

Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds

A.A. Zwane^{1,2#}, A. Maiwashe^{1,5}, M.L. Makgahlela¹, A. Choudhury³,
J.F. Taylor⁴ & E. van Marle-Köster²

¹Department of Animal Breeding and Genetics, ARC-API, P/Bag X2, Irene, 0062, ²Department of Animal and Wildlife Sciences, University of Pretoria, P/Bag X20, Hatfield, Pretoria, 0028, ³Sydney Brenner Institute of Molecular Bioscience, University of the Witwatersrand, P/Bag 3, Wits, Gauteng, 2050, ⁴Division of Animal Sciences, University of Missouri, 920 East Campus Drive, Columbia, MO 65211-5300, USA, ⁵Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein 9300, South Africa

(Received 1 June 2016; Accepted 31 July 2016; First published online 20 September 2016)

Copyright resides with the authors in terms of the Creative Commons Attribution 2.5 South African Licence.

See: <http://creativecommons.org/licenses/by/2.5/za>

Condition of use: The user may copy, distribute, transmit and adapt the work, but must recognise the authors and the South African Journal of Animal Science.

Abstract

Access to genotyping assays enables the identification of informative markers that discriminate between cattle breeds. Identification of these markers can assist in breed assignment, improvement and conservation. The objective of this study was to identify breed informative markers to discriminate between three South African indigenous cattle breeds. Data from BovineSNP50 and GeneSeek Genomic Profiler (GGP-80K) assays were generated for Afrikaner, Drakensberger and Nguni, and were analysed for their genetic differentiation. Hereford and Angus were included as outgroups. Breeds were differentiated using principal component analysis (PCA). Single-nucleotide polymorphisms (SNPs) within the breeds were determined when minor allele frequency (MAF) was ≥ 0.05 . Breed-specific SNPs were identified using Reynolds F_{st} and extended Lewontin and Krakauer's (FLK) statistics. These SNPs were validated using three African breeds, namely N'Dama, Kuri and Zebu from Madagascar. PCA discriminated among the breeds. A larger number of polymorphic SNPs was detected in Drakensberger (73%) than in Afrikaner (56%) and Nguni (65%). No substantial numbers of informative SNPs ($F_{st} \geq 0.6$) were identified among indigenous breeds. Eleven SNPs were validated as discriminating the indigenous breeds from other African breeds. This is because the SNPs on BovineSNP50 and GGP-80K assays were ascertained as being common in European taurine breeds. Lower MAF and SNP informativeness observed in this study limits the application of these assays in breed assignment, and could have other implications for genome-wide studies in South African indigenous breeds. Sequencing should therefore be considered to discover new SNPs that are common among indigenous South African breeds and also SNPs that discriminate among these indigenous breeds.

Keywords: Beef cattle, genetic differentiation, minor allele frequency, polymorphisms

Corresponding author: zwanea@arc.agric.za

Introduction

In southern Africa, livestock has always played a vital role in the agricultural economies of countries such as South Africa, Namibia and Botswana, where commercial livestock enterprises, smallholder and communal farming contribute to food production, social needs and the general wellbeing of rural households (Bettencourt *et al.*, 2013). The demand for animal protein is increasing. Meeting this demand will require more efficient production to ensure long-term sustainability of production and environmental conservation (Otten & Van den Weghe, 2011). Developing countries are often richly endowed with indigenous livestock resources that are well adapted to environmental challenges, but lack productivity (Mwai *et al.*, 2015) in milk and meat production compared with imported commercial breeds (Renaudeau *et al.*, 2012).

South African indigenous cattle have unique morphological features that distinguish them from other cattle breeds (Makina *et al.*, 2014). These breeds include Nguni, Afrikaner and Drakensberger, known as Sanga cattle, which belong to the subspecies *Bos taurus africanus*. Sanga cattle breeds originated from eastern and northern Africa, the home of Sanga and Zenga cattle. Sanga are possibly crossbreds between the indigenous humpless cattle (*Bos taurus*) and Zebu (*Bos indicus*), whereas Zenga are crossbreds between Zebu and Sanga (Rege, 1999). These breeds inhabit eastern and southern Africa and are known to

be well adapted to harsh environmental conditions (Okello & Sabiiti, 2006). Nguni cattle are recognized for their ability to survive when exposed to high temperatures and low-quality grass and for their resistance to parasites and tick-borne diseases (Scholtz, 1988; Mapholi *et al.*, 2014). The Afrikaner is a hardy beef cattle breed, known for its adaptation to harsh conditions. It was used in the development of the Bonsmara, a South African composite breed (Van Marle, 1974; Mason, 1996; Strydom, 2008). The Drakensberger is known for its adaptability, especially to Sourveld regions and is regarded as one of the local indigenous breeds of South Africa. Its origin and history have not been well documented (Scholtz, 2010). Other commercial breeds such as Angus and Hereford have been continuously selected for production traits including beef yield (Scholtz, 2010; Kugonza *et al.*, 2011). Hereford and Angus are common European taurine breeds that were introduced to South Africa in 1892 and 1895, respectively (Hanotte *et al.*, 2002; Scholtz, 2010). These breeds are well known internationally, and possess good mothering ability, good growth rate, early marketability, grazing performance and good temperament (Vasconcellos *et al.*, 2003). The Hereford has contributed to the development of the South African beef industry through its role in the development of the Bonsmara breed (Scholtz, 2010).

South African beef cattle are genetically diverse. However, certain populations have been identified as critically endangered, namely Pedi and Shangaan cattle (both South African Sanga) (Rege, 1999; Mwai *et al.*, 2015). Indigenous breeds are often subjected to indiscriminate crossbreeding with exotic breeds to improve production, especially in rural areas. However, this practice leads to the loss of genetic diversity. The future utilization of indigenous genetic resources depends on their conservation, promotion and improvement (Frese *et al.*, 2014). There is an urgent need to characterize South African indigenous cattle populations using genomic information to enhance their productivity and to inform on their utilization in breeding programmes (Hanotte *et al.*, 2010; Mwai *et al.*, 2015). However, there is limited knowledge of their genetic composition (e.g., Makina *et al.*, 2014).

