South African Journal of Animal Science 2021, 51 (No. 6)

Gene-set enrichment analysis of selective sweeps reveals phenotypic traits in Nguni cattle

A.A. Zwane[1#], K.S. Nxumalo[1], M.L. Makgahlela[1,3], E. Van Marle-Koster[2] & N. Maiwashe[1,3]

[1] Animal Breeding and Genetics, Agricultural Research Council-Animal Production, Private Bag X2, Irene, 0062, South Africa

[2]Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, South Africa

[3]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa

(Submitted 28 July 2021; Accepted 24 October 2021; Published 26 December 2021)

Copyright resides with the authors in terms of the Creative Commons Attribution 4.0 South African Licence.
See: http://creativecommons.org/licenses/by/4.0/za
Condition of use: The user may copy, distribute, transmit and adapt the work, but must recognise the authors and the South African Journal of Animal Science.

Abstract

Adaptation of animals to different environments is typically associated with structural and functional genomic variations. High throughput SNP genotyping and next-generation sequencing (NGS) have made it possible to study positive selection footprints and adaptation traits. Nguni is a small frame-size breed, mostly horned, and well known for being adapted to diverse South African environmental conditions. This study used previously identified selective sweeps to perform functional analysis of genes related to phenotypic characteristics in Nguni. Two hundred and sixty-four candidate selective sweeps were used for gene-set enrichment analysis in molecular functional categories (KEGG pathways) using the database for annotation, visualization, and integrated discovery (DAVID). In total, 107 genes were identified across all the chromosomes with 74 genes associated with eight phenotype queries, including fat content, milk production, walking ability, heat tolerance, meat production, reproduction, and bone and muscle development. Gene *CRHR*$_2$ was associated with meat quality (juiciness and flavour). The *IRAK3* gene was associated with decreased body size, feed intake and fatness in cattle, and *CARD15* with disease resistance. Gene annotation using phenotype queries identified four genes (*SPI*, *YWHAZ*, *RGS4,* and *RGS5*) that were associated with myometrial relaxation in cattle. Genes such as *NOD2* and *IL21R* were associated with inflammatory bowel diseases in cattle, whereas *CPLS* gene was associated with fat content. These genes are important to the phenotypic and adaptive characteristics present in South African Nguni cattle and hold potential for selection for traits of economic importance.

Keywords: annotation, genes, phenotypes, selective sweeps

[#]Corresponding author: Zwanea@arc.agric.za

Introduction

South Africa is the home of five indigenous cattle breeds, namely Afrikaner, Bonsmara, Drakensberger, Nguni, and Tuli. These breeds have played an important role in the history of the country as well as socio-cultural roles in many African societies (Ramsey *et al*., 2000; Mwai *et al*., 2015; Mapiye *et al*., 2019). They survive under harsh local conditions and have lower susceptibility to disease and parasites. In a world threatened by climate change, indigenous/local breeds are sources of irreplaceable genetic materials. They can survive unfavourable conditions posed by drought, extreme heat, and tropical diseases (Hoffmann, 2013). For example, Nguni is well known for its adaptive features, hardiness, and excellent resistance to internal and external parasites with natural immunity to tick-borne diseases (Marufu *et al*., 2013; Mapholi *et al*., 2015). It is a small-framed breed compared with most of the other large-framed beef breeds in South Africa, mostly horned, with a variety of vibrant red and black colour patterns, which make the breed unique and easily distinguishable from other breeds (Bester *et al*., 2003, Tada et al., 2013). Furthermore, Nguni is a multi-purpose breed used for meat, milk, leather, and socio-cultural functions such as dowries, communal feasts, and religious sacrifices (Bettencourt *et al*., 2015).

Natural and artificial selection played a major role in the phenotypic diversity of domesticated animals, which resulted in the formation of diverse livestock breeds adapted to a wide variety of environments and with special characteristics. The phenotypic variation in these animals has aided mapping of causal genes

URL: http://www.sasas.co.za
ISSN 0375-1589 (print), ISSN 2221-4062 (online)
Publisher: South African Society for Animal Science
http://dx.doi.org/10.4314/sajas.v51i6.9

and the mutations underlying a variety of traits (Casas & Kehrli, 2016; Müller *et al*., 2017; Jivanji *et al*., 2019). The genotype–phenotype map can increase the understanding of how changes in the genome can bring about alterations in both quantitative and discrete traits (Wright, 2015). Phenotypes that have been linked to domestication include meat and milk production, fertility, coat colour, decreased fearfulness, social motivation, and mild temperament (Qanbari *et al*., 2014). Selection that affected these phenotypes has left detectable signatures/sweeps within the genome of modern cattle (Bovine HapMap Consortium, 2009; Qanbari *et al*., 2014). These sweeps are beneficial mutations that arise and rapidly increase in frequency in the population, and can be detected as reduced local variability, deviations in the site frequency spectrum (SFS), increased linkage disequilibrium and extended haplotype structure (Stella *et al*., 2010; Carneiro *et al*., 2011; Qanbari *et al*., 2014). As a result, these sweeps can be used to scan for genes involved in recent adaptation.

Identification of selection sweeps is currently one of the areas of interest in evolutionary genetics since it can provide insights into the evolutionary processes involved in shaping the genomes of animals as well as functional information about genes and genomic regions (Gouveia *et al*., 2014). Recent genomic technologies have made it possible to study the genomes of large animals such as cattle to detect selection footprints and understand the genetic architecture of local adaptation, phenotypic variability, and gene function (Ramey *et al*., 2013; Huber *et al*., 2014; Naval-Sanchez, 2020). Although techniques for variant detection are now becoming routine, the key question remains for the functions of detected variants, genes, gene products, as well as their interactions and regulation in the genome (Gasperskaja & Kueinskas, 2017).

Growing evidence suggests that livestock breeding could benefit from a deep understanding of the regulatory mechanisms between the genotype and phenotype relationships. The location of these variants can result in phenotype differences, including the level of DNA methylation, transcription factor binding sites, alternative splicing, and protein translation and modification, since the function of variants depends on their location (Zhao *et al*., 2020). More novel functional variants are being confirmed with bioinformatics analysis, but most of them lack underlying molecular mechanisms (Crouch & Bodmer, 2020). Gene ontology captures statements of how genes function, where in the cell they function, and what biological processes dictate the outcome (Thomas, 2017). Gene-set enrichment and pathway-based analyses have been used to investigate the polygenic background of complex traits, such as leucosis, bull fertility, and meat quality (Abdalla *et al*., 2016). These enrichment analyses direct the focus from a single gene to a group-based analysis. Findings of these analyses can contribute to an improved understanding of the genetic and biological architecture of complex traits (Visscher *et al*., 2017).

Few studies have identified genomic regions and variants in selective sweeps and copy number variant (CNV) regions in South African indigenous cattle using bovine SNP array (Makina *et al*., 2015; Wang *et al*., 2015) and whole genome sequencing (Zwane *et al*., 2019). These studies provide the foundation for understanding genetic variation between the indigenous breeds and identifying genes of economic importance. The occurrence of selection has changed patterns of variation such that each form of selection causes specific changes in selected loci and in neutral loci linked to them (Kreitman, 2000; Gouveia *et al*., 2014). Locating these changes may enhance our understanding of the variations associated with adaptation in local breeds. The aim of this study was to perform the functional enrichment analysis to identify genes pertaining phenotypic variation in South African Nguni cattle using selective sweeps identified by Zwane *et al*. (2019).

Materials and Methods

Selective sweeps for Nguni cattle analysed in this study were identified in previous research (Zwane *et al*., 2019). In the study, 30 animals from Nguni were sequenced at 30x coverage to discover variants using Illumina HiSeq 2500 (Animal ethics approval EC: S4285-15 of University of Pretoria). Selective sweep regions were identified using the Z-transformed pooled heterozygosity (ZHp) scores computed from a 50% overlapping sliding window approach with 150 kb windows of SNP distribution. A total of 264 candidate selective sweep regions (ZHp Z-scores ≤ −4) were confirmed and used for gene enrichment analysis. Animal quantitative trait loci database (QTLdb) was used to retrieve and visualize the QTL information within the selective sweep regions as it provides tools for aligning various genome features to QTLs and enables comparison of QTL results within species and across experiments (Hu *et al*., 2013).

