

RESEARCH

Open Access



Genetic variation and demographic history of Sudan desert sheep reveal two diversified lineages

Bashir Salim^{1*}, Saeed Alasmari², Nouh Saad Mohamed³, Mohamed-Khair A. Ahmed⁴, Ryo Nakao⁵ and Olivier Hanotte^{6,7}

Abstract

More than 400 million sheep are raised on the African continent, the majority of which are indigenous and are primarily reared for sustenance. They have effectively adapted to various climatic and production environments, surviving and flourishing. The genetic relationships among these sheep populations remain understudied. Herein, we sequenced the entire mitochondrial DNA control region of 120 animals from Hamary and Kabashi and their crossbreed (Hamary x Kabashi) of Sudan desert sheep (SDS) to understand their maternal-inherited genetic variation and demographic history profiles and relate those to the history of sheep pastoralism on the African continent. The results show a diversified and predominant *D*-loop haplogroup B ($n = 102$, 85%), with all other sequences belonging to haplogroup A. Most of the maternal genetic variation was partitioned between haplogroup (76.3%) while within haplogroup accounted for 23.7% of the variation. However, little genetic differentiation was observed among the two breeds and their crosses, with our results supporting a Hamari maternal origin for the crossbreed. Bayesian coalescent-based analysis reveals distinct demographic history between the two haplogroups, two breeds and their crosses. Comparison of the two haplogroup showed that haplogroup B experienced an earlier expansion than haplogroup A. Unlike the breed-based comparison, the expansion of the two breeds started roughly at the same time, around 6500 years ago, with Kabashi having a slightly greater effective population size. The maternal ancestors of SDS may have diverged before their introduction to the African continent. This study provides novel insights into the early history of these two main breeds of Sudan desert sheep and their crosses.

Keywords mtDNA, Sheep, Haplogroup B and A, Hamary, Kabashi

*Correspondence:

Bashir Salim
bashirsalim@gmail.com

¹Department of Parasitology, Faculty of Veterinary Medicine, University of Khartoum, P.O Box 32, Khartoum-North, Sudan

²Department of Biology, Faculty of Arts and Sciences, Najran University, 1988 Najran, Kingdom of Saudi Arabia

³Molecular Biology Unit, Sirius Training and Research Center, Khartoum, Sudan

⁴Department of Genetics and Animal Breeding, Faculty of Animal Production, University of Khartoum, Khartoum, Sudan

⁵Laboratory of Parasitology, Faculty of Veterinary Medicine, Graduate School of Infectious Diseases, Hokkaido University, Sapporo, Japan

⁶Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, UK

⁷LiveGene – CTLGH, International Livestock Research Institute, Addis Ababa, Ethiopia



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Countless rural households and pastoralists across the African continent rely on indigenous sheep breeds for their livelihood. These breeds are thought to contain unique genetic variations that allow them to tolerate adverse environmental circumstances. They do well in ecologically marginal areas like the mountainous, desert, and semi-arid areas where other domestic animals might not economically survive [1]. Both genetic data from modern and archaeological specimens highlight our understanding of animal domestication. It is well documented that the main ancestor of domestic sheep (*Ovis aries*) belongs to a species found in the Fertile Crescent, the Asiatic mouflon *Ovis orientalis* [2]. Historical genetic profiles of sheep have been investigated by analysing maternally inherited mitochondrial DNA (mtDNA) in modern sheep breeds. Hitherto, there are at least five genetically different lineages. Sheep belonging to haplogroups A and B are present in many parts of the world, and haplogroups C, D, and E, have a much more restricted geographical range [3, 4]. These different lineages might represent spatially and temporally discrete “domestication events” in which diverse populations of animals were brought under domestication independently of one another [5]. According to a study by Tarekegn and his colleagues [6], sheep and goats first entered Egypt through the Sinai Peninsula, the Mediterranean and the Red Sea coast before spreading through the Nile Basin southward into Sudan and Ethiopia. Until now, no mutation rate of the sheep control region has been documented. However, complete mtDNA dating suggests 30 million years ago for the divergence between the bovine and ovine lineages [7]. Also, a recent study [8] showed that the male-specific region of the Y chromosome has 0.93×10^{-10} mutations per generation per site, which is roughly fifty times the one reported for the full mtDNA.

Sudan desert sheep belong to the thin-tailed hair sheep group and subgroup of African long-legged sheep. They are found strictly within the semi-arid climatic zone of Sudan, North of the 10 degree north latitude, extending eastwards into Eritrea and westwards into Chad (DAGRIS). The Sudan Desert sheep probably descended from ancient Egyptian stock [9]. These sheep originated in western Asia and entered Africa through the Isthmus of Suez. Until the third Millennium BC, the only type of sheep on the African continent was the hairy thin-tailed sheep. Domestic sheep had reached Egypt and other parts of North Africa by 5000 BC. The today observed coat colours of tribal breeds might have been already present in the ancestral population with selection toward colours preferred by particular groups or tribes leading to the near fixation of coat colour in some populations. For instance, Hamari breeds in south-western Kordofan and

south-eastern Darfur are predominately brown and dark brown, whereas the Kababesh sheep (Kabashi) of northern Kordofan and northern Darfur are multi-coloured [10]. According to the Veterinary Legislation Identification Mission Report, the sheep population of Sudan was estimated to be approximately 39.2 million in 2016 (REF). In 2017 40.752.000 heads of sheep have been reported [11] (www.ceicdata.com/en/sudan). Whereas a report published in 2018 indicated that Sudan in 2009 had about 51.5 million sheep with a total meat production of 313,000 tons [11].

