frontiers in Genetics

ORIGINAL RESEARCH
published: 29 March 2019
doi: 10.3389/fgene.2019.00273

Genome-Wide SNP Discovery in Indigenous Cattle Breeds of South Africa

Avhashoni A. Zwane[1,2], Robert D. Schnabel[3,4], Jesse Hoff[3], Ananyo Choudhury[5], Mahlako Linah Makgahlela[1,6], Azwihangwisi Maiwashe[1,6], Este Van Marle-Koster[2] and Jeremy F. Taylor[3]*

[1] Department of Animal Breeding and Genetics, Agricultural Research Council-Animal Production, Irene, South Africa, [2] Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, South Africa, [3] Division of Animal Sciences, University of Missouri, Columbia, MO, United States, [4] Informatics Institute, University of Missouri, Columbia, MO, United States, [5] Sydney Brenner Institute of Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa, [6] Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa

Single nucleotide polymorphism arrays have created new possibilities for performing genome-wide studies to detect genomic regions harboring sequence variants that affect complex traits. However, the majority of validated SNPs for which allele frequencies have been estimated are limited primarily to European breeds. The objective of this study was to perform SNP discovery in three South African indigenous breeds (Afrikaner, Drakensberger, and Nguni) using whole genome sequencing. DNA was extracted from blood and hair samples, quantified and prepared at 50 ng/μl concentration for sequencing at the Agricultural Research Council Biotechnology Platform using an Illumina HiSeq 2500. The fastq files were used to call the variants using the Genome Analysis Tool Kit. A total of 1,678,360 were identified as novel using Run 6 of 1000 Bull Genomes Project. Annotation of the identified variants classified them into functional categories. Within the coding regions, about 30% of the SNPs were non-synonymous substitutions that encode for alternate amino acids. The study of distribution of SNP across the genome identified regions showing notable differences in the densities of SNPs among the breeds and highlighted many regions of functional significance. Gene ontology terms identified genes such as *MLANA*, *SYT10,* and *CDC42EP5* that have been associated with coat color in mouse, and *ADAMS3, DNAJC3,* and *PAG5* genes have been associated with fertility in cattle. Further analysis of the variants detected 688 candidate selective sweeps (ZH_p Z-scores ≤ -4) across all three breeds, of which 223 regions were assigned as being putative selective sweeps (ZH_p scores ≤ -5). We also identified 96 regions with extremely low ZH_p Z-scores (≤ -6) in Afrikaner and Nguni. Genes such as *KIT* and *MITF* that have been associated with skin pigmentation in cattle and *CACNA1C*, which has been associated with biopolar disorder in human, were identified in these regions. This study provides the first analysis of sequence data to discover SNPs in indigenous South African cattle breeds. The information will play an important role in our efforts to understand the genetic history of our cattle and in designing appropriate breed improvement programmes.

Keywords: indigenous breeds, sequencing, novel variants, annotation, genes

OPEN ACCESS

Edited by:
Ino Curik,
University of Zagreb, Croatia

Reviewed by:
Fabyano Fonseca Silva,
Universidade Federal de Viçosa, Brazil
Eveline M. Ibeagha-Awemu,
Agriculture and Agri-Food Canada (AAFC), Canada

****Correspondence:***
Avhashoni A. Zwane
zwanea@arc.agric.za

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *17 August 2018*
Accepted: *12 March 2019*
Published: *29 March 2019*

Citation:
Zwane AA, Schnabel RD, Hoff J, Choudhury A, Makgahlela ML, Maiwashe A, Van Marle-Koster E and Taylor JF (2019) Genome-Wide SNP Discovery in Indigenous Cattle Breeds of South Africa. Front. Genet. 10:273. doi: 10.3389/fgene.2019.00273

INTRODUCTION

The development of next generation sequencing (NGS) technologies has enabled rapid and cost-effective generation of sequence data for SNP discovery in cattle (Le Roex et al., 2012; Mullen et al., 2012). These developments have also enabled the simultaneous estimation of SNP allele frequencies in a diverse range of reference populations (Van Tassell et al., 2008). Low and high-density SNP genotyping assays are available for performing genome-wide analyses in cattle (Matukumalli et al., 2009). However, while the available assays have been shown to be adequate for studies in European taurine breeds, they are less informative when applied to indicine or indigenous South African (SA) breeds (Gurgul et al., 2013; Zwane et al., 2016). Studies using the BovineSNP50 assay in indigenous SA breeds have shown substantially lower levels of linkage disequilibrium (LD) and lower minor allele frequencies (MAF) compared to those obtained in European taurine breeds (Edea et al., 2013; Makina et al., 2014). Furthermore, a study by Makina et al. (2015) using the BovineSNP50 assay for the detection of signatures of selection in indigenous SA breeds, also indicated reduced numbers of informative markers. Further analysis of these markers showed little evidence for the existence of breed-specific markers in indigenous SA cattle breeds (Zwane et al., 2016). Consequently, there is a limited utility for the implementation of these assays for genome-wide association studies (GWAS), quantitative trait locus (QTL) detection or for the identification of genes associated with economically important traits in indigenous SA breeds as observed by Albrechtsen et al. (2010).

South African indigenous cattle such as Afrikaner (AFR), Drakensberger (DRA), Nguni (NGI), Bonsmara, and Tuli have played a major role in traditional, social, and commercial history of the country (Scholtz, 2010). These breeds provide valuable farm animal genetic resources for beef production in combination with exotic beef breeds that were introduced many decades ago (Scholtz, 2010). The establishment of the SA indigenous cattle breeds and those from Africa was closely associated with human development and migration (Strydom, 2008). Africa's indigenous cattle incorporate various crosses between Hamitic longhorn cattle (*Bos taurus*), zebu cattle (*Bos indicus*), and shorthorn cattle (Strydom, 2008). As the movement of man proceeded southward through Africa, new cattle breeds (zebu and sanga-types) were developed (Strydom, 2008). The zebu-type cattle include the Boran, Masai, and Sokoto breeds, while sanga-type (also known as *Bos taurus africanus*) includes the Afrikaner, Nguni, Pedi, Mashona, and Tuli (Hanotte et al., 2000; Strydom, 2008). Sanga cattle were introduced to SA during the migration of the San and Sudanic Bantu tribes to southern Africa and the arrival of Europeans during the fifteenth-century (Bachmann, 1983). Their substantial body mass and greater production in tsetse-free areas have made these breeds more appealing to the local farmers, which somewhat explains the abundance of these breeds and wide distribution throughout Africa (Mwai et al., 2015).

Indigenous breed of SA have made major contributions to livestock production because of their ability to adapt and produce in different SA production systems (Abin et al., 2016). These breeds have been participating in animal recording programmes with an average complete pedigree recording varying from 88.5% for the Nguni to 92.5% for the Afrikaner (Abin et al., 2016). The availability of the pedigree records have been essential for genetic evaluation using BLUP model in determining the selection efficiency and actual genetic change (Mostert, 2007; Groeneveld et al., 2009). However, crossbreeding and inbreeding within cattle breeds has been reported to have negative effects on production and fitness traits (Nazokkarmaher, 2016), and have contributed to loss of diversity in most cattle populations (Pienaar et al., 2014). The use of a small number of selected genotypes increases the chance of having undesirable recessive genes within a population, which may result in inbreeding depression in the near future (Abin et al., 2016). Quantitative breeding methods such as artificial insemination has resulted in more intense selection pressure on a number of traits of economic importance, which could have contributed to an increase in production efficiency. Therefore, maintaining within-breed genetic diversity is essential for selection (Oltenacu and Broom, 2010). Currently, relatively little information is available on SA cattle breeds at the genome level, including sequence variation. Therefore, sequencing the genomes of indigenous SA cattle could be beneficial in animal production, in understanding the genetic architecture of traits of economic importance, animal health and welfare, and in understanding the genetic basis of diseases, as well as genomic selection-based breeding program. Genome sequencing also presents opportunities for increased knowledge of the evolutionary histories of these breeds (Pool and Waddell, 2002).

NGS technologies have identified a large number of SNPs and insertions-deletions (Indels), with many variants remaining to be detected, especially in cattle breeds that are phylogenetically distinct from the more extensively studied European breeds (Choi et al., 2013). More than 60,000 putative SNPs were identified from the sequencing of reduced representation DNA libraries generated for 66 cattle from three populations (Van Tassell et al., 2008). More than 2 million novel SNPs were discovered from resequencing of a Fleckvieh bull (Eck et al., 2009). Furthermore, Kawahara-Miki et al. (2011) re-sequenced the genome of a single Kuchinoshima-Ushi (Japanese native cattle) bull and identified 6.3 million SNPs, of which more than 5.5 million (87%) were novel. Choi et al. (2014) reported a total of 10.4 million SNPs identified in Korean Hanwoo, Jeju Heugu and Holstein cattle, and found 54.12% novel SNPs and also detected 1,063,267 Indels in these genomes. This indicates that NGS technologies are effective for SNP discovery projects and can also be applied to variant discovery in indigenous SA cattle.

The use of sequence data for variant discovery and genotyping has the advantage of reduced SNP ascertainment bias compared to the use of commercially available SNP assays (Nielsen et al., 2011). SNP ascertainment bias influences the extent to which polymorphisms are shared across populations

due to the distribution of allele frequencies within studied populations that may result in biases in measures of genetic differentiation, e.g., F_{st} estimates between populations affects the weighting of principal components, which in turn, can affect inferences about admixture in populations (McTavish and Hillis, 2015). Consequently, the sequencing of indigenous SA cattle genomes presents the potential to discover new SNPs for inclusion in existing SNP assays or for developing custom-made SNP chips for local SA populations. This information can also improve the accuracy of inferences made in population studies and the genome-wide detection of genes associated with complex traits such as disease resistance (Pool et al., 2010). It also holds potential for the identification of breed informative SNPs for breed assignment in SA populations (Ramos et al., 2011).

Fewer studies were done to understand the genetic variation of indigenous SA breeds at the genome level. Studies have focussed on the use of microsatellite markers and available bovine SNP assays to determine the extent of genetic diversity among Nguni, Bonsmara, Drakensberger, and Afrikaner cattle (Makina et al., 2014; Pienaar, 2014; Sanarana et al., 2016). Other studies include the identification of genes for tick resistance and copy number variation (CNV) in Nguni, as well as determining the extent of LD in SA cattle as compared to the European taurine breeds (Makina et al., 2015; Mapholi, 2015; Wang et al., 2015). These studies have provided a basis for understanding the genetic diversity and variation among these cattle breeds. To date, limited sequence data have been generated for indigenous SA cattle breeds. Breeds such as Brahman, Afrikaner and Tuli (African indicine), representing Australian populations, have been sequenced and analyzed resulting in 3.56 million new SNPs being submitted to dbSNP (Barris et al., 2012). The objective of this study was to search for novel SNPs the three indigenous SA cattle breeds [i.e., Afrikaner (AFR), Drakensberger (DRA), and Nguni (NGI)] using whole genome sequencing, and use identified SNPs perform functional enrichment analysis.

