Genome-wide variation and putative candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep

Abulgasim M. Ahbara[1, 2*], Hussain Bahbahani[3], Faisal Almathen[4], Mohammed A. Al Abri[5], Mukhtar O. Agoub[6], Ayelle Abeba[7], Adebabay Kebede[8, 9], Hassan H. Musa[10], Salvatore Mastrangelo[11], Fabio Pilla[12], Elena Ciani[13], Olivier H. Hanotte[1, 14], Joram M. Mwacharo[15*]

[1]Life Sciences, University of Nottingham, United Kingdom, [2]Department of Zoology, Misurata University, Libya, [3]Department of Biological Sciences, Faculty of Science, Kuwait University, Kuwait, [4]Department of Veterinary Public Health and Animal Husbandry, King Faisal University, Saudi Arabia, [5]Department of Animal and Veterinary Sciences, Sultan Qaboos University, Oman, [6]Agricultural Research Center, Libya, [7]Debre Berhan Research Centre, Ethiopia, [8]Amhara Regional Agricultural Research Institute, Ethiopia, [9]International Livestock Research Institute, Ethiopia, [10]Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan, [11]Dipartimento di Scienze Agrarie e Forestali, University of Palermo, Italy, [12]Dipartimento Agricoltura, Ambiente e Alimenti, Università degli Studi del Molise, Italy, [13]Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, Università degli Studi di Bari, Italy, [14]LiveGene, International Livestock Research Institute, Ethiopia, [15]Small Ruminant Genomics, International Center for Agriculture Research in the Dry Areas (ICARDA), Ethiopia

Submitted to Journal:
Frontiers in Genetics

Specialty Section:
Livestock Genomics

Article type:
Original Research Article

Manuscript ID:
387968

Received on:
20 Apr 2018

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

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

Author contribution statement

AA, JMM and OH conceived and designed the study/project, AA analysed the data and wrote the manuscript. JMM, OH revised the manuscript. HB provided logistical support in data analysis. SM, FP and EC contributed to genotyping and genotype data of non-Ethiopian breeds (Najdi, Omani, and Libyan Barbary) and provided critical inputs on data analysis and in the writing of the manuscript. FA, MA and MOA supported the sampling and genotyping of Najdi, Omani and Libyan sheep. AK and AA lead and coordinated the sampling of Ethiopian sheep HHM lead and coordinate the sampling of Sudanese sheep. All authors contributed to the interpretation of the results with their knowledge on local indigenous sheep genetic resources of their respective countries. All the authors read and approved the final manuscript.

Keywords

admixture, Africa, Fat-tail, Ovis aries, Thin-tail

Abstract

Word count: 268

Abstract

Variations in body weight and body fat distribution are associated with feed quality and quantity, thermoregulation and energy reserve. Ethiopia is characterised by distinct agro-ecologies and human ethnic farmer's diversity of ancient origin which have impacted on the variation of its livestock species. Here, we investigate the autosomal genome-wide profiles of 11 Ethiopian indigenous sheep populations using the Illumina Ovine 50K SNP BeadChip assay. Populations from The Caribbean, Europe, Middle East, China and western, northern and southern Africa were included addressing globally, the genetic variation and history of Ethiopian sheep populations. Population structure, PCA and phylogenetic analysis separate the Ethiopian indigenous fat-tail sheep from the North African and Middle Eastern fat-tailed sheep. It indicates two main genetic backgrounds and supports two distinct genetic history for the African fat-tailed sheep. Within Ethiopia, our results indicate that the short fat -tailed do not represent a monophyletic group. Four genetic backgrounds are present within Ethiopian sheep but at different proportions among fat-rump sheep, long fat-tailed sheep from western Ethiopia and long fat-tailed sheep from southern Ethiopia. Ethiopian fat-rump sheep were also found to share a common genetic background with Sudanese thin-tail sheep. Selection signature analysis identified eight candidate genomic regions that spanned genes influencing growth traits and fat deposition (NPR2, HINT2, SPAG8), embryonic development of tendons, bones and cartilage (EYA2, SULF2), regulation of body temperature (DIS3L2, LIN28B) and the control of lipogenesis and intracellular transport of long-chain fatty acids (SREBF1 and FABP3). Our findings indicate that Ethiopian indigenous fat-tail sheep represent a distinct genepool and an important resource for understanding the genetic control of fat metabolism and associated physiological processes.

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: Yes

Please provide the complete ethics statement for your manuscript. Note that the statement will be directly added to the manuscript file for peer-review, and should include the following information:

- Full name of the ethics committee that approved the study
- Consent procedure used for human participants or for animal owners
- Any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species

As per the Frontiers authors guidelines, you are required to use the following format for statements involving human subjects: This study was carried out in accordance with the recommendations of [name of guidelines], [name of committee]. The protocol was approved by the [name of committee]. All subjects gave written informed consent in accordance with the Declaration of

Helsinki.
For statements involving animal subjects, please use:
This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee'. The protocol was approved by the 'name of committee'.

If the study was exempt from one or more of the above requirements, please provide a statement with the reason for the exemption(s).
Ensure that your statement is phrased in a complete way, with clear and concise sentences.
Local regulations were observed during the sampling of all the populations analysed here. Skin tissues importation and/or exportation was permitted by the Ethiopian Ministry of Livestock and Fisheries (Certificate No: 14-160-401-16).

Genome-wide variation and putative candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep

Abulgasim Ahbara[1,2*], Hussain Bahbahani[3], Faisal Almathen[4], Mohammed Al Abri[5], Mukhtar Omar Agoub[6], Ayelle Abeba[7], Adebabay Kebede[8,9], Hassan Hussein Musa[10], Salvatore Mastrangelo[11], Fabio Pilla[12], Elena Ciani[13], Olivier Hanotte[1,9], Joram M. Mwacharo[14*]

[1]School of Life Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD UK

[2]Department of Zoology, Faculty of Sciences, Misurata University, Misurata, Libya

[3]Department of Biological Sciences, Faculty of Science, Kuwait University, Safat 13060, Kuwait

[4] Department of Veterinary Public Health and Animal Husbandry, College of Veterinary Medicine, King Faisal University, 400 Al-Ahsa 31982, Saudi Arabia

[5]Department of Animal and Veterinary Sciences, College of Agriculture and Marine Sciences, Sultan Qaboos University, PO box 34 Al Khod, Postal Code 123, Muscat, Oman

[6]Agricultural Research Center, Misurata, Libya

[7]Debre Berhan Research Centre, Debre Berhan, Ethiopia

[8]Amhara Regional Agricultural Research Institute, Bahir Dar, Ethiopia

[9]LiveGene, International Livestock Research Institute, P.O. Box 5689, Addis Ababa, Ethiopia

[10]Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan

[11]Dipartimento di Scienze Agrarie e Forestali, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy

[12]Dipartimento Agricoltura, Ambiente e Alimenti, Università degli Studi del Molise, Campobasso, Italy

[13]Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, Università degli Studi di Bari "Aldo Moro," Bari, Italy

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

*** Correspondence:** Abulgasim.ahbara@nottingham.ac.uk; j.mwacharo@cgiar.org

Keywords: Admixture, Africa, Fat-tail, Ovis aries, Thin-tail

Abstract

Variations in body weight and body fat distribution are associated with feed quality and quantity, thermoregulation and energy reserve. Ethiopia is characterised by distinct agro-ecologies and human ethnic farmer's diversity of ancient origin which have impacted on the variation of its livestock species. Here, we investigate the autosomal genome-wide profiles of 11 Ethiopian indigenous sheep populations using the Illumina Ovine 50K SNP BeadChip assay. Populations from The Caribbean,

Europe, Middle East, China and western, northern and southern Africa were included addressing globally, the genetic variation and history of Ethiopian sheep populations. Population structure, PCA and phylogenetic analysis separate the Ethiopian indigenous fat-tail sheep from the North African and Middle Eastern fat-tailed sheep. It indicates two main genetic backgrounds and supports two distinct genetic history for the African fat-tailed sheep. Within Ethiopia, our results indicate that the short fat - tailed do not represent a monophyletic group. Four genetic backgrounds are present within Ethiopian sheep but at different proportions among fat-rump sheep, long fat-tailed sheep from western Ethiopia and long fat-tailed sheep from southern Ethiopia. Ethiopian fat-rump sheep were also found to share a common genetic background with Sudanese thin-tail sheep. Selection signature analysis identified eight candidate genomic regions that spanned genes influencing growth traits and fat deposition (*NPR2*, *HINT2*, *SPAG8*), embryonic development of tendons, bones and cartilage (*EYA2*, *SULF2*), regulation of body temperature (*DIS3L2*, *LIN28B*) and the control of lipogenesis and intracellular transport of long-chain fatty acids (*SREBF1* and *FABP3*). Our findings indicate that Ethiopian indigenous fat-tail sheep represent a distinct genepool and an important resource for understanding the genetic control of fat metabolism and associated physiological processes.

1 Introduction

African indigenous sheep originate in the Near East, and reached North Africa first, *via* the Isthmus of Suez, by the seventh millennium before present (BP) (Marshall, 2000). These sheep were of thin-tail type and their dispersion to East Africa followed the Nile valley and the Red Sea coastline (Blench and MacDonald, 2006; Gifford-Gonzalez and Hanotte, 2011). The second introductory wave, brought fat-tail sheep into North and Northeast Africa with two entry points, the Isthmus of Suez and the Horn of Africa across the strait of Bab-el-Mandeb, respectively. Fat-rump sheep are believed to be of recent introduction and represent the third wave of arrival and dispersal of the species into East Africa (Epstein, 1971; Ryder, 1983; Marshall, 2000).

Sheep fulfill important socio-cultural and economic roles in the Horn of Africa and particularly in Ethiopia providing a wide range of products, such as meat, milk, skin, hair, and manure, and as a form of savings and investment (Assefa et al., 2015). Ethiopia hosts a large number of local breeds/ecotypes of sheep, with currently 14 recognized populations/types, based on their geographic location and/or the ethnic communities rearing them (Gizaw et al., 2008). Using structure analysis, Edea *et al.* (2017) showed that the five Ethiopian indigenous sheep populations they analysed tended to cluster together based on their geographic distribution and tail phenotypes.

Fat depots act as an energy reserve that allows sheep to survive harsh environments and extreme conditions such as prolonged periods of droughts, cold and food scarcity (Atti et al., 2004; Nejati-Javaremi et al., 2007; Moradi et al., 2012). Based on the combination of tail type and length, Ethiopian sheep can be assigned to four broad groups: short fat-tail, long fat-tail, thin-tail and fat-rump sheep. The short fat-tail types inhabit the sub-alpine regions, the long fat-tailed types predominate in mid- to high-altitude environments and the fat-rumped sheep are distributed in dry lowland environments (Gizaw et al., 2007). These populations are considered to be adapted to their production environments and they represent an important model species that can enhance our understanding of the genome profiles of adaptation to varied environments, tail morphology and fat localization.

Different approaches, that contrast groups of fat and thin-tailed sheep, have been used to identify candidate regions and genes that are associated with tail formation and morphotypes. For instance, Moradi et al. (2012) identified three regions on chromosomes 5, 7 and X that are associated with tail

fat deposition in Iranian sheep breeds. Using two fat-tailed (Laticauda and Cyprus fat-tailed) and 13 Italian thin-tailed breeds, Moioli et al. (2015) identified *BMP2* and *VRTN* as the most likely genes explaining fat-tail phenotype in the studied populations/breeds. Zhu et al. (2016) detected several copy number variations intersecting genes (*PPARA, RXRA and KLF11*) associated with fat deposition in three Chinese native sheep breeds (Large-tailed Han, Altay and Tibetan sheep). Several candidate genes that affect fat-tail development, i.e. *HOXA11, BMP2, PPP1CC, SP3, SP9, WDR92, PROKR1* and *ETAA1*, were identified using genome scans that contrasted fat-tailed and thin-tailed Chinese sheep (Yuan et al., 2016). A whole genome sequencing of extremely short-tailed Chinese sheep revealed the T gene as the best possible candidate, among other nine genes influencing tail size, following its association with vertebral development (Zhi et al. 2017). There is so far no information on the genetic basis of the tail phenotype (fat distribution and sizes) in African indigenous sheep.

In this study, using the Ovine 50K SNP BeadChip genotyping information, we address two main questions (i) the genetic relationships of indigenous Ethiopian sheep of different fat-tail morphotypes in relation to Eurasian as well as North, West and East African sheep, and (ii) the possible genetic control of fat deposition and tail morphotypes in African sheep. Eleven Ethiopian indigenous sheep of different fat-tail types and two thin-tailed sheep populations from Sudan were examined alongside The Caribbean, European, Middle Eastern, China and Africa.

2 MATERIALS and METHODS

2.1 DNA samples and SNP genotyping

The sampling strategy targeted populations/breeds of indigenous sheep from different geographic regions in Ethiopia using altitude as a proxy to define different agro-ecologies and climates (Table 1 and Fig. 1). Efforts were made to include populations representing different sheep tail phenotypes present in the country. Genomic DNA was extracted from 146 ear tissue punches from 11 indigenous sheep populations in Ethiopia. GPS coordinates were recorded for all populations. Twenty samples from two populations of thin-tailed sheep were collected from Sudan (Table 1). DNA was extracted from the ear tissues using the NucleoSpin® Tissue Kit (www.mn-net.com) following the manufacturers protocol. The 166 genomic DNA samples were genotyped using the Ovine 50K SNP BeadChip assay.