A database for annotation, visualisation, and integrated discovery (DAVID) v6.8 tool (Huang *et al*., 2008; Maiorano *et al*., 2018) was used to identify significant ($P < 0.05$) gene ontology (GO) terms and KEGG (Kyoto encyclopedia of genes and genomes) pathways using a gene list with significant SNPs based on ZHp Z-scores. The GO database defines biological descriptors to genes based on the features of their encoded products, and is partitioned into three components, that is, biological process, molecular function, and cellular component. The KEGG pathway database contains metabolic and regulatory pathways, representing

the actual knowledge on molecular interactions and reaction networks (Azizpour *et al*., 2020). The gene ID list was converted into gene names and symbols using DAVID for annotation.

The genes were examined for enrichment in molecular functional categories using KEGG pathways (a module in DAVID) with selected phenotype queries such as fat content, milk production, walking ability, heat tolerance, meat production, reproduction, and bone and muscle development. The queries were chosen as desirable traits of adaptation in Nguni cattle. The gene list with enriched functional annotations and their statistical *P*-values was obtained from DAVID. The *P*-values were calculated and adjusted to increase the statistical significance (from the default 0.025 to 0.05) using Benjamini–Hochberg method. A *P*-value of 0.05 was used as a criterion cutoff, whereby gene sets with *P*-values less than 0.05 were considered significantly enriched (Nguyen *et al*., 2019; Jafari & Ansari-Pour, 2019). All genes were associated to one or multiple process, function, and component annotations (Berardini *et al*., 2012). Nearly all of the genes were assigned to a GO term. The fold enrichment was used to select the top-listed gene GO terms associated with identified genes, with significance of $P < 0.05$. The false discovery rate (FDR) was used to determine significance of enrichment or overrepresentation of terms for each annotation (Lewin & Grieve, 2006).

Results and discussion

Bovine functional genomics tools have enhanced the discovery of genes underlying important phenotypic traits and their implementation in livestock selection programmes. This study used previously identified selective sweep regions to perform enrichment analysis to identify genes associated with phenotypes in South African Nguni cattle. Nguni possess desirable phenotypic and adaptive characteristics such as tolerance to heat, drought, diseases, and parasites (Mapholi *et al*., 2017; Mapiye *et al*., 2019), ability to reproduce under low input systems (Scholtz & Theunissen, 2010), grazing ability, and good milk and beef quality (Marufu *et al*., 2011). These characteristics require understanding of the genes involved, which may be important in sustainable production of Nguni and other local breeds in changing environments in view of climate change.

Figure 1A indicates the 264 selective sweep regions used in this study, and the heterozygosity levels in all chromosomes are displayed in Figure 1B. The regions with the most significant selective sweeps were identified in chromosome 8 and 19. Genomic regions subjected to high selective pressures can display signatures such as reduced nucleotide diversity, stretches of homozygous loci, shifted site frequency spectrum and reduced recombination rate, and a number of selective sweeps can be identified in these regions (Aslam *et al*., 2014; Purfield *et al*., 2017). The lowest heterozygosity in the selective sweep regions was observed in chromosome 8, 19, and 27 (Figure 1B). Heterozygosity shows the genetic variation in a population. However, studies indicated that a hard/soft sweep would leave a footprint of reduced heterozygosity in the area of the genome subjected to selection (Oleksyk *et al*., 2008; Burke, 2012; Purfield *et al*., 2017). Therefore, identification of selective sweeps in regions of low/reduced heterozygosity could aid the discovery of novel genes subjected to selection.

In this study, 13 selective sweeps with low heterozygosity were identified in chromosomes 2, 4, 5, 8, 17, 19, and 27 (Figure 1B). These regions had the most significant selective sweeps (ZHp Z-scores ≤ -7), showing the presence of possible strong/hard sweeps. Supplementary Table 1 shows the most significant selective sweep regions (ZHp Z-scores ≤ -6) identified in Nguni. Forty-seven (47) putative selective sweeps were identified. However, some regions were unknown with little or no QTL information.

Druet *et al*. (2013) used a hidden Markov model to identify selective sweeps with long stretches of reduced heterozygosity in cattle. They identified genes associated with coat colour and horn development in these regions. Analysis of pooled genome sequences from Djallonke and Sahelian sheep in Ghana revealed co-localization of regions of reduced heterozygosity with candidate genes for disease resistance and adaptation to a tropical environment (Yaro *et al*., 2019). Strong selective pressure can lead to an increase in the frequencies of favourable alleles, leading to a decreased genetic variation in the genes under selection. Owing to the physical association of the locus with nearby loci (genetic linkage), genetic variants located in the loci may also suffer a decrease in genetic variability (Aramburu *et al*., 2020). Soft selective sweeps (where more than one allele is targeted by selection) are much harder to detect owing to their resistance to heterozygosity decrease (Sun *et al*., 2014; Aramburu *et al.*, 2020). Strong selective sweeps are easier to detect since they are determined by the increase in the frequency of a specific allele (in some cases leading to complete fixation). These are therefore useful for detection of candidate genes responsible for important traits in response to selection (Sun *et al*., 2014; Aramburu *et al*., 2020).

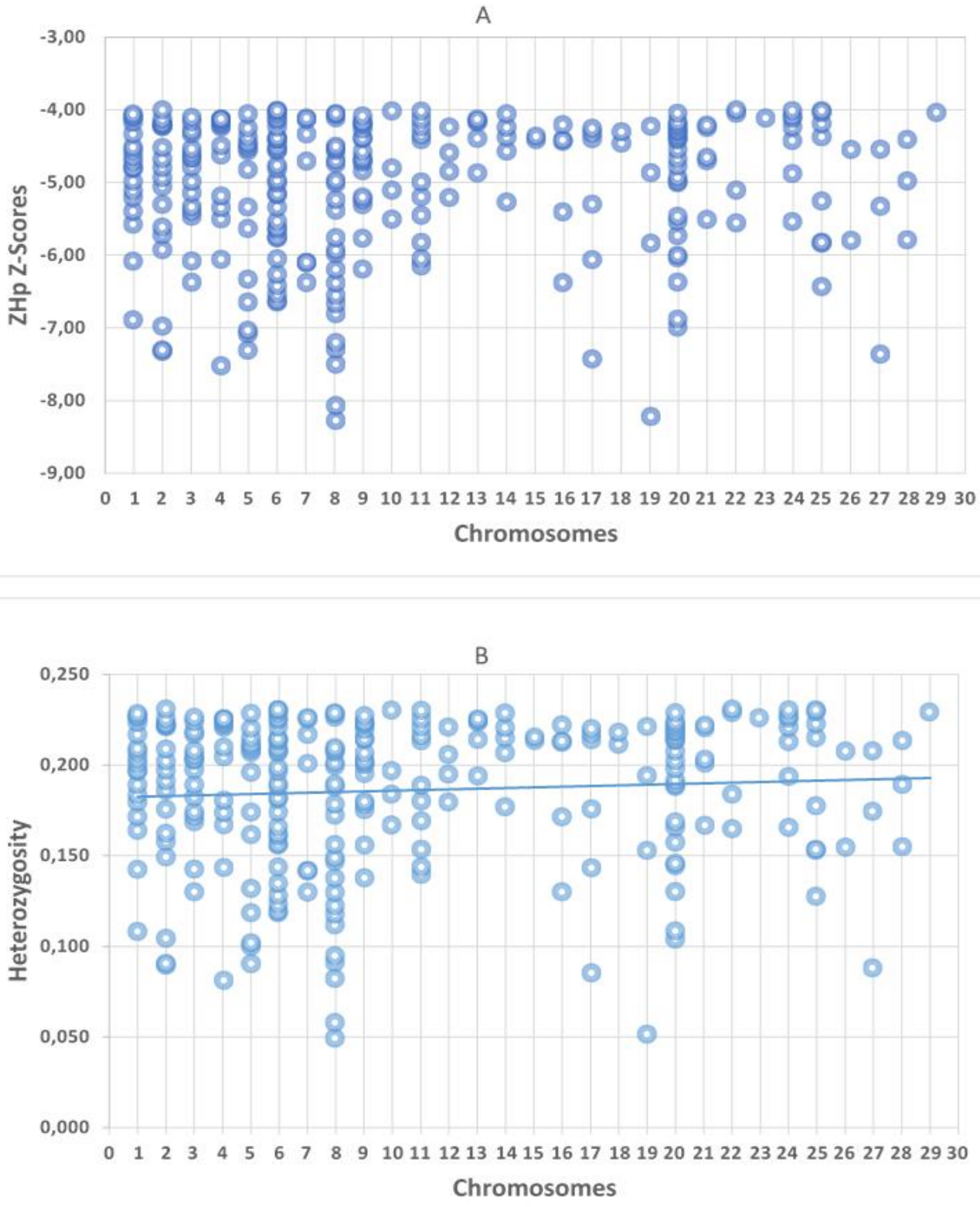

Figure 1 A Previously identified candidate selective sweep regions with ZHp Z-Scores ≤ -4 identified in South African Nguni cattle, **B** Heterozygosity distribution of selective sweep regions in all chromosomes as identified in Nguni cattle

Gene annotation of 264 selective sweeps identified 107 genes located in 29 chromosomes. These genes were previously associated with various functions in cattle, human and mice. Most of the gene associations were conducted in model animals such as mice, humans, and zebra fish rather than in cattle. However, the gene functions are mostly the same because of DNA similarities between humans and cattle (Zimin *et al*., 2009, Costilla *et al*., 2020). Previous studies discovered more than 200 homologous blocks of DNA between the two species (human and cattle), which could also enable the mapping of human to cattle sequences (Everts-van der Wind *et al*., 2005; Zimin *et al*., 2009; Larkin, 2011).