MATERIALS AND METHODS

Pedigree Analysis and Sample Identification

The available pedigree data for each breed were obtained from the Agricultural Research Council (ARC) Integrated Registration and Genetic Information System (INTERGIS) database. Pedigree analysis of Afrikaner (n = 251,964), Drakensberger (n = 198,237), and Nguni (n = 241,491) were performed within breed to identify the least related individuals in these populations. Relationship coefficients between individuals were estimated using the method of Meuwissen and Luo (1992) implemented in the PEDIG software (Boichard, 2002), where males born between 2006 and 2012 were considered to be the reference population. In total, 90 least related animals across breeds (i.e., 30 animals per breed) with average relationship coefficients of 0.006, 0.008, and 0.0008 for Afrikaner, Drakensberger, and Nguni, respectively were selected across all nine SA provinces for sequencing to span the cattle's genetic diversity, and breeder's consent was obtained from the animal owners.

Sample Collection, Library Construction, and DNA Sequencing

Sampling of blood and hair was performed with the approval of the Animal Ethics Committee of the University of Pretoria (EC: S4285-15), according to guidelines for the proper handling of animals during sample collection. Genomic DNA was extracted from whole blood (200 μl/sample) using the Roche DNA extraction Kit (Roche, Germany) following the standard protocol of the manufacturer. The procedure included a proteinase K digestion followed by column purification for the extraction of high quality DNA. The extraction of DNA from hair roots was performed using an optimized Phenol-Chloroform protocol (Sambrook and Russell, 2006), that included a Proteinase K and Dithiothreitol digestion followed by phenol-chloroform extraction and centrifugal dialysis with Centricon concentrators (Slikas et al., 2000). The quality of the extracted DNA samples was assessed using a NanoDrop UV/Vis Spectrophotometer (NanoDrop ND-1000) and verified using a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). All DNA samples were maintained at a concentration of 50 ng/μl in preparation for NGS sequencing at the ARC-Biotechnology Platform.

Equimolar DNA pools were prepared for each breed using 170 ng of DNA per animal, and each DNA pool contained 30 animals per breed. Genomic libraries were prepared with the Truseq DNA sample preparation kit v2 (Illumina, San Diego, CA, United States) using 1 μg of genomic DNA according to the manufacturer's instructions. DNA was fragmented using a Covaris E220 sonicator (350 bp), end-repaired and A-tailed followed by the ligation of adapters (TruSeq, Illumina) and 12 cycles of polymerase chain reaction (PCR) were performed. Quantities and the quality of usable material for each of the libraries were estimated by qPCR (KAPA Library Quantification Kit–Illumina Genome Analyzer-SYBR Fast Universal). The automated cBot Cluster Generation System (Illumina, San Diego, CA, United States) was used to generate clusters on the flow cell. Each DNA pool was then sequenced (paired-end; read length 125 bp) on a single lane of the Illumina HiSeq 2500. The resulting images were analyzed with the Bcl2fastq v2.0 (Illumina) to generate the raw fastq files (Ramos et al., 2009; Van et al., 2013; Boutet et al., 2016).

Sequence reads were filtered for base quality and adapter trimming using Trimmomatic v0.33 (Bolger et al., 2014). Reads were trimmed if four consecutive bases had an average Phred-like quality score of less than 20. After trimming, only pairs of DNA sequences for which each read exceeded 35 bp were retained for analysis. Sequence reads were aligned to the UMD3.1 reference genome using the Burrows-Wheeler aligner (BWA-MEM) v0.7, a software package for mapping lowly-divergent sequences against a large reference genome (Li and Durbin, 2009). The alignments were coordinate sorted and converted to the BAM format using SAMtools v1.2

(Ramirez-Gonzalez et al., 2012). Data were then formatted for variant calling using Picard v1.135, by marking duplicate reads (Li et al., 2009).

Variant Discovery and SNP Annotation

Variant discovery was performed within breed according to GATK Best Practices using the genomic variant call format (GVCF) workflow, using HaplotypeCaller (Van der Auwera et al., 2013). The workflow includes data pre-processing steps and calling variants separately for each population using a command that is specific for paired-end data. The pre-processing steps include realigner target creator to generate intervals for each chromosome for Indel realignment, depth of coverage estimation for each chromosome, base recalibration (using dbSNP build 143 as known variants), analyzing covariates/variables and printing reads. Genotype calling was performed separately for each chromosome to generate GVCF files for variant calling, using –sample ploidy of 60 for pooled samples. The workflow included a joint analysis step that empowers variant discovery by providing the ability to leverage population-wide information from a cohort of samples (in this case three populations), allowing the detection of variants with greater sensitivity and genotyping samples as accurately as possible (GATK Best Practices; Bareke et al., 2013). Variants were generated in VCF files, and the genotypes were called for each breed with a minimum genotype quality of 20 (Aslam et al., 2012).

To reduce the false discovery rate, hard filtering steps were conducted using the following criteria: Phred scaled polymorphism probability (QUAL) < 30.0, variant confidence normalized by depth (QD) < 2.0, mapping quality (MQ) < 40.0, strand bias (FS) > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, and ReadPosRank-Sum < -8.0 (GATK Best Practices; Choi et al., 2015). All SNPs that passed these criteria were consequently categorized into fixed (homozygous non-reference assembly nucleotide genotypes within the breed) or segregating (variable/heterozygous genotypes identified in the breed) (Aslam et al., 2012). The transition-to-transversion (Ti/Tv) ratio for each SNP call was calculated for each population as an indicator of potential sequencing errors (Choi et al., 2015) using VCFtools (Danecek et al., 2011). This is the ratio of the number of transitions (interchanges of either purines, A $<->$ G or pyrimidines, C $<->$ T) to the number of transversions (interchanges of purine for pyrimidine bases), for a pair of DNA sequences (Mitchell, 2015).

SNP annotation and the functional consequences of sequence variants were predicted using the Variant Effect Predictor (VEP) v2.0 tool, Ensembl Build 87 (Huang et al., 2009; McLaren et al., 2010). For all input variants, VEP provides detailed annotations for transcripts, proteins, and regulatory regions, and also provides phenotype information for known variants (McLaren et al., 2016). The functional effects of each SNP were estimated, and all SNPs were assigned with a diverse range of functional categories based on genomic coordinates, functional class, codon change, gene name, transcript biotype, gene coding, transcript ID, exon rank, and corresponding genotype (Choi et al., 2015). Annotation results were downloaded for further downstream analysis. The identified variants were verified by using data from European taurine or indicine breeds that were available from Run 6 of the 1000 Bull Genomes Project, consisting of more than 2,700 cattle and 86,474,165 million variants (Frischknecht et al., 2017) to identify novel SNPs present within the SA breeds.

Assessment of SNP Density

SNPs were examined to determine their distribution throughout the genome, identify regions enriched for novel and non-synonymous SNPs and identify genes associated with enriched regions. To identify genomic regions of exceptional SNP densities, we compared the distribution of four different categories of SNPs [all SNPs, missense SNPs, LoF SNPs (stop gain and stop loss) and novel SNPs] in 1 Mb non-overlapping windows across the genome in each breed. The choice of window was based on a recent publication by Das et al. (2015). An in-house script was used to compute the SNP densities and R-script was used to visualize the distribution (Turner, 2014). The top 1% windows for each breed and category were annotated with the Ensembl Cow database, Release 87[1] and when one or more genes were found in a window, the corresponding windows were ascribed the gene names.

Identification of Selective Sweeps

Identification of selective sweeps was performed using the approach of Rubin et al. (2012) that makes provision for the identification of variants from pooled whole genome sequence data. This method determines, for each pool and SNP, the numbers of reads corresponding to the most (n_{MAJ}) and least abundant alleles (n_{MIN}) and for each window in each breed pool, a pooled heterozygosity score is calculated as:

$$\mathrm{H_p} = 2\Sigma n_{\mathrm{MAJ}}\Sigma n_{\mathrm{MIN}}/(\Sigma n_{\mathrm{MAJ}} + \Sigma n_{\mathrm{MIN}})^2$$

where Σn_{MAJ} and Σn_{MIN} are the sums of n_{MAJ} and n_{MIN} for all SNPs in the window. Individual $\mathrm{H_p}$ values are then Z-transformed as follows:

$$\mathrm{ZH_p} = (H_\mathrm{p} - \mu H_\mathrm{p})/\sigma H_\mathrm{p}$$

where μH_p and σH_p are the mean and standard deviation for the $\mathrm{H_p}$ scores. To detect putative selective sweeps, a whole genome screen was performed to identify genomic regions with an excess of homozygosity (heterozygote deficiency) from the autosomes. SNPs were used to calculate Z-transformations of the pooled heterozygosity ($\mathrm{ZH_p}$) in each of the three breeds, and the number of sequence reads containing major and minor alleles were also counted. Subsequently, a 50% overlapping sliding window approach with 150 kb windows was used to compute $\mathrm{ZH_p}$ in each of the windows, and plot the distribution of SNP counts within these windows. Windows with $\mathrm{ZH_p}$ Z-scores of ≤ -4 were retained as candidate selective sweep regions and regions with ZHp Z-scores of ≤ -5 as putative selective sweeps. In addition, animal QTLdb was used to retrieve quantitative trait loci (QTL)

[1] www.ensembl.org/

information and visualize the QTL located within the putative selective sweep regions (Hu et al., 2013).

RESULTS

Sequencing and Mapping

Sequencing of AFR, DRA, and NGI generated approximately 1.8 billion (184 Gb) of high quality paired-end reads using an Illumina HiSeq 2500 sequencer, of which 99% of the reads were mapped to the bovine reference genome (UMD 3.1). PCR duplicates were marked and reads were realigned around insertion and deletion events resulting in approximately 1.7 billion sequence reads (90.2%) across the three breeds, with an average coverage of 21.1-fold across the reference genome (**Table 1**). The Ti/Tv ratio and heterozygous/homozygous variant ratios are computed in genetic studies as a quality control measure for sequence data. To evaluate the quality of the detected SNPs, the Ti/Tv ratio was computed and found to be similar for each breed (AFR:2.20, DRA:2.23, NGI:2.22).

Variant Detection

A total of 17.6 million variants were identified in the three studied breeds with the greatest number of variants in NGI and AFR and lowest in DRA (**Table 2**). The detected variants comprised 89% SNPs and 11% Indels. From the total number of identified SNPs, on average, 58% of the SNPs were shared among the three cattle populations (**Figure 1**) with the highest number of SNPs shared between AFR and NGI.

Validation of SNPs Using 1000 Bull Genomes Data

Run 6 of the 1000 Bull Genomes Project (Daetwyler et al., 2014; Frischknecht et al., 2017) was used to identify SNPs in the three SA breeds that are in common with other cattle breeds worldwide (i.e., taurine and indicine). On average, 93% of all SNPs identified in the three SA indigenous breeds were also shared among the breeds represented in the 1000 Bull Genomes Project data (**Table 3**). The remaining 7% of SNPs appear to be unique to SA indigenous breeds, AFR (7%), DRA (6%), and NGI (7%), using Run 6 of 1000 Bull Genomes Project data (**Figure 2**).

