Genomic diversity and autozygosity within the SA Drakensberger beef cattle breed

Lashmar, S.F.[1, 2], Visser, C.[1], van Marle-Köster, E.[1] & Muchadeyi, F.C[2]

[1] *Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa*

[2] *Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort, 0110, South Africa*

Corresponding author: Simon Lashmar. Email: LashmarS@arc.agric.za

Highlights

• Inter-chromosomal differences in MAF and LD were observed.

• Inter-chromosomal MAF and LD variation will influence future genomic endeavors.

• ROH, and other coefficients, indicated low inbreeding.

• Conserved ROH might reflect directional selection to maintain breed purity.

• ROH characteristics conveyed more ancient inbreeding.

Abstract

The SA Drakensberger, as a Sanga beef breed, is a composite of *Bos taurus* and *Bos indicus* subspecies. Variation within admixed genomes will influence downstream applications such as imputation and genomic selection (GS). Being an indigenous breed with unique characteristics, such as the black coat, within-breed selection of the SA Drakensberger has focused on maintaining breed purity, which furthermore predisposes the breed to inbreeding. This study aimed to primarily identify possible patterns of variation in population-specific parameters such as minor allele frequency (MAF) and linkage disequilibrium (LD) that might influence the accuracy of future genomic applications. Second, the study investigated possible patterns of genomic uniformity using runs of homozygosity (ROH) as a measure of

inbreeding. Average genome-wide MAF was 0.26 with chromosome-specific MAF ranging from 0.24 (Bos Taurus Autosome; BTA14) to 0.28 (BTA21). The proportion of low-MAF (<5%) SNPs supported average estimates, ranging from 6.6% for BTA23 to 16.0% for BTA14. The r^2 measure of LD was 0.14, 0.17 and 0.22, respectively, when SNPs separated by ≤1Mb, ≤0.1Mb and ≤0.05Mb were considered. LD was generally low, ranging from r^2=0.11 (BTA28) to r^2=0.17 (BTA14) for SNPs separated by ≤1Mb and r^2=0.20 extended only up to <30kb. LD was weaker between SNP pairs including low-MAF SNPs. The ROH identified were predominantly shorter in length, with more than 50% (54.5%) of ROH falling within the <4Mb length interval. Consensus ROH segments were identified and the most prevalent of these occurred on BTA14 and was identified in ~23% of the sampled population. All coefficients of inbreeding indicated low levels of inbreeding, which corresponded to 3% (F_{PED}), 1% (F_{SNP}) and 7% ($F_{ROH>1Mb}$). Correlations of F_{PED} with F_{SNP} and $F_{ROH>1Mb}$ were moderate equating to values of ~0.63 and ~0.64 (P<0.001), respectively. Such moderate correlations could be attributed to the incompleteness of pedigree records. The direct impact of MAF, LD and relatedness on the accuracy of within-breed genetic improvement strategies and its accompanying methodologies, such as imputation, may influence how different chromosomes are treated or accounted for in future genomic endeavors.

Keyword(s): Cattle, genome variation, runs of homozygosity (ROH), single-nucleotide polymorphism (SNP), within-breed diversity

1. Introduction

The genomics era has brought into existence many tools to aid in the genetic characterization and improvement of livestock species. High-throughput technologies such as next-generation sequencing (NGS) and SNP genotyping provide platforms for the identification and utilization of informative SNP markers in various methodologies such as genome-wide association studies (GWAS) and genomic selection (GS). For beef cattle, efforts to localize genomic regions of interest have been focused on economically important traits pertaining to

growth, feed efficiency as well as carcass and meat quality (Sharmaa et al., 2015). Genomic evaluations in the form of genomic breeding values (GEBVs), are furthermore already being implemented, or at least researched, for beef cattle in the Americas, Europe and Australasia (Berry et al., 2016). Genetic studies implementing these genomic techniques are, however, scarce in Africa. This is due in part to the complexity of beef industries in most African countries, such as South Africa (van Marle-Köster et al., 2013), as well as the myriad of indigenous, non-descript beef breeds of uncertain genomic architecture existing across the continent. A main limitation is, however, the inability to compete financially.

The economic status of a country is a main driver of its inclination to adopt genomic technologies (Dekkers & Hospital, 2002). The relative scale of genomics-based studies, with regards to the number of samples and markers required to draw meaningful conclusions, is a major attributor to the absence of studies of this caliber in the developing world. This highlights the need to identify efficient strategies to alleviate the financial burden involved in genotyping many animals for high densities of SNP markers. These strategies can include the establishment of large-scale collaborations, in the form of consortia, or the implementation of statistical methodologies such as genotype imputation. Imputation is primarily a statistical strategy to fulfill incomplete data or generate higher volumes of data by means of predictions that are based on probabilistic methodology (Marchini et al., 2007). This methodology presents an ideal opportunity for application in adapted and economically important local breeds. The accuracy of predictions that can be achieved by this method, however, is influenced by genomic characteristics that are not fully understood for African breeds or breed-groups such as the Sanga sub-species.

Sanga cattle (*Bos taurus africanus*) are indigenous to southern Africa (Schoeman, 1989) and are presumably taurine-indicine hybrids (Grigson, 1991). This was confirmed by initial genome-wide SNP data suggesting that the genome of Sanga cattle has selection signatures of both subspecies (Makina et al., 2016). Population genetic analyses indicate that whereas a breed such as the Afrikaner shows strong divergence from its ancestors, Bonsmara and SA Drakensberger genomes are more complex. The latter breeds have an admixture of European-

and African taurine as well as indicine footprints (Makina et al., 2016). It is common knowledge that the Bonsmara is a composite breed that resulted from experimental crosses (Bonsma, 1980); however, admixture in the SA Drakensberger genome was probably introduced unintentionally (Makina et al., 2016).

The modern SA Drakensberger is a medium-framed beef cattle breed with a sleek, black coat (Rege & Tawah, 1999). One of the most important qualities of this breed is its ability to adapt and perfom consistantly under even low quality grazing conditions (Bisschoff & Lotriet, 2013). The genomic makeup of this breed is largely unknown with regards to proportions of the genome descended from taurine- or indicine ancestors and proportions that are entirely unique. Bolormaa et al. (2011) showed that the origin of chromosomal segments, whether from taurine or indicine descent, has varying effects on traits of economic importance. Genomic heterogeneity will therefore impact on downstream applications for composite breeds when commercial SNP chips are used, as genomic segments with divergent ancestral origins will have to be treated differently to ensure reliability of these applications.