Table 1 shows the top 10 significant genes (P <0.05) identified using DAVID. Genes such as corticotropin-releasing hormone receptor 2 (*CRHR*$_2$), nucleotide binding oligomerization domain containing 2 (*NOD*$_2$), tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta (*YWHAZ*), interleukin 1 receptor associated kinase 3 (*IRAK3*) and interleukin 21 receptor (*IL21R*) were high-ranking genes identified in this study. These are protein-coding genes most viable in human, mouse, cattle, and dogs (Luisi *et al*., 2015). The *CRHR*$_2$ gene in chromosome 4 was associated with insulin sensitivity in this analysis. This gene has been associated with juiciness and flavour in cattle (Jiang *et al*., 2009). Together with *CRHR*$_1$, the *CRHR*$_2$ gene is expressed in the bovine adrenal gland and plays a role in the thyroid physiological function through autocrine/paracrine mechanisms (Squillacioti *et al*., 2012). The presence of this gene in Nguni cattle suggest the same regulatory pattern based on urocortin-like immunoreactivity (UCN-IR) and CRHR$_2$–IR that is found in the thyroid follicular and parafollicular cells (Squillacioti *et al*., 2012). In pigs, *CRHR*$_2$ was found in skeletal smooth and cardiac muscles, and in the brain (Choi *et al*., 2017). This gene has been associated with meat quality on Duroc x Pietrain populations (Casiro *et al*., 2017).

The *NOD*$_2$ gene, also known as caspase recruitment domain-containing protein 15 (*CARD15*), is a pattern recognition receptor for bacterial lipopolysaccharides, which release inflammatory mediators that recognize the muramyl dipeptide and activate NF-kb, which controls transcription of tumour necrosis factor-α, interferon-γ, interleukin-1β and interleukin-12 (Wang *et al*., 2015). The *CARD15* protein contributes to immune response of cells and plays an important role in antibody production.

CARD15 gene variations are associated with diseases such as bovine tuberculosis, Crohn's disease (inflammatory bowel disease), and paratuberculosis (Pinedo *et al*., 2009; Küpper *et al*., 2014; Wang *et al*., 2015). The gene's relationship to diseases and its conservation between species suggest that it may have a conserved role in bovine disease resistance. Comparative analysis of the *CARD15* gene between the bovine, mouse and human revealed a high level of conservation of this gene in sequence, genomic structure and protein domains between species (Taylor *et al.*, 2006).

The *CARD15* gene polymorphisms were associated with susceptibility to tuberculosis in Chinese Holstein cattle and haplotype markers such as TGGACA and CAGACA were identified (Wang *et al*., 2015). This showed that genetic markers for this gene can be identified in local breeds and used in breeding animals with high resistance to bovine tuberculosis, as was demonstrated in the Chinese Holstein cattle (Wang *et al*., 2015).

Parasitic diseases are major challenges to cattle production in South Africa (Katikati & Fourie, 2019). With no effective vaccination programmes available against most parasitic diseases, significant morbidity and mortality in the developing countries remain limitations to effective sustainable prediction. Based on human studies, the *IL21R* gene is highly expressed in parasitized organs of infected individuals and in murine models of parasitic diseases (Solaymani-Mohammadi *et al*., 2019). Nguni cattle are known to be less susceptible to parasite diseases that affect livestock in South Africa, such as ticks and tick-borne diseases (Mapholi *et al*., 2017). The presence of *IL21R* gene in Nguni could assist in the enhancement of their innate immunity. *IL21R* is expressed at high levels on intestinal epithelial cells and stomal fibroblasts in inflammatory bowel diseases. *IL21* induces macrophage inflammatory protein-3 alpha (MIP-3alpha), a T-cell chemoattractant in epithelial cells, and has been signalling between epithelial and immune cells in the gut (Caruso *et al*., 2007).

The *IRAK3* and specific protein 1 (*SP1)* genes have been associated with decreased body size in this analysis, as shown in Table 1. *IRAK3* is a cytoplasmic homeostatic mediator of inflammatory responses, which is potentially useful as a prognostic marker in inflammation (Freihat *et al*., 2019). It inhibits signalling cascades downstream of myddosome complexes associated with toll-like receptors and contains a death domain that interacts with other *IRAK* family members involved with tumour necrosis factor receptor associated factor 6 (*TRAF6*) (Freihat *et al*., 2019). The *IRAK3* was identified within large selective sweep spanning 430 kb on *Bos taurus* chromosome 5 (BTA5) (Naval-Sanchez *et al*., 2020), the largest region under selection, which also displayed the largest difference between indicine and taurine cattle.

Table 1 List of top 10 genes identified within genomic location (strand) and their phenotypic annotation.

Gene name	Genomic location	Gene Name	Phenotype	Selected Orthologues	Associated trait in cattle
CRHR2	4:66002072-66006203	Corticotropin releasing hormone receptor 2	Increased insulin sensitivity	*Mus musculus*,*Gallus gallus,Sus scrofa*	Reproduction, production
NOD2	18:19157447-19166134	Nucleotide binding oligomerization domain containing 2	Decreased tumour Necrosis factor Secretion	*Equus caballus,Sus scrofa, Pan troglodytes*	Meat and carcass, health
YWHAZ	14:65584487-65617329	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta	Abnormal skin morphology	*Macaca mulatta, Equus caballus, Sus scrofa*	Exterior, milk
IRAK3	5:47796256-47839675	Interleukin 1 receptor Associated kinase 3	Decreased body size	*Canis familiaris, Felis catus, Equus caballus*	Meat and carcass, milk, exterior
IL21R	5:25233186-5268874	Interleukin 21 receptor	Decreased T cell proliferation	*Camelus dromedarius, Catagonus wagneri, Balaenoptera musculus*	Milk, health
IMPDH1	4:66002072-66006203	Inosine monophosphate dehydrogenase 1	Developmental growth, increased rate	*Bison bison bison, Urocitellus parryii,*	Reproduction, production
RGS4	3:6287012-6292402	Regulator of G protein signalling 4	Decreased systemic arterial diastolic blood pressure	*Struthio camelus australis, Mus spretus, Bison bison*	Milk, meat and carcass
PRKAA1	20:33688291-33716677	Protein kinase AMP-activated catalytic subunit alpha 1	Slow postnatal weight gain	*Struthio camelus australis, Vicugna pacos, Bison bison*	Meat, milk, exterior
RGS5	3:6228349-6426528	Regulator of G protein signalling 5	Decreased total body fat amount	*Mus spretus,Marmota marmota marmot, Camelus dromedarius*	Milk, meat and carcass
SP1	5:26753574-26759487	Sp1 transcription factor	Decreased body size	*Mus musculus, Cavia porcellus*, *Microcebus murinus*	Milk, exterior

The *SP1* transcriptional factor has a putative binding site in cattle and has shown numerous polymorphisms in the promoter region of the bovine leptin gene. Polymorphisms in *SP1* gene have been associated with production traits such as feed intake and fatness in cattle (Adamowicz *et al*., 2006). The presence of these genes in Nguni could be associated with their thin covering of fat compared with other South African beef breeds (Chulayo & Muchenje, 2016). Nguni can fatten to high body mass on natural grazing, producing good quality carcasses, an even distribution of fat, and excellent marbling (Maciel *et al*., 2013; Katikati, 2017). The regulator of G-protein signalling 5 (*RGS5*) gene was associated with decreased body fat, which is an observed characteristic in Nguni cattle (Tada *et al*., 2013). These genes have been identified in various studies in cattle and have biological functions related to milk production, walking ability, bone and muscle development, heat tolerance, reproduction, and fat (Zimin *et al*., 2009; Underwood *et al*., 2015; Weldenegodguad *et al*., 2019; Leal-Gutiérrez *et al.,* 2019*;* Rebl *et al*., 2019; Kopke *et al.,* 2020).