Genotypes of Caribbean, European, Middle Eastern and Chinese sheep populations and from western, northern and southern African sheep (Supplementary Table S1) were included in the study to investigate the genetic origins, trajectories of introduction and dispersal of the species into Ethiopia.

2.2 Quality control and genetic diversity analyses

The genotype data consisted of 54240 SNPs composed of 52413, 1449 and 378 autosomal, X chromosome and mitochondrial SNPs, respectively. Input files for downstream analysis were generated by pruning individual SNPs and animals as follows: (1) all unmapped SNPs and those mapping to the sex chromosomes and the mitochondrial DNA; (2) SNPs with minor allele frequency (MAF) ≤ 0.01; (3) animals displaying $\geq 10\%$ missing genotypes, and (4) markers with $\geq 0.05\%$ missing genotypes.

A total of 160 animals and 45328 autosomal SNPs were retained for analysis. The proportion of polymorphic SNPs (*Pn*), representing the fraction of the total number of SNPs displaying both alleles, expected (*He*) and observed (*Ho*) heterozygosity and inbreeding coefficient (*F*) were estimated for each population and across all populations using PLINK (Purcell et al., 2007).

2.3 Population genetic analyses

Principal Components Analyses (PCA) were performed using PLINK (Purcell et al., 2007) to investigate the underlying genetic structure of the studied breeds/populations based on genetic correlations between individuals (Wang et al., 2009). A graphical display of the first two principal components (PC1 and PC2) was generated using GENESIS (Buchmann and Hazelhurst, 2014). Admixture analysis implemented in ADMIXTURE v1.3 software (Alexander et al., 2009) was used to further investigate the underlying genetic structure and estimate the shared genome ancestry proportions of the study populations. A 15 folds cross-validation procedure was used to determine the optimal number (*K*) of ancestral genomic clusters and proportions. To further infer population splits and admixtures, we generated a population-level phylogeny tree using the maximum likelihood (ML) approach implemented in TreeMix (Pickrell et al., 2012). Migration events were sequentially added to the tree to determine the maximum number of migrations explaining possible interactions between the populations.

2.4 Analysis of genome-wide selection signals

For this analysis, 12 of the 13 study populations were separated into four different genetic groups following the PCA analysis. The four groups are: western long fat-tail, southern long fat-tail and fat-rump sheep from Ethiopia and thin-tail sheep from Sudan. One population of short fat-tail (Molale) sheep was included with the fat-rump group and the other population (Gafera) which was distinct from the other populations was excluded from the analysis of selection signatures. Equal numbers of samples were then chosen at random to represent each genetic group. Three comparisons were then performed which contrasted the fat-rump (E1), western Ethiopian long fat-tail (E2) and southern Ethiopian long fat-tail sheep (E3) with the Sudanese thin-tail sheep (S). The F_{ST} values (Weir and Cockerham, 1984) for each SNP and between the genetic groups were calculated using HIERFSTAT package (Goudet, 2005). The F_{ST} values were standardized into *Z*-scores using the formula:

$$\boldsymbol{ZFst} = \frac{\boldsymbol{Fst} - \boldsymbol{\mu Fst}}{\boldsymbol{\sigma Fst}}$$

Where μFst is the overall average weighted F_{ST} value and σF_{ST} is the standard deviation for all windows within the group.

The hapFLK statistics were calculated using the hapFLK v1.2 package (Fariello et al., 2013) to detect selection signatures based on differences in haplotype frequencies between groups of populations. Reynolds distances were converted into a kinship matrix using an R script supplied with the hapFLK v1.2 package. The hapFLK program was then run using the genotypes (*.PED and *.MAP files). The hapFLK statistic and kinship matrix were calculated assuming 10 clusters in the fastPHASE model (-K 10), and the hapFLK statistic was then computed as the average value across 20 expectation maximization (EM) runs to fit the LD model (--nfit = 20). The *P*-values for hapFLK were obtained by running a python script "Scaling_chi^2_hapFLK.py" available on the hapFLK software web page (https://forge-dga.jouy.inra.fr/documents/588) which fits a chi-squared distribution to the empirical distribution.

Using haplotype information, we computed *Rsb* and P_{Rsb} using the R package *rehh* (Gautier and Vitalis, 2012). Haplotypes were estimated with SHAPEIT (Delaneau et al., 2014). Per-SNP *Rsb* scores were transformed into $P_{Rsb} = -\log[\Phi(Rsb)]$. Assuming that the *Rsb* values are normally distributed (under neutrality), P_{Rsb} can be interpreted as $\text{Log}_{10}(1/P)$, where *P* is the one-sided *P*-value associated with the

neutral hypothesis. For each approach, the significant regions were determined as those that had SNPs with the highest *p*-values falling within the top 0.5 % of the empirical distribution.

2.5 Gene annotation

Candidate regions that overlapped between the three selection detection approaches (F_{ST}, hapFLK and *Rsb*) were identified and compared using *intersectBed* function of the *Bed Tools* software (Quinlan and Hall, 2010). Considering the average distance between markers of between 60 to 200 Kb (Moioli et al., 2015), candidate regions and genes were considered by exploring 100 Kb up- and down-stream of the most significant SNPs. The Oar v3.1 reference genome assembly (Jiang et al., 2014) was used to annotate the candidate regions. Enriched functional terms were determined by processing all the identified genes using the functional annotation tool in *DAVID* Bioinformatics resources (Huang et al., 2008). Gene functions were determined using the National Centre for Biotechnology Information (NCBI) Gene database (http://www.ncbi.nlm.nih.gov/gene/), the OMIM online database (http://www.ncbi.nlm.nih.gov/omim/) and a review of literature.

3 RESULTS

3.1 Genetic diversity and population structure

Four metrics were used to evaluate the levels of within-breed genetic diversity (Table 2). The proportion of polymorphic SNPs (*Pn*), expected (*He*) and observed (*Ho*) heterozygosity, and the inbreeding coefficient (*F*) among the 13 populations range from 79.59 (Bonga) to 90.29 (Molale-Menz), 0.277 (Bonga) to 0.319 (Hammary and Kabashi), 0.293 (Bonga) to 0.334 (Arabo), and 0.017 (Gafera-Washera) to 0.071 (Adane), respectively. With the exception of Bonga, Gesses, Kido and Loya whose *Pn* values were below 84%, all the other populations had *Pn* values higher than 87%.

To examine and visualise the distribution of genetic variation within and among Ethiopian populations in the context of the global ovine genepool, we included populations from the Caribbean, Europe, Middle East and China and from western, northern and southern Africa in the PCA analysis. The first two PCs account for 28.15% and 14.42% of the total variation, respectively (Figure 2). PCA1 separates the Ethiopian sheep, South African (Namaqua, Ronderib) fat-tail sheep, thin-tailed (Sudanese) sheep, the West African Djallonke and the Algerian Sidaoun breeds from the other sheep populations. Populations from the Middle East (Najdi, Local Awassi, Omani) and North Africa (Libyan Barbary and Egyptian Ossimi and Barki) are found at the center of the PCA1-PCA2 graph and together with the fat tail sheep from Cyprus and the Chinese sheep they are separated by PCA2 from the African Dorper, Barbados Blackbelly and European (Italian Babaresca, Icelandic, Dorset Horn, Soay) sheep. The Ethiopian short fat-tail sheep are separated into two genetic groups; one group is close to the Ethiopian long fat-tailed and the other group clusters together with the Ethiopian fat-rump sheep. The West African Djallonke sheep clusters close to the South African Ronderib and Namaqua sheep, the two Algerian sheep (Sidaoun and Berber) are separated from each other, and the Cyprus fat-tailed clusters closely to the Chinese sheep (Figure 2).

To obtain a clearer picture of the variation within the fat-tail sheep, we performed the PCA analysis while excluding the thin-tail sheep (Figure 3). PCA1 explains 20.79% of the variation. It separates the Ethiopian fat-tailed sheep from their Middle East (Local Awassi, Najdi, Omani), North Africa (Libyan Barbary, Algerian Berber and Egyptian Ossimi and Barki), Mediterranean (Cyprus fat-tailed, Italian Babaresca) and Chinese counterparts. PCA2, which accounts for 9.6% of the variation, differentiated the three South African breeds from two broad clusters of Ethiopian sheep. Here, as for the global PCA analysis, one short fat-tailed population (Gafera) clusters with the Ethiopian long-fat tailed sheep and

the other (Molale) forms a cluster with the Ethiopian fat-rump sheep. The Middle East sheep cluster together with the North African ones while the ones from the Mediterranean (Italy and Cyprus) unexpectedly cluster together with the Chinese sheep (both thin and fat tailed) considering their geographic distributions (Figure 3).

To further illustrate the distribution of genetic variation in the sheep populations from East Africa, we performed a PCA including the Ethiopian and Sudanese thin-tailed sheep only (Figure 4). PC1, which explains 6.54% of the variation, separates the Ethiopian fat-rump sheep (Adane, Arabo, Kefis), one population of Ethiopian short-fat tailed sheep (Molale) and the thin tailed sheep (Hammari, Kabashi) from the Ethiopian long fat-tailed sheep (Loya, Shubi Gemo, Bonga, Doyogena, Geses, Kido) and one population of Ethiopian short-fat tailed sheep (Gafera). Overall, PCA1 separates populations of fat-rump sheep from those found in the western and southern parts of the country. Further separation of these populations is revealed by PCA2, which explains 3.87% of the genetic variation. In particular, PCA2 separates Molale, Adane and some Arabo animals from Kefis, the remaining Arabo as well as Gafera, Kido and Gesses from Doyogena, Shubi Gemo, Bonga and Loya populations.

The genomic composition of Ethiopian sheep was investigated using ADMIXTURE. Two to six hypothetical ancestral clusters (K) were investigated. The lowest cross-validation error suggests $K = 4$ as the most likely number of genetic backgrounds defining the dataset. The proportion of each genetic background in the study populations at $K = 4$ is shown in Figure 5. We refer to the four genetic backgrounds as A, B, C, and D (Figure 5, Table S2). They occurred with the highest proportions (> 90%) in Loya (A), Bonga, Kido and Gesses (B), Molale and a few individuals of Adane (C) and in the thin-tail sheep (D). Three backgrounds (A, B and C) defined Shubi Gemo and Doyogena. The B and C backgrounds defined Gafera and Molale. Some individuals of Adane were defined by B, C and D background while all individuals of Arabo and Kefis had the C and D genetic backgrounds. ADMIXTURE analysis also show that Gafera, Adane, Molale, Arabo and Kefis shared the "C" genetic background, while Kabashi and Hammari had the "D" genetic background in common with Arabo and Kefis. Shubi Gemo, Loya and Doyogena, all long fat-tail sheep from southern Ethiopia shared the A genetic background.

To further infer the likelihood of gene flow, and provide support for admixture, between populations, we constructed the ML tree (Figure 6) of the study populations using TreeMix. Gafera is revealed as the most genetically distant population. Three broad population clusters that correspond to the geographic location and the type of fat-tail can be observed. The first cluster puts together Adane, Molale, Arabo, Kefis, Kabashi and Hammari. Within this cluster, three sub-clusters can be observed; the first one involves Adane and Molale, the second, Arabo and Kefis and the third Kabashi and Hammari. Adane, Arabo and Kefis are fat-rump, Molale is short fat-tail and Kabashi and Hammari are thin-tail. The third cluster incorporates three populations, Kido, Bonga and Gesses, are all long fat-tail sheep from western Ethiopia. Migration events between the three fat-rump sheep (Adane and both Arabo and Kefis), between the two short-fat tail populations Gafera and Molale, between Shubi Gemo (long fat-tail from southern Ethiopia) and Arabo and Kefis (fat-rump), between Bonga (long fat-tail from western Ethiopia) and Doyogena (long fat-tail from southern Ethiopia) and between Gesses (long fat-tail from western Ethiopia) and the two thin-tail sheep (Kabashi and Hammari) from Sudan are revealed.

3.2 Signature of selection

The PCA (Supplementary Figure S1) revealed three broad groups in Ethiopian sheep (fat-rump (E1), long fat-tail from western Ethiopia (E2) and long fat-tail from southern Ethiopia (E3). One population

of short fat-tail (Molale) sheep was clustered with the fat-rump group and another population (Gafera) was distinct from the other populations. The three groups are genetically distinct from the thin-tailed sheep (Figure 4). The two short fat-tail sheep (Molale and Gefera) clustered separately from each other (see Figure 4). Molale shwed close genetic affinity to the fat-rump sheep while Gefera appeared distinct. For selection signature analysis, we included the former (Molale) with the fat-rump but excluded the latter (Gefera) from the analysis due to small sample size. We selected at random 20 samples to represent the four broad genetic groups and performed selection signature analysis using *hapFLK,* F_{ST} and *Rsb*. The objective was to identify candidate regions associated with tail morphology, fat deposition and possible eco-climatic adaptation. We compared the different groups of Ethiopian fat-tailed sheep (E1, E2, E3) with the Sudanese thin-tailed sheep (S) (see Section 2.4 in Materials and methods). The top 226 SNPs (Table S3), which passed the significance threshold, for each method (E1*S: *hapFLK* > 3.25, *ZFst* > 4.49, *Rsb* > 2.42, 2.75; E2*S: *hapFLK* > 3.14, *ZFst* > 4.27, *Rsb* > 2.87, 2.53; E3*S: *hapFLK* > 3.20, *ZFst* > 4.27, *Rsb* > 2.82, 2.39; top 0.5%) were used to define the candidate regions. Genes present within 100 kb up- and down-stream of the extreme-most significant SNPs defining the candidate regions that overlapped between the three approaches in each pairwise comparisons were considered as potential candidate genes under selection.

