the **genetics**society

Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations

K. Mdladla[1,2] · E. F. Dzomba [iD][2] · F. C. Muchadeyi [iD][1]

Abstract

In Africa, extensively raised livestock populations in most smallholder farming communities are exposed to harsh and heterogeneous climatic conditions and disease pathogens that they adapt to in order to survive. Majority of these livestock species, including goats, are of non-descript and uncharacterized breeds and their response to natural selection presented by heterogeneous environments is still unresolved. This study investigated genetic diversity and its association with environmental and geographic conditions in 194 South African indigenous goats from different geographic locations genotyped on the Illumina goat SNP50K panel. Population structure analysis revealed a homogeneous genetic cluster of the Tankwa goats, restricted to the Northern Cape province. Overall, the Boer, Kalahari Red, and Savanna showed a wide geographic spread of shared genetic components, whereas the village ecotypes revealed a longitudinal distribution. The relative importance of environmental factors on genetic variation of goat populations was assessed using redundancy analysis (RDA). Climatic and geographic variables explained 22% of the total variation while climatic variables alone accounted for 17% of the diversity. Geographic variables solitarily explained 1% of the total variation. The first axis (Model I) of the RDA analysis revealed 329 outlier SNPs. Landscape genomic approaches of spatial analysis method (SAM) identified a total of 843 (1.75%) SNPs, while latent factor mixed models (LFMM) identified 714 (1.48%) SNPs significantly associated with environmental variables. Significant markers were within genes involved in biological functions potentially important for environmental adaptation. Overall, the study suggested environmental factors to have some effect in shaping the genetic variation of South African indigenous goat populations. Loci observed to be significant and under selection may be responsible for the adaption of the goat populations to local production systems.

Introduction

South Africa (SA) has diverse indigenous livestock breeds and populations raised under heterogeneous production systems and climatic environments. For example, there are over 6.8 million goats in SA categorized into commercial meat breeds of the Boer, Kalahari Red, and Savanna, feral

goat breed of the Tankwa and non-descript village ecotypes farmed by smallholder communal farmers (DAFF 2014). Over 63% of South African goats are farmed by smallholder farmers in the arid zones of the country (DAFF 2014). Goats farmed under smallholder farming systems are non-descript populations generally defined as "ecotypes" according to the geographical region and agro-ecological zone they are found, e.g., Nguni ecotype goats inhabit the arid and subtropical wet agro-ecological zones of KwaZulu-Natal where the Nguni ethnic groups reside, whereas the Xhosa lobbed goats occupy the arid zones of the Eastern Cape province of the country. The Northern Cape Skilder goats are found in the desert-like agro-ecological zones of the Northen Cape province. The Venda, Tswana, and Zulu ecotypes are similarly raised by Venda, Tswana, and Zulu ethnic communities in Limpopo, North West, and KwaZulu-Natal provinces, respectively. The Tankwa goats are a unique feral breed found in the hot and drier desert-like agro-climatic zones of the Northern Cape. The

✉ F. C. Muchadeyi
MuchadeyiF@arc.agric.za

1   Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa

2   Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, South Africa

smallholder goat ecotypes and feral goat populations occur in extensive production systems with limited human management and are therefore not buffered from the effects of adverse climatic and environmental conditions. The populations are exposed to extreme climatic conditions and disease pathogens. Together with other forces of evolution such as founder effects, genetic isolation, and drift, the different environmental factors, diseases, and disease pathogen impose diverse natural selection pressures that influence the population genetic structure and distribution of genotypes, which define the level of local adaptation. Indigenous smallholder goats and other livestock species are considered to be generally highly adapted to local environmental conditions contrary to exotic breeds (Harper and Penzhorn 1999; Mdladla et al. 2016b; Onzima et al. 2017). Village goats in South Africa for example are known to survive and produce optimally even under harsh environmental and production conditions (Gwaze et al. 2009). Assessing the natural selection pressure and affected genetic loci of different breeds and populations will give insight into genetic mechanisms underlying local adaptation in marginal goat populations, which is crucial for guiding their breed characterization, conservation, and selection strategies and programs.

Advances in genomics and bioinformatics are facilitating higher resolution of livestock genomes and the identification of loci that are potentially of ecological significance. Landscape genomics combine population genetics, spatial statistics, and landscape ecology to decipher the geographic and environmental processes responsible for population genetic structuring of breeds or populations (Manel et al. 2010). It has proven a valuable tool in understanding the effects of the geographic and environmental factors and in identifying production landscape variables that influence the genetic structure of indigenous livestock populations. Contrary to F_{ST} or haplotype based of methods evaluating signatures of selection, landscape genomics approaches screen for loci under selection and their association with production variables simultaneously through the incorporation of GIS information. Pariset et al. (2009) found distinct patterns of genetic variation using 27 SNP markers along environmental gradients and reported a correlation between 16 loci and the environmental parameters in North-East Mediterranean goat breeds. Colli et al. (2014) used AFLP markers and found associations with diurnal temperature range, frequency of precipitation, relative humidity, and solar radiation in European and Western Asian goat breeds. These environmental association studies are important for unraveling the genetic basis for local adaptation (Joost et al. 2007).

This study investigated the genetics of adaptation in South African goat populations using the Illumina SNP50K data and landscape genomics approaches. The Illumina SNP50K was developed from SNPs generated from meat, milk, and mixed breeds and validated in diverse breeds, some of which included South African meat breeds (Tosser-Klopp et al. 2014). The chip has been used in genetic diversity and signatures of selection studies in other populations (Kijas et al. 2013; Brito et al. 2017) and its utility was validated for South African breeds (Lashmar et al. 2015; Mdladla et al. 2016a; Visser et al. 2016). The aims of the study were therefore to: (i) reveal the geographic population genetic structure; (ii) assess and quantify the relative contribution of geographic and environmental factors to genetic variation using a combination of genetic and site-averaged environmental data, (iii) reveal evidence of association between production environment and genetic variation, and (iv) investigate the biological pathways affected by selection to ensure adaptation of indigenous goats to local environments.