Enrichment analysis of genes was performed using phenotype queries such as milk production, walking ability, bone and muscle development, heat tolerance, reproduction and fat. Figure 2 shows the top 10 genes ($P < 0.05$) that were associated with the phenotype queries as identified in Nguni. Genes were mostly associated with walking ability, fat digestion, bone and muscle development, reproduction and heat tolerance. Some of these genes were identified as highly significant ($P >0.025$) in Table 1, showing their significant molecular functions. These genes were previously associated with phenotypes in *Bos taurus* (Weldenegodguad *et al*., 2019) and species such as mice, chicken and other cattle breeds. Genes such as *NOD2* and *IL21R* were associated with inflammatory bowel disease (IBD), which has been reported to be common in all ruminants, and pasteurized milk could be the potential source of IBD in humans and in dairy cows. It is characterized by numerous acid-fast *Mycobacterium avium paratuberculosis* (MAP) (McNees *et al*., 2016), a real threat to both cattle and human beings, and can survive in contaminated pastures for a year, remain viable after milk pasteurization, and resists the ripening process of cheese (Fawzy *et al*., 2013).

The colipase (*CLPS*) gene identified in this study was associated with fat digestion in Nguni cattle. It consists of enterostatin, which has a biological activity as a satiety signal, and is known to suppress intake of a high fat diet in rat (Lin *et al*., 2007). In pigs, the *CLPS* gene was located in chromosome 7 and considered a candidate gene for QTL, which affects carcass fatness (Jankowiak *et al*., 2008).

Figure 2A illustrated that four genes were associated with myometrial relaxation and contraction pathways in Nguni cattle, namely *SP1*, *YWHAZ*, *RGS4,* and *RGS5*. These genes were associated with fat digestion, bone and muscle development, reproduction, and walking ability in this study, which confirmed earlier results. This was in line with the observable phenotypes in Nguni cattle such as walking ability (Matjuda, 2012). The *RGS5* gene was previously associated with ruminal vascularity and lean growth (Kern *et al*., 2017). In humans, this gene is transcribed in myometrial muscle associated with contractile activity and muscle relaxation during pregnancy, and acts on the myometrium to regulate muscle contraction. This was in line with the studies by Gunaje *et al*. (2011) and Daniel *et al*. (2016), which showed that the regulators of G protein signalling genes (*RGS4* and *RGS4*) were critical mediators of vascular smooth muscle contraction and potentially arterial remodelling. The *RGS5* gene has also been associated with heat tolerance in cattle (Stronen *et al*., 2019). More information on gene molecular functions and biological processes is given in Supplementary Table 2.

In Table 2 the top GO terms ($P <0.05$) with molecular functions in Nguni cattle are summarized. The GO enrichment analysis revealed three molecular functions associated with kinase activity (GO:0004674~protein serine/threonine kinase activity, GO:0004672~protein kinase activity, and GO:0016301~kinase activity). The protein kinase activity has been reported to be negatively associated with intramuscular fat content in *longissimus dorsi* muscle of beef cattle (Underwood *et al*., 2008). This could explain the lacking of marbling in meat produced by Nguni cattle (Chulayo *et al*., 2016). Two GO molecular processes related to activatory and regulatory activity (GO:0005096~GTPase activator activity; GO:0008047~enzyme activator activity; GO:0030695~GTPase regulator activity; GO:0060589~nucleoside-triphosphatase regulator activity, respectively) were also identified.

#	Symbol	Description	Type	Matched Phenotypes	Matched Phenotypes Count	Direct / Indirect	Score	Average Disease Causing Likelihood
1	IMPDH1	Inosine Monophosphate Dehydrogenase 1	Protein	fat; bone; muscle; reproduction	4	Direct	29.36	67.9
2	NOD2	Nucleotide Binding Oligomerization Domain Containing 2	Protein	fat; bone; muscle	3	Direct	18.79	18.3
3	YWHAZ	Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta	Protein	fat; bone; muscle; walking; reproduction; heat	6	Indirect	10.75	80.3
4	IRAK3	Interleukin 1 Receptor Associated Kinase 3	Protein	fat; bone; muscle; walking; reproduction	5	Indirect	10.49	29.5
5	IL21R	Interleukin 21 Receptor	Protein	fat; bone; muscle	3	Direct	8.17	48.0
6	CRHR2	Corticotropin Releasing Hormone Receptor 2	Protein	fat; milk; bone; muscle; walking; reproduction	6	Indirect	7.65	25.0
7	RGS4	Regulator Of G Protein Signaling 4	Protein	fat; milk; bone; muscle; reproduction; walking	6	Indirect	7.17	60.0
8	PRKAA1	Protein Kinase AMP-Activated Catalytic Subunit Alpha 1	Protein	muscle	1	Direct	2.57	83.4
9	RGS5	Regulator Of G Protein Signaling 5	Protein	muscle	1	Direct	1.81	59.3
10	SP1	Sp1 Transcription Factor	Protein	bone	1	Direct	0.78	83.3

Figure 2 Top 10 genes that were highly associated with cattle phenotypes in Nguni based on their enrichment scores

Table 2 Top gene ontology terms with molecular functions ranked by fold enrichment

Term	Gene count	Adjusted *P*-value	Fold enrichment
GO:0004674~protein serine/threonine kinase activity	7	6.1E-3	4.137
GO:0004672~protein kinase activity	8	1.3E-2	3.075
GO:0016773~phosphotransferase activity, alcohol group as acceptor	8	2.8E-2	2.630
GO:0016301~kinase activity	8	4.2E-2	2.423
GO:0005096~GTPase activator activity	4	5.0E-2	4.728
GO:0008047~enzyme activator activity	5	5.5E-2	3.419
GO:0030695~GTPase regulator activity	4	6.5E-2	4.238
GO:0060589~nucleoside-triphosphatase regulator activity	4	7.8E-2	3.933
GO:0005524~ATP binding	10	9.4E-2	1.784

The GTPase activator activity (GO:0005096) is a family of regulatory proteins that can bind to activate G-proteins and stimulate their essential GTPase activity. The G-protein is involved in important cellular processes and physiological functions such as senses of light, taste and smell, neurotransmission, metabolism, endocrine and exocrine secretion (Syrovatkina *et al*., 2016; Wang, 2018). The phosphotransferase activity orthology related term GO:0016773 (alcohol groupacceptor) was also enriched. This term has been enriched significantly for bovine digital dermatitism, which is a foot disease related to several pains and discomfort and other problems that are prevalent in cattle kept indoors and to lameness in dairy cows (Kopke *et al.,* 2020, Salano *et al*., 2016).

Table 2 also shows that many of the significantly enriched terms were related to gene kinase and protein kinase activity and transcription. The GO terms were significantly enriched for different genome functions in Nguni cattle. Among the highest ranking molecular functions was kinase activity, in which genes such as *IRAK3*, *PRKAA1*, *YWHAZ* and *RGS5* were identified. Here the importance of *IRAK3* gene in immune response to infections is emphasized (Rebl *et al*., 2019), which could explain Nguni cattle being mostly adapted and resitant to a wide range of diseases (Mapholi *et al*., 2017), and *PRKAA1* gene as a tumor supressor in mice (Vara-Ciruelos *et al*., 2019). McKay *et al*. (2018) reported conserved DNA mythylation levels and patterns in *PRKAA1* and *PRKAB1* genes (also known as 5' AMP-activated protein kinase (AMPK) genes) at various methylation sites in cattle breeds including bison, and highlighted the need to maintain the functioning of these genes as vital for biological functions. *YWHAZ* has been identified as the most stable gene in bovine neutrophils (Crookenden *et al*., 2017) and *RGS5* has been associated with feed efficiency in cattle (Kern *et al*., 2017). These genes are most significant in South African Nguni cattle.

Conclusion

Gene-set enrichment and functional annotation were used successfully to associate selective sweeps with phenotypic variation in Nguni cattle. Genes identified in this study were highly associated with molecular functions that may underlie phenotypic and adaptive characteristics in Nguni cattle. These include genes that were associated with body size, feed intake, meat quality, heat tolerance, and disease resistance. These results provide the basis for understanding genetic mechanisms involved in economic traits and adaptation to local conditions. Further studies are needed on signalling pathways and expression of these genes. Similar studies are needed for other indigenous breeds as efforts to characterize them genetically for improvement and enhanced food production in the era of climate change.

Acknowledgements

The authors acknowledge the Red Meat Research and Development of South Africa (RMRDSA) and the National Research Foundation (NRF) the financial support for this research, Dr Ananyo Choudhury for assistance with data analysis, and Prof Jeremy Taylor for his involvement in this work. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the RMRDSA and the NRF.