For the E1*S pairwise comparison, the Ethiopian fat-rump sheep were differentiated from the Sudanese sheep in twelve candidate regions that overlapped between the three approaches. These were located on Oar2 (three regions), Oar3, Oar5 (three regions), Oar10, Oar11 (three regions) and Oar13. These regions spanned 47 candidate genes (Figure 7, Table 3). Similarly, 18 candidate genes were present across eight candidate regions that overlapped between the three approaches in the E2*S comparison (western Ethiopian long fat-tailed sheep against the Sudanese ones), from which, 10 genes were identified in four overlapping candidate regions on Oar13, five genes on two candidate regions on Oar2 and one gene each on Oar3 and Oar9, respectively (Figure 8, Table 4). Six overlapping candidate regions, that spanned 9 candidate genes, found on Oar2, Oar3, Oar9, Oar11 and Oar20 differentiated southern Ethiopian fat-tailed sheep from the Sudanese ones (E3*S) (Figure 9, Table 5).

We performed gene ontology (GO) enrichment analysis for the candidate genes revealed in each pairwise comparison (Table 6). Candidate genes in the E1*S comparison are involved in embryonic skeletal system morphogenesis (GO:0009952, GO:0048704, GO:0030224, GO:0048706), response to cold (GO:0009409), innervation (GO:0060384), stem cell population maintenance (GO:0019827) and positive regulation of cell adhesion (GO:0045785). The GO terms associated with the E2*S candidate genes included cellular response to heat (GO:0034605), lipid binding (GO:0008289), magnesium ion binding (GO:0000287) and response to gamma radiation (GO:0000287). The GO terms for the genes from the E3*S comparison included skin development (GO:0043588), regulation of actin cytoskeleton reorganization (GO:2000249) and wound healing (GO:0042060). In general, these biological functions likely play pertinent roles associated with tail formation and size, fat localization and local adaptation in Ethiopian sheep.

4 DISCUSSION

We used the Ovine 50K SNP generated genotype data to investigate at the autosomal level the genetic diversity of Ethiopian indigenous sheep populations. Including populations from other regions of the world and the African continent allowed us to assess this diversity in a global geographic context. Our results showed that the Ethiopian indigenous sheep are genetically differentiated from the other populations including the other fat-tailed sheep found in Africa (Figures 2, 3, 4). In particular, the findings at genome-wide level, that the Ethiopian fat-tailed populations are clearly distinct from those found in the North of Africa, support the presence of at least two distinct genetic groups of fat-tailed

sheep in the African continent and two separate introductions *via* Northeast Africa and the Mediterranean Sea coastline and *via* the Horn of Africa through the strait of Bab-el-Mandeb, respectively. Likewise, the distinct clustering of the Sudanese thin-tailed sheep support its independent introduction into the continent. Also, the fact that the South African Ronderib and Namaqua sheep occur on the same PCA planar axis with the Ethiopian sheep (Figure 2) may suggest, most likely, a common genetic heritage between the two rather than with the North African breeds.

These results are in agreement with previous work using microsatellite loci (Muigai, 2003) and 50K SNP genotype data (Mwacharo et al 2017). They are also in line with archeological and anthropological findings indicating that the introduction first, of thin-tail sheep into the continent followed by the introduction of fat-tail sheep, initially through the Sinai Peninsula and later through the Horn of Africa (Gifford-Gonzalez and Hanotte 2011, Muigai and Hanotte 2013).

Interestingly, the PCA results including only Ethiopian and Sudanese sheep separate the Ethiopian populations into three groups while ADMIXTURE analysis revealed four genetic backgrounds in the Ethiopian populations irrespective of the geographic origin of the populations within Ethiopia. Despite of sharing the same tail phenotypes, the two short fat-tail Ethiopian populations examined here do not cluster as a single group. TreeMix also reveals five migration events between the different populations analysed. Taken together, these results indicate likely current and historical intermixing of sheep, as was also observed by Tarekegn et al. (2018) in the case of Ethiopian goats, as a result of socio-cultural links and economic activities. In particular, we propose that the common D genetic background in the short fat-tail and the fat-rump sheep represent the historical introgression of the thin-tail gene pool into the short fat-tail and fat-rump sheep.

Our findings regarding the relationships and differentiation between Ethiopian sheep populations are in agreement with previous studies findings, using either microsatellites (Gizaw et al. 2008) or SNP genotyping (Edea et al. 2017) data which indicate a grouping of Ethiopian indigenous sheep populations based on their tail phenotypes. However, uniquely in our study the long fat-tail populations were further subdivided into two secondary groups representing sheep populations found in the western and southern of Ethiopia (Figure 4). These two groups were also defined by different genetic backgrounds in our admixture analysis (Figure 5) and they clustered separately in TreeMix (Figure 6). Geographic isolation and adaptation to different eco-climates, as well as ethnic, cultural and religious practices and differences that may act as barriers to gene flow may have shaped this population genetic sub-structuring (Madrigal et al., 2001; Gizaw et al., 2007).

In this study, we contrasted groups of Ethiopian indigenous sheep that showed variation in tail lengths and fat distribution with African thin-tail sheep (see materials and methods). Our results identified several genes as strong candidates for tail morphology and fat localization. A large number of the candidate genes occurred within candidate regions that overlapped between at least two of the approaches used to detect signatures of selection (hapFLK, F_{ST} and *Rsb*).

In the E1*S comparison, three genes associated with growth traits were identified on the candidate region located on Oar2, i.e. histidine triad nucleotide binding protein 2 (*HINT2*), sperm associated antigen 8 (*SPAG8*) and natriuretic peptide receptor 2 (*NPR2*). Previous studies reported these genes to be associated with birth and carcass weights and fat depth, respectively, in cattle (Casas et al., 2000; McClure et al., 2010) and sheep (Moradi et al., 2012; Wei et al., 2015). We also identified two genes on Oar5 (*ANGPTL8, INSR*) which might be responsible for fat accumulation in adipose tissues. Angiopoietin-like 8 (*ANGPTL8*), when induced by insulin receptor (*INSR*), inhibits lipolysis and controls post-prandial fat storage in white adipose tissue and directs fatty acids to adipose tissue for

storage during the fed state (Mysore et al., 2017). Our enrichment analysis for the E1*S genes revealed a cluster of genes (*BMP4, PDE1B, PPARG, MED1*) with functions that could possibly be related to tail formation. Bone Morphogenetic Protein 4 and 7 (*BMP4, BMP7*), which are important paralogs of *BMP2*, were revealed by *Rsb* to be on a candidate region on Oar7 and Oar13, respectively; they have been implicated in tail formation (Moioli et al., 2015). *PDE1B (*Phosphodiesterase 1B*)* was reported to be involved in carcass fat distribution in cattle (Stone et al., 2005) while Peroxisome Proliferator Activated Receptor Gamma (*PPARG*) expression has been associated with back-fat thickness in sheep (Dervish et al., 2011). Ge et al., (2008) reported *MED1* (Mediator Complex Subunit 1) was essential for optimal function of *PPARG*.

In the same comparison (E1*S), we identified a cluster of genes (*CDH8, ADRB3, THRA, TRPM8,* and *PLAC8*) that are associated with the GO biological process of response to cold. Indeed, Adreno receptor Beta 3 (*ADRB3*) plays a major role in energy metabolism and regulation of lipolysis and homeostasis (Wu et al., 2012) and is associated with birth weight, growth rate, carcass composition and survival in various sheep breeds (Horrell et al., 2009). The ion channel *TRPM8* has been reported to play a major role in eliciting cold defence thermoregulation, metabolic and defence immune responses in humans (Kozyreva and Voronova, 2015).

Several other genes occurring in the candidate regions identified in the E1*S comparison and associated with the GO term embryonic skeletal system development (GO:0048706) included *HOXC6, T, SULF2, WNT11* and *HOXB9*. The Brachybury homolog (T gene) was identified by *Rsb* on a region on Oar8, *WNT11* identified by *ZFst* on Oar15 and *HOXC6* and *HOXB9* were revealed by *hapFLK* on Oar3 and Oar13, respectively. The WNT gene family and the *T* gene have been reported to be involved in vertebral development in laboratory mice (Greco et al., 1996), and to be related to the short-tail phenotype in sheep (Zhi et al., 2017). In addition, the role of the WNT gene family in lipid metabolic processes in fat-tailed sheep have also been reported (Kang et al., 2017). The *HOX* genes represent transcriptional regulatory proteins that control axial patterning in bilaterians (Garcia-Fernandez, 2005), where the inactivation of one of the *HOX* genes often causes transformations in the identity of vertebral elements (Mallo et al., 2010). *HOX* genes are able to control morphologies along the anteroposterior axis (Lewis et al., 1978). Furthermore, *HOXC11 HOXC12* and *HOXC13* developmental genes were found to be expressed in the tail region indicating their strong associations with tail size and fat development in fat-tailed sheep (Kang et al., 2017).

The candidate regions revealed by the E2*S comparison (Table 4), spanned 18 candidate genes. Three genes of the BPI fold Containing Family B (*BPIFB3, BPIFB4 and BPIFB6*) were present in a candidate region on Oar13. These, along with other paralogs (*BPIFB1, BPIFA3, BPIFB2* and *BPIFA1*), formed a cluster of functional genes related to the GO term lipid binding functional process (Table 6). In contrast to the E1*S comparison, the cluster of genes identified in the E2*S comparison were associated with the GO terms magnesium ion binding, response to gamma radiation and cellular response to heat. This suggests most likely the propensity of this group of sheep to adapt to the eco-climatic conditions prevailing in their home-tract. This is consistent with the humid highland and moist lowland conditions of the geographic area where the populations representing the E2 group (Bonga, Gesses, Kido) were sampled. High fecundity and prolificacy is a common reproductive trait in the Bonga sheep (field observations by the last author). This may explain the occurrence of the *CIB4* and *PRKAA1* in a candidate region in the E2*S comparison. The *CIB4* gene was suggested to be linked, in some way, to high fecundity in the small Tail Han sheep (Yu et al., 2010) and *PRKAA1* is involved in ewe's follicular development process (Foroughinia et al., 2017).

The third comparison (E3*S) resulted in twenty one candidate genes that occurred in candidate regions that were revealed by at least two of the methods used to detect selection signatures (Table 5). Fatty acid binding protein 3 (*FABP3*) and sterol regulatory element binding transcription factor 1 (*SREBF1)* found on candidate regions located on Oar2 and Oar11, respectively are the genes that relate most closely to fat deposition. *SREBF1* along with *PPARG* are the main transcription factors controlling lipogenesis in adipose tissue and skeletal muscle (Ferre et al., 2010), and are mainly regulated by fatty acid-binding proteins (*FABP*) (Lapsys et al., 2000). Recently, Bahnamiri et al., (2018) evaluated the effects of negative and positive energy balances on the expression of these genes in fat-tailed and thin-tailed lambs. They observed differential transcriptional regulation of lipogenesis and lipolysis during negative and positive energy balances in the two groups of lambs. In general, the cluster of genes identified in this comparison were significantly enriched for GO functional term clusters relating to skin development, wound healing and regulation of actin cytoskeleton reorganization (Table 6).

The overlapped genes between all comparisons are shown in Figure 10. The commonest gene between the three comparisons was DIS3 like 3'-5' exoribonuclease 2 (*DIS3L2*). The gene has also been identified in a candidate region in cattle (Gautier et al., 2009) and sheep (de Simoni Gouveia et al., 2017). *DIS3L2* has reportedly been associated with the Perlman syndrome, which is characterized by overweight in humans (Astuti et al., 2012). Seven genes (*GNE*, *CLTA, CCIN, RECK, Hammerhead_HH9, SULF2, NCOA3)* were common between the E1*S and E2*S comparisons. For instance, on Oar2, three genes were identified within the overlapping candidate region, i.e. *CLTA* associated with prion protein deposition in sheep (Filali et al., 2014), *GNE* which is important for the metabolism of sialated oligosaccharides in bovine milk (Wickramasinghe et al., 2011) and *RECK* which encodes an inhibitor of the angiogenesis, invasion and metastasis, DNA methylation, and increased mRNA in cell lines in human (Su 2012). The *SULF2* gene, which most likely affects tail formation, occurred in a candidate region on Oar13. Both *SULF1* and *SULF2* are secreted and expressed in numerous fetal tissues, including bone and cartilage. They play major roles in modulating growth factor signalling during embryo development by removing 6-O-sulfate from heparan sulfate (Ratzka et al., 2010; Dawson 2011), and causes malformations in the sternum, lumbar and tail vertebrae, and an advance in bone differentiation indicating their redundancy in modulating skeletal development (Ratzka et al., 2008).

Furthermore, six genes (*PHB, B4GALNT2, IGF2BP1, UBE2Z, ATP5G1, CALCOCO2)* were in candidate regions that overlapped between the E1*S and E3*S comparisons. Most of these genes were reported to be associated with prolificacy in Lacaune sheep (Drouilhet et al., 2013). The insulin-like growth factor 2 mRNA-binding protein 1 (*IGF2BP1*) plays a major role in regulating skeletal muscle growth, differentiation, and in maintaining homeostasis in adult muscle tissues (Duan et al., 2010). The ATP synthase lipid-binding protein, mitochondrial (*ATP5G1*) was found to be positively correlated with fat thickness in gene transcription profiles in the *Longissimus dorsi* muscle of four genetic groups of Brazilian hair sheep (Lôbo et al., 2012). Finally, two common genes (*TSPAN8, CPQ*) were identified between the E2*S and E3*S comparisons. The *TSPAN8* (Tetraspanin 8) occurred in the candidate region on Oar3; it is among the genes that are reported to be associated with insulin release, insulin sensitivity, and obesity in humans (Grarup et al., 2008).

