Sahelian pastoralism from the perspective of variants associated with lactase persistence.

Edita Priehodová, Frederic Austerlitz, Martina Čížková, Jana Nováčková, Francois-Xavier Ricaut, Zuzana Hofmanová, Carina Schlebusch, Viktor Černý

To cite this version:

Edita Priehodová, Frederic Austerlitz, Martina Čížková, Jana Nováčková, Francois-Xavier Ricaut, et al.. Sahelian pastoralism from the perspective of variants associated with lactase persistence.. American Journal of Physical Anthropology, 2020, pp.e24116. 10.1002/ajpa.24116 . hal-02919786

HAL Id: hal-02919786
https://hal.science/hal-02919786v1

Submitted on 13 Nov 2020

Title: Sahelian pastoralism from the perspective of variants associated with lactase persistence

Running title: Lactase persistence variants in the Sahel

Authors: Edita Priehodová,[1] Frédéric Austerlitz,[2] Martina Čížková,[1] Jana Nováčková,[1] François-Xavier Ricaut,[3] Zuzana Hofmanová,[1] Carina M. Schlebusch,[4,5,6] Viktor Černý[1*]

[1] Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Czech Republic
[2] UMR 7206 EcoAnthropologie et Ethnobiologie, CNRS/MNHN/Université Paris Diderot, Musée de l'Homme, Paris, France
[3] Department of Evolution and Biological Diversity (UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, Toulouse, France
[4] Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden
[5] Palaeo-Research Institute, University of Johannesburg, P.O. Box 524, Auckland Park, 2006, South Africa
[6] SciLifeLab, Uppsala

*Corresponding author Viktor Černý, cerny@arup.cas.cz

Abstract

Objectives. Archaeological evidence shows that first nomadic pastoralists came to the African Sahel from northeastern Sahara, where milking is reported by ~7.5 ka. A second wave of pastoralists arrived with the expansion of Arabic tribes in 7th–14th century CE. All Sahelian pastoralists depend on milk production but genetic diversity underlying their lactase persistence is poorly understood.

Materials and methods. We detected SNP variants associated with lactase persistence in 1,148 individuals from 29 mostly pastoralist populations in the Sahel. Then we analyzed six SNPs in the neighboring fragment (419 kb) in the Fulani and Tuareg with mutation –13910*T, reconstructed haplotypes, and calculated expansion age and growth rate of this variant.

Results. Our results reveal a geographic localization of two different lactase persistence variants in the Sahel: –13910*T west of Lake Chad (Fulani and Tuareg pastoralists) and –13915*G east of there (mostly Arabic-speaking pastoralists). We show that –13910*T has a more diversified haplotype background among the Fulani than among the Tuareg and that the age estimate for expansion of this variant among the Fulani (~8.5 ka) corresponds to introduction of cattle to the area.

Conclusions. This is the first study showing that the 'Eurasian' lactase persistence allele –13910*T is widespread both in northern Europe and in the Sahel, where it is, however, limited to pastoralists. Since the Fulani haplotype with –13910*T is shared with contemporary Eurasians, its origin could be in a region encompassing the Near East and northeastern Africa in a population ancestral to both Saharan pastoralists and European farmers.

Keywords: African Sahel, lactase persistence, subsistence, pastoralism

Introduction

Various food-production strategies first emerged during the Neolithic. This remarkable cultural innovation significantly changed human populations in both a social (Hodder 1990) and biological sense (Wolfe et al. 2007). Sub-Saharan Africa with its remarkable wild food resources (National Research Council 2008) has always been viewed as the home of diverse hunter-gatherer populations (Cummings et al. 2014). Its genetic structure, however, has been significantly reshaped by several genetic adaptations related to food-production strategies in the Holocene and pathogen-driven selection related to culturally modified environments (Podgorná et al. 2015; Sanchez-Mazas et al. 2017; Tishkoff et al. 2007; Triska et al. 2015). With respect to food-production strategies found in Africa, it is believed that nomadic pastoralism preceded sedentary farming (Marshall and Hildebrand 2002), but both of these subsistence strategies contributed to population expansions and admixture (Schlebusch and Jakobsson 2018).

In sub-Saharan Africa, only few animal species were eligible for domestication (Diamond 2002; Diamond 1997) but local environment provided plentiful palatable grasses suitable for grazing domesticated animals, which were introduced into this region soon after the beginning of the Neolithic in the Near East. In Early and Middle Holocene, the Sahara was green, covered by vast expanses of grassland. Due to subsequent climatic changes, by ~5 ka this environmental setting receded to the Sahel/Savannah belt, where both pastoralism and farming have been widely practiced ever since (Batello et al. 2004; Boutrais 1988; Sougnabe 2011). Some models even suggest that ancient forms of pastoralism may have delayed the collapse of Middle Holocene Saharan ecosystem, which was brought about by climatic desiccation (Brierley et al. 2018). Traditional pastoralism can thus be viewed as a suitable cultural adaptation to the Sahelian environment and it has been reported that children of Sahelian pastoralists are better nourished than the children of farmers (Pedersen and Benjaminsen 2008; Sheik-Mohamed and Velema 1999).

Population genetic studies show that the African Sahel/Savannah belt formed a bidirectional corridor of human migrations between western and eastern Africa, with Lake Chad acting as a ‘center of gravity’, where in the course of the Holocene new haplogroups emerged and diversified and where the subsistence pattern had a more significant impact on population structure than linguistic affiliation did (Černý et al. 2009; Černý et al. 2011; Černý et al. 2007; Nováčková et al. 2020; Podgorná et al. 2013). The origins of food production strategies in the Sahel/Savannah belt are not yet understood in detail but archaeological investigations show that the first food-producers in this area were pastoralists coming from the Sahara at ~5 ka (Ozainne 2013). Somewhat later, they domesticated pearl millet in the Tilemsi Valley (northern affluent of the Niger River in Mali), while some of the groups became mixed agro-pastoralists and spread to other Sahelian locations in western Africa. Other Sahelian areas, such as the Chad Basin were not, however, settled by these agro-pastoralists from Middle Niger but by herders coming from outside the first pearl millet domestication area. These herders possibly originated in either Northern Niger/Chad or Eastern Sahara (Ozainne et al. 2014; Ozainne et al. 2017).

For the Sahelian pastoralists, the most important animal has always been cattle, whose milk – as reported in several anthropological and ethnological studies (Bernus 1981; Dupire 1962; Evans-Pritchard 1940; Homewood 2008; Lhote 1955) – was consumed fresh. Although the auroch, ancestor of the domestic cattle, also lived in northeastern Africa, as evidenced by rock art and archaeological findings (Le Quellec 1997), its domestication took place in Asia. The auroch was domesticated independently in the Near East and in India and two different breeds were subsequently introduced to Africa (Mwai et al. 2015). The route these Near Eastern cattle keepers took went through the Nile Valley and then from ~8 ka on via Wadi Howar to Central Sahara (Jesse et al. 2013), while Indian cattle was imported to East Africa by a sea route across the Indian Ocean and then via Southern Arabia at ~4 ka (Boivin et al. 2009). It is still unclear whether Near Eastern cattle (*Bos taurus* or taurin) and/or the Indian cattle (*Bos indicus* or zebu) had subsequently mixed with aurochs living in Africa. Nevertheless, the high diversity of currently kept indigenous African cattle breeds and selection signals in their gene pool attest to these animals' genetic adaptation to Sahelian tropical environmental settings (Kim et al. 2017).

In Europe, animal husbandry during the initial phases of the Neolithic merely supplemented widespread cereal cultivation, while milking appeared as a later innovation during the 'Secondary Product Revolution' (Sherratt 1981). In sub-Saharan Africa (Smith 2005) and Arabia (McCorriston et al. 2012; McCorriston and Martin 2009), on the other hand, the first food-producing societies were nomadic pastoralists. These pastoralists gradually adapted to the specific environment of the Sahel/Savannah belt (Linseele 2013) by adopting a system of seasonal movements (transhumance) that is incompatible with the sedentary lifestyle of traditional farmers (Pedersen and Benjaminsen 2008). It has been suggested that these African cattle keepers supplemented their diet – based initially on hunting wild game and harvesting of wild plants (Cancellieri and di Lernia 2014) – by milking already by around 7,5 ka (Dunne et al. 2012). The amount of milk used at that time for human consumption is, however, still uncertain (Le Quellec 2016).

Milk drinking is of vital importance to all mammals that depend on breast milk as a food source before weaning. In addition to other nutritional components, it contains a sugar, lactose, which is the main source of energy for infants (Ségurel and Bon 2017). To obtain energy from lactose, almost all mammals synthetize an enzyme called lactase-phlorizin hydrolase (LPH), encoded by the lactase gene (*LCT*). In populations where dairy production is common, downregulation of LPH production after weaning – the original physiological state in all LPH-producing mammals – has been counterselected. In these groups, milk consumption became no longer limited to the first years of life. Especially in groups highly dependent on pastoralism, fresh milk became the main constituent of their diet, and ability to produce lactase after weaning (i.e. lactase persistence, henceforth LP) became an advantageous trait. In these populations, specific mutations which maintained LPH production have been under strong positive selection, as evidenced by high selection coefficients and hard sweep signals associated with LP, which have been found in many populations (Bersaglieri et al. 2004; Enattah et al. 2008; Jones et al. 2013; Ranciaro et al. 2014; Schlebusch et al. 2013; Tishkoff et al. 2007).

Several variants associated with LP are found in current populations. There are three different mutations in eastern Africa (–14010*C, –14009*G, and –13907*G), one in western Eurasia and northwestern Africa (–13910*T), and one in Arabia and northeastern Africa (–13915*G). These mutations can be viewed as evidence of convergent genetic adaptation of humans to pastoralism (Tishkoff et al. 2007). It has been estimated that allele –13910*T is the oldest among these variants. Simulation models show that it emerged somewhere between central Europe and northern Balkans at ~7.5 ka (Itan et al. 2009), but its presence has not yet been attested by ancient DNA studies of European human remains from this period (Burger et al. 2007; Krüttli et al. 2014; Malmström et al. 2010). Currently prevailing view is that this allele had been present in the population at a very low frequency and became strongly selected for much later, perhaps during the Bronze Age, in conjunction with other alleles associated with lipid metabolism (Mathieson and Mathieson 2018).