Identifying inter-chromosomal differences in SNPs, whether markers are segregating within a population or whether there are significant correlations between markers, might aid in how individual chromosomes are dealt with in downstream processes. These aspects would be reflected by differences in population-specific parameters such as MAF and inter-SNP LD. SNPs with low MAF, for example, have been shown to negatively influence imputation accuracy, as alleles occurring in low frequencies are not necessarily presented in the haplotypes identified (Schrooten et al., 2014). Ascertainment bias disadvantages indigenous, non-discovery populations and currently available genotyping platforms, such as the Bovine SNP50 bead chip, are expected to include SNPs displaying low MAF in these populations. Cattle populations in Africa furthermore seem to follow a trend of high haplotype diversity and low LD (Mwai et al., 2015). This needs to be taken into consideration given a general consensus that stronger LD improves imputation accuracy (e.g. Pei et al., 2008; Hickey et al., 2012).

Genomic uniformity becomes detectable in the form of long stretches of consecutively homozygous marker genotypes referred to as runs of homozygosity (ROH) (Falconer & Mackay, 1996; Purfield et al., 2012). These regions can be the result of consanguineous mating in the distant or recent past and, alternatively, long-term artificial selection. The number of breeding animals within the SA Drakensberger breed is gradually declining (Abin et al., 2016), with increasing utilization of sires across herds. Based on available pedigree information, the level of inbreeding has steadily increased since the year 2000 albeit lower than the previously suggested 0.5-1% cut-off per generation (Abin et al., 2016). It is therefore expected that their genomes harbour some signatures of inbreeding, which if prevalent will also impact on imputation and other applications. Imputation accuracy will increase with closer within-breed genetic relationships thereby improving the chances of encountering shared haplotypes from common ancestors (Pei et al., 2008). Coupled to this and given the prioritization of conserving local animal genetic resources (AnGR), knowledge of inbreeding can further aid in effectively managing breeding practices to prevent inbreeding depression.

This study aimed to investigate diversity within the SA Drakensberger genome. The main objective was to quantify inter-chromosomal variation using genetic parameters such as MAF and inter-SNP LD. Knowledge of these metrics may help underpin appropriate imputation strategies in the future. The study furthermore investigated chromosomes harbouring uniformly homozygous regions and used ROH to determine the level of genomic inbreeding within the SA Drakensberger breed. Inferences were made on the implications of these breed parameters on genetic diversity and possible downstream genomic strategies.

2. Materials and methods

2.1 Animals sampled

A total of 620 SA Drakensberger cattle were sampled, consisting of 184 bulls and 436 cows. Samples formed part of a cohort that was specifically selected to include half-sib families and family trios for imputation. The sample constituted animals ranging in birth date from 1982 to

2016 and originated from approximately 40 breeders. All breeders, and therefore animals, included are registered with South African Stud Book (SA Stud Book). Ethical clearance was obtained from the University of Pretoria's Faculty of Natural and Agricultural Science (EC151106-024) and written consent was given by the Drakensberger Breeders' Society to perform this study. Pedigree data up to five generations deep is currently ~65% and ~30% complete, respectively, for animals born within- and longer than 25 years ago (Abin, 2014).

2.2 Genotyping and quality control

A cohort of 414 SA Drakensberger samples was genotyped at the Agricultural Research Council's Biotechnology Platform (ARC-BTP) as part of the Beef Genomics Project (BGP) currently underway in SA. Hair and semen samples were received from individual breeders. DNA was isolated and SNP genotypes generated with the GeneSeek® Genomic Profiler (GGP) 150K-bead chip according to the standard infinium protocol. Hair samples for the remaining animals (206 animals) were sent to GeneSeek® (Lincoln, USA) where DNA isolation and genotyping were performed according to standard protocols. Samples and markers were subjected to standard quality control (QC) procedures using PLINK (Purcell et al., 2007). Individuals with high proportions of uncalled genotypes (call rate<90%) were filtered from analysis. Only autosomal, mapped (UMD 3.1 bovine reference genome) markers were considered. Of these markers, SNPs that displayed low call rates (≤95%) and low MAF (<1%) were removed from further analysis. SNPs with duplicated genomic positions were also excluded. After these QC procedures, 606 animals and 120 218 SNPs remained and were used for downstream analyses. To minimize the effect of shared genetic structure on certain analyses, such as principal component analysis (PCA) and ROH estimation, SNPs were additionally filtered based on high LD, removing SNP in LD exceeding r^2=0.5 with another SNP. The LD-pruned (LDP) data set consisted of 76 581 SNPs.

2.3 Genetic relatedness between animals

Relatedness between sampled animals was assessed with PCA as implemented in GCTA software (Genome-wide Complex Trait Analysis; Yang et al., 2011). A genomic relationship matrix (GRM) was constructed based on LDP data and used to estimate eigenvectors and -values.

2.4 Inter-chromosomal variation in SNP parameters

The MAF per SNP was estimated using PLINK software and incorporated all autosomal SNPs with call rates above 95%. No other QC procedures, except for filtering on call rates, were performed before MAF estimation. Summary statistics per chromosome were calculated using R (R Development Core Team, 2013).

The r^2 measure, as proposed by Hill and Robertson (1968), was used to quantify LD and was calculated using the following formula implemented in PLINK software;

$$r^2(p_a, p_b, p_{ab}) = \left(\frac{(p_{ab} - p_a p_b)^2}{p_a(1-p_a)p_b(1-p_b)}\right), \quad [1]$$

where p_{ab} represented the frequency of the haplotype consisting of 2 SNPs; p_a and p_b represented the frequency of allele *a* at the first locus and allele *b* at the second locus, respectively. In PLINK software, no restrictions were set for the minimum r^2 (*--ld-window-r2* 0) and inter-SNP distance (*--ld-window-kb* 99999) allowed for LD estimation and this enabled all possible pairwise comparisons. Mean r^2 values were then calculated for SNP pairs separated by ≤0.05Mb, ≤0.1Mb and ≤1Mb. Averages were also calculated for bins (10kb bins, 0-100kb; 100kb bins, 100kb-1Mb; 1Mb bins, 1-4Mb) to observe LD decay. The effect of MAF on LD was furthermore investigated by estimating the difference in the extent of LD decay when MAF-filtering thresholds (<1%, <5%, <10% and <20%) were adjusted. Post-PLINK calculations were performed using custom scripts.

