Ethiopian indigenous goats offer insights into past and recent demographic dynamics and local adaptation in sub-Saharan African goats

Tarekegn GM[1,2#], Khayatzadeh N[3#], Liu B[4], Osama S[5], Haile A[6], Rischkowsky B[6], Zhang W[4], Tesfaye K[7], Dessie T[8], Okeyo AM[9], Djikeng A[10], Mwacharo JM[6,10*]

[1]Department of Animal Production and Technology, School of Animal Sciences and Veterinary Medicine, Bahir Dar University, Ethiopia

[2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden

[3]Department of Sustainable Agricultural Systems, Division of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria

[4]Inner Mongolia Agricultural University, Hohhot, China

[5]The University of Queensland, Queensland, Australia

[6]Small Ruminant Genomics, International Centre for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia

[7]Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Ethiopia

[8]International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia

[9]International Livestock Research Institute (ILRI), Nairobi, Kenya

[10]Animal and Veterinary Sciences Group, SRUC and Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, Easter Bush, Midlothian, UK

[#]Contributed equally to this work

[*]Correspondence: TGM: yafetgetinet@gmail.com; MJM: j.mwacharo@cgiar.org

Abstract

Knowledge on how adaptive evolution and human socio-cultural and economic interests shaped livestock genomes particularly in sub-Saharan Africa remains limited. Ethiopia is in a geographic region that has been critical in the history of African agriculture with ancient and diverse human ethnicity and bio-climatic conditions. Using 52K genome-wide data analysed in 646 individuals from 13 Ethiopian indigenous goat populations, we observed high levels of genetic variation. Although runs of homozygosity (ROH) were ubiquitous genome-wide, there were clear differences in patterns of ROH length and abundance and in

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/eva.13118

effective population sizes illustrating differences in genome homozygosity, evolutionary history, management and, past and recent population demographic history and dynamics. Phylogenetic analysis incorporating patterns of genetic differentiation and gene flow with ancestry modeling, highlighted past and recent intermixing and possible two deep ancient genetic ancestries that could have been brought by humans with the first introduction of goats in Africa. We observed four strong selection signatures that were specific to Arsi-Bale and Nubian goats. These signatures overlapped genomic regions with genes associated with morphological, adaptation, reproduction and production traits due possibly to selection under environmental constraints and/or human preferences. The regions also overlapped uncharacterised genes, calling for a comprehensive annotation of the goat genome. Our results provide insights into mechanisms leading to genome variation and differentiation in sub-Saharan Africa indigenous goats.

Keywords: Autozygosity, Diversity, Effective population size, Genome dynamics, LD decay, Runs of homozygosity, Selection Signatures

Introduction

Threats to biodiversity, and projected future food demands and climatic conditions, underscores the urgent need to characterize, monitor and maintain agricultural biodiversity. Data on molecular changes has been critical in understanding genome architecture, maintaining biodiversity and fitness and exploring impacts of genome evolution. This has been made possible through the investigation of genetic diversity and structure, demographic dynamics (Bosse et al. 2012), and assessment of autozygosity under varying degrees of reproductive isolation and inbreeding (Kirin et al. 2010; Peripolli et al. 2016; Ceballos et al. 2018). As livestock populations dispersed from their centers of domestication, they encountered diverse environments with unique climatic, anthropological and biophysical limitations. These animals responded to the niche-specific environmental pressures through behavioural and biological adjustments. The former provided short-term buffer against many of the stressors, and the latter ensured long-term survival through physiological acclimatization and/or genetic adaptation.

Among domestic ungulates, goats have the longest socio-cultural and economic co-existence with humans. They were domesticated around 11,000 years ago in the Fertile Crescent of Southwest Asia and adjacent areas (Zeder and Hesse 2000). Since their domestication, goats have been contributing to human cultural and socio-economic transformations which have shaped ancient and modern human civilizations and the goat's genomes. Their rusticity and resilience have allowed them to cope and adapt to a wide range of environments. However, the evolutionary history of goats is complex with two contrasting hypotheses explaining their domestication and little geographic pattern in their maternal diversity. Luikart et al. (2001) observed at least six distinct mtDNA haplogroups with little geographic partitioning. They interpreted this to be the consequence of multiple domestication events and ease of translocation of goats through commerce and trade. Naderi et al. (2008) on the other hand suggested that such diversity of mtDNA haplogroups was compatible with a single domestication event, followed by a phase of human management of wild or semi-domesticated goats comprising multiple mtDNA lineages, prior to their global dispersion. Evidence from the analysis of ancient goat genomes suggest the domestication of multiple divergent wild goats in a dispersed manner resulting in distinct Neolithic populations that contributed disproportionately to modern goat genomes (Daly et al. 2018). The initial diffusion of goats to Africa is also complex. Archeological evidence suggests at least three entry points into the continent (Gifford-Gonzalez and Hanotte 2011). At least two mtDNA haplogroups have also been observed in Kenyan (Kibegwa et al. 2015), Ethiopian (Tarekegn et al. 2018), Sudanese (Sanhory et al. 2014) and Egyptian (Naderi et al 2008) indigenous goats, suggesting the influence of at least two independent maternal lineages.

Paleogenomic evidence has shown that the northeastern and Horn of Africa region were at the center of a vast web of maritime and terrestrial routes of ancient and modern trade and thus a major entry point of domesticates into the continent (Marshall 2000; Fuller and Boivin 2009). The eastern and Horn of Africa landscape exhibit considerable changes in elevation within short geographic distances due to complex volcano-tectonic activity over the past millennia (Mohr 1971). In particular, the Ethiopian highlands, which range from 125 meters below sea level in the Afar depression to altitudes exceeding 4,000 meters above sea level in the Arsi-Bale mountains, form an extensive uplifted plateau, that is delimited by pronounced escarpments (Umer et al. 2007). The altitudinal gradients result in a wide range of agro-eco-climates and production environments. The region is also characterized by human ethnic diversity of ancient origin that has always been associated with livestock husbandry. Paleo-climatic data has also revealed periods of prolonged and severe droughts in the region with profound impacts (Verschuren et al. 2000). These factors make the region particularly attractive for investigating how the genomes of indigenous livestock might have been shaped by past and recent events. Here, we generated 52K SNP genotypes in 13 populations of Ethiopian indigenous goats. The data were used to investigate past and recent demographic dynamics through the analysis of genome-wide diversity and admixture, autozygosity and evidence for selection.

Materials and Methods

Animals and genotypes: The animals used in this study were provided by farmers and pastoralists who participated in the study by agreeing to have their animals sampled. The sampling was done following standard guidelines and procedures of the Ministry of Livestock and Fisheries of the Federal Democratic Republic of Ethiopia. No ethical approvals were required at the time.

In total, 646 unrelated goats from 13 populations (FARM-Africa 1996; Table 1) were sampled by collecting whole blood from two mature unrelated animals per flock via jugular venipuncture into EDTA vacutainers. Prior to sampling, and in the absence of written pedigree records, flock owners were interviewed in detail on the extent of genetic relationships between their animals. Genomic DNA was recovered with the salting out procedure (Shinde et al. 2008). Genotyping was performed with the Illumina® Caprine SNP 52K panel featuring 53,347 SNPs (single nucleotide polymorphisms). Using PLINK v1.9 (Purcell et al. 2007), SNPs with call rates < 95%, minor allele frequency (MAF) < 0.01, unknown and redundant genome coordinates, non-autosomal location and those not in Hardy-Weinberg equilibrium (HWE; $P < 0.001$), and samples with > 5% missing genotypes were filtered from the dataset. This left 44,723 SNPs and 628 goats

for analysis. A subset of these samples were previously analysed for mtDNA D-loop variation (Tarekegn et al. 2018).

Data analysis

To estimate genetic diversity, observed (H_O) and expected (H_E) heterozygosity, proportion of polymorphic SNPs (P_N) and average MAF were calculated with PLINK v1.9. The average pair-wise genetic distances between individuals within a population were also calculated with PLINK v1.9. Higher values indicate elevated genetic distances and thus low genetic relationships between individuals. The average proportion of alleles shared between two individuals was calculated as D_{ST} using PLINK v1.9 with the "–genome" command line:

$$D_{ST} = \frac{IBS_2 + (0.5 \times IBS_1)}{N}$$

IBS_1 and *IBS_2* represent the number of loci which share either one or two alleles that are identical by state in pair-wise comparisons between individuals, respectively and N is the number of loci tested. The genetic distance between all possible pairwise combinations of individuals was calculated as D = 1 - D_{ST}.

Detection of ROH was performed by invoking the "*--homozyg*" option in PLINK v1.9. To account for medium marker density, the following set of conditions were imposed: minimum length of ROH = 1 Mb, number of SNPs present in the RoH = 5, number of missing SNP in the ROH = 1, minimum allowed density of SNPs within a run = 1 SNP/100 Kb, number of heterozygous SNPs in each ROH = 1, and maximum gap between consecutive homozygous SNPs = 1 Mb. Descriptive statistics for the total ROH were also calculated and Pearson correlation coefficients between average ROH length and the average number of ROH per population were computed with R (https://www.r-project.org/). The ROH were classified into four length categories 2 Mb, 4 Mb, 8 Mb and 16 Mb.

For each population, runs of homozygosity (F_{ROH}; Kim et al. 2013) and excess of homozygosity (F_{HOM}; Wright 1922) statistics, were calculated as estimators of genomic inbreeding. F_{ROH} was derived for each individual following McQuillan et al. (2008):

$$F_{\mathrm{ROH}} = \frac{\sum_{\mathrm{K}}(\mathrm{Length}(\mathrm{ROH}_{\mathrm{K}}))}{\mathrm{L}}$$

the numerator represents the sum of ROH per animal above a certain criteria length and the denominator (L) is the total length of the genome covered by autosomal markers (McQuillan et al. 2008). The F_{HOM} statistic was derived using the formulae:

$$F_{HOM} = {}^{O_{HOM} - E_{HOM}}/_{1 - E_{HOM}}$$

O_{HOM} and E_{HOM} represent the observed and expected homozygosity for each population, respectively.

To determine the extent of linkage disequilibrium (LD) between adjacent SNPs, the r^2 statistic was calculated for each pair of loci using the formulae:

$$r^2(fA, fB, fAB) = \frac{[f(AB) - f(A)f(B)]^2}{fA(1-fA)fB(1-fB)}$$

where *f(AB)* is the frequency of haplotypes having allele A at locus 1 and allele B at locus 2 and *f(A)* and *f(B)* are the observed allele frequencies at each locus (Hill and Robertson 1968). For the calculation, we used the "*--ld-window 9999 --ld-window-kb 1000 --ld-window-r² 0*" command line in PLINK v1.9. These settings allowed the analysis of SNPs that were not more than 9,999 SNPs apart, set the window size to 1,000 kb and used all SNPs in the analysis.

The r^2 values were used to model changes in effective population sizes (N_e) over generations for each population and across the 13 populations using the SNeP tool (Barbato et al. 2015). Sved (1971) described the relationship between LD (r^2), N_e and c (recombination rate) using the equation:

$$E(r2) = 1/(\alpha + KNec) + 1/n$$

$$Ne = 1/(r2 - 1/n)kc - 2/kc$$

Here r^2 is the LD between different markers, N_e is the effective population size, c is the genetic distance between various markers measured in Morgans, n is the chromosome experimental sample size, α is a correction for the occurrence of mutations (Ohta and Kimura 1971; α = 1 in the absence of mutation and α = 2 if mutation is considered), k = 4 for autosomes and k = 2 for the X chromosome. In contemporary studies, physical distance is used instead of genetic distance to estimate population size. A physical distance of 100 Kb is approximately equivalent to a genetic distance of 0.1 cM (or 1 cM ≈ 1 Mbp; Dumont and Payseur, 2008). When population size is changing linearly, the expectation of chromosome segment homozygosity is ~1/(4Ntc+1), where Nt is the population size 1/(2c) generations ago (Hayes et al. 2003). SNPs with a MAF > 0.05 were used to estimate N_e.