As mentioned above, the –13910*T variant occurs not only in western Eurasia but also in northwestern Africa (Coelho et al. 2005; Lokki et al. 2011; Ranciaro et al. 2014). Based on high frequencies of –13910*T and Eurasian Y chromosome haplotypes in the Fulani (Fulbe) from Cameroon, it has been suggested that this LP variant may have been introduced in their population from outside Africa (Mulcare et al. 2004). Based on (probably different) Fulani DNA samples but again from northern Cameroon, it has been demonstrated that their –13910*T variant shares its haplotype backgrounds with Europeans (Coelho et al. 2005; Ranciaro et al. 2014), which suggests a single mutation event for this mutation and its introduction via admixture of non-African sources and through gene flow. A recent study of Fulani pastoralists from Ziniaré, Burkina Faso (Vicente et al. 2019) confirmed these suggestions and based on very high molecular resolution showed that this variant shares its haplotype background with Eurasians and that ancestors of the Fulani probably acquired the haplotype via very recent admixture events which took place in northern Africa.

The Fulani are, however, spread through a much larger area of the Sahelian Africa and samples from the groups that live in Cameroon and Burkina Faso cannot represent them fully. Other Fulani populations are concentrated around the rivers of Senegal, and Niger, along Hausaland in Niger and Nigeria, as well as around Lake Chad. From these locations, no data on LP variation of this originally nomadic people have been available so far. The aim of this article is to fill this gap in our knowledge and gain a deeper understanding of the history of (Fulani) pastoralism in sub-Saharan Africa by providing new data for LP variants from 31 Sahelian, mostly Fulani, populations. Thanks to such a large LP dataset, we can also test the hypothesis regarding their mutual prehistorical contacts and past gene flows between nomadic pastoralists and sedentary farmers in the Sahel/Savannah belt previously indicated by the results of studies on uniparental markers (Černý et al. 2011; Čížková et al. 2017; Nováčková et al. 2020).

Materials and methods

We analyzed LP variants in 31 African populations living in the Sahel/Savannah belt (see Table 1 for their geographical locations, lifestyles, and language affiliations). Sampling authorizations were obtained from the relevant state institutions of countries involved in the study prior to sample collection from participants. Ethical approval was obtained from the Charles University in Prague (Approval Number 2019/12).

We focused primarily on nomadic pastoralists, where a higher frequency of LP alleles was to be expected. Additionally, however, we included all existing linguistic families in the Sahel/Savannah belt, such as the Fulani who belong to the Niger-Congo family, the Tuareg, Afar, Beja, Kababish, and Moors who all belong to the Afro-Asiatic family, and the Zaghawa who belong to the Nilo-Saharan family. Moreover, we also analyzed populations traditionally seen as sedentary farmers, such as the Gurunsi, Gourmantche, and Mossi of the Niger-Congo family, the Masa, Kotoko, Buduma, and Dangaléat of the Afro-Asiatic family, and the Daju, who are affiliated with Nilo-Saharan languages.

By doing so, we included all three linguistic groups and both lifestyles currently found in the Sahel/Savannah belt, but a majority of our samples come from two pastoral populations: the Fulani and the Arabs. While Fulani are autochthonous sub-Saharan population groups whose ancestors admixed with population(s) of Eurasian ancestry (Kulichová et al. 2017; Triska et al. 2015; Vicente et al. 2019), Arabs are an originally Eurasian population which arrived from Arabia from the 7th century CE onwards. They were present in the Lake Chad basin in the 14th century (Zeltner 2002) and nowadays, they are admixed with local sub-Saharan populations (Priehodová et al. 2014; Priehodová et al. 2017).

In total, we screened 1,241 individuals (2,482 chromosomes) for variants associated with LP. We sequenced a 359-bp fragment located in intron 13 of the *MCM6* gene, where the five main LP-associated mutations are located. We used primers reported in a previous study (Coelho et al. 2009). PCR products were sequenced with forward primers and in case of ambiguity, we generated reverse complements. The variants were identified using BioEdit software (Hall 1999) and allele frequencies and homozygous and heterozygous variants were calculated by GenAlEx software (Peakall and Smouse 2012). Arlequin software (Excoffier and Lischer 2010) was used to perform analyses of molecular variance (AMOVA). We have also produced a non-metric multidimensional scaling plot (nMDS) computed by R-package *vegan 2.5-6* (Oksanen et al. 2019) based on Reynolds' distances matrix from Arlequin (Excoffier and Lischer 2010; Reynolds et al. 1983).

LP variant frequencies of the 31 analyzed Sahelian populations were subsequently compared with previously published LP variant frequencies in other Sahelian and neighboring populations. In this manner, we obtained a dataset including 101 populations from the Sahel/Savannah belt or its close vicinity (Supplementary Table S1). To visualize the distribution of LP variants across this dataset, we used an open source geographic information system, the QGIS ver. 3.4.9 LTR Madeira, with free maps from Natural Earth quick start packages. Frequencies of LP variants, geographical coordinates, and information about lifestyle were used to produce pie charts. Variance structure of the four main LP genotypes was computed and visualized by principal component analysis (PCA) using allele frequencies available for the whole dataset. We applied the *prcomp* method from *stats v3.6.1* package and further plotting was done using the *autoplot* function in *survMisc v0.4.2* (both found in the R). Groups of populations were differentiated by specific color palettes. To determine and visualize the geographical distribution of LP variants, we drew interpolation maps using ArcView software.

A clinal pattern of LP mutations within the Sahel/Savannah belt was tested by correlogram autocorrelation analysis (Moran 1950), implemented in PASSaGE software, version 2 (Rosenberg 2001), and applied with 10,000 permutations. We tried different numbers of classes with equal observations but results presented in this study were achieved with 10 classes. These spatial analyses included other previously published datasets, which were based on a total of 101 populations from the Sahel/Savannah belt and neighboring locations (see Supplementary Table S1).

The age of expansion and growth rate of the –13,910*T allele was estimated by a method described by Austerlitz et al. (Austerlitz et al. 2003), which was previously used to estimate the expansion age and growth rate of –13915*G in Arabic-speaking populations (Priehodová et al. 2017). For this analysis, we genotyped six SNPs in a 419 kb-long fragment around –13910*T (Heyer et al. 2011). We decided to apply a High Resolution Melting analysis that needs samples of over 40 ng of DNA, i.e. more than we had available in all our samples. For this reason, we worked with two smaller sub-subsets: a Fulani (n=286) and a Tuareg (n=45) one.

Subsequently, we reconstructed the haplotypes using the Phase program (Stephens and Scheet 2005; Stephens et al. 2001) with settings identical to those used by Priehodová et al. (2017), and repeated these results using Arlequin software (Epsilon value = 1e-12 and 10,000 iterations) (Excoffier and Lischer 2010). Finally, we estimated the age of expansion (g) and geometric growth rate (*r*) of the –13,910*T allele in the Fulani and Tuareg populations separately using a method developed in Austerlitz et al. (2003). This method takes as inputs the current number of copies (*k*) of the allele in a population and the frequency (p_{nr}) of the most frequent haplotype for a set of markers around the locus of interest. It assumes that all current copies of the allele trace back to the most recent common ancestor (MRCA) and that since that time, the allele has been spreading at a constant rate (*r*), with g being the time separating all current copies from their MRCA, i.e. their coalescent times. The probability of reaching the observed values of *k* and p_{nr} can be obtained from eqs (1) and (5) respectively in Austerlitz et al. (2003). These probabilities can be combined into a likelihood formula that allows us to estimate *g* and *r* from the values of *k* and p_{nr}. The value of *k* can be obtained as $2Np$, where N is the current census size of the population and *p* is the allelic frequency of the mutant allele in the population. While we could estimate *p* using our sample, the current population size (*N*) is difficult to estimate in nomadic pastoralists. For the Fulani, we assumed a population size ranging between 10 and 30 million and for the Tuareg, 1.3 million (Eberhard et al. 2019). p_{nr} was obtained from the haplotypes reconstructed with PHASE (see above). The method requires an estimate of the total recombination rate (θ) of the haplotype, which we estimated using HAPMAP map build 37.

Since the method focuses on the number of copies of the allele, tracing all the way back to the ancestral copy of this allele, it is affected neither by the population size at the time of occurrence of this copy, nor by the fact that the population had since then expanded in size. It assumes that the number of copies followed a geometric growth at a constant rate *r* since that time. This growth rate combines the demographic growth rate of the population and specific growth rate for the gene linked for instance to selection. If there was, for instance, first a period where the number of copies evolved only through drift followed then by selection of standing variation, the estimated *r* would be an average of the growth rate of the two periods, while the expansion age (g) would be in any case the estimated coalescence time of all copies of the allele. It would also apply to the case of gene flow, where *r* would correspond to the growth rate of the allele in the sets of connected populations, while g would still be an estimate of this coalescence time. The method also assumes that the mutant allele appeared through a unique mutation event.

Results

We screened the sequences for five variants (–13907*G, –13910*T, –13915*G, –14009*G, and –14010*C) associated with LP. The most frequent variant in our dataset, –13910*T, was found in all analyzed western Sahelian pastoralists such as the Fulani, Tuareg, and Moors in more or less similar frequencies but was not present in other groups (Table 2). The second most widespread variant in our dataset, –13915*G, was found in seven populations living in the eastern part of the region. It occurred at the highest frequency among the Kababish Arabs (33.3%), while in other populations, its frequency was rather low. The third most frequent variant was –14009*G, found almost exclusively among the Beja (42.6%): in other groups, it was present at much lower frequencies. Variant –13907*G was found in seven Afar, five Beja, and one Kababish individual, and variant –14010*C was not found in our dataset at all. The LP variants of the Sahelian and neighboring populations are shown in Figure 1.

There are, however, also other mutations not currently associated with LP but observed in the sequences we obtained. An interesting case is that of –13913*T, which had been proven not to be associated with LP (Jones et al. 2013). In our dataset, however, it was found only in nomadic pastoralists, such as the Fulani, Tuareg, and Zaghawa. Another, and more frequent, variant is the –14107*G, found in many Sahelian populations, such as the Afar, Buduma, Fulani, Gurunsi, Kotoko, and Masa. It has likewise been shown that it is not an LP-associated variant (Torniainen et al. 2009) and in our dataset, it failed to correlate with subsistence patterns. Another interesting finding was a T to C change in position –13915, found in two Fulani from Diafarabe, Mali.

Molecular variance within and among populations was evaluated by AMOVA. The overall variance for the most frequent variants (–13907*G, –13910*T, –13915*G, –14009*G, –13913*T, and –14107*G) expressed as a percentage was 73.0% within populations and 27.0% among populations. For variant –13910*T alone, values were 70.6% within populations and 29.4% among populations. We also used MDS to visualize affinities among the populations analyzed for the six main variant positions (Supplementary Figure S1), but population clustering turned out to be governed mainly by frequency of the –13910*T. When only –13910*T was considered, all eastern Sahelian and sedentary populations cluster together, showing that this variant does not contribute to differentiation in this area. It displays a striking geographical division between the distribution of variants –13910*T and –13915*G and differences in frequencies of these variants between pastoralists and farmers.