This study investigated the maternal genetic variations and demographic histories of three indigenous and important Sudan desert sheep breeds by analysing the mitochondrial DNA (mtDNA) control region. To better understand sheep pastoralism in North-East Africa, its origins and evolution, we particularly sought to assess the maternal genetic diversity, and its variation within and among the Hamary, Kabashi and their crossbreed (Hamary x Kabashi) breeds.

Materials and methods

Sampling and DNA extraction

A total of 120 blood samples from Sudan desert sheep breeds (Hamary, N=72; Kabashi, N=25; and crossbred, N=23) from North Kordofan State was collected. These animals were owned by nomads who had no records of their pedigree or book registration. The owner, however, knew off-hand the breed origin. To avoid sampling sibling or related animals, we sampled different herds. Informed consent has been obtained from all the owners, and all efforts were made to avoid sampling closely related individuals. The sampling protocol was approved by the Faculty of Veterinary Medicine, University of Khartoum, according to their guidelines for sampling domestic animals in Sudan and in accordance with ARRIVE guidelines (<https://arriveguidelines.org>). Genomic DNA was extracted using DNeasy® Blood and Tissue Kit (Qiagen, Germany), following the manufacturer’s instructions.

PCR amplification and sequencing

Complete mtDNA D-loop region (1180 bp) was amplified using forward primer CsumF was 5’GGCTGGGACCAAACCTAT –3’, and the reverse primer CsumR was 5’-GAACAACCAACCTCCCTAAG –3’ as described by [12]. PCR reactions were performed in a 25 µl-reaction mixture containing 12.5 µl of 2 × Gflex PCR Buffer (Mg²⁺, dNTP plus) (TaKaRa.

Bio Inc., Shiga, Japan), 0.5 µl of Tks Gflex DNA polymerase (1.25 units/µl) (TaKaRa Bio Inc.), 200 nM of each primer, and 1.0 µl of template DNA. The thermal reaction conditions consisted of an initial denaturation step at 95 °C (3 min), followed by 35 cycles of 95 °C for 1 min, 56 °C for 30 s, and 68 °C for 90 s, and a final extension

step at 68 °C for 5 min. PCR products were purified by using a NucleoSpin Gel and PCR Clean-Up Kit (Takara Bio Inc.) and sequenced directly by the two PCR primers using the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). The sequencing was analyzed on an ABI Prism 3130 x genetic analyzer (Applied Biosystems) according to the manufacturer's instructions.

Sequence data analysis

Prior to analysis, all the chromatograms were visually inspected, and sequence fragments were manually edited using ATGC software version 9.1 (GENETYX Corporation, Tokyo, Japan), to correct base-calling errors. Multiple sequences alignments were performed using MUSCLE algorithms implemented in MEGA 7 [13], reference sequence to each haplogroup was utilized ([7] haplogroup A and B; AF039578 and AF039577 [14] for haplogroups C, D and E; HM236178, HM236180 and HM236182). These were subsequently joined to reconstruct a 1180 bp fragment spanning the entire ovine mtDNA D-loop. The haplotypes were determined with DnaSP v5 [15]. The data processing was performed based on haplogroups and breeds. The level of genetic diversity was determined by the number of haplotypes, haplotype diversity, nucleotide diversity, and mean numbers of nucleotide differences between haplotypes. This was computed for haplogroup and breed datasets using Arlequin 3.5 [16]. To gain insight into the genetic relationships between the haplotypes and determine the number of distinct mtDNA D-loop haplogroups present in the dataset, a median-joining (MJ) haplotype network [17] was created using PopArt software 1.7 (<https://popart.maths.otago.ac.nz>). All the mutations and character states were weighted equally.

The Analysis of Molecular Variance (AMOVA) was performed in Arlequin v3.5 with 1,000 permutations to partition the genetic variation among populations and sub-populations. Phi (ϕ) statistics representing haplotype correlations at various hierarchical levels (ϕ_{CT} , ϕ_{SC} , ϕ_{ST}) were calculated. The significance levels of the variance components associated with the different hierarchical clusters were evaluated with 1000 non-parametric coalescent simulations in Arlequin v3.5 [16]. The sequences obtained and analysed in the study were submitted to the DNA Data Bank of Japan (<http://www.ddbj.nig.ac.jp>) under accession numbers LC456425 – LC456544.

Mismatch distribution, tests of neutrality, and bayesian inferences

Each population's historical dynamics and demographic profiles were inferred from mismatch distribution patterns [17]. The chi-square test of goodness of fit and

Harpending's raggedness index " r " [18] statistics were used to evaluate the significance of the deviations of the observed sum of squares differences (SSD) from the simulated model of expansion (demographic or spatial) following 1,000 coalescent simulations. Fu's F_s [19] and Tajima's D [20] statistics were also calculated using the infinite sites model in Arlequin v3.5 to supplement the mismatch distributions. To further explore the evolutionary relationships between breeds, the unrooted neighbour-joining (NJ) phylogenetic was reconstructed using MEGA 7.

The demographic dynamics and history of the two breeds and their crosses were further investigated by generating Bayesian Skyline Plots (BSP) [21] using the piecewise constant function implemented in BEAST 2.0 [22] following [23]. In brief, the HKY+G+1 nucleotide substitution model was used for the analysis, and each Markov Chain Monte Carlo simulation (MCMC) run was performed for 2000 million generations that were sampled every 2,000 generations. The initial two million generations served as burn-in. Convergence of the posterior estimates of the N_e to the likelihood stationary distribution was evaluated with TRACER v1.6 (<http://tree.bio.ed.ac.uk/software/treestat/>). Since there is no available mutation rate for sheep D-loop, we calibrated the BSPs using the molecular rate of evolution (μ) of cattle mtDNA D-loop of 6.94×10^{-7} substitutions/site/year [s/s/y; 95% highest posterior density interval (HPD) 4.52×10^{-7} – 9.35×10^{-7} s/s/y] [24]. The final BSP plot was generated using outputs from TRACER v1.5 and displayed using MICROSOFT EXCEL (Microsoft Corporation).

