Signatures Analysis of Divergent Selection Revealed Genes Associated with Different Biological Aspects of Goats in Kenya

Kamidi C.M.[2], Waineina R.W.*[1,2], Wasike C.B[3]. Ilatsia E.D.[2], and Ngeno K.[1]

[1]Department of Animal Sciences, Animal Breeding and Genomics Group, Egerton University, P.O. Box 536-20115, Egerton, Kenya
[2]Dairy Research Institute, Kenya Agricultural and Livestock Organization, P.O. Box 2520117, Naivasha, Kenya
[3]Livestock Efficiency Enhancement Group (LEEG), Department of Animal and Fisheries Sciences, Maseno University, P.O. Box Private Bag 40105, Maseno, Kenya

*Corresponding author e-mail: *waineinaruth@gmail.com*

Abstract

The identification of genes and mutations linked to phenotypic features significant for livestock species may be revealed through the detection of selection signatures. This study aimed to decipher the selection signatures and mutational load of different dairy goat populations kept at KALRO-research station Naivasha. Analysis was done on the following breeds; Galla (n=12), and exotic dairy goats imported from South Africa (n=16). DNA was extracted from whole blood and genotyped using the Illumina GoatSNP50 BeadChip where 49773 single-nucleotide polymorphisms (SNPs) were generated. Analysis of selection signatures was carried out using fixation index (Fst) and nucleotide diversity (Pi). Two breeds, the Galla and imported dairy goats from South Africa were assessed based on the 0.05% top and lowest Fst and Pi respectively. The genes linked to pathways that are either directly or indirectly related to environmental adaptability were found to be more abundant in certain locations. These was revealed by genes associated with the immune response, organ development (TBC1D32, ADAMTS18), memory (SORCS3), sensory receptors and signaling (CDH23, CEP350, KCND2). We found genomic regions that could offer background information to comprehend the mechanisms influencing economic traits in the different goat breeds. Furthermore, the findings of this study are useful in furthering the genetic improvement of goat breeds kept at the KALRO research station and Kenya as a whole.

Keywords: Annotation, Capra hircus, Gene ontology

Introduction

Natural selection is crucial in the process of selecting individuals best adapted to changing environmental conditions. However, artificial selection has been frequently used to produce more desirable/profitable phenotypes in livestock species (Brito *et al.,* 2017). Artificial selection has produced goat specialised breeds that are raised in various parts of the world for the production of milk, fibre, meat and leather. Pressure from natural and artificial selection strategies has an impact on the genomic regions that control these traits (i.e. milk, meat, and fibre) as well as other crucial traits like environment adaptation, reproduction, body conformation, behaviour, and resistance to diseases and parasites. Signatures of selection are the distinctive genetic patterns that natural and/or artificial selection has left in the genomes of individual livestock species. These signatures are typically found in areas of the genome that include functionally significant sequence variants. The detection of selection signatures in breeds of livestock species can contribute to the identification of regions of the genome that are, or have been, functionally important and, as a consequence, have been targeted by selection (Makina *et al.,* 2015). Molecular genetic markers, such as high-density SNPs, are often used to analyse genetic variation at the genome-wide level caused by natural and artificial selection of species.

Exotic dairy goats were first introduced in Kenya in the 1950s by British farmers. The

recent population estimate is about 175,000 with mostly exotic breeds such as Toggenburg, Anglo-Nubian, German Alpines, Saanen, and Boer and their hybrids with local goats. Indigenous and exotic goats have significant diversity, but little is known about the genetic variance that has resulted in diverse phenotypic characteristics and adaptation to environmental stresses (water, heat, disease and parasite stress). The Illumina Goat 50K SNP Bead Chip (Tosser-Klopp *et al.,* 2014) which was made available by recent developments in genomic technologies, has made it possible to look for genomic areas that may have undergone selection. These investigations in cattle (Makina *et al.,* 2015) and cattle have all been discovered genes that have the positive selection and are probably directly responsible for variations. In this study, we investigated the selection signatures of goat populations in Kenya. Understanding these populations and integrating their genomic information will undoubtedly lead to improvement in productivity, better management and sustainable utilisation of these genetic resources. Information from this study will advance our understanding of ecological adaptation and its potential application in functional genomics and selective breeding as well as the design of management programs to conserve livestock genetic diversity to cope with the current and future predicted effects of climate change for the benefit of the farmer.

Materials and methods

Ethics Statement

The study was undertaken in accordance with the Institute of Primate Research (IPR) ethical guidelines on animal care and the use of laboratory animals (approval reference ISERC/03/2020). A qualified veterinary officer handled the animals during blood collection to minimise pain and discomfort.