Knowledge of breed composition may enable better understanding of the basis of adaptive traits of cattle in their own production environments, which is critical for genome-wide association studies (Kuehn *et al.*, 2011) and for assigning individuals to their population of origin (Sanz *et al.*, 2014). Furthermore, understanding the breed composition of these cattle populations could be useful in predicting heterosis (Kuehn *et al.*, 2011), and assisting with the proper management of genetic resources for long-term sustainability (Gorbach *et al.*, 2010).

Several studies have shown the utility of single-nucleotide polymorphisms (SNP) markers for breed differentiation and individual assignment (Yoon *et al.*, 2008; Negrini *et al.*, 2009; Pariset *et al.*, 2010; Kuehn *et al.*, 2011; Lewis *et al.*, 2011; Wilkinson *et al.*, 2011; Dimauro *et al.*, 2013; Hulsegge *et al.*, 2013). Individual assignment uses genetic information to allocate an individual to a population and to determine the origin of unknown individuals (Negrini *et al.*, 2008). Methods of selecting informative markers to discriminate among breeds and assign individuals to their population of origin have been described (Negrini *et al.*, 2009; Ramos *et al.*, 2011; Wilkinson *et al.*, 2011; Opara *et al.*, 2012). A relatively small number of SNPs can be used to elucidate the genetic structure among breeds (Wilkinson *et al.*, 2011), and only a small set of informative SNPs, if chosen appropriately, may be needed for accurate breed assignment (Mackay *et al.*, 2008; Hulsegge *et al.*, 2013; Martinez-Cambor *et al.*, 2014). High-density SNP assays, such as the BovineSNP50 and bovine high-density (BovineHD) are now available with large numbers of SNPs from which the most informative SNPs can be selected for breed assignment (Matukumalli *et al.*, 2009). The objective of this study was to identify breed informative SNPs for differentiating among three South African cattle breeds using genotype data generated with the BovineSNP50 and GGP-80K assays.

Materials and Methods

Genotype data generated from previous projects were available for this study (Makina *et al.*, 2014). Data from five breeds were studied, including three indigenous South African breeds (20 Afrikaner, 48 Drakensberger and 47 Nguni) and two exotic British breeds (31 Angus and 33 Hereford), genotyped with the Illumina BovineSNP50 (Illumina, San Diego, Calif, USA) and GGP-80K assay (Neogen, Lincoln, Nebr., USA). Angus and Hereford were considered outgroups in this study because these breeds were included in the development of the BovineSNP50 chip. Afrikaner and Hereford, genotyped with the GGP-80K assay, were provided by the Department of Animal & Wildlife Sciences at the University of Pretoria, South Africa while Angus, Drakensberger and Nguni, genotyped with the BovineSNP50 assay, were provided by the Agricultural Research Council (ARC), South Africa.

BovineSNP50 BeadChip genotypes for 54 609 SNPs and GGP-80K assay genotypes for 88 683 SNPs were available, respectively. These chips contain highly informative SNPs that are evenly distributed throughout the autosomal genome of the major European cattle breeds (Michelizzi *et al.*, 2011). The GGP-80K assay consists of SNPs common to taurine cattle that were derived from the BovineSNP50 and BovineHD assays, but includes variants derived from these assays that are common in *Bos indicus* (Edea

et al., 2015). Genotypes from the two genotyping platforms were merged and SNPs that were common to both BeadChips (i.e., 28 261 SNPs) were used for the analysis. The markers were next filtered within breeds to remove those with call rates of less than 98% and samples with more than 10% missing genotypes. After applying these filters, 26 472 SNPs remained for further analysis. To evaluate that the genotypes had been called with the same Illumina format for both assays, sets of Afrikaner ('afr', n = 48), Hereford ('hfd', n = 20) and Angus ('ang', n = 20) animals genotyped with the Illumina BovineSNP50 chip and called with the Illumina A/B format were included in the analysis (Makina *et al.*, 2014; Decker *et al.*, 2014). Furthermore, three breeds originating from Africa (N'Dama (*Bos taurus taurus*), Kuri (*Bos taurus taurus*) and Zebu from Madagascar (*Bos taurus indicus*)) were used to compare the indigenous South African breeds with these other African breeds. These breeds were among the several African breeds in the dataset, but were selected based on their demographic location since no breeds originate from neighbouring countries to South Africa.

Allele frequency estimates for each SNP marker were used to determine its utility for breed differentiation. To examine the basic indices of genetic variability between the breeds, allele frequency distributions and the proportion of SNPs were estimated within a breed using PLINK version 1.09 (Purcell *et al.*, 2007). SNP proportions by minor allele frequency (MAF) ≥ 0.05 were determined for each breed (Edea *et al.*, 2012; Grasso *et al.*, 2014; Edea *et al.*, 2015). Within-population genetic diversity was estimated by calculating observed heterozygosity (H_o), expected heterozygosity (H_e) and mean inbreeding for each population using GoldenHelix SNP & Variation Suite (SVS) software (GoldenHelix Inc., Bozeman, Mont, USA) (Grasso *et al.*, 2014).

For genetic structure analysis with PCA, 26 472 SNPs were further filtered across breeds to remove SNPs with call rates $\leq 98\%$, $MAF \leq 0.01$ or with P -value for a chi-square test for Hardy-Weinberg equilibrium (HWE) ≤ 0.0001 (Lee *et al.*, 2013). SNPs were filtered to avoid the effects of ascertainment bias on diversity indexes and genetic distances (Edea *et al.*, 2015). The average proportion of alleles shared between animals was calculated using PLINK using the commands - - cluster and - - distance-matrix and the resulting matrix was used to generate the PCA plots (Kijas *et al.*, 2012). The genotypes from 'afr', 'hfd' and 'ang' were used to check for the possibility of assay or genotyping call effects between the two datasets (BovineSNP50 and GGP-80K), using principal component analysis (PCA) analysis (Gurdasani *et al.*, 2015).

Breed-specific markers were determined by identifying markers with $MAF \geq 0$ (Grasso *et al.*, 2014). A SNP was declared to be breed specific when it possessed an allele that was present in only one breed (Ramos *et al.*, 2011). To find the SNPs that distinguished between the breeds, pairwise F_{st} (Weir & Cockerham, 1984; Weir, 1996) and Lewontin and Krakauer's (FLK) statistics (Lewontin & Krakauer, 1973) were calculated between each pair of breeds (Wilkinson *et al.*, 2011; Fariello *et al.*, 2013). Pairwise F_{st} was calculated using SVS software and the FLK statistics were calculated with the haplotype-based method in hapFLK (Bonhomme *et al.*, 2010). The method uses the genetic distance of Reynolds *et al.* (1983), and builds the population relationship tree using the neighbour-joining algorithm applied to the matrix of Reynolds distances for a specified outgroup (e.g., Angus in this study). SNP pairs with high FLK values ($P < 0.001$) and F_{st} values of 0.60 or greater were selected for each breed comparison (Nishimura *et al.*, 2012).