Author's Contributions

AAZ and KSN carried out the analysis and drafted the manuscript. MLM, EVM, and AM assisted with structuring and reviewing scientific content.

Conflict of Interest Declaration

The authors declare that they have no competing interests.

References

Abdalla, E.A., Penagaricano F., Byrem, T.M., Weigel, K.A. & Rosa, G.J.M., 2016. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. Anim. Genet. 47, 395-407. DOI: 10.1111/age.12438

Adamowicz, T., Flisikowski, K., Starzyński, R., Zwierzchowski, L., & Switonski, M., 2006. Mutation in the Sp1 motif of the bovine leptin gene affects its expression. Mamm. Genome 17, 77-82. DOI: 10.1007/s00335-005-0068-1

Aramburu, O., Ceballos, F., Casanova, A., Le Moan, A., Hemmer-Hansen, J., Bekkevold, D., Bouza, C. & Martínez, P., 2020. Genomic signatures after five generations of intensive selective breeding: Runs of homozygosity and genetic diversity in representative domestic and wild populations of turbot (Scophthalmus maximus). Front. Genet. 11, 1-14. DOI: 10.3389/fgene.2020.00296

Aslam, M.L., Bastiaansen, J.W.M., Megens, H.J., Crooijmans, R.P.M.A., Nasreen, F., Blomberg, L.A., Van Tassell, C.P., Sonstegard, T.S., Schroeder, S.G., Groenen, M.A.M. & Long, J.A., 2014. Genome-wide candidate regions for selective sweeps revealed through massive parallel sequencing of DNA across ten turkey populations. BMC Genet. 15, 1-10. DOI: 10.1186/s12863-014-0117-4

Azizpour, N., Farahani, K.A.H., Moradi, M.H. & Mohammadi, H., 2020. Genome-wide association study based on gene-set enrichment analysis associated with milk yield in Holstein cattle. Anim. Sci. Res. 30(1), 79-92. https://www.sid.ir/en/journal/ViewPaper.aspx?ID=790666

Berardini, T.Z., Li, D., Muller, R., Chetty, R., Ploetz, L., Singh, S., Wensel, A. & Huala, E., 2012. Assessment of community-submitted ontology annotations from a novel database-journal partnership. Database 2012, bas030. https://doi.org/10.1093/database/bas030

Bester, J., Matjuda, L.E., Rust, J.M. & Fourie, H.J., 2003. The Nguni: Case study in FAO community-based management of animal's genetics resources. UNDP, GTZ, CTA, FAO, Rome. Pp. 45-6. https://www.fao.org/3/y3970e/y3970e.pdf

Bettencourt, E.M.V., Tilman, M., Narciso, V., Carvalho, M.L.S. & Henriques, P.D.S., 2015. The role of livestock functions in the wellbeing and development of Timor-Leste rural communities. Livest. Res. Rural. Dev. 26, 1-19. https://core.ac.uk/download/pdf/62472131.pdf

Bovine HapMap Consortium, 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Sci. 324, 528-532. DOI: 10.1126/science.1167936

Burke, M.K., 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. Proc. Royal Society B-Biol. Sci. 279, 5029-5038. DOI: 10.1098/rspb.2012.0799

Carneiro, M., Afonso, S., Geraldes, A., Garreau, H. & Bolet, G., 2011. The genetic structure of domestic rabbits. J. Mol. Biol. Evol. 28, 1801-16. DOI: 10.1093/molbev/msr003.

Caruso, R., Fina, D., Peluso, I., Fantini, M.C., Tosti, C., Del Vecchio Blanco, G., Paoluzi, O.A., Caprioli, F., Andrei, F., Stolfi, C., Romano, M., Ricci, V., MacDonald, T.T., Pallone, F. & Monteleone, G., 2007. IL-21 is highly produced in Helicobacter pylori-infected gastric mucosa and promotes gelatinases synthesis. J. Immunol. 178, 5957-5965. DOI: 10.4049/jimmunol.178.9.5957

Casas, E. & Kehrli, M.E. Jr., 2016. A review of selected genes with known effects on performance and health of cattle. Front. Vet. Sci. 3, 1-11. DOI: 10.3389/fvets.2016.00113

Casiro, S., Velez-Irizarry, D., Ernst, C.W., Raney, N.E., Bates, R.O., Charles, M.G. & Steibel, J.P., 2017. Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits. J. Anim. Sci. 95, 545-558. DOI: 10.2527/jas.2016.1003

Choi, I., Bates, R.O., Raney, N.E. & Ernst, C.W., 2017. Association of a corticotropin-releasing hormone receptor 2 (*CRHR2*) polymorphism with carcass merit, meat quality, and stress response traits in pigs. Can. J. Anim. Sci. 97, 536-540. https://doi.org/10.1139/cjas-2016-0168

Chulayo, A.Y. & Muchenje, V., 2016. Effects of animal class and genotype on beef muscle nanostructure, pHu, colour and tenderness. S. Afr. J. Sci. 112, 1-9. https://doi.org/10.17159/sajs.2016/20150056

Costilla, R., Kemper, K.E., Byrne, E.M., Porto-Neto, L.R., Carvalheiro, R., Purfield, D.C., Doyle, J.L., Berry, D.P., Moore, S.S., Wray, N.R. & Hayes, B.J., 2020. Genetic control of temperament traits across species: association of autism spectrum disorder risk genes with cattle temperament. Genet. Sel. Evol. 52, 1-14. DOI: 10.1186/s12711-020-00569-

Crookenden, M.A., Walker, C.G., Kuhn-Sherlock, B., Murray, A., Dukkipati, V.S.R., Heiser, A. & Roche, J.R., 2017. Evaluation of endogenous control gene expression in bovine neutrophils by reverse-transcription quantitative PCR using microfluidics gene expression arrays. J. Dairy Sci. 100, 6763-6771. DOI: 10.3168/jds.2016-12460

Crouch, D.J.M. & Bodmer, W.F., 2020. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. Proc. Natl. Acad. Sci. USA 117(32),18924-18933. DOI: 10.1073/pnas.2005634117

Daniel, J.M., Prock, A., Dutzmann, J., Sonnenschein, K., Thum, T., Bauersachs, J. & Sedding, D.G., 2016. Regulator of G-protein signaling 5 prevents smooth muscle cell proliferation and attenuates neointima formation. Arterioscler. Thromb. Vasc. Biol. 36, 317-327. DOI: 10.1161/ATVBAHA.115.305974

Druet, T., Pérez-Pardal, L., Charlier, C. & Gautier, M., 2013. Identification of large selective sweeps associated with major genes in cattle. Anim. Genet. 44, 758-762. DOI: 10.1111/age.12073

Everts-van der Wind, A., Kata, S.R., Band, M.R., Rebeiz, M., Larkin, D.M., Everts, R.E., Green, C.A., Liu, L., Natarajan, S., Goldammer, T., Lee, J.H., McKay, S., Womack, J.E. & Lewin, H.A., 2004. A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. Genome Res. 14, 1424-1437. DOI: 10.1101/gr.2554404

Fawzy, A., Zschöck, M., Ewers, C., & Eisenberg, T., 2018. Genotyping methods and molecular epidemiology of *Mycobacterium avium* subsp. paratuberculosis (MAP). Int. J. Vet. Sci. Med. 6, 258-264. DOI: 10.1016/j.ijvsm.2018.08.001

Freihat, L.A., Wheeler, J.I., Wong, A., Turek, I., Manallack, D.T. & Irving, H.R., 2019. IRAK3 modulates downstream innate immune signalling through its guanylate cyclase activity. Sci. Rep. 9, 1-12. DOI: 10.1038/s41598-019-51913-3

Gasperskaja, E. & Kucinskas, V., 2017. The most common technologies and tools for functional genome analysis. Acta. Med. Litu. 24, 1-11. DOI: 10.6001/actamedica.v24i1.3457

Gouveia, J.J.D.S., Silva, M. V.G.B.D., Paiva, S.R. & Oliveira, S.M.P.D., 2014. Identification of selection signatures in livestock species. Genet. Mol. Biol. 37, 330-342. DOI: 10.1590/s1415-47572014000300004

Gunaje, J.J., Bahrami, A.J., Schwartz, S.M., Daum, G. & Mahoney Jr, W.M., 2011. PDGF-dependent regulation of regulator of G protein signaling-5 expression and vascular smooth muscle cell functionality. A. J. Physiol. Cell Physiol., 301, C478-C489.Am. J. Physiol. Cell Physiol. 301(2),C478-89. DOI: 10.1152/ajpcell.00348.2010

Hoffmann, I., 2013. Adaptation to climate change – exploring the potential of locally adapted breeds. Anim. 7, 346-362. DOI: 10.1017/S1751731113000815

Hu, Z., Park, C.A., Wu, X. & 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. & Lempicki, R.A., 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4, 44-57. DOI: 10.1038/nprot.2008.211

Huber, C.D., Nordborg, M., Hermisson, J. & Hellmann, I., 2014. Keeping it local: Evidence for positive selection in Swedish *Arabidopsis thaliana*. Mol. Biol. Evol. 31, 3026-3039. DOI: 10.1093/molbev/msu247

Jafari, M. & Ansari-Pour, N., 2019. Why, when and how to adjust your P-values? Cell J. 20, 604-607. DOI: 10.22074/cellj.2019.5992

Jankowiak, H., Kapelański, W., Wilkanowska, A., Jelińska, A. & Biegniewska, M., 2008. Slaughter value of pigs carcasses as related to colipase genotype (*CLPS*) considering *RYR1* gene effect. J. Cent. Eur. Agric. 9, 803-808.