Sample collection

A total of 28 blood samples were collected from the goat population. The breeds sampled included imported exotic dairy breeds (n=16) and Galla (n=12). Blood samples were collected via jugular venipuncture from each animal into EDTA vacutainer tubes. The collected blood samples were kept in iceboxes until refrigerated at 4°C before DNA extraction. The study was conducted in strict accordance with the recommendations of the Institute of Primate Research (IPR) ethical guidelines on animal care and the use of laboratory animals.

DNA extraction and genotyping

Genomic DNA was extracted from the whole blood using the DNeasy Blood and tissue kit 149 (Qiagen®, Hilden, Germany). Genomic DNA was quantified using Nanodrop Spectrophotometer 150 (Nanodrop ND-1000) and genotyped using the GoatSNP50 Bead Chip (Illumina, Inc. San Diego, 151 CA 92122 USA) developed by the International Goat Genome Consortium (IGGC).

Quality control

Quality control procedures were performed in PLINK v 1.9 using the following parameters; SNPs with less than 95% call rate, less than 0.05 Minor Allele Frequency (MAF< 0.05), Hardy-Weinberg Equilibrium <0.001 and SNPs with more than 10% missing genotypes were filtered. SNPs that had high linkage disequilibrium (LD) were pruned using the indep-pairwise command parameters (SNP window size: 50, SNPs shifted per step:5, r2 thresholds: 0.2).

Data Analysis

Signatures of selection were analysed between two different populations' i.e. Galla and imported exotic dairy genotypes. A three-step approach was adopted to identify those regions that are affected by selection. In step one, the calculation of fixation index (Fst) values for the two populations will be done. Using VCFtools in bins of a 100kb window sliding 50kb each time over the whole genome. Step two involved the calculation of heterozygosity within each 100kb window for each IC cluster using VCFtools. To reduce of the number of false positives, all windows with less than 10 variants in the two steps were filtered (Ngeno *et al.,* 2015). Using Z-transformation using $ZFst=(Fst-\mu Fst)/\sigma Fst$ and $ZPi=(Pi-\mu Pi)/\sigma Pi$ where μ and σ are the mean and standard deviation. Finally, regions affected by selection were identified as 0.05% windows with high transformed Fst and lowest

nucleotide diversity respectively. The graph was drawn using the ggplot2 package in the R environment (http://www.r-projects.org/).

Using Variant Effect Predictor VEP75, functional annotations of genomic variants in the high transformed Fst regions were determined. SIFT predictions option within VEP was used in the prediction of whether substitution of amino acid has an effect on protein function through the use of the annotations in Ensembl 170. Missense variants within the elevated Fst regions were extracted for the Gene Ontology (GO) enrichment analysis for their associated genes. Overrepresentation of GO terms enrichment for biological processes was carried out using BinGO v2.44 within Cytoscape v.2.8.3 and gProffiler. significance at 0.05 was tested using Benjamini and Hochberg correction method designed for multiple.

Results

As presented in Figure 1 the results of show genetic signatures of selection in dairy goats with a total of 49773 out of 53347 SNPs that were retained for downstream analysis after quality control checks. A fixation index (Fst) of >0.93 was found in the 57600001 to 57750000 region of chromosome 6. This shows that there is very little genetic variation in this region, which suggests that it has been under strong selection. The strongest region affected by selection (Fst-0.93 and ZFst = 6.06) is also on chromosome 6. A cut-off threshold of the top 0.05% of ZFst values was used to identify regions that may have been under selection. 23 regions were identified across 12 different chromosomes (2, 4, 5, 6, 8,11,12,14,16,18,19 and 21).. The study found that there are a number of regions in the dairy goat genome that have been under strong selection. These regions may contain genes that are important for traits such as milk production or disease resistance.

Selection hotspots were searched by identifying genomic regions with the lowest nucleotide diversity (Fig. 2). Screening of genomes revealed four distinct chromosomes with the lowest (0.0500%, 24 regions).

Functional annotation of selection signature genes

The strongest selection hotspot (ZFst = 0.68) harboured varied variants (Figure 4). Results for variant (located in elevated Fst regions) annotation performed using VEP, showed that 60% of coding variants were missense and 40% synonymous (Fig 3). All the missense variants predicted were not deleterious but rather significant (SIFT) < 0.05) modifiers of protein structure. Elevated Fst candidate selection regions overlapped with APBA2, ARMC7, CLASP1, DMRT1, N4BP2L2, NELFB, NRARP and PGM2 genes whose size and location are shown in Appendix.