2.5 Runs of homozygosity

Contiguous homozygous segments were called using PLINK's sliding window approach. This analysis was performed on the full data set of 606 animals after sample filtering procedures, considering independent, LD-filtered SNPs. Segments were called as ROH if: 1) it was a minimum of 1Mb in length; 2) it included no more than one heterozygous SNP, but included up to two missing SNPs; 3) had a minimum SNP density of one SNP every 75kb; and 4) the maximum gap between consecutive SNPs was no longer than 1Mb. The thresholds for these parameters were based on PLINK defaults and consensus with previous ROH studies on cattle in order to allow comparison. The minimum number of consecutive homozygous SNPs that constituted a ROH segment was calculated using the following formula as implemented by Purfield et al. (2012),

$$l = \frac{log_e \frac{\alpha}{n_s . n_i}}{log_e(1 - \overline{het})} \quad [2]$$

where n_s and n_i were the number of SNPs and individuals, respectively, α represented the proportion of false-positive identifications (set to 0.05) and $\overline{het}$ was the average SNP heterozygosity. Using the formula the minimum number of SNPs constituting a ROH was set to 50.

2.6 Inbreeding coefficients

Three methods were utilized to estimate inbreeding: 1) F_{PED} represented a pedigree-derived estimate, 2) F_{SNP} represented a SNP-by-SNP excess in homozygosity and 3) F_{ROH} represented genome-wide ROH coverage. F_{PED} per animal was estimated by SA Stud Book, which the SA Drakensberger is registered with, and forms part of standard breed evaluations (SA Stud Book). All available pedigree information was utilized. F_{SNP}, and heterozygosity, was calculated on LDP data in PLINK. The ROH-based inbreeding coefficient (F_{ROH}) was estimated per individual as follows,

$$F_{ROH} = \frac{S_{ROH}}{L_{GEN}}$$

[3]

where S_{ROH} was the summed length of ROH segments for an individual and L_{GEN} represented the total length of the genome covered by the SNPs on the specific bead chip. Box plots were generated for each inbreeding coefficient using R. Pearson correlations between coefficients were also calculated using R.

3. Results

3.1 Genomic relationships between individuals within the breed

Within breed genomic relationships were estimated from a set of LDP SNPs and the resulting eigenvectors (EVs) did not indicate separation into different clusters, but rather one cluster with dispersion. Correlations between the a) first- and second as well as b) first- and third PCs are illustrated in **Figure 1**.

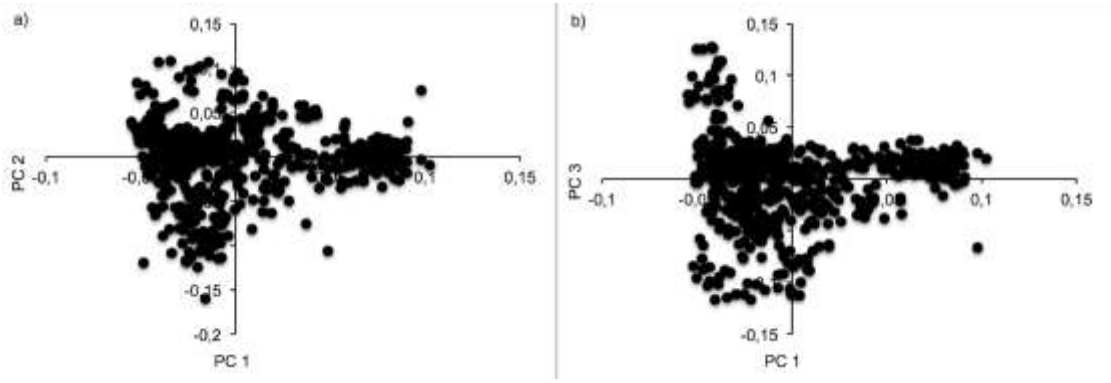

Figure 1 Principal component analysis (PCA) of genetic relatedness between SA Drakensberger animals sampled

PC1 (EV range: -0.055-0.103), PC2 (EV range: -0.160-0.108) and PC3 (EV range: -0.117-0.128) accounted for 46.3%, 28.0% and 25.7% of the variation estimated for the first three principal components. No outliers were identified for EVs estimated for the first PC,

however, 59 and 49 outliers were identified based on EVs estimated for PC2 and PC3, respectively.

3.2 Population-specific SNP parameters

The mean MAF across all autosomes was 0.26±0.14 (median: 0.27) with BTA14 and BTA21 having the lowest (0.24) and highest (0.28) mean MAF, respectively. The highest percentage of SNP displaying low MAF (less than 5%) was observed for BTA14 (16.0%), while BTA23 had the smallest percentage of low-MAF SNPs (6.6%). This was consistent with the percentage monomorphic SNPs observed (BTA14=1.3%; BTA23=0.3%). Across all autosomal markers with sufficient call rates (123 505 SNPs) there were only 0.6%, 2.6% and 9.3% SNPs with MAF of 0%, <1% and <5%, respectively. MAF-related trends are illustrated in **Figure 2**.

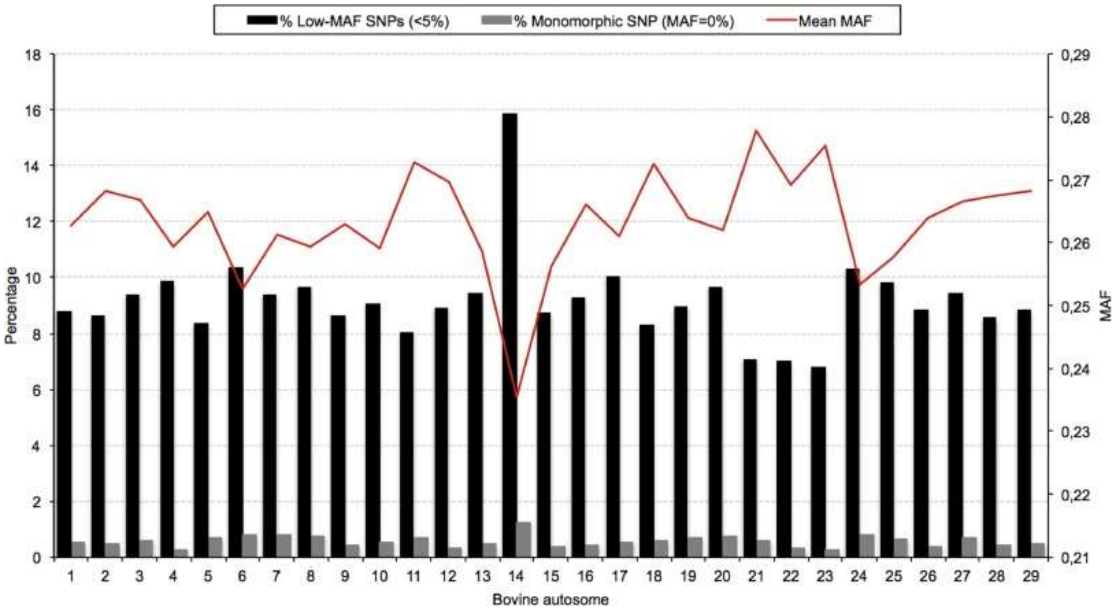

Figure 2 Variation in SNP minor allele frequency (MAF) between autosomal chromosomes