Jiang, Z., Michal, J.J., Chen, J., Daniels, T.F., Kunej, T., Garcia, M.D., Gaskins, C.T., Busboom, J.R., Alexander, L.J., Wright, R.W., Jr. & MacNeil, M.D., 2009. Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. Int. J. Biol. Sci. 5, 528-542. DOI: 10.7150/ijbs.5.528

Jivanji, S., Worth, G., Lopdell, T.J. Yeare, A., Couldrey, C., Reynolds, E., Tiplady, K., McNaughton, L., Johnson, T.J.J., Davis, A.R., Harris, B., Spelman, R., Snell, R.G., Garrick, D. & Littlejohn, M.D., 2019. Genome-wide association analysis reveals QTL and candidate mutations involved in white spotting in cattle. Genet. Sel. Evol. 51, 1-18. DOI: 10.1186/s12711-019-0506-2

Katikati, A., 2017. Assessment of production practices of emerging cattle farmers in the selected districts of the Eastern Cape Province, South Africa. MSc thesis, Faculty of Health and Environmental Sciences, Central University of Technology, Free State, South Africa. Pp 1-143.

Katikati, A. & Fourie, P.J., 2019. Improving management practices of emerging cattle farmers in selected areas of the Eastern Cape Province: The role of agricultural extension. S. Afr. J. Agric. Ext. 47, 1-11. DOI:10.17159/2413-3221/2019/V47N1A492

Kern, R.J., Zarek, C.M., Lindholm-Perry, A.K., Kuehn, L.A., Snelling, W.M., Freetly, H.C., Cunningham, H.C. & Meyer, A.M., 2017. Ruminal expression of the *NQO* 1, *RGS* 5, and *ACAT* 1 genes may be indicators of feed efficiency in beef steers. Anim. Genet. 48, 90-92. DOI: 10.1111/age.12490

Kopke, G., Anklam, K., Kulow, M., Baker, L., Swalve, H.H., Lopes, F.B., Rosa, G.J.M. & Döpfer, D., 2020. The identification of gene ontologies and candidate genes for digital dermatitis in beef cattle from a genome-wide association study. Intern. J. Vet. Sci. Res. 6, 027-037. DOI: 10.17352/ijvsr.000050

Küpper, J.D., Brandt, H.R. & Erhardt, G., 2014. Genetic association between NOD2 polymorphism and infection status by Mycobacterium avium ssp. paratuberculosis in German Holstein cattle. Anim. Genet. 45, 114-116. DOI: 10.1111/age.12097

Kreitman, M., 2000. Methods to detect selection in populations with applications to the human. Annu. Rev. Genom. Hum. Genet. 1, 539-559. DOI: 10.1146/annurev.genom.1.1.539

Larkin, D.M., 2011. Status of the cattle genome map. Cytogenet. Genome Res. 134, 1-8. DOI: 10.1159/000324695

Leal-Gutiérrez, J.D., Elzo, M.A., Johnson, D.D., Hamblen, H. & Mateescu, R.G., 2019. Genome wide association and gene enrichment analysis reveal membrane anchoring and structural proteins associated with meat quality in beef. BMC Genom. 20, 1-18. DOI: 10.1186/s12864-019-5518-3

Lewin, A. & Grieve, A.C., 2006. Grouping gene ontology terms to improve the assessment of gene set enrichment in microarray data. BMC Bioinform. 7, 1-9. DOI: 10.1186/1471-2105-7-426

Lin, L., Park, M. & York, D.A., 2007. Enterostatin inhibition of dietary fat intake is modulated through the melanocortin system. Peptides 28, 643-649. DOI: 10.1016/j.peptides.2006.10.003

Luisi, P., Alvarez-Ponce, D., Pybus, M., Fares, M.A., Jaume, B. & Laayouni, H., 2015. Recent positive selection has acted on genes encoding proteins with more interactions within the whole human interactome. Genome Biol. Evol. 7, 1141–1154. DOI: 10.1093/gbe/evv055

Maciel, S.M.A., Amino, J.O., Martins, M., Okeyo, A.M., Scholtz, M.M. & Neser, F., 2013. Feedlot performance of the Nguni ecotypes in southern Mozambique. Livest. Res. Rural Dev. 25, 1-15. http://www.lrrd.org/lrrd25/6/maci25111.htm

Matjuda, E.L., 2012. Developing breeding objectives for the Nguni cattle breed of South Africa. PhD thesis, Faculty of Science and Agriculture, University of Limpopo, South Africa. Pp 1-129.

Maiorano, A.M., Lourenco, D.L., Tsuruta, S., Ospina, A.M.T., Stafuzza, N.B., Masuda, Y., Filho, A.E.V., Cyrillo, J.N.G., Curi, R.A. & Silva, J.A.D. 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. Plos One, pone.0200694, 1-24. DOI: 10.1371/journal.pone.0200694

Makina, S.O., Muchadeyi, F.C., Marle-Köster, E., Taylor, J.F., Makgahlela, M.L. & Maiwashe, A., 2015. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genet. Sel. Evol. 47, 1-14. DOI: 10.1186/s12711-015-0173-x

Mapholi, N.O., 2015. Exploring genetic architecture of tick resistance in South African Nguni cattle. Doctoral dissertation, Department of Animal Science, Stellenbosch University, South Africa. Pp. 1-120.

Mapholi, N.O., Maiwashe, A., Matika, O., Riggio, V., Banga, C., MacNeil, M.D., Muchenje, V., Nephawe, K. & Dzama, K., 2017. Genetic parameters for tick counts across months for different tick species and anatomical locations in South African Nguni cattle. Trop. Anim. Health Prod. 49.1201-1210. DOI: 10.1007/s11250-017-1336-2

Mapiye, C., Chikwanha, O.C., Chimonyo, M. & Dzama, K., 2019. Strategies for sustainable use of indigenous cattle genetic resources in southern Africa. Diversity 11, 1-14. https://doi.org/10.3390/d11110214

Marufu, M.C., Chimonyo, M., Mans, B.J. & Dzama, K., 2013. Cutaneous hypersensitivity responses to Rhipicephalus tiek larval antigens in pre-sensitized cattle. Ticks Tick Borne. Dis. 4, 311-316.

Marufu, M.C., Qokweni, L., Chomonyo, M. & Dzama, K., 2011. Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semi-arid rangelands in South Africa. Ticks Tick-Borne Dis. 2, 172-177. DOI: 10.1016/j.ttbdis.2012.12.001

McKay, S., Betancourt, F., Bhattarai, S., Buttolph, T., White, S., Lachance, H., Quijada, D., Friedman, S., Perlee, S., Cantrell, B. & Murdoch, B., 2018. 115 Profiling conservation of DNA methylation in cattle. J. Anim. Sci. 96, 370-370.