The frequencies of LP variants found in our populations were compared with those obtained in previously analyzed African populations (see Supplementary Table S1). Unfortunately, only four variants were available but the entire dataset contains a much larger number of populations. PCA showed that principal components 1 and 2 are directed by –13910*T and –13915*G, respectively, and jointly they explain 93.4% of variation in the dataset (Figure 2). Another striking observation of the study is the virtual absence of LP variants among Nilotic pastoralists, such as the Dinka, Nuer, and Shilluk.

To assess whether frequencies of the mutations as shown in geographical space (Figure 3) correspond to a clinal pattern, we applied Moran’s I correlograms for the –13910*T and –13915*G variants (based on the entire dataset) and showed that frequency distribution of these variants is not random. In fact, –13910*T shows significantly positive autocorrelation in all distance classes up to

1,800 km and significantly negative estimates above 2,600 km (Supplementary Figure S2), thus clearly corresponding to a clinal pattern (Barbujani 2000). Variant –13915*G showed a similar pattern except for displaying positive autocorrelations only up to 800 km with several non-significant values, therefore corresponding to a less clear clinal pattern.

After genotyping six SNPs around the –13910*T position in the Fulani and Tuareg samples, we took all 198 Fulani individuals bearing –13910*T and identified 22 haplotypes in the first three runs of the Phase software. For one haplotype, the fourth run differed from the previous three runs, and the fifth run produced one haplotype less than the other four runs. We repeated the phasing with the algorithm implemented in Arlequin: this yielded the same haplotypes as in the first three runs of Phase and we have therefore decided to use this haplotype set for further calculations. For the Tuareg subset, all five runs in Phase were congruent. We obtained 11 haplotypes from 26 samples bearing –13910*T.

Using the method of Austerlitz et al. (2003), we estimated the expansion age of –13910*T in the Fulani population at 7,534–9,686 ya (Table 3), along with a slow growth rate of 1.033–1.053. Conversely, for the Tuareg we estimated a much younger date of 1,541–2,200 ya, associated with a faster growth rate of 1.16–1.25. These estimates reflected the different haplotype distributions in these two populations. We found a low haplotype diversity in the Tuareg, with a strongly predominant single haplotype (23 out of 26), while in the Fulani, several haplotypes were present at comparable frequencies. This difference was also striking when we compared the haplotype network (Supplementary Figure S3). These patterns were thus consistent with a rapid expansion of the –13910*T haplotype in the Tuareg, which did not leave much time for recombination, while expansion of this haplotype was much slower in the Fulani, leaving more time for recombination.

Discussion

Our investigation of LP variant frequencies revealed new and interesting results related to the origins of pastoralism and subsequent gene flow between pastoralists and farmers in the Sahel/Savannah belt of Africa. We observed a clear distinction between regions west and east of Lake Chad: while variant –13910*T prevails in western Sahel, where we found it only in pastoralists such as the Fulani, Tuareg, and Moors (it is virtually absent in sedentary farmers analyzed in our dataset). On the other hand, eastern Sahel is much more diverse and several LP-associated variants, such as –13907*G, –13915*G, and 14010*C, are found in various populations.

The almost zero frequency of the –13910*T variant in sedentary farmers is consistent with a hypothesis of very low gene flow in the western part of the Sahel/Savannah belt. Low gene flow in this region is further attested by analyses of mitochondrial DNA variation (Čížková et al. 2017; Kulichová et al. 2017) and supported by recent geometric morphometrics analyses (Kleisner et al. 2019). Presence of the "Arabian" –13915*G variant in many populations of the eastern Sahel (including several sedentary farmers) is, on the other hand, consistent with the idea of a higher gene flow in this part of Africa (Čížková et al. 2017), especially between the Arabic-speaking Baggara and their neighbors (Priehodová et al. 2017). The –14009*G variant has been associated with LP in eastern Africa (Jones et al. 2015; Jones et al. 2013) and frequency at which we found it in the Beja near Kassala, Sudan (42.6%) is much higher than reported by other studies (Hassan et al. 2016; Ranciaro et al. 2014). Such isolated occurrence and high frequency of this mutation among the Beja specifically could be explained by East African origin of this Cushitic-speaking population and subsequent genetic drift with a limited gene flow, which has been attested also by an mtDNA study (Čížková et al. 2017).

Interestingly, in all Nilotic pastoralists analyzed in this study we found zero frequencies of the main known LP variants. It is possible that these populations also carry other, yet unknown, LP variant since tests for lactose tolerance based on breath hydrogen determination (done before molecular genetic testing became available) showed that least 25% of persons in this population are lactose absorbers (Bayoumi et al. 1982); note, however, that in this research only the Dinka were examined in sufficient numbers. Nevertheless, at least some Nilotic populations are highly dependent on fresh milk consumption, as reported for the Nuer, where only surplus milk is used for cheese-making (Evans-Pritchard 1940). Non-problematic digestion of large quantities of lactose in these populations might be accounted for by the presence of a certain kinds of bacteria (including Bifidobacteria) in their gut microbiome, which could possibly provide lactose tolerance (Goodrich et al. 2017).

Low genetic structure of the Sahel/Savannah belt, revealed by uniparental markers, inspired the notion of a bidirectional migration corridor situated between the Sahara and tropical rainforests (Černý et al. 2009; Černý et al. 2011; Černý et al. 2007). Other studies, however, stress the importance of pastoralism as a barrier of gene flow (Bučková et al. 2013; Kleisner et al. 2019; Nováčková et al. 2020; Shriner and Rotimi 2018). In fact, a high degree of endogamy can be observed among the Fulani pastoralists even today (Hampshire and Smith 2001), especially in fully nomadic groups such as *Wodaabe* (sing. *Bodaaɗo*), where finding of a partner is mostly limited to ceremonial meetings irregularly organized by certain lineages (Loncke 2015). A split between full-time pastoralism and sedentary farming thus hindered gene flow between populations practicing two subsistence systems and exchange of alleles associated with LP was limited.

Our study shows that the age of expansion of –13910*T obtained for the Fulani pastoralists in western Sahel/Savannah belt (7,534–9,686 years ago) is much older than that obtained for the –13915*G in Arabic pastoralists in eastern Sahel/Savannah belt (Priehodová et al. 2017), when applying the same method (Austerlitz et al. 2003). Priehodová et al. 2017 estimated age of –13915*G at between 1,274–1,782 years ago, which is consistent with the time of Arabic expansion to Africa. The estimate obtained in this study for the Fulani pastoralists conforms to estimates for introduction of pastoralism in Africa which are based on archeological evidence. There is clear evidence of early Middle Holocene migration(s) of pastoralists from the Near East to northeast Africa (Jesse et al. 2013; Kuper and Riemer 2013) as well as of the fact that these pastoralists later established themselves in western Africa (Linseele 2013; Ozainne et al. 2014). It has been suggested that presence of the –13910*T variant in North African Berber population could be a sign of past gene flow from Near Eastern pastoralists (Myles et al. 2005).

This first Saharan/Sahelian influx of population with LP individuals was followed by a later migration of other pastoralists from the Arabian Peninsula throughout northern Africa and Sahara up to the Lake Chad Basin, which was led by Bedouin Arabic-speaking tribes (Braukämper 1994; Levy and Holl 2002; Zeltner 2002). This migration probably added to this region the –13915*G variant which arose in Arabia as an adaptation to the consumption of camel milk (Enattah et al. 2008; Imtiaz et al. 2007; Priehodová et al. 2014).

In this context, we would like to mention the Baggarization theory, which explains how the incoming Arabic tribes abandoned camels in favor of cows, possibly through contact with Fulani pastoralists. This phenomenon was revealed by both linguistic and ethnographic data (Braukämper 1994; Manfredi 2012; Owens 1994), which indicate an incorporation of not only Fulani linguistic and cultural items but also individuals into the Arabic population. Our last study also revealed that this theory can be supported by genetics (Nováčková et al. 2020). The only slight disagreement between the various versions of Baggarization hypothesis lies in the dating and localization of the Arab-Fulani contact. While Owens suggests it took place in the Bagirmi region prior to 1500 CE, Braukämper indicates a later time, around 1635 CE, and localizes the process further to the east, towards Waddai (both regions are located in Chad, east of the Lake). Our study revealed several populations bearing both –13910*T and –13915*G that live as far as in Sudan, but most populations which have both of these mutations are nowadays located within Lake Chad Basin, i.e. in the Bagirmi area.

It has been suggested that the Fulani acquired the –13910*T haplotype by admixing with a North African population which had some Eurasian ancestry, so that this variant in the Fulani population has the same descent as the variant found in Eurasia (Coelho et al. 2005; Mulcare et al. 2004; Ranciaro et al. 2014). Furthermore, it has been estimated that two distinct admixture events took place, one 1,828 and the other 302 years ago (Vicente et al. 2019). Our estimates of the growth and selection of –13910*T haplotype in the Fulani indicate an earlier dating than the abovementioned admixture times (Vicente et al. 2019) and are consistent rather with the estimates of Coelho et al, (2005), who proposes a period of 10,125–6,060 years ago.

The age of the second main LP Sahelian variant –13915*G, which we in our last study (Priehodová et al. 2017) estimated at 1,5 ka, perfectly matches the timing of migration of Bedouin tribes from Arabia to Africa (Zeltner, 2002). A previous estimate of the age of –13915*G, which was up to 4 ka (Enattah et al. 2008), concerns its origin in Arabia, probably Central Arabia. From there, it then expanded at the beginning of modern history, much latter than the –13910*T which was already established in the Fulani western Sahelian populations.

This discrepancy between the age estimates of –13910*T in the Fulani population can be explained by the fact that the two studies, i.e. our present study and Vicente et al.'s (2019), targeted different parts of the genome (a local haplotype age of growth estimate vs. a genome-wide estimate of admixture), used different dating methods, and the samples on which they were based covered different populations. Moreover, Vicente et al. (2019) analyzed only one Fulani population (from Ziniaré, Burkina Faso), while our results are based on a larger Fulani dataset stretching geographically from Senegal to Chad. Nevertheless, we want to stress that our results do not exclude the possibility of a later admixture and introgression as suggested by Vicente et al. 2019. In fact, the *TAS2R* genes alleles may have also been introduced to the Fulani via a recent admixture with a Eurasian population (Triska et al. 2015). In short, these results reflect the complexity of the population history of Africa.

Sampling coverage, too, may have affected the results of our present study. While our Fulani dataset covers almost all places where this nomadic population lives, the three Tuareg samples are only from the southern part of their geographic range. We cannot therefore exclude the possibility that greater diversity would have been captured had we analyzed also Saharan Tuareg samples (from Algeria and Libya) and not only the Sahelian ones we had at our disposal (from Mali, Burkina Faso, and Niger). Further research is needed to resolve this issue.