After QC, the mean genome-wide SNP density was 1 SNP/20.9kb and ranged, per chromosome, from 1 SNP/17.9kb (BTA14) to 1 SNP/22.1kb (BTA8). Five autosomes harboured outlying high SNP densities namely BTA6, 7, 14, 20 and 24. Considering SNP pairs separated by ≤1Mb, r^2 ranged from 0.11 (BTA28) to 0.17 (BTA14) with BTA14

identified as displaying outlying high LD. Mean r^2 for shorter inter-SNP distances of up to 100kb and 50kb, respectively, ranged from 0.14 (BTA28) to 0.22 (BTA14) and 0.17 (BTA28) to 0.28 (BTA14). Inter-chromosomal SNP density and LD statistics are depicted in **Figure 3**.

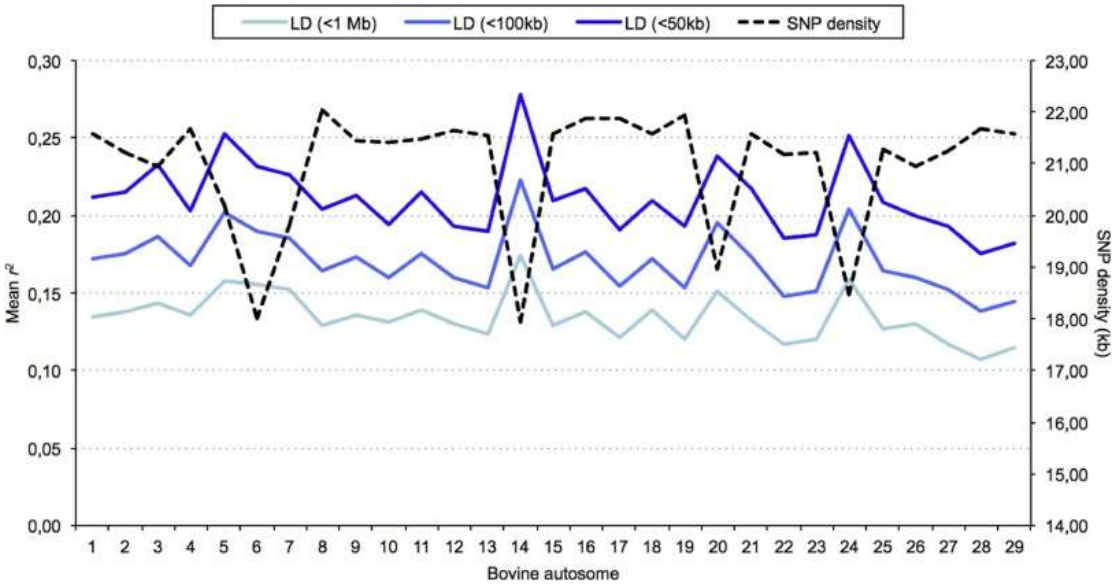

Figure 3 Variation in SNP density- and LD between autosomal chromosomes

On average, the proportion of SNP pairs showing LD of $r^2 \geq 0.2$ increased with 7.1% when only high-MAF SNP (>20%) were included as opposed to lower-MAF SNPs (>1%). Including SNPs with MAF>1%, high LD persisted between SNP pairs <30kb apart. Estimated r^2 of 0.32, 0.24 and 0.21 were observed for SNPs separated by 0-10kb, 10-20kb and 20-30kb, respectively. When only high-MAF (>20%) SNPs were included, LD extended up to approximately double the distance (<60kb). LD decay is illustrated in **Figure 4**.

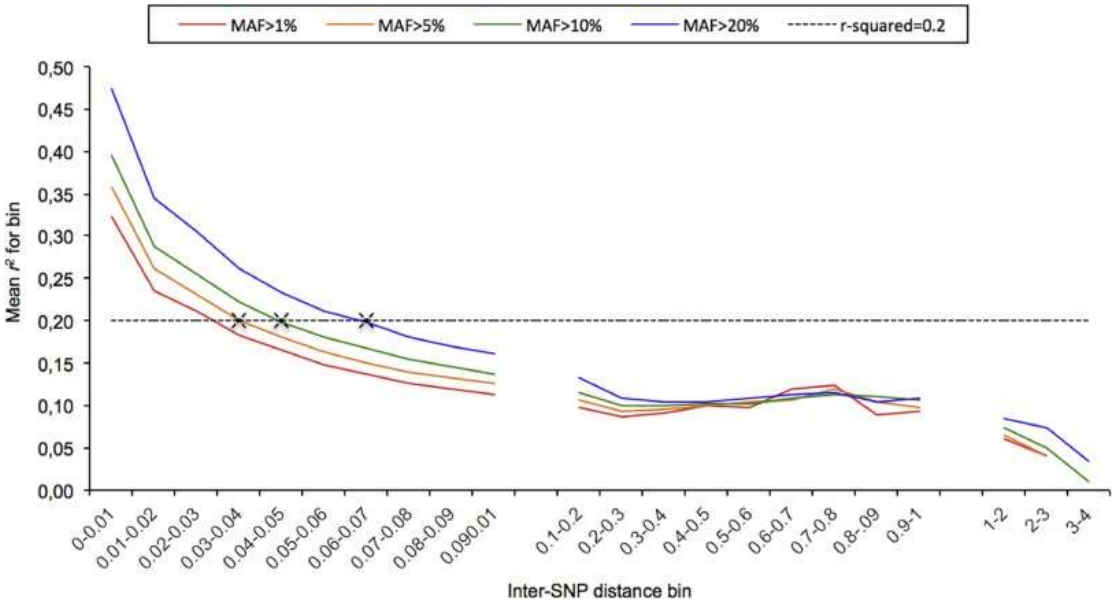

Figure 4 The decay of LD with increasing inter-SNP distances

3.3 ROH analysis

The mean number of ROH (n_{ROH}) per animal was ~33 segments (min=0; max=152). The majority of these segments were between 2-4Mb in length; approximately 18.8%, 35.7%, 25.6%, 14.6% and 5.2% of the segments belonged to the 0-2Mb, 2-4Mb, 4-8Mb, 8-16Mb and >16Mb length categories. Long ROH (>16Mb) were only identified in ~62% of the population and these animals on average had only ~3 of these long segments. On average, the segments identified were composed of 171 SNPs (min=50 SNPs; max=2735 SNPs); the mean distance between homozygous SNPs was 32.78kb (min=11.22kb; max=73kb). Specific clusters of consecutively homozygous SNPs were conserved within varying proportions of the sampled population, and this is illustrated in **Figure 5**.