The underlying population genetic structure was assessed with principal component analysis (PCA; Jolliffe 2002), ADMIXTURE v1.3.0 (Alexander et al. 2009), TreeMix v1.13 (Pickrell and Pritchard 2012) and NetView P v0.4.2.5 (Steinig et al. 2016). PCA was performed with the "--pca" command line in PLINK v1.9. To investigate past and recent population admixture, the number of genetic clusters inherent in the

genomes of the study populations was determined with ADMIXTURE v1.3.0. The number of genetic clusters tested ranged between 2 and 13, and CLUMP v1.1.2 and DISTRUCT v1.1 were used to process and generate the graphics, respectively. To investigate patterns of population splits and cross-population gene flow, the maximum-likelihood tree-based approach incorporating admixture events and implemented in TreeMix v1.13 was used. The phylogenetic tree was built without rooting, and the migration events "m" modelled as edges were added to the phylogeny until the model explained at least 99% of the variance in ancestry. The value of "m" with the highest log-likelihood following six replicate runs of TreeMix v1.13 was chosen as the optimal.

To reveal fine-scale population stratification independent of *a priori* ancestry information, network analysis was carried out using the NetView v.0.4.2.5 (Neuditschko et al 2012; Steinig et al 2016). NetView explores network topologies using a single user-defined threshold parameter, the number of mutual nearest neighbors (k). Fewer individuals are considered nearest neighbors at small values of k, leading to only genetically more similar individuals being connected and highlighting fine-scale structure in the dataset. We generated a population network based on shared allele distance matrix (1-identity-by-state (IBS)) generated with PLINK v.1.9. The network was constructed with the super-paramagnetic clustering (SPC) algorithm and Sorting Points Into Neighbourhoods (SPIN) software, which computes the maximum number of nearest neighbours for a given individual (Neuditschko et al. 2012; Steinig et al 2016). The network was visualized and edited in the Cytoscape v.2.8.3 network construction package (Smoot et al. 2011). SPC and CYTOSCAPE are implemented in NetView. NetView requires a specification of the maximum number of nearest neighbours (k-NN) that an individual can have. In this study, the number of k-NN values tested were between 5 and 120.

The software hapFLK v.1.2 (Fariello et al. 2013) was used to implement the hapFLK algorithm, which can be applied to un-phased genotypic data, to detect signatures of selection while accounting for haplotype structure and varying N_e. The implementation required the construction of a neighbor-joining (NJ) tree using a kinship matrix. Pairwise Reynolds genetic distances (Reynolds et al. 1983) were calculated and converted to a kinship matrix with R scripts provided in the hapFLK webpage (https://forge-dga.jouy.inra.fr/projects/hapflk). In the construction of the NJ tree, no outgroup population was defined but the software was prompted to use all populations and the midpoint as outgroup. The kinship matrix captured the population structure, which was used to model covariance matrix of allele frequencies whereas a multi-point LD model was used to create haplotype clusters on each chromosome (set to 13 (-k, 13) per chromosome). This was determined using cross-validation based estimation in fastPHASE 1.4.0

(Scheet and Stephens 2006) using the setting -KL10 -KU20 -Ki3. The hapFLK statistic was calculated for each SNP as the average of 20 expected maximization runs fitting the LD model. *P*-values were then computed based on a Chi-square distribution using a python script (https://forge-dga.jouy.inra.fr/document/588). To limit false positives, a q-value threshold of 0.01 was applied to control for false discovery rate (FDR). Putative selection signatures were defined by the regions with a threshold of $P < 0.001$.

To further pinpoint loci under selection, we also used the cross-population extended haplotype homozygosity (XP-EHH) test (Sabeti et al. 2002; 2007) to make comparisons between the 13 populations. The XP-EHH assesses haplotype differences between two populations and is designed to detect alleles that have increased in frequency to the point of fixation or near-fixation in one of the populations (Sabeti et al. 2007; Pickrell et al. 2009). The test uses the integrated EHH (*iHH*) of a core SNP in two populations, A and B, rather than two alleles in a single population. The unstandardized XP-EHH statistic is calculated as:

$$unstandardized\ XP-EHH\ =\ ln(iHHA/iHHB)$$

where iHH_A and iHH_B are the integrated EHH of a given core SNP in population A and B, respectively. A large positive value of XP-EHH at a locus suggests selection in population A and in the case of a negative value, selection in population B. Here, we used the software developed by Pickrell et al. (2009) to estimate the unstandardized XP-EHH statistics for all SNPs in all the 13 populations with cross-population comparison of each population against the other 12. The unstandardized XP-EHH statistics were standardized using their means and variances in each comparison. Because previous studies showed that the standardized XP-EHH statistics follow standard normal distribution (Sabeti et al. 2007; Ma et al. 2014), *P*-values were estimated using the standard normal distribution. For each cross-population comparison, we determined the regions under selection based on the threshold *P*-value < 0.001.

Using the NCBI map viewer, the regions revealed by hapFLK and XP-EHH were annotated using the ARS1 release 102 goat reference genome assembly (Bickhart et al. 2017). Functional enrichment and gene ontology (GO) analysis were performed with Panther14.1 (Mi et al. 2019) using *Bos taurus* as the background. We performed text mining with STRING 11.0 (https://string-db.org/) to identify phenotypes and protein-protein interaction networks associated with the candidate genes.

Results

The genetic diversity indices (mean ± standard deviation (SD); Table 1) that were calculated for each population show that Keffa had the lowest values of H_O (0.348 ± 0.143), H_E (0.361 ± 0.130), P_N (0.978) and

D_{ST} (0.289 ± 0.018). The highest values were observed in Afar (H_O = 0.382 ± 0.124), Nubian (H_E = 0.391 ± 0.106; D_{ST} = 0.315 ± 0.019) and Hararghe Highland (P_N = 0.997). The lowest values of inbreeding were observed in Long-eared Somali (F_{HOM} = -0.002 ± 0.03) and Gondar (F_{ROH} = 0.006 ± 0.014), respectively. Overall, Nubian had the highest values of F_{HOM} (0.054 ± 0.118) and F_{ROH} (0.087 ± 0.116). F_{HOM} showed a strong positive correlation (r = 0.978) with F_{ROH}.

The highest and lowest proportion of SNPs with low and high MAF, respectively were observed in Keffa (Figure 1a). Woyto-Guji and Western lowland/Gumez had the lowest variation in MAF. Figure 1b shows the proportion of ROH for the four genome length categories (2 Mb, 4 Mb, 8 Mb, 16 Mb) examined in this study. Gondar had the highest proportion of short ROH (2 Mb), followed by Western highland/Agew, Western lowland/Gumez, Woyto-Guji, Keffa and Long-eared Somali, respectively. Populations with the highest proportion of long ROH (16 Mb) were Small-eared Somali, Hararghe Highland and Nubian. Abergelle, Ambo, Afar and Arsi-Bale had a high proportion of ROH segments of intermediate length.

The average number of ROH segments (mean ± standard deviation) per animal ranged from 2.46 ± 2.99 (Gondar) to 16.85 ± 19.84 (Nubian) (Table 2). Woyto-Guji had the next lowest mean number (3.51 ± 2.99) and Keffa presented the next highest mean number (12.13 ± 12.67). The maximum number of ROH segments ranged from 14 (Woyto-Guji) to 70 (Nubian). Keffa presented the next highest maximum number (44) of ROH segments. The average length of the genome comprising of ROH segments ranged from 4,554.92 ± 2,240.02 Kb (Gondar) to 10,072.83 ± 8,267.14 Kb (Short-eared Somali). The shortest and longest lengths of the genome comprising of ROH segments were observed in Hararghe Highland (2,236.95) and Short-eared Somali (40,669.67), respectively. The average number of SNPs within an ROH segment was lowest in Gondar (84.61 ± 40.18) and highest in Short-eared Somali (187.53 ± 151.62). This presented the minimum and maximum values of 40 SNPs in Afar and 749 SNPs in Short-eared Somali, respectively. The average SNP density (SNPs per Kb) were similar across the 13 populations (53-54 SNPs/Kb), although the range was large. The proportion of homozygous sites averaged 99% across all populations (Table 2).

The average r^2 value (Table 3) ranged from 0.028 (Woyto-Guji) to 0.051 (Keffa). The trends in LD decay over genomic distances reveal high LD over short distances, which decays rapidly with distance (Figure 2a). LD decays more rapidly in Woyto-Guji but slower in Keffa, Western highland/Agew and Western lowland/Gumez.

One thousand (1000) generations ago, Nubian and Keffa had the highest and lowest N_e, respectively (Table 3). Thirteen (13) generations ago, the lowest N_e was observed in Gumez and highest in Hararghe Highland. The overall composite trend over the past 1000 generations (Figure 2b) reveals a gradual increase in N_e to about 600 generations followed by a rapid increase to about 150 generations, and then a sharp decline to present time. Each population however showed contrasting trends (Supplementary Figure S1). A trend similar to the composite characterizes Woyto-Guji, Short-eared Somali and Hararghe Highland. A continuous and gradual decline characterizes Arsi-Bale, Keffa, Western lowland/Gumez, Western highland/Agew and Nubian. The N_e for Abergelle, Gondar, Ambo and Long-eared Somali increases gradually to 200-300 generations ago after which it declines rapidly to present time. The N_e of the Afar population increases gradually to about 300 generations ago, stabilizes to 150 generations, then declines rapidly to present time.

PCA projected seven genetic clusters (Figure 3a). Cluster-1 comprised Abergelle, Ambo, Gondar, Western highland/Agew and Western lowland/Gumez. Cluster-2 comprised Hararghe Highland, Long- and Short-eared Somali. Clusters-3, -4, -5, -6 and -7 comprised respectively Afar, Arsi-Bale, Keffa, Nubian and Woyto-Guji which were depicted as distinct genetic entities. In agreement with the high D_{ST} value, Nubian individuals spread out across the fourth quadrant of the PCA suggesting high intra-population variation. TreeMix (Figure 3b) replicated the PCA clusters and showed two gene flow events, one from Long-eared Somali to Woyto-Guji and the other from Woyto-Guji to Keffa.

The optimal value of *K* following ADMIXTURE analysis could not be determined with certainty. At the first instance, the CV error was lowest at $4 \leq K \leq 5$. It then increased slightly at $K = 6$ before declining to the same lowest value observed at $4 \leq K \leq 5$, at $K = 7$, 8, 9 and 10 (Figure 3c inset). If the increase in CV error at $K = 6$ reflects lack of stability in allocating the genetic backgrounds, then the values of *K* at 7, 8, 9 and 10 may explain the variation in the dataset. Taking $K = 7$ as the most optimal, it supports the PCA and TreeMix but with higher resolution. At this *K* value, Keffa is the only population with one uniform genetic background which is shared with Woyto-Guji. The other major genetic backgrounds are observed in Abergelle, Ambo, Long-eared Somali, Nubian and Gumez. The seventh genetic background is unique to a few individuals of Nubian. Other than these individuals of Nubian and Keffa, the other populations exhibit variable proportions of genome admixture of at least two genetic backgrounds. Although this admixture pattern is repeated at $8 \leq K \leq 10$, the genomes of Afar and Woyto-Guji show gradual reduction in admixture while the opposite is observed for Arsi-Bale and Nubian. By revealing a common genome

background between Long-eared Somali and Woyto-Guji and between Woyto-Guji and Keffa at $7 \leq K \leq 10$, ADMIXTURE corroborates TreeMix which showed migration events between these populations.