McNees, A.L., Markesich, D., Zayyani, N.R. & Graham, D.Y., 2015. Mycobacterium paratuberculosis as a cause of Crohn's disease. Expert Rev. Gastroenterol. Hepatol. 9, 1523-1534. DOI: 10.1586/17474124.2015.1093931

Müller, M.P., Rothammer, S., Seichter, D., Russ, I., Hinrichs, D., Tetens, J., Thaller, G. & Medugorac, I., 2017. Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. J. Dairy Sci., 100, 1987-2006. DOI: 10.3168/jds.2016-11506

Mwai, O., Hanotte, O., Kwon, Y.J. & Cho, S., 2015. African indigenous cattle: unique genetic resources in a rapidly changing world. Asian Australas. J. Anim. Sci. 28, 911-921. DOI: 10.5713/ajas.15.0002R

Naval-Sanchez, M., McWilliam, S., Evans, B., Yanez, J.M., Houston, R.D. & Kijas, J.W., 2020. Changed patterns of genomic variation following recent domestication: 495 selection sweeps in farmed Atlantic salmon. Front. Genet. 11, 1-13. DOI: 10.3389/fgene.2020.00264

Nguyen, T.M., Shafi, A., Nguyen, T. & Draghici, S., 2019. Identifying significantly impacted pathways: A comprehensive review and assessment. Genome Biol. 20, 1-15. DOI: 10.1186/s13059-019-1790-4

Oleksyk, T.K., Zhao, K., De La Vega, F.M., Gilbert, D.A., O'Brien, S.J. & Smith, M.W., 2008. Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. PLoS One, e1712, 1-15. DOI: 10.1371/journal.pone.0001712

Pinedo, P.J., Buergelt, C.D., Donovan, G.A., Melendez, P., Morel, L., Wu, R., Langaeee, T.Y. & OwenRaea, D., 2009. Association between CARD15/NOD2 gene polymorphisms and paratuberculosis infection in cattle. Vet. Microbiol. 134, 346-352. DOI: 10.1016/j.vetmic.2008.09.052

Purfield, D.C., McParland, S., Wall, E. & Berry, D.P. 2017. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One, e0176780, 1-23. DOI: 10.1371/journal.pone.0176780

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., Nielsen, R. & Simianer, H., 2014. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 10, 1-13. DOI: 10.1371/journal.pgen.1004148

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D. & Taylor, J.F., 2013. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genom. 14, 1-18. DOI: 10.1186/1471-2164-14-382

Ramsay, K., Harris, L. & Kotze, A., 2000. Landrace Breeds: South Africa's indigenous and locally developed farm animals. Farm Animal Conservation Trust, Pretoria. Pp 38-39.

Rebl, A., Rebl, H., Verleih, M., Haupt, S., Köbis, J.M., Goldammer, T. & Seyfert, H.M., 2019. At least two genes encode many variants of Irak3 in rainbow trout, but neither the full-length factor nor its variants interferes directly with the TLR-mediated stimulation of inflammation. Front. Immunol. 10, 1-19. DOI: 10.3389/fimmu.2019.02246

Scholtz, M.M. & Theunissen, A., 2010. The use of indigenous cattle in terminal cross-breeding to improve beef cattle production in sub-Saharan Africa. Anim. Genet. Resour. 46, 33-39. DOI:10.1017/S2078633610000676

Solaymani-Mohammadi, S., Eckmann, L. & Singer, S.M., 2019. Interleukin (IL)-21 in inflammation and immunity during parasitic diseases. Front. Cell. Infect. Microbiol. 9, 1-14. DOI: 10.3389/fcimb.2019.00401

Squillacioti, C., De Luca, A., Ali, S., Paino, S., Liguori, G. & Mirabella, N. 2012. Expression of urocortin and corticotropin-releasing hormone receptors in the horse thyroid gland. Cell and Tissue Res. 350, 45-53. DOI: 10.1007/s00441-012-1450-4

Stella, A., Ajmone-Marsan, P., Lazzari, B. & Boettcher, P., 2010. Identification of selection signatures in cattle breeds selected for dairy production. Genetics 185, 1451-1461. DOI: 10.1534/genetics.110.116111

Stronen, A.V., Pertoldi, C., Iacolina, L., Kadarmideen, H.N. & Kristensen, T.N., 2019. Genomic analyses suggest adaptive differentiation of northern European native cattle breeds. Evol. Appl. 12, 1096-1113. DOI: 10.1111/eva.12783

Sun, L., Liu, S., Wang, R., Jiang, Y., Zhang, Y., Zhang, J., Bao, L., Kaltenboeck, L., Dunham, R., Waldbieser, G. & Liu, Z., 2014. Identification and analysis of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (*Ictalurus punctatus*). PLoS One 9, e109666, 1-10. DOI: 10.1371/journal.pone.0109666

Syrovatkina, V., Alegre, K.O., Dey, R. & Huang, X.Y., 2016. Regulation, signaling, and physiological functions of G-proteins. J. Mol. Biol. 428, 3850-3868. DOI: 10.1016/j.jmb.2016.08.002

Tada, O., Muchenje, V. & Dzama, K., 2013. Preferential traits for breeding Nguni cattle in low-input in-situ conservation production systems. Springerplus 2, 1-7. DOI: 10.1186/2193-1801-2-195

Taylor K.H., Taylor J.F., White S.N. & Womack J.E., 2006. Identification of genetic variation and putative regulatory regions in bovine CARD15. Mamm. Genome17(8),892-901. DOI: 10.1007/s00335-005-0148-2

Thomas, P.D., 2017. The gene ontology and the meaning of biological function. In: C. Dessimoz & N. Škunca (eds), The gene ontology handbook. Methods Mol. Biol., Humana Press, New York. Pp. 15-24. https://doi.org/10.1007/978-1-4939-3743-1_2

Vara-Ciruelos, D., Dandapani, M. & Hardie, D.G., 2020. AMP-activated protein kinase: Friend or foe in cancer? Annu. Rev. Cancer Biol. 4, 1-16. DOI:10.1146/annurev-cancerbio-030419-033619

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. & Yang, J., 2017. 10 Years of GWAS discovery: Biology, function, and translation. Am. J. Hum. Genet. 101, 5-22. DOI: 10.1016/j.ajhg.2017.06.005

Wang, Y., Wang, S., Liu, T., Tu, W., Li, W., Dong, G., Xu, C., Qin, B., Liu, K., Yang, J., Chai, J., Shi, X. & Zhang, Y., 2015. CARD15 gene polymorphisms are associated with tuberculosis susceptibility in Chinese Holstein cows. PloS One, 10, 1-10. DOI: 10.1371/journal.pone.0135085

Wang, D., 2018. The essential role of G protein-coupled receptor (GPCR) signaling in regulating T cell immunity. Immunopharmacol. Immunotoxicol. 40, 187-192. DOI: 10.1080/08923973.2018.1434792

Weldenegodguad, M., Popov, R., Pokharel, K., Ammosov, I., Ming, Y., Ivanova, Z. & Kantanen, J., 2019. Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. Front. Genet. 9, 1-13. DOI: 10.3389/fgene.2018.00728

Wright, D., 2015. The genetic architecture of domestication in animals. Bioinform. Biol. Insights. 9(Suppl. 4), 11-20. DOI: 10.4137/BBI.S28902

Yaro, M., Munyard, K.A., Morgan, E., Allcock, R.J.N., Stear, M.J. & Groth, D.M., 2019. Analysis of pooled genome sequences from Djallonke and Sahelian sheep of Ghana reveals co-localisation of regions of reduced heterozygosity with candidate genes for disease resistance and adaptation to a tropical environment. BMC Genom. 20, 1-14. DOI: 10.1186/s12864-019-6198-8

Zimin, A., Delcher, A., Florea, L., Kelley, D., Schatz, M., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C., Sonstegard, T., Marçais, G., Roberts, M., Subramanian, P., Yorke, J. & Salzberg, S., 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10, 1-10.DOI: 10.1186/gb-2009-10-4-r42

Zhao, H., Wu, M., Yi, X., Tang, X., Chen, P., Wang, S. & Sun, X. Functional analysis of haplotypes in bovine PSAP gene and their relationship with beef cattle production traits. Anim. 11, 1-15. DOI: 10.3390/ani11010049

Zwane, A.A., Schnabel, R.D., Hoff, J., Choudhury, A., Makgahlela, L.M., Maiwashe, A., Van Marle-Koster, E. & Taylor, J.F., 2019. Genome-wide SNP discovery in indigenous cattle breeds of South Africa. Front. Genet. 10, 1-16. DOI: 10.3389/fgene.2019.00273

Supplementary Table 1 Most significant selective sweeps (ZHp Z-scores ≤ -6) and their associated genes.