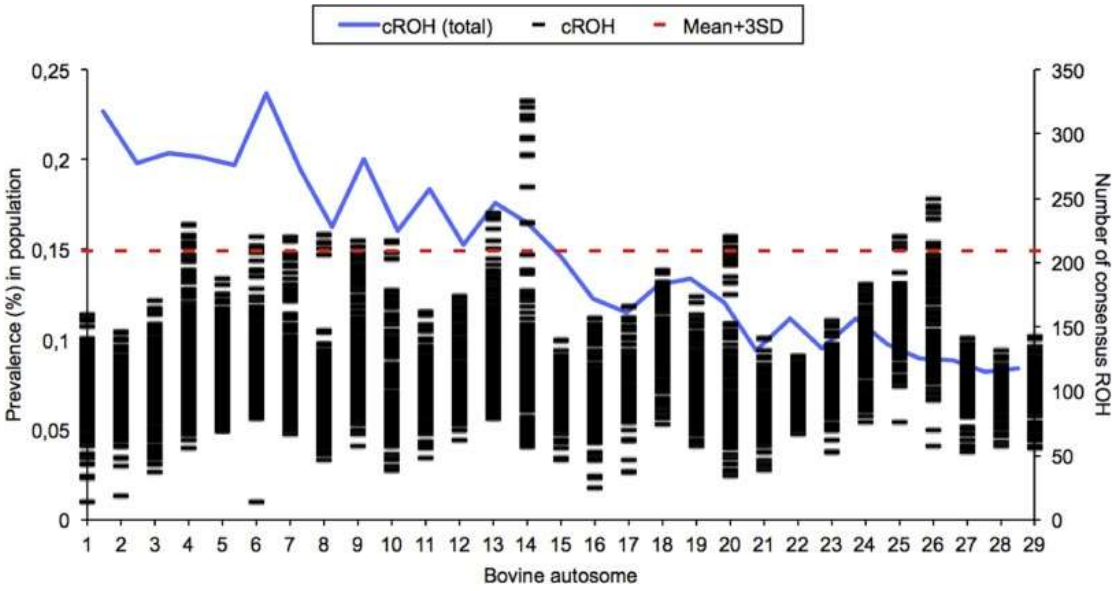

Figure 5 Consensus runs of homozygosity (cROH) within the Drakensberger population

Consensus ROH were on average 86.86kb in length and were composed of ~4 SNPs. BTA6 and BTA28 harbored the highest (331) and lowest (115) number of consensus ROH segments. The largest consensus ROH segment was located on BTA15 and was 1723.83kb in length, consisting of 32 consecutively homozygous SNPs. The most prevalent consensus ROH segment was found in 141 (23.3%) of the sampled animals; this segment was located on BTA14 and constituted 5 SNPs stretching over 225.82kb. Consensus segments occurring in >100 of the sampled animals were also observed on BTA13 and BTA26.

3.4 Inbreeding coefficients

All measures of inbreeding indicated positive inbreeding at a low level. Box plots for each measure are illustrated in **Figure 6**. Only animals with non-zero pedigree-based inbreeding (586 animals) were considered. The mean F_{PED}, F_{SNP} and $F_{ROH>1Mb}$ were calculated as 0.03, 0.01 and 0.07, respectively. Due to the nature of the F_{SNP} measure, the SNP-by-SNP coefficient showed the most variation. F_{SNP} did not change significantly when estimated before- or after LDP and was in agreement with genome-wide heterozygosity estimates that showed a slight loss in genetic diversity (Before LDP: $H_O=0.351<H_E=0.354$; after LDP: $H_O=0.344<H_E=0.347$). Mean F_{ROH} decreased with increasing ROH-length intervals, estimated

as 0.07, 0.06 and 0.05 when only ROH>4Mb, >8Mb and >16Mb were included in calculations.

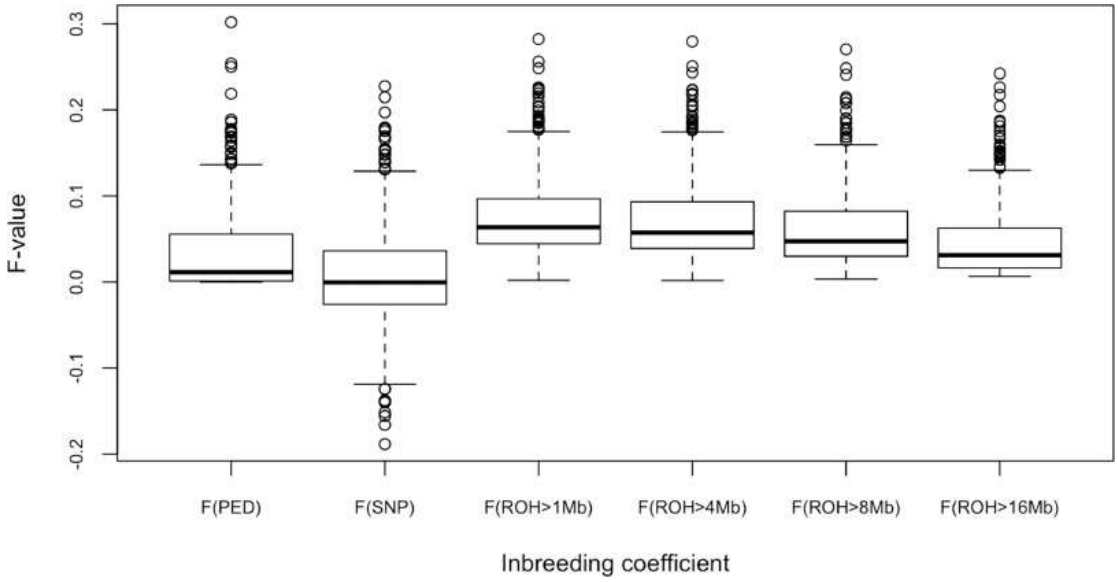

Figure 6 Box plots of pedigree-, SNP- and ROH-based inbreeding coefficients

Pearson correlations between F_{PED} and both F_{SNP} and F_{ROH}, respectively, were moderate. Correlations with F_{ROH} were estimated for different length classes and are indicated in **Table 1**. The highest correlation with F_{PED} was observed when all ROH>8Mb were included in the calculation. Given the fact that ROH were estimated based on SNP data, high correlation between F_{SNP} and F_{ROH} were expected.

