'Dama breeds. The Ankole animals do not cluster as tightly as the reference breeds. While a couple of Ankole outliers are clearly animals that have a small proportion of European *Bos taurus* content (they lie on the axis between Ankole and European taurine), it is not clear whether the remaining Ankole are simply more diverse than reference breeds or they might contain a low proportion of contamination with other breeds. The same picture is evident for the Nganda and SEAZ breeds, with diversity being substantially higher again. The SEAZ breed contains several animals that clearly have a substantial European taurine content. The SEAZ is not a single breed but consists of a range of small framed zebu breeds. The SEAZ position closer to the Nelore end of the Nelore to N'Dama axis, but not as distant from Ankole as their appearance might suggest: SEAZ look like *Bos indicus* while Ankole look substantially taurine.

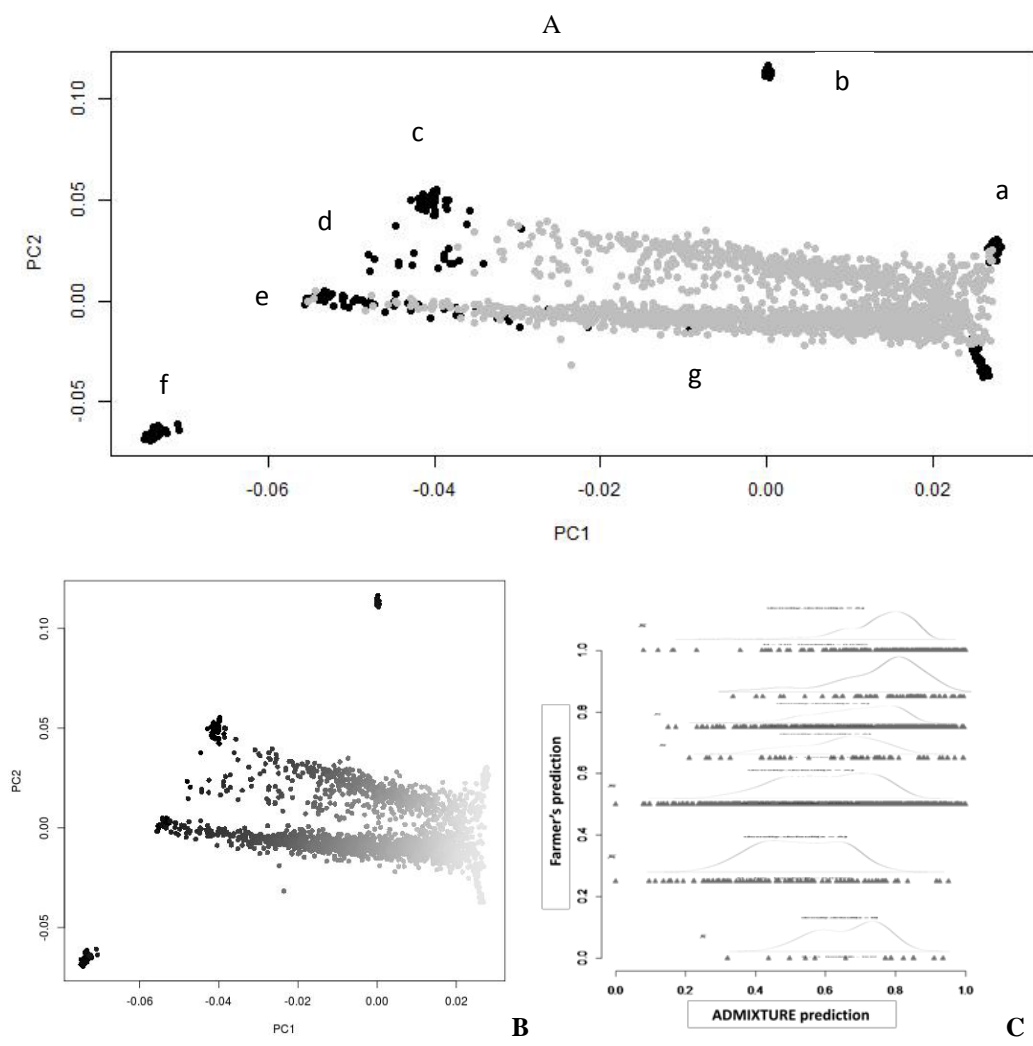


Figure 1. A. PC1 plotted against PC2 to visualise the genetic structure of East African cattle samples (a-European *Bos taurus*, b-N'Dama, c-Ankole, d-Nganda, e-SEAZ, f-Nelore, g-crossbred). **B.** Same as A, but with estimated dairy breed percentage from ADMIXTURE represented on a scale from black = 0% dairy to light grey = 100% dairy. **C.** Distribution of ADMIXTURE predictions of dairy breed proportion for individual animals shown for each class of farmer predicted breed composition.

The results for SEAZ agree with the study of Rege *et al.* (2001) where microsatellite allele data on autosomes indicated that all the Kenyan zebu breeds have a mixture of both taurine and indicine genes. Nganda animals are known to result from relatively recent (approximately 100 years) crosses between Ankole and SEAZ. Consistent with this origin, they lie in the middle of the axis between Ankole and SEAZ.

The estimated breed proportions of small holder crossbred dairy cattle vary from 0% to 100 %. Figure 1B shows the dairy proportion estimated using K=4 on a scale representing dairy