Materials and methods

Goat populations and genotyping

A total of 239 goats were sampled from five provinces, Eastern Cape, KwaZulu-Natal, Limpopo, North West, and Northern Cape, that fall under three agro-ecological zones representing subtropical wet, arid, and desert conditions and representing commercial locally developed Boer ($n = 33$), Kalahari Red ($n = 40$) and Savanna ($n = 31$), local ecotypes ($n = 100$) and a feral Tankwa goat breed ($n = 25$). The subtropical wet agro-ecological zone consists of the wet and humid eastern coastal regions of the Eastern Cape and KwaZulu-Natal provinces while dry conditions are experienced in the northwestern region (arid zone) of the Eastern Cape and KwaZulu-Natal and the Limpopo and the North West provinces. The Northern Cape is located in the desert zone of the country with a maximum summer temperature of 48 °C. The same individuals were previously used in a 50K SNP genetic diversity study (Mdladla et al. 2016a). Of these 239 animals, we excluded 25 individuals without GPS coordinates (latitude and longitude), 9 with >5% missing data and 11 related individuals (IBD > 0.45). The filtering criterion resulted in 194 goats representing South African indigenous goat populations from 19 sampling locations in the form of villages, commercial farmers, and research stations. Breed representation in the final data set included goats from locally developed meat-type breeds of the Boer ($n = 23$), Kalahari Red ($n = 30$), and Savanna ($n = 27$); non-descript village ecotype populations from smallholder farmers ($n = 100$) from Eastern Cape (Xhosa; $n = 20$), KwaZulu-Natal (Zulu; $n = 25$), Limpopo (Venda; $n = 25$), North West (Tswana; $n = 19$); Nguni breed from KwaZulu-Natal ($n = 10$); and a feral Tankwa ($n = 15$) goat population

from the Carnarvon, Northern Cape. IBD in the resultant data is shown in Supplementary Table S1 The random sampling represented an extensive coverage of major agro-ecological zones and diversity of the environmental conditions in the country (sampling locations and coordinates for each individual is provided in Table S2. SNPs with >5% missing data, call rate of 95%, MAF > 0.05, and HWE P > 0.001 were retained resulting in 48,126 SNPs.

Environmental variables

Environmental variables related to temperature, precipitation, solar radiation, wind speed, and water vapor pressure were downloaded from WorldClim (Hijmans et al. 2005) with a 30-arc seconds (~1 km^2) resolution (Table S3) and extracted using *raster* function in *R* package (R Development Core Team). Data collected were over a period of 30 years from 1970 to 2000. The variables represented annual trends (e.g., mean annual temperature and annual precipitation), seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). Values for each climate variable were extracted for each population using the geographical coordinates (longitude and latitude) data for each goat sampled (Table S2).

Inference of geographic population structure

ADMIXTURE 1.21 (Alexander et al. 2009) was used to investigate admixture of subpopulations from $K = 2$ to $K = 10$ independent runs using unsupervised default settings as described in Mdladla et al. (2016a). The membership coefficients for the chosen K value for each individual were plotted as pie charts using *maps* function in *R* package (R Development Core Team) onto the South African geographic map. In cases where numerous individuals were sampled in the same geographic location; pie charts were plotted near their original geographic location for ease of visualization.

Partitioning of genomic variation among geographic and climatic variables

A subset of 36,594 SNPs with a genotyping call rate of 100% (including MAF > 0.05 and HWE P > 0.001) was used in the multivariate regression redundancy analysis (Van Den Wollenberg 1977; Legendre and Legendre 1998) to estimate the degree to which geographic coordinates, climate variables, and their combination explained the genomic variation in the indigenous goat populations. Genetic data (dependent matrix) and individual-specific

climate and geographic variables (explanatory matrix) were compared in a partial RDA using three models implemented in *vegan* function in *R* package (R Development Core Team). Model I was the full RDA model and it incorporated all geographic and climatic variables as explanatory variables. Model II controlled for geographic variables (latitude and longitude) conditioned on climate, while Model III used geographic variables as explanatory variables conditioned on geography (Gugger 2015). To estimate the percentage of variance accounted for by each component and the direction of the effect, RDA scores were plotted for each individual and vectors depicting the direction of the predictor variables relative to the RDA axis. An outlier analysis of RDA results was also done to determine SNPs that were strongly linked with multivariate environmental gradients. Outliers were identified as SNPs with the greatest squared scores (99% CI) along the first RDA axis of Model I, Model II, and Model III using PROC UNIVARIATE analysis of statistical analysis system (SAS Institute Inc. 2013).

Locus-specific landscape genomics analysis to identify environment-associated SNPs

Landscape genomics approaches were used to determine SNPs significantly associated with the geographic variables (latitude and longitude) and environmental variables (Table S1). The first landscape genomics analysis method used an individual-based spatial analysis method (SAM), developed by Joost et al. (2007) and implemented in SAMβADA software program (https://lasig.epfl.ch/sambada) to investigate the effects of the environment on locus-specific differentiation. The SAM method uses a logistic regression model, whereby individuals were coded as either presence/absence for each SNP allele and the association between the allele and the environmental parameters was measured across sampling sites. The model was considered significant when both the G and Wald tests were significant following a Bonferroni correction at 99% confidence level (Joost et al. 2007).

The second method used a non-spatial latent factor mixed model (Frichot et al. 2013) to evaluate signals of environmental adaptation. The analysis controlled for population structure using ADMIXTURE ($K = 5$) genetic structure. An MCMC algorithm was used for each of the variables on the LEA package in *R* (http://membres-timc.imag.fr/Olivier.Francois/LEA/index.htm), for 10 runs with a burn in of 5000 and 10,000 iterations to compute LFMM parameters (|z|-scores) for all loci (Frichot et al. 2013). A false discovery rate (FDR) q value was calculated for each locus based on the p values in R (R Development Core Team).

