

RESEARCH ARTICLE

Open Access



# Whole genome population genetics analysis of Sudanese goats identifies regions harboring genes associated with major traits

Siham A. Rahmatalla<sup>1,2</sup>, Danny Arends<sup>1</sup> , Monika Reissmann<sup>1</sup>, Ammar Said Ahmed<sup>1</sup>, Klaus Wimmers<sup>3</sup>, Henry Reyer<sup>3</sup> and Gudrun A. Brockmann<sup>1\*</sup>

## Abstract

**Background:** Sudan is endowed with a variety of indigenous goat breeds which are used for meat and milk production and which are well adapted to the local environment. The aim of the present study was to determine the genetic diversity and relationship within and between the four main Sudanese breeds of Nubian, Desert, Taggar and Nilotic goats. Using the 50 K SNP chip, 24 animals of each breed were genotyped.

**Results:** More than 96% of high quality SNPs were polymorphic with an average minor allele frequency of 0.3. In all breeds, no significant difference between observed (0.4) and expected (0.4) heterozygosity was found and the inbreeding coefficients ( $F_{IS}$ ) did not differ from zero.  $F_{ST}$  coefficients for the genetic distance between breeds also did not significantly deviate from zero. In addition, the analysis of molecular variance revealed that 93% of the total variance in the examined population can be explained by differences among individuals, while only 7% result from differences between the breeds. These findings provide evidence for high genetic diversity and little inbreeding within breeds on one hand, and low diversity between breeds on the other hand. Further examinations using Nei's genetic distance and STRUCTURE analysis clustered Taggar goats distinct from the other breeds. In a principal component (PC) analysis, PC1 could separate Taggar, Nilotic and a mix of Nubian and Desert goats into three groups. The SNPs that contributed strongly to PC1 showed high  $F_{ST}$  values in Taggar goat versus the other goat breeds. PCA allowed us to identify target genomic regions which contain genes known to influence growth, development, bone formation and the immune system.

**Conclusions:** The information on the genetic variability and diversity in this study confirmed that Taggar goat is genetically different from the other goat breeds in Sudan. The SNPs identified by the first principal components show high  $F_{ST}$  values in Taggar goat and allowed to identify candidate genes which can be used in the development of breed selection programs to improve local breeds and find genetic factors contributing to the adaptation to harsh environments.

**Keywords:** SNP analysis, Genetic diversity, Adaptation, Desert goat, Nilotic goat, Nubian goat, Taggar goat

\* Correspondence: gudrun.brockmann@agrar.hu-berlin.de

<sup>1</sup>Albrecht Daniel Thaer-Institut für Agrar- und Gartenbauwissenschaften, Humboldt-Universität zu Berlin, Invalidenstraße 42, D-10115 Berlin, Germany  
Full list of author information is available at the end of the article



## Background

Compared to other African countries, goats in Sudan constitute a large part of the livestock population. With an estimated number of 31 million goats (out of 365 million in whole Africa) that produced about 1.532 million tons of milk in 2013, Sudan was the largest producer of goat milk in Africa and the third largest producer in the world [1]. Goats in Sudan have an important contribution to food security by producing milk and meat. Beyond that, manure and skins provide a source of income for farmers. Thus, goats constitute an important source of livelihood, social security and rural economy. Therefore, the improvement of the productivity of local breeds contributes to rural development. Even under the harsh environments, selection of favorable genome variants could sustainably improve productivity. Systematic genetic diversity studies of indigenous adapted breeds would be necessary to understand the acquired unique features of these breeds.

In Sudan, different indigenous goat populations are distributed across all agro-ecological zones from the arid region in the North to the fertile Savannah in the South. The adaptation to the harsh climate and the limited feed resources has led to natural selection of goats for minimal maintenance and low water requirements [2]. They developed a high water economy utilizing their body fluid more effectively which ensures the maintenance of an appropriate dry matter intake during periods of water scarcity and thereby, nonetheless, an adequate level of productivity [3].

Sudanese goats are classified into Nubian, Desert, Nilotic and Taggar goats [4]. The main classifications of these goats can be shortly summarized as follows: Nubian, a normal size dairy goat commonly found in the semi-arid areas in Sudan. Desert is a dual purpose goat kept by the nomadic tribes of Sudan, and is commonly found in the desert areas of Sudan. The Nilotic goat is mostly used for meat production, after Sudan was split into North and South, it is mostly found in the border area between North and South Sudan. Taggar goats are a species of dwarf goats commonly found in mountainous areas all over Sudan.

The aim of this study was to investigate the genetic diversity and population structure of the four above mentioned important indigenous Sudanese goat breeds living in an arid climate and suffering harsh feeding conditions. For the genetic analyses, we used the 50 K goat SNP chip (Illumina, San Diego, CA). The chip provides markers across the whole genome, allowing for precise identification of the whole genome diversity within and among the different populations. Furthermore, the whole genome data made it feasible to identify genomic regions that differ between the breeds and contain genes likely linked to the typical traits of the Sudanese goat breeds.

## Methods

### Animals and sampling

Blood samples were collected for genotyping from each of the four goat breeds, Nubian (NU), Desert (D), Taggar (T), and Nilotic (NI), 24 representative female animals were sampled from different regions of Sudan (Additional file 1: Table S1).

Nubian goats were collected at six locations in four different states along the river Nile, the Dongola area (Northern State), Abu Hamad, Aldamir, Shendi (River Nile State), Khartoum (Khartoum State) and Aljazirah (Aljazirah State). In total Nubian samples were obtained from 10 villages, three districts, one university farm, and three research stations.

Desert goat samples were collected from Bara and Abu Zabad area in the North Kordofan state (eight villages, one university farm).

Taggar goat samples were obtained from the Nuba Mountains and Dalang area in the South Kordofan state (five villages, one research station).

Nilotic goat samples were collected in the Kosti and Rabak areas in the White Nile state (eight villages).

Livestock experts, herdsman and owners of the animals were consulted to ensure representative sampling of unrelated animals.

The characterization of the different Sudanese goat breeds are the following:

### *Nubian goat*

Nubian is a highly productive dairy goat compared to other Sudanese goat breeds. Nubian goats are widely distributed in arid and extreme arid areas [4]. Besides Sudan, Nubian goats are found widespread in North Eastern Africa and the Mediterranean coastal belt. They have likely their origin in Sudan [5]. These goats are commonly black, but pure brown and multi-colorations of black and white also exist [6, 7].