Results

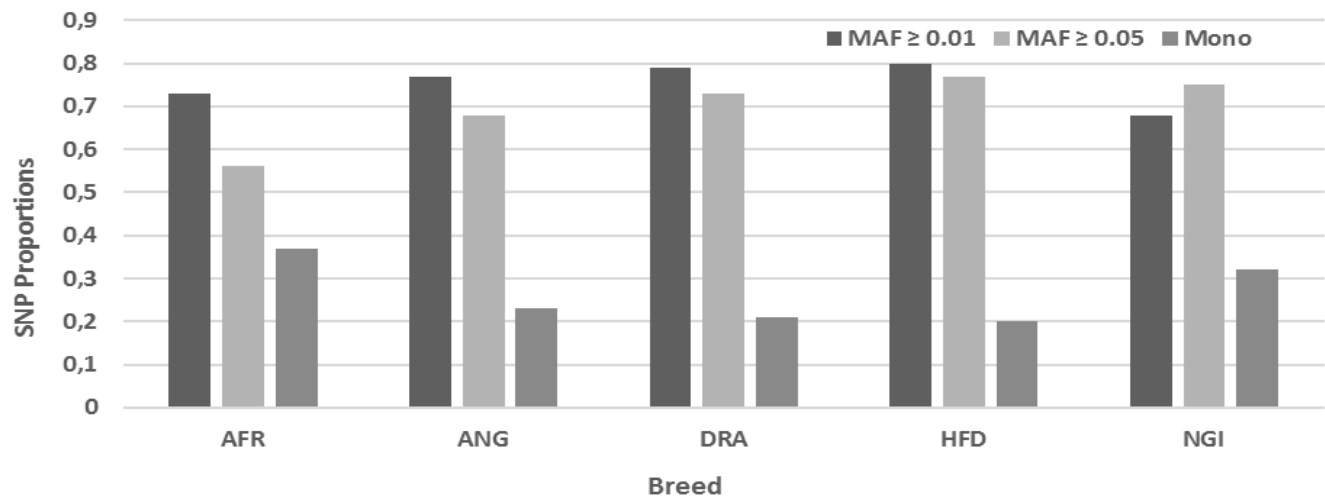
The average MAF observed for Afrikaner, Drakensberger, Nguni, Hereford and Angus is shown in Table 1. There were no differences between the indigenous and the exotic breeds for the proportion of SNP with $MAF \geq 0.01$. A small difference was observed between the Afrikaner and Nguni populations for the proportion of SNP with $MAF \geq 0$ and $MAF \geq 0.01$. However, for both MAF criteria, the Drakensberger and Angus (a taurine breed) were similar. Less than 1% of the SNPs were fixed within the South African indigenous cattle (data not shown).

As shown in Figure 1, approximately 80% of the markers had $MAF \geq 0.01$, while 65% of markers were highly polymorphic with $MAF \geq 0.05$ across the South African indigenous breeds. In the Drakensberger, 73% of the markers were highly polymorphic, which was greater than in Angus (68%). The Afrikaner had the lowest proportion of polymorphic SNPs compared with the other breeds. Angus and Hereford generally had high levels of polymorphism. Table 2 presents the measure of H_o , H_e and average inbreeding coefficients. There was low observed heterozygosity (H_o) in Hereford compared with the other breeds (Table 2).

A principal component analysis was performed to evaluate the genetic structure and affinities among the five populations included in this study. Figure 2 illustrates the clustering of the five breeds, showing the separation of indigenous South African breeds and exotic breeds (Angus and Hereford). The cluster also shows the relationship between the Drakensberger and exotic breeds. The cluster of the five breeds with 'afr', 'hfd' and 'ang' showed that the observed pattern of clustering separated these populations based on their relatedness rather than on their genotyping platforms (BovineSNP50 versus GGP-80K) or sample batches, as shown in Figure 3. These PCA results, as expected, show that the indigenous populations cluster closer to each other in comparison with the exotic breeds.

Table 1 Average minor allele frequency (MAF) and standard deviations (SD) in Afrikaner (AFR), Drakensberger (DRA), Nguni (NGI), Angus (ANG) and Hereford (HFD) cattle breeds

Population	Breed	No. of animals	Mean \pm SD (MAF \geq 0)	Mean \pm SD (MAF \geq 0.01)	Mean \pm SD (MAF \geq 0.05)
Indigenous breeds					
	Afrikaner	20	0.20 \pm 0.159	0.24 \pm 0.146	0.25 \pm 0.138
	Drakensberger	48	0.26 \pm 0.145	0.26 \pm 0.142	0.28 \pm 0.130
	Nguni	47	0.21 \pm 0.158	0.23 \pm 0.151	0.27 \pm 0.133
Outgroup taurine breeds					
	Hereford	33	0.28 \pm 0.142	0.29 \pm 0.136	0.30 \pm 0.126
	Angus	31	0.26 \pm 0.147	0.27 \pm 0.139	0.29 \pm 0.126

**Figure 1** Single-nucleotide polymorphism and monomorphism as determined by minor allele frequency = 0, MAF \geq 0.01 and MAF \geq 0.05 thresholds for each breed. (AFR: Afrikaner; DRA: Drakensberger; NGI: Nguni; ANG: Angus; HFD: Hereford).**Table 2** Indexes of genetic diversity in South African cattle breeds

Breed	Breed group	Observed heterozygosity (Ho)	Expected heterozygosity (He)	Mean inbreeding coefficient (fi)
AFR	Sanga	0.22	0.22	0.00
ANG	Taurine	0.25	0.24	0.03
DRA	Sanga	0.26	0.25	0.01
HFD	Taurine	0.20	0.20	-0.02
NGI	Sanga	0.24	0.23	-0.01

AFR: Afrikaner; DRA: Drakensberger; NGI: Nguni; ANG: Angus; HFD: Hereford.