Gene annotation and pathway analysis

Significant SNPs were annotated to corresponding genes within the candidate region intervals that were retrieved from the National Centre for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov) genome browser using the *Capra hircus* genome assembly ASM170441v1 (accessed October 2016). A SNP was considered to be associated with a gene if the chromosomal position of the marker was within the 100 kb chromosomal position of the gene. A marker distance of 100 kb was based on the low linkage disequilibrium of the village goats that was below 0.1 for SNPs markers that were >200 kb apart (Mdladla et al. 2016a). Enriched functional annotation clusters were defined using a functional annotation tool implemented in Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa et al. 2004) database.

Data availability

The scripts and the input files used to filter SNPs and individuals, run ADMIXTURE, and generate maps and figures are publicly available in the Dryad Digital Repository: https://doi.org/10.5061/dryad.3402n.

Results

Geographic representation of population structure

The ADMIXTURE results from $K = 2$ to $K = 10$ were described in Mdladla et al. (2016a) and presented as Supplementary Figure S1. In summary, the commercial breeds (Boer, Kalahari Red, and Savanna), the Tankwa feral goat, and the village goat ecotypes clustered separately as early as $K = 3$. Adding additional clusters just increased the resolution of the clustering by either (i) defining substructures within these core groups or (ii) demonstrating increased diversity within a given core group. $K = 5$ was the optimal population cluster based on the lowest cross-validation standard error of 0.62 (Mdladla et al. 2016a) and was used to investigate spatial representation of population structure. The geographic distribution of the $K = 5$ population clusters across and along the geographical coordinates of South Africa is shown in Fig. 1. The Tankwa goats represented as blue circles clustered as a small homogeneous genetic structure restricted to the Northern Cape province. The ecotypes were admixed to varying degrees, corresponding to two genetic substructures, one consisting of the Venda, Zulu, and Nguni goats, and the other of Tswana and Xhosa ecotype populations. The Zulu, Venda, and Nguni genetic component was geographically distributed in Limpopo and KwaZulu-Natal provinces while the Tswana and Xhosa

genetic component was predominantly found in North West and Eastern Cape provinces. Commercial breeds (Boer, Kalahari Red and Savanna) shared genetic components spread across geographic regions of Eastern Cape, Kwa-Zulu-Natal, North West and parts of the Northern Cape provinces. The few Savanna and Kalahari Red goats from the Northern Cape (29.0313 latitude, 23.1826 longitude) appeared as two independent and list admixed populations compared to the ones from other geographical locations that were admixed. ADMIXTURE maps plotted for the Boer, Kalahari Red, and Savanna are in Supplementary Figure S2.

Redundancy analysis (RDA) models and the genomic variation partitioning

The partitioning of variation in genetic diversity by environmental and geographical variables is shown in Table 1. The Monte Carlo test permutation showed that the environmental and geographical variables explained a significant ($p < 0.001$) proportion of the genetic variation. Climate and geographic factors together (Model I) explained 22% ($R^2_{adj} = 11\%$) of the total explainable variation. Model II accounted for 17% ($R^2_{adj} = 7\%$), while Model III explained 1% ($R^2_{adj} = 1\%$) of the total variance (Table 1).

In Model I, variables strongly associated with the genetic variation were BIO9, AMWV, longitude, BIO11, BIO6 together with AMSR, BIO4, altitude, and BIO2 albeit antagonistically (Fig. 2a). For Model II, BIO9, BIO8, BIO10, BIO1, and BIO5 accounted for more variation than other variables when geographic variables were controlled (Fig. 2b). Model III showed that altitude accounted for more of the variation than longitude when the effects of environmental variables were removed. The first RDA axis for Model I detected 329 outlier SNPs using a threshold of squared scores ≤0.13 (Table S4), while Model II detected 328 SNPs (Table S5), and Model III detected 204 SNPs (Table S6) with squared scores ≤0.12 and ≤0.07, respectively. For Model I, the top ten SNPs were located on chromosomes 1, 3, 4, 5, and 6. The top ten SNPs for Model II were located on chromosomes 1, 2, 3, 4, 6, 28 and 29, while for Model III SNPs were located on chromosome 2, 4, 5, 6, and 8.

Locus-specific landscape genomic approach for SNPs associated with the environmental variables

SAM analysis detected a total of 843 SNPs (1.75%) showing significant association with one or more geographic and climatic variables. Longitude was associated with the highest number of SNPs ($n = 319$) while altitude was associated with six SNPs. No significant SNPs were observed for latitude, BIO1, BIO3, BIO5, BIO8, BIO10, BIO14, BIO15, BIO17, BIO19, and AMWS. The
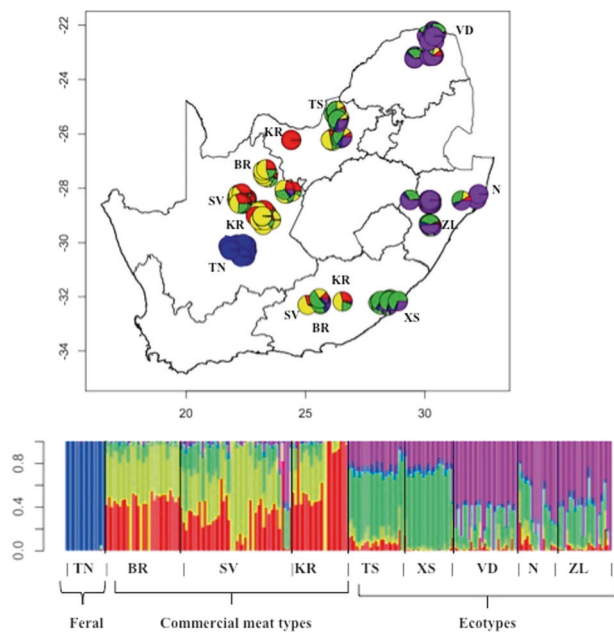
Fig. 1 Map of South Africa. The ADMIXTURE plot represents average membership coefficients resulting from the genetic structure analysis (best fit model, $K = 5$). Each color represents a different gene pool. The included barplot represents each accession as a single vertical bar broken into K color segments, with lengths proportional to the estimate probability of membership in each inferred cluster. The breeds are represented as: TN Tankwa, BR Boer, KR Kalahari Red, SV Savanna, TS Tswana, VD Venda, XS Xhosa, N Nguni, and ZL Zulu

significant SNPs and associated genes for the different environmental variables are presented in Table S7.