### *Desert goat*

Sudanese Desert goats are dual purpose goats. They are mainly found in semi-arid areas of the West of Sudan. They also move to extreme arid areas during nomadic migration. The efficient transformation of low quality feed into body mass makes the Desert goat valuable for the production of meat [8]. Desert goats are phenotypically similar to West African long-legged goats, and are possibly related to the Nubian goats [4]. Coat color is variable and mixed colors exist [9, 10].

### *Taggar goat*

Taggar is a meat-type goat that has adapted to survive under harsh environmental conditions [11–13]. It is kept in many parts of Sudan with the highest density in the Nuba Mountain area close to the border to South Sudan.

Taggar is a dwarf goat with disproportionally short legs, plump body and short head. The short stature is thought to result from achondroplastic dwarfism with lack of ossification at the cartilage joints [13, 14]. It is assumed that natural selection for the recessive dwarfism gene was favorable in response to the humid and hot climate conditions [4]. The most common coat colors for Taggar goats are dark or grey brown [7, 12].

### **Nilotic goat**

Nilotic goat is another meat-type goat that produces high muscle mass under good feeding conditions [15]. Nilotic goats have a high reproductive potential since they reach sexual maturity at an early age. Nilotic goats live in the border region between Sudan and South Sudan. They distinguish from other breeds through their resistance against *Trypanosomiasis* [16]. These goats are small in size; the body is compact, but has normal proportions [6, 7, 17, 18]. Though, different from Taggar goats, achondroplasia does not occur in this breed. Almost all colors occur, but the predominant color is a mixture of black and white.

### **Genotyping**

Blood samples were collected from the jugular vein using vacutainer tubes containing EDTA as anticoagulant. Blood was stored at  $-20^{\circ}\text{C}$  until DNA was extracted using the Puregen core kit A (Qiagen Sciences, Maryland, USA). All animals were genotyped with the Goat SNP52 BeadChip (Illumina, San Diego, CA), developed by the International Goat Genome Consortium (IGGC) [19]. The raw signal intensities of the 53,347 SNPs on the chip were imaged using the IlluminaScan Reader and converted into genotype calls with GenomeStudio software suite (version 2011.1) by using the SNP genomic locations and cluster file made available by IGGC ([ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/goat\\_9925/viewBatch/snpBatch\\_IGGC\\_1057128.gz](ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/goat_9925/viewBatch/snpBatch_IGGC_1057128.gz)). Locations of the SNP probes were remapped using BLASTN towards the CHIR 1.0, CHIR 2.0 and LWT01 genome version of *Capra hircus*. This was done to provide future researchers with an easy way to compare our results with their own work, since other researchers can find the corresponding locations on the genome version of their choice. Locations reported in this paper are based on the CHIR1.0 genome build. Locations of protein coding genes on the CHIR1.0 genome were obtained from The National Center for Biotechnology Information (NCBI) [<http://www.ncbi.nlm.nih.gov/>]; the version used is available as supplemental file (Additional file 2: locations of Protein coding genes).

### **SNP quality control**

The R language for statistical computing v3.2.3 was used for quality control of the called SNP data [20]. From the entire set of 53,347 SNPs, only those with an Illumina GenTrain score  $\geq 0.6$  were kept (1441 SNPs failed). Furthermore, all SNPs with a minor allele frequency (MAF) lower than 5% (2273 SNPs) as well as SNPs which showed more than 5% missing data (150 SNPs) were excluded from further analysis. The deviation from Hardy-Weinberg equilibrium (HWE) was calculated and we found only 42 SNPs not in HWE. Since HWE deviations only concerns a very limited amount of SNPs  $<0.1\%$  of the total, we decided to not remove these SNPs from further analysis. SNP probe sequences were mapped against the CHIR1.0 genome assembly using the Basic Local Alignment Search Tool (BLAST) [21, 22]. Probe sequences that were not mapping in the CHIR1.0 genome (41 SNPs) or probes which showed multiple hits (937 SNPs) against the reference genome were dropped from further analysis. 48,505 SNPs passed the quality control. Based on quality control of the data, one sample (Nubian goat) was excluded from further analysis.

### **Statistical analyses**

#### **Genetic diversity assessment**

In order to assess the genetic diversity, the minor allele frequency, distribution and proportion of polymorphic SNPs were computed using the R language for statistical computing v3.2.3 [20]. To measure the genetic variation within a population, mean observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_E$ ) for each breed of goats and the total population were analyzed using the *diversity* R package v.1.9.89 [23]. The average population inbreeding coefficient ( $F_{IS}$ ) using Sewall Wright's method [24] and the pairwise genetic differentiation between populations ( $F_{st}$ ) were also calculated using the *diversity* package. To detect the level of genetic variation among samples within populations and among populations at different hierarchical levels, the Analysis of MOlecular VAriance (AMOVA) [25] was performed using *StAMPP* R package [26]. *StAMPP* calculates an AMOVA based on the Nei's genetic distance matrix using the *amova()* function from the package *PEGAS* for exploring within and between population variation. *StAMPP* uses the formula: distance = populations, to calculate a hierarchical AMOVA as described in Excoffier et al. [25] to explore population differentiation and within/between population variation.

#### **Hierarchical clustering**

To measure the genetic distance between the four goat populations, hierarchical clustering of SNP data was performed. Pairwise Nei's genetic distances [27] between all individuals were calculated from the SNP data by using

the StAMPP R package [26]. Additionally, we used Reynolds [28] and Manhattan [29] distances between individuals (Additional file 3: Figure S1). Distances were clustered using the hclust function in R. After clustering of genetic distances between individuals, the ape package [30] was used to convert the resulting dendrogram into a phylogenetic tree. For visualization of the phylogenetic tree a standard ape plotting functions was used.

### STRUCTURE analysis

Population structure was determined by using a model-based clustering for assigning individuals from multi locus genotypes to a population using the STRUCTURE 2.3.3 software suite [31–33]. STRUCTURE uses a maximum likelihood method to infer the genetic ancestry of each individual from a mixture of  $K$  pre-defined ancestral groups. STRUCTURE analysis was carried out using an admixture model and correlated allele frequencies. Under the hypothesis of two to five sub-populations  $K$  was set from 2 to 5 and the length of the burn-in period was set to 100,000 iterations, followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations. The whole analysis was then repeated five times in STRUCTURE to prevent the model from producing local optimal results. In our analysis, all five runs gave similar results. Results from STRUCTURE were loaded back into the R environment to visualize resulting groups. The most likely hierarchical level of genetic structure was determined by log probability of the data ( $\ln Pr(X|K)$ ) [33] and observed variance of this probability.