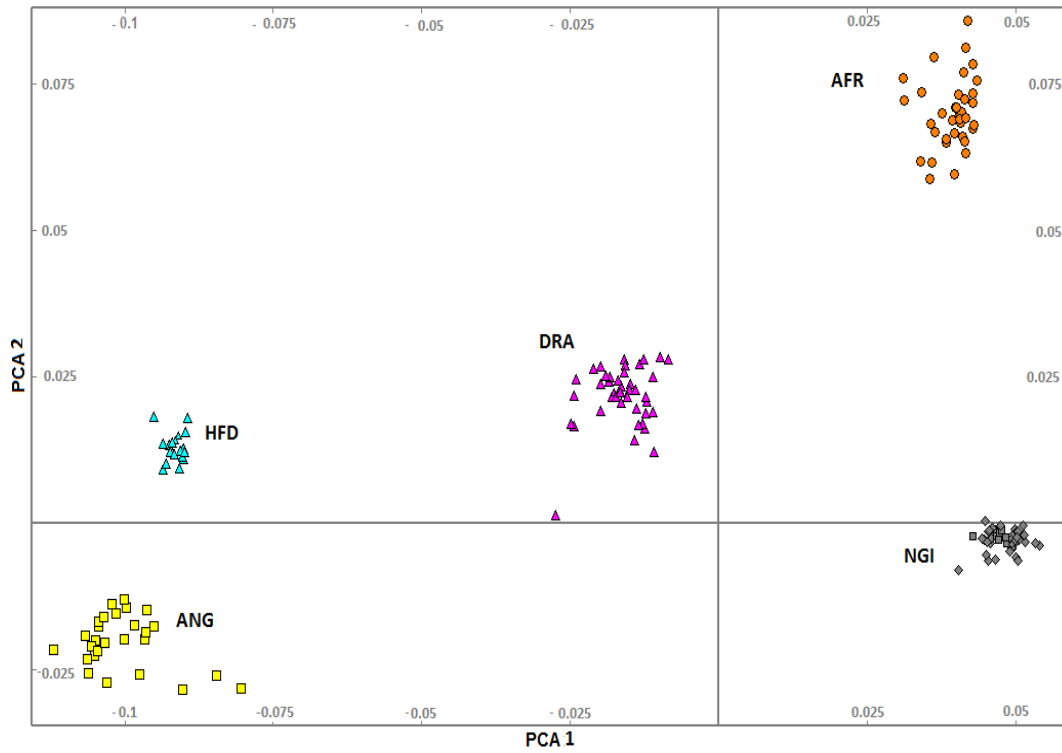


Figure 2 Principal component analysis for population structure in South African cattle populations showing first two principal components for all the breeds. (AFR: Afrikaner; DRA: Drakensberger; NGI: Nguni; ANG: Angus; HFD: Hereford).

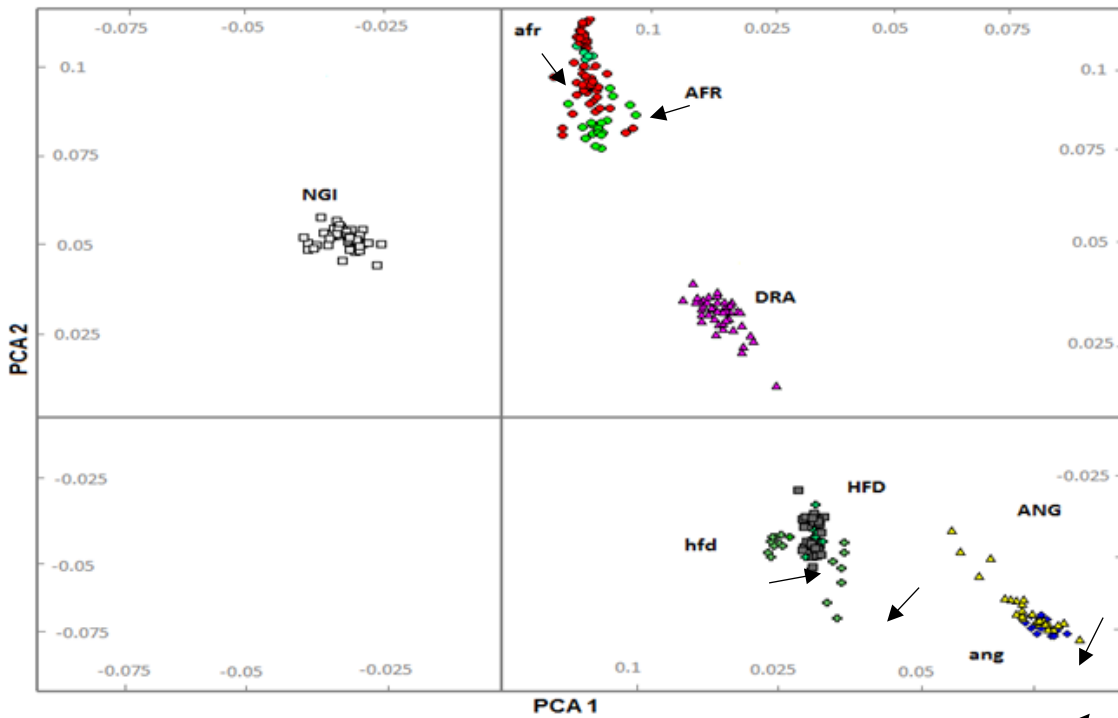


Figure 3 First two principal components for the eight breeds (five breeds including second Afrikaner (afr), Hereford (hfd) and Angus (ang) group genotyped with BovineSNP50K). (AFR: Afrikaner; DRA: Drakensberger; NGI: Nguni; ANG: Angus; HFD: Hereford).

To examine the influence of indicine introgression in the indigenous South African breeds, PCA of allele sharing was performed between South African and the other African populations (N'Dama (NDAM), Kuri (KUR) and Zebu from Madagascar (ZMA) breeds) (Figure 4). This analysis showed a clear separation between the South African and other African breeds. The clustering still placed Drakensberger on the diagonal axis between European and African taurine breeds. Regardless of the analytical method or subset of breeds analysed, these three groups were consistently observed to be highly differentiated. SNPs with highly differentiated allele frequencies were identified using pairwise Reynolds F_{st} and hapFLK analyses. From this analysis, 325 informative loci, for example SNPs with $F_{st} \geq 0.6$, were identified between South African breeds, but the hapFLK analysis provided little evidence of the existence of highly breed informative SNPs in these data ($P > 0.001$).