The NetView results (Supplementary Figure S2) showed that from 75 ≤ k-NN ≤ 120 only four individuals, two each of Keffa and Long-eared Somali, were unassigned. The major clusters became evident, at the first instance, from k-NN = 55, when the 13 populations appeared to separate into two broad groups, hereby designated as G1 and G2 (Figure 4a). G1 comprised of Abergelle, Ambo, Gondar, Western highland/Agew and Western lowland/Gumez. G2 comprised Afar, Arsi-Bale, Hararghe Highland, Keffa, Long-eared Somali, Nubian, Short-eared Somali and Woyto-Guji. These two groups are consistently retained up to k-NN = 120. This prompted us to look at the results of ADMIXTURE at $K = 2$ (Figure 4b; 4c). As expected, it revealed two genome backgrounds, hereby named GB1 and GB2. GB1 occurs at a frequency of >70% in Abergelle, Ambo, Western highland/Agew, Western lowland/Gumez, Short and Long-eared Somali. Abergelle, Ambo, Western highland/Agew and Western lowland/Gumez comprise G1 of NetView. GB2 is found at a frequency of >70% in Arsi-Bale, Afar, Gondar, Hararghe Highland and Nubian, all the populations found in G2 of NetView except Gondar which occurs in G1. Long- and Short-eared Somali occur in G2 of NetView but in GB1 of ADMIXTURE. Although Keffa and Woyto-Guji occur in G2, in ADMIXTURE they show almost an equal proportion of GB1 and GB2 in their genomes. The minor discordance between NetView and ADMIXTURE could be due to differences in the way they are structured to allocate populations into clusters. Despite this minor difference, the NetView results led us to hypothesize that the analysis is revealing, possibly two ancestral genomes that could have arrived with the initial introduction of goats in the Horn of Africa.

Seventy-eight pair-wise comparisons were performed with hapFLK v1.2 software and XP-EHH test, assuming that each population was adapted to specific circumstances in its home range. Although the two approaches identified multiple regions under selection across the genome (Supplementary Figure S3 and S4), here we focused only on the regions that overlapped between the two approaches. The regions revealed by XP-EHH fell within the boundaries of the ones identified by hapFLK. The most significant and consistent signals were revealed in pairwise comparisons involving Arsi-Bale and Nubian with the other 11 populations (Supplementary Figure S3 and S4). The ones involving Arsi-Bale were on chromosome (CHI) 6 (8,533,828 – 17,357,213 bp; average size: 3,621,429 ± 2,264,393 bp) and CHI12 (53,982,423 – 62,977,761 bp; average size: 6,978,324 ± 952,467.8 bp). The ones involving Nubian were on CHI8 (54,717,960 – 63,857,851 bp; average size: 4,912,265 ± 1,773,863 bp) and CHI13 (49,975,695 – 67,760,577 bp; average size: 10,147,644 ± 2,916,617 bp). Comparative analysis between Arsi-Bale and Nubian revealed the four

putative signatures (Figure 5a; 5b), suggesting they are specific to the two populations. Arsi-Bale goats carry dense hairy coats and reside at an altitude of ≥ 4,000 meters above sea level (www.dagris.info/). Nubian goats have long haired coats and reside at low altitude drylands in northwestern Ethiopia, northern Sudan and western Eritrea (www.dagris.info/). High altitudes present a hypoxic cold environment. Low altitude drylands present an environment that is characterized by grazing stress. We define grazing stress as a term that encompasses complex interacting biophysical stressors including heat, physical exhaustion, direct solar radiation, and unavailability of feed (quality and quantity) and water. The selection signatures observed in Arsi-Bale and Nubian led to the hypothesis that putative selection may have resulted in adaptive and phenotypic divergence in the two populations.

We annotated the four candidate regions based on the ARS1 release 102 goat reference genome assembly. We found a total of 123 (Arsi Bale = 29; Nubian = 94) unannotated genes i.e. prefixed with "LOC/ENSCHIG" in the candidate regions. Although potentially under selection, these unannotated genes were not included in the functional enrichment and ontology analysis. In total, 20 (CHI6) and 45 (CHI12) genes were present in the Arsi-Bale regions, while 62 (CHI8) and 244 (CHI13) were observed in the Nubian regions (Supplementary Table S1). Functional enrichment and GO analysis were performed separately for each population-specific gene-sets using Panther and STRING. The 65 Arsi-Bale and 306 Nubian candidate genes showed greater enrichment in biological processes relating to cellular processes (GO:0009987), metabolic processes (GO:0008152) and biological regulation (GO:0065007). Text mining was performed with STRING to determine gene functions from published literature. It identified that several of the candidate genes are linked with traits of adaptive, economic and functional significance (Table 4). Some genes such as *CDX2*, *NBEA* and *PITX* (Arsi-Bale) and *PROCR*, *CTCFL*, *NNAT* and *PDYN* (Nubian), are involved in multiple functions.

Discussion

Genetic variation and marker polymorphisms

We present findings from an analysis of genome diversity, structure and dynamics of indigenous goats in Ethiopia, at the route of one of the most ancient gateways of domesticates into Africa. Ethiopia is characterized by ancient and modern human ethnic diversity that has long been associated with livestock husbandry and diverse agro-eco-climates, which may have influenced the genome architecture of indigenous livestock. The 13 study populations retained high levels of genome diversity (mean H_O and H_E above 0.348) and within population genetic variation (D_{ST} > 0.289). Although these values are high, they are close to those reported in other indigenous goats (Nicoloso et al. 2015; Kim et al. 2016; Manunza et al.

2016; Mdladla et al. 2016; Onzima et al. 2018; Berihulay et al. 2019). Keffa was the least diverse and Nubian was the most diverse population. Extensive genome admixture incorporating at least two genetic backgrounds at $7 \leq K \leq 10$, was observed in all but Keffa. This admixture could explain the high diversity and genome variability in the Ethiopian indigenous goats. It is the result of past and modern-day socio-economic and cultural exchanges involving livestock and natural selection which retains genetic variation.

ROH and inbreeding

Genomic inbreeding coefficients, such as F_{ROH} and F_{HOM}, are more accurate at estimating autozygosity and detecting past and recent inbreeding compared to estimates derived from pedigree data (Ferenčaković et al. 2013; Curik et al. 2014; Knief et al. 2017). Furthermore, in the absence of genealogical information, molecular data can be used to infer population history and inbreeding (Kim et al. 2015; Zavarez et al. 2015). The latter is important especially for African indigenous livestock which often lack written pedigree records. The strong and positive correlation (r = 0.978) between F_{ROH} and F_{HOM} observed in our study corroborates previous findings in pigs (Zhang et al. 2014) and cattle (Mastrangelo et al. 2016). It confirms that the extent of a genome under ROH can be used to predict the proportion of the genome that is identical by descent (IBD). A high and positive correlation between F_{ROH} and conventional pedigree-based estimates of inbreeding (F_{PED}) has been reported (Purfield et al. 2012; Ferenčaković et al. 2013; Martikainen et al. 2017) confirming F_{ROH} as an appropriate estimator of IBD alleles. The importance of understanding and quantifying genome-wide autozygosity has also been highlighted through correlations of F_{ROH} with inbreeding depression for a range of production (Bjelland et al. 2013; Pryce et al. 2014; Kim et al. 2015) and fertility (Kim et al. 2015; Ferenčaković et al. 2017; Martikainen et al. 2017) traits. In this study, F_{ROH} and F_{HOM} revealed low levels of genomic inbreeding in Ethiopian indigenous goats, results that are consistent with findings for Egyptian Barki (Kim et al. 2016) and Ugandan goats (Onzima et al. 2018). Colli et al. (2018) suggested that significant introgression of other breeds in African goats explains their low inbreeding. We however suggest that low inbreeding levels in African goats is due to extensive outcrossing within and between diverse populations and individuals considering that these flocks are communally grazed and watered, and mating is mostly uncontrolled.

Demographic history and dynamics

Genome-wide ROH, LD and N_e are valuable sources of information on how population structure, demography and management evolve over time (Kirin et al. 2010; Ceballos et al. 2018). Generally, short ROH are most likely correlated to ancestral inheritance, ancient bottlenecks or consanguinity, whereas long ROH are associated with recent inbreeding (Purfield et al. 2012; Browning and Browning 2013). The

distribution of ROH can also provide information on specific selection events. The distribution of the sizes of ROH segments varied across the 13 study populations; with six reporting high proportion of short (< 2Mb) ROH segments, three with a high proportion of long (≥ 8-16Mb) ROH segments and four with intermediate (4Mb) ROH segments. The variable distribution of ROH lengths with frequent short and intermediate length ROH segments has been observed in other goats (Kim et al. 2016; Brito et al. 2017; Onzima et al. 2018), sheep (Purfield et al. 2017), cattle (Mastrangelo et al. 2016) and pigs (Bosse et al. 2012). The high proportion of short ROH observed in Gondar, Western highland/Agew, Western lowland/Gumez, Woyto-Guji, Keffa and Long-eared Somali may indicate more ancient inbreeding and/or bottleneck in these populations. The high number of long ROH in the Small-eared Somali, Hararghe Highland and Nubian suggest a large effect of recent inbreeding. However, for all the 13 populations analysed in this study, low levels of genomic inbreeding were observed and thus ancient and contemporary bottleneck may explain our results, a possibility that we do not favour. Although selection coupled with a smaller N_e can account for differences in ROH lengths, we also do not favour this as a possible explanation of our results as artificial selection for any trait is either weak or absent among Ethiopian goats. We suggest that the high proportion of short ROH segments in Gondar, Western highland/Agew, Western lowland/Gumez, Woyto-Guji, Keffa and Long-eared Somali may indicate an initial small founder population while the long ROH segments in Small-eared Somali, Hararghe Highland and Nubian were possibly derived from large founding flocks.

Although we observed an overall decrease in N_e from 1000 to 13 generations ago, each population had a unique trend (Supplementary Figure S1). Assuming a generation time of three years for traditionally managed indigenous goats, it can be inferred from the composite plot that the gradual increase in N_e begun around 1860 years ago (ya) (620 generations). It was followed by a rapid increase starting around 1140 ya (380 generations) and then by a sharp decline from around 480 ya (160 generations) that persists until now (Figure 2b). Within the estimated time frame (240–1140 ya), three exceptionally favourable climatic periods, corresponding to the African Humid Period (AHP) (Wright 2017), interspersed with short dry spells were experienced in eastern Africa (Verschuren et al. 2000; 2007). Indigenous goats seem to have thrived when conditions were favourable but drastically shrunk during the subsequent drought (Verschuren et al. 2000; 2007) or following the abrupt or gradual termination of the AHP. Each population, however, seems to have responded differently to the climatic events as deduced from the individual trends (Supplementary Figure S1). It suggests the historical demographic dynamics of the indigenous goats may have been more complex than is reflected by the composite trend thus supporting further investigation possibly using the PSMC algorithm (Nadachowska-Brzyska et al. 2016) and full

genome sequences. The PSMC capitalizes on the combined pattern of the distributions of the time to the most recent common ancestor (TMRCA) between two alleles in an individual at large number of loci spread across the genome. It therefore gives more detailed information on historical-ancient N_e dynamics. Full genome sequences offer the advantage of sampling large numbers of unlinked loci across the genome. This may be necessary to accurately estimate population genetic parameters and determine the timing of demographic events.

Past and recent population genetic structure

The use of several methods to investigate population structure allowed us to reveal and describe past and recent population genetic structure in indigenous goats in Ethiopia and by extension sub-Saharan Africa. NetView revealed two broad genetic groups lacking a clear phylogeographic structure that corresponded slightly to the genetic backgrounds revealed by ADMIXTURE at K = 2. These two genetic groups may possibly represent two deep ancient ancestries that arrived with the first introduction of goats in eastern Africa. It mirrors findings of mitochondrial DNA which identified two haplogroups that also lacked a clear phylogeographic structure in Ethiopian (Tarekegn et al. 2018), Kenyan (Kibegwa et al. 2015), Sudanese (Sanhory et al. 2014) and Egyptian (Naderi et al. 2008) indigenous goats. With the current dataset, we nevertheless cannot determine the origin and route of entry and dispersal of the two deep ancestries and whether they arrived together or independently. However, their occurrence in all the 13 populations, albeit at different frequencies, points to a relatively early dispersal in the region, most likely facilitated by human socio-cultural and trade interactions as inferred from anthropologic, linguistic and human genetic studies (Pagani et al. 2012).