5 Conclusion

Overall, our results revealed a high level of genetic diversity in Ethiopian indigenous sheep that may be explained by four distinct genetic background (A, B, C, D). However, with the genotypes of the majority of the individuals to be of at least two genetic backgrounds, some levels of either current or historical admixing between the populations are likely. Selection signature analysis identified

candidate genomic regions that spanned genes associated with skeletal structure and morphology, fat deposition and possibly adaptation to environmental selection pressures. Our results indicate that Ethiopian indigenous sheep are a valuable animal genetic resource that can be used to understand genetic mechanisms associated with body fat metabolism and distribution. This is especially important because fat deposits are a crucial component of adaptive physiology in extreme environments and excessive fat deposition in adipose tissue can result in obesity and overweight, and energy metabolism disorders in humans. These ailments are already a major problem in the developed world and are increasing in developing countries.

6 ETHICS STATEMENT

Local regulations were observed during the sampling of all the populations analysed here. Skin tissues importation and/or exportation was permitted by the Ethiopian Ministry of Livestock and Fisheries (Certificate No: 14-160-401-16).

7 CONFLICT OF INTEREST

The authors declare no conflicts of interest.

8 AUTHOR CONTRIBUTIONS

AA, JMM and OH conceived and designed the study/project, AA analysed the data and wrote the manuscript. JMM, OH revised the manuscript. HB provided logistical support in data analysis. SM, FP and EC contributed to genotyping and genotype data of non-Ethiopian breeds (Najdi, Omani, and Libyan Barbary) and provided critical inputs on data analysis and in the writing of the manuscript. FA, MA and MOA supported the sampling and genotyping of Najdi, Omani and Libyan sheep. AK and AA lead and coordinated the sampling of Ethiopian sheep HHM lead and coordinate the sampling of Sudanese sheep. All authors contributed to the interpretation of the results with their knowledge on local indigenous sheep genetic resources of their respective countries. All the authors read and approved the final manuscript.

9 ACKNOWLEDGMENT

This study was conducted during Abulgasim Ahbara's PhD study which is sponsored by the Libyan Ministry of Higher Education and Scientific Research and the University of Misurata. Sampling of Ethiopian sheep was supported by the CGIAR Research Program on Livestock and …….. ICARDA and ILRI thank the donors supporting the CGIAR Research Program on Livestock (Livestock CRP). This study from part of our on-going efforts to understand the adaptation of local indigenous livestock to improve their productivity.

10 REFERENCES

Alexander, D.H., Novembre, J. and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, 19(9), 1655-1664.

Assefa, F., Tigistu, T. and Lambebo, A. (2015). Assessment of the Production System, Major Constraints and Opportunities of Sheep Production in Doyogena Woreda, Kembata Tambaro Zone, Southern Ethiopia. *Assessment*, 5(21).

464 Astuti, D., Morris, M.R., Cooper, W.N., Staals, R.H., Wake, N.C., Fews, G.A., et al. (2012). Germline
465 mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nat.*
466 *Genet*. 44(3), 277.

467 Atti, N., Bocquier, F. and Khaldi, G. (2004). Performance of the fat-tailed Barbarine sheep in its
468 environment: adaptive capacity to alternation of underfeeding and re-feeding periods. A
469 review. *Animal Res*. 53(3), 165-176.

470 Bahnamiri, H.Z., Zali, A., Ganjkhanlou, M., Sadeghi, M., Shahrbabak, H.M. (2018). Regulation of
471 lipid metabolism in adipose depots of fat-tailed and thin-tailed lambs during negative and positive
472 energy balances. *Gene* 641, 203-211.

473 Blench, R. and MacDonald, K. eds., (2006). *The origins and development of African livestock:*
474 *archaeology, genetics, linguistics and ethnography*. Routledge.

475 Buchmann, R., Hazelhurst, S. (2014). Genesis manual. *University of the Witwatersrand,*
476 *Johannesburg.* (available at: http://www. bioinf. wits. ac. za/software/genesis/Genesis.pdf).

477 Casas, E., Shackelford, S.D., Keele, J.W., Stone, R.T., Kappes, S.M., Koohmaraie, M. (2000).
478 Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms
479 of myostatin. *J Anim. Sci*. 78(3), 560-569.

480 Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R.R., et al, (2009).
481 Revealing the history of sheep domestication using retrovirus integrations. Sci. 324(5926),532-536.

482 Dawson, P.A. (2011). Sulfate in fetal development. In *Seminars in Cell & Developmental*
483 *Biology* (Vol. 22, No. 6, 653-659). Academic Press.

484 de Simoni Gouveia, J.J., Paiva, S.R., McManus, C.M., Caetano, A.R., Kijas, J.W., Facó, O., et al.
485 (2017). Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep
486 breeds. *Livest. Sci*. 197, 36-45.

487 Delaneau, O., Marchini, J., McVean, G.A., Donnelly, P., Lunter, G., Marchini, J.L., et al. (2014).
488 Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference
489 panel. *Nat. Commun* 5, 3934.

490 Dervish, E., Serrano, C., Joy, M., Serrano, M., Rodellar, C., Calvo, J.H. (2011). The effect of feeding
491 system in the expression of genes related with fat metabolism in semitendinous muscle in sheep. *Meat*
492 *Sci*. 89(1), 91-97.

493 Devendra, C. and McLeroy, G.B., (1982). Goat and sheep production in the tropics. Intermediate
494 tropical agriculture series.

495 Drouilhet, L., Mansanet, C., Sarry, J., Tabet, K., Bardou, P., Woloszyn, F., Lluch, J., Harichaux, G.,
496 Viguié, C., Monniaux, D. and Bodin, L., 2013. The highly prolific phenotype of Lacaune sheep is
497 associated with an ectopic expression of the B4GALNT2 gene within the ovary. *PLoS Genetics*, 9(9),
498 p.e1003809.

499 Duan, C., Ren, H., Gao, S. (2010). Insulin-like growth factors (IGFs), IGF receptors, and IGF-binding
500 proteins: roles in skeletal muscle growth and differentiation. *General Comparat.*
501 *Endocrinology* 167(3), 344-351.

502 Edea, Z., Dessie, T., Dadi, H., Do, K.T., Kim, K.S. (2017). Genetic diversity and population structure
503 of Ethiopian sheep populations revealed by high-density SNP markers. *Front. Genet*. *8*:218.doi:
504 10.3389/fgene.2017.00218

505 Epstein, H. (1971). *The Origin of Domestic Animals of Africa*, Vol. 2. New York, NY: Africana
506 Publication Corporation, 719.

507 Fariello, M.I., Boitard, S., Naya, H., San Cristobal, M., Servin, B. (2013). Detecting signatures of
508 selection through haplotype differentiation among hierarchically structured
509 populations. *Genet*. 193(3), 929-941.

510 Ferre, P., Foufelle, F. (2010). Hepatic steatosis: a role for de novo lipogenesis and the transcription
511 factor SREBP-1c. *Diabetes, Obesity Metabolism* 12(s2), 83-92.

512 Filali, H., Martín-Burriel, I., Harders, F., Varona, L., Hedman, C., Mediano, D.R., et al. 2014. Gene
513 expression profiling of mesenteric lymph nodes from sheep with natural scrapie. *BMC Genom*. 15(1),
514 59.

515 Foroughinia, G., Fazileh, A., Eghbalsaied, S. (2017). Expression of genes involved in BMP and
516 estrogen signaling and AMPK production can be important factors affecting total number of antral
517 follicles in ewes. *Theriogenology* 91, 36-43.

518 Gaouar, S.B.S., Lafri, M., Djaout, A., El-Bouyahiaoui, R., Bouri, A., Bouchatal, A., Maftah, A., Ciani,
519 E. and Da Silva, A.B., (2017). Genome-wide analysis highlights genetic dilution in Algerian
520 sheep. *Heredity*, 118(3), p.293.

521 Garcia-Fernandez, J. (2005). Hox, ParaHox, ProtoHox: facts and guesses. *Heredity* 94(2), 145.

522 Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., Foulley, J.L.
523 (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC*
524 *Genomics* 10(1), 550.

525 Gautier, M., Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide
526 SNP data from haplotype structure. *Bioinformatics* 28(8), 1176-1177.

527 Ge, K., Cho, Y.W., Guo, H., Hong, T.B., Guermah, M., Ito, M., et al. (2008). Alternative mechanisms
528 by which mediator subunit MED1/TRAP220 regulates peroxisome proliferator-activated receptor γ-
529 stimulated adipogenesis and target gene expression. *Molecul Cellular Bio*. 28(3), 1081-1091.

530 Gifford-Gonzalez, D., Hanotte, O. (2011). Domesticating animals in Africa: Implications of genetic
531 and archaeological findings. *J. World Prehistory* 24(1),1-23.

532 Gizaw, S., (2008). Sheep resources of Ethiopia: genetic diversity and breeding strategy. *A PhD Thesis*
533 *Presented to Wageningen University. 145p*.

534 Gizaw, S., Van Arendonk, J.A., Komen, H., Windig, J.J., Hanotte, O., (2007). Population structure,
535 genetic variation and morphological diversity in indigenous sheep of Ethiopia. *Anim. Genet*. 38(6),
536 621-628.

537 Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular*
538 *Ecology Resources 5*(1), 184-186.

539 Grarup, N., Andersen, G., Krarup, N.T., Albrechtsen, A., Schmitz, O., Jørgensen, T., Borch-Johnsen,
540 K., Hansen, T. and Pedersen, O., (2008). Association testing of novel type 2 diabetes risk alleles in the
541 JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release,
542 insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged
543 Danes. *diabetes*, 57(9),2534-2540.

544 Greco, T.L., Takada, S., Newhouse, M.M., McMahon, J.A., McMahon, A.P., Camper, S.A. (1996).
545 Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial
546 development. *Genes Develop*. 10(3), 313-324.

547 Horrell, A., Forrest, R.H.J., Zhou, H., Fang, Q., Hickford, J.G.H. (2009). Association of the ADRB3
548 gene with birth weight and growth rate to weaning in New Zealand Romney sheep. *Anim. Genet*. 40(2),
549 251-251.

550 Horsburgh, K.A., Prost, S., Gosling, A., Stanton, J.A., Rand, C., Matisoo-Smith, E.A. (2013). The
551 genetic diversity of the Nguni breed of African Cattle (Bos spp.): complete mitochondrial genomes of
552 haplogroup T1. *PloS One* 8(8), e71956.

553 Huang, D.W., Sherman, B.T., Lempicki, R.A. (2008). Systematic and integrative analysis of large gene
554 lists using DAVID bioinformatics resources. *Nat. Protocols* 4(1), 44.

555 Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J.F., Faraut, T., Wu, C., et al. (2014). The sheep
556 genome illuminates biology of the rumen and lipid metabolism. *Sci*. 344(6188), 1168-1173.

557 Kang, D., Zhou, G., Zhou, S., Zeng, J., Wang, X., Jiang, Y., Yang, Y. Chen, Y. (2017). Comparative
558 transcriptome analysis reveals potentially novel roles of Homeobox genes in adipose deposition in fat-
559 tailed sheep. *Scientific Rep* 7(1), 14491.

560 Kozyreva, T.V., Voronova, I.P. (2015). Involvement of neurogenomic regulation in maintenance of
561 temperature homeostasis in the cold. *Russian J Genet. Appl Res. 5*(6), 569-576.

562 Krings, M., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., et al. (1999). mtDNA analysis
563 of Nile River Valley populations: A genetic corridor or a barrier to migration? *The American J. Human*
564 *Genet*. 64(4), 1166-1176.

565 Lapsys, N.M., Kriketos, A.D., Lim-Fraser, M., Poynten, A.M., Lowy, A., Furler, S.M., et al. (2000).
566 Expression of genes involved in lipid metabolism correlate with peroxisome proliferator-activated
567 receptor γ expression in human skeletal muscle. *The J Clin. Endocr. Metab*. *85*(11), 4293-4297.

568 Leta, S., Mesele, F. (2014). Spatial analysis of cattle and shoat population in Ethiopia: growth trend,
569 distribution and market access. *Springer Plus*, *3*(1), 310.

570 Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. In *Genes, Development*
571 *and Cancer* (pp. 205-217). Springer, Boston, MA.

572 Lôbo, A.M.B.O., Guimarães, S.E.F., Paiva, S.R., Cardoso, F.F., Silva, F.F., Júnior, G.A.F., Lôbo,
573 R.N.B. (2012). Differentially transcribed genes in skeletal muscle of lambs. *Livestock Sci*. 150(1), 31-
574 41.

575 Madrigal, L., Ware, B., Miller, R., Saenz, G., Chavez, M., Dykes, D. (2001). Ethnicity, gene flow, and
576 population subdivision in Limon, Costa Rica. *American J. Phys. Anthrop*. 114(2), 99-108.

577 Mallo, M., Wellik, D.M., Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate
578 body plan. *Develop. Biol*. *344*(1),7-15.

579 Marshall, F., 2000. The origins and spread of domestic animals in East Africa. *The origins and*
580 *development of African livestock: Archaeology, genetics, linguistics and ethnography*, pp.191-221.