### Principal component analysis (PCA)

As a different approach to characterize divergence, the genetic relationship between animals was analyzed by principal component analysis (PCA) using the 'prcomp' function in the R language. PCA is a statistical method for exploring and interpretation of big data by reducing the dimension of the data to the few principal components which capture the majority of the genetic variation observed in the genotypes. We visualized the contribution of individuals to the first 10 principal components; from our data we observe that a combination of PC1 and PC2 shows a good separation between different goat breeds. As such we investigate which SNPs allow PCA to separate between the four populations of Sudanese goats; we investigated which SNPs contribute highly to PC1.

To select high impact SNPs we calculated the variable correlations with PC1 by multiplying the loading factors with the component standard deviations. Quality of representation for all variables on the factor map ( $\cos^2$ ) was calculated as the squared variable correlation. Afterwards, we summed up all the  $\cos^2$  values for PC1 and

expressed each SNP contribution as percentage of total variation. We then looked for SNPs which have more than 10 times the expected contribution to PC1 (contribution  $\geq 0.02\%$ , expected 0.0021%). These SNPs will capture most of the differentiation among breeds separated by PC1. We extracted genes within a region of 500 kb cis-window around these high contributing SNPs according to SNP Annotation and Proxy Search (SNAP) [34].

### Results

The main result from this study is that Taggar goats show significant genetic differences from the other three goat breeds. In all our analysis e.g.  $F_{ST}$ , STRUCTURE, clustering and PCA we observe that Taggar goats show significant genetic differences compared to the other breeds.

### Genetic diversity within breeds

The percentage of polymorphic SNPs within each breed among all 48,505 SNPs that passed the quality control was very high, ranging from 96.9% (Taggar goat) to 98.2% (Desert goat) (Table 1). After removing SNPs with MAF below 0.05, the average minor allele frequency, was about the same for Nubian ( $0.31 \pm 0.12$ ), Desert ( $0.31 \pm 0.12$ ), Nilotic ( $0.31 \pm 0.13$ ), and Taggar ( $0.30 \pm 0.13$ ) goats. The distribution of minor allele frequencies across the Sudanese goat populations are represented in Fig. 1a. Rare variant MAFs were observed in about 3.1%, 2.2%, 2% and 1.8% in Taggar, Nilotic, Nubian, and Desert goat breeds respectively.

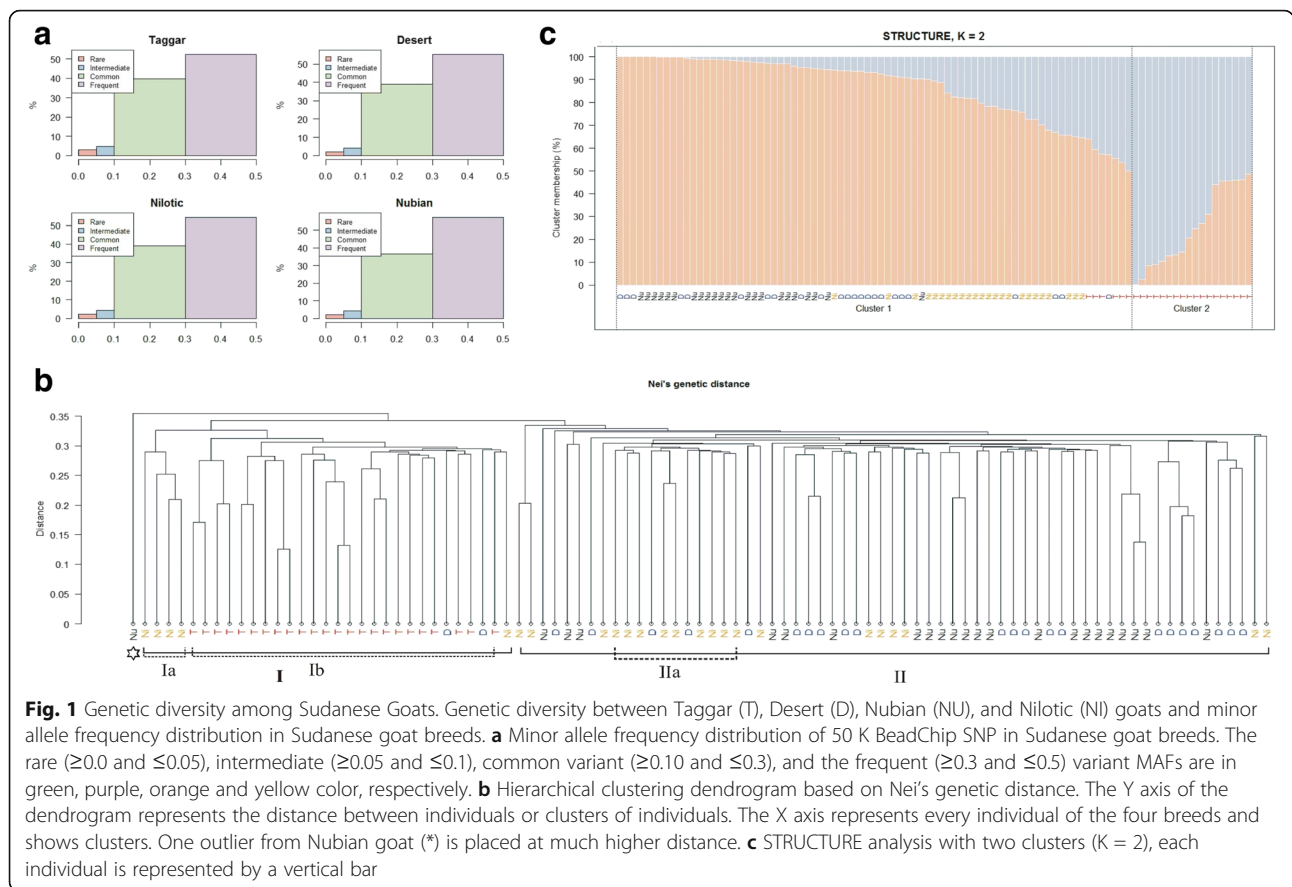
Additionally, observed and expected heterozygosity values were the same for the examined Nubian, Desert, and Taggar populations (0.39). For Nilotic goats, the expected heterozygosity (0.40) was slightly higher than observed (0.39). The high degree of heterozygosity and the coincidence of expected and observed heterozygosity indicate high genetic diversity and near to zero inbreeding within each breed. This was further confirmed by the inbreeding coefficients ( $F_{IS}$ ). The population inbreeding