The association of SNP markers with geographic and climatic variables was also analyzed using LFMM analysis. The population structure-based association analysis revealed that 714 (1.48%) SNPs (FDR q value ≤0.01) were associated with geographic and environmental variables. The background population structure was modeled for five latent factors (K), which corresponded to the number of neutral genetic structure deteced by population structure analysis. The highest number of outlier SNPs was associated with mean temperature of the wettest quarter (BIO8; $n = 81$), followed by those SNPs associated by precipitation of driest quarter and precipitation of warmest quarter ($n = 51$). The lowest number of SNPs ($n = 1$) was observed for annual mean water vapor pressure (AMWV). A complete summary of the LFMM analysis is presented in Table S8.

Of the 329 SNPs detected by Model I, 28 loci (Table 2) were common to those discovered using the SAM approach (Table 2) and associated with longitude, AMWV, AMSR, BIO4, and BIO6. The allelic composition of outlier SNPs was further analyzed to determine its distribution across populations as illustrated using an example of snp46885-scaffold654-729925 in Fig. 3. The most prevalent major allele (A) of snp46885-scaffold654-729925, with a

frequency of 73.71%, was observed in most provinces of the country. The minor allele (C) was observed in two provinces of Limpopo and KZN.

In addition, four outlier SNPs associated with different geographic and climatic variables were common between SAM and LFMM (Table 3). Among the SNPs commonly detected was snp9199-scaffold1335-368697 (chromosome 19), snp14653-scaffold1591-11752 (chromosome 20), snp37582-scaffold460-108115 (chromosome 3), and snp36146-scaffold431-10361406 (chromosome 12).

Associated candidate genes and functional analysis

Redundancy analysis Model I detected 111 genes, while Models II and III detected 106 and 66 genes, respectively. Hundred and forty genes were detected by SAM and 55 were detected from LFMM analysis (Tables S4–S8). Only one gene (DGKB) was common for SAM, LFMM, and RDA (Model I), while 17 (CCSER1, CNTNAP2, COL6A6, CSMD1, JAK2, JAKMIP2, LMO4, LOC102188913, MARCH1, MBD5, NAV2, NLGN1, NTNG1, PRR16, SLCO1C1, SPATA16, and VAV3) were common between RDA Model I and SAM analyses. Two associated genes (LOC102181667 and GRID2) were common between LFMM and RDA Model I, while none were common for LFMM and SAM (Fig. 4). The functional implications of the associated goat candidate genes were investigated using the KEGG pathways (http://www.kegg.jp; last accessed 09 July 2016). The detected genes were involved in 205 pathways (Table S9).

Overall, the 478 adaptive genes associated with the most significant SNPs were characterized, and numerous pathways were considered particularly relevant in adaptation of indigenous goats to their environment and production system (Table S9).

Discussion

The study used the landscape genomic approach which, although attracting growing interest in studying population genetic structures, is still uncommon in investigating the genetics of local adaptation in indigenous livestock species. South Africa experiences considerable variation in climate and topography and is mainly classified as semi-arid. Population structure typically results from a combination of demographic and adaptive factors (Ometto et al. 2015). Mdladla et al. (2016a) demonstrated strong patterns of genetic differentiation of the South African goat population but the actual causative factors responsible for this structuring were unexplained. This study evaluated the contribution of the environmental and geographical factors to the genetic variation. Further, we investigated the genomic

regions associated with temperature and rainfall and provided new insights on the adaptive potential of South African goat populations.

Results of geographical clustering detected fluctuating levels of genetic differentiation within the goat populations across the geographic regions. Among the inferred genetic groups, a distinct cluster of Tankwa goats was aligned with a single geographic region of the Northern Cape province. This breed is genetically distinct from the commercial and domestic village goat populations (Kotze et al. 2014; Mdladla et al. 2016a) and showed low within-population diversity. The Tankwa have been in the Karoo Desert for over 80 years and, in 2009, about 60 animals were moved to the Carnavon Research Station for conservation (AGRINC 2013). The single genetic cluster to which the Tankwa goats were found could therefore be a result of founder effects when a small effective population size was used to establish

the Carnavon population. These founders have endured years of genetic isolation from farmed goat populations after the establishment of the conservation population. Mdladla et al. (2016a) reported low-diversity indices, i.e., observed heterozygosities (H_O) of 0.35 ± 0.33 and inbreeding coefficients (F_{IS}) of 0.15 ± 0.05 supporting expectations of small and inbred populations.

Almost all village goat ecotypes shared genomic clusters, which could be a result of the village goat production system that is characterized by communal breeding leading to high genetic diversity and gene flow among villages and populations in close proximity. Within population diversity indices of village goat ecotypes ranged from $H_O = 0.40 \pm 0.14$ to 0.41 ± 0.18 and $F_{IS} = 0.01 \pm 0.05$ to 0.07 ± 0.08 (Mdladla et al. 2016a). A mantel test for isolation by distance revealed a positive but low correlation between geographic and genetic distance ($r = 0.22$, $p = 0.04$). The weak impact of latitude on the genetic structure of ecotype populations concurs with clustering based on morphological characteristics reported by Mdladla et al. (2017). Detected clusters revealed geographically structured populations biased toward the longitude. The Zulu and Venda goats fall in the 30° longitude while the Tswana and Xhosa cluster is spread between 26° and 28° longitude. The commercial meat breeds clustered together probably as a result of the effects of similar breeding objectives and selection criteria.