581 Mastrangelo, S., Moioli, B., Ahbara, A., Latairish, S., Portolano, B., Pilla, F.,Ciani, E. (2018). A
582 genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation.
583 *Anim Prod Sci*. just accepted (http://www.publish.csiro.au/AN/justaccepted/AN17753)

584 Mastrangelo, S., Portolano, B., Di Gerlando, R., Ciampolini, R., Tolone, M., Sardina, M. T. (2017).
585 Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. Animal. 11, 1107-
586 1116.

587 McClure, M.C., Morsci, N.S., Schnabel, R.D., Kim, J.W., Yao, P., Rolf, M.M., et al. (2010). A genome
588 scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in
589 commercial Angus cattle. *Anim. Genet*. 41(6), 597-607.

590 Moioli, B., Pilla, F., Ciani, E. (2015). Signatures of selection identify loci associated with fat tail in
591 sheep. *J Anim. Sci.* 93(10), 4660-4669.

592 Moradi, M.H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K.G., McEwan, J.C. (2012).
593 Genomic scan of selective signals in thin and fat tail sheep breeds for identifying of candidate regions
594 associated with fat deposition. *BMC Genet*. 13(1), 10.

595 Muigai, A. W. T and Hanotte O. (2013). Then origin of African sheep: Archaeological and genetics
596 perspectives. Afr Archaeo.l Rev (2013) 30:3 9–50. DOI 10.1007/s10437-013-9129-0

597 Muigai, A.W.T., 2003. Characterization and conservation of indigenous animal genetic resources:
598 Genetic diversity and relationships of fat-tailed and thin-tailed sheep of Africa. *Jomo Kenyatta*
599 *University of Agriculture and Technology Juja, Kenya*.

600 Mwacharo JM, Kim E-S, Elbeltagy AR, Aboul_naga AM, Rischkowsky BA, Rothschild MF (2017).
601 Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from
602 East Africa and western Asia cohorts. *Scientific Rep* 7, 17647.

603 Mysore, R., Liebisch, G., Zhou, Y., Olkkonen, V.M., Haridas, P.N. (2017). Angiopoietin-like 8
604 (Angptl8) controls adipocyte lipolysis and phospholipid composition. *Chem. Phys Lipids* 207, 246-
605 252.

606 Nejati-Javaremi, A., Izadi, F., Rahmati G., Moradi. M. (2007). Selection in fat-tailed sheep based on
607 two traits of fat-tail and body weight versus single-trait total body weight. *Int. J Agri. Biol*. 9(4), 645-
608 648.

609 Pickrell, J.K., Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide
610 allele frequency data. *PLoS Genet*. 8(11), e1002967.

611 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., et al. (2007). PLINK:
612 a tool set for whole-genome association and population-based linkage analyses. *The American J.*
613 *Human Genet.* 81(3), 559-575.

614 Quinlan, A.R., Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
615 features. *Bioinformatics* 26(6), 841-842.

616 Ratzka, A., Kalus, I., Moser, M., Dierks, T., Mundlos, S., Vortkamp, A. (2008). Redundant function
617 of the heparan sulfate 6-O-endosulfatases Sulf1 and Sulf2 during skeletal development. *Developmental*
618 *Dynamics* 237(2), 339-353.

619 Ratzka, A., Mundlos, S., Vortkamp, A. (2010). Expression patterns of sulfatase genes in the developing
620 mouse embryo. *Developmental Dynamics* 239(6), 1779-1788.

621 Ryder, M.L. (1983). *Sheep and man*. Gerald Duckworth & Co. Ltd.

622 Spangler, G.L., Rosen, B.D., Ilori, M.B., Hanotte, O., Kim, E.S., Sonstegard, T.S., Burke, J.M.,
623 Morgan, J.L., Notter, D.R. and Van Tassell, C.P. (2017). Whole genome structural analysis of
624 Caribbean hair sheep reveals quantitative link to West African ancestry. *PloS One*, *12*(6), p.e0179021.

625 Stone, R.T., Casas, E., Smith, T.P.L., Keele, J.W., Harhay, G., Bennett, et al. (2005). Identification of
626 genetic markers for fat deposition and meat tenderness on bovine chromosome 5: Development of a
627 low-density single nucleotide polymorphism map 1 2. *J Anim. Sci.* 83(10), 2280-2288.

628 Su, L.J., (2012). Diet, epigenetics, and cancer. In *Cancer Epigenetics* (pp. 377-393). Humana Press,
629 Totowa, NJ.

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

633 Wang, D., Sun, Y., Stang, P., Berlin, J.A., Wilcox, M.A., Li, Q. (2009). Comparison of methods for
634 correcting population stratification in a genome-wide association study of rheumatoid arthritis:
635 principal-component analysis versus multidimensional scaling. In *BMC proceedings* Vol. 3, No. 7, p.
636 S109.

637 Wei, C., Wang, H., Liu, G., Wu, M., Cao, J., Liu, Z., et al. (2015). Genome-wide analysis reveals
638 population structure and selection in Chinese indigenous sheep breeds. *BMC Genom.* 16(1), 194.

639 Weir, B.S., Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population
640 structure. *Evolution* 38(6), 1358-1370.

641 Wickramasinghe, S., Hua, S., Rincon, G., Islas-Trejo, A., German, J.B., Lebrilla, C.B., Medrano, J.F.
642 (2011). Transcriptome profiling of bovine milk oligosaccharide metabolism genes using RNA-
643 sequencing. *PLoS One* 6(4), e18895.

644 Wilson, R.T., (1991). Small ruminant production and the small ruminant genetic resource in tropical
645 Africa (Animal Production and Health Paper No 88). *Food and Agriculture Organization: Rome,*
646 *Italy*, *106114*.

647 Wu, J., Qiao, L., Liu, J., Yuan, Y., Liu, W. (2012). SNP variation in ADRB3 gene reflects the breed
648 difference of sheep populations. *Molecular Biol. Rep* 39(8), 8395-8403.

649 Yu, Y., Zhang, Y., Song, X., Jin, M., Guan, Q., Zhang, Q., et al. (2010). Alternative splicing and tissue
650 expression of CIB4 gene in sheep testis. *Anim. Repro. Sci.* 120(1-4), 1-9.

651 Yuan, Z., Liu, E., Liu, Z., Kijas, J.W., Zhu, C., Hu, S., et al. (2017). Selection signature analysis reveals
652 genes associated with tail type in Chinese indigenous sheep. *Anim. Genet*. 48(1), 55-66.

653 Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins,
654 diffusion, and impact. *Proc. Natl. Acad. Sci. USA* 105, 11597-604.

655 Zhi, D., Da, L., Liu, M., Cheng, C., Zhang, Y., Wang, et al. (2017). Whole Genome Sequencing of
656 Hulunbuir Short-Tailed Sheep for Identifying Candidate Genes Related to the Short-Tail
657 Phenotype. *G3: Gene Genom. Genet*. 3, 300307.

658 Zhi, D., Da, L., Liu, M., Cheng, C., Zhang, Y., Wang, X., et al. (2018). Whole Genome Sequencing of
659 Hulunbuir Short-Tailed Sheep for Identifying Candidate Genes Related to the Short-Tail
660 Phenotype. *G3: Gene Genom. Genet*. 8(2), 377-383.

661 Zhu, C., Fan, H., Yuan, Z., Hu, S., Ma, X., Xuan, J., et al. (2016). Genome-wide detection of CNVs in
662 Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP
663 arrays. *Scientific Rep.* 6, 27822.

664

665

666

667

668

669

670

671

672

673

674 **Table 1** Description of the sampled populations

Origin	Population	Zone	Latitude (N)	Longitude (E)	Altitude	N	Tail Type	Agro-ecology
Ethiopian	Kefis	Zone 3	9°30’	40°10’	890 M	14	Fat-Rump	Arid lowland
	Adane	South Wollo	11°14’	39°50’	2450 M	12	Fat-Rump	Cool highland
	Arabo	South Wollo	11°31’	36°54’	1500 M	10	Fat-Rump	Cool highland
	Gafera (Washera)	Agew Awi	11°31’	36°54’	2500 M	15	Short, fat tail	Wet, warmer mid-highlands
	Molale (Menz)	North Shewa	10°70’	39°39’	3068 M	15	Short, fat tail	Sub-alpine
	Bonga	Keffa	7°16’	36°15’	1788 M	15	Long, fat tail	Humid mid-highland
	Gesses	Metekel	10°50’	36°14’	1300 M	10	Long, fat tail	Moist lowlands
	Kido	Metekel	10°71’	36°19’	1300 M	10	Long, fat tail	Moist lowlands
	Doyogena	Kembata Tembara	7°21’	37°47’	2324 M	15	Long, fat tail	Cool, wet highlands
	Shubi Gemo	East Shewa	8°80’	38°51	1600 M	15	Long, fat tail	Cool, wet highlands
	Loya	Sidama	6°29’	38°24’	1900 M	15	Long, fat tail	Cool, wet highlands

Sudan	Hammari	North Kurdufan	13°09’	29°22’	620 M	11	Long, tail	thin	Arid lowland
	Kabashi	North Kurdufan	13°09’	29°22’	620 M	9	Long, tail	thin	Arid lowland
Total						166			

675

676 **Table 2** Measures of genetic diversity for each of the 13 populations analysed

Breed	***N***	***P_n*** **(%)**	***H_e***	***H_o***	***F***
Kefis	14	89.95	0.316	0.328	0.035
Adane	12	88.85	0.315	0.319	0.071
Arabo	10	88.69	0.317	0.334	0.050
Molale (Menz)	15	90.29	0.316	0.319	0.055
Gafera (Washera)	15	87.54	0.303	0.318	0.017
Bonga	9	79.59	0.277	0.293	0.038
Kido	10	82.18	0.290	0.310	0.038
Gesses	10	83.09	0.294	0.317	0.027

Doyogena	15	87.17	0.302	0.308	0.044
Loya	15	83.58	0.286	0.294	0.039
Shubi Gemo	15	88.40	0.304	0.313	0.037
Hammari	11	89.93	0.319	0.332	0.038
Kabashi	9	88.64	0.319	0.328	0.025

677

678 **Table 3** Candidate genes within overlapping *hapFLK*, *ZFst* and *Rsb* candidate regions under positive selection (Fat-Rump versus Sudanese
679 thin-tailed)

Chr	Overlapping region	Gene location	Candidate gene	Annotation
2	232644986-232845013	232749221-233048136	DIS3L2	DIS3 like 3'-5' exoribonuclease 2
	51904499-52092919	51989342-52042116	GNE	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase
	52104988-52323079	52048202-52065307	CLTA	clathrin light chain A
		52087650-52089416	CCIN	Calicin

		52128947-52210749	RECK	reversion inducing cysteine rich protein with kazal motifs
		52181262-52181338	Hammerhead_HH9	Hammerhead ribozyme HH9
	52413944-53104335	52421111-52423389	HINT2	histidine triad nucleotide binding protein 2
		52423298-52426475	SPAG8	sperm associated antigen 8
		52423842-52445175	NPR2	natriuretic peptide receptor 2
		52480200-52481163	MSMP	microseminoprotein, prostate associated
		52480334-52485038	RGP1	RGP1 homolog, RAB6A GEF complex partner 1
		52485320-52495944	GBA2	glucosylceramidase beta 2
		52496387-52500153	CREB3	cAMP responsive element binding protein 3
		52506528-52531560	TLN1	talin 1

		52537459-52544952	TPM2	tropomyosin 2
		52546134-52551851	CA9	carbonic anhydrase 9
		52560703-52563910	ARHGEF39	Rho guanine nucleotide exchange factor 39
		52564548-52567161	CCDC107	coiled-coil domain containing 107
		52572730-52573775	SIT1	signaling threshold regulating transmembrane adaptor 1
		52594675-52607206	CD72	CD72 molecule
		52603605-52607846	TESK1	testis-specific kinase 1
		52616756-52618641	FAM166B	family with sequence similarity 166 member B
		52619243-52632387	RUSC2	RUN and SH3 domain containing 2
		52817902-53036532	UNC13B	unc-13 homolog B

		53056098-53059144	FAM214B	family with sequence similarity 214 member B
		53061224-53067598	STOML2	stomatin like 2
		53070391-53079125	PIGO	phosphatidylinositol glycan anchor biosynthesis class O
		53079030-53084363	FANCG	Fanconi anemia complementation group G
		53089776-53099744	VCP	valosin containing protein
3	131513798-131666499	131500322-131523998	USP44	ubiquitin specific peptidase 44
5	13304764-13358703	13290740 13330992	DOCK6	dedicator of cytokinesis 6
		13317907-13320915	ANGPTL8	angiopoietin like 8
	13377005-13404764	13376942-13392573	TSPAN16	tetraspanin 16
	13707943-13849403	13733596-13879145	INSR	insulin receptor

10	29164065-29264065	28986741 29188660	FRY	FRY microtubule binding protein
11	36653681-36777181	36736432-36746262	PHB	Prohibitin
	36978484-37078484	36929322-36992982	B4GALNT2	beta-1,4-N-acetyl-galactosaminyltransferase 2
		37058898-37099743	IGF2BP1	insulin like growth factor 2 mRNA binding protein 1
	37162531-37262531	37146942-37164597	UBE2Z	ubiquitin conjugating enzyme E2 Z
		37173130-37175267	ATP5G1	ATP synthase, H+ transporting, mitochondrial Fo complex subunit C1
		37227823-37243185	CALCOCO2	calcium binding and coiled-coil domain 2
	37949285-38049285	37972076-37981743	NFE2L1	nuclear factor, erythroid 2 like 1
		37992980-38001708	COPZ2	coatomer protein complex subunit zeta 2
		38037788-38047808	CDK5RAP3	CDK5 regulatory subunit associated protein 3