Cryptic population genetic structure, represented by seven genetic groups, was revealed by PCA and TreeMix. These results were supported by the data obtained using ADMIXTURE: the optimal number of clusters was K=7. Except Keffa, the genomes of the other populations are a mosaic of at least two genetic backgrounds. This pattern was repeated at $K \geq 7$ except for Abergelle, Woyto-Guji and Afar which showed one predominant genetic background. How these seven backgrounds evolved from the two ancestries remains a subject of speculation. Geo-specific local founder events accentuated by genetic drift arising from past reproductive isolation could be a possible source of the observed evolutionary hepta-groupings. With extensive human migrations across Ethiopia dating back to the 16th century (Yelma, 1967; Habitamu, 2014), it is possible that human cultural and socio-economic interactions dispersed the seven backgrounds across the country resulting in admixed genomes. These results, are similar to the observation of similar levels of genome admixture in Ethiopian cattle (Dadi et al. 2008), sheep (Edea et al.

2017) and humans (Plaster 2011; Pagani et al. 2012) and in other African livestock (Benjelloun et al. 2015; Mbole-Kariuki et al. 2014; Mwacharo et al. 2017), and is explained by the fact that the three livestock species were and still are often kept together.

The unique genetic background in Keffa goats at $K \geq 7$ is noteworthy. This population is found exclusively in the Keffa region of southwestern Ethiopia, which is also the home tract of Sheko cattle. Amongst Ethiopian cattle, Dadi et al. (2008) observed a homogeneous and unique genetic background in the Sheko. The Keffa region is tsetse-infested and trypanosomosis, a potential driver of selection in African cattle (Smetko et al. 2015), is endemic. Sheko cattle are reported to have some degree of trypanotolerance (Lamecha et al. 2006; Mekonnen et al. 2019). During informal conversations with Keffa goat owners, they indicated that their goats also exhibit a degree of trypanotolerance. If confirmed, it would explain the genome uniformity in the two species which could be the result of natural selection towards trypanotolerance. Trypanosomosis acts as a natural barrier for the dispersal of other goat and cattle populations in the Kaffa region and, until recently when control measures against the disease have been enhanced and the ecology drastically modified for human activities and increased settlements, reproductively isolated the Keffa goats and Sheko cattle. The correlation between the uniform genetic background and reduced susceptibility to trypanosomosis in the two species need to be confirmed.

Genome-wide signatures of selection

Four strong selection signatures were identified, two each in Arsi-Bale and Nubian goats. Despite the observation of a unique genetic background in Keffa that we suggest could be the outcome of trypanotolerance, no selection signatures were observed in this population. This was rather surprising and difficult to explain but may suggest that reproductive isolation and the resulting genome autozygosity could be the cause. We attribute the four strong selection signatures to differential adaptation to contrasting environments namely, high altitude and the associated low temperatures experienced by the Arsi-Bale population, and low dry altitude and high temperatures experienced by the Nubian population. One Arsi-Bale candidate region spanned *ALOX5AP* a gene that was identified in sheep as a potential candidate for climate-mediated adaptation (Lv et al. 2014). In humans, a mutation in *ALOX5AP* was associated with lung function (Ro et al. 2012). Given the high altitude and restricted oxygen concentration in the Arsi-Bale mountains, *ALOX5AP* may play a role in adaptation to high altitude, especially in so far as respiration is concerned. Several candidate genes with critical roles in maintaining genome integrity through DNA repair processes (*BRCA2*, *PDS5B*, *RAD51*) (Prakash et al. 2015; Couturier et al. 2016) overlapped with the Arsi-Bale candidate genomic regions. In the tropics high altitude regions receive

much higher annual levels of UV radiation. Therefore maintaining genome integrity against possible DNA damage induced by ionizing radiation is essential. Other candidate genes found in selective sweeps included *KATNAI1*, *FRY* and *RXFP2*. *KATNAI1* has been associated with variation in fiber diameter in domestic sheep breeds (Zhang et al 2013; Seroussi et al 2017). The *FRY* has been linked with variations in coat pigmentation in many sheep breeds (Gracia-Gamez et al. 2011; Zhang et al. 2013; Wei et al. 2015; Seroussi et al. 2017) and is also involved in growing wing hairs and bristles in Drosophila (Cong et al. 2001; Fang et al. 2010; He et al. 2015). Mutations in *RXFP2* were linked to horn type and development in sheep (Johnston et al. 2011; Wang et al. 2014), reproductive success and survival in Soay sheep (Johnston et al. 2013) and the genomic region spanning *RXFP2* was also identified to be under selection in Creole (Gautier and Naves 2011) and West African Borgou cattle (Flori et al. 2014). We also observed genes (*STARD13*/*CCNA1*, *FLT1*, *DCLK1*, *NBEA*) which are reported to be associated with female reproduction in bovines (Kfir et al. 2018) and chicken (Sun et al. 2015; Shen et al. 2017). The *PITX2*, *CDX2* have been associated with puberty in cattle (Cánovas et al. 2014) and embryogenesis in the mouse (Lu et al. 2018). Evidence drawn from animal experiments suggest that conditions in high altitude environments and prolonged exposure to altitude-initiated stress, including cold and hypoxia can have direct negative effects on reproductive function (Heath and Williams 1989). The exposure can have significant negative effects on female reproduction through reduction of fertility, increasing fetal loss and/or reducing fecundability (Adekilekun et al. 2019).

The Nubian is a goat breed that is adapted to arid environments. It is therefore not surprising that hapFLK and XP-EHH identified selective sweeps overlying regions that spanned genes associated with adaptation to environmental stress (*PCK1*, *CTCFL*, *SPO11*, *BMP7*), oxidative stress (*PLAGL2*, *SRXN1*), adaptation to different ecological environments (*ZBTB46*, *ARFRP1*, *STMN3*, *GMEB2*), and mitochondrial homeostasis (*TDRD7*) including classical innate immunity and fever induction (*HCK*, *PROCR*, *CTSZ*) (Supplementary Table S2). Coat and skin colour are an integral part of adaptation. The coat of Nubian goats is light coloured and most often white. In one of the candidate regions, there were two genes which have been associated with colouration (*GNE*, *EDN3*). Surprisingly, we observed genes that have been associated with dairy (*PROCR*, *AURKA*, *FOXS1*, *CD72*, *AVP, PXMP4*, *PIGU*, *ZNF341*, *NCOA6*, *ACSS2*, *E2F1*), beef (*MYLK2*, *MYL9*, *MYH7B*, *TNFRSF6B*, *PROCR*, *ERGK3*), and reproduction (*CTCFL*, *SPO11*, *RBM38*, *PMEPA1*, *CTNNBL1*, *NNAT*, *BLCAP*, *CPNE1*, *SPAG4*, *CTCFL*, *TGFBR1*) traits (Supplementary Table S2). With candidate regions overlapping these genes our results demonstrate the potential of the Nubian to be used in breeding programs to increase production to meet future demand for proteins of animal origin under unpredictable climatic conditions.

Conclusion

The genetic history of African indigenous goats is complex. It has obviously been closely intertwined with the history of local human communities. The need to adapt to diverse African environments may also have shaped present-day African goat genomes. Here, have provided insights that improve our understanding of the genomic landscape and demographic history of sub-Sahara African indigenous goats in Ethiopia. Our results provide a foundation to formulate and test biological hypotheses relating to population demographic profiles and genome dynamics in African livestock. Further studies are needed to refine and confirm our interpretations and the proposed hypotheses. These include among others, the association between the uniform genetic background in Keffa and reduced susceptibility to trypanosomosis, the origin and route of entry and dispersal of the two deep ancestries into the region, and the suggestion that the historical demographic dynamics of the indigenous goats may be more complex than is reflected by the composite trend.

Acknowledgements

Special thanks to flock owners who volunteered their animals for sampling. We acknowledge the assistance of the district agriculture and rural development offices during sampling. This project was supported by the BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) program and partially by the CGIAR Research Program on Livestock. The ABCF Program is funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill & Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID) and; the Swedish International Development Cooperation Agency (Sida).

Data availability

The data used here will be made available upon request.

Conflict of interest.

The authors declare no conflict of interest regarding this manuscript

References

Adekilekun, A.M., Aboua, Y.G., Oyeyipo, I.P., & Oguntibeju, O.O. (2019). High altitude and human

reproductive function. *African Journal of Biomedical Research* 22, 1-9.

Alexander, D.H., Novembre, J. & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19, 1655-1664.

Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M.W. (2015). SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics* 6, 109.

Benjelloun, B., Alberto, F.J., Streeter, I., Boyer, F., Coissac, E., Stucki, S., BenBati, M., Ibnelbachyr, M., Chentouf, M., Bechchari, A., Leempoel, K., Alberti, A., Engelen, S., Chikhi, A., Clarke, L., Flicek, P., Joost, S., Taberlet, P., Pompanon, F. & NextGen Consortium. (2015). Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Frontiers in Genetics* 6:107.

Berihulay, H., Li, Y., Liu, X., Gebreselassie, G., Islam, R., Liu, W., Jiang, L. & Ma, Y. (2019). Genetic diversity and population structure in multiple Chinese goat populations using a SNP panel. *Animal Genetics* 50, 242-249.

Bickhart, D.M., Rosen, B.D., Koren, S., Sayre, B.L., Hastie, A.R., Chan, S. et al. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics* 49, 643-650.

Bjelland, D., Weigel, K., Vukasinovic, N. & Nkrumah, J. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science* 96, 4697-4706.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdalla, J., Blott, S. & SanCristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics* 186, 241-262.

Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L.A., Schook, L.B., et al. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genetics.* 8, e1003100.

Brito, L.F., Kijas, J.W., Ventura, R.V., Sargolzaei, M., Porto-Neto, L.R., Cánovas, A., Feng, Z., Jafarikia, M. & Schenkel, F.S. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics* 18, 229.

Browning, B. & Browning, S. (2013). Improving the accuracy and efficiency of identity by descent detection in population data. *Genetics* 194, 459-471.

Cánovas, A., Reverter, A., DeAtley, K.L., Ashley, R.L., Colgrave, M.L., Fortes, M.R., Islas-Trejo, A., Lehnert, S., Porto-Neto, L., Rincón, G., Silver, G.A., Snelling, W.M. et al. (2014). Multi-tissue omics analysis reveal molecular regulatory networks for puberty in composite beff cattle. *PloS One* 9, e102551.

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M. & Wilson, J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics* 19, 220-234.

Chudasama, P., Mughal, S.S., Sanders, M.A., Hübschmann, D., Chung, I. et al. (2018). Integrative genomic and transcriptomic analysis of leiomyosarcoma. *Nature Communications* 9, 144.

Colli, L., Milanesi, M., Talenti, A., Bertolini, F., Chen, M. et al. (2018). Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genetics Selection Evolution* 50, 58.

Cong, J., Geng, W., He, B., Liu, J., Charlton, J. et al. (2001). The furry gene of Drososphila is important for maintaining the integrity of cellular extensions during morphogenesis. *Development* 128, 2793-2802.

Couturier, AM., Fleury, H., Patenaude, A-M., Bentley, V.L., Rodrigue, A., Coulombe, Y., et al. (2016). Roles of APRIN (PDS5B) in homologous recombination and in ovarian cancer prediction. *Nucleic Acids Research* 44, 10879-10897.

Curik, I., Ferenèakoviæ, M. & Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science* 166, 26-34.

Dadi, H., Tibbo, M., Takahashi, Y., Nomura, K., Hanada, H. & Amano, T. (2008). Microsatellite analysis reveals high genetic diversity but low genetic structure in Ethiopian indigenous cattle populations. *Animal*

Genetics 39, 425-431.