The study further examined genomic regions under selection in the local goat populations. Our hypothesis was

Table 1 RDA of the contribution of both climate and geographic variables (Model I); climate alone (Model 1I), and geographic variables only (Model III) in genetic variation

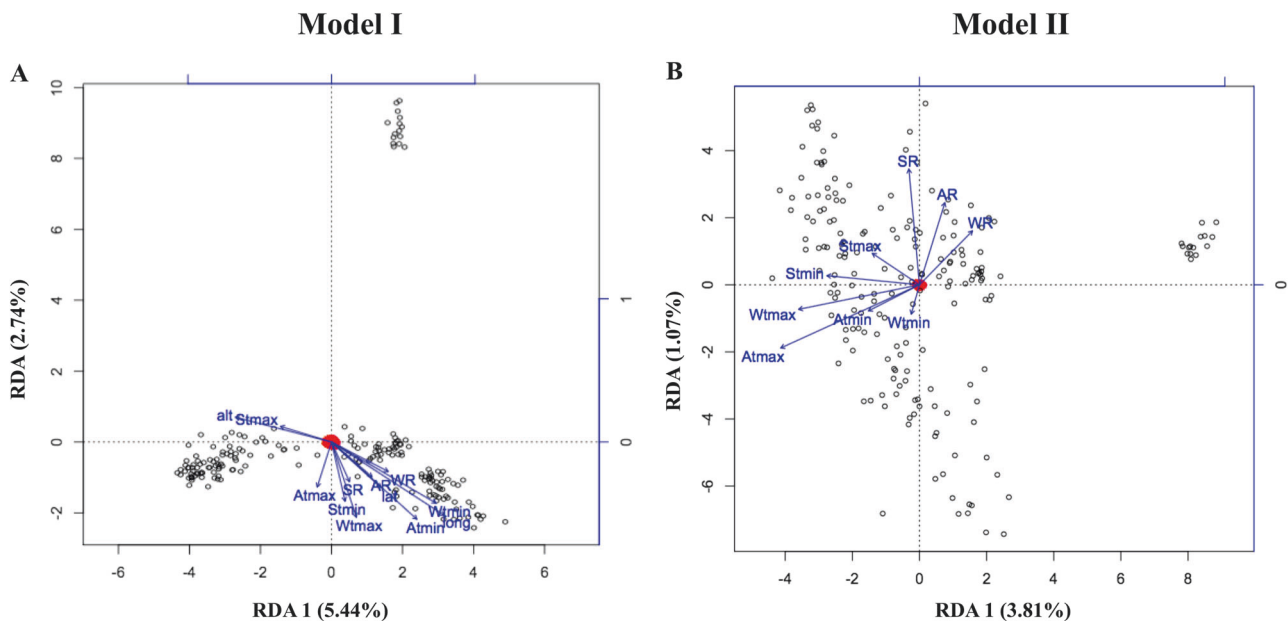| Models	Proportion			R^2_{adj}	P value
	Controlled	Uncontrolled	Conditioned		
Model I	0.22	0.78	–	0.11	0.001
Model II	0.17	0.78	0.05	0.07	0.001
Model III	0.01	0.78	0.21	0.01	0.001

Fig. 2 Redundancy analyses of contribution of **a** both climate and geographic variables (Model I); **b** climate alone (Model 1I) to genetic variation. Atmax = Maximum annual temperature; Atmin = Minimum annual tempera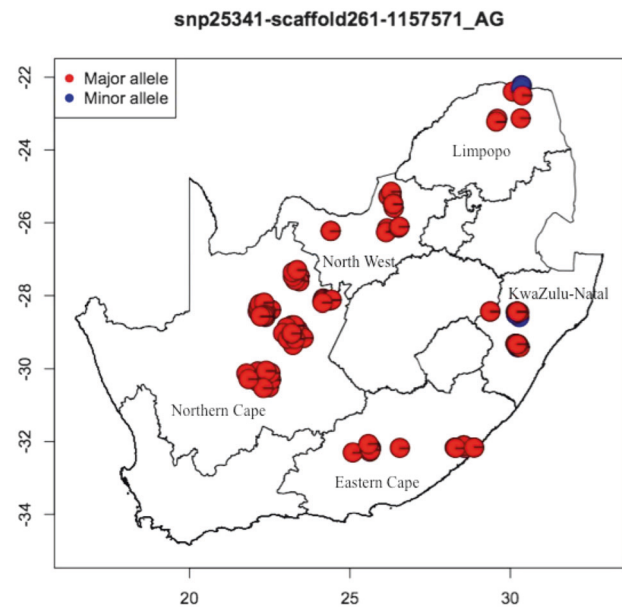ture; Stmax = maximum summer temperature; Stmin = minimum summer temperature; Wtmax = maximum winter temperature; Wtmin = minimum winter temperature; AR annual rainfall, WR winter rainfall, SR summer rainfall