Function and pathway analysis for the RNP gene was done, and gene set enrichment analysis was performed using BINGO. The gene was significantly enriched in the cellular process (FDR = 4.07E-02) GO terms.

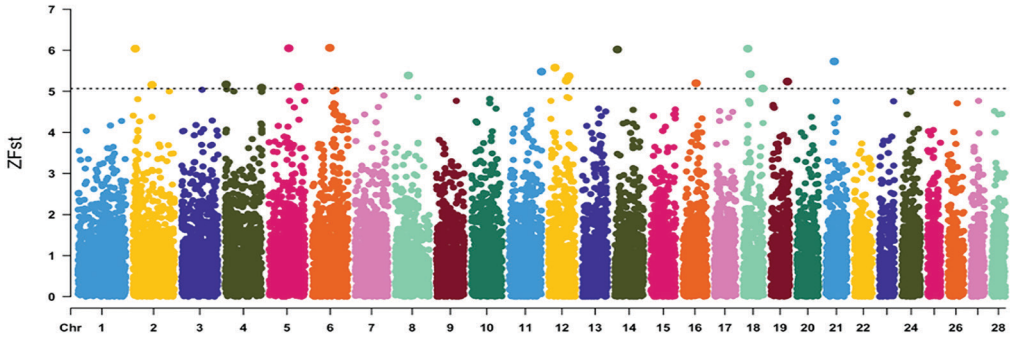

Figure 1: Genome-wide distributions of Z score of Fst (ZFst) across all the autosomal chromosomes. The y-axis values are ZFst, and the x-axis is locations within chromosomes

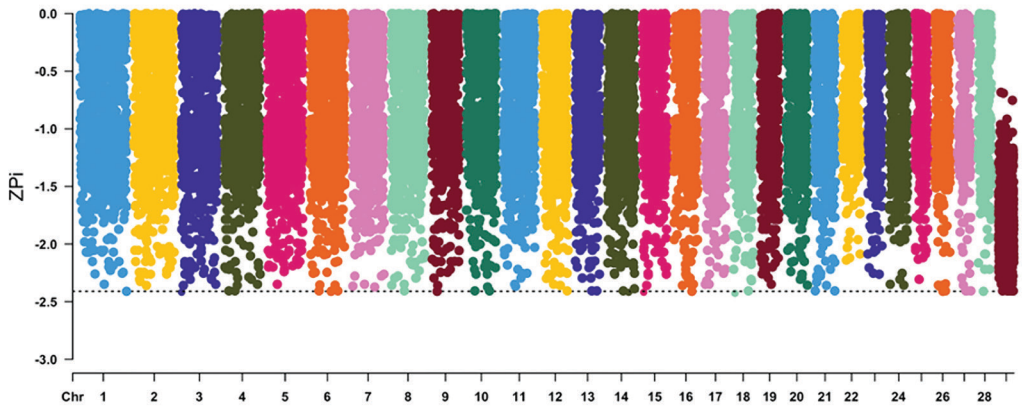

Figure 2: Genome-wide Z score of nucleotide diversity (ZPi) plot. The y-axis values are −ZPi, and the x-axis are locations within chromosomes

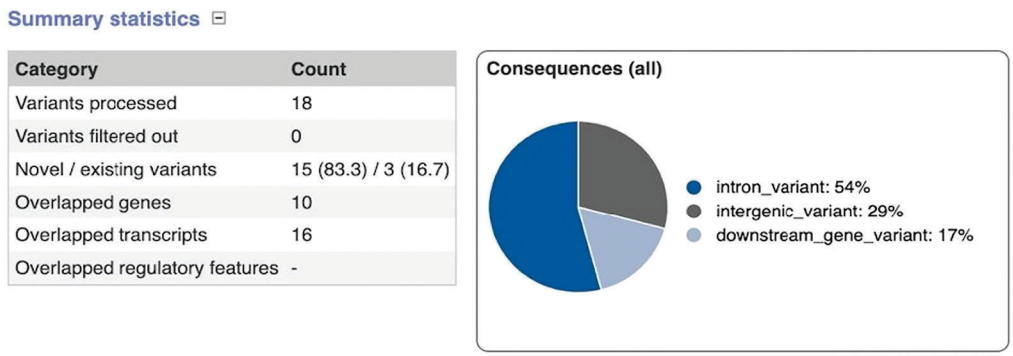

Figure 3: Summary Statistics per category for variant types using VEP Ensemble gene annotation. (A) Pie plot for all the variant consequences. (B) Pie plot of the coding consequence