Daly, K.G., Delser, P.M., Mullin, V.E., Scheu, A., Mattiangeli, V. et al. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* 361, 85-88.

Dumont, B.L. & Payseur, B.A. (2008). Evolution of the genomic rate of recombination in mammals. *Evolution* 62, 276-294.

Edea, Z., Dessie, T., Dadi, H., Do, K.-T. & Kim, K.-S. (2017) Genetic Diversity and Population Structure of Ethiopian Sheep Populations Revealed by High-Density SNP Markers. *Frontiers in Genetics* 8:218.

Fang, X., Lu, Q., Emoto, K., Adler, P.N. (2010). The Drosophila FRY protein interacts with Trc and is highly mobile in vivo. *BMC Developmental Biology* 10, 40.

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M. & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* 193, 929-941.

Ferenčaković, M., Hamziæ, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I. et al. (2013). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics* 130, 286-293.

Flori, L., Thevenon, S., Dayo, G.-K., Senou, M., Sylla, S., Berthier, D., Moazami-Goudarzi, K. & Gautier M. (2014). Adaptive admixture in the West African bovine hybrid zone: insight from the Borgou population. *Molecular Ecology* 23, 3241-3257.

Fuller, D.Q. & Boivin, N.L. (2009). Crops, cattle and commensals across the Indian Ocean: Current and potential archaeobiological evidence. *Etudes Ocean Indien*, 42, 13-46.

Garcia-Gamez, E., Reverter, A., Whan, V., McWilliam, S.M., Arranz, J.J., Kijas, J. (2011). Using regulatory and epistatic networks to extend the findings of a genome scans identifying the gene drivers of pigmentation in Merino sheep. *PLoS One*. 6, e21158.

Gautier, M. & Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole

cattle breed. *Molecular Ecology* 20, 3128-3143.

Gifford-Gonzalez, D. & Hanotte, O. (2011). Domesticating animals in Africa: implications of genetic and archaeological findings. *Journal of World Prehistory*. 24, 1-23.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long-term response. *Genetica* 136, 245-257.

Guan, D., Luo, N., Tan, X., Zhao, Z., Huang, Y., Na, R., Zhang, J. & Zhao, Y. (2016). Scanning of selection signature provides a glimpse into important economic traits in goats (Capra hircus). *Scientific Reports* 6, 36372.

Habitamu, A. (2014). (2006 E.C.). Country and politics (Amharic version, 1st ed.). Addis Ababa, Ethiopia.

Hayes, B., Visscher, P., McPartlan, H. & Goddard, M. (2003) Novel multilocus measure of Linkage Disequilibrium to estimate past effective population size. *Genome Research* 13, 635-643.

Heath, D. & Williams, D.R. (1989). Fertility and pregnancy. In *High-Altitude Medicine and Pathology,* Chapter 21, pp223-231. London Butterworths.

He, Y., Fang, X., Emoto, K., Jan, Y.N. & Adier, P.N. (2005). The tricornered Ser/Thr protein kinase is regulated by phosphorylation and interacts with furry in Drosophila wing hair development. *Molecular Biology of the Cell*. 16,689-700.

Hill, W.G. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical Applied Genetics* 38, 226-231.

Johnston, S.E., Gratten, J., Berenos, C., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* 502, 93-95.

Johnston, S., McEwan, J.C., Pickering, N.K., Kijas, J.W., Beraldi, D., Pilkington, J.G., Pemberton, J.M. & Slate, J. (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative

variation in sexual weaponry in a wild sheep population. *Molecular Ecology* 20, 2555-2566.

Jolliffe, I.T. (2002). Principal component analysis. Second Edition, New York, Springer-Verlag New York Inc.

Kfir, S., Basavaraja, R., Wigoda, N., Ban-Dor, S., Orr, I. & Meidan, R. (2018). Genomic profiling of bovine corpus-luteum maturation *PLoS One* 13, e0194456.

Kibegwa, F.M., Githui, K.E., Jung'a, J.O., Badamana, M.S. & Nyamu, M.N. (2015). Mitochondrial DNA variation of indigenous goats in Narok and Isiolo counties of Kenya. *Journal of Animal Breeding and Genetics*, 133, 1-10.

Kim, E.-S., Cole, J.B., Huson, H., Wiggans, G.R., Van Tassell, C.P., Crooker, B.A., Liu, G., Da, Y. & Sonstegard, T.S. (2013). Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PLoS One* 8, e80813.

Kim, E.-S., Sonstegard, T.S., Van Tassell, C.P., Wiggans, G. & Rothschild, M.F. (2015) The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. *PLoS One* 10, e0129967.

Kim, E.S., Elbeltagy, A.R., Aboul-Naga, A.M., Rischkowsky, B., Sayre, B., Mwacharo, J.M. & Rothschild, M.F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* 116, 255-264.

Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M. & Wilson, J.F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5, e13996.

Knief, U., Kempenaers, B. & Forstmeier, W. (2017) Meiotic recombination shapes precision of pedigree- and marker-based estimates of inbreeding. *Heredity* 118, 239-248.

Lemecha, H., Mulatu, W., Hussein, I., Rege, E., Tekle, T. & Abdicho, S. (2006). Response of four indigenous cattle breeds to natural tsetse and trypanosomosis challenge in the Ghibe valley of Ethiopia. *Veterinary Parasitology* 141, 165-176.

Liu, N., Li, H., Liu, K., Yu, J., Cheng, M., De, W. et al. (2014). Differential expression of genes and proteins associated with wool follicle cycling. *Molecular Biology Reports* 41, 5343-5349.

Lu, X., Zhao, Z.A., Wang, X., Zhang, X., Zhai, Y., Deng, W., Yi, Z., Li, L. (2018). Whole-transcriptome splicing profiling of E7.5 mouse primary germ layers reveals frequent alternative promoter usage during mouse early embryogenesis. *Biology Open* 7, bio032508.

Luikart, G., Gielly, L., Excoffier, L., Vigne, J.D., Bouvet, J. & Taberlet, P. (2001). Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences* 98, 5927-5932.

Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J.W. et al (2014). Adaptations to climate-mediated selective pressures in sheep. *Molecular Biology Evolution* 31, 3324-3343.

Ma, Y., Zhang, H., Zhang, Q. & Ding, X. (2014). Identification of selection footprints on the X Chromosome in pig. *PLoS One* 9, e94911.

Manunza, A., Noce, A., Serradilla, J.M., Goyache, F., Martinez, A. et al. (2016). A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. *Genetics Selection Evolution* 48, 52.

Marshall, F. (2000).The origins of domesticated animals in Eastern Africa. In Blench, R. M., and MacDonald, K.C. (eds.), *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics, and Ethnography*, UCL Press, London, pp. 191–221.

Martikainen, K., Tyrisevä, A.M., Matilainen, K., Pösö, J. & Uimari, J.P. (2017). Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. *Journal of Animal Breeding and Genetics* 134, 383-392.

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M.T. & Portolano, B. (2016). Genomic inbreeding estimation in small populations: evaluations of runs of homozygosity in three local dairy cattle breeds. *Animal* 10, 746-754.

Mekonnen, Y.A., Gültas, M., Effa, K., Hanotte, O. & Schmitt, A.O. (2019). Identification of candidate signature genes and key regulators associated with trypanotolerance in the Sheko breed. *Frontiers in*

Genetics 10, 1095.

Mbole-Kariuki, M.N., Sonstegard, T., Orth, A., Thumbi, S.M., Bronsvoort, B.M.D.C. et al. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity* 113, 297-305.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M. et al. (2008). Runs of homozygosity in European populations. *American Journal of Human Genetics* 83, 359-372.

Mdladla, K., Dzomba, E.F., Huson, H.J. & Muchadeyi, F.C. (2016). Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. *Animal Genetics* 47, 471-482.

Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* 47, D419-D426.

Mohr, P.A. (1971). The Geology of Ethiopia. University College of Addis Ababa Press, Ethiopia. 268pp.

Mwacharo, J.M., Kim, E.-S., Elbeltagy, A.R., Aboul-Naga, A.M., Rischkowsky, B.A. & Rothschild, M.F. (2017). Genomic footprints of dryland stress adaptation in Egyptian fat-Tail sheep and their divergence from East African and western Asia cohorts. *Scientific Reports* 7, 17647.

Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. (2016). PSMC analysis of effective population size in molecular ecology and its application to black-and-white *Ficedula* flycatchers *Molecular Ecology* 25, 1058-1072.

Naderi, S., Rezaei, H.R., Pompanon, F.O., Blum, M.G.B., Negrini, R. et al. (2008). The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences* 105, 17659-17664.

Neuditschko, M., Khatkar, M.S. & Raadsma, H.W. (2012). NetView: A high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS One* 7,

e48375.

Nicoloso, L., Bomba, L., Colli, L., Negrini, R., Milanesi M. et al. (2015). Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genetics Selection Evolution* 47, 1-10.

Nie, C., Zhang, Z., Zheng, J., Sun, H., Ning, Z., Xu, G., Yang, N. & Qu, L. (2016). Genome-wide association study revealed genomic regions related to white/red earlobe colour trait in the Rhode Island Red chickens. *BMC Genetics* 17, 115.

Ohta, T. & Kimura, M. (1971). Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics*. 68, 571-580.

Onzima, R.B., Upadhyay, M.R., Doekes, H.P., Brito, L.F., Bosse, M., Kanis, E., Groenen, M.A.M. & Crooijmans, R.P.M.A. (2018). Genome-wide characterisation of selection signatures and runs of homozygosity in Ugandan goat breeds. *Frontiers in Genetics* 9, 318.

Pagani, L., Kivislid, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., Bekele, E., Bradman, N., Balding, D.J. & Tyler-Smith, C. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *American Journal of Human Genetics* 91, 83-96.

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R. & Baldi, F. (2016). Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics* 48, 255-271.

Pickrell, J.K. & Pritchard JK. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics* 8, e1002967.

Pickrell, J.J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z. et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Research*. 19, 826-837.

Plaster, C.A. (2011) Variation in Y chromosome, mitochondrial DNA and labels of identity on Ethiopia PhD Thesis, University College London (UCL) 204pp.

Prakash, R., Zhang, Y., Feng, W. & Jasin, M. (2015). Homologous recombination and human health: The roles of BRCA1, BRCA2 and associated proteins. *Cold Spring Harbour Pespectives in Biology*. 7, a016600.

Pryce, J.E., Haile-Mariam, M., Goddard, M.E. & Hayes, B.J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution* 46, 71.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R. et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*. 81, 559-575.

Purfield, D.C., Berry, D.P., McParland, S. & Bradley, D.G. (2012). Runs of homozygosity and population history in cattle. *BMC Genetics*. 13, 70.

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* 12, e0176780.

Ro, M., Kim, S., Pyun, J.A., Shin, C., Cho, N.H. et al. (2012). Association between arachidonate 5-lipoxygenase-activating protein (ALOX5AP) and lung function in a Korean population. *Scandinavian Journal of Immunology* 76, 151-157.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J. et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832-837.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913-918.

Sanhory, E.A., Giha, R.S., Ebrahim, Z.H., Eldin, M.T. & Gornas, N.A. (2014). Mitochondrial DNA diversity in three Sudanese goat breeds. *Open Access Library Journal* 1, 1-10.

Scheet, P. & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* 78, 629-644.

Seroussi, E., Rosov, A., Shirak, A., Lam, A., Gootwin, E. (2017). Unveiling genomic regions that underlie differences between Afec-Assaf sheep and its parental Awassi breed. *Genetics Selection Evolution*. 49, 19.

Shen, M., Sun, H., Qu, L., Ma, M., Dou, T., Lu, J., Guo, J., Hu, Y., Wang, X., Li, Y., Wang, K. & Yang, N. (2017). Genetic architecture and candidate genes identified for follicle number in chicken. *Scientific Reports* 7, 16412.