Results

Sequence variability and diversity analysis of the two lineages and the breeds

One hundred and twenty sequences, spanning the 1180 bp of the ovine mtDNA D-loop, were generated (Hamary, Kabashi, and crossbreed (Hamary x Kabashi). Following their alignment against the reference sheep sequence of haplogroup lineages, two haplogroups, A and B, were identified. The complete mtDNA control region sequences were obtained, spanning the *Ovis aries* reference for 120 sequences. These sequences show 175 polymorphic sites, a transversion to transitions rate of 133:4 and two indels for haplogroup B, and a transversion to transitions rate of 35:2 and two indels for haplogroup A. The total haplotype and nucleotide diversities were 0.993 and 0.08, respectively. The analysis of mtDNA lineages A and B revealed a high level of nucleotide diversity differences between haplogroups ($K=44.748$) and a low level of nucleotide substitution per site between haplogroups (0.03792). The predominant haplogroup B included 102 individuals and 79 haplotypes, whereas haplogroup A consisted of 18 individuals and 17 haplotypes (Table 1).

Table 1 Complete D-loop region of the mtDNA diversity between two lineages and the three of Sudan desert sheep

Population	N	S	H	Hd (SD)	π (SD)	K	D	F _s	SSD (P-value)	'r' (P-value)	K ^{b*}	D ^c
Haplogroup A	18	36	17	0.994 (0.021)	0.007 (0.004)	11.266 (5.155)	-0.829 (0.221)	-8.556 (0.001)	0.019 (0.060)	0.031 (0.300)	44.748	0.03792
Haplogroup B	102	139	79	0.992 (0.003)	0.009 (0.005)	8.569 (4.155)	-1.917 (0.005)	-24.334 (0.000)	0.0007 (0.570)	0.004 (.680)		
Total	120	175	96	0.993	0.008							

N, sample size; S, number of polymorphic sites; H, number of haplotypes; Hd, haplotype diversity; π , nucleotide diversity; K, mean number of nucleotide differences; SD, standard deviation. D, Tajima's D; F_s, Fu's F_s; SSD, Sum of square deviation, 'r', Harpending's raggedness index, K^{b*}; Average number of nucleotide differences between Haplogroups, D^c; Average number of nucleotide substitution per site between Haplogroups

Table 2 Complete D-loop region of the mtDNA diversity between the two of Sudan desert sheep and crossbreed

Breed	N	Hap-A	Hap-B	S	H	Hd (SD)	π (SD)	K (SD)	D	F _s	SSD (P-value)	'r' (P-value)
Hamary	72	12	60	151	64	0.996 (0.003)	0.018 (0.009)	21.415 (9.555)	-1.088 (0.127)	-24.064 (0.000)	0.012 (0.100)	0.004 (0.770)
Kabashi	25	5	20	100	24	0.997 (0.009)	0.019 (0.009)	22.160 (10.103)	-0.73 (0.241)	-7.561 (0.007)	0.018 (0.310)	0.015 (0.270)
Crossbreed	23	1	22	78	17	0.968 (0.006)	0.011 (0.006)	13.130 (6.135)	-1.569 (0.041)	-3.298 (0.099)	0.008 (0.390)	0.020 (0.440)
Total	120	18	102									

N, sample size; S, number of polymorphic sites; H, number of haplotypes; Hd, haplotype diversity; π , nucleotide diversity; K, average number of nucleotide differences; SD, standard deviation. D, Tajima's D; F_s, Fu's F_s; SSD, Sum of square deviation, 'r', Harpending's raggedness index

The number of haplotypes detected in each Sudan desert sheep population was 64 (88.88%), 24 (96%), and 17 (74%) for Hamary, Kabashi, and the Crossbreed, respectively. High haplotype and low nucleotide diversity were observed in the three breeds (Table 2), supporting high levels of maternal genetic diversity for the three Sudan desert sheep populations examined.

Population phylogenetic analysis and partitioning of genetic variation

We constructed a median-joining haplotype network to understand the phylogenetic relationships of Sudan desert sheep based on the complete mtDNA D-loop sequences of 120 individuals. Using the reference sequences for the five sheep haplogroups (A, B, C, D and E) (Fig. 1), the sequences were clustered into two main haplogroups, A and B with a total of 96 distinct haplotypes.

Haplogroup B was the predominant haplogroup. A total of 96 haplotypes were identified, of which 69 haplotypes were singletons and eight haplotypes were shared among the breeds within haplogroup B, whereas in haplogroup A, 16 haplotypes were singletons, and one shared haplotype was within Kabashi breed (Table 1). The commonest haplotype included five individuals (3 Crossbreed and 2 Hamary). The next most common haplotype was composed of four individuals each (Hamary, Kabashi, and Crossbreed) (Fig. 1). The number of shared mutations between haplogroup is 21, and the number of net nucleotide substitutions per site between haplogroup was $D_a=0.02948$. As is shown in (Figure S1),

the NJ phylogenetic tree revealed that 79 haplotypes of SDS sequences clustered into haplogroup B, and the remaining 17 into haplogroup A. The haplotype network analysis showed a star-like structure for haplogroup B, suggesting population expansion.