SNP Annotation and Analysis of Functional Enrichment

SNP annotation using VEP Ensembl gene annotation and dbSNP indicated that 62% of the SNPs were located in intergenic regions (AFR:62%, DRA:61%, NGI:62%), 29% were located in genic regions including introns, splice sites, and exons. Fewer SNPs (9%) were located in upstream/downstream regions (i.e., 5′ and 3′ untranslated regions; UTR). **Tables 4**, **5** indicate the distribution of SNPs and Indels detected within each functional class within genic regions.

Of the total number of Indels, 61% were located in intergenic regions, 28% in genic regions including introns, exons and splice sites, and 1% were located in untranlsted regions. In AFR, there were 433,495 (4.4%) SNPs located within 5 kb upstream of a transcription start site and 437,355 (4.4%) SNPs within 5 kb downstream of a transcription stop site; 3,974 (0.04%) SNPs were located in a 5′UTR and 18,999 (0.2%) in a 3′UTR. These totals were slightly different in other two breeds, and were slightly lower in NGI. A total of 20,508 SNPs across the three breeds were located in splice sites, and 498 SNPs were in splice/donor sites. A total of 109,067 non-synonymous single nucleotide polymorphisms (nsSNP) substitutions were observed. There were 868 SNPs predicted to cause premature stop codons and 75 to cause gains in coding sequence across the breeds.

Distribution of SNPs and Their Associated Genes

Figure 3 summarizes the distribution of four different categories of SNPs in the three breeds. The figure shows that while most of regions found to be enriched for these four categories were shared, some were breed specific. For example, regions spanning multiple 1 Mb windows in chr 12, chr 18, chr 23 were found to

TABLE 1 | Sequencing results for indigenous Afrikaner, Drakensberger, and Nguni cattle breeds.

Breed	Animals pooled	Raw reads	Non-duplicated reads	Properly paired reads	Mapped reads	High quality mapped reads	Average coverage
AFR	30	537,681,018	518,717,587	500,986,036	536,215,468	424,043,570 (79%)	21.2X
DRA	30	540,797,394	498,063,449	502,707,076	537,486,252	385,388,748 (71%)	15.4X
NGI	30	682,407,201	646,078,421	640,580,750	680,935,451	528,151,411 (77%)	26.6.X
Total	90	1,760,885,613	1,662,859,457	1,644,273,862	1,754,637,171	1,337,583,729 (76%)	21.1X

TABLE 2 | Summary of SNPs and Indels identified in Afrikaner (AFR), Drakensberger (DRA), and Nguni (NGI).

		SNPs		Indels	
Breed	No. variants	No. SNPs	Proportion SNPs	No. Indels	Proportion indels
AFR	11,165,172	9,950,384	0.89	1,212,231	0.11
DRA	7,049,789	6,327,515	0.90	721,628	0.10
NGI	12,514,952	11,164,415	0.89	1,347,215	0.11
Total	17,243,304	15,442, 314	0.89	1,908,137	0.11

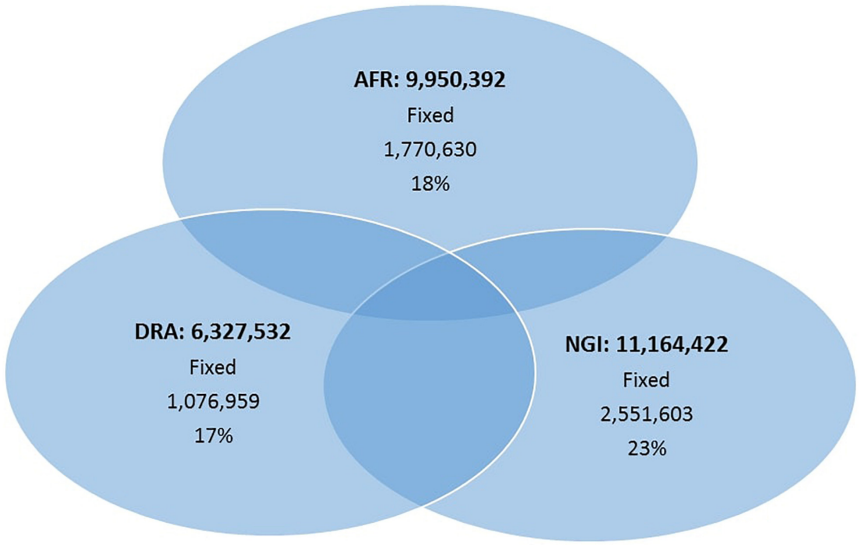

FIGURE 1 | The number of SNPs identified among the three indigenous South African breeds.

TABLE 3 | Novel variants identified in the three breeds through comparison to 1000 Bull Genomes Project Run 6 data.

	All Variants				SNPs			
Breed	**Known variants**	**Novel variants**	**Total**	**Proportion novel variants**	**Known SNPs**	**Novel SNPs**	**Total**	**Proportion novel SNPs**
AFR	9,381,545	614,536	9,996,081	0.07	8,576,732	617,296	9,194,028	0.07
DRA	6,307,154	381,743	6,688,897	0.06	5,764,627	413,795	6,178,422	0.06
NGI	10,693,999	631,412	11,325,411	0.07	9,793,635	647,269	10,440,904	0.07
Total	26,382,698	1,627,691	28,010,389	0.07 (Av)	24,134,994	1,678,360	25,813,354	0.07 (Av)

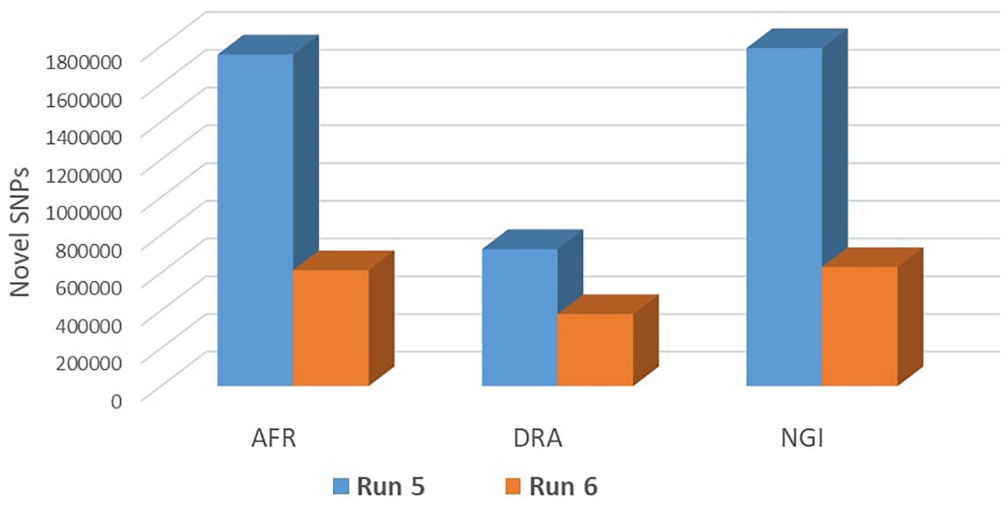

FIGURE 2 | Validation of novel SNPs using Run 6 of 1000 Genomes Project data.

show high overall SNP density in all three breeds, while a window on chr 28 (44,000,000–45,000,000) appeared among the top 1% regions in AFR and NG but did not appear even in the top 10% in DRA. Similar trends were also observed in the distribution of missense variants where regions on chr 4, chr 10, chr 15, chr 18, chr 19, chr 23, chr 25, and chr 29 were observed to be enriched

TABLE 4 | Counts of SNPs within each functional class for gene regions.

SNP class	Count						Total
	AFR	%	DRA	%	NGI	%	
Downstream_gene	437,355	4.4	288,515	4.6	440,357	3.9	1,166,227
Stop_lost	318	0.003	200	0.003	350	0.003	868
Stop_gain	38	0.0004	22	0.0003	15	0.0001	75
Splice_site	7,650	0.08	5,305	0.008	7,553	0.07	20,508
Upstream_gene	433,495	4.4	435,935	6.9	435,955	3.9	1,305,385
Intronic	2,726,502	27.4	1,800,155	28.4	2,731,530	24.5	7,258,187
miRNA	32,911	0.33	21,670	0.34	33,544	0.3	88,125
Synonymous_coding	38,537	0.4	29,836	0.47	40,694	0.36	109,067
Nonsynonymous_coding	31,205	0.31	22,395	0.35	31,130	0.28	84,730
3′_UTR	18,999	0.2	13,163	0.21	18,968	1.7	51,130
5′_UTR	3,974	0.04	3,055	0.05	3,805	0.034	10,834
Within_non_coding_gene	8,561	0.09	5,608	0.09	8,725	0.08	22,894
Essential_splice_site	182	0.002	124	0.002	192	0.002	498
Total	3,739,545	37.6	2,625,859	41.5	3,752,626	33.6	10,033,300

TABLE 5 | Counts of Indels by functional class for gene regions.

Indel class	Count						Total
	AFR	%	DRA	%	NGI	%	
Downstream_gene	126,159	10.4	73,669	10.2	50,823	3.8	250,651
Stop_lost	43	0.004	49	0.007	26	0.002	118
Stop_gain	82	0.007	115	0.016	34	0.003	231
Splice_site	2481	0.2	1667	0.23	952	0.007	5,100
Upstream_gene	123,341	10.2	71,747	10.4	48,080	3.6	243,168
Intronic	745,500	61.5	431,225	59.8	317,114	23.5	1,493,839
miRNA	10,296	0.85	5,644	0.8	3,816	0.28	19,756
Synonymous_coding	1,004	0.08	855	0.12	449	0.33	2,308
Non-synonymous_coding	2,943	0.24	2,293	0.32	1,145	0.008	6,381
3′_UTR	5,574	0.46	3,165	0.44	2,166	0.16	10,905
5 ′_UTR	842	0.07	660	0.01	376	0.028	1,878
Within_non_coding_gene	2,141	0.18	1,311	0.18	545	0.04	3,997
Total	1,020,406	84.1	592,400	82.1	425,526	31.6	2,038,332

with these variants in all three of the breeds. On the other hand two consecutive windows chr 5 (containing *ATN1*, *C3AR1,* and other genes) were found to be among top 1% missense variants in AFR but were not observed among the top 10% regions for either NGI or DRA.