Shinde, D.B., Gujar, B.V., Patil, H.B., Satpute, S.A. & Kashid, U.B. (2008). Protocol for isolation of genomic DNA from livestock blood for microsatellite analysis. *Indian Journal of Animal Research* 42, 279-281.

Smetko, A., Soudre, A., Silbermayr, K., Müller, S., Brem, G., Hanotte, O., Boettcher, P.J., Stella, A., Mészáros, G., Wurzinger, M., Curik, I., Müller, M., Burgstaller, J. & Sölkner, J. (2015). Trypanosomosis: potential driver of selection in African cattle. *Frontiers in Genetics* 6, 137.

Steinig, E.J., Neuditschko, M., Khatkar, M.S., Raadsma, H.W. & Zenger, K.R. (2016). Netview p: A network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources* 16, 216-227.

Sun, C., Lu, J., Yi, G., Yuan, J., Duan, Z., Qu, L., Xu, G., Wang, K. & Yang, N. (2015). Promising loci and genes for yolk and ovary weight in chickens revealed by a genome-wide association study. *PLoS One* 10, e0137145.

Sved, J.A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical and Population Biology* 141, 125-141.

Tarekegn, G.M., Tesfaye, K., Mwai, O.A., Djikeng, A., Dessie, T., Birungi, J., Osama, S., Zergaw, N., Alemu, A., Achieng, G., Tutah, J., Mutai, C., Njuguna, J. & Mwacharo, J.M. (2018). Mitochondrial DNA variation reveals maternal origins and demographic dynamics of Ethiopian indigenous goats. *Ecology and Evolution* 8, 1543-1553.

Umer, M., Legesse, D., Gasse, F., Bonnefille, R., Lamb, H.F., Leng, M.J. & Lamb, A.A. (2007). Late quaternary climate changes in the Horn of Africa. In: Battarbee R.W., Gasse F., Stickley C.E. (eds) Past

Climate Variability through Europe and Africa. *Developments in Paleoenvironmental Research*, vol 6. Springer, Dordrecht.

Verschuren, D., Laird, K.R. & Cumming, B.F. (2000). Rainfall and drought in equatorial East Africa during the past 1,100 years. *Nature* 403, 410-414.

Verschuren, D. (2007). Decadal and century-scale climate variability in tropical Africa during the past 2000 years. In: Battarbee R.W., Gasse F., Stickley C. E. (eds), Past Climate Variability through Europe and Africa. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 139-158.

Wang, X., Zhou, G., Li, Q., Zhao, D. & Chen, Y. (2014). Discovery of SNPs in RXFP2 related to horn types in sheep. *Small Ruminant Research* 116, 133-136.

Wei, C., Wang, H., Liu, G., Wu, M., Cao, J., Liu, Z. et al. (2015). Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* 16, 194.

Wright, S. (1922). Coefficients of -nbreeding and relationship. *American Naturalist* 56:330-338.

Wright, D.K. (2017). Humans as agents in the termination of the Africa Humid Period. *Frontiers in Earth Science* 5, 4.

Yelma, D. (1966). History of Ethiopia in the sixteenth century. Berhanena Selam, Addis Ababa. 248pp

Zavarez, L.B., Utsunomiya, Y.T., Carmo, A.S., Neves, H.R.H., Carvalheiro, R., Ferenčaković, M., O'Brien, A.M.P., Curik, I., Cole, J.B., Van Tassell, C.P., da Silva, M.V.G.B., Sonstegard, T.S., Sölkner, J. & Garcia, J.F. (2015). Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. *Frontiers in Genetics* 6, 5.

Zhang, L., Mousel, M.R., Wu, X., Michal, J.J., Zhou, X., Ding, B., Dodson, M.V. et al. (2013). Genome-wide genetic diversity and differentially selected regions among Suffolk, Rambouillet, Columbia, Polypay and Targhee sheep. *PloS One*. 8, e65942.

Zhang, Y., Young, J., Wang, C., Sun, X., Wolc, A. & Dekkers. J. (2014). Inbreeding by pedigree and genomic

markers in selection lines of pigs. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock Production: WCGALP-2014; Vancouver, Canada.

Zeder, MA. & Hesse, B. (2000). The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science* 287, 2254-2257.

Figure titles and legends

Figure 1a. Minor allele frequency (MAF) distribution and 1b. Frequency of runs of homozygosity segments (2, 4, 8 and 16 Mb) in each Ethiopian goat population

Figure 2a. Trend in LD decay over genomic distances for each population and 2b. The trend in composite effective population size modelled across the 13 Ethiopian goat population

Figure 3a. Principal component analysis plot. 3b. Tree Mix phylogenic tree. 3c. Plots of CV error against *K* and ADMIXTURE displays for $7 \leq K \leq 10$.

Figure 4a. NetView P plot for K-NN = 75 showing clustering of the 13 populations into two groups. 4b. Geographic sampling regions in Ethiopia together with pie charts showing ancestry proportions for *K* = 2 generated in ADMIXTURE. 4c. ADMIXTURE bar plot showing proportion of ancestries for *K* = 2

Figure 5. Genome–wide hapFLK statistics between Arsi-Bale and Nubian goat populations. Green and red lines correspond to *p*-values < 0.005 and < 0.001, respectively

Supplementary Figure S1. Trends in Ne in each population analysed in the current study

Supplementary Figure S2. NetView P plots for K-NN = 5 to K-NN = 120.

Supplementary Figure S3. Genome–wide hapFLK statistics for all pair-wise comparisons performed in this study. Green and red lines correspond to *p*-values < 0.005 and < 0.001, respectively.

Table 1 Genetic diversity and variation statistics for 13 Ethiopian indigenous goat populations (mean ± standard deviation)

Population	Abbreviation	Regions	N	H_O	H_E	P_N	D_{ST}	F_{HOM}	F_{ROH}
Abergelle	Abe	Lowland	52	0.369±0.131	0.373±0.119	0.982	0.295±0.008	0.010±0.045	0.023±0.039
Gondar	Gon	Central highland	54	0.376±0.131	0.373±0.118	0.984	0.292±0.011	-0.007±0.019	0.006±0.014
Ambo	Amb	Central highland	71	0.370±0.125	0.375±0.117	0.988	0.296±0.014	0.011±0.050	0.029±0.044
Western-Highland/Agew	WeH	Highland	45	0.371±0.134	0.372±0.120	0.988	0.292±0.008	0.001±0.027	0.019±0.022
Western-Lowland/Gumez	WeL	Lowland	41	0.370±0.138	0.370±0.122	0.987	0.291±0.019	0.000±0.062	0.021±0.044
Keffa	Kaf	Middle land	36	0.348±0.143	0.361±0.130	0.978	0.289±0.018	0.035±0.105	0.057±0.099
Woyto-Guji	WoG	Lowland	51	0.375±0.129	0.375±0.117	0.985	0.295±0.009	0.002±0.028	0.011±0.015
Arsi-Bale	ArB	Highland	46	0.365±0.130	0.374±0.118	0.991	0.297±0.009	0.022±0.056	0.044±0.051
Afar	Afa	Lowland	49	0.382±0.124	0.386±0.110	0.996	0.304±0.009	0.009±0.061	0.028±0.070
Hararghe Highland	HaH	Highland	44	0.379±0.124	0.385±0.109	0.997	0.304±0.015	0.014±0.062	0.043±0.072
Short-Eared-Somali	SeS	Lowland	41	0.379±0.127	0.381±0.111	0.995	0.300±0.007	0.006±0.045	0.028±0.051
Long-Eared-Somali	LeS	Lowland	47	0.375±0.132	0.374±0.118	0.992	0.294±0.012	-0.002±0.030	0.016±0.024
Nubian	Nub	Lowland	51	0.369±0.115	0.391±0.106	0.994	0.315±0.019	0.054±0.118	0.087±0.116

Note: N, H_O, H_E, P_N and D_{ST}, F_{HOM} and F_{ROH} refers to sample size, observed and expected heterozygosity, proportion of polymorphic SNPs, average pair-wise genetic distance, inbreeding coefficients based on excess of homozygosity and runs of homozygosity, respectively.

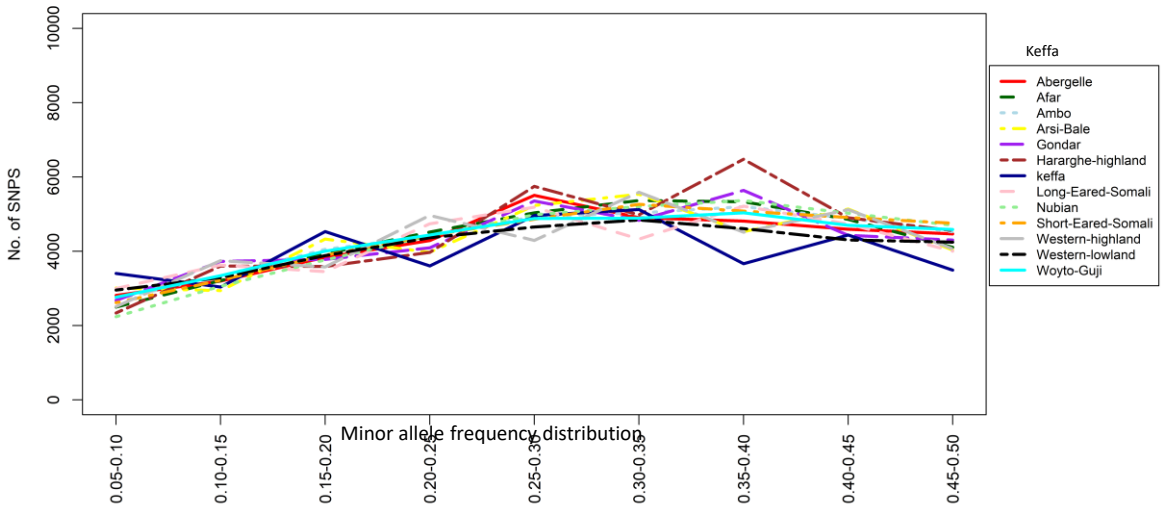
Table 2 Descriptive statistics for runs of homozygosity (ROH) segments within each of the 13 Ethiopian indigenous goat populations

	No. of ROH segments		Size (Kb)		No. of SNP		SNP Density (SNP/Kb)		Prop. of homozygous
	Mean ± SD	Max	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD
Abergelle	5.50 ± 7.45	35	7152.31 ± 4247.39	2420.37-16840.43	132.46 ± 76.94	48-303	53.64 ± 4.04	42.48-62.34	0.99 ± 0.01
Gondar	2.46 ± 2.99	16	4554.92 ± 2240.02	2438.59-13010.61	84.62 ± 40.18	49-241	53.58 ± 3.33	47.35-61.09	0.99 ± 0.01
Ambo	6.64 ± 7.29	33	7031.41 ± 4647.14	2324.55-21353.34	131.22 ± 84.69	48-391	53.16 ± 3.10	43.04-63.05	0.99 ± 0.01
WH/Agew	6.09 ± 4.83	20	5962.93 ± 3008.47	2385.04-14184.03	111.34 ± 54.85	48-263	53.19 ± 2.83	45.53-60.64	0.99 ± 0.01
WL/Gumez	5.69 ± 6.25	38	6386.99 ± 3944.45	2247.98-19014.48	118.72 ± 71.91	52-348	53.44 ± 3.17	43.23-59.64	0.99 ± 0.01
Keffa	12.14 ± 12.67	44	6881.78 ± 5043.47	2822.41-19047.42	128.80 ± 93.20	55-353	53.37 ± 2.04	49.78-57.66	0.99 ± 0.01
WoG	3.51 ± 2.99	14	6435.37 ± 3240.10	2368.95-13882.20	118.09 ± 60.34	50-271	54.61 ± 3.52	47.38-64.31	0.99 ± 0.01
Arsi-Bale	10.12 ± 9.88	34	8414.97 ± 3762.29	2579.62-16555.66	156.59 ± 68.49	53-304	53.63 ± 2.02	48.67-60.36	0.99 ± 0.01
Afar	5.12 ± 7.95	37	6635.72 ± 5722.43	2340.03-26593.98	123.54 ± 106.88	40-489	53.82 ± 3.31	47.79-64.60	0.99 ± 0.01
HaH	7.23 ± 9.68	33	8217.44 ± 6080.78	2236.95-23376.38	162.93 ± 110.61	49-415	53.24 ± 2.76	44.73-57.04	0.99 ± 0.01
SeS	5.41 ± 6.19	28	10072.83 ± 8267.15	2893.13-40669.67	187.53 ± 151.63	54-749	53.58 ± 2.98	48.39-64.00	0.99 ± 0.01
LeS	4.59 ± 5.01	17	6207.37 ± 3281.42	2474.84-12859.00	114.78 ± 60.63	51-241	54.20 ± 3.43	45.96-62.83	0.99 ± 0.01
Nubian	16.85 ± 19.84	70	9669.93 ± 5017.69	2839.61-21506.99	179.99 ± 91.74	55-396	53.63 ± 2.70	44.62-60.09	0.99 ± 0.01

Note: WH: Western highland, WL: Western lowland, WoG: Woyto-Guji, HaH: Hararghe Highland, SeS: Small-eared Somali, Les: Long-eared Somali.