**Table 1** Diversity indices comparing Sudanese goat breeds

| Breed name | Breed acronym | N  | $H_E$ | $H_O$ | $F_{IS}$ | PIC (%) |
|------------|---------------|----|-------|-------|----------|---------|
| Nubian     | NU            | 23 | 0.39  | 0.39  | 0.0010   | 98.0    |
| Desert     | D             | 24 | 0.39  | 0.39  | 0.0041   | 98.2    |
| Taggar     | T             | 24 | 0.39  | 0.39  | -0.0129  | 96.9    |
| Nilotic    | NI            | 24 | 0.40  | 0.39  | 0.0094   | 97.8    |
| Total      |               | 95 | 0.40  | 0.39  | N.A.     | N.A.    |

N: number of animals;  $H_E$ : expected heterozygosity based on the observed genotype frequencies;  $H_O$ : observed heterozygosity;  $F_{IS}$ : Wright's inbreeding coefficient, significance was determined using the diversity package, after 1000 bootstraps the 95% bias corrected confidence interval was used to determine if the estimated  $F_{IS}$  is significantly different from 0; PIC: polymorphic information content; N. A.: Not applicable





**Fig. 1** Genetic diversity among Sudanese Goats. Genetic diversity between Taggar (T), Desert (D), Nubian (NU), and Nilotic (NI) goats and minor allele frequency distribution in Sudanese goat breeds. **a** Minor allele frequency distribution of 50 K BeadChip SNP in Sudanese goat breeds. The rare ( $\geq 0.0$  and  $\leq 0.05$ ), intermediate ( $\geq 0.05$  and  $\leq 0.1$ ), common variant ( $\geq 0.10$  and  $\leq 0.3$ ), and the frequent ( $\geq 0.3$  and  $\leq 0.5$ ) variant MAFs are in green, purple, orange and yellow color, respectively. **b** Hierarchical clustering dendrogram based on Nei's genetic distance. The Y axis of the dendrogram represents the distance between individuals or clusters of individuals. The X axis represents every individual of the four breeds and shows clusters. One outlier from Nubian goat (\*) is placed at much higher distance. **c** STRUCTURE analysis with two clusters ( $K = 2$ ), each individual is represented by a vertical bar

coefficients ( $F_{IS}$ ) ranged from  $-0.0129$  in Taggar goats to  $0.0094$  in Nilotic goats and did not significantly differ from zero (Table 1). These values indicate that there is neither inbreeding nor strong kinship within the studied populations. To further confirm this observation, we calculated the kinship matrix using the 'kinship' function from the EMMA package [35], EMMA calculates kinship based on identity by state (IBS), meaning that two unrelated individuals measured using SNP markers on average show a kinship coefficient of 0.5, when looking at the values calculated we that kinship between most individuals ranges between 0.64 and 0.71. Between two pairs of individuals we find a slightly increased kinship of 0.81. EMMA analysis confirms that no strong kinship exists between individuals (Additional file 4: Figure S2).

**Genetic diversity between breeds**

The estimates for pairwise genetic differentiation between populations ( $F_{st}$ ) were low and varied between 0.0053 (Nubian vs. Desert) and 0.0229 (Nubian vs. Taggar) (Table 2).  $F_{st}$  values between Desert, Nilotic and Nubian goats were below 0.0076. The highest genetic distance was found between Taggar and the three other breeds ( $F_{st} \geq 0.0134$ ). Only the Taggar breed showed significant genetic differentiation towards the three other

breeds. However, pairwise  $F_{st}$  values for the other breeds did not significantly deviate from zero. As such  $F_{st}$  analysis only provides evidence for population differentiation of Taggar goats versus the other three goat breeds, and no such evidence for population differentiation between the Desert, Nilotic or Nubian. We investigated the SNP by SNP  $F_{st}$  values (Additional file 5: Figure S3.) from the supplemental figure it becomes clear that using  $F_{st}$  values we are unable to observe regions which explain inter-breed differences. This motivated us to perform additional PCA analysis to further look for other ways to find regions which show inter-breed differences. AMOVA confirmed the high within population variation

**Table 2** Estimated pairwise fixation indices ( $F_{st}$ )

| Population | Taggar        | Desert  | Nilotic | Nubian  |
|------------|---------------|---------|---------|---------|
| Desert     | <b>0.0179</b> | x       |         |         |
| Nilotic    | <b>0.0134</b> | 0.0076  | x       |         |
| Nubian     | <b>0.0229</b> | 0.0053  | 0.0072  | x       |
| Combined*  | 0.0032        | -0.0024 | -0.0031 | -0.0012 |

\*Combined: The breed mentioned in the column against the other three breeds combined, bold values indicate significance after 1000 bootstraps, with an alpha level of 0.005 (Bonferonni adjusted multiple testing  $p$ -value), *curvsive values* indicate suggestive significance after 1000 bootstraps, with an alpha level of 0.01

and low differentiation between populations. The largest proportion of the variation (93.04%) was attributable to the variation within individuals (Additional file 6: Table S2). Differences between populations accounted for only 6.96% of the variance, but was highly significant ( $P < 1.0 \times 10^{-6}$ ).

Hierarchical clustering using Nei's genetic distances clearly shows two different clusters (Cluster I and II) (Fig. 1b). Within Cluster I, the Taggar population forms a distinct group (Cluster I group b); however, we observed two Desert goat individuals positioned within the Taggar cluster. In addition, four Nilotic goat individuals clustered closely together and formed a distinct subgroup (Cluster I group a) which shows a closer relationship to Taggar goats (Cluster I group b). Cluster II comprises all Nubian goats, except one outlier (marked with a \* in Fig. 1b), and the remaining Desert and Nilotic goats. Within cluster II, Nilotic goats clustered in a smaller sub-group (Cluster II group a) with an approximately Nei's genetic distance of 0.31 between individuals belonging to this cluster. None of the Taggar goats was assigned to cluster II. Similar clusters were observed when SNP data was clustered using different distances measurements, such as Manhattan or Reynold distances (Additional file 3: Figure S1).