13	75263522-75839049	75066765-75328455	EYA2	EYA transcriptional coactivator and phosphatase 2
		75666854-75730764	NCOA3	nuclear receptor coactivator 3
		75726734-75771128	SULF2	sulfatase 2

680

681 **Table 4** Candidate genes within overlapping *hapFLK, ZFst* and *Rsb* candidate regions under positive selection (Western Long Fat-tailed
682 versus Sudanese thin-tailed)

Chr	Overlapping region	Gene location	Candidate gene	Annotation
2	232644986-232845013	232749221-233048136	DIS3L2	DIS3 like 3'-5' exoribonuclease 2
	51992919-52092919	51989342-52042116	GNE	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase
		52048202-52065307	CLTA	clathrin light chain A
		52087650-52089416	CCIN	Calicin
	52104988-52204988	52128947-52210749	RECK	reversion inducing cysteine rich protein with kazal motifs
		52181262-52181338	Hammerhead_HH9	Hammerhead ribozyme HH9

3	106881919-107331750	107108271-107174474	TSPAN8	tetraspanin 8
9	79952652-80052652	79583307-80022406	CPQ	carboxypeptidase Q
13	61450582-61594570	61459737-61515972	DNMT3B	DNA methyltransferase 3 beta
		61523883-61574930	EFCAB8	EF-hand calcium binding domain 8
		61581681-61607701	SUN5	Sad1 and UNC84 domain containing 5
	61642091-61742091	61641482-61656002	BPIFB6	BPI fold containing family B member 6
		61665357-61680683	BPIFB3	BPI fold containing family B member 3
		61689117-61711550	BPIFB4	BPI fold containing family B member 4
	63949422:64077915	63957714 63986272	EDEM2	ER degradation enhancing alpha-mannosidase like protein 2
		64012722-64016764	PROCR	protein C receptor
	75532759-75756932	75666854-75730764	NCOA3	nuclear receptor coactivator 3
		75726734-75771128	SULF2	sulfatase 2

683

684 **Table 5** Candidate genes within overlapping *hapFLK, ZFst* and *Rsb* candidate regions under positive selection (Southern Long Fat-tailed
685 versus Sudanese thin-tailed)

Chr	Overlapping region	Gene location	Candidate gene	Annotation
2	232644986-232845013	232749221-233048136	DIS3L2	DIS3 like 3'-5' exoribonuclease 2
	234978683-235052832	234989907-235002810	TINAGL1	tubulointerstitial nephritis antigen like 1
	235131414-235231414	235135457-235145925	FABP3	fatty acid binding protein 3
3	106881919-107331750	107108271-107174474	TSPAN8	tetraspanin 8
9	79952652-80052652	79583307-80022406	CPQ	carboxypeptidase Q
11	34145242-34245242	34176887-34191779	SREBF1	sterol regulatory element binding transcription factor 1
	36365602-36465602	36354359-36388790	KAT7	lysine acetyltransferase 7
		36435040-36455430	FAM117A	family with sequence similarity 117 member A
		36455228-36464252	SLC35B1	solute carrier family 35 member B1

	36653681-36777181	36736432-36746262	PHB	Prohibitin
	36978484-37078484	36929322-36992982	B4GALNT2	beta-1,4-N-acetyl-galactosaminyltransferase 2
		37058898-37099743	IGF2BP1	insulin like growth factor 2 mRNA binding protein 1
	37162531-37262531	37146942-37164597	UBE2Z	ubiquitin conjugating enzyme E2 Z
		37227823-37243185	CALCOCO2	calcium binding and coiled-coil domain 2
	37162531-37262531	37173130-37175267	ATP5G1	ATP synthase, H+ transporting, mitochondrial Fo complex subunit C1
20	9547911-9785788	9523663-9535319	FANCE	Fanconi anemia complementation group E
		9551273-9570868	TEAD3	TEA domain transcription factor 3
		9574110-9588018	TULP1	tubby like protein 1
		9692040-9766541	FKBP5	FK506 binding protein 5
	10877621:10977621	10858811-10881027	PPIL1	peptidylprolyl isomerase like 1
		10946568-10957116	PI16	peptidase inhibitor 16

686

687 **Table 6** Significantly enriched functional term clusters and their enrichment scores following DAVID analysis for genes identified in
688 Ethiopian and Sudanese sheep

ID	Term	P value	Associated genes	Comparison
GO:0009952	anterior/posterior pattern specification	0.0001	*HOXB3, HOXC6, HOXB4, HOXB1, HOXC8, HOXB2, HOXB7, HOXB5, HOXC4, HOXB6, HOXB9*	*Fat-Rump vs Sudanese*
GO:0048704	embryonic skeletal system morphogenesis	0.0010	*HOXB3, HOXB4, HOXB1, HOXB2, HOXB7, HOXB5, HOXB6*	*Fat-Rump vs Sudanese*
GO:0009409	response to cold	0.0040	*CDH8, ADRB3, THRA, TRPM8, PLAC8*	*Fat-Rump vs Sudanese*
GO:0030224	monocyte differentiation	0.0045	*BMP4, PDE1B, PPARG, MED1*	*Fat-Rump vs Sudanese*
GO:0048706	embryonic skeletal system development	0.0096	*HOXC6, T, SULF2, WNT11, HOXB9*	*Fat-Rump vs Sudanese*
GO:0060384	Innervation	0.0149	*SULF2, RNF165, LRIG2, UNC13B*	*Fat-Rump vs Sudanese*
GO:0019827	stem cell population maintenance	0.0161	*MED28, NODAL, DIS3L2, MED24, LEO1, FZD7*	*Fat-Rump vs Sudanese*
GO:0045785	positive regulation of cell adhesion	0.0209	*VAV3, ERBB2, ITGAV, ANGPT1, SKAP1*	*Fat-Rump vs Sudanese*
GO:0034605	cellular response to heat	0.0033	*TFEC, CLPB, NF1, SLC52A3, MYOF*	*Western vs Sudanese*
GO:0008289	lipid binding	0.0051	*BPIFB1, BPIFA3, BPIFB2, BPIFB3, BPIFB4, BPIFA1, BPIFB6*	*Western vs Sudanese*
GO:0000287	magnesium ion binding	0.0069	*GSS, CIB4, EYA2, GTPBP10, SNCA, ATP10A, DIS3L2, ERN1, ITPK1, STK3, ADPRH*	*Western vs Sudanese*
GO:0010332	response to gamma radiation	0.0162	*BRCA2, TRIM13, PRKDC, PRKAA1*	*Western vs Sudanese*
GO:0043588	skin development	0.0027	*COL3A1, ITGA3, PTCH2, ARRDC3, COL5A2, DHCR24*	*Southern vs Sudanese*
GO:2000249	regulation of actin cytoskeleton reorganization	0.0072	*GMFG, SEMA3E, RAPGEF3, ARHGDIB*	*Southern vs Sudanese*

GO:0042060	wound healing	0.0232	*PPARD, COL3A1, NF1, GRHL3, PAK1*	*Southern vs Sudanese*

689

690 **FIGURES LEGEND;**

691 **Figure 1** Ethiopian and Sudanese sampling location

692 **Figure 2** Genetic variation among the Ethiopian population in a global geographic context

693 **Figure 3** Distribution of the genetic variation among the worldwide fat-tailed sheep

694 **Figure 4** Distribution of the genetic variation among the East African sheep populations (PC1 and
695 PC2)

696 **Figure 5** Admixture analysis of the studied populations. (K4*= K-value with the lowest cross-
697 validation error)

698 **Figure 6** Tree topology with five migration events inferred by TreeMix

699 **Figure 7** Manhattan plots of genome-wide autosomal *hapFLK*, *ZFst* and *Rsb* analyses of Ethiopian
700 Fat-rump (E1) *versus* Sudanese (S) sheep

701 **Figure 8** Manhattan plots of genome-wide autosomal *hapFLK, ZFst* and *Rsb* analyses of long fat-tailed
702 sheep from western Ethiopia (E2) *versus* Sudanese (S) sheep

703 **Figure 9** Manhattan plots of genome-wide autosomal *hapFLK, ZFst* and *Rsb* analyses of Ethiopian
704 long fat-tailed (E3) versus Sudanese (S) sheep

705 **Figure 10** Venn diagram showing distribution and sharing of genes identified by the different
706 comparisons of sheep