Figure 1

(1a)



(1b)

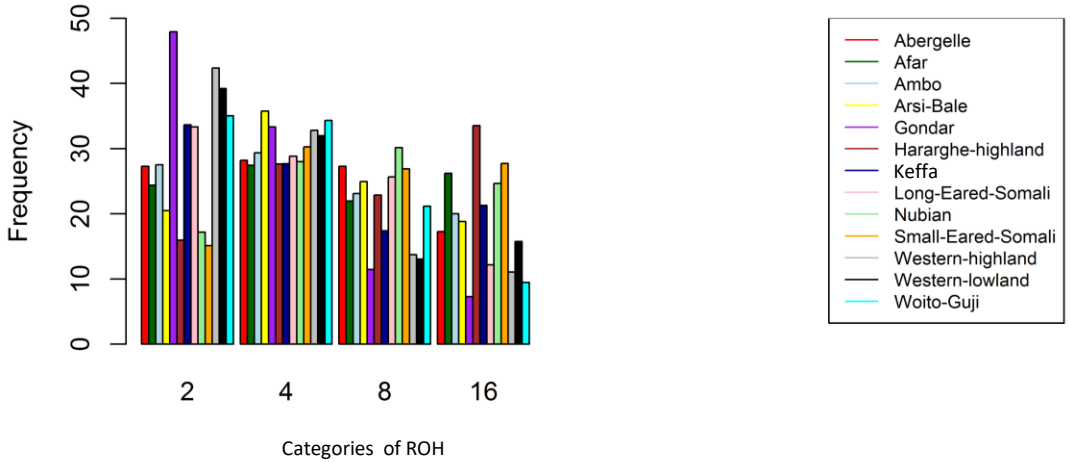
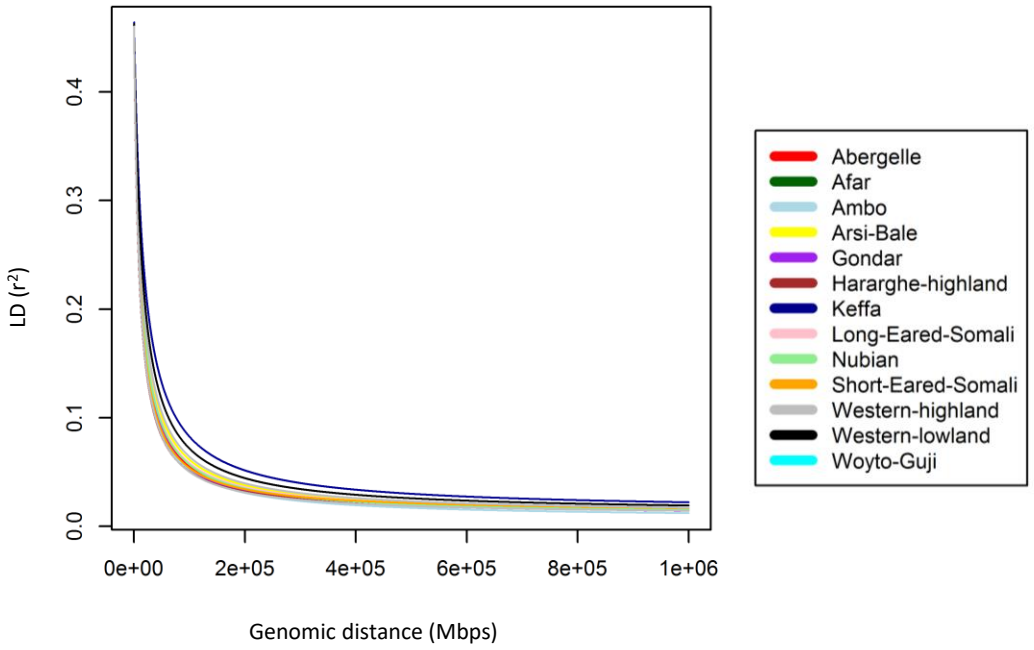
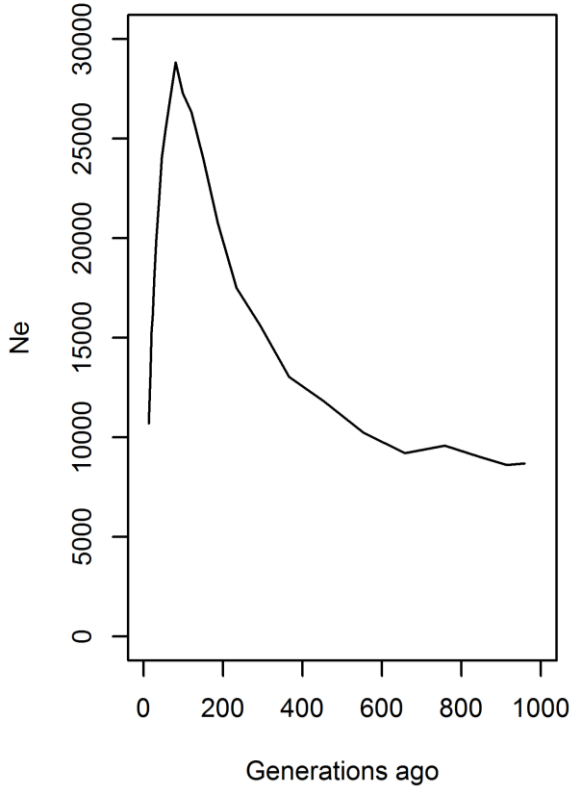


Figure 2

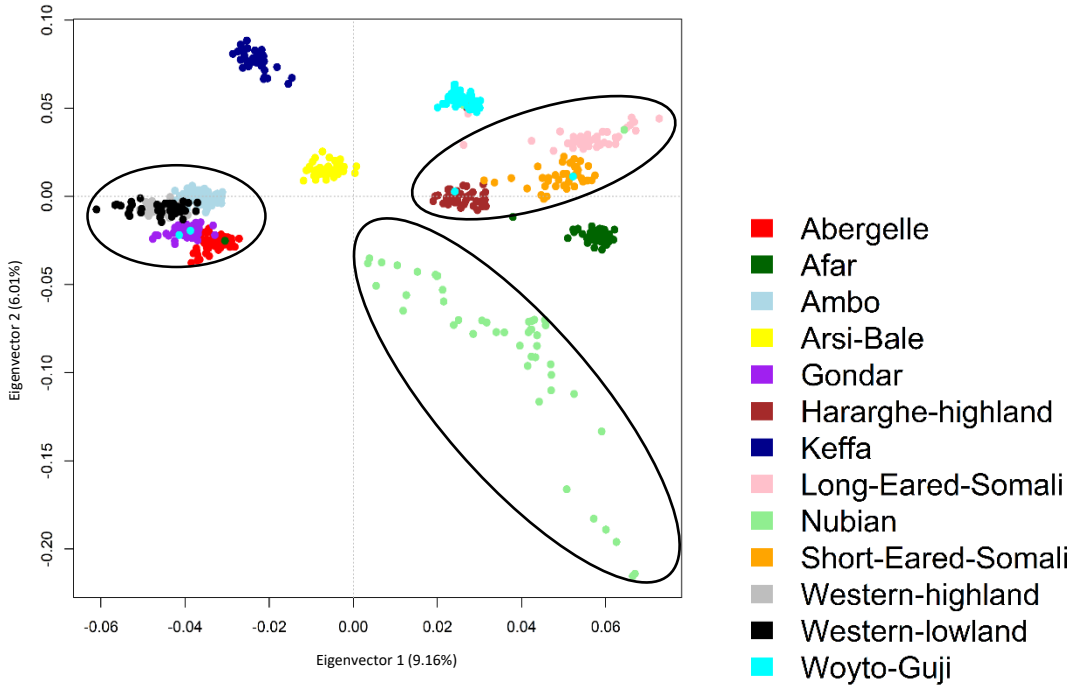
(2a)



(2b)

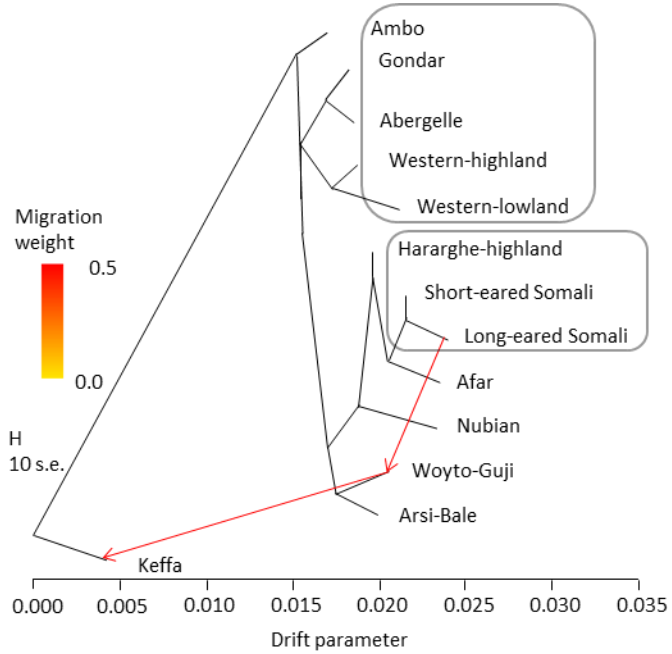


(3a)

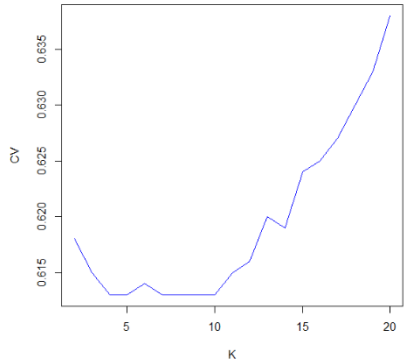


Abergelle
Afar
Ambo
Arsi-Bale
Gondar
Hararghe-highland
Keffa
Long-Eared-Somali
Nubian
Short-Eared-Somali
Western-highland
Western-lowland
Woyto-Guji

(3b)



(3c)