Table 2 Summary of the SNP detected by both SAM and RDA model I

SNP	SAM	RDA score	CHR
snp46885-scaffold654-729925	Longitude	0.2172	3
snp50006-scaffold716-1302734	Longitude	0.1841	7
snp14775-scaffold1596-57767	AMWV	0.1811	29
snp59828-scaffold995-301815	AMSR	0.1802	7
snp37581-scaffold460-60946	AMWV	0.1722	3
snp12149-scaffold145-736257	Longitude	0.1697	1
snp28114-scaffold300-4928606	Longitude	0.1679	8
snp7370-scaffold1269-4675266	Longitude	0.1668	7
snp45013-scaffold614-1411491	AMWV	0.1666	2
snp54499-scaffold830-4996258	AMWV	0.1656	8
snp49500-scaffold706-1143916	Longitude	0.1625	4
snp46893-scaffold654-1069751	AMWV	0.1609	3
snp1672-scaffold1047-1386839	BIO6	0.1581	3
snp5866-scaffold1205-1993299	Longitude	0.1569	4
snp30755-scaffold34-2405052	Longitude	0.1562	27
snp46920-scaffold654-2228032	Longitude	0.1529	3
snp49503-scaffold706-1271483	AMSR	0.1528	4
snp926-scaffold1025-560411	AMWV	0.1527	6
snp54098-scaffold825-43831	BIO4	0.1491	29
snp12577-scaffold148-3245543	BIO4	0.1488	6
snp51082-scaffold740-2613701	Longitude	0.1462	5
snp12969-scaffold15-95710	Longitude	0.1457	5
snp45014-scaffold614-1443654	AMWV	0.145	2
snp30577-scaffold339-825223	BIO12	0.1448	7
snp28078-scaffold300-3412212	Longitude	0.1439	8
snp26382-scaffold276-3252186	BIO11	0.1401	1
snp7319-scaffold1269-2530432	AMSR	0.1384	7
snp8927-scaffold1321-128489	Longitude	0.1361	6

Fig. 3 Geographic locations of individuals with the major allele "A" (red) and minor "G" allele (blue) of the outlier SNP (snp46893-scaffold654-1069751) on a South African geographic map

Table 3 Summary of the SNP detected by both SAM and latent factor mixed model and the associated environmental variables

SNP	Environmental variables		Chromosome
	SAM	LFMM	
snp9199-scaffold1335-368697	AMWV	AMWS	19
snp14653-scaffold1591-11752	AMWV	Altitude	20
snp37582-scaffold460-108115	AMWV	BIO11	3
snp36146-scaffold431-10361406	Longitude	BIO12	12

that extreme and heterogeneous environmental parameters facilitate adaptive genetic variance that largely determines the survival of natural populations (McGaughran et al. 2014). Climatic variables, such as temperature and rainfall, have been shown to both directly and indirectly affect livestock production (Yang et al. 2016). Evidence of adaptation to the different environmental parameters, such as temperature, rainfall, and altitude, has been suggested in goats (Kim et al. 2016; Song et al. 2016) and sheep (Lv et al. 2014; Yang et al. 2016).

Variance partitioning using the two partial RDA models (Model II and III) indicated a higher contribution of climatic variables, highlighting that rainfall, temperature, and other climatic variables shape population genetic structures as a result of either their imposing selection pressure or by their historical association during the establishment of the populations. In other non-livestock species, genetic variation was highly associated with climatic than geographical

variables for example in *Arabidopsis thaliana* (Lasky et al. 2012), *Pristionchus pacificus* (McGaughran et al. 2014), and *Hordeum vulgare* L. (Abebe et al. 2015). The RDA full model (Model I) showed that BIO9, AMWV, longitude, BIO11, BIO6, AMSR, BIO4, altitude, and BIO2 have significant effects on genetic variation. With Model II, BIO9, BIO8, BIO10, BIO1, and BIO5 were the main contributors to the total genetic variation. Temperature of the driest quarter was observed as the first significant explanatory variables in Models I and II, indicating the importance of temperature as an environmental selective pressure in these goats. Further, an outlier method using the first RDA axes (Hancock et al. 2012; Lasky et al. 2012) was employed to investigate significant SNPs for all three models and 861 significant SNPs were detected that are located within 283 genes.
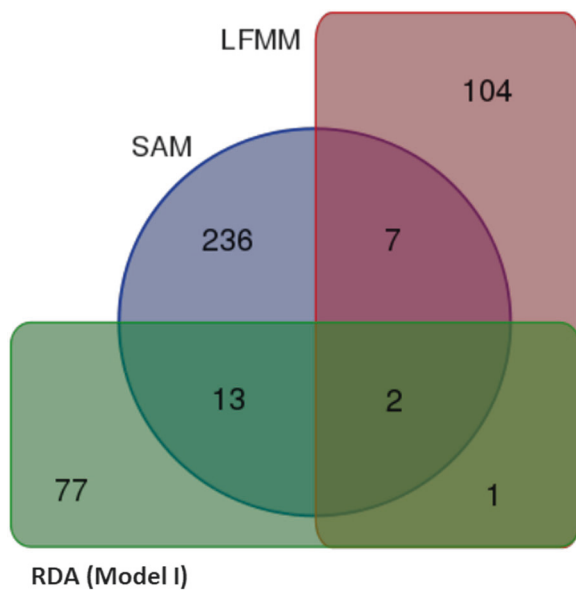
Fig. 4 Venn diagram for genes detected among the three methods SAM, LFMM, and RDA (Model I)

The two correlation-based landscape association genomic approaches, namely, SAM (Joost et al. 2007) and LFMM (Frichot et al. 2013), reported several significant loci associated with geographic and environmental variables. LFMM analysis returned significant loci mostly associated with temperature of the wettest quarter. The SAM analysis revealed that the highest number of SNPs were associated with longitude. The population structure analysis showed different genetic patterns of divergence among the goat populations to be associated with longitude. With LFMM analysis, mean temperature of the wettest quarter, precipitation of driest quarter, and precipitation of warmest quarter had the highest number of associated SNPs, suggestive of geographic regions as an important selection pressure.

By identifying SNPs associated with climatic variables, altitude, and geographic location, we were able to make inferences on the genetic basis of response to climatic change in indigenous goat populations. Kim et al. (2016) found eight selection sweep regions, distributed across chromosomes 3, 6, 7, 11, 12, 14, and 17 in Barki goats, associated with response to hot arid environment, using the integrated Haplotype Score (iHS) approach. Similarly, outlier SNPs associated with low rainfall and high temperature ranges spanning the same chromosomes were reported by Song et al. (2016) who emphasized the role of altitude in assembling adaptive genetic differentiation in Tibetan cashmere goats inhabiting high altitudes.