Many of the regions which showed top 1% SNP densities in one or more of the categories have been found to be associated to disease or critical functional processes in previous studies, highlighting possible evolutionary rationale for their unusual densities. A window on Chr 4 corresponding to the *TRB* genes (*TRBV15, 16, 29, 30, and TRBV6*) was found to be in the top 1% both for overall SNP density as well as missense SNPs density in AFR and DRA (but not in NGI). *TRB* genes have been associated with domestication in cattle, showing strong evolutionary pressure (Connelley et al., 2009). The *CLEC5A* gene also from the same region of Chr 4 has been associated with diseases (increased level of swine influenza) in pigs (Fraser, 2018). *CDC42EP5* gene on Chr 18 (region detected to be enriched with missense variants in all three breeds) that has been associate with coat/hair pigmentation in mouse, is associated with meat tenderness in Nellore cattle (Carvalho et al., 2017). A region on chr 12 containing the *CLDN10, DNAJC3, and DZIP1* genes was found to show top 1% overall SNP density as well top 1% novel SNP density in both AFR and DRA. *DNAJC3* is a heat shock protein gene previously associated with embryonic development in Zebra fish (Soares et al., 2012). A region on chr 28 enriched in LoF variants in DRA and NGI contains the *TTC13* gene that has been associated with domestication – adaptation in cattle (Chen N. et al., 2018). Similarly, the *BTBD17* gene in Chr 9 (corresponding region shows enrichment for missense variants in all the three breeds) has been associated with network of genes that regulate milk yield in Holstein cattle (Chen Z. et al., 2018), and also responsible for abortion – embryonic lethalty in dogs. *PAG5* gene in Chr 29 (corresponding region showing non-synonymous variants enrichment only in NGI) have been

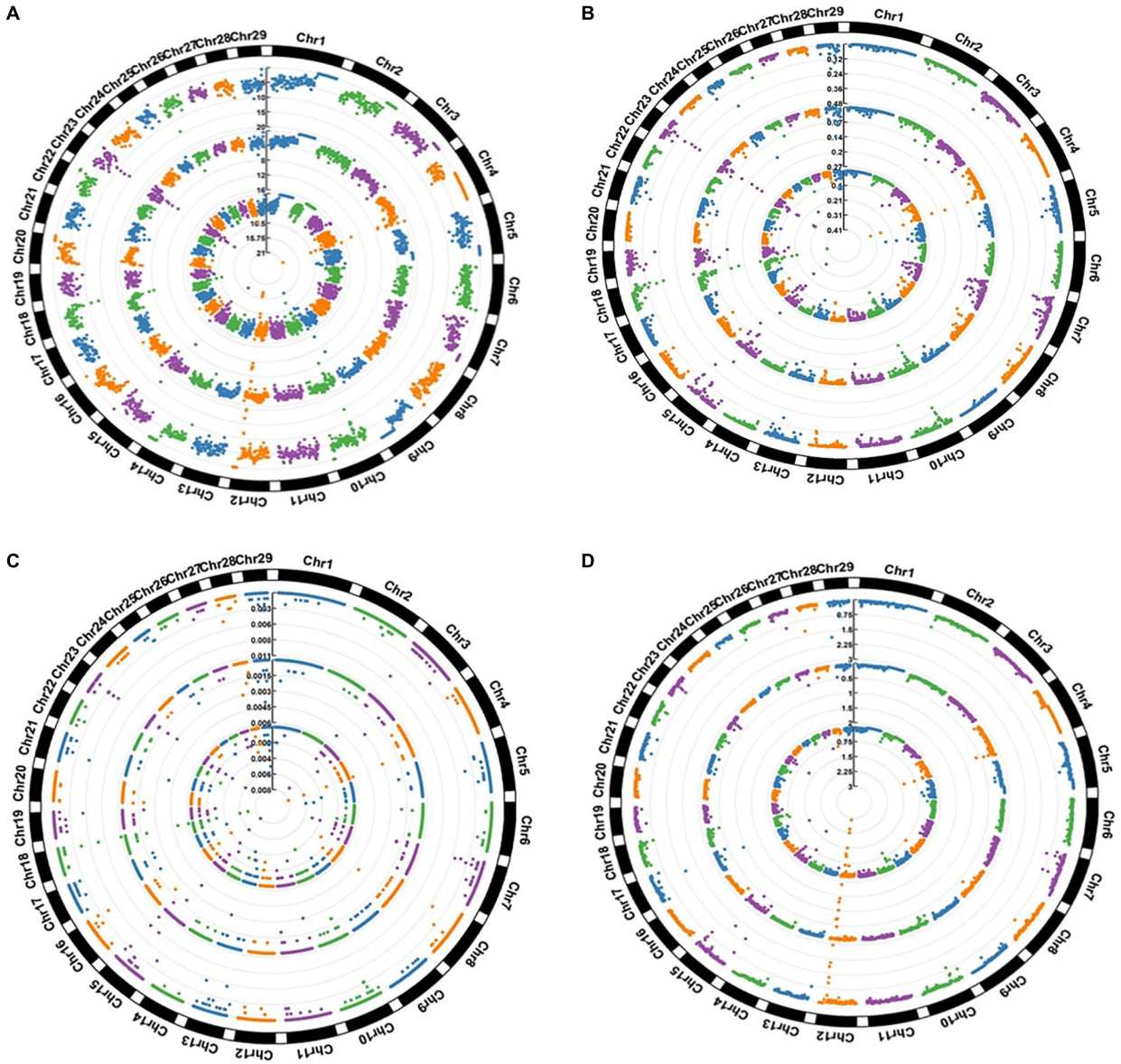

FIGURE 3 | Distribution of SNPs in 1 Mb windows **(A)** All SNPs **(B)** Missense SNPs **(C)** LoF (stop gain and stop loss) SNPs **(D)** Novel SNPs. The values expressed are Number of SNPs per kilobase. The inner circle represents AFR, the middle circle represent DRA and the outer circle represent NGI.

associated with pregnancy (formation of placenta) in cattle while *R8I2* in Chr 15 have been associated with taste in rat, as well as sensory perception of smell in humans (Ignatieva et al., 2014; Wallace et al., 2015).

Identification of Selective Sweeps

A total of 33,467 150 kb sliding windows were used to calculate the Z-transformed pooled heterozygosity (ZHp) scores to identify putative selective sweep regions. The ZHp Z-scores ranged from −10.26 to 2.18, from −5.27 to 1.37, and from −8.27 to 1.94 in AFR, DRA, and NGI, respectively. Thus, there appeared to be regions of excess homozygosity but not excess heterozygosity in the genomes of these animals. **Figure 4** show the distributions of the ZHp Z-scores genome-wide for the three indigenous SA breeds. The most noteworthy regions of homozygosity were observed in a region spanning <50 Mb on chromosomes 8 and 19 in NGI, chromosomes 13 and 19 in AFR, and on chromosome 7 in DRA.

The genome-wide screening of the breeds revealed 113 distinct loci with ZHp Z-scores ≤ -5, and 157 loci with ZHp Z-scores ≤ -4 in AFR, 2 and 152, respectively in DRA, and 108 and 156, respectively in NGI. In total, 465 candidate selective sweeps with ZHp Z-scores ≤ -4 were identified across the genomes of AFR, DRA, and NGI and 223 regions were identified as putative selective sweeps (ZHp Z-scores ≤ -5). The lowest number of putative selective sweeps was observed

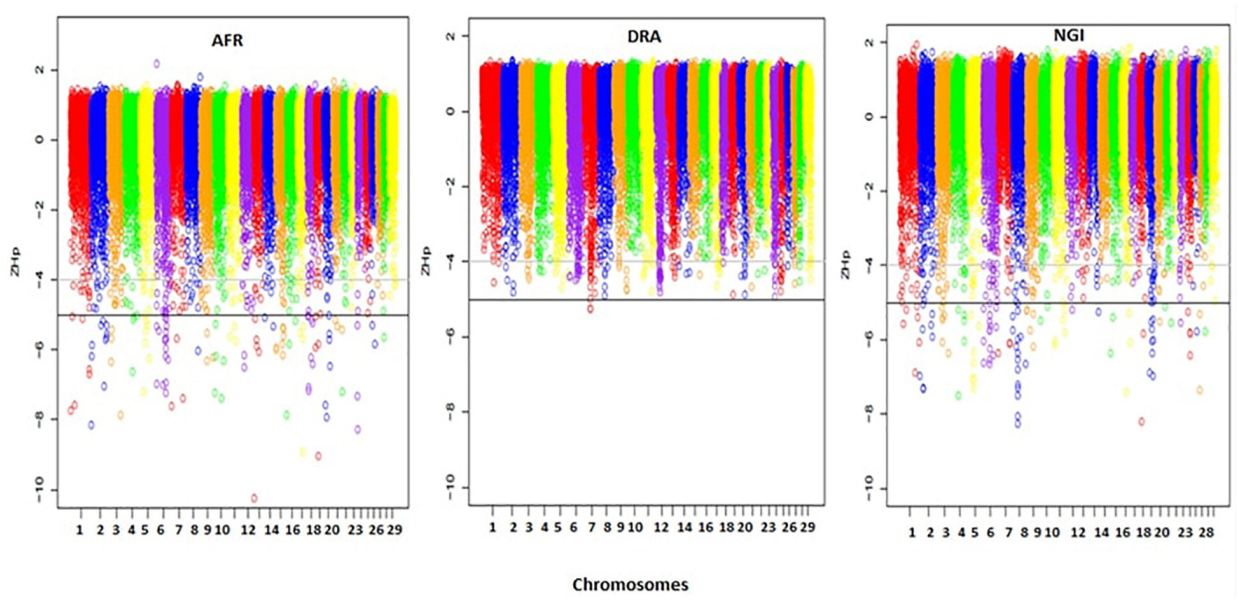

FIGURE 4 | Distribution of ZHp Z-scores across all 29 autosomes for Afrikaner (AFR), Drakensberger (DRA), and Nguni (NGI). The horizontal lines indicate ZHp Z-score thresholds of −4 and −5 used to define candidate and putative selective sweep regions in this study.

in DRA. The regions identified as candidate selective sweeps (ZHp Z-scores ≤ -4) in DRA were located on 26 different chromosomes. We also identified 93 selective sweep regions with extremely low ZHp Z-scores (ZHp-scores ≤ -6) as indicated in **Table 6**.

A locus with extremely low ZHp Z-score of −10.26 was found in AFR on chromosome 13, but no annotated genes were identified in this region. A protein coding gene, family with sequence similarity 101, member B (*FAM101B*) was identified in a sweep region with a ZHp Z-score of −9.05 in AFR on chromosome 19, and was also found in NGI with a ZHp Z-score of −8.2. This gene is involved in the regulation of the perinuclear actin network and nuclear shape through interaction with filamins, and plays an essential role in the formation of cartilaginous skeletal elements in human. There were also a few overlapping common genes that were identified across all the three breeds that could have been associated with breed formation in cattle. The *KIT* and *MITF* genes on chromosomes 6 and 12 respectively, have been associated with pigmentation in cattle, *KDR* on chromosome 6 is a tyrosine kinase receptor, *ERBB4* on chromosome 2 is associated with a signaling pathway involved in the development and progression of melanocytes in human (Choi et al., 2010). Other genes include *CACNA1C* on BTA5, *LAMC3* on BTA11, *TAS2R16* on BTA4, *UNC93A* on BTA9, *TNFRSF9* on BTA16, *CAV2* on BTA4, and *DCST1* on BTA3. These genes have previously been identified in selective sweep regions in cattle and have been associated with: (1) major depression, (2) the development of brain cortex and formation of axons, (3) dietary habits, (4) associated with Herpes simplex encephalitis type 1, (5) induced by lymphocyte activation, (6) involved in Cystic Fibrosis, and (7) implicated in Down syndrome, respectively (Qanbari et al., 2014). The keratin genes *KRT24, KRT25, KRT26, KRT27,* and *KRT28*; and the heat shock protein gene *HSPB9* found on chromosome 19, which have previously been associated with adaptation to tropical environment in Zebu cattle, were detected in selective sweep regions common to all three breeds. Other associated genes including *ATP2B*, *FMOD*, *WNT5B,* and *PRELP* on chromosome 16, have also previously been identified as being under positive selection in cattle, and were located in sweep regions shared across the three breeds.