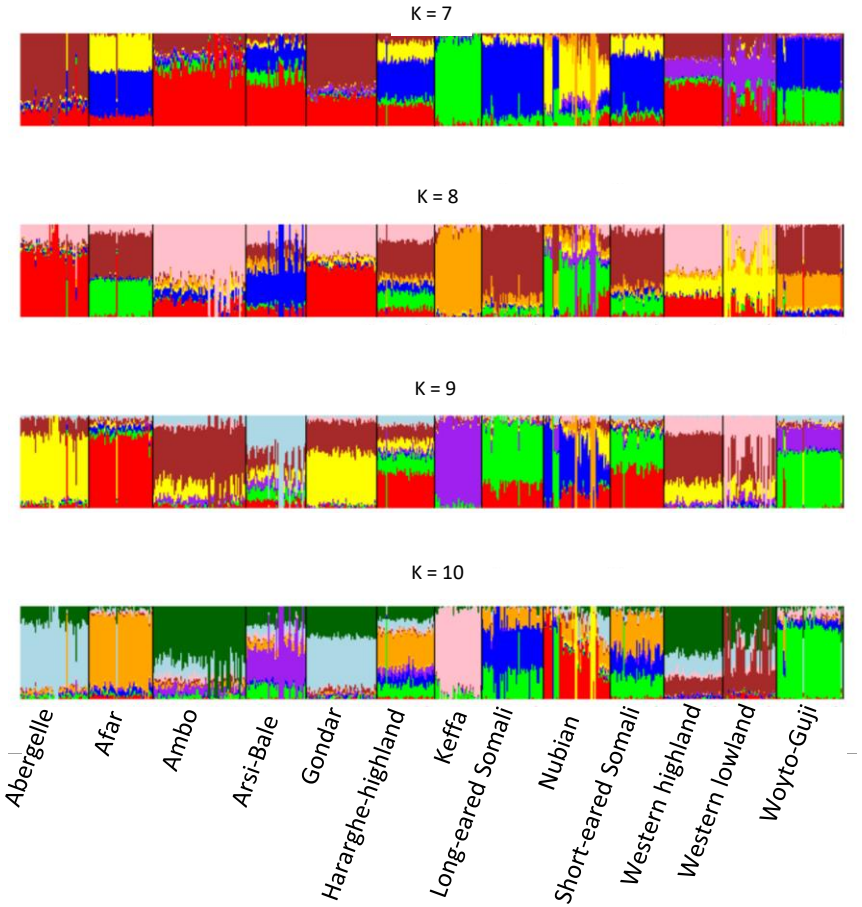
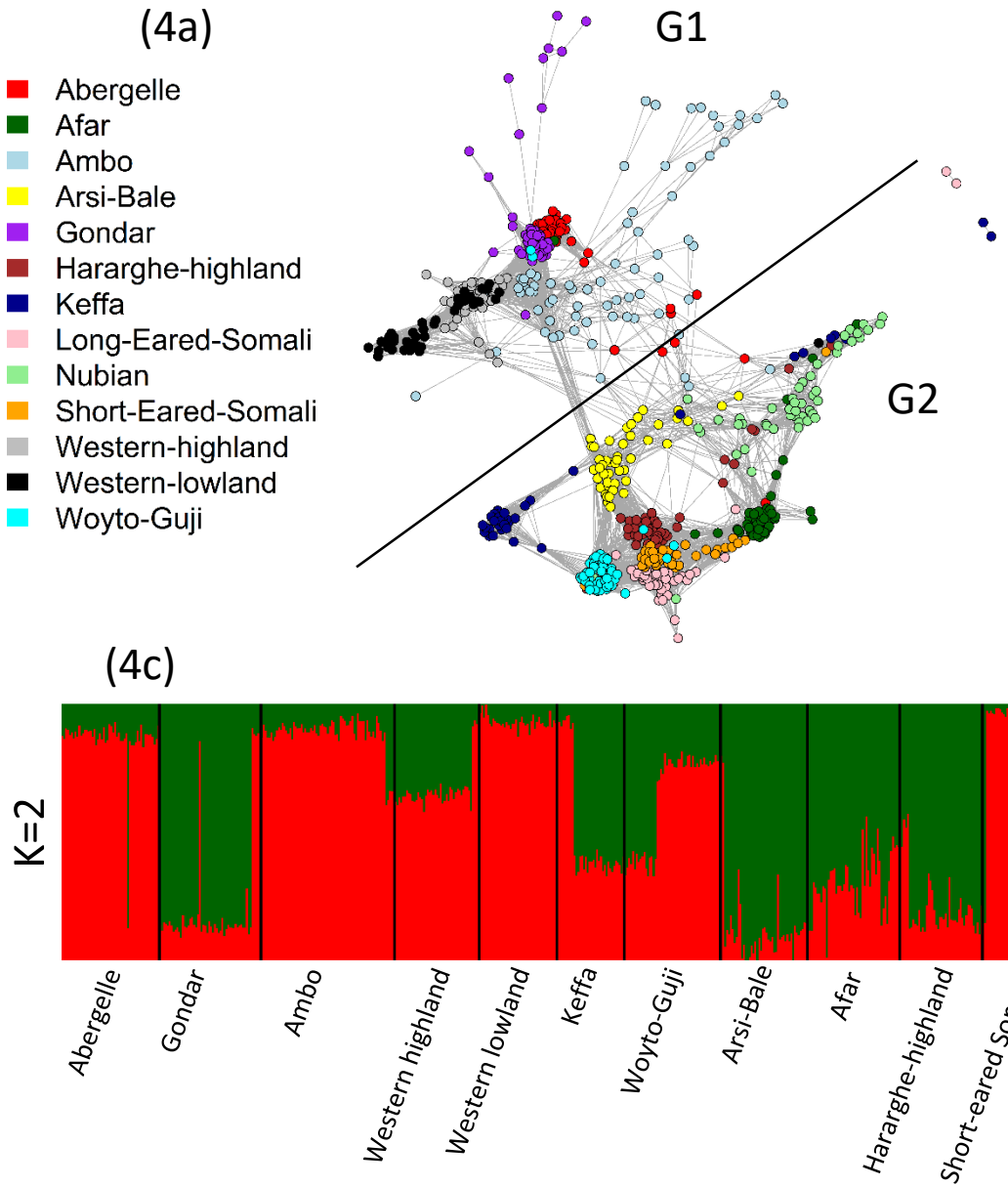


Figure 3

Figure 4


(4a)
G1
G2
Abergelle
Afar
Ambo
Arsi-Bale
Gondar
Hararghe-highland
Keffa
Long-Eared-Somali
Nubian
Short-Eared-Somali
Western-highland
Western-lowland
Woyto-Guji
(4c)
K=2
Abergelle
Gondar
Ambo
Western highland
Western lowland
Keffa
Woyto-Guji
Arsi-Bale
Afar
Hararghe-highland
Short-eared Somali
Long-eared Somali
Nubian

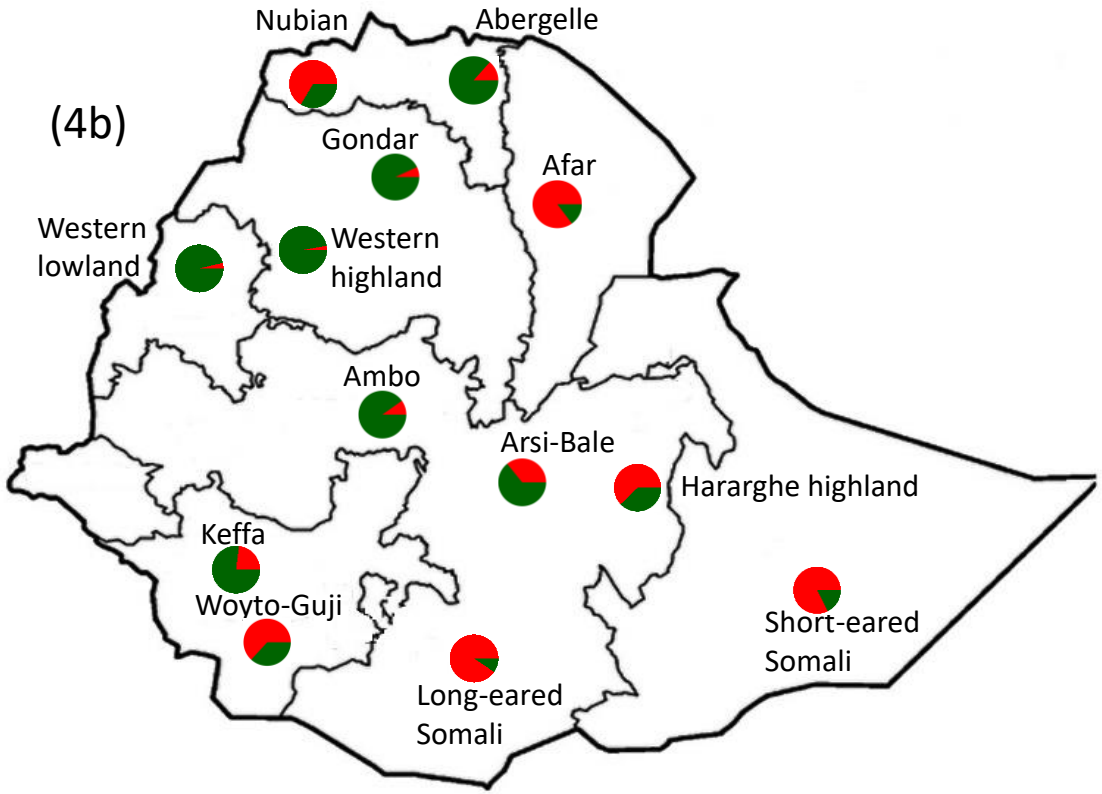
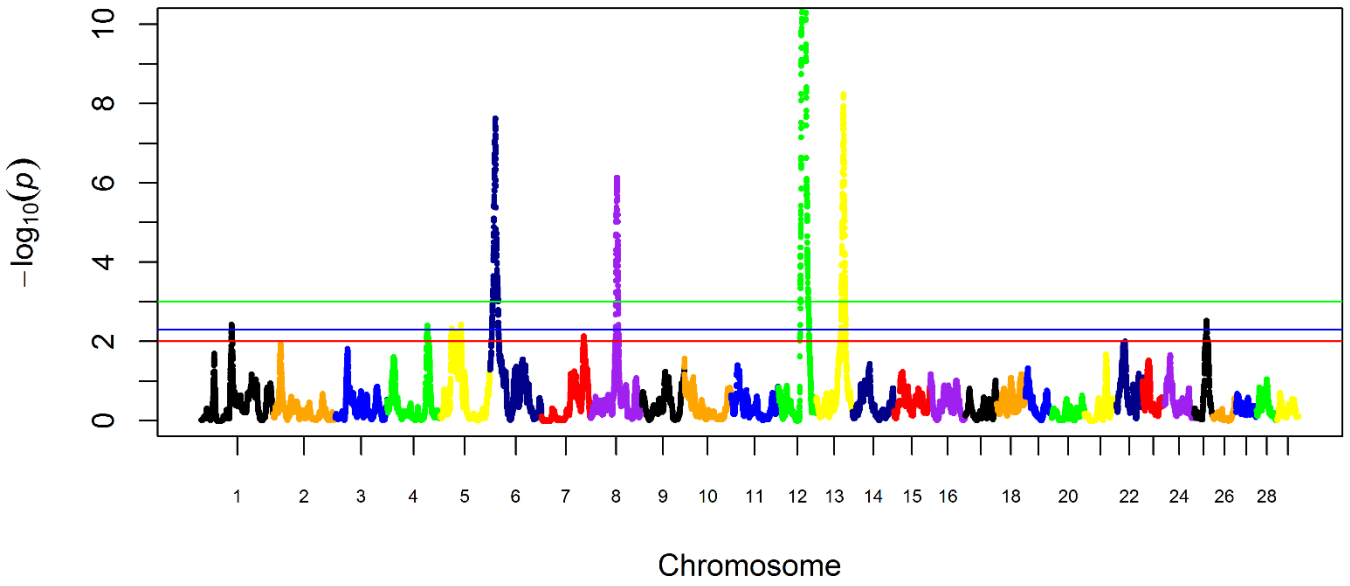

(4b)
Nubian
Abergelle
Gondar
Afar
Western lowland
Western highland
Ambo
Arsi-Bale
Hararghe highland
Keffa
Woyto-Guji
Short-eared Somali
Long-eared Somali

Figure 5.

5a.



5b.