One may wonder whether some processes, such as gene flow or selection on a standing variation, may have also affected our estimates of growth rates and ages of the mutation. Gene flow is unlikely in the Tuareg population. It would have introduced other haplotypes into the population, which seems rather improbable in the light of results of relatively recent study on uniparental markers in the Tuareg (Pereira et al. 2010). Similarly, there would be more haplotype diversity in the Tuareg population had there been standing variation when the selection started.

Regarding the Fulani, we cannot completely exclude the possibility that their high haplotype diversity results partly from gene flow and/or selection on standing variation. However, as noted above, there are several pieces of genetic and ethnological evidence to the effect that gene flow with neighboring populations was rather limited, which indicates that this factor is unlikely to have biased our estimates. Regarding standing variation, several haplotypes may have indeed been present already in the founding population. That, however, would not have a substantial impact on our estimate of the age of the mutation, which is based on the estimated coalescent time of all carriers of the mutant allele in this population, i.e. on the time to their common ancestor. This ancestor may but need not have lived within the territory where the Fulani currently reside. What is important is that our estimate of the age of this mutation is consistent with the emergence of pastoralism. This would have held even in the case of extensive gene flow, since even then all lineages carrying the mutant allele would still coalesce at the time of emergence of pastoralism.

Because the –13910*T haplotype associated with LP is shared between the Fulani from Ziniaré, Burkina Faso, and Europeans (Vicente et al. 2019) and since we found an old age of expansion for this mutation within the African population of the Fulani (dataset covering much larger area of the Sahel), it is possible that the –13910*T may have originated in a food-producing population that lived in the Near East, including northeastern Africa, as suggested by Myles et al. (2005). Subsequently, though, the first Neolithic farmers who lived in the Near East/northeastern Africa were replaced by various population movements (Černý et al. 2016; Vyas et al. 2016), with the results that nowadays only the originally 'Arabian' –13915*G LP variant is present in this region (Priehodová et al. 2014).

Interestingly, genetically the nearest population to the first Near Eastern farmers has been conserved in Sardinia (Chiang et al. 2018; Sikora et al. 2014), where the frequency of –13910*T is as low as elsewhere in the Mediterranean (Gerbault 2013; Meloni et al. 2001). Since there is archaeological evidence of milking among pastoralists in Central Sahara by at least around ~7.5 ka (Dunne et al. 2012), which corresponds to our molecular estimates for the Fulani, it is possible that the first Neolithic sedentary populations consumed not fresh milk but rather dairy products such as cheese, where lactose content is low. Fresh milk consumption is, after all, a habit more common among nomadic pastoralists originating in northeastern Sahara than among sedentary farmers whose origin is in the Near East.

Although there is genetic evidence of gene flow from the Near East to North Africa (but not vice versa) in Late Pleistocene, ~15 ka (van de Loosdrecht et al. 2018), several studies based on uniparental markers show that a bidirectional gene flow occurred between these regions later, during the Holocene (Arredi et al. 2004; Ottoni et al. 2009; Ottoni et al. 2010; Pereira et al. 2010). Moreover, links between Africa and Eurasia are documented in studies of several mitochondrial haplogroups, such as U5 and H1 in Africa and L1b in Eurasia (Achilli et al. 2005; Cerezo et al. 2012; Kulichová et al. 2017), as well as in studies of Y chromosome haplogroups, such as E-M35 in Eurasia and R1b-V88 in Africa (Cruciani et al. 2010; Trombetta et al. 2015). Other, later migrations between Arabia, Near East, and North Africa (Černý et al. 2016; Fernandes et al. 2015) may be responsible for the fact that variant –13910*T is nowadays almost entirely absent from the Near East (Gerbault et al. 2011). More data on Eurasian and especially African populations is needed to fully understand the evolution of food-producing lifestyles and history of lactase persistence among human populations.

Acknowledgements

This research was supported by Czech Science Foundation grant #19-09352S-P505. FXR was supported by French ANR grant ANR-14-CE31-0013-01. CMS was supported by the Swedish Research Council (no. 621-2014-5211) and the European Research Council (ERC – no. 13 759933). We are grateful to Laure Ségurel from Musée de l'Homme for commenting on an earlier version of the manuscript. The authors would like to thank Anna Pilátová for proofreading the manuscript.

Data Availability Statement

Data are provided in the Table 2 and Supplementary Table S1.

References

Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., ... Torroni, A. (2005). Saami and Berbers - an unexpected mitochondrial DNA link. *American Journal of Human Genetics*, 76, 5, 883-886.

Arredi, B., Poloni, E.S., Paracchini, S., Zerjal, T., Fathallah, D.M., Makrelouf, M., … Tyler-Smith, C. (2004). A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *American Journal of Human Genetics*, 75, 2, 338-345.

Austerlitz, F., Kalaydjieva, L., & Heyer, E. (2003). Detecting population growth, selection and inherited fertility from haplotypic data in humans. *Genetics*, 165, 3, 1579-1586.

Barbujani G. (2000). Geographic patterns: how to identify them and why. *Human Biology*, 72, 1, 133-153.

Batello, C., Marzot, M., Touré, A.H., & Kenmore, P.E. (2004). *The future is an ancient lake: traditional knowledge, biodiversity, and genetic resources for food and agriculture in Lake Chad Basin ecosystems*. Rome: Food and Agriculture Organization of the United Nations.

Bayoumi, R.A., Flatz, S.D., Kuhnau, W., & Flatz, G. (1982). Beja and Nilotes: nomadic pastoralist groups in the Sudan with opposite distributions of the adult lactase phenotypes. *American Journal of Physical Anthropology*, 58, 2, 173-178.

Bernus, E. (1981). *Touaregs Nigeriens: unite culturelle et diversite regionale d'un peuple pasteur*. Paris: Mémoires ORSTOM no. 94.

Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., ... Hirschhorn, J.N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, 74, 6, 1111-1120.

Boivin, N., Blench, R., & Fuller, D.Q. (2009). Archaeological, Linguistic and Historical Sources on Ancient Seafaring: A Multidisciplinary Approach to the Study of Early Maritime Contact and Exchange in the Arabian Peninsula. In M.D. Petraglia, & J.I. Rose (Eds.), *The evolution of human populations in Arabia Paleoenvironments, prehistory and genetics* (pp. 251-278). New York: Springer.

Boutrais, J. (1988). *Des Peuls en savanes humides : développement pastoral dans l'ouest centrafricain*. Paris: ORSTOM.

Braukämper, U. (1994). Notes on the origin of Baggara Arab culture with special reference to the Shuwa. In J. Ownes (Ed.), *Arabs and Arabic in the Lake Chad region* (pp. 13-46). Köln: Rüdiger Köppe Verlag.

Brierley, C., Manning, K., & Maslin, M. (2018). Pastoralism may have delayed the end of the green Sahara. *Nature Communications*, 9, 1, 4018.

Bučková, J., Černý, V., & Novelletto, A. (2013). Multiple and differentiated contributions to the male gene pool of pastoral and farmer populations of the African Sahel. *American Journal of Physical Anthropology*, 151, 1, 10-21.

Burger, J., Kirchner, M., Bramanti, B., Haak, W., & Thomas, M.G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences*, 104, 10, 3736-3741.

Cancellieri, E., & di Lernia, S. (2014). Re-entering the central Sahara at the onset of the Holocene: A territorial approach to Early Acacus hunteregatherers (SW Libya). *Quaternary International*, 320, 43-62.

Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gomez-Carballa, A., Brisighelli, F., ... Salas, A. (2012). Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Research*, 22, 5, 821-826.

Coelho, M., Luiselli, D., Bertorelle, G., Lopes, A.I., Seixas, S., Destro-Bisol, G., & Rocha, J. (2005). Microsatellite variation and evolution of human lactase persistence. *Human Genetics*, 117, 4, 329-339.

Coelho, M., Sequeira, F., Luiselli, D., Beleza, S., & Rocha, J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evolutionary Biology*, 9, 80.

Cruciani, F., Trombetta, B., Sellitto, D., Massaia, A., Destro-Bisol, G., Watson, E., ... Scozzari, R. (2010). Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *European Journal of Human Genetics*, 18, 7, 800-807.

Cummings, V., Jordan, P., & Zvelebil, M. (2014). *The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers*. Oxford: Oxford University Press.

Černý, V., Čížková, M., Poloni, E.S., Al-Meeri, A., & Mulligan, C.J. (2016). Comprehensive view of the population history of Arabia as inferred by mtDNA variation. *American Journal of Physical Anthropology*, 159, 4, 607-616.

Černý, V., Fernandes, V., Costa, M.D., Hájek, M., Mulligan, C.J., & Pereira, L. (2009). Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evolutionary Biology*, 9, 63.

Černý, V., Pereira, L., Musilová, E., Kujanová, M., Vašíková, A., Blasi, P., ... Novelletto A. (2011). Genetic structure of pastoral and farmer populations in the African Sahel. *Molecular Biology and Evolution*, 28, 9, 2491-2500.

Černý, V, Salas, A, Hájek, M, Žaloudková, M, & Brdička, R. (2007). A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. *Annals of Human Genetics*, 71, 4, 433-452.

Čížková, M., Munclinger, P., Diallo, M.Y., Kulichová, I, Mokhtar, M.G., Dème, A., … Černý, V. (2017). Genetic Structure of the Western and Eastern African Sahel/Savannah Belt and the Role of Nomadic Pastoralists as Inferred from the Variation of D-Loop Mitochondrial DNA Sequences. *Human Biology*, 89, 4, 281–302.

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature*, 418, 6898, 700-707.

Diamond, J.M. (1997). *Guns, germs, and steel: the fates of human societies*. New York: W.W. Norton & Co.

Dunne, J., Evershed, R.P., Salque, M., Cramp, L., Bruni, S., Ryan, K., … di Lernia, S. (2012). First dairying in green Saharan Africa in the fifth millennium BC. *Nature*, 486, 7403, 390-394.

Dupire M. (1962). *Peuls nomades: Étude descriptive des Wodaabe du Sahel nigérien*. Paris: Karthala.

Eberhard, D.M., Simons, G.F., & Fennig, C.D. (2019). *Ethnologue: Languages of the World*. Dallas: Texas: SIL International.

Enattah, N.S., Jensen, T.G., Nielsen, M., Lewinski, R., Kuokkanen, M., Rasinpera, H., … Peltonen, L. (2008). Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *American Journal of Human Genetics*, 82, 1, 57-72.

Evans-Pritchard, E.E. (1940). *The Nuer. A description of the modes of livelihood and political institutions of a Nilotic people*. Oxford: Clarendon press.

Excoffier, L., & Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, 10, 3, 564-567.