DISCUSSION

Sequencing of individuals can identify millions of SNPs that differ between any two individual genomes (Bischoff et al., 2008), while good coverage ensures better identification of sequencing errors, increasing sequencing accuracy (Borisevich et al., 2017). Pools of DNA were sequenced for Afrikaner, Drakensberger and Nguni to discover new SNPs. The average sequence depth of 21-fold was obtained in this study, a coverage efficient to make accurate SNP calls (Wang et al., 2011). The sequence depth was relative to studies by Eck et al. (2009) and Stothard et al. (2011), but higher than the study by Choi et al. (2015) with an average coverage of 10.71X for Hanwoo and Yanbian cattle, and lower than the 27X mean coverage obtained by Das et al. (2015) for Danish Holstein dairy cattle. The Ti/Vi ratio are computed for quality control of sequence data and are helpful for understanding patterns of DNA sequence evolution (Wang et al., 2014). The quality of sequence data obtained from this study was comparable to the studies of Gayal, Red Angus, and Japanese Black cattle where the Ti/Tv values were 2.32, 2.17, and 2.18, respectively (Mei et al., 2016). These results suggest

TABLE 6 | List of genes within SNP enriched genomic regions in the top 100 kb window.

Gene	CHR	Function	Species	Reference
Afrikaner				
MOV10	3	Gene silencing by miRNA	Human	Goodier et al., 2012
MPV17	11	Abnormal coat/hair pigmentation, thin skin, decreased body weight, kidney failure, anemia, hypertension, increased heart rate	Mouse	Weiher et al., 1990; Viscomi et al., 2009
UCN	11	Increased anxiety, feeding behavior, heart failure, decreased drinking behavior, parkinsonian disorders	Mouse, rat	Vetter et al., 2002
TRIM54	11	Premature death, abnormal heart morphology	Mouse	Hwang et al., 2010
DNAJC5G	11	Cardiovascular system phenotype, decreased anxiety-related response	Mouse	Rovelet-Lecrux et al., 2012
WNT4	2	Serkal syndrome, female sex determination, kidney failure, male sex differentiation	Mammals, mouse	Vainio et al., 1999; Brisken et al., 2000
CDC42	2	Negative regulation of gene expression, hair follicle placode formation, spinal cord injuries, bipolar disorder, epilepsy arthritis	Mouse, rat	Erschbamer et al., 2005; Park et al., 2009
MLANA	8	Diluted coat color, hair morphology	Mouse	Steingrimsson et al., 2006
KIAA1549	4	Decreased total body fat amount, pilocytic astrocytoma (brain tumor)	Human, mouse	Hughes, 1998; Antonelli et al., 2015
HECTD3	3	Decreased lean body mass, length, increased total body fat amount	Mouse	Zhang et al., 2009
Drakensberger				
YTHDC2	10	Prostatic neoplasms	Rat	Arambula et al., 2016
DCLRE1B	3	Decreased embryo size, neonatal lethality, cell cycle checkpoint	Mouse, human	Liu et al., 2009; Dronkert et al., 2000
AP4B1	3	Spastic paraplegia, autosomal recessive	Mouse, human	Tuysuz et al., 2014
PTPN22	3	Autoimmune diseases, enlarged spleen, diabetes mellitus, insulin-dependent	Human, mouse, rat	Bottini et al., 2006; Michou et al., 2007
ZC3HAV1	4	Suppression by virus of host molecular function, endosome to lysosome transport	Mouse	Lee et al., 2009
PSMB11	10	Increased T-cell proliferation, abnormal self-tolerance	Mouse	Anderson and Takahama, 2012
AJUBA	10	Gene silencing by miRNA, wound healing, spreading of epidermal cells, heart contraction, decreased rate, abnormal cell migration	Human, zebrafish, mouse	Bergantinos et al., 2010; Wilkinson et al., 2014
SLC7A8	10	Decreased susceptibility to pharmacologically induced seizures	Mouse	Dai et al., 2007
IFT74	8	Abnormal lung lobe morphology, notch signaling involved in heart development, cilium assembly	Human, mouse	Kwong et al., 2007; Bhogaraju et al., 2013
SUPT7L	11	Abnormal hair texture, decreased body weight, embryonic lethality	Mouse	Bardot et al., 2016
Nguni				
RAB33B	17	Skeletal system morphogenesis	Human	Bonafe et al., 2015
SYT10	5	Shortened circadian period (sleep disorder), sensory perception of smell	Mouse	Anda et al., 2016
STT3B	22	Congenital disorder of glycosylation	Human	Scott et al., 2014
CEACAM16	18	Deafness, autosomal dominant 4b	Human, mouse	Zheng et al., 2011; Lukashkin et al., 2012
SRGAP2	16	Dendritic spine development	Mouse	Charrier et al., 2012
TMEM98	19	Nanophthalmia, hemorrhage	Human, mouse	Liao et al., 2016
CCL17	18	Staphylococcal pneumonia, bronchiolitis obliterans	Mouse	Montgomery and Daum, 2009
TXN	8	Fatty liver, myocarditis, diabetes mellitus	Rat	Chung et al., 2011
COG5	4	Congenital disorder	Human	Wu et al., 2004
AIRE	1	Reduced fertility, thyroid, and eye inflammation	Mouse	Schaller et al., 2008

that the majority of SNPs identified in this study were accurately identified (Choi et al., 2015).

Due to the history of human migration and trading, it is expected that indigenous breeds will often have multiple genetic signatures of origin and admixture, and this has been confirmed by analyses using available molecular data (Hanotte and Jianlin, 2006; Decker et al., 2014; Makina et al., 2014). These analyses have suggested that several ancestral lineages have contributed to today's genetic pool of livestock (Hanotte et al., 2000; Qi, 2004). The proportion of novel SNPs found

in this study was low compared to what was reported by Choi et al. (2013) where 29.4% of SNPs were found to be novel in Korean Black Cattle when compared to the dbSNP version 137. This likely reflects the large number of SNPs that have now been discovered in 1000 Bull Genomes Project (80 million SNPs in Run 6). This shows the effort made by researchers to address the issue of SNP biasness reflected on the current SNP assays of which mostly European breeds were used in the design of these assays. The availability of this data will allow for imputation of genetic variants for genomic prediction and genome wide association studies in all cattle breeds.

Higher number of novel SNPs identified in the study of Mei et al. (2016) could reflect different SNP filtering criteria, the comparison set used (dbSNP Build 140), and the increased divergence of Gayal cattle from the reference genome, Hereford, relative to the SA breeds. The detected SNPs were validated using dbSNP Build 140, which also represents a smaller validation set than was used in this study. The greater number of novel SNPs found in AFR and DRA cattle likely reflects the extent of genetic diversity that exists between these breeds and also their phylogenetic distance from the reference genome. Novel variants characterize the extent of genetic differentiation that exists between individuals and populations (Choudhury et al., 2014). The lower number of novel SNPs found in DRA suggests that the breed might be more closely related to European breeds than the AFR or NGI (Zwane et al., 2016). The complex origins of cattle are associated with both natural and artificial selection, which gave rise to numerous different breeds displaying a broad spectrum of phenotypes. This happened after the global partitioning of the world-wide cattle genetic diversity by three distinct events, two of which involved domestication, and that resulted in European taurines, West African taurines and Zebu from India spreading all over the world through the migration of different tribes (Gautier et al., 2010; Decker et al., 2014).

The identification of functional variants such as missense variants, and variants within upstream and downstream genic regions in indigenous SA cattle will enable the testing of these variants for their effects on complex traits (Koufariotis et al., 2014). While the roles of variation in overlapping genes is less clear, studies have suggested that this could be a mechanism allowing the regulation of key genes in eukaryotes (Park et al., 2009). Further studies of overlapping genes will enable an understanding of the tissue- and developmental-stage regulation of each strand and will provide insight into their mechanisms of evolution (Nakayama et al., 2006). Genetic variants such as insertions, deletions and structural variants may also be applied in association studies and in genomic prediction (Koufariotis et al., 2014).

The number of SNPs that were located on the splice sites were higher than found by Stothard et al. (2011) in Holstein and Black Angus, and Choi et al. (2013) in Heugu cattle. This could be due to the number of samples, breeds used for sequencing and different genetic backgrounds. The number of functional and non-functional genes differs depending on the breeds and the method used for annotation (Das et al., 2015). The number of functionally annotated Indels was slightly higher than the number of detected Indel loci, because a SNP or Indel locus may have multiple annotations (Choi et al., 2015). The numbers of SNPs and Indels identified in this study were slightly greater in NGI and AFR than in DRA due to the higher indicine percentage present in their genomes (Makina et al., 2016). The numbers of nsSNPs segregating in these breeds were greater than for Danish Jutland Cattle, which had 34,257 nonsynonymous substitutions (34,183 missense and 74 initiator codon variants) identified from four cows (Das et al., 2015). The ability to identify non-neutral substitutions could help targeting diseases caused by detrimental mutations, and SNPs that increase the fitness of particular phenotypes (Bromberg and Rost, 2007). In human, among all types of variants, nsSNPs are believed to be the major contributors to heritable diseases. They constitute more than half of the disease-causing genetic changes deposited in the Human Gene Mutation Database (HGMD) (Stenson et al., 2009).

The analysis of SNP density distribution of the four different categories of SNPs was able to identify genomic regions showing distinctive trends among the breeds. Genes showing high densities of different SNP categories were identified to be associated to phenotypes (coat color), adaptation, fertility (embryonic development, placenta formation), production (meat tenderness) and diseases (abortion, induced viral infection), which represent some of the desired traits in livestock production. More validation of these associations are still needed, especially in farm animals. The global study of the 1000 Bulls Genomes Project has mined genes that influence complex genetic traits in cattle, opening the door for researchers to use the same approach to map high-value traits including those important for beef and milk production. We expect the current data would serve as an important resource, for a comprehensive analysis comparing genomic distribution and densities of SNPs on a global scale.

It has been suggested that rare or low-frequency variants may explain a substantial proportion of the heritability of many complex diseases, most of which have previously not been fully captured in GWAS studies (Bang et al., 2014). The power to identify variants associated with traits, particularly those of small effect, could be increased if certain regions of the genome are known to be enriched for trait associations (Koufariotis et al., 2014). However, given the typical genetic architecture of complex traits, such regions are likely to be very few as also observed in this study. Variants in regions of the genome for which the sequence is strongly conserved across species have been proposed as an important annotation class for prioritization since they are potentially regulatory. The majority of these variants are found in non-coding regions, and it is believed that at least some of these are cis regulators for genes (Knight et al., 2011).