707

708

709

710

711

712

713

714

715

716

Figure 1.JPEG

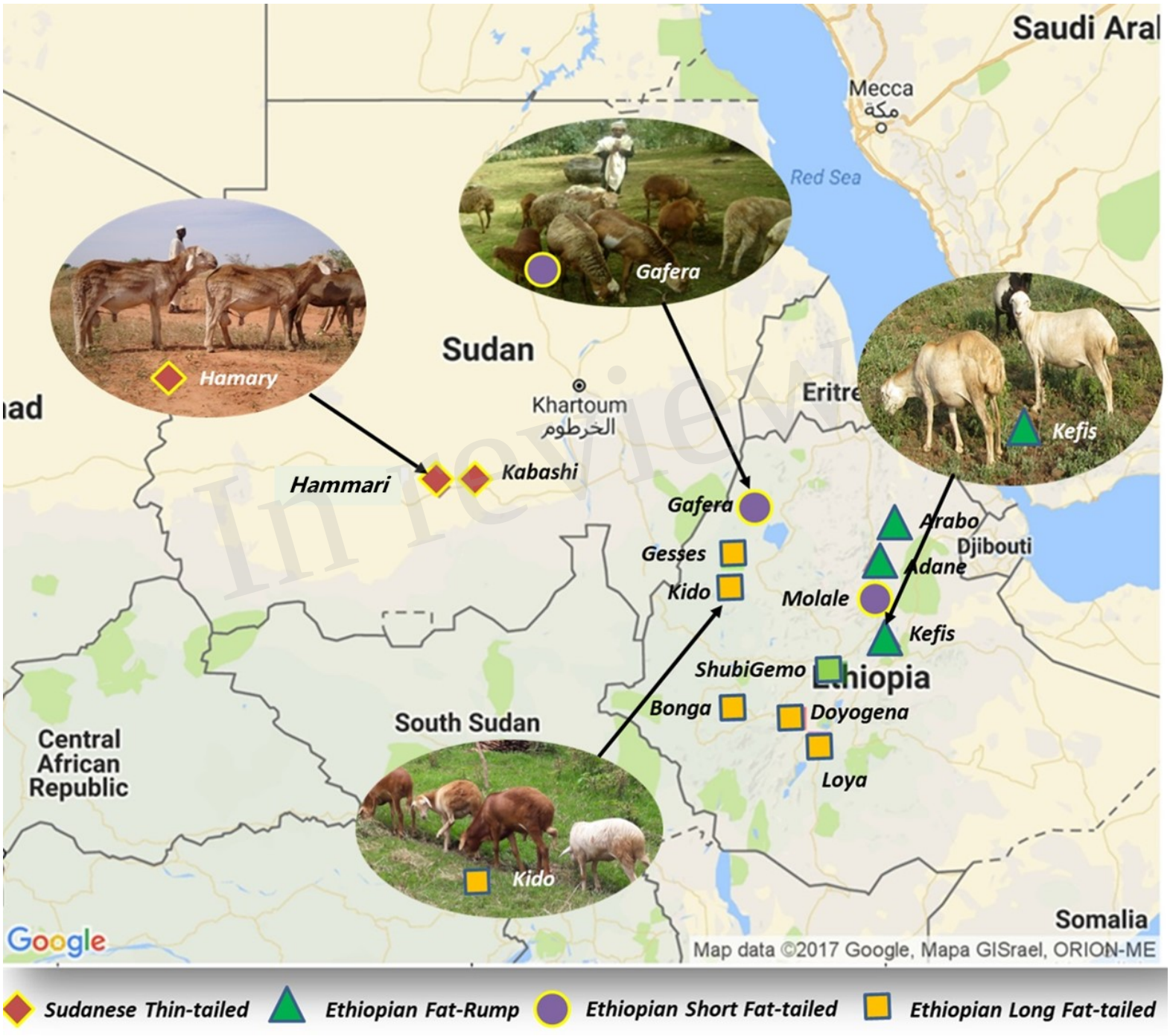

Figure 2.JPEG

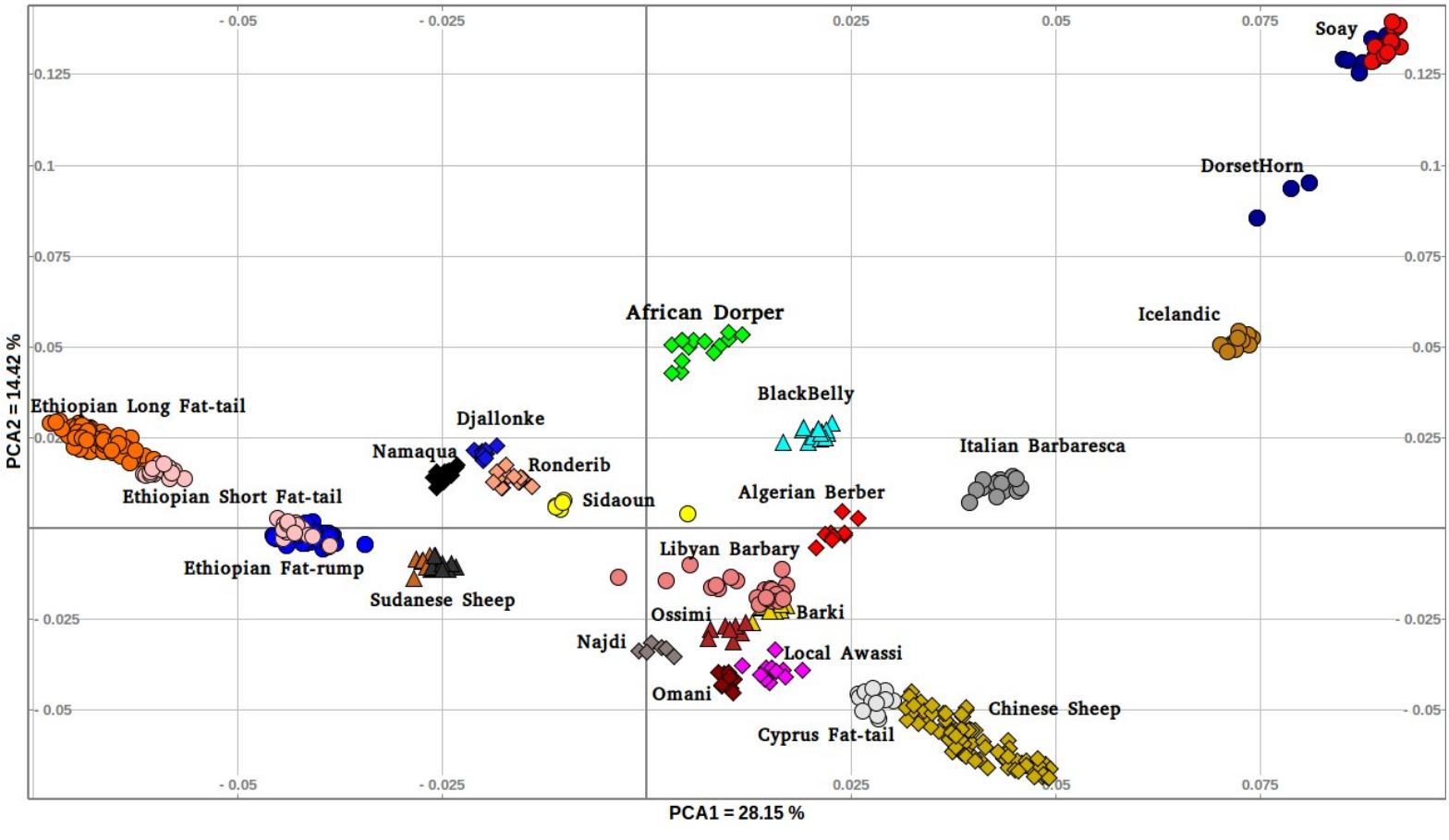

Figure 3.JPEG

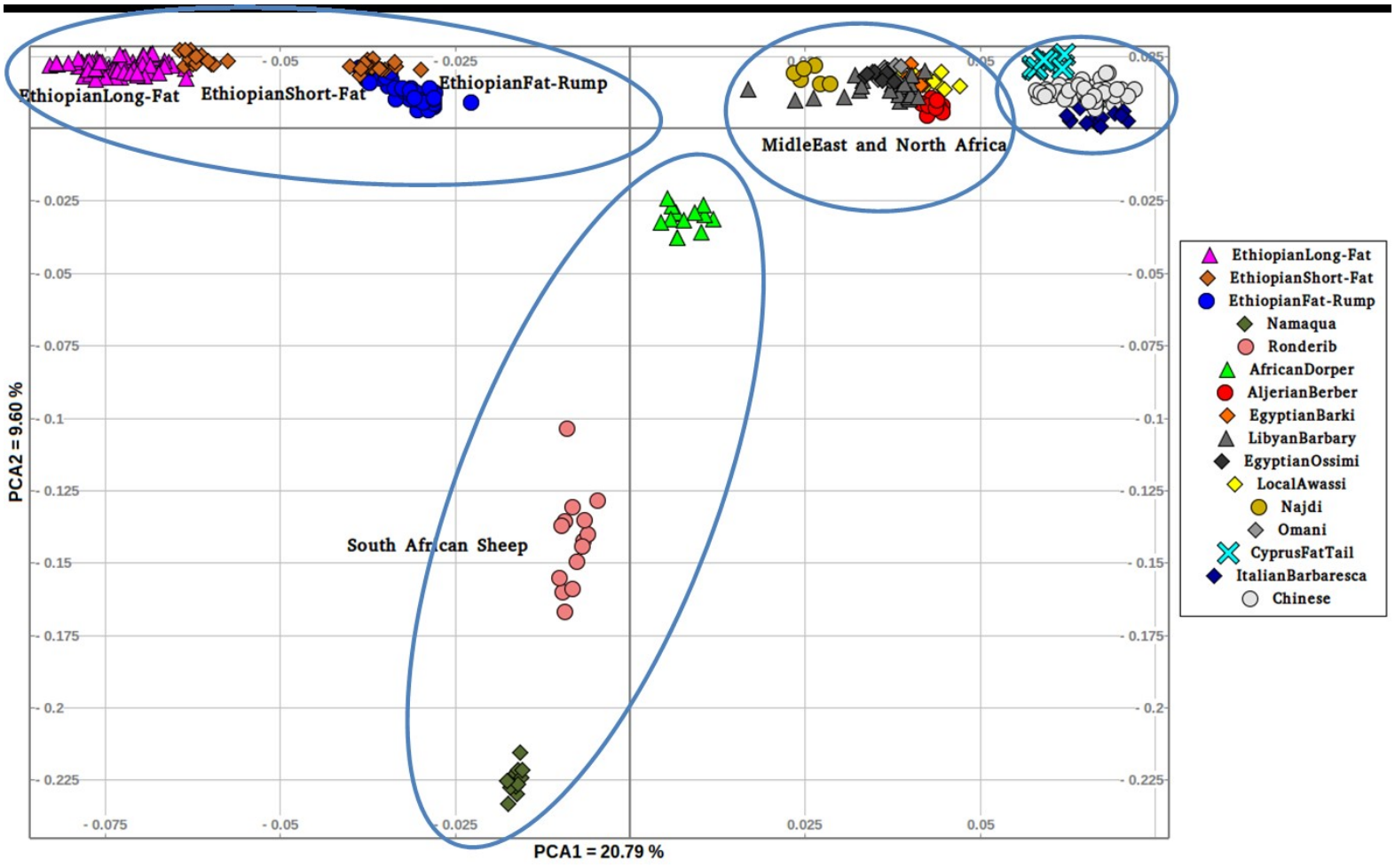

Figure 4.JPEG

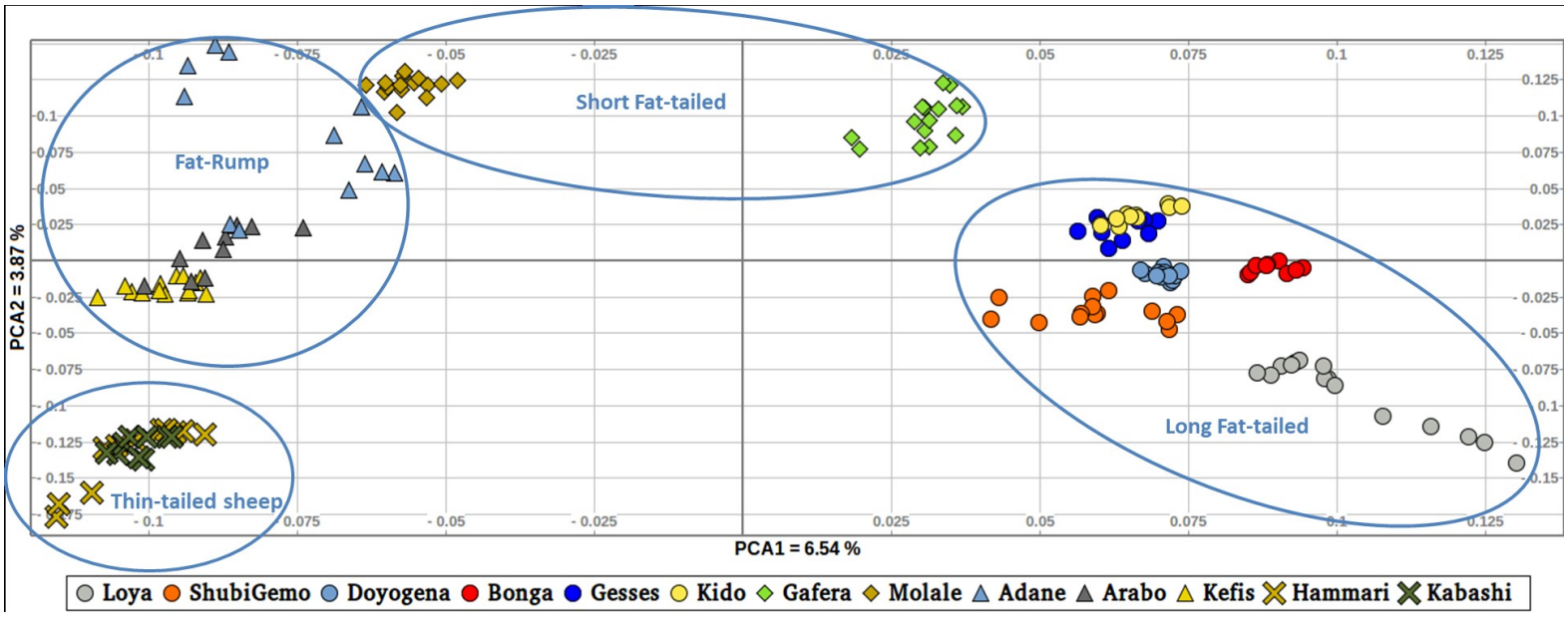

Figure 5.JPEG