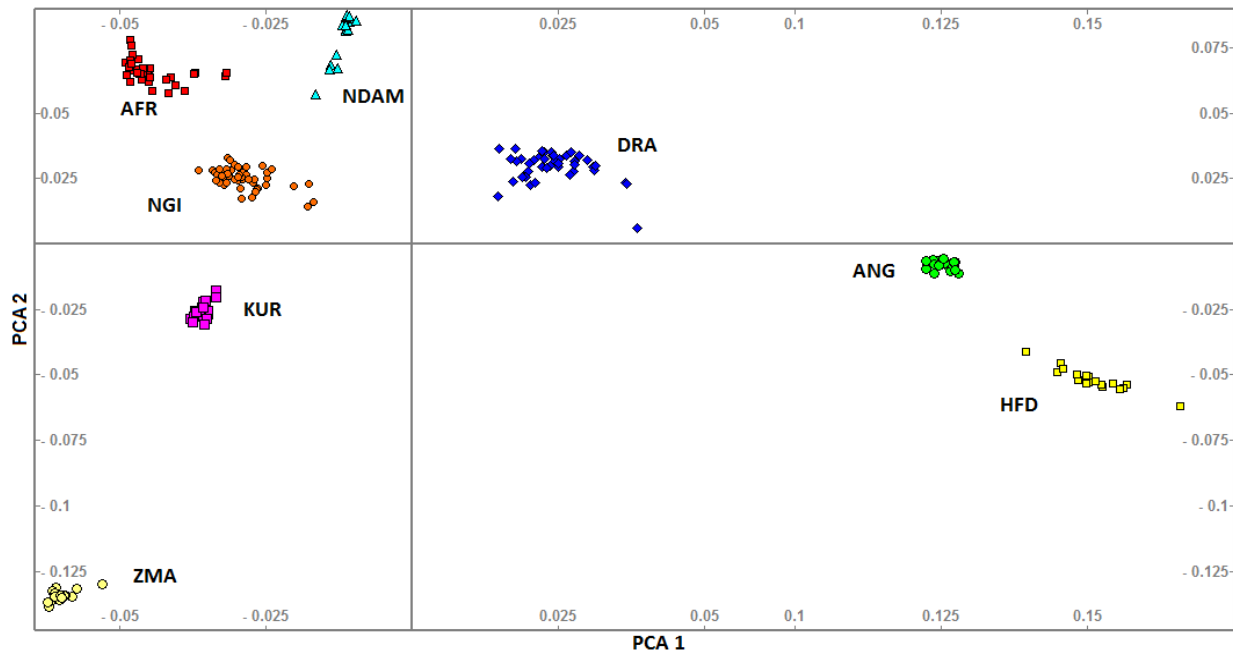


Figure 4 Clustering of five South African cattle breeds and three African breeds showing separation among the breeds. (AFR: Afrikaner; DRA: Drakensberger; NGI: Nguni; ANG: Angus; HFD: Hereford; NDAM: N'Dama; KUR: Kuri; ZMA: Zebu from Madagascar).

Analyses of South African breeds with African breeds (N'Dama, Kuri and Zebu from Madagascar) revealed a set of informative SNPs that differentiate the South African indigenous and other African (taurine \times indicine) breeds. Differentiated regions were identified on nine chromosomes (Table 3), containing SNPs with high FLK values. Furthermore, these SNPs were checked against a SNP allele frequency database at the University of Missouri to determine whether they were fixed only in the African breeds. This revealed that both alleles at each SNP segregated in at least 30 other cattle breeds for all SNPs (data not shown). Consequently, the SNPs producing high FLK values did not possess alleles that were specific to the African breeds, but are sufficiently skewed in frequency to differentiate South African cattle breeds from other African breeds.

Discussion

The analyses performed in this study were conducted to identify breed informative markers for use in discriminating among indigenous South African cattle breeds using the BovineSNP50 and GGP-80K data. A number of studies have shown the usefulness of SNP data for identifying breed informative SNPs for discrimination among breeds (Negrini *et al.*, 2009; Wilkinson *et al.*, 2011; Edea *et al.*, 2012; Nishimura *et al.*, 2012). Although the BovineSNP50 and GGP-80K assays were designed to contain variants that were common to taurine breeds, the authors decided to test their usefulness in identifying informative SNPs to discriminate between South African indigenous cattle breeds. The differences between cattle breeds for the mean MAFs ranged from 0.20 to 0.29 with an average of 0.24 (SD = 0.143) and were similar to those

Table 3 Informative single-nucleotide polymorphisms that discriminate between South African and African breeds

SNP	Chromosome No	Position	MAF
BTB-00187975	4	57553315	0.13
BTB-00432889	10	62653672	0.11
BTB-01642403	24	6512738	0.48
BTB-00363099	8	84738093	0.11
ARS-BFGL-NGS-14285	1	30024945	0.40
BFGL-NGS-109801	17	55713369	0.37
ARS-BFGL-NGS-34121	21	32113699	0.18
BTA-70284-no-rs	4	41895490	0.46
Hapmap48127-BTA-93939	1	142370512	0.14
ARS-BFGL-BAC-27254	20	44861949	0.42
BTA-118486-no-rs	19	21716537	0.32

SNP: Single nucleotide polymorphism; MAF: Minor allele frequency.

observed in previous studies (Chan *et al.*, 2008; Edea *et al.*, 2012). Studies have shown that MAF limits the utility of markers in association studies owing to the effects of rare alleles that are difficult to estimate (Gurgul *et al.*, 2013). Although rare and fixed alleles could be used to explain the distinct loci in a particular population, they may account for the reduced percentage of informative markers within the breeds (Dadi *et al.*, 2011). SNPs with low MAF have a frequency imbalance between the two allelic groups, which may reflect functional importance (Cargill *et al.*, 1999).