proportion from black (= 0% dairy) to light grey (=100% dairy). The Nelore, indigenous animals and N'Dama all are 0% dairy, while Holstein, Jersey and Brown Swiss are 100% dairy. The estimated proportion of dairy from ADMIXTURE is shown for each of classes of farmer predicted breed composition in Figure 1C. The farmer's predictions were based on a combination of phenotypic appearance and varying degrees of knowledge of the ancestry of each animal. The class of 100% dairy includes all animals said to be of very high proportion dairy breed (>85%), rather than just absolutely pure animals; in practice the mean dairy breed proportion was 81%, and approximately 70% are less than 85% dairy. For cows predicted by farmers to be close to 50% dairy, the average dairy percentage is 61% and the range was from almost 0% to almost 100%. For the farmer's predicted classes of 65%, 75% and 85% dairy, the ADMIXTURE average estimates were 66%, 72% and 79%. The overall correlation of farmer's prediction and ADMIXTURE prediction is 0.41.

CONCLUSIONS

The population structure of East African smallholder dairy cattle is clearly illustrated using PCA and the results indicate that Ankole and SEAZ animals are quite similar composites of *Bos indicus* and *Bos taurus*, with Ankole having somewhat higher *Bos taurus* content but not so different as has been generally assumed given their physical appearance. All indigenous breeds have some animals that are clearly not purebred. Farmers' assessments/assumptions about the breed composition of their cows are poorly correlated with the cow's actual breed composition. This will be a problem when it comes to recommending the best breeding options for such cows, when the goal is produce a replacement female of a breed composition best suited to that farm environment.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of other Dairy Genetics East Africa project team members, the collaborating farmers in Kenya and Uganda and the International Bovine Hap Map project for sharing their data. This work was funded by the Bill and Melinda Gates Foundation.

REFERENCES

- Alexander D. H., Novembre J. and Lange K. (2009) *Genome Research*. doi: 10.1101/gr.094052.109
- Gondro C. and Gibson J. P. (2012) Centre for Genetic Analysis and Applications, University of New England, Armidale, Australia.
- Rege J. E. O., Kahi A., Okomo-Adhiambo M., Mwacharo J. and Hanotte O. (2001) *Animal Genetic Resources Research 1* (pp. 103). Nairobi, Kenya: International Livestock Research Institute.
- VanRaden P. M. (2008) *Journal of Dairy Science*, **91**; 4414.