We also examined the genetic distance between the two haplogroups and among breeds, measured in nucleotide substitutions per site, by dividing the three breeds of Sudan desert sheep populations into two groups using the neighbour-joining phylogenetic tree constructed from the 120 sequences of the mtDNA control region (Figure S1). The AMOVA analysis at the breed level resulted in little genetic differentiation among the three breeds (Table S1). However, AMOVA revealed a clear genetic distinction between the two haplogroups with 76.3% of the variation between haplogroup and 23.7% within haplogroup (Table 3). These results support a high maternal genetic differentiation between the haplogroups for Sudan desert sheep. The comparison also revealed 15 polymorphic sites in haplogroup A, monomorphic in haplogroup B and 116 polymorphic sites in haplogroup B, but monomorphic in haplogroup A.

In an exponentially growing population, the distribution of pairwise differences can provide useful information if the distribution is a Poisson distribution [25]. The gene haplotype network in this scenario resembles a star with all the nodes clustered in time, implying that all coalescent events will take place close to the root and few, if any, will take place later.

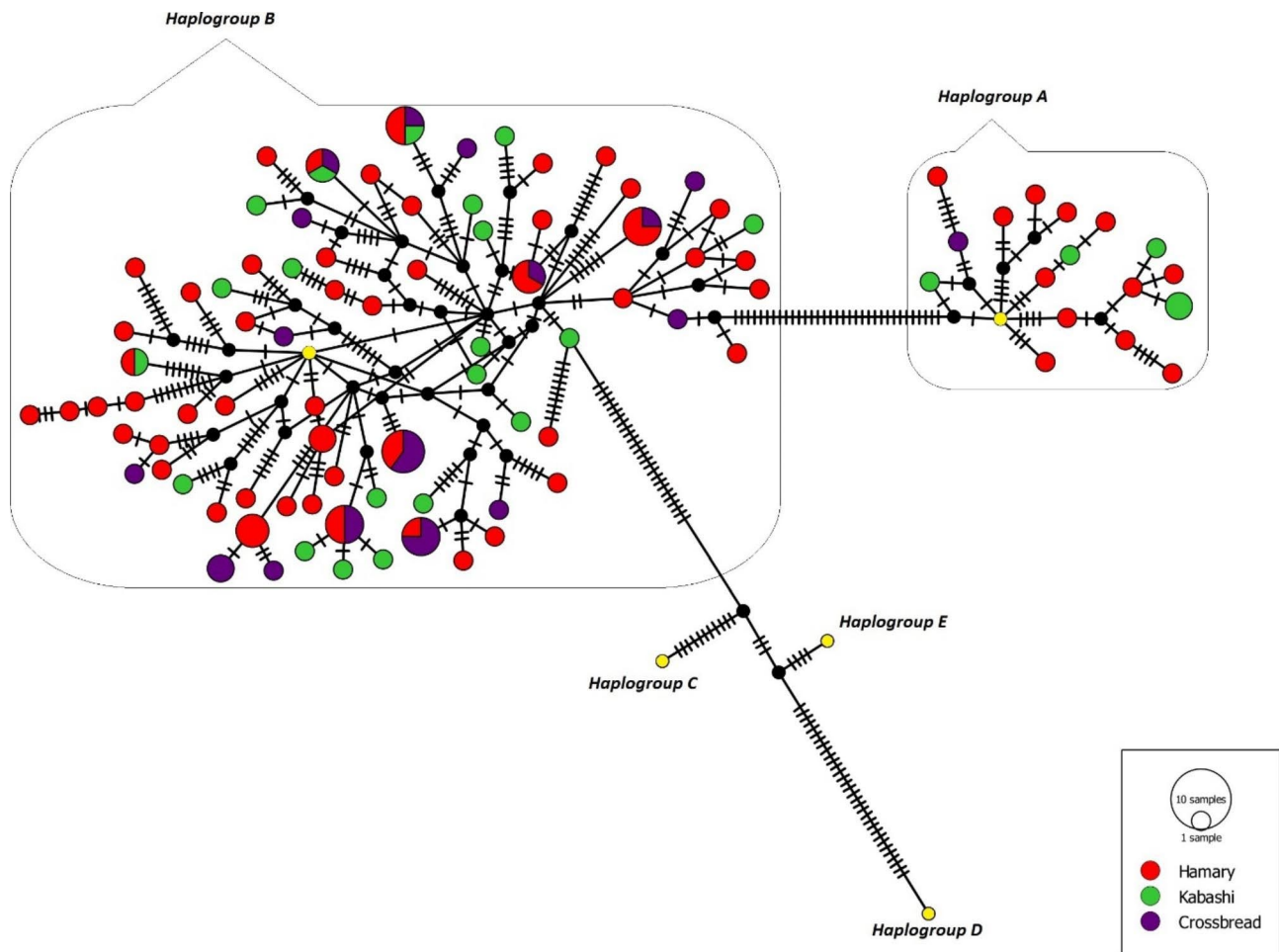


Fig. 1 Median joining network showing the relationships among 96 Sudan desert sheep haplotypes. 17 belongs to Haplogroup A, 79 belongs to Haplogroup B. Reference sequences are represented in yellow colour, red, green and purple colours denoted for Hamari, Kabashi and Crossbreed. None of the them belongs to Haplogroups either C, D E