Table 1 Correlations between pedigree- and molecular-based inbreeding coefficients

	Correlation
$r(F_{PED})$	
F_{SNP}	0.633***
$F_{ROH>1Mb}$	0.642***
$F_{ROH>4Mb}$	0.639***
$F_{ROH>8Mb}$	0.655***
$F_{ROH>16Mb}$	0.523***
$r(F_{SNP})$	
$F_{ROH>1Mb}$	0.954***

*** *P*-value < 0.001.

4. Discussion

The utility of a specific SNP genotyping platform is influenced by its development and the application of such platforms can adversely impact on population-specific SNP parameters estimated for breeds that were not represented in its design. Breeds with unclear- but presumably diverse ancestry may therefore display variation in these parameters due to the origin of the SNP investigated. The development of the Illumina SNP50 bovine bead chip was based exclusively on the selection of SNPs occurring in taurine beef- and dairy cattle genomes (Matukumalli et al., 2009). Conversely, the GGP 80K bovine bead chip was developed to incorporate more SNP of indicine descent (Edea et al., 2015). Even though no details regarding the exact composition of the 150K chip have been published, it is believed that SNP selection was also biased towards taurine cattle albeit that indicine cattle were also included (personal communication).

SNP-based genetic studies have observed lower MAF estimates for African breeds when only taurine- as opposed to indicine-derived SNPs were studied. Edea et al. (2015), for example, observed significantly lower MAF in Ethiopian cattle using SNP50-genotypes (0.15) as opposed to SNP80-genotypes (0.32). Estimates were similarly low for South African Sanga breeds based on SNP50 genotypes (Nguni: 0.17; Qwabe et al., 2013). Given that indigenous Sanga breeds are believed to be hybrids (Grigson, 1991) that harbour signatures of both sub-species, albeit in unknown proportions (Makina et al., 2015b), average MAF was expected to be higher when indicine-derived SNP were also included. The average MAF obtained in this study (0.26±0.143) is comparable with estimates obtained by Zwane et al. (2016) for the SA Drakensberger breed (0.26 ± 0.145 for MAF≥0%). The latter authors used a set of SNPs that were common between the Illumina SNP50 and GGP 80K chips. Given, firstly, the improvement in SNP density and, secondly, the inclusion of indicine SNPs, ascertainment bias seemed to influence the utility of the SNP150 chip to a lesser extent than the SNP50 chip for a non-discovery, composite breed such as the SA Drakensberger.

Identifying inter-chromosomal variation in MAF is important firstly for its direct impact on downstream analyses and secondly for its influence on local LD. Low-MAF SNPs display

lower imputation accuracy as alleles occurring in low frequencies are not represented in reference haplotypes (Schrooten et al., 2014). These SNPs furthermore negatively impact on the accuracy of GEBVs (Weng et al., 2012). In this study, autosomes with up to ~16% (BTA14) low-MAF SNPs were identified. This will need to be accounted for when imputation, or eventually genomic selection, is implemented especially considering the fact that most imputation methods apply algorithms on a chromosome-by-chromosome basis. For non-discovery breeds it might necessitate the identification of specific markers segregating within these populations by identifying evenly distributed, high-MAF SNPs, and developing breed-specific low-density SNP panels. The relationship between MAF and LD was further investigated and, in agreement with previous research on Sanga- (Makina et al., 2015a) and international breeds (eg. Khatkar et al., 2008; Qanbari et al., 2010), showed low MAF to locally diminish LD. This relationship will be useful for populations, such as the SA Drakensberger, where LD does not persist over long genomic distances (<30kb).

LD is an important population-specific parameter in genomic studies. It serves as a predictor of the density of SNPs required to produce accurate GEBVs and powerful associations in GWAS (Qanbari et al., 2010). The strength of local LD furthermore influences the achievable imputation accuracy of specific genomic regions (Hickey et al., 2012). Many studies have investigated inter-chromosomal differences in LD in cattle populations (Sargolzaei et al., 2008; Bohmanova et al., 2010; Qanbari et al., 2010; Cañas-Álvarez et al., 2014; Edea et al., 2015) and the general consensus has been that there is a chromosome-effect influencing LD. In concordance with previous research, the density of SNPs per autosome seemed to be a primary contributor to high local LD. BTA14, which showed outlying high LD, was the most densely covered autosome. High LD on this autosome might therefore be an artifact of close inter-SNP distance and not necessarily a true reflection of strong relationships between SNPs overall. Developing a breed-specific low-density panel that is optimized for imputation would therefore necessitate the selection of high-LD SNP pairs per autosome, while assuring even distribution of SNPs across autosomes. Investigating inter-chromosomal differences in LD is important as high-LD autosomes are expected to produce higher imputation and GEBV

accuracies. LD for the sampled population was relatively high when considering only SNP pairs separated by ≤50kb (mean r^2=0.22). At short inter-SNP distances (10-20kb), estimates were comparable with composite Brazilian beef breeds (0.24 versus 0.25; Mokry et al., 2014). Furthermore, short distance (~10kb) estimates were on par with what was found for indicine breeds like Brahman (0.25) and Nellore (0.27), but lower than values obtained for taurine breeds (eg. Angus, 0.46; Porto-Neto et al., 2014).

LD of r^2=0.2 was found to persist only for short inter-marker distances (<30kb), albeit higher than previously suggested by Makina et al. (2015a) for the SA Drakensberger breed (10-20kb). This was expected considering the improvement in SNP density of the SNP150- compared to the SNP50 bead chip used by Makina et al. (2015a). In the afore-mentioned study, LD was also estimated in a significantly smaller sample size of SA Drakensberger animals (±40 versus 606). Results were in agreement with Edea et al. (2014) who showed LD decay after 20-40kb in indigenous Ethiopian cattle. Rapid LD decay in the SA Drakensberger was furthermore not surprising, as previous studies have suggested shorter persistence of high LD in admixed populations (Toosi et al., 2010; Mokry et al., 2014). This phenomenon is attributed to more distant common ancestry and therefore the sharing of short haplotype structures within these populations (Mokry et al., 2014). Makina et al. (2015a) observed the smallest average haplo-block length for the SA Drakensberger breed compared to other Sanga breeds.