The differences in allele frequencies among the breeds may be caused by genetic drift, selection to adaptation to the local South African environment or ancient divergence among founder populations (MacEachern *et al.*, 2009, Dadi *et al.*, 2012). The lowest average MAF was observed in Afrikaner (0.20). Drakensberger had a higher average MAF compared with the Afrikaner and Nguni. This agrees with the study by Makina *et al.* (2014), which revealed the closer relationship of Drakensberger to the European taurine breeds. The overall MAF in South African indigenous breeds was lower compared with the European taurine breeds, which may be due to ascertainment bias in the design of the BovineSNP50 assay. McKay *et al.* (2008) also found a lower average MAF for *Bos indicus* than for *Bos taurus* breeds. The lower MAF in *Bos indicus* could again reflect the lower representation of indicine populations in the design of these assays such that common loci identified in taurines are generally not the most common in indicines (Chan *et al.*, 2008; Edea *et al.*, 2012; Espigolan *et al.*, 2013). Therefore, it is possible that the SNPs that have been identified as being useful in one population may not necessarily be as useful in another (Allen *et al.*, 2010). The differences in observed allele frequencies among breeds show the genetic diversity that exists within and between the breeds (Allen *et al.*, 2010).

The proportion of SNP polymorphisms that was common to South African indigenous cattle breeds was generally lower than for the two British breeds, except for Drakensberger. This result was expected because these breeds were included in the design of the bovine SNP assays. The higher level of polymorphisms in Drakensberger is likely related to the admixture that occurred in the development of the breed, which was observed by Makina *et al.* (2014). The higher degree of SNP polymorphisms observed in indigenous breeds could have contributed to the inability to find SNPs with alleles that are specific to South African indigenous breeds, and could have other significant impacts on the design and application of marker association studies in South African populations. Studies have indicated that the majority of the SNP markers on the BovineSNP50 BeadChip were discovered in Angus, Holstein and Hereford (Van Tassell *et al.*, 2008) and could have influenced the level of SNP informativeness in such a way that the breeds used in the discovery process show higher MAFs and SNP variability, manifested in this study in Angus and Hereford compared with the South African indigenous breeds.

The separation of indigenous South African breeds from Angus and Hereford populations was consistent with the current understanding of ancestry and population structure in South African populations (Makina *et al.*, 2014). It was also in agreement with insights from previous studies that included partially overlapping populations (Decker *et al.*, 2014). Since subsets of animals from the same breed that were

genotyped on assays clustered closely (Figure 3), this indicates that there were no assay or genotyping effects from differences in ascertainment bias, genotyping accuracy, allele calling or other technical variables between the genotyping platforms (Gurdasani *et al.*, 2015).

The F_{st} measure for genetic differentiation (Willing *et al.*, 2015) and the FLK test, which accounts for unequal population size and the hierarchical structure of relationships among the populations (Bonhomme, 2010), provided little evidence of the existence of breed informative SNPs from these genotyping platforms in indigenous South African breeds. This could be because of lower MAF and high SNP polymorphisms observed between these breeds (Tabangin *et al.*, 2009). It is generally considered that uninformative markers (i.e. low MAF loci) add variability and noise to the results and compromise the power of population genetic studies (Liu *et al.*, 2005). However, effective exploration of other SNP identification methods, such as genome resequencing, could help to identify the most informative markers, and produce an optimal minimum set of markers that could accurately and efficiently differentiate among populations (Ding *et al.*, 2011).

The cluster of South African indigenous breeds and other African breeds sampled from Decker *et al.* (2014) enabled the identification of the regions in the genome that discriminate between the populations. Eleven identified SNPs with higher F_{st} values (>0.6) and higher FLK best described the distinctiveness of the breeds. These SNPs were useful in segregating the South African breeds from the other African breeds. The numbers of SNPs, however, were lower than the 18 informative SNPs found between Japanese Black and Holstein by Nishimura *et al.* (2012) using BovineSNP50 data. This is consistent with the development of the assay in which SNPs with a high MAF across taurine breeds were preferentially selected in the assay design. Consequently, sets of randomly chosen SNP markers in taurine breeds may have sufficient genetic information to produce moderate levels of power to assign individuals to other taurine breeds (Wilkinson *et al.*, 2011). Therefore, the allele frequency distribution within other breeds reveals that the 11 identified SNPs do not possess breed-specific alleles. This suggests that the sequencing of indigenous South African breeds should be considered to identify a large number of informative SNPs specific to discriminating among South African breeds.

Conclusion

The levels of genetic variation for SNPs on the BovineSNP50 and GGP-80K assays identified in this study indicate that these assays have utility for genetic studies in South African populations. The lower average MAF in the indigenous South African breeds reduced the effectiveness of the assays for the selection of breed-informative markers. This may affect their utility in downstream genomic applications. The assays were not adequate for identifying breed informative markers allowing for a small subset of markers to be used to differentiate between the South African indigenous breeds and African breeds. Therefore, identification of SNPs with breed-specific fixation of alternate alleles appears to require the whole genome sequencing of pools of DNA from individuals from the local cattle breeds to avoid the biases inherent to SNP assays. This would help to overcome the challenge of ascertainment bias, and would improve the MAF distribution of variants available for genotyping South African indigenous breeds.

Acknowledgements

The authors acknowledge the financial support from the Red Meat Research and Development of South Africa (RMRDSA). The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged by the first author. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

Authors' Contributions

AAZ designed the experiment, carried out the analysis and drafted the manuscript. AC, MLM assisted with statistical analysis. AM, EVM and JFT structured scientific content. All authors provided editorial suggestions and revisions, read and approved the final draft.

Conflict of Interest Declaration

The authors declare that they have no competing interests.

References

- Allen, H.L., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S. & Ferreira, T., 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nat.* 467, 832-838.
- Bettencourt, E.M.V., Tilman, M., Henriques, P.D.D.S., Narciso, V. & Carvalho, M.L.D.S., 2013. The economic and socio-cultural role of livestock in the wellbeing of rural communities of Timor-Leste. CEFAGE-UE Working Paper 2013/01. pp. 1-18.
- Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S. & SanCristobal, M., 2010. Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genet.* 186, 241-262.