Table 3 Analysis of molecular variance within and between Haplogroup A and B

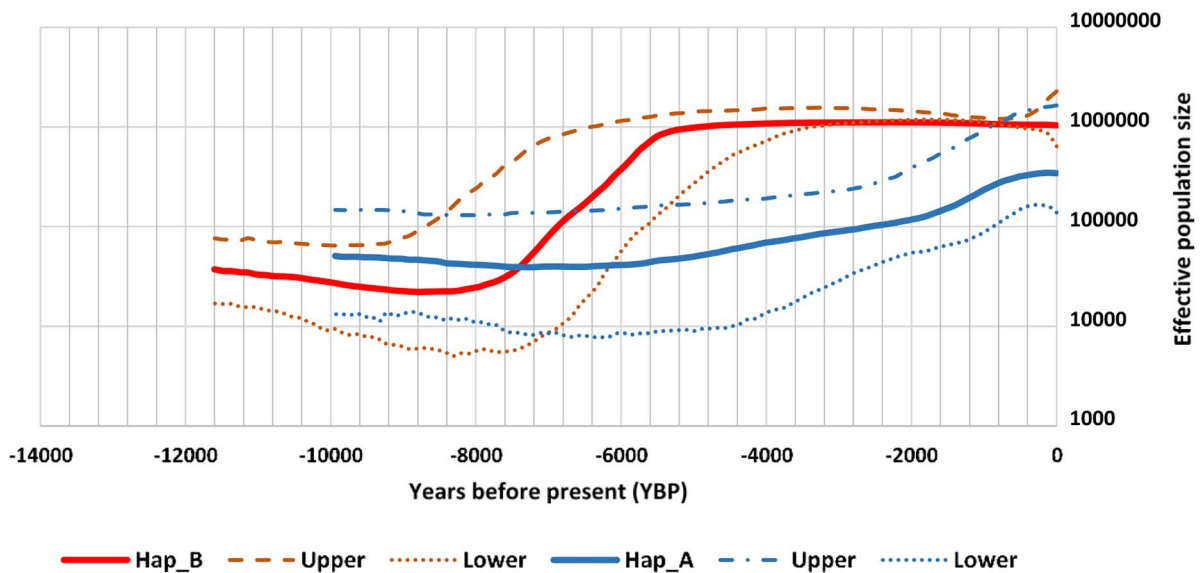
Source of variation	Sum of squares	Variance components	Percentage of variation	P-value
Between haplogroups	546.088	17.66663	76.3	0.001
Within haplogroup	647.745	5.48937	23.7	0.001
Total	1193.833	23.15599		

Historical and demographic profile of Sudan desert sheep

The mismatch analysis for all haplotypes gave negative values of Tajima's D and Fu's F_s with significant values for all Fu's F_s results. The histograms of mismatch distribution revealed two distinct peaks (bimodal) for all except the haplogroup B (Table 1 and Figure S2). These findings support the recent expansion of Sudan desert sheep breeds. We obtained a better resolution of the demographic history and profile of the study populations by modelling changes in effective population size (N_e) through time with the generation of Bayesian Skyline Plots (BSP) for each breed (Hamary, Kabashi and Crossbreed) and the two haplogroups (A and B). As indicated in the materials and methods, we calibrated the

BSPs using the cattle mtDNA control region's molecular rate of evolution (μ). The profiles of the skyline plot for the haplogroups showed that haplogroup B had the highest effective population size. It started to coalesce earlier, at around 11,000 YBP, and started its expansion earlier, at around 8000 YBP, compared to haplogroup A, which started to coalesce at about 10,000 YBP and started to expand at about 6500 YBP (Fig. 2A). Moreover, haplogroup B reached a plateau at around 2000 YBP compared to 150 YBP for haplogroup A. The combined dataset of SDS revealed coalescence, the start of the expansion, and reaching a plateau occurred at around 3700, 700, and 4000 YBP, respectively (Fig. 2B).

(a)



(b)

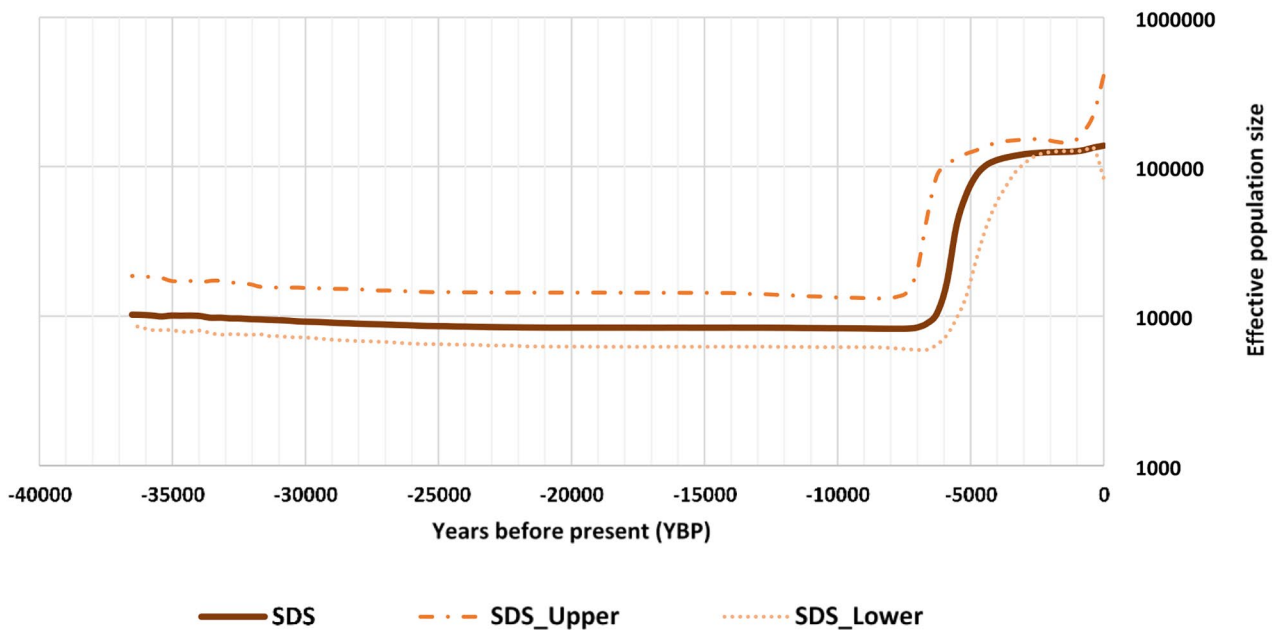


Fig. 2 Coalescent Bayesian skyline plots for the a, haplogroup A & B; b All dataset (Sudan desert sheep). Solid lines show median estimate of effective population size. Dotted lines indicate 95% highest posterior density interval (HPD) curves

On the other hand, the three sheep breed Hamary, Kabashi, and Crossbreed started to expand at around 8000 YBP. Crossbreed was the earliest to coalesce around 6500 YBP, then Hamary at 5900 YBP and Kabashi at 4900 YBP. The N_e of all populations remains constant to the present time, except in Crossbreed population

which shows a gradual declining trend from ~100 YBP. The highest effective population size was observed in Kabashi, Hamari, and the lowest one in Crossbreed (Fig. 3 and Figure S3).