Searching for genomic regions of reduced variability as signatures of strong positive selection can also help in identifying causal mutations controlling selected phenotypes (Voight et al., 2006). Selective sweeps and their associated genes provide an insight into the genomic footprints left by natural and artificial selection in indigenous SA breeds. While identifying a

selective sweep in the same region in different breeds provides support that a particular genomic region has undergone selection for a given trait, many selection signatures appear to be breed-specific (Gutiérrez-Gil1 et al., 2015). This study identified 465 candidate selective sweeps with ZHp scores ≤ -4 on 29 chromosomes and 223 regions were identified as putative selective sweeps (ZHp Z-scores ≤ -5) on 17 chromosomes. Using BovineSNP50 data, Ramey et al. (2013) identified 28 genomic regions on 15 chromosomes as putatively harboring selective sweeps in 14 breeds. They also identified 85 putative selective sweep regions from 200 to 846 kb in size using the very high density AFFXB1P assay. Only 11 regions were validated as putative selective sweeps using both assays and no selective sweeps overlapped between the taurine and indicine breeds. For several of the detected sweep regions, Ramey et al. (2013) were able to identify the phenotypes and genes that were likely subjected to selection. However, for many of these regions, the selected genes and phenotypes were unclear. But when using NGS, Qanbari et al. (2014) identified 146 regions of positive selection in non-overlapping 40 kb windows across the genome. They were able to localize regions/genes harboring phenotypic characteristics such as patterned pigmentation, brain development and neurobehavioral functioning, sensory perception, immune system, genetic disorders, and blood coagulation. This shows that the amount of data used and the analytical method employed both impact the identification of the number of regions of positive selection. In this study, the number of identified selective sweeps was even higher. A total of 93 putative selective sweeps with extremely low ZHp Z-scores (ZHp Z-scores ≤ -6.0) were identified across the genomes of Afrikaner and Nguni. This information will help in the discovery of disease resistance alleles and for the inference of the events that molded the genetic structure of these populations. These imprints of historic selection/adaptation episodes left in cattle genomes allow one to interpret modern and ancestral gene origins and modifications (Qanbari et al., 2014).

CONCLUSION

The SNPs and Indels identified in this study will serve as useful genetic tools, and as candidates in searches for phenotype-altering DNA differences. Novel SNPs provide an insight into the genomic regions that are unique to each breed. Identification of nsSNPs provides the potential for the detection of genes and variants underlying variation in traits of economic importance in these breeds, in particular environmental adaptation. Genes located in genomic regions that are enriched for variation suggests their potential for selection due to effects on phenotypic characteristics. Identification of selective sweeps provides a broader insight into the events that happened during recent selection events and artificial selection processes that have shaped the livestock genome in SA indigenous cattle breeds. These results provide a framework for further genetic association and QTL fine-mapping studies in indigenous SA cattle.

AUTHOR CONTRIBUTIONS

AZ designed the experiments, carried out the analysis, and drafted the manuscript. RS and JH assisted with structuring the methodology, data handling, and data management. AC and MM assisted with statistical analysis. AM, EVM-K, and JT structured scientific content. All authors provided editorial suggestions and revisions, read and approved the manuscript.

ACKNOWLEDGMENTS

We would like to acknowledge the financial support from the Red Meat Research and Development of South Africa (RMRDSA), South Africa Beef Genomic Project (BGP), and financial assistance of the National Research Foundation (NRF) toward this research. Opinions expressed and conclusions arrived at are those of the authors.

REFERENCES

Abin, S., Theron, H. E., and Van Marle-Koster, E. (2016). Population structure and genetic trends for indigenous African beef cattle breeds in South Africa: short communication. *S. Afr. J. Anim. Sci.* 46, 152–156. doi: 10.4314/sajas.v46i2.5

Albrechtsen, A., Nielsen, F. C., and Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Mol. Biol. Evol.* 27, 2534–2547. doi: 10.1093/molbev/msq148

Anda, F. C., Madabhushi, R., Rei, D., Meng, J., Gräff, J., Durak, O., et al. (2016). Cortical neurons gradually attain a post-mitotic state. *Cell Res.* 26, 1033–1047. doi: 10.1038/cr.2016.76

Anderson, G., and Takahama, Y. (2012). Thymic epithelial cells: working class heroes for T cell development and repertoire selection. *Trends Immunol.* 33, 256–263. doi: 10.1016/j.it.2012.03.005

Antonelli, M., Badiali, M., Moi, L., Buttarelli, F. R., Baldi, C., Massimino, M., et al. (2015). KIAA1549: BRAF fusion gene in pediatric brain tumors of various histogenesis. *Pediatr. Blood Cancer* 62, 724–727. doi: 10.1002/pbc.25272

Arambula, S. E., Belcher, S. M., Planchart, A., Turner, S. D., and Patisaul, H. B. (2016). Impact of low dose oral exposure to Bisphenol A (BPA) on the neonatal rat hypothalamic and hippocampal transcriptome: a CLARITY-BPA consortium study. *Endocrinology* 157, 3856–3872. doi: 10.1210/en.2016-1339

Aslam, M. L., Bastiaansen, J. W., Elferink, M. G., Megens, H. J., Crooijmans, R. P., Blomberg, L. A., et al. (2012). Whole genome SNP discovery and analysis of genetic diversity in Turkey (*Meleagris gallopavo*). *BMC Genomics* 13:391. doi: 10.1186/1471-2164-13-391

Bachmann, M. (1983). Early origins of cattle. *Farmers Weekly* 23:18.

Bang, S. Y., Na, Y. J., Kim, K., Joo, Y. B., Park, Y., Lee, J., et al. (2014). Targeted exon sequencing fails to identify rare coding variants with large effect in rheumatoid arthritis. *Arthritis Res. Ther.* 16:447. doi: 10.1186/s13075-014-0447-7

Bardot, P., Vincent, S. D., Fournier, M., Hubaud, A., Joint, M., Tora, L., et al. (2016). The TAF10-containing TFIID and SAGA transcriptional complexes are dispensable for early somitogenesis in the mouse embryo. *bioRxiv* [preprint]. doi: 10.1101/071324

Bareke, E., Saillour, V., Spinella, J. F., Vidal, R., Healy, J., Sinnett, D., et al. (2013). Joint genotype inference with germline and somatic mutations. *BMC Bioinformatics* 14:S3. doi: 10.1186/1471-2105-14-S5-S3

Barris, W., Harrison, B. E., McWilliam, S., Bunch, R. J., Goddard, M. E., and Barendse, W. (2012). Next generation sequencing of African and Indicine cattle to identify single nucleotide polymorphisms. *Anim. Prod.* 52, 133–142. doi: 10.1071/AN11095

Bergantinos, C., Corominas, M., and Serras, F. (2010). Cell death-induced regeneration in wing imaginal discs requires JNK signaling. *Development* 137, 1169–1179. doi: 10.1242/dev.045559

Bhogaraju, S., Cajanek, L., Fort, C., Blisnick, T., Weber, K., Taschner, M., et al. (2013). Molecular basis of tubulin transport within the cilium by IFT74 and IFT81. *Science* 341, 1009–1012. doi: 10.1126/science.1240985

Bischoff, S. R., Tsai, S., Hardison, N. E., York, A. M., Freking, B. A., Nonneman, D., et al. (2008). Identification of SNPs and INDELS in swine transcribed sequences using short oligonucleotide microarrays. *BMC Genomics* 9:252. doi: 10.1186/1471-2164-9-252

Boichard, D. (2002). "PEDIG: a fortran package for pedigree analysis suited for large populations," in *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Vol. 32, Montpellier, 525–528.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bonafe, L., Cormier-Daire, V., Hall, C., Lachman, R., Mortier, G., Mundlos, S., et al. (2015). Nosology and classification of genetic skeletal disorders: 2015 revision. *Am. J. Med. Genet.* 167, 2869–2892. doi: 10.1002/ajmg.a.37365

Borisevich, D. I., Krasnenko, A. Yu., Stetsenko, I. F., Plakhina, D. A., and Ilinsky, V. V. (2017). The impact of sequencing depth on accuracy of single nucleotide variant calls. *Bull. Russia* 3, 48–52. doi: 10.24075/brsmu.2017-03-06

Bottini, N., Vang, T., Cucca, F., and Mustelin, T. (2006). Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin. Immunol.* 18, 207–213. doi: 10.1016/j.smim.2006.03.008

Boutet, G., Carvalho, S. A., Falque, M., Peterlongo, P., Lhuillier, E., Bouchez, O., et al. (2016). SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genomics* 17:121. doi: 10.1186/s12864-016-2447-2

Brisken, C., Heineman, A., Chavarria, T., Elenbaas, B., Tan, J., Dey, S. K., et al. (2000). Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes. Dev.* 14, 650–654.

Bromberg, Y., and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* 35, 3823–3835. doi: 10.1093/nar/gkm238

Carvalho, M. E., Baldi, F. S., Santana, M. H. A., Ventura, R. V., Oliveira, G. A., Bueno, R. S., et al. (2017). Identification of genomic regions related to tenderness in Nellore beef cattle. *Adv. Anim. Biosci.* 8, s42–s44. doi: 10.1371/journal.pone.0157845

Charrier, C., Joshi, K., Coutinho-Budd, J., Kim, J. E., Lambert, N., De Marchena, J., et al. (2012). Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* 149, 923–935. doi: 10.1016/j.cell.2012.03.034

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Comm.* 9:2337. doi: 10.1038/s41467-018-04737-0

Chen, Z., Yao, O., Ma, P., Wang, Q., and Pan, Y. (2018). Haplotype-based genome-wide association study identifies loci and candidate genes for milk yield in Holsteins. *PLoS One* 13:e0192695. doi: 10.1371/journal.pone.0192695

Choi, J. W., Choi, B. H., Lee, S. H., Lee, S. S., Kim, H. C., Yu, D., et al. (2015). Whole-genome resequencing analysis of Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection. *Mol. Cells* 38, 466–473. doi: 10.14348/molcells.2015.0019

Choi, J. W., Liao, X., Park, S., Jeon, H. J., Chung, W. H., Stothard, P., et al. (2013). Massively parallel sequencing of Chikso (Korean brindle cattle) to discover genome-wide SNPs and Indels. *Mol. Cells* 36, 203–211. doi: 10.1007/s10059-013-2347-0

Choi, J. W., Liao, X., Stothard, P., Chung, W. H., Jeon, H. J., Miller, S. P., et al. (2014). Whole-genome analyses of Korean native and Holstein cattle breeds by massively parallel sequencing. *PLoS One* 9:e101127. doi: 10.1371/journal.pone.0101127

Choi, W., Wolber, R., Gerwat, W., Mann, T., Batzer, J., Smuda, C., et al. (2010). The fibroblast-derived paracrine factor neuregulin-1 has a novel role in regulating the constitutive color and melanocyte function in human skin. *J. Cell Sci.* 123, 3102–3111. doi: 10.1242/jcs.064774