**STRUCTURE analysis**

The genetic structure among all goats was studied using a Bayesian model-based approach that assigns each individual to one or more populations based on the allele frequencies detected at different loci. Following STRUCTURE analysis the posterior probability ( $\ln P(D)$ ) of the

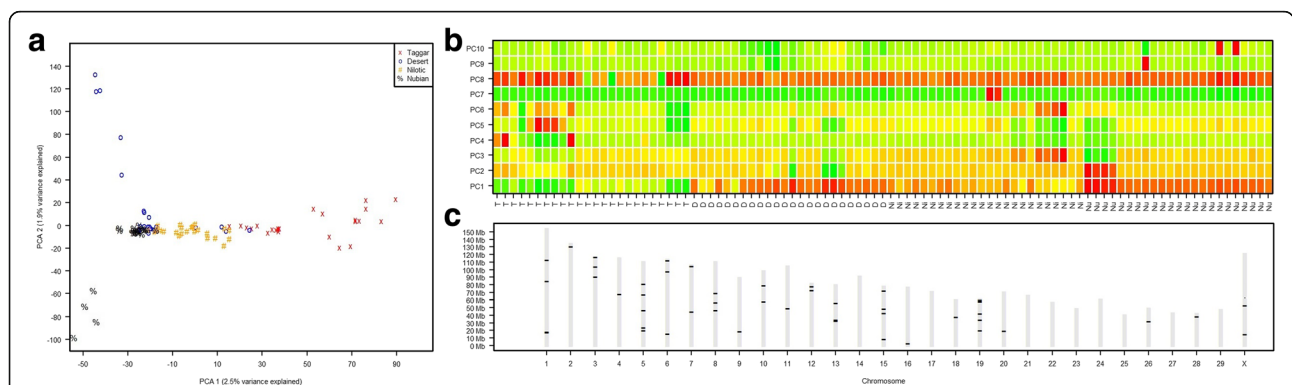
data indicated the most optimal K value to be 2 (Additional file 7: Table S3).

For  $K > 2$ , there is a higher variance of the log likelihood values, while the mean log likelihood of models  $K > 2$  is worse compared to the  $K = 2$  model. With  $K = 2$ , individuals were assigned into two main genetic clusters. When looking at a default clustering threshold of 50%, cluster 2 contained 18 out of 24 Taggar goats, while cluster 1 comprised of Nubian, Desert, Nilotic, and six Taggar goats (Fig. 1c). However, when we put the clustering threshold at 60% contribution compared to the default 50% used by STRUCTURE all Taggar goats, and a single Desert goat is in the second cluster.

Results from the STRUCTURE analysis with  $K = 2$  to 5 is summarized in Additional file 7: Table S3, the graphical STRUCTURE plots of the Sudanese goat breeds from  $K = 2$  to 5 are shown in Additional file 8: Figure S4.

**PCA analysis**

PCA was used to assess how different the four Sudanese goat populations are. Figure 2a shows the individual loadings of the first and second principal component (PC) against each other. We observed that the loading of PC1 resulted in tight clustering by breed. PC1 could separate three out of four different subpopulations, namely Taggar, Nilotic and a mix of Desert and Nubian goats, this is also seen in Fig. 2b (contribution of individuals to each PC) where the Taggar goats are colored green (high contribution), the Nilotic goats are mostly yellow (average contribution), and the other 2 breeds have a low contribution to the first principal component. Although this separation is not perfect, we assume this could be



**Fig. 2** Principal component analysis of all individuals. **a** PC1 (X axis) versus PC2 (Y axis) are shown. Individuals of different breeds are differently colored. PC1 clearly separates between Taggar, Nilotic and a mix between Desert and Nubian goats. The second component allows separating between Desert and Nubian goats. Desert goats are scattered throughout all clusters. The first two principal components, PC1 and PC2, account for 2.5% and 1.9% of the total variance, respectively. **b** Visual representation of the contribution of individuals to the first 10 principal components. The contribution of each individual to each PC1 through 10 is shown using a heat map image. Green color indicates a high contribution to the PC, while yellow shows average contribution, orange is low contribution, and the red color indicates the lowest contribution. **c** Genomic location of SNP contributing to the principle component 1 (PC1) – PC1 allows for the separation of the Taggar, Nubian and Nilotic goat breeds

caused by misclassification of individuals by the owners. We do not observe another principal component in the first 10 components investigated that cause this clear of a separation between breeds, which is expected since PCs are ordered by their variance explained. PC2 was able to separate some of the Nubian ( $PC2 \leq 0$ ) from the Desert goat population ( $PC2 > 0$ ) this is not clearly observable in Fig. 2b, since the classification here seems to depend on 5 Desert goats in green (high contribution) versus 4 Nubian goats in red (low contribution). The first two principal components explained 2.5% (PC1) and 1.9% (PC2) of the genetic variation observed between all individuals. PC1 and PC2 together seemed to be able to almost classify samples into their respective breed, we do observe overlap of some of the Desert goat samples with Taggar, Nilotic and Nubian goats. PCA using 48,505 SNPs allowed us to detect which SNPs contributed highly to PC1 and thereby to the genetic differentiation between Taggar, Nilotic and Nubian goats. To investigate the SNPs which have a high contribution to PC1, we generated a list of SNPs which had a 10 times higher contribution to PC1 compared to the expected contribution (Additional file 9: Sheet 1). We found that 49 SNPs out of 48,505 contributed highly to the PC1, which are visualized in Fig. 2c.

**Relationship between  $F_{st}$  and contribution to PC1**

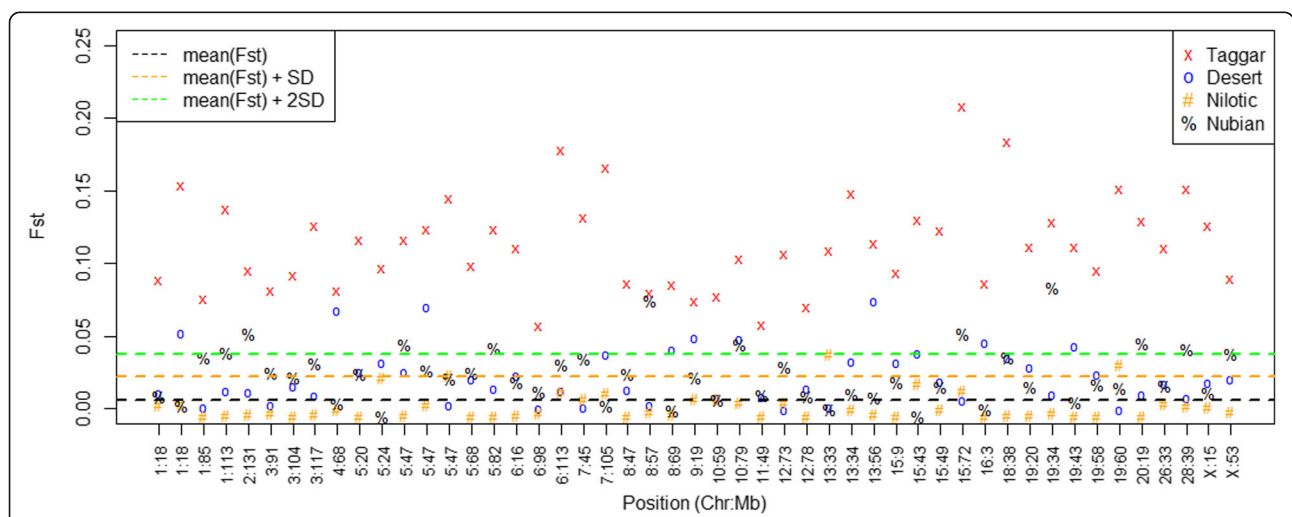
To further investigate our results that SNPs contributing highly to PC1 are biologically important in the distinction between Taggar goats versus the other Sudanese breeds. We explored this by investigating if these SNPs show higher than average differentiation. We used the  $F_{st}$  values of the SNPs which contributed highly to PC1. Since this would give us an independent line of evidence