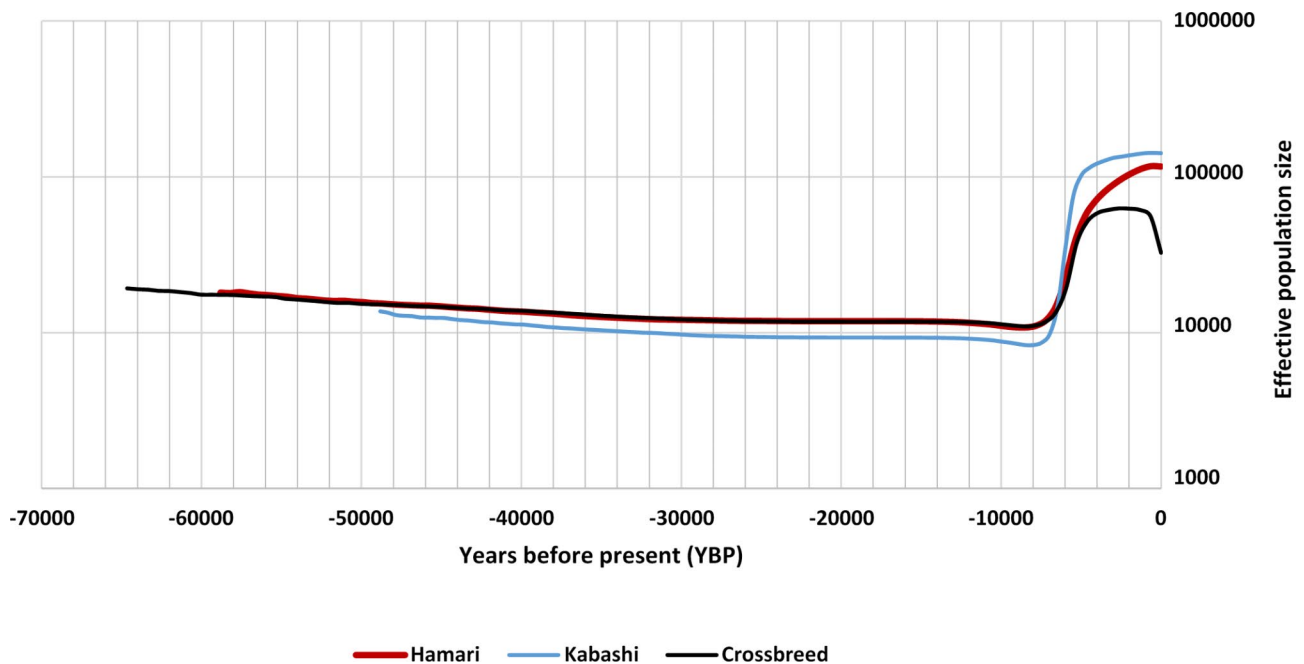


Fig. 3 Coalescent Bayesian skyline plots for the Hamari, Kabashi and Crossbreed of Sudan desert sheep. Solid lines show median estimate of effective population size. Dotted lines indicate 95% highest posterior density interval (HPD) curves

Discussion

An analysis of the complete mitochondrial control region sequences of 120 sheep belonging to three Sudan desert sheep (SDG) breeds (Hamary, Kabashi, and Crossbreed) was presented in this study. All SDG are classified as thin-tailed sheep and have been reported to likely share an ancestry with both European and Asian sheep [26].

Our results provide interesting insights about the genetic origin of the crossbred Sudan desert sheep breed, with mtDNA D-loop data supporting predominantly female Hamary origins for the Crossbreed. Indeed, only shared mtDNA D-loop haplotypes were observed between Hamary and Crossbreed, with none observed between the Crossbreed and Kabashi. Thus, though the crossbreeding between Hamary and Kabashi may appear random, it appears to follow a crossbreeding pattern selected by the shepherds.

It is widely acknowledged that domestic sheep have five maternal mitochondrial DNA (mtDNA) lineages (i.e., A, B, C, D, and E), some with distinct geographic distributions. This study revealed widespread occurrences of haplogroup B and, to a smaller extent, of haplogroup A in Sudan desert sheep. Similar results were obtained in a previous study that screening 231 Sudan sheep using restriction fragment length polymorphism, where the majority of the sequences belonged to haplogroup B, with only around 10% to haplogroup A [27]. Additionally, a mtDNA control region analysis of 91 domestic sheep from Kenya identified 90 haplogroup B and only one haplogroup A haplotype [28]. A study of 31 Ethiopian

domestic sheep identified five (16.12%) haplogroup A and 26 (83.88%) haplogroup B sequences [29]. Interestingly, in Algeria, 87% of Algerian sheep had sequences within haplogroup B, with the remaining belonging to haplogroup C rather than A [30].

The signature of a population expansion in Sudan desert sheep was revealed through a mismatch distributions analysis under spatial expansion assumptions. A negative and significant F_s value indicated an abundance of rare haplotypes, which is consistent with a recent population expansion or background selection [19]. This finding was further supported by an association between one common haplotype and others with lower frequencies or private haplotypes [17, 25].