Pathways associated with metabolism, immunity, and response to heat and water scarcity were reported indicating adaptive mechanisms that allow indigenous goats to tolerate environmental pressure in their local production systems. Goats generally possess an efficient heat stress thermoregulatory response mechanism. Several genes associated with the heat stress response pathways, including circadian entrainment and circadian rhythm, were observed under selection in this study. The *PLCB1* gene involved in numerous pathways was also reported by Kim et al. (2016) for thermoregulation in hot arid environment. The other eight genes (*PRKG1*, *PLCB1*, *ADRA1D*, *EDNRA*, *CALCRL*, *ITPR2*, *BRAF*, and *CALD1*) that showed evidence of strong signals of selection are involved in the vascular smooth muscle contraction pathway. The cGMP-dependent protein kinase 1 (*PRKG1*) plays a role in the adaptive mechanisms to plateau environments in sheep (Yang et al. 2016).

Numerous candidate genes involved in energy regulation and metabolism were identified in the present study, including protein digestion and absorption, amino acid metabolism processes, secretion and taste transduction. An important attribute that goats present for surviving and producing in semi-arid and arid areas is their ability to utilize low-grade roughage (Silanikove et al. 1993; Daramola and Adeloye 2009) due to an efficient system of digesting fiber to enable maximal food intake and utilization and high tolerance for bitter substances (Goatcher and Church 1970; Casey and Van Niekerk 1988; Silanikove et al. 1996; Silanikove 2000). Indigenous goats also have the capacity to withstand prolonged periods of water deprivation (Silanikove et al. 1994) because of their ability to withstand and minimize water loss through urine and feces (Daramola and Adeloye 2009). In this study, genes involved in vasopressin-regulated water reabsorption and salivary secretion were observed to be under selection.

Response to disease and disease pathogens is important in developing countries, where populations are frequently exposed to various endemic diseases as well as disease outbreaks. Indigenous breeds usually display enhanced resistance to endemic diseases as compared to exotic ones reared in the same environment (Baker and Gray 2004). Recent evidence indicates that a number of indigenous goat breeds have natural resistance or tolerance to specific diseases (Malan 2000; Agyemang 2005; Ahmed and Othman 2006; Morrison 2007). The relative tolerance of the South African goats to prevalent heartwater in these populations was established in Mdladla et al. (2016b). Genes involved in response to diseases were correlated to environmental conditions indicative of the importance of immune response and disease resistance/tolerance as a survival mechanism in the South African goat population.

This study used a combination of multivariate statistics, landscape genomics, and population genomics-based analyses in order to increase accuracy and draw from the multidimensional effects of environmental factors in shaping the genetic variation in local indigenous goat

populations. The three statistical methods (SAM, LFMM, and RDA analysis) were used as a strategy to offer complementary comparisons between the different detection methods. Using multiple methods of population and landscape analyses, the study provided evidence of genetic subdivision within the village ecotype populations and demonstrated the presence of genetic subpopulations associated with specific geographic localities. Given that South African goats inhabit arid and semi-arid areas characterized by conditions of high temperature, water scarcity, and frequent droughts, the observed evidence for natural selection in our study was therefore envisaged. Overall, the study provided an insight into the genetic structure of South African goat populations, which could be due to a combination demographic and other evolutionary factors. The study further provided evidence of some environmental and geographic selection pressures driving evolution and local adaptation of South African goats to important prevailing production conditions. Further experiments involving targeted gene sequencing and expression analysis are necessary to confirm the causal mutations and the precise role of candidate genes in the process of genetic differentiation and local adaptation.

Data archiving

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.3402n.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

Abebe TD, Naz AA, Leon J (2015) Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.). Front Plant Sci 6:1–17

Agyemang K (2005) Trypanotolerant livestock in the context of trypanosomiasis intervention strategies. PAAT Technical and Scientific, FAO, Rome, Italy

AGRINC (2013) Tankwa feral goats a real cliffhanger or the end of the road. Retrieved from http://www.karoofoundation.co.za/images/Parliament/The%20feral%20goats%20of%20Tankwa%20-11%20October%202013.pdf

Ahmed S, Othman OE (2006) A PCR-RFLP methods for the analysis of Agyptian goat MHC class II DRB gene. Biotechnology 5:58–61

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664

Baker R, Gray G (2004) Appropriate breeds and breeding schemes for sheep and goats in the tropics. Worm control for small ruminants in tropical Asia, 63. ACIAR, Canberra, Australia, p 63–95

Brito LF, Kijas JW, Ventura RV, Sargolzaei M, Porto-Neto LR, Cánovas A, Feng Z, Jafarikia MSchenkel FS (2017) Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC Genome. https://doi.org/10.1186/s12864-017-3610-0

Casey N, Van Niekerk W (1988) The Boer goat. I. Origin, adaptability, performance testing, reproduction and milk production. Small Rumin Res 1:291–302

Colli L, Joost S, Negrini R et al. (2014) Assessing the spatial dependence of adaptive loci in 43 European and Western Asian goat breeds using AFLP markers. PloS ONE 9:e86668

DAFF (2014) Newsletter: livestock statistics, estimated livestock numbers in RSA (August 2013 and February 2014). http://www.daff.gov.za/daffweb3/Resource-Centre. Accessed 22 Sept 2014

Daramola J, Adeloye A (2009) Physiological adaptation to the humid tropics with special reference to the West African Dwarf (WAD) goat. Trop Anim Health Prod 41:1005–1016

Frichot E, Schoville SD, Bouchard G et al. (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol 30:1687–99

Goatcher WD, Church DC (1970) Taste responses in ruminants. IV Reaction of pygmy goats, normal goats, sheep and cattle to acetic acid and quinine hydrochloride. J Anim Sci 31:373–382

Gugger PF (2015). Exercises for landscape genomics workshop at UNAM-Morelia, Mexico. https://github.com/pgugger/LandscapeGenomicsWorkshop_Morelia/. Accessed 21 Jan 2018