Choudhury, A., Hazelhurst, S., Meintjes, A., Achinike-Oduaran, O., Aron, S., Gamieldien, J., et al. (2014). Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. *BMC Genomics* 15:437. doi: 10.1186/1471-2164-15-437

Chung, J. H., Choi, H. J., Kim, S. Y., Hong, K. S., Min, S. K., Nam, M. H., et al. (2011). Proteomic and biochemical analyses reveal the activation of unfolded protein response, ERK-1/2 and ribosomal protein S6 signaling in experimental autoimmune myocarditis rat model. *BMC Genomics* 12:250. doi: 10.1186/1471-2164-12-520

Connelley, T., Aerts, J., Law, A., and Morrison, W. I. (2009). Genomic analysis reveals extensive gene duplication within the bovine TRB locus. *BMC Genomics* 10:192. doi: 10.1186/1471-2164-10-192

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R. F., et al. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46, 858–865. doi: 10.1038/ng.3034

Dai, Z., Huang, Y., Sadee, W., and Blower, P. (2007). Chemoinformatics analysis identifies cytotoxic compounds susceptible to chemoresistance mediated by glutathione and cystine/glutamate transport system xc. *J. Med. Chem.* 50, 1896–1906. doi: 10.1021/jm060960h

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Das, A., Panitz, F., Gregersen, V. R., Bendixen, C., and Holm, L. E. (2015). Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes. *BMC Genomics* 16:1043. doi: 10.1186/s12864-015-2249-y

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Alcala, A. M., Sonstegard, T. S., et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10:e1004254. doi: 10.1371/journal.pgen.1004254

Dronkert, M. L. G., De Wit, J., Boeve, M., Vasconcelos, M. L., van Steeg, H., Tan, T. L. R., et al. (2000). Disruption of mouse SNM1 causes increased sensitivity to the DNA interstrand cross-linking agent mitomycin C. *Mol. Cell. Biol.* 20, 4553–4561. doi: 10.1128/MCB.20.13.4553-4561.2000

Eck, S. H., Benet-Pages, A., Flisikowski, K., Meitinger, T., Fries, R., and Strom, T. M. (2009). Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome Biol.* 10:R82. doi: 10.1186/gb-2009-10-8-r82

Edea, Z., Dadi, H., Kim, S. W., Dessie, T., Lee, T., Kim, H., et al. (2013). Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front. Genet.* 4:35. doi: 10.3389/fgene.2013.00035

Erschbamer, M. K., Hofstetter, C. P., and Olson, L. (2005). RhoA, RhoB, RhoC, Rac1, Cdc42, and Tc10 mRNA levels in spinal cord, sensory ganglia, and corticospinal tract neurons and long-lasting specific changes following spinal cord injury. *J. Comp. Neurol.* 484, 224–233. doi: 10.1002/cne.20471

Fraser, R. S. (2018). *Investigation of Genetic Variation in the Collagenous Lectins of Livestock with and Without Infectious Diseases*. Ph.D. thesis, University of Guelph, Guelph, ON.

Frischknecht, M., Pausch, H., Bapst, B., Signer-Hasler, H., Flury, C., Garrick, D., et al. (2017). Highly accurate sequence imputation enables precise QTL mapping in Brown Swiss cattle. *BMC Genomics* 18:999. doi: 10.1186/s12864-017-4390-2

Gautier, M., Laloe, D., and Moazami-Goudarzi, K. (2010). Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One* 5:e13038. doi: 10.1371/journal.pone.0013038

Goodier, J. L., Cheung, L. E., and Kazazian, H. H. Jr. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.* 8:e1002941. doi: 10.1371/journal.pgen.1002941

Groeneveld, E., Van der Westhuizen, B., Maiwashe, A., Voordewind, F., and Ferraz, J. B. S. (2009). POPREP: a generic report for population management. *Genet. Mol. Res.* 8, 1158–1178. doi: 10.4238/vol8-3gmr648

Gurgul, A., Zukowski, K., Pawlina, K., Zqbek, T., Semik, E., and Bugno-Poniewierska, M. (2013). The evaluation of bovine SNP50 BeadChip assay performance in Polish Red cattle breed. *Folia Biol.* 61, 173–176. doi: 10.3409/fb61_3-4.173

Gutiérrez-Gil1, B., Arranz, J. J., and Wiener, P. (2015). An interpretive review of selective sweep studies in *Bos taurus* cattle populations: identification of unique and shared selection signals across breeds. *Front. Genet.* 6:167. doi: 10.3389/fgene.2015.00167

Hanotte, O., and Jianlin, H. (2006). "Genetic characterization of livestock populations and its use in conservation decision-making," in *The Role of Biotechnology in Exploring and Protecting Agricultural Genetic Resources*, eds J. Ruane and A. Sonnino (Rome: FAO), 89–96.

Hanotte, O., Tawah, C. L., Bradley, D. G., Okomo, M., Verjee, Y., Ochieng, J., et al. (2000). Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Mol. Ecol.* 9, 387–396. doi: 10.1046/j.1365-294x.2000.00858.x

Hu, Z. L., Park, C. A., Wu, X. L., and Reecy, J. M. (2013). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 41, D871–D879. doi: 10.1093/nar/gks1150

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Hughes, S. G. (1998). Prescribing for the elderly patient: why do we need to exercise caution? *Br. J. Clin. Pharmacol.* 46, 531–533. doi: 10.1046/j.1365-2125.1998.00842.x

Hwang, C. Y., Holl, J., Rajan, D., Lee, Y., Kim, S., Um, M., et al. (2010). Hsp70 interacts with the retroviral restriction factor TRIM5α and assists the folding of TRIM5α. *J. Biol. Chem.* 285, 7827–7837. doi: 10.1074/jbc.M109.040618

Ignatieva, E. V., Levitsky, V. G., Yudin, N. S., Moshkin, M. P., and Kolchanov, N. A. (2014). Genetic basis of olfactory cognition: extremely high level of DNA sequence polymorphism in promoter regions of the human olfactory receptor genes revealed using the 1000 Genomes Project dataset. *Front. Psychol.* 5:247. doi: 10.3389/fpsyg.2014.00247

Kawahara-Miki, R., Tsuda, K., Shiwa, Y., Arai-Kichise, Y., Matsumoto, T., Kanesaki, Y., et al. (2011). Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle Kuchinoshima-Ushi. *BMC Genomics* 12:103. doi: 10.1186/1471-2164-12-103

Knight, J., Barnes, M. R., Breen, G., and Weale, M. E. (2011). Using functional annotation for the empirical determination of bayes factors for genome-wide association study analysis. *PLoS One* 6:e14808. doi: 10.1371/journal.pone.0014808

Koufariotis, L., Chen, Y. P. P., Bolormaa, S., and Hayes, B. J. (2014). Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. *BMC Genomics* 15:436. doi: 10.1186/1471-2164-15-436

Kwong, L. K., Neumann, M., Sampathu, D. M., Lee, V. M. Y., and Trojanowski, J. Q. (2007). TDP-43 proteinopathy: the neuropathology underlying major forms of sporadic and familial frontotemporal lobar degeneration and motor neuron disease. *Acta Neuropathol.* 114, 63–70. doi: 10.1007/s00401-007-0226-5

Le Roex, N., Noyes, H., Brass, A., Bradley, D. G., Kemp, S. J., Kay, S., et al. (2012). Novel SNP discovery in African buffalo, Syncerus caffer, using high-throughput sequencing. *PLoS One* 7:e48792. doi: 10.1371/journal.pone.0048792

Lee, S. M., Gardy, J. L., Cheung, C. Y., Cheung, T. K., Hui, K. P., Ip, N. Y., et al. (2009). Systems-level comparison of host-responses elicited by avian H5N1 and seasonal H1N1 influenza viruses in primary human macrophages. *PLoS One* 4:e8072. doi: 10.1371/journal.pone.0008072

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liao, X., Lan, C., Liao, D., Tian, J., and Huang, X. (2016). Exploration and detection of potential regulatory variants in refractive error GWAS. *Sci. Rep.* 6:33090. doi: 10.1038/srep33090

Liu, L., Akhter, S., Bae, J. B., Mukhopadhyay, S. S., Richie, C. T., Liu, X., et al. (2009). SNM1B/Apollo interacts with astrin and is required for the prophase cell cycle checkpoint. *Cell Cycle* 8, 628–638. doi: 10.4161/cc.8.4.7791

Lukashkin, A. N., Legan, P. K., Weddell, T. D., Lukashkina, V. A., Goodyear, R. J., Welstead, L. J., et al. (2012). A mouse model for human deafness DFNB22 reveals that hearing impairment is due to a loss of inner hair cell stimulation. *Proc. Natl. Acad. Sci.* 109, 19351–19356. doi: 10.1073/pnas.1210159109

Makina, S. O., Muchadeyi, F. C., Marle-Köster, E., Taylor, J. F., Makgahlela, M. L., and Maiwashe, A. (2015). Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genet. Sel. Evol.* 47:92. doi: 10.1186/s12711-015-0173-x

Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., MacNeil, M. D., and Maiwashe, A. (2014). Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front. Genet.* 5:333. doi: 10.3389/fgene.2014.00333

Makina, S. O., Whitacre, L. K., Decker, J. E., Taylor, J. F., MacNeil, M. D., Scholtz, M. M., et al. (2016). Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide. *Genet. Sel. Evol.* 48:88. doi: 10.1186/s12711-016-0266-1

Mapholi, N. O. (2015). *Exploring Genetic Architecture of Tick Resistance in South African Nguni Cattle*. Doctoral dissertation, Stellenbosch University, Stellenbosch.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., et al. (2009). Development and characterization of a high-density SNP genotyping assay for cattle. *PLoS One* 4:e5350. doi: 10.1371/journal.pone.0005350

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics* 26, 2069–2070. doi: 10.1093/bioinformatics/btq330

McTavish, E. J., and Hillis, D. M. (2015). How do SNP ascertainment schemes and population demographics affect inferences about population history? *BMC Genomics* 16:266. doi: 10.1186/s12864-015-1469-5

Mei, C., Wang, H., Zhu, W., Wang, H., Cheng, G., Qu, K., et al. (2016). Whole-genome sequencing of the endangered bovine species Gayal (*Bos frontalis*) provides new insights into its genetic features. *Sci. Rep.* 6:19787. doi: 10.1038/srep19787

Meuwissen, T. H. E., and Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* 24, 305–313. doi: 10.1186/1297-9686-24-4-305

Michou, L., Lasbleiz, S., Rat, A. C., Migliorini, P., Balsa, A., Westhovens, R., et al. (2007). Linkage proof for PTPN22, a rheumatoid arthritis susceptibility gene and a human autoimmunity gene. *Proc. Natl. Acad. Sci.* 104, 1649–1654. doi: 10.1073/pnas.0610250104

Mitchell, K. J. (2015). *Using High-Throughput DNA Sequencing and Molecular Phylogenies to Investigate the Evolution and Biogeography of the Southern Hemisphere*. Doctoral dissertation, University of Adelaide, Adelaide, SA.