- Bradley, D.G., MacHugh, D.E., Loftus, R.T., Sow, R.S., Hoste, C.H. & Cunningham, E.P., 1994. Zebu-taurine variation in Y chromosomal DNA: A sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Anim. Genet.* 25, 7-12.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J. & Ziaugra, L., 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22, 231-238.
- Chan, E.K.F., Hawken, R. & Reverter, A., 2008. The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. *Anim. Genet.* 40, 149-156.
- Dadi, H., Kim, J.J., Yoon, D. & Kim, K.S., 2011. Evaluation of single nucleotide polymorphisms (SNPs) genotyped by the Illumina Bovine SNP50K in cattle focusing on Hanwoo breed. *Asian Austral. J. Anim. Sci.* 25, 28-32.
- Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Alcalá, A.M., Sonstegard, T.S., Hanotte, O., Gotherstrom, A., Seabury, C.M., Praharani, L. & Babar, M.E., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10 (3), e1004254, 1-14.
- Dimauro, C., Cellesi, M., Steri, R., Gaspa, G., Sorbolini, S., Stella, A. & Macciotta, N.P.P., 2013. Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes. *Anim. Genet.* 44, 377-382.
- Ding, L., Wiener, H., Abebe, T., Altaye, M., Go, R.C., Kerckmar, C., Grabowski, G., Martin, L.J., Hershey, G.K.K., Chakorborty, R. & Baye, T.M., 2011. Comparison of measures of marker informativeness for ancestry and admixture mapping. *BMC Genomics* 12 (622), 1-18.
- Edea, Z., Dadi, H., Kim, S.W., Dessie, T. & Kim, K.S., 2012. Comparison of SNP variation and distribution in indigenous Ethiopian and Korean Cattle (Hanwoo) populations. *Genomics Inform.* 10, 200-205.
- Edea, Z., Bhuiyan, M.S.A., Dessie, T., Rothschild, M.F., Dadi, H. & Kim, K.S., 2015. Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. *Anim.* 9, 218-226.
- Espigolan, R., Baldi, F., Boligon, A.A., Souza, F.R., Gordo, D.G., Tonussi, R.L., Cardoso, D.F., Oliveira, H.N., Tonhati, H., Sargolzaei, M. & Schenkel, F.S., 2013. Study of whole genome linkage disequilibrium in Nelore cattle. *BMC Genomics* 14, 1-8.
- Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M. & Servin, B., 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genet.* 193, 929-941.
- Frese, L., Palme, A. & Kik, C., 2014. On the sustainable use and conservation of plant genetic resources in Europe. Report from Work Package 5, 1-34.
- Gorbach, D.M., Makgahlela, M.L., Reecy, J.M., Kemp, S.J., Baltenweck, I., Ouma, R., Mwai, O., Marshall, K., Murdoch, B., Moore, S. & Rothschild, M.F., 2010. Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. *J. Anim. Breed. Genet.* 127, 348-351.
- Grasso, A.N., Goldberg, V., Navajas, E.A., Iriarte, W., Gimeno, D., Aguilar, I., Medrano, J.F., Rincón, G. & Ciappesoni, G., 2014. Genomic variation and population structure detected by single nucleotide polymorphism arrays in Corriedale, Merino and Creole sheep. *Genet. Mol. Biol.* 72, 389-395.
- Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A. & Ritchie, G.R., 2015. The African genome variation project shapes medical genetics in Africa. *Nat.* 517 (7534), 327-332.
- Gurgul, A., Żukowski, K., Pawlina, K., Ząbek, T., Semik, E. & Bugno-Poniewierska, M., 2013. The evaluation of bovine SNP50 BeadChip assay performance in Polish Red cattle breed. *Folia Biol.* 61, 173-176.
- Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W. & Rege, J.E.O., 2002. African pastoralism: Genetic imprints of origins and migrations. *Science* 296 (5566), 336-339.
- Hanotte, O., Dessie, T. & Kemp, S., 2010. Time to tap Africa's livestock genomes. *Sci. (Washington)* 328, 1640-1641.
- Hayes, B. & Goddard, M., 2010. Genome-wide association and genomic selection in animal breeding. *Genome* 53 (11), 876-883.
- Heaton, M.P., Leymaster, K.A., Kalbfleisch, T.S., Kijas, J.W., Clarke, S.M., McEwan, J., Maddox, J.F., Basnayake, V., Petrik, D.T., Simpson, B. & Smith, T.P., 2014. SNPs for parentage testing and traceability in globally diverse breeds of sheep. *PLoS One.* 9, e94851, 1-10.
- Hulsegge, B., Calus, M.P.L., Windig, J.J., Hoving-Bolink, A.H., Maurice-van Eijndhoven, M.H.T. & Hiemstra, S.J., 2013. Selection of SNP from 50K and 777K arrays to predict breed of origin in cattle. *J. Anim. Sci.* 91, 5128-5134.
- Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Neto, L.R.P., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K. & Paiva, S., 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10 (2), e1001258, 1-14.
- Kuehn, L.A., Keele, J.W., Bennett, G.L., McDanel, T.G., Smith, T.P.L., Snelling, W.M., Sonstegard, T.S. & Thallman, R.M., 2011. Predicting breed composition using breed frequencies of 50,000 markers from the US Meat Animal Research Center 2,000 Bull Project. *J. Anim. Sci.* 89, 1742-1750.
- Kugonza, D.R., Nabasiye, M., Mpairwe, D., Hanotte, O. & Okeyo, A.M., 2011. Productivity and morphology of Ankole cattle in three livestock production systems in Uganda. *Anim. Genet. Res.* 48, 13-22.
- Lee, S.H., Choi, B.H., Lim, D., Gondro, C., Cho, Y.M., Dang, C.G., Sharma, A., Jang, G.W., Lee, K.T., Yoon, D. & Lee, H.K., 2013. Genome-wide association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). *PLoS One.* 8, e74677, 1-9.
- Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P. & Drineas, P., 2011. Tracing cattle breeds with principal components analysis ancestry informative SNP. *PLoS One.* 6 (4), 1-8.
- Lewontin, R.C. & Krakauer, J., 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genet.* 74, 175-195.