Fernandes, V., Triska, P., Pereira, J.B., Alshamali, F., Rito, T., Machado, A., ... Pereira L. (2015). Genetic Stratigraphy of Key Demographic Events in Arabia. *PLoS One*, 10, 3, e0118625.

Gerbault, P. (2013). The onset of lactase persistence in Europe. *Human Heredity*, 76, 3-4, 154-161.

Gerbault, P., Liebert, A., Itan, Y. Powell, A., Currat, M., Burger, J. … Thomas, M.G. (2011). Evolution of lactase persistence: an example of human niche construction. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, 366, 1566, 863-877.

Goodrich, J.K., Davenport, E.R., Clark, A.G., & Ley, R.E. (2017). The Relationship Between the Human Genome and Microbiome Comes into View. *Annual Review of Genetics*, 51, 413-433.

Hall, T.A. (1999). *BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT/2K/XP/7*. Carlsbad: Ibis Biosciences.

Hampshire, K.R., & Smith M.T. (2001). Consanguineous marriage among the Fulani. *Human Biology*, 73, 4, 597-603.

Hassan, H.Y., van Erp, A., Jaeger, M., Tahir, H., Oosting, M., Joosten, L.A., & Netea, M.G. (2016). Genetic diversity of lactase persistence in East African populations. *BMC Research Notes*, 9, 1, 8.

Heyer, E., Brazier, L., Ségurel, L., Hegay, T., Austerlitz, F., Quintana-Murci, L., ... Veuille M. (2011). Lactase persistence in central Asia: phenotype, genotype, and evolution. *Human Biology*, 83, 3, 379-392.

Hodder, I. (1990). *The domestication of Europe: structure and contingency in Neolithic societies*. Oxford: Basil Blackwell.

Homewood, K. (2008). *Ecology of African pastoralist societies*. Oxford: James Currey.

Chiang, C.W.K., Marcus, J.H., Sidore, C., Biddanda, A., Al-Asadi, H., Zoledziewska, M., ... Novembre, J. (2018). Genomic history of the Sardinian population. *Nature Genetics*, 50, 10, 1426-1434.

Imtiaz, F., Savilahti, E., Sarnesto, A., Trabzuni, D., Al-Kahtani, K., Kagevi, I., … Jarvela, I. (2007). The T/G 13915 variant upstream of the lactase gene (LCT) is the founder allele of lactase persistence in an urban Saudi population. *Journal of Medical Genetics*, 44, 10, e89.

Itan, Y., Powell, A., Beaumont, M.A., Burger, J., & Thomas, M.G. (2009). The origins of lactase persistence in Europe. *PLOS Computational Biology*, 5, 8, e1000491.

Jesse, F., Keding, B., Lenssen-Erz, T., & Pöllath, N. (2013). 'I Hope Your Cattle Are Well' Archaeological Evidence for Early Cattle-centred Behaviour in the Eastern Sahara of Sudan and Chad. In M. Bollig, M. Schnegg & H.-P. Wotzka (Eds.), *Pastoralism in Africa: Past, Present and Future* (pp. 66-103). New York: Berghan Books.

Jones, B.L., Oljira, T., Liebert, A., Zmarz, P., Montalva, N., Tarekeyn, A., ... Swallow, D.M. (2015). Diversity of lactase persistence in African milk drinkers. *Human Genetics*, 134, 8, 917-925.

Jones, B.L., Raga, T.O., Liebert, A., Zmarz, P., Bekele, E., Danielsen, E.T., … Swallow, D.M. (2013). Diversity of lactase persistence alleles in Ethiopia: signature of a soft selective sweep. *American Journal of Human Genetics*, 93, 3, 538-544.

Kim, J., Hanotte, O., Mwai, O.A., Dessie, T., Bashir, S., Diallo, B., … Kim, H. (2017). The genome landscape of indigenous African cattle. *Genome Biology*, 18, 1, 34.

Kleisner, K., Pokorný, Š., Čížková, M., Froment, A., & Černý V. (2019). Nomadic pastoralists and sedentary farmers of the Sahel/Savannah Belt of Africa in the light of geometric morphometrics based on facial portraits. *American Journal of Physical Anthropology*, 169, 4, 632-645.

Krüttli, A., Bouwman, A., Akgül, G., Della Casa, P., Rühli, F., & Warinner C. (2014). Ancient DNA analysis reveals high frequency of European lactase persistence allele (T-13910) in medieval central Europe. *PLoS One*, 9, 1, e86251.

Kulichová, I., Fernandes, V., Deme, A., Nováčková, J., Stenzl, V., Novelletto, A., ... Černý, V. (2017). Internal diversification of non-Sub-Saharan haplogroups in Sahelian populations and the spread of pastoralism beyond the Sahara. *American Journal of Physical Anthropology*, 164, 2, 424-434.

Kuper, R., & Riemer, H. (2013). Herders before Pastoralism. Prehistoric Prelude in the Sahara. In M. Bollig, M. Schnegg & H.-P. Wotzka (Eds.), *Pastoralism in Africa: Past, Present and Future* (pp. 31-65), New York: Berghan Books.

Le Quellec, J.-L. (1997). *Art rupestre et préhistoire au Sahara*. Paris: Payot.

Le Quellec, J.-L. (2016). Recent Work on Saharan rock art (2010-2014). In P. Bahn, N. Franklin, M. Strecker, & E. Devlet (Eds.), *Rock Art Studies: News of the World V* (pp. 55-74), Oxford: Archaeopress Publishing.

Levy T.E., & Holl A.F.C. (2002). Migrations, ethnogenesis and settlement dynamics: Israelites in Iron Age Canaan and Shuwa-Arabs in the Chad Basin. *Journal of Anthropological Archaeology*, 21, 83-118.

Lhote, H. (1955). *Les Touaregs du Hoggar (Ahaggar)*. Paris: Payot.

Linseele, V. (2013). From the first stock keepers to specialised pastoralists in the West African Savannah. In M. Bollig, M. Schnegg & H.-P. Wotzka (Eds.), *Pastoralism in Africa: Past, Present and Future* (pp. 145-170). New York: Berghan Books.

Lokki, A.I., Jarvela, I., Israelsson, E., Maiga, B., Troye-Blomberg, M., Dolo, A., ... Holmberg, V. (2011). Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. *Malaria Journal*, 10, 1, 9.

Loncke, S. (2015). *Geerewol. Musique, danse et lien social chez les Peuls nomades woɗaaße du Niger*. Nanterre: Société d'ethnomusicologie.

Malmström, H., Linderholm, A., Liden, K., Storå, J., Molnar, P., Holmlund, G., … Götherström A. (2010). High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evolutionary Biology*, 10, 89.

Manfredi, S. (2012). Dialect Mixing and Dialect Levelling in Kordofanian Baggara Arabic (Western Sudan). In A. Barontini, C. Pereira, Á. Vicente, & K. Ziamari (Eds.), *Dynamiques langagières en Arabophonies : variations, contacts, migrations et créations artistiques* (pp. 141-162). Colección Estudios de Dialectología Árabe: Universidad de Zaragoza.

Marshall, F., & Hildebrand, E. (2002). Cattle before Crops: The Beginnings of Food Production in Africa. *Journal of World Prehistory*, 16, 2, 99-143.

Mathieson, S., & Mathieson, I. (2018). FADS1 and the Timing of Human Adaptation to Agriculture. *Molecular Biology and Evolution*, 35, 12, 2957-2970.

McCorriston, J., Harrower, M., Martin, L., & Oches, E. (2012). Cattle cults of the Arabian Neolithic and early territorial societies. *American Anthropologist*, 114, 1, 45-63.

McCorriston, J., & Martin, L. (2009). Southern Arabia's Early Pastoral Population History: Some Recent Evidence. In M.D. Petraglia, & J.I. Rose (Eds.), *The evolution of human populations in Arabia Paleoenvironments, prehistory and genetics* (pp. 237-250), New York: Springer.

Meloni, G.F., Colombo, C., La Vecchia, C., Pacifico, A., Tomasi, P., Ogana, A., ... Meloni, T. (2001). High prevalence of lactose absorbers in Northern Sardinian patients with type 1 and type 2 diabetes mellitus. *The American Journal of Clinical Nutrition*, 73, 3, 582-585.

Moran, P.A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37, 1-2, 17-23.

Mulcare, C.A., Weale, M.E., Jones, A.L., Connell, B., Zeitlyn, D., Tarekegn, A. … Thomas, M.G. (2004). The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *American Journal of Human Genetics*, 74, 6, 1102-1110.

Mwai, O., Hanotte, O., Kwon, Y.J., & Cho, S. (2015). African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. *Asian-Australasian journal of animal sciences*, 28, 7, 911-921.

Myles, S., Bouzekri, N., Haverfield, E., Cherkaoui, M., Dugoujon, J.M., & Ward, R. (2005). Genetic evidence in support of a shared Eurasian-North African dairying origin. *Human Genetics*, 117, 1, 34-42.

National Research Council. (2008). *Lost Crops of Africa*. Washington, DC: The National Academies Press.

Nováčková, J., Čížková, M., Mokhtar, M.G., Duda, P., Stenzl, V., Tříska, P., … Černý, V. (2020). Subsistence strategy was the main factor driving population differentiation in the bidirectional corridor of the African Sahel. *American Journal of Physical Anthropology*, 171, 6, 496–508.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner H. (2019). *vegan: Community Ecology Package. R package version 2.4-2*.

Ottoni, C., Martinez-Labarga, C., Loogvali, E.L., Pennarun, E., Achilli, A., De Angelis, F., … Rickards, O. (2009). First genetic insight into Libyan Tuaregs: a maternal perspective. Annals of Human Genetics, 73, 4, 438-448.

Ottoni, C., Primativo, G., Hooshiar Kashani, B., Achilli, A., Martinez-Labarga, C., Biondi, G., … Rickards, O. (2010). Mitochondrial haplogroup H1 in north Africa: an early Holocene arrival from Iberia. *PLoS One*, 5, 10, e13378.

Owens, J. (1994). *Arabs and Arabic in the Lake Chad region*. Köln: Rüdiger Köppe Verlag.

Ozainne, S. (2013). *Un Néolithique Ouest-Africain. Cadre chrono-culturel, économique et environnemental de l'Holocène récent en Pays dogon (Mali)*. Frankfurt am Main: Africa Magna.

Ozainne, S., Lespez, L., Garnier, A., Ballouche, A., Neumann, K., Pays, O., & Huysecom, E. (2014). A question of timing: spatio-temporal structure and mechanisms of early agriculture expansion in West Africa. *Journal of Archaeological Science*, 50, 359-368.