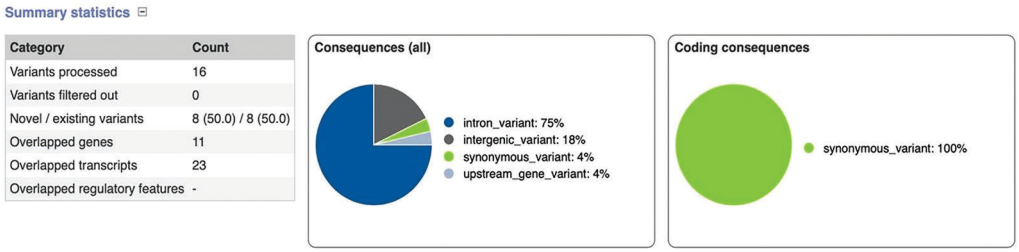

Figure 4: Summary statistics per category for variant types using VEP Ensemble gene annotation

Further analysis of PRNP Gene in relation to phenotypes was performed using human phenotype ontology in g: Profiler. Phenotype enrichment analysis revealed seven significant (FDR < 0.05) human phenotype terms that are associated with the following phenotype terms; Vestibular nystagmus, Motor impersistence, Simultanapraxia, EEG with persistent abnormal rhythmic activity, EEG with constitutional variants, Jaw pain and Progressive forgetfulness. The lowest nucleotide diversity regions have Z score transformed nucleotide values

(ZPi) ranging from -2.41 to – 2.43 and host different types of variants (Fig 4). these. regions overlapped with PP2D1, OSER1, GGH, CEP350, ADAMTS18, SORCS3, CDH23, OPCML, KCND2 and TBC1D32 genes. The activities of the GGH gene include omega peptidase, hydrolase, gamma-glutamyl-peptidase and cellular anatomical entity. The CEP350 role is microtubule binding, protein localization to the centrosome and non-motile cilium assembly. ADAMTS18 activities are eye development, metalloendopeptidase, proteolysis, hydrolase, extracellular matrix organisation, metal ion binding and negative regulation of platelet aggregation.

Discussion

Results of the phylogenetic tree established Galla and imported exotic populations as genetically distinct. Differentiations in the two populations would infer the existence of distinctive genomic architectures possibly due to divergent evolution and selection pressures.

Goat farmers in arid and semi-arid lands (ASALs) face many challenges, including diseases, pests, poor quality and inadequate forage, and heat and water stress. Understanding how goats adapt to these harsh environments could help to improve their performance. Unravelling of genomic basis for the ability to adapt to the harsh ASAL production environments is vital for enhancing goat performances. Analysis of selection signatures has been a proven approach for dissecting animal genome and understanding their adaptation. The objective of the analysis of selection signatures in this study was to uncover genomic regions that have undergone favoured selection in Galla and exotic goat genomes. To achieve this, Fst and nucleotide diversity statistical approaches were used in detecting regions with high and low allele frequency differences due to selection between the two goat breeds. This procedure identified 23 candidate genomic regions across 12 chromosomes that underwent selection. The selection hotspots mapped to APBA2, ARMC7, CLASP1, DMRT1, N4BP2L2, NELFB, NRARP and PGM2 protein-coding genes. Evidence of diverged selection for sex determination and gonadal development has been mapped to an intron variant in the DMRT1 gene at chromosome 8. ARMC7 is a protein-coding gene associated with Atresia ani (imperforate anus) disease in goats, where the anus is blocked or missing due to the breakdown failure of anal membrane (Kumar 2009). The gene has an allele (A) positioned at 55549012 on chromosome 19 that has undergone strong selection. The NRARP gene is associated with caprine lymphoma disease. Disruption of interactions of the genetic or acquired immune system with various pathogens and environmental elements has been associated with the aetiology of lymphoma disease. The downstream gene variant (allele T) in the sweep region is located in 105952406 of chromosome 11.

Nucleotide diversity analysis revealed candidate regions with low variation (low Pi) that overlapped different genes including SORCS3 which is linked with Amyloidosis disease. This heritable disease is caused by insoluble protein accumulation. The OPCML gene is associated with ovarian cancer. Tumours in goats mostly affect the genital system (ovaries, uterus, cervix and vagina (Agnew and MacLanchlan, 2016).

Conclusion

Genomic analysis of the phylogenetic tree established Galla and imported exotic populations revealed two distinct genetic pools. Galla and exotic dairy goats contrasting evolutionary patterns could be caused by genes affected by selection. Galla, which are mainly kept in a harsh scavenging system for meat, and exotic dairy goats, which primarily provide milk, could evolve differently, and disease-linked GO genes have played some roles in their evolution.