High LD reflects genomic regions lacking recombination. This absence of recombination is a key factor in the existence of ROH. The length of a ROH segment is arguably its most important characteristic as this can be used to infer population history (McQuillan et al., 2008). Short ROH were significantly more frequent ($ROH_{2\text{-}4Mb}$=35.7%) than longer ROH ($ROH_{>16Mb}$=5.2%) in the SA Drakensberger population sampled. ROH of 16Mb in length has been proposed to represent recent inbreeding of up to ~6 generations ago, whereas ROH of 1Mb corresponds to more ancient inbreeding of up to ~50 generations ago (Ferenčaković, 2015). Given a generation interval (L) of ~6 years for the SA Drakensberger breed (Abin, 2014), this represents inbreeding of up to 36 and 300 years ago, respectively. The ROH

identified in the SA Drakensberger breed therefore implies that inbreeding is predominantly the result of more ancient consanguinity. Hybridization between taurine and indicine cattle is expected to increase diversity within the genomes and may have lead to the interruption of homozygous stretches in African cattle (Purfield et al., 2012). The sharing of homozygous segments between up to 23% of the sampled population may also point towards selection-driven fixation of some segments. It would therefore be beneficial to explore ROH as a trait-association tool, using appropriate phenotypic information, in the future.

Inbreeding has traditionally been quantified by tracing back consanguineous mating through the use of pedigree information. The reliability of this method is, however, dependent on the degree of pedigree recording and hence the completeness of records (McParland et al., 2007). Considering that the birth dates of the animals included in this study ranged from 1982 to 2016 and the fact that not all animals had equal depths of pedigree data available, F_{PED} was not the most robust method of inbreeding estimation. This was confirmed by moderate Pearson correlations (~64%) between F_{PED} and F_{ROH}. The relevance of F_{PED} is further brought into question due to the fact that it does not account for recombination within the genome (McQuillan et al., 2008; Mastrangelo et al., 2016). The insufficiency of in-depth pedigree records therefore deemed F_{ROH} based on larger ROH segments (~66% correlated), which are indicators of more recent inbreeding, a more relevant substitute for F_{PED} than F_{ROH} based on short segments (~64% correlated). Molecular-based measures, such as F_{ROH}, have therefore gained popularity. SNP-based measures such as PLINK's F_{SNP}, however, merely estimates excessive marker homozygosity while the ROH-based measure is representative of the age of inbreeding. Genomic measures of inbreeding were strongly correlated (~95%) and this was in agreement with Mastrangelo et al. (2016) who showed correlations ranging from ~83% to ~95% between these measures.

All coefficients indicated low-level inbreeding (1-7%) within the SA Drakensberger breed. This was expected considering a decline in breeding animals observed within the breed (van der Westhuizen & Groeneveld, 2004; Abin, 2014) and the fact that sires are increasingly being used across breeders. PCA results, however, showed that there is still some dispersion,

indicating weaker relations, between individuals and herds within the population. Currently within-breed selection focuses on a "standard of excellence", placing emphasis on the maintenance of breed purity with regards to coat colour, physique and growth (Drakensberger Cattle Breeders' Society of SA, 2011). The preservation of specific breed characteristics such as the black coat is presumed to have decreased genetic diversity, as was supported by observed- (0.344) versus expected (0.347) heterozygosity, through directional changes in the genotypic frequencies of the loci selected for. Certain historic events in the breed timeline are also believed to have diminished genomic diversity. One such genetic bottleneck was the "Great Trek" in 1834 during which the breed accompanied the migration of Dutch settlers (Drakensberger Cattle Breeders' Society of SA, 2011).

5. Conclusion

The origin and genetic composition of the SA Drakensberger is uncertain. Previous molecular research has identified *Bos taurus* and -*indicus* signatures within the genome of this breed, albeit in unknown proportions. Considering the fact that modern SNP genotyping platforms incorporate SNPs discovered in predominantly *Bos taurus* breeds, it is uncertain how genomic variation within the admixed genome of this breed will influence downstream applications. Inter-chromosomal differences in MAF and LD conformed to expectations, but suggest that these differences need consideration in future genomic endeavors. The relationship between MAF and LD can be exploited in the selection of informative SNP for the possible development of optimized low-density panels. Inbreeding coefficients indicated low levels of inbreeding, which was expected due to artificial selection practices to maintain breed purity (with regards to characteristics such as the all-black coat colour and adaptability). ROH length characteristics furthermore pointed towards more ancient inbreeding, reflecting known historic bottleneck events. The economic importance of the SA Drakensberger as an adaptable indigenous breed in the SA beef industry has sparked interest in genomics-based breed improvement. Results of this study suggest that genomic

applications such as imputation and GS can be further explored if genomic diversity is accounted for. Inferences made on the effect of a heterogeneous genome, such as the SA Drakensberger genome, on downstream applications may apply to other local genetic resources, for example non-descript or composite African breeds, with similarly complex genome architecture.

Statement on conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

Project funding by the Beef Genomics Program (BGP) and the Red Meat Research and Development Trust of South Africa (RMRDT-SA) is hereby acknowledged. The authors would also like to thank the SA Drakensberger Breeders' Society for providing their consent to proceed with this research project. The financial assistance of the National Research Foundation (NRF) towards this research is hereby also acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

References

1. Abin, S.A., 2014. Animal recording as a tool for improved genetic management in African beef cattle breeds. MSc thesis, University of Pretoria.
2. Abin, S.A., Theron, H.E., van Marle-Köster, E., 2016. Short communication: Population structure and genetic trends for indigenous African beef cattle breeds in South Africa. S. Afr. J. Anim. Sci. 46, 152-156.
3. Berry, D.P., Garcia, J.F., Garrick, D.J., 2016. Development and implementation of genomic predictions in beef cattle. Anim. Front. 6, 32-38.

4. Bisschoff, C., Lotriet, R., 2013. The Drakensberger as competitive breed of cattle in the South African beef industry. In: *The 19th International Farm Management Congress*. Warsaw, Poland, 2013. pp1-10.
5. Bohmanova, J., Sargolzaei, M., Schenkel, F.S., 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics 11, 421.
6. Bolormaa, S., Hayes, B.J., Hawken, R.J., Zhang, Y., Reverter, A., Goddard, M.E., 2011. Detection of chromosome segments of zebu and taurine origin and their effect on beef production and growth. J. Anim. Sci. 89, 2050-2060.
7. Bonsma, J.C., 1980. Cross-breeding, breed creation and the genesis of the Bonsmara. Livestock Production - a Global Approach. Tafelberg, Cape Town, 90–110.
8. Cañas-Álvarez, J.J., Mouresan, E.F., Varona, L., Díaz, C., Avilés, C., Baró, J.A., Altarriba, J., Casellas, J., Piedrafita, J., 2014. Linkage disequilibrium and persistence of phase in five spanish local beef cattle breeds. In: *Proceedings of 10th World Congress of Genetics Applied to Livestock Production*. Vancouver, BC, Canada, 2014. Poster presentation, 502.
9. Dekkers, J.C.M., Hospital, F., 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3, 22-32.
10. Drakensberger Cattle Breeders' Society of SA, 2011. Drakensberger Handbook, first ed. Volksrust, Mpumalanga, South Africa.
11. Edea, Z., Dadi, H., Kim, S.W., Park, J.H., Shin, G.H., Dessie, T., Kim, K.S., 2014. Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in Ethiopia. J. Anim. Breed. Genet. 131, 358-366.
12. Edea, Z., Dadi, H., Dessie, T., Lee, S-H., Kim, K-S., 2015. Genome-wide linkage disequilibrium analysis of indigenous cattle breeds of Ethiopia and Korea using different SNP genotyping BeadChips. Genes Genom. 37, 759-765.
13. Falconer, D.S., Mackay, T.F.C., 1996. Introduction to quantitative genetics, fourth ed. Longman, New York.