- Liu, N., Chen, L., Wang, S., Oh, C. & Zhao, H., 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genet.* 6, S26, 1-5.
- MacEachern, S., Hayes, B., McEwan, J. & Goddard, M., 2009. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in domestic cattle. *BMC Genomics* 10, 1-19.
- Makina, S.O., Muchadeyi, F.C., Van Marle-Köster, E., MacNeil, M.D. & Maiwashe, A., 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front Genet.* 5, 1-7.
- Mapholi, N.O., Marufu, M.C., Maiwashe, A., Banga, C.B., Muchenje, V., MacNeil, M.D., Chimonyo, M. & Dzama, K., 2014. Towards a genomics approach to tick (*Acari: Ixodidae*) control in cattle: A review. *Ticks and Tick-Borne Dis.* 5 (5), 475-483.
- Martinez-Cambor, P., Carleos, C., Baro, J.Á. & Canon, J., 2014. Standard statistical tools for the breed allocation problem. *J. Appl. Statist.* 41, 1848-1856.
- Mason, I.L., 1996. *A World Dictionary of Livestock Breeds, Types and Varieties* (4th ed.), C.A.B. International, Wallingford, Oxfordshire, UK. ISBN 0-85199-102-5.
- 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. & Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One.* 4, e5350, 1-11.
- McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Neto, D.E., Gill, C.A., Gao, C. & Mannen, H., 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genet.* 9, 1-9.
- Michelizzi, V.N., Wu, X., Dodson, M.V., Michal, J.J., Zambrano-Varon, J., McLean, D.J. & Jiang, Z., 2011. A global view of 54,001 single nucleotide polymorphisms (SNPs) on the Illumina BovineSNP50 BeadChip and their transferability to water buffalo. *Int. J. Biol. Sci.* 7, 18-27.
- Mwai, O., Hanotte, O., Kwon, Y.J. & Cho, S., 2015. African indigenous cattle: unique genetic resources in a rapidly changing world. *Asian Australas. J. Anim. Sci.* 28, 911-921.
- Negrini, R., Nicoloso, L., Crepaldi, P., Milanese, E., Colli, L., Chegiani, F., Pariset, L., Dunner, S., Leveziel, H., Williams, J.L. & Ajmone Marsan, P., 2009. Assessing SNP markers for assigning individuals to cattle populations. *Anim. Genet.* 40, 18-26.
- Nishimura, S., Watanabe, T., Ogino, A., Shimizu, K., Morita, M., Sugimoto, Y. & Takasuga, A., 2013. Application of highly differentiated SNPs between Japanese Black and Holstein to a breed assignment test between Japanese Black and F₁ (Japanese Black x Holstein) and Holstein. *Anim. Sci. J.* 84, 1-7.
- Okello, S. & Sabiti, E.N., 2006. Milk production of indigenous Ankole cattle in Uganda as influenced by seasonal variations in temperature, rainfall and feed quality. *Makerere Univ. Res. J.* 1, 73-92.
- Opara, A., Razpet, A. & Logar, B., 2012. Breed assignment test of Slovenian cattle breeds using microsatellites. *Acta. Agric. Slov.* 3, 167-170.
- Otten, D. & Van den Weghe, H.F., 2011. The Sustainability of Intensive Livestock Areas (ILAS): Network system and conflict potential from the perspective of animal farmers. *Int. J. Food Syst. Dyn.* 2, 36-51.
- Pariset, L., Mariotti, M., Nardone, A., Soysal, M.I., Ozkan, E., Williams, J.L., Dunner, S., Leveziel, H., Maroti-Agots, A., Bodo, I. & Valentini, A., 2010. Relationships between Podolic cattle breeds assessed by single nucleotide polymorphisms (SNPs) genotyping. *J. Anim. Breed. Genet.* 127, 481-488.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. & Sham, P.C., 2007. PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81, 559-575.
- Ramos, A.M., Megens, H.J., Crooijmans, R.P.M.A., Schook, L.B. & Groenen, M.A.M., 2011. Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing. *Anim. Genet.* 42 (6), 613-620.
- Rege, J.E.O. & Tawah, C.L., 1999. The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains. *Anim. Genet. Res. Inf.* 26, 1-25.
- Renaudeau, D., Collin, A., Yahav, S., De Basilio, V., Gourdine, J.L. & Collier, R.J., 2012. Adaptation to hot climate and strategies to alleviate heat stress in livestock production. *Anim.* 6, 707-728.
- Reynolds, J., Weir, B.S. & Cockerham, C.C., 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genet.* 105, 767-779.
- Sanz, A., Martin-Burriel, I., Cons, C., Reta, M., Poblador, A., Rodellar, C. & Zaragoza, P., 2014. Genetic diversity, structure and individual assignment of Casta Navarra cattle: a well-differentiated fighting bull population. *J. Anim. Breed. Genet.* 131 (1), 11-18.
- Scholtz, M.M., 1988. Selection possibilities of hardy beef breeds in Africa: The Nguni example. In: 3. *Congres Mondial de Reproduction et Selection des Ovins et Bovins a Viande*, Paris (France), 19-23 Jun 1988. INRA.
- Scholtz, M.M., 2010. *Beef breeding in South Africa* (2nd ed.). Asikhulume pixArt, Rooihuiskraal, Pretoria, South Africa.
- Strydom, P.E., 2008. Do indigenous Southern African cattle breeds have the right genetics for commercial production of quality meat? *Meat Sci.* 80, 86-93.
- Tabangin, M.E., Woo, J.G. & Martin, L.J., 2009. The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proceedings.* 3 (7), S41.
- Van Marle, J., 1974. The breeding of beef cattle in South Africa: Past, present and future. *S. Afr. J. Anim. Sci.* 4, 297-304.

- Van Tassell, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C. & Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Meth.* 5, 247-252.
- Vasconcellos, L.P.D.M.K., Tambasco-Talhari, D., Pereira, A.P., Coutinho, L.L. & Regitano, L.C.D.A., 2003. Genetic characterization of Aberdeen Angus cattle using molecular markers. *Genet. Mol. Biol.* 26, 133-137.
- Yoon, D., Kwon, Y.S., Lee, K.Y., Jung, W.Y., Sasazaki, S., Mannen, H., Jeon, J.T. & Lee, J.H., 2008. Discrimination of Korean cattle (Hanwoo) using DNA markers derived from SNPs in bovine mitochondrial and SRY genes. *Asian Austral. J. Anim. Sci.* 21 (1), 25-28.
- Weir, B.S., 1996. *Genetic Data Analysis II: Methods for discrete population genetic data*. Sinauer Associates Inc, Sunderland, MA.
- Weir, B.S. & Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evol.* 38, 1358-1370.
- Wilkinson, S., Wiener, P., Archibald, A.L., Law, A., Schnabel, R.D., McKay, S.D., Taylor, J.F. & Ogden, R., 2011. Evaluation of approaches for identifying population informative markers from high density SNP chip. *BMC Genet.* 12, 1-14.
- Willing, E.M., Dreyer, C. & Van Oosterhout, C., 2012. Estimates of genetic differentiation measured by F_{st} do not necessarily require large sample sizes when using many SNP markers. *PLoS One.* 7 (8), e42649, 1-7.