Montgomery, C. P., and Daum, R. S. (2009). Transcription of inflammatory genes in the lung after infection with community-associated methicillin-resistant Staphylococcus aureus: a role for Panton-Valentine leukocidin? *Infect. Immun.* 77, 2159–2167. doi: 10.1128/IAI.00021-09

Mostert, B. E. (2007). *The Suitability of Test-Day Models for Genetic Evaluation of Dairy Cattle in South Africa*. Ph.D. thesis, University of Pretoria, Pretoria.

Mullen, M. P., Creevey, C. J., Berry, D. P., McCabe, M. S., Magee, D. A., Howard, D. J., et al. (2012). Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples. *BMC Genomics* 13:16. doi: 10.1186/1471-2164-13-16

Mwai, O., Hanotte, O., Kwon, Y. J., and Cho, S. (2015). African indigenous cattle: unique genetic resources in a rapidly changing world. *Asian Australas. J. Anim. Sci.* 28, 911–921. doi: 10.5713/ajas.15.0002R

Nakayama, K., Iwata, H., Kim, E.-Y., Tashiro, K., and Tanabe, S. (2006). Gene expression profiling in common cormorant liver with an oligo array: assessing the potential toxic effects of environmental contaminants. *Environ. Sci. Technol.* 40, 1076–1083. doi: 10.1021/es051386m

Nazokkarmaher, M. (2016). *The Effect of Inbreeding on Holstein-Friesian Breed*. Logan, UT: Utah State University.

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986

Oltenacu, P. A., and Broom, D. M. (2010). The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Anim. Welf.* 19, 39–49.

Park, S. Y., Lee, J. H., Ha, M., Nam, J. W., and Kim, V. N. (2009). miR-29 miRNAs activate p53 by targeting p85α and CDC42. *Nat. Struct. Mol. Biol.* 16, 23–29. doi: 10.1038/nsmb.1533

Pienaar, L. (2014). *Genetic Diversity in the Afrikaner Cattle Breed*. MSc thesis, University of Freestate, Bloemfontein.

Pienaar, L., Grobler, J. P., Neser, F. W. C., Scholtz, M. M., Swart, H., Ehlers, K., et al. (2014). Genetic diversity in selected stud and commercial herds of the Afrikaner cattle breed. *S. Afr. J. Anim. Sci.* 44, 1–5.

Pool, J. E., Hellmann, I., Jensen, J. D., and Nielsen, R. (2010). Population genetic inference from genomic sequence variation. *Genome Res.* 20, 291–300. doi: 10.1101/gr.079509.108

Pool, R., and Waddell, K. (2002). *Exploring Horizons for Domestic Animal Genomics*. Washington, DC: National Academic Press.

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., et al. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genetics* 10:e1004148. doi: 10.1371/journal.pgen.1004148

Qi, X.-B. (2004). *Genetic Diversity, Differentiation and Relationship of Domestic Yak Populations - a Microsatellite and Mitochondrial DNA Study*. Ph.D. thesis, Lanzhou University, Lanzhou.

Ramey, H. R., Decker, J. E., McKay, S. D., Rolf, M. M., Schnabel, R. D., and Taylor, J. F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14:382. doi: 10.1186/1471-2164-14-382

Ramirez-Gonzalez, R., Bonnal, R. J., Caccamo, M., and MacLean, D. (2012). Biosamtools: Ruby bindings for Samtools, a library for accessing bam files containing high-throughput sequence alignments. *Source Code Biol. Med.* 7:6. doi: 10.1186/1751-0473-7-6

Ramos, A. M., Crooijmans, R. P., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., et al. (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One.* 4:e6524. doi: 10.1371/journal.pone.0006524

Ramos, A. M., Megens, H. J., Crooijmans, R. P. M. A., Schook, L. B., and 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, 613–620. doi: 10.1111/j.1365-2052.2011.02198.x

Rovelet-Lecrux, A., Legallic, S., Wallon, D., Flaman, J. M., Martinaud, O., Bombois, S., et al. (2012). A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. *Eur. J. Hum. Genet.* 20, 613–617. doi: 10.1038/ejhg.2011.225

Rubin, C. J., Megens, H. J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci.* 109, 19529–19536. doi: 10.1073/pnas.1217149109

Sambrook, J., and Russell, D. W. (2006). Amplification of cDNA generated by reverse transcription of mRNA. *CSH Protoc.* 2006:pdb.prot3837. doi: 10.1101/pdb.prot3837

Sanarana, Y., Visser, C., Bosman, L., Nephawe, K., Maiwashe, A., and van Marle-Köster, E. (2016). Genetic diversity in South African Nguni cattle ecotypes based on microsatellite markers. *Trop. Anim. Health Prod.* 48, 379–385. doi: 10.1007/s11250-015-0962-9

Schaller, C. E., Wang, C. L., Beck-Engeser, G., Goss, L., Scott, H. S., Anderson, M. S., et al. (2008). Expression of Aire and the early wave of apoptosis in spermatogenesis. *J. Immunol.* 180, 1338–1343. doi: 10.4049/jimmunol.180.3.1338

Scholtz, M. M. (2010). *Beef Breeding in South Africa*, 2nd Edn. Pretoria: Asikhulume pixArt.

Scott, K., Gadomski, T., Kozicz, T., and Morava, E. (2014). Congenital disorders of glycosylation: new defects and still counting. *J. Inherit. Metab. Dis.* 37, 609–617. doi: 10.1007/s10545-014-9720-9

Slikas, B., Jones, I. B., Derrickson, S. R., and Fleischer, R. C. (2000). Phylogenetic relationships of Micronesian white-eyes based on mitochondrial sequence data. *Auk* 117, 355–365. doi: 10.1642/0004-8038(2000)117[0355:PROMWE]2.0.CO;2

Soares, A. R., Reverendo, M., Pereira, P. M., Nivelles, O., Pendeville, H., Bezerra, A. R., et al. (2012). Dre-miR-2188 targets Nrp2a and mediates proper intersegmental vessel development in zebrafish embryos. *PLoS One* 7:e39417. doi: 10.1371/journal.pone.0039417

Steingrimsson, E., Copeland, N. G., and Jenkins, N. A. (2006). Mouse coat color mutations: from fancy mice to functional genomics. *Dev. Dynam.* 235, 2401–2411. doi: 10.1002/dvdy.20840

Stenson, P. D., Ball, E. V., Howells, K., Phillips, A. D., Mort, M., and Cooper, D. N. (2009). The human gene mutation database: providing a comprehensive central mutation database for molecular diagnostics and personalised genomics. *Hum. Genomics* 4, 69–72. doi: 10.1186/1479-7364-4-2-69

Stothard, P., Choi, J. W., Basu, U., Sumner-Thomson, J. M., Meng, Y., Liao, X., et al. (2011). Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics* 12:559. doi: 10.1186/1471-2164-12-559

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. doi: 10.1016/j.meatsci.2008.04.017

Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *biorXiv* 005165. doi: 10.1101/005165

Tuysuz, B., Bilguvar, K., Koçer, N., Yalçinkaya, C., Çaglayan, O., Gul, E., et al. (2014). Autosomal recessive spastic tetraplegia caused by AP4M1 and AP4B1 gene mutation: expansion of the facial and neuroimaging features. *Am. J. Med. Genet.* 164, 1677–1685. doi: 10.1002/ajmg.a.36514

Vainio, S., Heikkila, M., Kispert, A., Chin, N., and McMahon, A. P. (1999). Female development in mammals is regulated by Wnt-4 signaling. *Nature* 397, 405–409. doi: 10.1038/17068

Van, K., Kang, Y. J., Han, K. S., Lee, Y. H., Gwag, J. G., Moon, J. K., et al. (2013). Genome-wide SNP discovery in mungbean by Illumina HiSeq. *Theor. Appl. Genet.* 126, 2017–2027. doi: 10.1007/s00122-013-2114-9

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinf.* 43, 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43

Van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., et al. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Meth.* 5, 247–252. doi: 10.1038/nmeth.1185

Vetter, C. S., Groh, V., thor Straten, P., Spies, T., Brocker, E. B., and Becker, J. C. (2002). Expression of stress-induced MHC class I related chain molecules on human melanoma. *J. Invest. Dermatol.* 118, 600–605. doi: 10.1046/j.1523-1747.2002.01700.x

Viscomi, M. T., Oddi, S., Latini, L., Pasquariello, N., Florenzano, F., Bernardi, G., et al. (2009). Selective CB2 receptor agonism protects central neurons from remote axotomy-induced apoptosis through the PI3K/Akt pathway. *J. Neurosci.* 29, 4564–4570. doi: 10.1523/JNEUROSCI.0786-09.2009

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72. doi: 10.1371/journal.pbio.0040072

Wallace, R. M., Pohler, K. G., Smith, M. F., and Green, J. A. (2015). Placental PAGs: gene origins, expression patterns, and use as markers of pregnancy. *Reproduction* 149, R115–R126. doi: 10.1530/REP-14-0485

Wang, J., Raskin, L., Samuels, D. C., Shyr, Y., and Guo, Y. (2014). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics* 31, 318–323. doi: 10.1093/bioinformatics/btu668

Wang, M. D., Dzama, K., Hefer, C. A., and Muchadeyi, F. C. (2015). Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC Genomics* 16:894. doi: 10.1186/s12864-015-2122-z

Wang, W., Wei, Z., Lam, T. W., and Wang, J. (2011). Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Sci. Rep.* 55, 1–7. doi: 10.1038/srep00055

Weiher, H., Noda, T., Gray, D. A., Sharpe, A. H., and Jaenisch, R. (1990). Transgenic mouse model of kidney disease: insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. *Cell* 62, 425–434. doi: 10.1016/0092-8674(90)90008-3

Wilkinson, R. N., Jopling, C., and Van Eeden, F. J. (2014). Zebrafish as a model of cardiac disease. *Prog. Mol. Biol. Transl. Sci.* 124, 65–91. doi: 10.1016/B978-0-12-386930-2.00004-5

Wu, X., Steet, R. A., Bohorov, O., Bakker, J., Newell, J., Krieger, M., et al. (2004). Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. *Nat. Med.* 10, 518–523. doi: 10.1038/nm1041

Zhang, L., Kang, L., Bond, W., and Zhang, N. (2009). Interaction between syntaxin 8 and HECTd3, a HECT domain ligase. *Cell Mol. Neurobiol.* 29, 115–121. doi: 10.1007/s10571-008-9303-0

Zheng, J., Miller, K. K., Yang, T., Hildebrand, M. S., Shearer, A. E., DeLuca, A. P., et al. (2011). Carcinoembryonic antigen-related cell adhesion molecule 16 interacts with α-tectorin and is mutated in autosomal dominant hearing loss (DFNA4). *Proc. Natl. Acad. Sci.* 108, 4218–4223. doi: 10.1073/pnas.1005842108

Zwane, A. A., Maiwashe, A., Makgahlela, M. L., Choudhury, A., Taylor, J. F., and 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.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.