that these SNPs are important in the differentiation of the Taggar goat versus the other goat breeds, and that our PCA approach allows us to identify these regions. We therefore plot the  $F_{st}$  values of the four different breeds against the combination of the three other breeds at the positions of the SNPs contributing to PC1 (as can be seen in Fig. 3). We observe significant higher  $F_{st}$  values in the Taggar versus the other goat breeds at these SNPs. All 49 SNPs selected by PCA analysis, show  $F_{st}$  values higher than at least  $mean(F_{st}) + 2 SDs$ . This means that these SNPs show signs of differentiation in Taggar goats versus the other breeds. This is not observed for any of the other three goat breeds. Desert goats versus the other breeds showed only 5 SNPs above threshold, Nilotic goats versus the other breeds showed no SNPs above the threshold, and Nubian versus the rest showed 8 SNPs above the threshold, this means that the regions contributing to PC1 are regions of high differentiation in Taggar goats, but are not regions of high differentiation in the other Sudanese breeds.

**Genes near SNPs contributing to PC1**

To investigate regions which show a high contribution to PC1 and have a high  $F_{st}$  values in Taggar versus the other goats, we selected genes from a region of 500 kb around those SNPs. We extracted 208 genes (Additional file 9: Sheet 2). Among these genes we found genes known to be involved in the following physiological traits:

- Bone formation (*sclerostin*, *epiphycan*, *bone morphogenetic protein receptor type-1A precursor*, *arylsulfatase G*, *mesenchyme homebox 1*, *ClpB homolog mitochondrial AAA ATPase chaperonin*)



**Fig. 3**  $F_{st}$  values of SNPs contributing to PC1. The  $F_{st}$  values of the four different breeds against the combination of the three other goat breeds at the positions of the SNPs contributing to PC1. The SNPs selected by PCA analysis, show  $F_{st}$  values higher than at least  $mean(F_{st}) + 2 SDs$

- Blood (water balance, glucose, and salts) homeostasis (*vasopressin*, *haptoglobin*, *hepatocyte growth factor activator*), sodium/potassium/calcium exchanger (solute carrier family 24 member 5)
- Heart and muscle (*phospholamban*, myosin phosphatase, mitochondrial cardiolipin hydrolase)
- Growth / dwarfism (*Stanniocalcin-1*, *inhibitor of growth protein 5*)
- Eye development (*keratocan*, *lumican*, *melanopsin*, *epiphygan*, *retinal guanylyl cyclase 2*, *beta-crystallin A3*, *all-trans-retinol 13,14-reductase*)
- Coat color (*sodium/potassium/calcium exchanger 5*)

Unfortunately, gene-ontology (GO) is not available for goats, so we are unable to perform GO testing on these groups of genes. Their location around the SNPs that contributed highly to PC1 and the high  $F_{st}$  values observed in Taggar goats makes these genes together with the other positional candidate genes interesting targets for the analysis of adaptation found between the different Sudanese goat breeds.

## Discussion

This study was conducted to contribute to the genetic characterization of the economically most interesting indigenous goat breeds in Sudan. Using the 50 k goat SNP chip, we examined the genetic diversity within and between breeds, the breed differentiation, and the genetics contributing to breed differences.

High levels of genetic diversity within the examined goat populations were observed. More than 96% of SNPs that passed the quality check were polymorphic in each breed albeit SNPs on the chip were selected from breeds such as Saanen, Alpine, Creole, Boer, Kacang, and Savanna and did not consider indigenous East African breeds [19]. Since the expected and observed heterozygosity in the examined Sudanese goat breeds were similar (0.39), we did not find heterozygote deficiency. Using the same SNP chip, heterozygosity in Sudanese goat breeds was similar to Bakri goats in Egypt (0.40) [36], Ethiopian goats breeds (0.38) [37] and Angora goats from South Africa (0.37) [38]. Consistent with low heterozygote deficiency and high allelic diversity, inbreeding within Sudanese goat breeds was also close to zero. The inbreeding coefficient estimated for Sudanese Nubian goats ( $F_{IS} = 0.001$ ) was consistent with previous findings for Nubian goats in Ethiopia ( $F_{IS} = 0.073$ ) [37]. This outcome indicates the highly diverse genetic reservoir of Sudanese goats.

The low  $F_{st}$  values among Sudanese breeds indicate low genetic differentiation among these populations, which in turn mirrors the population history with a likely common origin and the recent husbandry system in connection with nomadic traditions. AMOVA further

supported this finding by providing evidence that most variation was distributed within individuals and to a lesser extent genetic variation (6.96%) was explained by differences among the Sudanese goat populations.

By applying different genetic distance clustering methods, STRUCTURE and PCA, we observed clear separation of Taggar goats towards the other goat breeds. Taggar, which is a dwarf goat, and was identified as the most genetically distinct group in respect to other goat populations in this study. The output of the STRUCTURE analysis at  $K = 2$  further supported this finding. It clearly distinguished Taggar from the other Sudanese goat breeds. This finding could be explained by the fact that Taggar goats are geographically isolated in the mountain regions of Sudan. Mountain regions could have caused natural selection of small animals which are more nimble and feed efficient, which over the course of many generations could have led to the dwarfism phenotype we observe currently in Taggar goats. This natural selection might have caused genetic signatures in Taggar not observed in the other three breeds. Additionally, since Taggar goats are geographically separated (mountain versus low-lands) this might have played a limiting role in their ability to mate with the other three low-land goat breeds in Sudan, thus explaining why they show a clear separation from the other goat breeds.