Ozainne, S., Mayor, A., & Huysecom, E. (2017). Chronology of human occupation during the Holocene in West Africa: The Dogon Country radiocarbon record. In N. Rupp, C. Beck, G. Franke, & K.P. Wendt (Eds.), *Winds of change* (pp. 211-227). Bonn: Rudolf Habelt Verlag.

Peakall, R., & Smouse, P.E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, 28, 19, 2537-2539.

Pedersen, J., & Benjaminsen, T.A. (2008). One Leg or Two? Food Security and Pastoralism in the Northern Sahel. *Human Ecology*, 36, 43–57.

Pereira, L., Černý, V., Cerezo, M., Silva, N.M., Hájek, M., Vašíková, A., ... Salas, A. (2010). Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. *European Journal of Human Genetics*, 18, 8, 915-923.

Podgorná, E., Diallo, I., Vangenot, C., Sanchez-Mazas, A., Sabbagh, A., Černý, V., & Poloni, E.S. (2015). Variation in NAT2 acetylation phenotypes is associated with differences in food-producing subsistence modes and ecoregions in Africa. *BMC Evolutionary Biology*, 15, 263.

Podgorná, E., Soares, P., Pereira, L., & Černý, V. (2013). The Genetic Impact of the Lake Chad Basin Population in North Africa as Documented by Mitochondrial Diversity and Internal Variation of the L3e5 Haplogroup. *Annals of Human Genetics*, 77, 513–523.

Priehodová, E, Abdelsawy, A, Heyer, E, & Černý, V. (2014). Lactase Persistence Variants in Arabia and in the African Arabs. *Human Biology*, 86, 1, 7–18.

Priehodová, E., Austerlitz, F., Čížková, M., Mokhtar, M.G., Poloni, E.S., & Černý, V. (2017). The historical spread of Arabian Pastoralists to the eastern African Sahel evidenced by the lactase persistence -13,915*G allele and mitochondrial DNA. *American Journal of Human Biology*, 29, 3.

Ranciaro, A., Campbell, M.C., Hirbo, J.B., Ko, W.Y., Froment, A., Anagnostou, P., … Tishkoff S.A. (2014). Genetic origins of lactase persistence and the spread of pastoralism in Africa. *American Journal of Human Genetics*, 94, 4, 496-510.

Reynolds, J., Weir, B.S., & Cockerham C.C. (1983). Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics*, 105, 3, 767-779.

Rosenberg, M. (2001). *PASSAGE, Pattern Analysis, Spatial Statistic, and Geographic Exegesis. Version I.I*. Tempe, AZ: Department of Biology, Arizona State University.

Sanchez-Mazas, A., Černý, V., Da, D., Buhler, S., Podgorná, E., Chevallier, E., ... Nunes J.M. (2017). The HLA-B landscape of Africa: Signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. *Molecular Ecology*, 26, 6238–6252.

Ségurel, L., & Bon, C. (2017). On the Evolution of Lactase Persistence in Humans. *Annual Review of Genomics and Human Genetics*, 18, 297-319.

Sheik-Mohamed, A., & Velema, J.P. (1999). Where health care has no access: the nomadic populations of sub-Saharan Africa. *Tropical Medicine and International Health*, 4, 10, 695–707.

Sherratt, A. (1981). Plough and pastoralism: aspects of the secondary products revolution. In I. Hodder, G. Isaac, & N. Hammond (Eds.), *Pattern of the Past: Studies in honour of David Clarke* (pp. 261–305). Cambridge: Cambridge University Press.

Shriner, D., & Rotimi, C.N. (2018). Genetic history of Chad. *American Journal of Physical Anthropology*, 167, 4, 804-812.

Schlebusch, C.M., & Jakobsson M. (2018). Tales of Human Migration, Admixture, and Selection in Africa. *Annual Review of Genomics and Human Genetics*, 19, 405-428.

Schlebusch, C.M., Sjodin, P., Skoglund, P., & Jakobsson M. (2013). Stronger signal of recent selection for lactase persistence in Maasai than in Europeans. *European Journal of Human Genetics*, 21, 5, 550-553.

Sikora, M., Carpenter, M.L., Moreno-Estrada, A., Henn, B.M., Underhill, P.A., Sanchez-Quinto, F., ... Bustamante, C.D. (2014). Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS Genetics*, 10, 5, e1004353.

Smith, A.B. (2005). *African herders: emergence of pastoral traditions*. Walnut Creek: AltaMira Press.

Sougnabe, P. (2011). *Pastoralisme sans terrain de parcours en savane tchadienne. Des Peul autour de la forêt classée de Yamba Berté*. Berlin: Univ Européenne.

Stephens, M, & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, 76, 3, 449-462.

Stephens, M., Smith, N.J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68, 4, 978-989.

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., … Deloukas P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, 39, 1, 31-40.

Torniainen, S., Parker, M.I., Holmberg, V., Lahtela, E., Dandara, C., & Jarvela I. (2009). Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. *BMC Genetics*, 10, 31.

Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., & Pereira, L. (2015). Extensive admixture and selective pressure across the Sahel Belt. *Genome biology and evolution*, 7, 12, 3484-3495.

Trombetta, B., D'Atanasio, E., Massaia, A., Ippoliti, M., Coppa, A., Candilio, F., ... Cruciani, F. (2015). Phylogeographic Refinement and Large Scale Genotyping of Human Y Chromosome Haplogroup E Provide New Insights into the Dispersal of Early Pastoralists in the African Continent. *Genome biology and evolution*, 7, 7, 1940-1950.

van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., ... Krause J. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science*, 360, 6388, 548-552.

Vicente, M., Priehodová E., Diallo I., Podgorná E., Poloni E.S., Černý V., & Schlebusch, C.M. (2019). Population history and genetic adaptation of the Fulani nomads: inferences from genome-wide data and the lactase persistence trait. *BMC Genomics*, 20, 1, 915.

Vyas, D.N., Kitchen, A., Miro-Herrans, A.T., Pearson, L.N., Al-Meeri, A., & Mulligan, C.J. (2016). Bayesian analyses of Yemeni mitochondrial genomes suggest multiple migration events with Africa and Western Eurasia. *American Journal of Physical Anthropology*, 159, 3, 382-393.

Wolfe, N.D., Dunavan, C.P., & Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, 447, 7142, 279-283.

Zeltner, J.C. (2002). *Histoire des Arabes sur les rives du lac Tchad*. Paris: Karthala.

Figure legends

Figure 1. Frequency of the different LP variants found among the examined populations. Positions of pie charts correspond to the approximate geographical coordinates of the populations. Star symbol indicates pastoral populations in the dataset. Numbers in pie charts are numeric codes referring to particular populations 1 Afar, 2 Beja, 3 Buduma, 4 Daju, 5 Dangaleat, 6 Fulani Abalak, 7 Fulani Ader, 8 Fulani Balatungur, 9 Fulani Banfora, 10 Fulani Bongor, 11 Fulani Diafarabe, 12 Fulani Diffa, 13 Fulani Timbi-Touni, 14 Fulani Ferlo, 15 Fulani Linia, 16 Fulani Mauritania, 17 Fulani Tcheboua, 18 Fulani Tindangou, 19 Fulani Zinder, 20 Fulani Ziniare, 21 Gurunsi, 22 Gourmantche, 23 Kababish, 24 Kotoko, 25 Masa, 26 Moors, 27 Mossi, 28 Tuareg Gorom-Gorom, 29 Tuareg Gossi, 30 Tuareg Tanut, 31 Zaghawa, 32 Afar, 33 Algerian, 34 Amharic, 35 Amizmiz, 36 Arabic Baggara, 37 Arabs, 38 Arabs Baggara, 39 Arabs Baggara, 40 Arabs Baggara, 41 Bani Amir, 42 Beja Banuamir, 43 Beja Hadandawa, 44 Beni Amer, 45 Berber, 46 Berber Moyen-Atlas, 47 Berber Mzab, 48 Beta Israel (Ethiopian Jews), 49 Boulala, 50 Burji, 51 Cameroonian, 52 Cameroonian, 53 Copts, 54 Darfurian, 55 Dinka, 56 Dinka, 57 Egyptian, 58 Ethiopian, 59 Ethiopian, 60 Ethiopian, 61 Ethiopian, 62 Ethiopian, 63 Fulani, 64 Fulani, 65 Fulani, 66 Gaali, 67 Gaalien, 68 Ghanaian, 69 Halfawien, 70 Hausa, 71 Jaali, 72 Konso, 73 Laka, 74 Mahas, 75 Mahas, 76 Mambila, 77 Mambila, 78 Mandara, 79 Manjak, 80 Mozabite Berbers, 81 Ngambay, 82 Nilotes, 83 Nso, 84 Nuba, 85 Nuer, 86 Nyimang, 87 Ouldoume, 88 Phenotyped Somali, 89 Rashaayda Arabs, 90 Shilluk, 91 Shokrya, 92 Shuwa Arab, 93 Shuwa Arab, 94 Shwaiga, 95 Somali, 96 Somali, 97 Sudanese, 98 Sudanese, 99 Wolof, 100 Yamba, and 101 Yoruba. For population details, see Supplementary Table S1.

Figure 2. Principal coordinate analysis (PCA) of LP mutation frequencies in Sahelian and neighboring populations; for population abbreviations, see Supplementary Table S1.