Out of all 96 observed haplotypes, 87.5% were unique, indicating significant maternal diversity in the studied populations. Furthermore, most haplotypes were one mutation step away from each other, suggesting recent expansions. The star-like median-joining network, which had several median vectors, indicates the presence of unsampled genotypes or extinct ancestral sequences. This, in association with extensive single haplotypes presence, support little maternal genetic structure within the SDS breeds.

Recent analysis of the control region of mitochondrial DNA (D loop) in 11 indigenous Indian sheep breeds revealed the presence of maternal haplogroups A, B, and C as well as evidence of population expansion [31]. In contrast, in the Mediterranean region and eastern Europe, haplogroups A, B, and C were reported in three

sheep breeds from Egypt and two from Italy [32], as well as in two breeds from Hungary [33], with the absence of haplogroups D and E. However, haplogroup D was found in 2.2% of seven Italian sheep breeds, according to [34].

As expected, the NJ phylogenetic tree formed two separate clades representing haplogroup B as the more frequent than haplogroup A. Two major lineages, A and B, and three minor lineages, C, D, and E, have been identified in sheep breeds worldwide [35]. We rationalized the absence of the minor lineages (C, D, and E) by the fact that lineage C is thought to have a limited distribution in semi-desert and steppe regions between 30° and 45° north latitudes. However, lineage E is present in Algeria which opens the door for further discussion. Additionally, lineage C co-occurs with native fat-tailed breeds, suggesting that the geographic distribution of fat-tailed breeds may be related to the predominance of this lineage [3, 36]. However, lineages D and E in domestic sheep are exceptionally rare and were only reported in the North Caucasus region [4].

Latest evidence on the diversity of the mitochondrial DNA control region, the phylogenetic relationships among African sheep breeds, and their demographic histories reveal that thin-tailed sheep primarily dominate haplogroup B, which has been further subdivided into B1, B2, and B3, with the Sudan haplogroup belonging to B1. According to [37], the sub-haplogroup B1, primarily from West Africa and Sudan, appears to have had higher dispersal characteristics than other sub-haplogroups. The same study suggests that Sudan may have played a significant role in the dispersion of B1, both southward and westward.

According to [26], the thin-tailed sheep were the first sheep to be introduced into Africa, followed by the fat-tailed sheep through the north-eastern part of the continent and the Horn of Africa. The thin-tailed sheep from the Sudan desert displayed various historical demographic characteristics of interest. Our findings indicate that haplogroup B coalesced before haplogroup A, supporting higher diversity and larger coalescent effective population sizes of haplogroup B compared to haplogroup A. Remarkably, the expansion of Hamary, Kabashi, and their Crossbreed all occurred around the same period. These can be accounted for using the mutation rate of the cattle mtDNA control region, as the mutation rate for sheep is currently unavailable.

Conclusion

This study has revealed the widespread presence of haplogroup B, low mtDNA differentiation of the three Sudan desert sheep, and high maternal diversity among breeds. The results also demonstrate that three breeds, Hamary, Kabashi, and Crossbreed, and the two major haplogroups, A and B, have undergone population expansions in the

past, suggesting differences in their demographic histories. The knowledge gained in this study may help improve sheep genetic resource conservation and utilisation. Indeed, they suggest that the Sudan desert sheep may represent a unique genetic resource with two main maternal influences: an ancient one (haplogroup B) and a more recent one (haplogroup A). However, further research is needed to investigate the diversity and linkages between contemporary populations of African sheep and their ancient counterparts to further support the history of Sudan desert sheep proposed here. It is also recommended to identify the genetic and phenotypic characteristics of other local sheep populations from various geographical regions to understand their adaptation to local environmental circumstances.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09231-6>.

Additional file 1: Table S1: Analysis of molecular variance within and among breeds of Sudan desert sheep. **Figure S1.** Unrooted NJ tree for Haplogroup A and B of Sudan desert sheep mtDNA D-loop haplotypes. **Figure S2.** Mismatch distribution of pairwise nucleotide differences at the haplogroup level (a, b) and at breed level (c, Hamary, d, Kabashi; e, Crossbreed). **Figure S3.** Coalescent Bayesian skyline plots at the breed level, a, Hamary; b, Kabashi; c, crossbreed.

Acknowledgements

We acknowledge the partial financial support from Japan Society for the Promotion of Science (JSPS) KAKENHI [22H02505] to the last author. We are also grateful to Sheep keepers in Sudan for their assistance and permission to sample their herds.

Author contributions

BS, designed, conceptualized research and supervised the work. BS, SA contributed samples. BS, RN carried out the experiments. BS, NSM analyzed data. BS, MA, RN funding acquisition. BS, NSM, MA, OH contributed to the interpretation of the results. BS, MA, RN administrated project. BS, NSM. SA writing original draft. BS, NSM. SA and OH writing, reviewing and editing. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Funding

Not applicable.

Availability of data and material

The sequences obtained were deposited to the DNA Data Bank of Japan (<http://www.ddbj.nig.ac.jp>) under accession numbers LC456425 – LC456544.

Declarations

Conflicts of interest

The authors declare no competing interests.

Ethics approval and consent to participate

All experimental protocols were approved with the experimental protocols and approval by the, Faculty of Veterinary Medicine, University of Khartoum, research review board.

Informed consent has been obtained from all the owners.