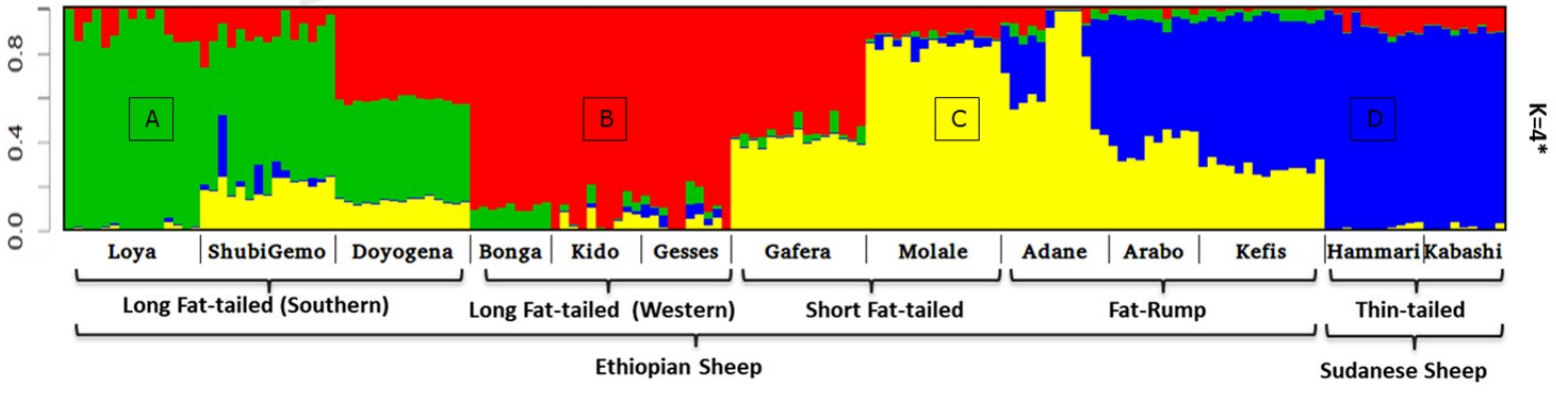

Figure 6.JPEG

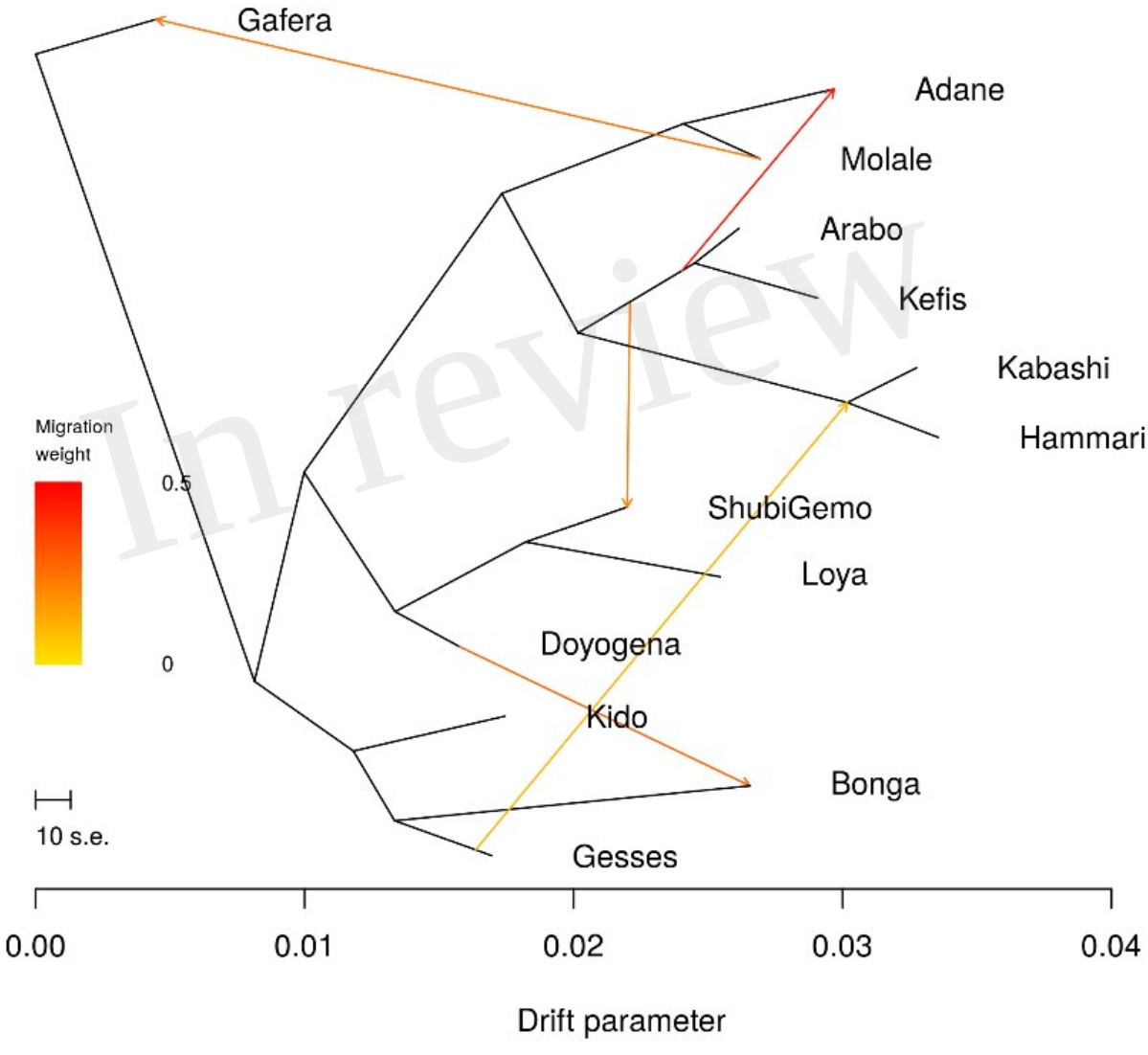

Figure 7.JPEG

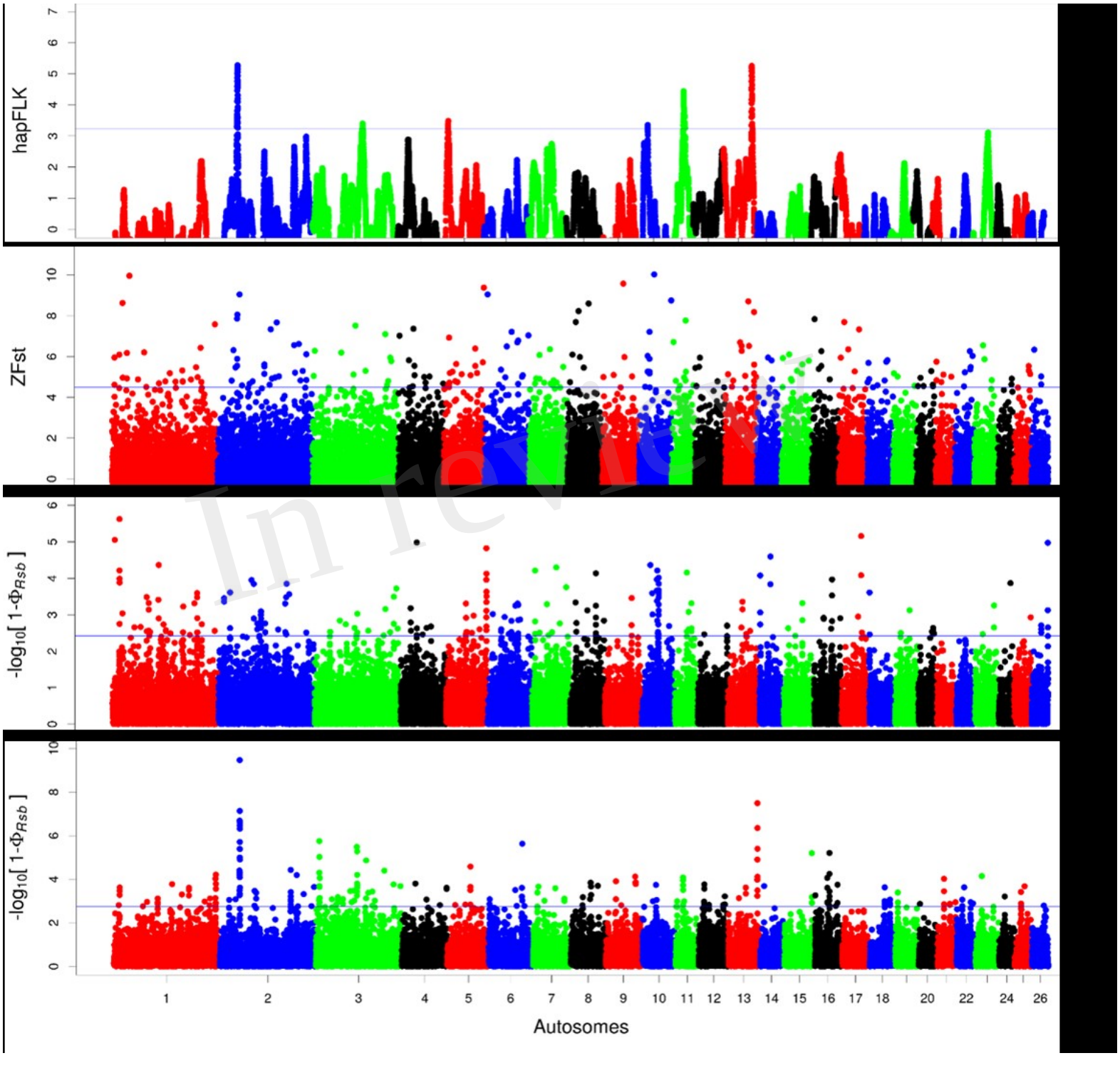

Figure 8.JPEG

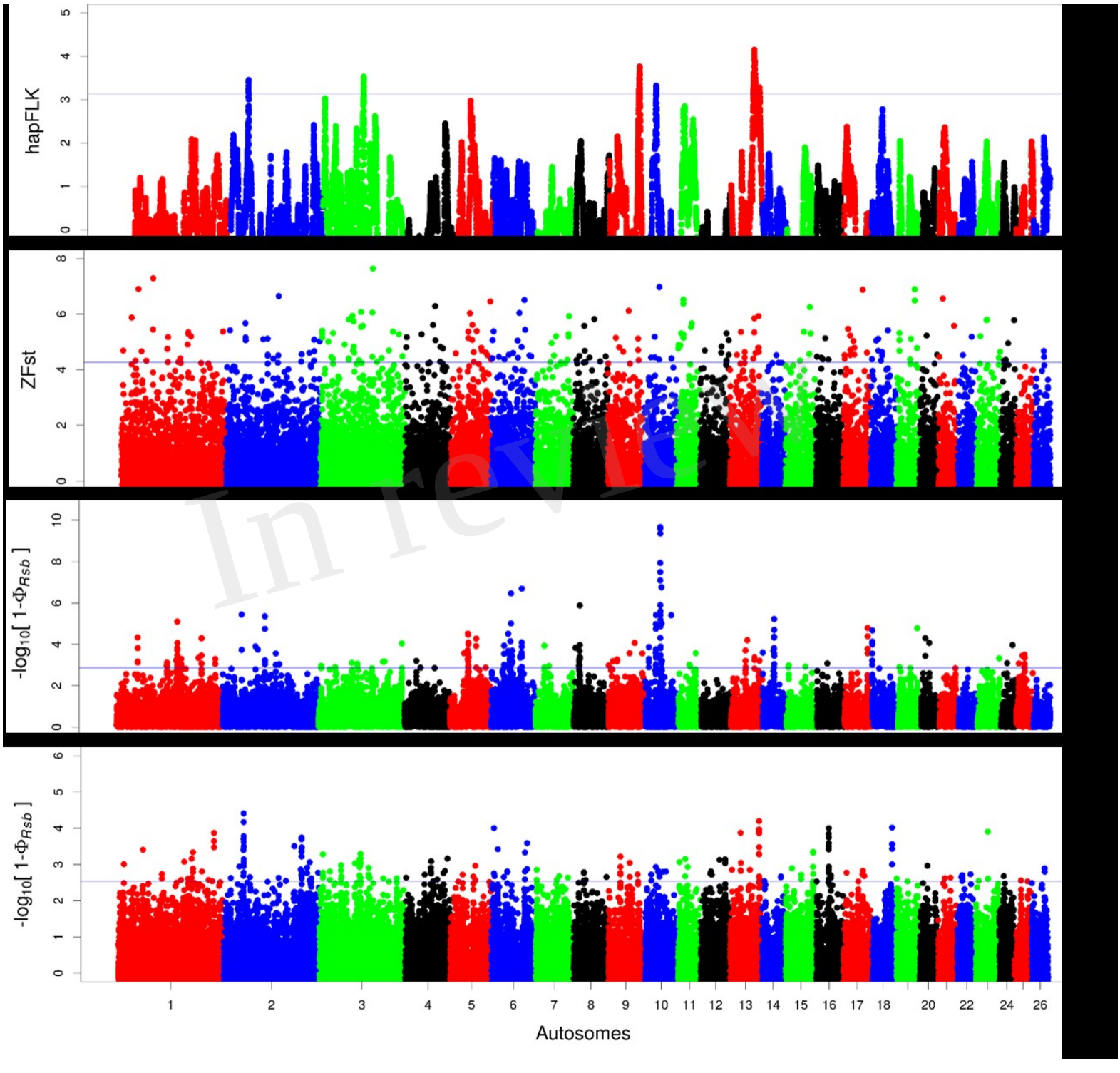

Figure 9.JPEG

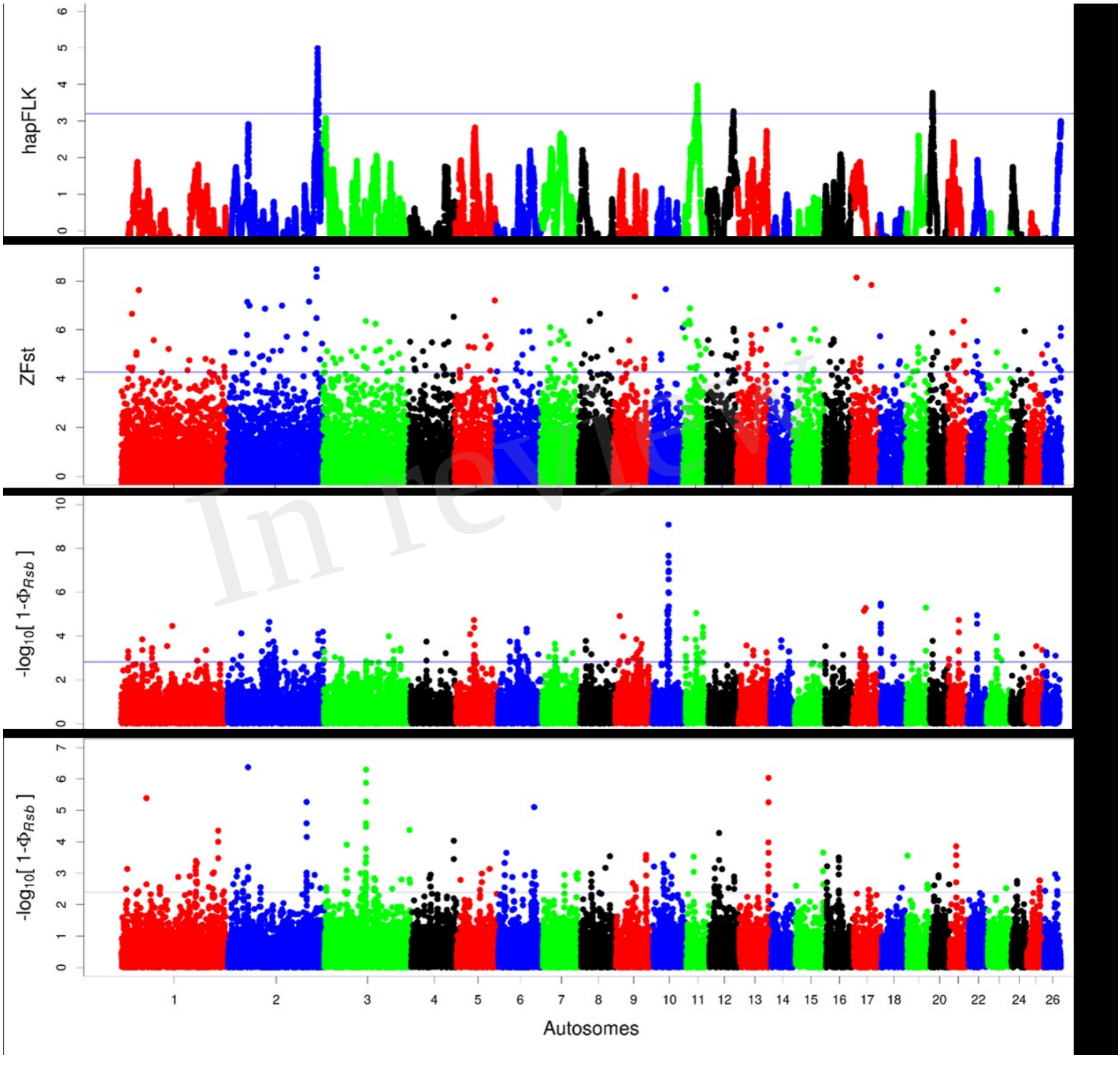

Figure 10.JPEG