References

Agnew, D.W. and MacLachlan, N.J., 2016. Tumors of the genital systems. Tumors in domestic animals, pp.689-722.

Brito, L.F., Kijas, J.W., Ventura, R.V., Sargolzaei, M., Porto-Neto, L.R., Cánovas, A., ... & Schenkel, F.S. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC genomics, 18(1), 1-20.

Makina, S.O., Muchadeyi, F.C., van 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. Genetics Selection Evolution, 47(1): 1-14.

Kumar, H., Sharma, A.K., Dass, L.L. and Anand, A., 2009. Atresia ani with scrotal anomaly in a goat. *Veterinary World,* 2(2), p.68.

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., ... & International Goat Genome Consortium. (2014). Design and characterization of a 52K SNP chip for goats. *PloS one,* 9(1), e86227.

Appendix

Table 1: GO terms

Chrom	Allele	Consequence	SYMBOL	GO
21	T	intron variant	APBA2	amyloid-beta binding, protein binding, chemical synaptic transmission and vesicle
21	T	intron variant	APBA2	amyloid-beta binding, protein binding, chemical synaptic transmission and vesicle
21	T	intron variant	APBA2	amyloid-beta binding, protein binding,: chemical synaptic transmission,: synaptic vesicle
21	T	intron variant	APBA2	amyloid-beta binding, protein binding, chemical synaptic transmission,: synaptic vesicle
19	A	intron variant	ARMC7	:protein binding
2	A	intron variant	CLASP1	kinetochore microtubule, cytoplasmic microtubule, kinetochore binding, microtubule plus-end binding
2	C	intron variant	CLASP1	kinetochore microtubule, cytoplasmic microtubule, GO:0043515:kinetochore binding, microtubule plus-end binding
8	G	intron variant	DMRT1	negative regulation of transcription by RNA polymerase II, cell morphogenesis, RNA polymerase II transcription regulatory region sequence-specific DNA binding, cis-regulatory region sequence-specific DNA binding, DNA-binding transcription activator activity - RNA polymerase II-specific, female germ cell nucleus, male germ cell proliferation, developmental process involved in reproduction, chromatin binding, nucleus, cytoplasm, regulation of transcription - DNA-templated, meiosis I, spermatogenesis, germ cell migration, male gonad development, male sex determination, identical protein binding, sequence-specific DNA binding, negative regulation of meiotic nuclear division, positive regulation of mitotic nuclear division, positive regulation of transcription by RNA polymerase II, male sex differentiation, oocyte development, Sertoli cell differentiation, Sertoli cell development, positive regulation of meiosis I, regulation of nodal signaling pathway, sequence-specific double-stranded DNA binding, positive regulation of male gonad development
8	G	intron variant	DMRT1	nucleus, regulation of transcription - DNA-templated, sequence-specific DNA binding

Chrom	Allele	Consequence	SYMBOL	GO
12	G	downstream gene variant	N4BP2L2	negative regulation of transcription by RNA polymerase II, in utero embryonic development, blastocyst development, transcription corepressor activity, nucleus, transcription repressor complex, enzyme binding, positive regulation of hematopoietic stem cell proliferation, negative regulation of hematopoietic stem cell differentiation
11	G	intron variant	NELFB	nucleus, nucleoplasm, cytoplasm, cell population proliferation, NELF complex, negative regulation of transcription elongation from RNA polymerase II promoter, negative regulation of transcription - DNA-templated, stem cell differentiation, negative regulation of stem cell differentiation
11	T	downstream gene variant	NRARP	negative regulation of transcription by RNA polymerase II, branching involved in blood vessel morphogenesis, endothelial cell proliferation, sprouting angiogenesis, blood vessel endothelial cell proliferation involved in sprouting angiogenesis, protein binding, Notch signaling pathway, regulation of cell-cell adhesion,T cell differentiation, somite rostral/caudal axis specification, negative regulation of T cell differentiati on,GO:0045746:negative regulation of Notch signaling pathway,positive regulation of canonical Wnt signaling pathway,v ascular endothelial cell proliferation, negative regulation of Notch signaling pathway involved in somitogenesis,:positive regulation of vascular endothelial cell proliferation
6	T	intron variant	PGM2	magnesium ion binding, phosphoglucomutase activity, cytosol,carbohydrate metabolic process, glucose metabolic process, phosphopentomutase activity, intramolecular transferase activity - phosphotransferases, organic substance metabolic process
6	T	intron variant	PGM2	magnesium ion binding, carbohydrate metabolic process,intramolecular transferase activity - phosphotransferases, organic substance metabolic process