CHR	BP (Kb)	ZH	Loci	Coordinates	Trait information
1	422	-6.890	-	-	
1	264	-6.084	-	-	
2	147	-7.327	-	-	
2	294	-7.304	-	-	
2	368	-6.980	*LMLA*	2:2026957-2027063	Meat and carcass
3	495	-6.375	-	-	
3	259	-6.081	-	-	
4	504	-7.523	-	-	
4	154	-6.060	*FAM71F1, INSINT,AD*	4:93460874-93462935, 4:93475459-93486895, 4:93508704-93524304	Reproduction, health
5	210	-7.309	-	-	
5	512	-7.082	-	-	
5	424	-7.036	*CALEASE,LMA*	5:48773272-48844474	Production, reproduction, meat & carcass
5	389	-6.646	-	-	
5	148	-6.331	-	-	
6	109	-6.647	-	-	
6	77	-6.628	-	-	
6	138	-6.537	-	-	
6	708	-6.428	-	-	
6	273	-6.265	-	-	
6	70	-6.051	*CONA,ALD-C18*	6:6013172-6020467	Health, meat
7	66	-6.378	*CNOT6,RBCC*	7:493450-574753	Health
7	244	-6.105	*SCS*	7:96124692-96354407	Health
7	480	-6.098	-	-	
8	132	-8.274	-	-	
8	51	-8.072	-	-	
8	12	-7.500	*DMRT1, RFI,HEATT*	8:43916605-43972570	Production, health, carcass
8	222	-7.293	-	-	
8	78	-7.204	-	-	
8	65	-6.806	KANK1,CALEASE,CWT	8:44046426-44076904	Reproduction, exterior, meat and carcass
8	131	-6.667	-	-	
8	517	-6.557	-	-	
8	209	-6.383	*DOCK8*	8:44310613-44545537	Reproduction
8	96	-6.197	-	-	
9	65	-6.195	-	-	
11	361	-6.148	-	-	
11	157	-6.051	*TTC27*	11:15207843-15381634	Milk, health
16	903	-6.377	-	16:659397-659500	Production
17	177	-7.430	*ASIC5,SCRCIR,WWT*	17:44427939-44483562	Production, reproduction
17	161	-6.063	*ZNF74,,TSSK1B, TSSK2, DGCR14,GSC2*	17:74560555-74570229, 17:74581054-74598153, 17:74607593-74609032, 17:74612301-74613377, 17:74613480-74620081, 17:74622501-74623699	Reproduction, production

CHR	BP (Kb)	ZH	Loci	Coordinates	Trait information
19	256	-8.222	*FAM101B, SIGG,LMA*	19:22824635-22824710, 19:22835876-22841764, 19:22856214-22856325, 19:22856388-22859271, 19:22872042-23011338	Health, production
20	716	-6.991	-	-	
20	233	-6.882	*GZMK,ESM1,CALVIND*	20:24097290-24107687, 20:24131984-24140817	Milk
20	311	-6.372	-	-	
20	401	-6.036	*NPR3, PTAT,UTH*	20:40967082-41041629	Health, milk, reproduction, exterior
20	363	-6.004	-	-	
25	48	-6.434	-	-	
27	250	-7.364	*HTC,HG*	27:4996159-4999264, 27:5024117-5031818	Health

CHR: chromosome, BP: base pairs, ZH: Z-score

Supplementary Table 2 List of genes associated with selected phenotype queries

Symbol	Description	Matched phenotypes	Matched phenotype count	Score
RAB11A	Tectonic family member 2	Fat, bone, muscle, walking, reproduction	5	1.38
PLK1	Tectonic family member 2	Fat, bone, muscle, walking, reproduction	5	1.23
CLPS	Pancreatic lipase related protein 2 (gene/pseudogene)	Fat, milk, bone, muscle, reproduction	2	0.93
GPI	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.78
NDUFV3	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.7
DPYS	Sterol-C5-Desaturase	Fat, bone, muscle, reproduction	4	0.66
TNNC1	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.66
SNRPA1	cwf19-like cell cycle control factor 1	Bone, muscle, walking	3	0.64
PRKAA1	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.61
YWHAZ	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.6
IMPDH1	Sterol-C5-desaturase	Fat, bone, muscle, reproduction	4	0.58
RIF1	Heat shock protein family b (small) member 3	Muscle, walking, bone, heat	3	0.5
CNOT6	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.45
DNAH1	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.44
TAGLN2	Heat shock protein family B (small) member 3	Muscle, walking, heat	3	0.44
KCTD10	Chloride intracellular channel 2	Fat, bone, muscle, reproduction	4	0.43
FLNC	Heat shock protein family B (small) member 3	Muscle, walking, heat	3	0.43
ZNF74	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.38

Symbol	Description	Matched phenotypes	Matched phenotype count	Score
RIOK2	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.37
TDRD9	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.36
MCOLN3	Chloride intracellular channel 2	Fat, bone, muscle, reproduction	4	0.35
MCOLN2	Chloride intracellular channel 2	Fat, bone, muscle, reproduction	4	0.35
ASIC5	FXYD domain containing ion transport regulator 4	Bone	1	0.33
PABPC1	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.32
PCBP2	Cytochrome C oxidase subunit 7B	Fat, bone, muscle, reproduction	4	0.32
AMHR2	Zinc finger DHHC-type containing 15	Fat, bone, muscle	3	0.31
IRAK3	Chloride intracellular channel 2	Fat, bone, muscle, reproduction	4	0.23
AQP8	Chloride intracellular channel 2	Muscle	4	0.22
SLC30A7	Arylsulfatase F	Heat, fat, bone	1	0.21
RGS5	N-Acyl phosphatidylethanolamine phospholipase D	Heat, muscle	1	0.2
CRHR2	Chloride intracellular channel 2	Bone, muscle, heat, reproduction	4	0.18
ZNF407	Zinc finger MYND-type containing 15	reproduction	1	0.1
LHFPL5	LHFPL tetraspan subfamily member 2	Reproduction	1	0.09
ZKSCAN2	Zinc finger MYND-type containing 15	Reproduction	1	0.08
GSC2	Keratin associated protein 10-12	Fat, bone	2	0.08
TSSK2	Myosin light chain 5	Muscle, fat	1	0.07
TSSK1B	Myosin light chain 5	Muscle, fat	1	0.07
OPN1SW	N-Acyl phosphatidylethanolamine phospholipase D	Heat, reproduction	1	0.07
ERCC6L2		Bone	1	0.07
KCP	Keratin 85	Muscle	1	0.07
KANK1		bone	1	0.06
C7	Serine protease 53	Fat	1	0.06
SPRR3	Keratin associated protein 10-12	Fat, bone	2	0.06
PPIL2	Leucine rich repeat containing 46	Fat	1	0.05
AFF1		Bone	1	0.05
RASAL2	Keratin associated protein 10-12	Fat, bone	2	0.05
TBCD	Arylsulfatase F	Heat	1	0.05
DCTN5	ER degradation enhancing alpha-mannosidase-like protein 2	Fat, bone	2	0.05
SNX20	Sorting nexin family member 21	Bone	1	0.05

Symbol	Description	Matched phenotypes	Matched phenotype count	Score
PHF7	Zinc finger protein 467	Bone	1	0.05
PRR13	CUGBP elav-like family member 5	Muscle	1	0.04
CALU	MicroRNA 6852	Bone	1	0.04
OTULIN	er degradation enhancing alpha-mannosidase-like protein 2	Fat. bone	2	0.04
NCKAP1	N-Acyl phosphatidylethanolamine phospholipase D	Heat	1	0.04
TMED5	N-Acyl phosphatidylethanolamine phospholipase D	Heat	1	0.04
WIF1	N-Acyl phosphatidylethanolamine phospholipase D	Heat	1	0.04
ARMC12	Uncharacterized LOC285847	Bone	1	0.04
NPR3	Myosin light chain 5	Muscle	1	0.04
ERN2	ER degradation enhancing alpha-mannosidase-like protein 2	Fat. bone	2	0.03
DUSP23	EF-hand calcium binding domain 9	Bone. muscle	1	0.03
KIAA1191	Zinc finger protein 286B (Pseudogene)	Bone	1	0.03
IGSF9	Immunoglobulin superfamily DCC subclass member 3	Bone	1	0.03
TTC33	DnaJ heat shock protein family (Hsp40) member C22	Fat. bone	1	0.03
SEMA3G	Myosin light chain 5	Muscle	1	0.02
C1QTNF8	Complement C1q-like 4	Bone	1	0.02
DMRT1	Ectonucleotide Pyrophosphatase/Phosphodiesterase 6	Bone	1	0.02
NELL2	Immunoglobulin superfamily DCC subclass member 3	Bone	1	0.02
UVSSA		Bone	1	0.02
FAM71F1	Family with sequence similarity 71 member B	Bone	1	0.02
TTC27	Armadillo-like helical domain containing 3	Bone	1	0.02
STRIP2	Chromosome 12 open reading frame 74	Bone	1	0.02
DR1	STT3A Antisense RNA 1	Bone	1	0.01
HEPHL1	Procollagen C-Endopeptidase enhancer 2	Bone	1	0.01
MEGF11	Leucine rich repeats and IQ motif containing 1	Bone	1	0