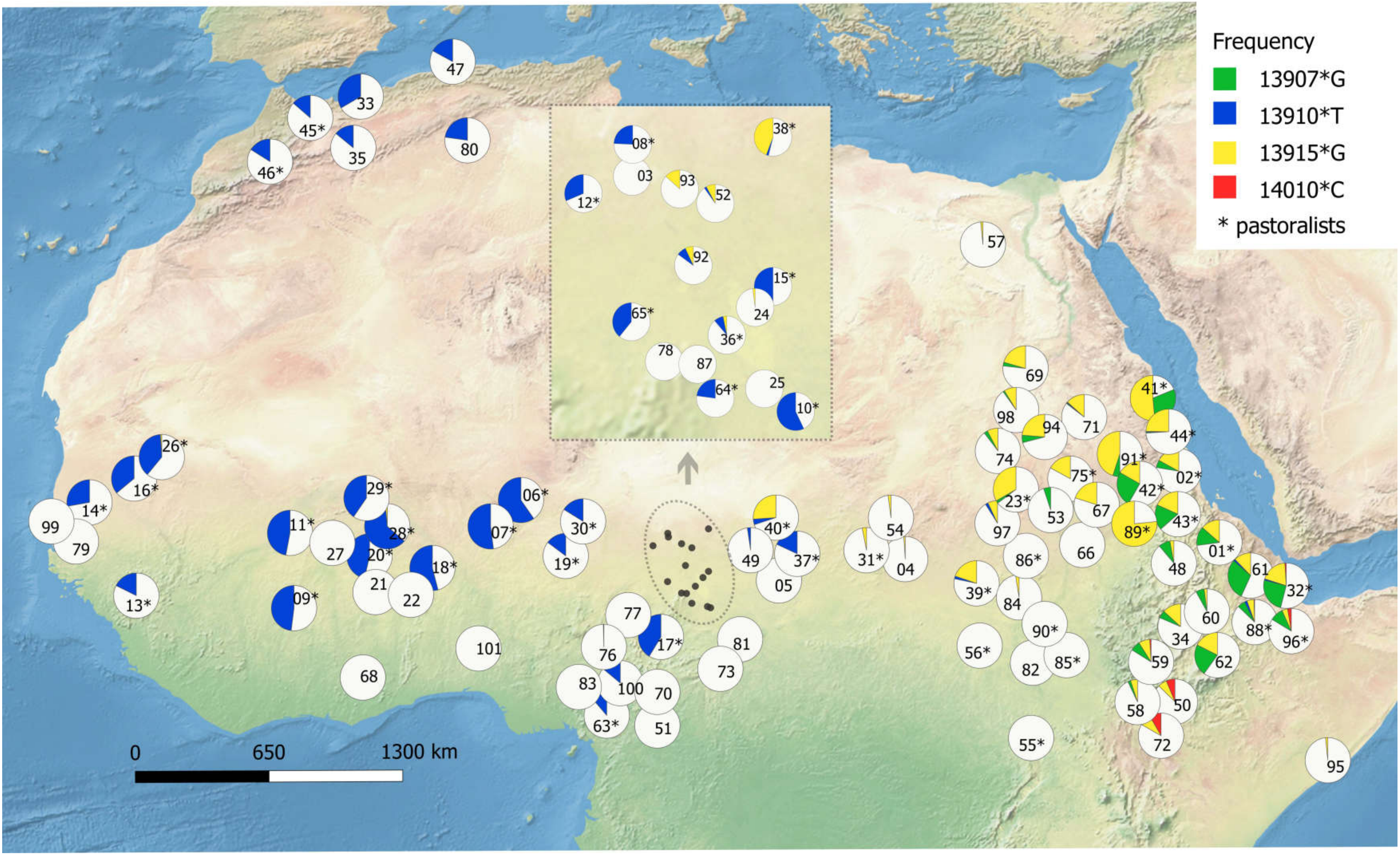
Figure 3. Interpolation maps of allelic frequencies of the main LP variants (–13910*T upper row, –13915*G middle row) between sedentary farmers (left column) and nomadic pastoralists (right

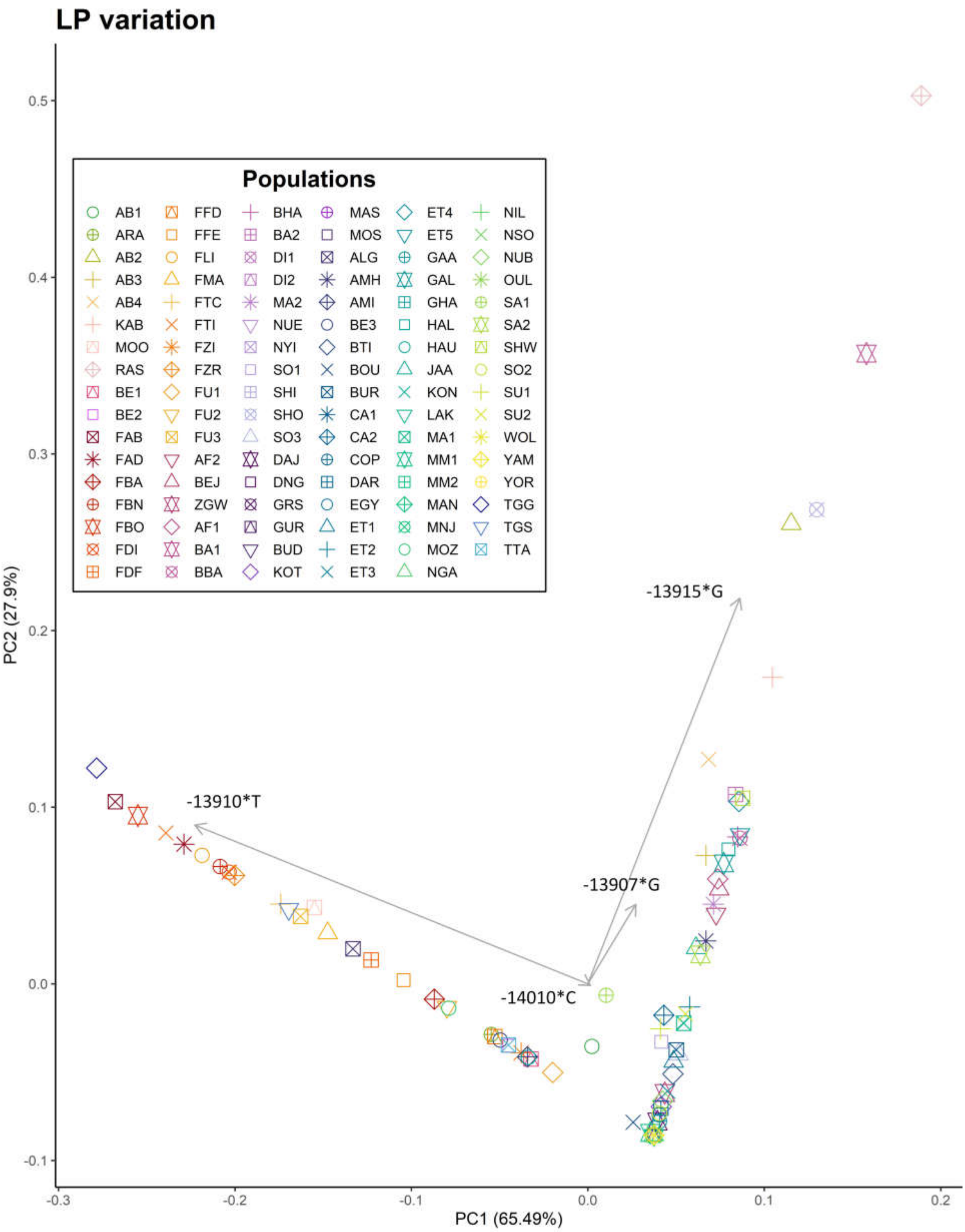
column) in the Sahel/Savannah belt and neighboring sites. The bottom row shows combined frequencies of all LP variants identified and collected so far.

Supplementary Figure S1. Multidimensional scaling plot showing the level of similarity among populations of the Sahel/Savannah belt based on LP variants.

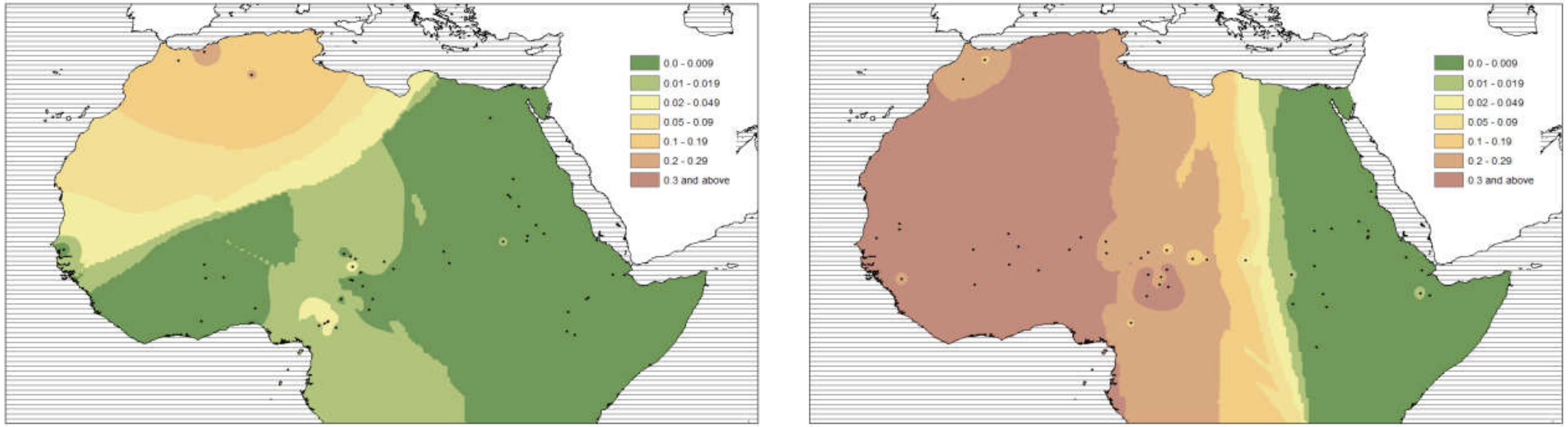
Supplementary Figure S2. Spatial autocorrelation analyses for –13910*T (top) and –13915*G (bottom). Filled points are statistically significant at 5% level.

Supplementary Figure S3. Seven SNPs network of –13910*T haplotypes in Fulani (top) and Tuareg (bottom) populations. Haplotypes are represented by circles filled with colors corresponding to the sites where individuals were sampled. Circle sizes are proportional to haplotype frequency in the sample.