It is observed that Desert goats are most scattered amongst the other two goats (Nilotic and Nubian). This could be due to Desert goat husbandry: (1) forced by changing market conditions, they began to shift from dual purpose to dairy goats by crossing their Desert goats with Nubian goats [39], and (2) Desert goats are owned by nomadic tribes who use communal grazing lands and watering points, where different herds meet and randomly mate which leads to gene flow.

Based on Nei's genetic distance we find a closer relationship between Taggar and a sub-group of Nilotic goats, which could be attributed to the low geographical distance between the two populations in the mountain area of the Southern part of Sudan.

Principal component analysis showed that the first two principal components could be used to differentiate between Sudanese goat breeds and to assign individuals to a particular breed. PC1 shows separation of Taggar goats, Nilotic goats, and a mixed of Desert and Nubian goats. PC2 seems to be able to differentiate between Nubian and Desert goats, though some misclassifications still remain. SNPs contributing highly to PC1 allowed us to define regions of the genome at which the Taggar goats are significantly different from the other three breeds. Among the genes in the vicinity of the SNPs contributing to PC1, we identified genes which might be interesting candidates when looking at breed characteristics and adaptation



differences between Sudanese breeds studied here. Among them we found genes known to contribute to bone formation, blood homeostasis, heart and muscle development, growth, eye development, and coat color.

Within the list of the genes there are genes such as *sclerostin* and *bone morphogenetic protein receptor type-1A precursor* known to have an effect on the bone morphology and formation. Studying differences in these genes could elucidate the underlying genetics of the differences between Sudanese goat breeds in regard to bone formation and body measurements characteristics. The *Stanniocalcin-1* gene on chromosome 8 is one of the genes, which is interesting for the discrimination among breeds, since Taggar goats are a species of achondroplastic dwarf goats with lack of ossification at the cartilage joints [13, 14]. *Stanniocalcin-1* acts as a paracrine regulator of growth plate chondrogenesis [40]. This gene is known to cause dwarfism in mice when it is over expressed [41].

Several other genes that are important for blood metabolism and homeostasis fell within regions in the PC1. This might reflect the adaptation of the Sudanese goat breeds to different environmental conditions. An example is the Desert goat which can survive in the areas where the lack of water resources is dominant. In a similar way, we detected candidate genes in regions which differ between the goat breeds for heart and muscle characteristics. Since natural variation is present between the breeds differences in these genes could provide targets for genomic breeding to improve meat quality.

The detection of genomic differences based on the principle component 1 (PC1) which allows for the separation of the Sudanese goat breeds can shed light on the genes underlying the adaptations. Natural variation present between these breeds provide a unique opportunity to improve local breeds through breeding using genomic markers or to breed improved resistance to harsh climates into imported high production breeds. Therefore, further research is required to identify the genomic regions which are associated with different important economical traits in Sudanese goat breeds.

## Conclusions

Based on our genome-wide analysis of SNPs in Sudanese goats, this study shows that Taggar goats show significant genetic differences from the other three breeds studied. We further conclude that the first principal components allow us to differentiate between Sudanese goat breeds. Furthermore  $F_{st}$  values of these SNPs show high differentiation for Taggar goats, but no significant differentiation for the other breeds under study. Genes in the proximity of these SNPs contributing highly to the first principal component might be interesting candidates when looking at breed characteristics and adaptation differences in Taggar goats.

## Additional files

**Additional file 1: Table S1.** Sample locations. (DOC 44 kb)

**Additional file 2:** Locations of protein coding genes. (TXT 3463 kb)

**Additional file 3: Figure S1.** Reynolds and Manhattan distances. (PNG 40 kb)

**Additional file 4: Figure S2.** Kinship. (PNG 71 kb)

**Additional file 5: Figure S3.**  $F_{st}$  (PNG 52 kb)

**Additional file 6: Table S2.** AMOVA. (DOCX 11 kb)

**Additional file 7: Table S3.** STRUCTURE. (DOC 33 kb)

**Additional file 8: Figure S4.** STRUCTURE analysis of Sudanese goat breeds. (PNG 102 kb)

**Additional file 9:** SNPs and genes contributed to PCA. (XLSX 36 kb)

**Additional file 10:** Raw genotype data obtained from the Goat SNP52 BeadChip. (XLSX 18918 kb)

## Abbreviations

AMOVA: Analysis of molecular variance; BLAST: Basic Local Alignment Search Tool; GO: Gene-ontology; He: Heterozygosity expected; Ho: Heterozygosity observed; HWE: Hardy-Weinberg equilibrium; IBS: Identity by state; IGGC: International Goat Genome Consortium; MAF: Minor Allele Frequency; PC: Principal Component; PCA: Principal Component analysis; SD: Standard deviation; SNAP: SNP Annotation and Proxy Search; SNP: Single Nucleotide Polymorphism

## Acknowledgments

Siham Rahmatalla acknowledges the financial support of the Alexander von Humboldt Foundation, Germany. The authors thank the goat owners in Sudan, the management staff of the Goat Research Stations Wad Medani, Kuku and Dongola, as well as the Sudanese farms of Bahri and Sudan University for providing goat samples.

## Funding

This study was supported by a Georg Foster Research Fellowship provided by the Alexander von Humboldt Foundation, Germany.

## Availability of data and materials

All relevant data are available as supplemental files to the manuscript, raw genotyping data are available in Additional file 10.

## Authors' contributions

Conceived and designed the experiments: SAR, GAB. Performed the experiments: SAR, MR, HR, WK. Analyzed the data: SAR, DA. Wrote the paper: SAR, DA, GAB. Critical revision of the manuscript: ASA, MR, HR. All authors read and approved the final manuscript.

## Ethics approval

All samples were collected with permission from the owners of the different animals.

## Consent for publication

Not applicable

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Albrecht Daniel Thae-Institut für Agrar- und Gartenbauwissenschaften, Humboldt-Universität zu Berlin, Invalidenstraße 42, D-10115 Berlin, Germany. <sup>2</sup>Department of Dairy Production, Faculty of Animal Production, University of Khartoum, P.O. Box 32, 13314 Khartoum North, Shambat, Sudan.

<sup>3</sup>Leibniz-Institut für Nutztierbiologie (FBN), Institut für Genombiologie, Wilhelm-Stahl-Allee 2, 18196, Dummerstorf, Germany.