All methods were carried out in accordance with the Faculty of Veterinary Medicine, University of Khartoum, according to their guidelines for sampling domestic animals in Sudan and in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

Received: 17 October 2022 / Accepted: 6 March 2023

Published online: 16 March 2023

References

- Pereira F, Queiros S, Gusmao L, Nijman IJ, Cuppen E, Lenstra JA, et al. Tracing the history of goat pastoralism: new clues from mitochondrial and Y chromosome DNA in North Africa. *Mol Biol Evol.* 2009;26:2765–73.
- Bruford MW, Townsend SJ. Mitochondrial DNA diversity in modern sheep. Documenting domestication: New genetic and archaeological paradigms. 2006;306–16.
- Tapio M, Marzanov N, Ozerov M, Činkulov M, Gonzarenko G, Kiselyova T, et al. Sheep mitochondrial DNA variation in European, caucasian, and central Asian areas. *Mol Biol Evol.* 2006;23:1776–83.
- Meadows JR, Cernal I, Karaca O, Gootwine E, Kijas JW. Five ovine mitochondrial lineages identified from sheep breeds of the Near East. *Genetics.* 2007;175:1371–9.
- Dobney K, Larson G. Genetics and animal domestication: new windows on an elusive process. *J Zool.* 2006;269:261–71.
- Tarekegn GM, Tesfaye K, Mwai OA, Djikeng A, Dessie T, Birungi J, et al. Mitochondrial DNA variation reveals maternal origins and demographic dynamics of Ethiopian indigenous goats. *Ecol Evol.* 2018;8:1543–53.
- Hiendleder S, Mainz K, Plante Y, Lewalski H. Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. *J Hered.* 1998;89:113–20.
- Deng J, Xie X-L, Wang D-F, Zhao C, Lv F-H, Li X, et al. Paternal origins and migratory episodes of domestic sheep. *Curr Biol.* 2020;30:4085–95.
- Epstein H. The origin of the domestic animals of Africa. 1971.
- Abualazayim M. Animal Wealth and Animal Production in Sudan. 2004.
- Wilson RT. Livestock in the Republic of the Sudan: Policies, production, problems and possibilities. *Animal Husbandry, Dairy and Veterinary Science.* 2018;2:1–12.
- Liu J, Ding X, Zeng Y, Yue Y, Guo X, Guo T, et al. Genetic diversity and phylogenetic evolution of Tibetan sheep based on mtDNA D-Loop sequences. *PLoS ONE.* 2016;11:e0159308.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
- Meadows J, Hiendleder S, Kijas J. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity.* 2011;106:700–6.
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25:1451–2.
- Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10:564–7.
- Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol.* 1992;9:552–69.
- Harpending H. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human biology.* 1994;5:591–600.
- Fu Y-X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics.* 1997;147:915–25.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123:585–95.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 2005;22:1185–92.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution.* 2012;29:1969–73.
- Salim B, Taha KM, Hanotte O, Mwacharo JM. Historical demographic profiles and genetic variation of the East African Butana and Kenana indigenous dairy zebu cattle. *Anim Genet.* 2014;45:782–90.
- Ho SY, Lanfear R, Phillips MJ, Barnes I, Thomas JA, Kolokotronis S-O, et al. Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Syst Biol.* 2011;60:366–75.
- Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics.* 1991;129:555–62.
- Muigai AW, Hanotte O. The origin of African sheep: archaeological and genetic perspectives. *Afr Archaeol Rev.* 2013;30:39–50.
- Gornas N, Weimann C, El Hussien A, Erhardt G. Genetic characterization of local Sudanese sheep breeds using DNA markers. *Small Ruminant Research.* 2011;95:27–33.
- Resende A, Gonçalves J, Muigai AW, Pereira F. Mitochondrial DNA variation of domestic sheep (*Ovis aries*) in Kenya. *Anim Genet.* 2016;47:377–81.
- Nigusie H, Mwacharo JM, Osama S, Agaba M, Mekasha Y, Kebede K, et al. Genetic diversity and matrilineal genetic origin of fat-rumped sheep in Ethiopia. *Trop Anim Health Prod.* 2019;51:1393–404.
- Ghernouti N, Bodinier M, Ranebi D, Maftah A, Petit D, Gaouar S. Control Region of mtDNA identifies three migration events of sheep breeds in Algeria. *Small Ruminant Research.* 2017;155:66–71.
- Sharma R, Ahlawat S, Sharma H, Sharma P, Panchal P, Arora R, et al. Microsatellite and mitochondrial DNA analyses unveil the genetic structure of native sheep breeds from three major agro-ecological regions of India. *Sci Rep.* 2020;10:1–13.
- Othman OE, Pariset L, Balabel EA, Mariotti M. Genetic characterization of Egyptian and Italian sheep breeds using mitochondrial DNA. *J Genetic Eng Biotechnol.* 2015;13:79–86.
- Gáspárdy A, Zenke P, Kovács E, Annus K, Posta J, Sáfár L, et al. Evaluation of maternal genetic background of two Hungarian autochthonous sheep breeds coming from different geographical directions. *Animals.* 2022;12:218.
- Mariotti M, Valentini A, Marsan PA, Pariset L. Mitochondrial DNA of seven Italian sheep breeds shows faint signatures of domestication and suggests recent breed formation. *Mitochondrial DNA.* 2013;24:577–83.
- Singh S, Kumar S Jr, Kolte AP, Kumar S. Extensive variation and sub-structuring in lineage A mtDNA in Indian sheep: genetic evidence for domestication of sheep in India. *PLoS ONE.* 2013;8:e77858.
- Bruford M. Molecular approaches to understanding animal domestication: what have we learned so far. 2005. p. 6–8.
- Wanjala G, Bagi Z, Kusza S. Meta-analysis of mitochondrial DNA Control Region Diversity to Shed Light on phylogenetic relationship and demographic history of African Sheep (*Ovis aries*) breeds. *Biology.* 2021;10:762.

Publisher's Note

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