Gwaze FR, Chimonyo M, Dzama K (2009) Prevalence and loads of gastrointestinal parasites of goats in the communal areas of the Eastern Cape Province of South Africa. Small Rumin Res 84:132–134

Hancock AM, Brachi B, Faure N et al. (2012) Adaptation to climate across the Arabidopsis thaliana genome. Science 334:83–86

Harper CK, Penzhorn BL (1999) Occurrence and diversity of coccidia in indigenous, Saanen and crossbred goats in South Africa. Vet Parasitol 82:1–9. https://doi.org/10.1016/S0304-4017(98)00266-0

Hijmans Robert J, Cameron Susan E, Parra Juan L., Jones Peter G., Andy Jarvis (2005) Very high resolution interpolated climate surfaces for global land areas Int Journal Climatol 25 (15):1965–1978

Joost S, Bonin A, Bruford MW et al. (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol 16:3955–3969

Kanehisa M, Goto S, Kawashima S et al. (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32:D277–D280

Kotze A, Grobler JP, van Marle-Köster E et al. (2014) The Tankwa Karoo National Park feral goat population: a unique genetic resource. S Afr J Anim Sci 44:43–48

Kijas JW, Ortiz JS, McCulloch R, James A, Brice B, Swain B, Tosser-Klopp G (2013) Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. Anim Genet 44:325–335

Kim E, Elbeltagy A, Aboul-Naga A et al. (2016) Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity 116:255–264

Lasky JR, Des Marais DL, McKay JK et al. (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. Mol Ecol 21:5512–5529

Lashmar SF, Visser C, van Marle-Köster E (2015) Validation of the 50k Illumina goat SNP chip in the South African Angora goat. S Afr J Anim Sci 45:56–59

Legendre P, Legendre L (1998) Numerical ecology. Elsevier, New York, NY

Lv FH, Agha S, Kantanen J et al (2014) Adaptations to climate-mediated selective pressures in sheep. Mol Biol Evol 31:324–343

Malan SW (2000) The improved Boer goat. Small Rluminant Res 36:165–170

Manel S, Joost S, Epperson BK et al. (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol Ecol 19:3760–72

McGaughran A, Morgan K, Sommer RJ (2014) Environmental variables explain genetic structure in a beetle-associated nematode. PLoS ONE 9:e87317

Mdladla K, Dzomba E, Huson H et al. (2016a) Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. Anim Genet 47:471–482

Mdladla K, Dzomba EF, Muchadeyi FC (2016b) Seroprevalence of Ehrlichia ruminantium antibodies and its associated risk factors in indigenous goats of South Africa. Prev Vet Med 125:99–105

Mdladla K, Dzomba EF and Muchadeyi FC (2017) Characterization of the village goat production systems in the rural communities of the Eastern Cape, KwaZulu-Natal, Limpopo and North West Provinces of South Africa. Trop Anim Health Prod 49:515–527. https://doi.org/10.1007/s11250-017-1223-x

Morrison JW (2007) A guide to the identification of the natural african indigenous veld goats of southern africa. Compiled for the Indigenous Veld Goat Breeders Society, Pretoria, South Africa

Ometto L, Li M, Bresadola L et al. (2015) Demographic history, population structure, and local adaptation in Alpine populations of cardamine impatiens and cardamine resedifolia. PLoS One 10: e0125199

Onzima RB, Mukiibi R, Ampaire A et al (2017) Between-breed variations in resistance/resilience to gastrointestinal nematodes among indigenous goat breeds in Uganda. Trop Anim Health Prod 49:1763–1769. https://doi.org/10.1007/s11250-017-1390-9

Pariset L, Joost S, Marsan PA et al. (2009) Landscape genomics and biased F ST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. BMC Genet 10:1

SAS Institute Inc (2013) SAS/ETS user's guide, version 13.1. SAS Institute Inc., Cary, NC

Silanikove N (2000) The physiological basis of adaptation in goats to harsh environments. Small Rumin Res 35:181–193

Silanikove N, Gilboa A, Perevolotsky A et al. (1996) Goats fed tannin-containing leaves do not exhibit toxic syndromes. Small Rumin Res 21:195–201

Silanikove N, Nitsan Z, Perevolotsky A (1994) Effect of daily supplementation of polyethylene glycol on intake and digestion of tannin-containing leaves (Ceratonia siliqua) by sheep. J Agric Food Chem 42:2844–7

Silanikove N, Tagari H, Shkolnik A (1993) Comparison of rate passage fermentation rate and efficiency of digestion of high fiber diet in desert black Bedouin goats as compared to Swiss Saanen goats. Small Rumin Res 12:45–60

Song S, Yao N, Yang M et al. (2016) Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus). BMC Genome 17:122

Tosser-Klopp Gwenola, Bardou Philippe, Bouchez Olivier, Cabau Cédric, Crooijmans Richard, Dong Yang, Donnadieu-Tonon Cécile, Eggen André, Heuven Henri C. M., Jamli Saadiah, Abdullah Johari Jiken, Klopp Christophe, Cynthia T. Lawley, John McEwan, Patrice Martin, Carole R. Moreno, Mulsant Philippe, Nabihoudine Ibouniyamine, Pailhoux Eric, Palhière Isabelle, Rupp Rachel, Sarry Julien, Sayre Brian L., Tircazes Aurélie, Wang Jun, Wang Wen, Zhang Wenguang, Liu Zhanjiang (2014) Design and Characterization of a 52K SNP Chip for Goats PLoS One 9:e86227

Van Den Wollenberg AL (1977) Redundancy analysis an alternative for canonical correlation analysis. Psychometrika 42:207–219

Visser C, Lashmar SF, Van Marle-Köster E, Poli MA, Allain D (2016) Genetic diversity and population structure in South African, French and Argentinian angora goats from genome-wide SNP data PLoS One 11:e015435310.1371/journal.pone.0154353

Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, Liu MJ (2016) Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol 33:2576–2592