Received: 16 May 2017 Accepted: 1 October 2017  
Published online: 23 October 2017

## References

- FAOSTAT: Food and Agriculture Organization of the United Nations: <http://www.fao.org/faostat/en/> (2014). Accessed 05 Sep 2015.
- Silanikove N. The physiological basis of adaptation in goats to harsh environments. *Small Rumin Res.* 2000;35:181–93. [https://doi.org/10.1016/S0921-4488\(99\)00096-6](https://doi.org/10.1016/S0921-4488(99)00096-6).
- Salem HB. Nutritional management to improve sheep and goat performances in semi arid regions. *Rev Bras Zootec.* 2010;39:337–47. <https://doi.org/10.1590/S1516-35982010001300037>.
- Wilson T. Small ruminant production and the small ruminant genetic resource in tropical Africa. *FAO Anim Prod Health Pap.* 1991;88
- Ballal KME, M-KA A, LMA M. Estimates of phenotypic and genetic parameters of growth traits in the Sudanese Nubian goat. *Res J Anim Vet Sci.* 2008;3:9–14.
- AOAD. Arab organization for agricultural development. Goats resources in Arab states. II-Sudan (Arabic). AOAD printing press. Sudan: Khartoum; 1990.
- El-Naim YA. Some reproductive and productive traits of Sudan nubian goats. In: MVSc dissertation. Sudan: University of Khartoum; 1979.
- Ismail AM, Yousif IA, Fadlelmoula AA. Phenotypic variations in birth and body weights of the Sudanese Desert goats. *Livest Res Rural Dev.* 2011;23
- Epstein H, Mason IL. The origin of the domestic animals of Africa. New York: Africana Publishing Corporation; 1971.
- Mason IL, Maule JP. The indigenous livestock of eastern and southern Africa. In: Technical communication N0, vol. 14. Edinburgh, UK: Commonwealth Bureau of Animal Breeding and Genetics; 1960.
- Bushara I, Abu Nikhaila MAAA. Productivity performance of Taggar female kids under grazing condition. *J Anim Prod Adv.* 2012;2:74–9.
- Muffarah MB. Goat breeds and varieties in Sudan. Proceeding of Training Course, Sheep and Goat Production, Arab Centre for Studies of arid and dryland (ACSAD), Khartoum 17-27 January, Sudan 1995.
- Valerie Porter LA, Stephen JG, Hall D. Phillip Sponenberg. Mason's world encyclopedia of livestock breeds and breeding. UK: CAB; 2016.
- Mason IL: Evolution of domesticated animals Longman London and New York; 1984.
- Zeinelabdeen WB, Atta M, Khidir OAE, Adam AA. Effect of two different diets on growth from birth to sexual maturity of Nilotic does. *Research opinions in animal & Veterinary Sciences.* 2011;1:562–6.
- Osman M, Nadia JK, Ghada HAE, Rahman AHA. Susceptibility of Sudanese Nubian goats, Nilotic dwarf goats and Garag ewes to experimental infection with a mechanically transmitted *Trypanosoma vivax* stock. *Pak J Biol Sci.* 2008;11:472–5.
- Devendra C, Burns M. Technical communication of the commonwealth Bureau of Animal Breeding and Genetics. In: Goat production in the tropics, vol. 19; 1983.
- Mason IL. The classification of west African livestock. Technical communication commonwealth Bureau of Animal Breeding and Genetics 1951;7.
- Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, Donnadieu-Tonon C, Eggen A, Heuven HC, Jamli S, et al. Design and characterization of a 52K SNP chip for goats. *PLoS One.* 2014; <https://doi.org/10.1371/journal.pone.0086227>.
- The R Core Team. R: a language and environment for statistical computing. In *R Foundation for Statistical Computing*, Vienna, Austria, vol. ISBN 3–900051–07-0; 2008.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
- Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. *diversity*: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol Evol.* 2013;4:782–8.
- Wright S, W. J. *Evolution and the genetics of populations, volume 2: the theory of gene frequencies.* Science. 1969;168:722–3.
- Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 1992;131:479–91.
- Pembleton L, Cogan N, Forste J. Statistical analysis of mixed ploidy populations. *Mol Ecol Resour.* 2013;13:946–52. <https://doi.org/10.1111/1755-0998.12129>.
- Nei M. Genetic distance between populations. *Am Nat.* 1972;106:283–92.
- Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics.* 1983;105:767–79.
- Craw S. Manhattan distance. In: *Encyclopedia of machine learning* (Sammut C, Webb GI eds.). Pp, vol. 639. New York: Springer US; 2010. p. 639.
- Paradis E, Claude J, Strimmer KAPE. Analyses of Phylogenetics and evolution in R language. *Bioinformatics.* 2004;20:289–90.
- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes.* 2007;7:574–8. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour.* 2009;9:1322–32. <https://doi.org/10.1111/j.1755-0998.2009.02591.x>.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics Society of America.* 2000;155:945–59.
- Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008;24:2938–9. <https://doi.org/10.1093/bioinformatics/btn564>.
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics.* 2008;178:1709–23. <https://doi.org/10.1534/genetics.107.080101>.
- Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, Rothschild MF. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity (Edinb).* 2016;116: 255–64. <https://doi.org/10.1038/hdy.2015.94>.
- Mekuriaw G, Mwacharo J, Tesfaye K, Tadelde D, Okeyo M, Djikeng A, Liu B, Osama S, Grossen C, Zhang W. High Density SNP Chips Array Uncovers Genetic Diversity and Population Structure of 16 Ethiopian and Chinese Goat Populations. *Plant and Animal Genome Conference XXIV.* In: January 9–13. San Diego: CA; 2016. p. 2016.
- Visser C, Lashmar SF, Van Marle-Köster E, Poli MA, Allain D. Genetic diversity and population structure in south African, French and Argentinian angora goats from genome-wide SNP data. *PLoS One.* 2016;11:e0154353. <https://doi.org/10.1371/journal.pone.0154353>.
- Mohamed Ali MA, Eldaw AS. Study of flock structure and some morphological, productive and reproductive characters of Sudanese Desert goats in north Kordofan state-Sudan. *Journal of Novel Applied Sciences.* 2015;4:1155–8.
- Shufang W, Yoshiko Y, Luca FD. Stanniocalcin 1 acts as a paracrine regulator of growth plate Chondrogenesis. *J Biol Chem.* 2006;281:5120–7.
- Jiang WQ, Chang AC, Satoh M, Furuichi Y, Tam PP, Reddel RR. The distribution of stanniocalcin 1 protein in fetal mouse tissues suggests a role in bone and muscle development. *J Endocrinol.* 2000;165:457–66.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