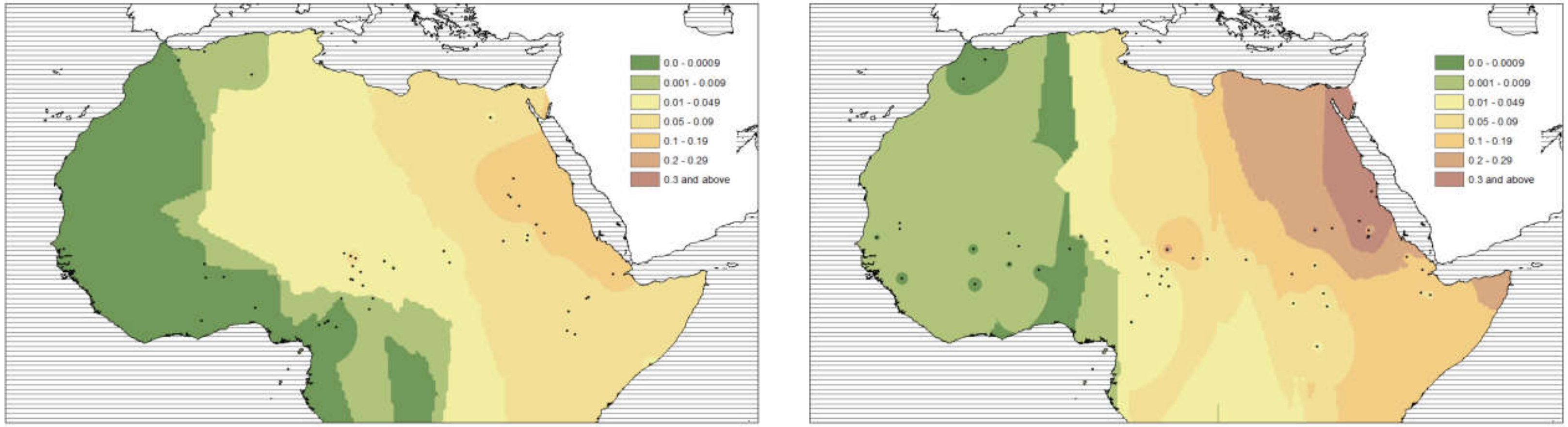

Frequency
13907*G
13910*T
13915*G
14010*C
* pastoralists
47
33
45*
35
80
46*
08*
38*
03
93
52
12*
57
92
15*
24
65*
36*
78
87
69
41*
25
64*
10*
98
94
71
44*
74
91*
02*
26*
16*
75*
42*
29*
06*
23*
14*
53
67
43*
89*
99
30*
40*
54
97
11*
28*
07*
79
27
20*
18*
19*
49
37*
31*
04
86*
66
01*
48
61
32*
21
05
39*
13*
22
09*
84
60
77
90*
88*
96*
34
17*
81
101
76
56*
59
62
68
73
85*
82
83
100
70
50
58
63*
51
72
55*
95
0
650
1300 km


LP variation
Populations
AB1
ARA
AB2
AB3
AB4
KAB
MOO
RAS
BE1
BE2
FAB
FAD
FBA
FBN
FBO
FDI
FDF
FFD
FFE
FLI
FMA
FTC
FTI
FZI
FZR
FU1
FU2
FU3
AF2
BEJ
ZGW
AF1
BA1
BBA
BHA
BA2
DI1
DI2
MA2
NUE
NYI
SO1
SHI
SHO
SO3
DAJ
DNG
GRS
GUR
BUD
KOT
MAS
MOS
ALG
AMH
AMI
BE3
BTI
BOU
BUR
CA1
CA2
COP
DAR
EGY
ET1
ET2
ET3
ET4
ET5
GAA
GAL
GHA
HAL
HAU
JAA
KON
LAK
MA1
MM1
MM2
MAN
MNJ
MOZ
NGA
NIL
NSO
NUB
OUL
SA1
SA2
SHW
SO2
SU1
SU2
WOL
YAM
YOR
TGG
TGS
TTA
-13915*G
-13910*T
-13907*G
-14010*C
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
PC2 (27.9%)
-0.3
-0.2
-0.1
0.0
0.1
0.2
PC1 (65.49%)

-13910*T



-13915*G



LP

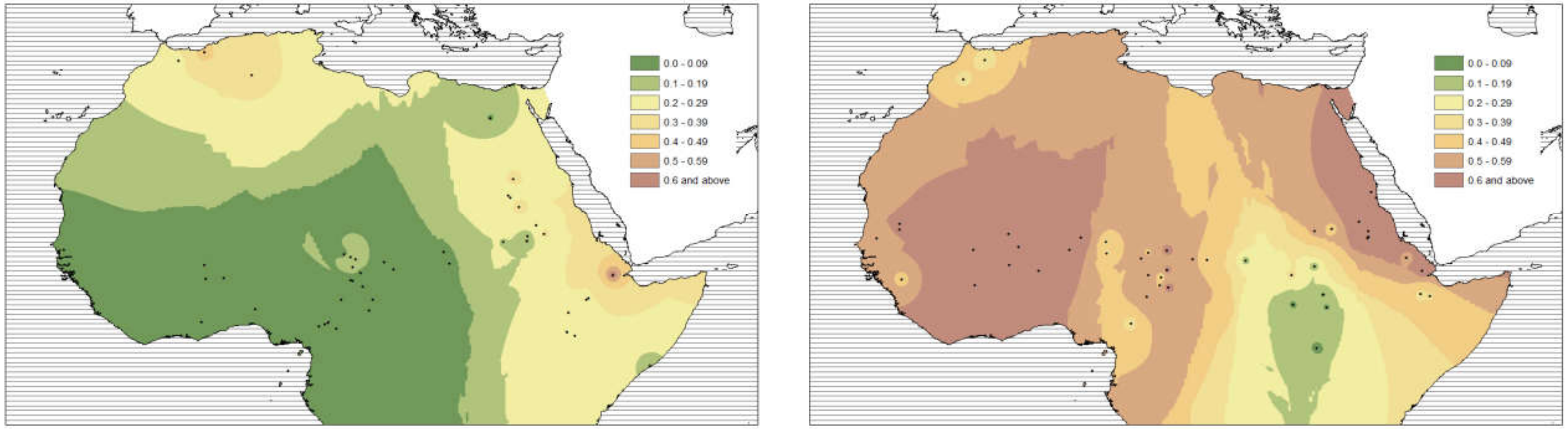


Table 1. Population samples used in the study

Population	n	Sampling location	Country of origin	Family	Lifestyle
Afar	26	Paris, Brussels	Djibouti, Ethiopia, Eritrea	Afro-Asiatic	pastoralists
Beja	47	Mackali	Sudan	Afro-Asiatic	pastoralists
Buduma	30	around Balatungur	Niger	Afro-Asiatic	farmers
Daju	50	Daju	Sudan	Nilo-Saharan	farmers
Dangaléat	49	west of Mongo	Chad	Afro-Asiatic	farmers
Fulani Abalak	26	around Abalak	Niger	Niger-Congo	pastoralists
Fulani Ader	49	around Ader	Niger	Niger-Congo	pastoralists
Fulani Balatungur	23	around Balatungur	Niger	Afro-Asiatic	pastoralists
Fulani Banfora	49	around Banfora	Burkina Faso	Niger-Congo	pastoralists
Fulani Bongor	49	east of Bongor	Chad	Niger-Congo	pastoralists
Fulani Diafarabe	50	Diafarabe	Mali	Niger-Congo	pastoralists
Fulani Diffa	32	around Diffa	Niger	Niger-Congo	pastoralists
Fulani Timbi-Touni	51	Timbi-Touni	Guinea	Niger-Congo	pastoralists
Fulani Ferlo	56	near Linguere	Senegal	Niger-Congo	pastoralists
Fulani Linia	27	around Linia	Chad	Niger-Congo	pastoralists
Fulani Mauretania	47	Brakna, Gorgol, Assaba	Mauretania	Niger-Congo	pastoralists
Fulani Tcheboua	40	Tcheboua	Cameroon	Niger-Congo	pastoralists
Fulani Tindangou	50	around Tindangou	Burkina Faso	Niger-Congo	pastoralists
Fulani Zinder	34	around Zinder	Niger	Niger-Congo	pastoralists
Fulani Ziniare	56	northeast of Ziniare	Burkina Faso	Niger-Congo	pastoralists
Gurunsi	32	Pô	Burkina Faso	Niger-Congo	farmers
Gourmantche	37	Nnear Pama	Burkina Faso	Niger-Congo	farmers
Kababish	27	Al-Obeid	Sudan	Afro-Asiatic	pastoralists
Kotoko	47	Holouf and around	Cameroon	Afro-Asiatic	farmers
Masa	35	Yagoua and around	Cameroon	Afro-Asiatic	farmers
Moors	46	Brakna, Gorgol, Assaba	Mauretania	Afro-Asiatic	pastoralists
Mossi	34	east of Ziniare	Burkina Faso	Niger-Congo	farmers
Tuareg 1	37	around Gorom-Gorom	Burkina Faso	Afro-Asiatic	pastoralists
Tuareg 2	26	around Gossi	Mali	Afro-Asiatic	pastoralists
Tuareg 3	31	around Tanut	Niger	Afro-Asiatic	pastoralists
Zaghawa	48	Al-Obeid	Sudan	Nilo-Saharan	pastoralists

Table 2. Counts (in parentheses) and allelic frequencies of mutations found in the analyzed segment of MCM6 gene; LP variants in bold

Population	**–13907*G**	**–13910*T**	**–13915*G**	**–14009*G**	–13913*C	–13915*C	–13980*A	–14107*A	–14156*A	–14159*G
Afar	(7) 0.135		(7) 0.135					(3) 0.058		
Beja	(5) 0.053		(16) 0.170	(40) 0.426						
Buduma								(3) 0.050		
Daju			(1) 0.01							
Dangaléat										
Fulani Abalak		(31) 0.596						(1) 0.019		
Fulani Ader		(51) 0.520								
Fulani Balatungur		(25) 0.543								
Fulani Banfora		(47) 0.480								
Fulani Bongor		(56) 0.571			(1) 0.010			(1) 0.010		
Fulani Diafarabe		(47) 0.470				(2) 0.020				
Fulani Diffa		(20) 0.313								
Fulani Timbi-Touni		(18) 0.176					(1) 0.010	(2) 0.020		
Fulani Ferlo		(31) 0.277						(1) 0.009		
Fulani Linia		(27) 0.500			(2) 0.037			(1) 0.019		
Fulani Mauretania		(34) 0.361								
Fulani Tcheboua		(33) 0.413			(2) 0.025			(2) 0.025		
Fulani Tindangou		(54) 0.540								
Fulani Zinder		(10) 0.147								
Fulani Ziniare		(52) 0.464			(1) 0.009					
Gurunsi								(3) 0.047		
Gourmantche									(1) 0.014	
Kababish	(1) 0.019		(18) 0.333	(6) 0.111						
Kotoko			(2) 0.021					(2) 0.021	(1) 0.011	
Masa								(3) 0.043		(1) 0.029
Moors		(35) 0.380	(1) 0.011							
Mossi										
Tuareg 1		(46) 0.622	(1) 0.014		(2) 0.027					
Tuareg 2		(21) 0.404								
Tuareg 3		(10) 0.161								
Zaghawa			(3) 0.031		(1) 0.010					

Table 3. Estimation of growth rate and age of expansion of the –13910*T allele in Fulani and Tuareg population

Population	Size	Allelic frequency	Number of chromosomes carrying the majoritary haplotypes	Total number of haplotypes	Recombination rate	Estimated growth rate (with CI)	Estimated age of the mutation in generations (with CI)	Estimated age of the mutation in years (with CI)
Fulani	10,000,00	0.455	187	260	0.00129807	1.038 (1.033, 1.048)	340.1 (305.6, 387.4)	8503.5 (7639.0, 9685.8)
Fulani	30,000,00	0.455	187	260	0.00129807	1.043 (1.038, 1.053)	333.3 (301.4, 378.5)	8333.5 (7534.4, 9461.4)
Tuareq	1,300,000	0.433	36	39	0.00129807	1.19 (1.16, 1.25)	70.6 (61.6, 88.0)	1765.7 (1540.6, 2200.4)