14. Ferenčaković, M., 2015. Molecular dissection of inbreeding depression for semen quality traits in cattle. Doctoral thesis, University of Zagreb.
15. Grigson, C., 1991. An African origin for African cattle? – some archaeological evidence. Afr. Archaeol. Rev. 9, 119-144.
16. Hickey, J.M., Crossa, J., Babu, R., De los Campos, G., 2012. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. Crop Sci. 52, 654-663.
17. Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226-231.
18. Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A.L., Barris, W., Schnabel, R.D., Taylor, J.F., Raadsma, H.W., 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics 9, 187.
19. Makina, S.O., Taylor, J.F., van Marle-Köster, E., Muchadeyi, F.C., Makgahlela, M.L., MacNeil, M.D., Maiwashe, A., 2015a. Extent of linkage disequilibrium and effective population size in four South African Sanga cattle breeds. Front. Genet. 6, 1-12.
20. Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E., Taylor, J.F., Makgahlela, M.L., Maiwashe, A., 2015b. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genet. Select. Evol. 47, 92.
21. Makina, S.O., Whitacre, L.K., Decker, J.E., Taylor, J.F., MacNeil, M.D., Scholtz, M.M., van Marle- Köster, E., Muchadeyi, F.C., Makgahlela, M.L., Maiwashe, A., 2016. Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide. Genet. Select. Evol. 48, 88.
22. Marchini, J., Howie, B., Myers, S., McVean, G., Donnelly, P., 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906-913.

23. Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M.T., Portolano, B., 2016. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. Anim. 10, 746-754.
24. 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.L., Sonstegard, T.S., Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS ONE 4, e5350.
25. McParland, S., Kearney, J.F., Rath, M., Berry, D.P., 2007. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. J. Anim. Sci. 85, 322-331.
26. McQuillan, R., Leutenegger, A-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H., Wilson, J.F., 2008. Runs of homozygosity in European populations. Am. J. Hum. Genet. 83, 359-372.
27. Mokry, F.B., Buzanskas, M.E., 2, Mudadu, M.D., Grossi, D.D., Higa, R.H., Ventura, R.V., de Lima, A.O., Sargolzaei, M., Meirelles, S.L.C., Schenkel, F.S., da Silva, M.V.G.B., Niciura, S.C.M., de Alencar, M.M., Munari, D.P., Regitano, L.C.dA., 2014. Linkage disequilibrium and haplotype block structure in a composite beef cattle breed. BMC Genomics 15, S6.
28. Mwai, O., Hanotte, O., Kwon, Y-J., Cho, S., 2015. African indigenous cattle: unique genetic resources in a rapidly changing world. Asian Austral. J. Anim. Sci. 28, 911-921.
29. Pei, Y-F., Li, J., Zhang, L., Papasian, C.J., Deng, H-W., 2008. Analyses and comparison of accuracy of different genotype imputation methods. PloS ONE 3, e3551.
30. Porto-Neto, L.R., Kijas, J.W., Reverter, A., 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genet. Select. Evol. 46, 22.

31. Purcell, S., Neale, B., Todd-Brown, K., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559-575.
32. Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13, 70.
33. Qanbari, S., Pimentel, E.C., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R., Simianer, H., 2010. The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet. 41, 346-356.
34. Qwabe, S.O., van Marle-Köster, E., Maiwashe, A., Muchadeyi, F.C., 2013. Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. S. Afr. J. Anim. Sci. 43, 64-67.
35. 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. Resour. Info. 26, 1-26.
36. Sargolzaei, M., Schenkel, F.S., Jansen, G.B., Schaeffer, L.R., 2008. Extent of linkage disequilibrium in Holstein cattle in North America. J. Dairy Sci. 91, 2106-2117.
37. Schoeman, S.J., 1989. Recent research into the production potential of indigenous cattle with special reference to the Sanga. S. Afr. J. Anim. Sci. 19, 55-61.
38. Schrooten, C., Dassonneville, R., Ducrocq, V., Brøndum, R.F., Lund, M.S., Chen, J., Liu, Z., González-Recio, O., Pena, J., Druet, T., 2014. Error rate for imputation from the Illumina BovineSNP50 chip to the Illumina BovineHD chip. Genet. Select. Evol. 46, 10.
39. Sharmaa, A., Leea, J.S., Dang, C.G., Sudrajad, P., Kim, H.C., Yeon, S.H., Kang, H.S., Lee, S.-H., 2015. Stories and challenges of genome wide association studies in livestock - a review. Asian Austral. J. Anim. Sci. 28, 1371-1379.
40. Toosi, A., Fernando, R.L., Dekkers, J.C.M., 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88, 32-46.

41. Van Marle-Köster, E., Visser, C., Berry, D.P., 2013. A review of genomic selection - Implications for the South African beef and dairy cattle industries. S. Afr. J. Anim. Sci. 43, 1-17.
42. Van der Westhuizen, R.R., Groeneveld, E., 2004. Population and pedigree analysis of indigenous South African beef breeds, in: Van Der Honing, Y., (Ed.), Book of abstracts of the 55th Annual Meeting of the Europian Association for Animal Production. Bled, Slovenia, pp. 37.
43. Weng, Z., Zhang, Z., Ding, X., Fu, W., Ma, P., Wang, C., Zhang, Q., 2012. Application of imputation methods to genomic selection in Chinese Holstein cattle. J. Anim. Sci. Biotech. 3, 6.
44. Yang, J., Hong Lee, S., Goddard, M.E., Visscher, P.M., 2011. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am. J. Hum. Genet. 88, 76-82.
45. Zwane, A.A., Maiwashe, A., Makgahlela, M.L., Choudhury, A., Taylor, J.F., van Marle-Köster, E., 2016. Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds. S. Afr. J. Anim. Sci. 46, 302-312.