View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Digital Commons@Wayne State University

DIGITALCOMMONS @WAYNESTATE

Wayne State University

Human Biology Open Access Pre-Prints

WSU Press

6-5-2018

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 mtDNA sequences

Martina Čížková
Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic, martina.ciz@email.cz

Pavel Munclinger
Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic, pavel.munclinger@natur.cuni.cz

Mame Yoro Diallo
Département d'Histoire, Faculté des Lettres et Sciences humaines, Université Cheikh Anta Diop de Dakar, Senegal, yorodiallo90@yahoo.fr

Iva Kulichová
Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic, iwakulichova@gmail.com

Mohammed Gassim Mokhtar
Arabic Department, Faculty of Arts, University of Kordofan, Al-Ubayyid, Sudan, mgmokhtar@yahoo.com

Recommended Citation

Čížková, Martina; Munclinger, Pavel; Diallo, Mame Yoro; Kulichová, Iva; Mokhtar, Mohammed Gassim; Deme, Alioune; Pereira, Luísa; and Černý, Viktor, "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 mtDNA sequences" (2018). *Human Biology Open Access Pre-Prints*. 129.
https://digitalcommons.wayne.edu/humbiol_preprints/129

This Article is brought to you for free and open access by the WSU Press at DigitalCommons@WayneState. It has been accepted for inclusion in Human Biology Open Access Pre-Prints by an authorized administrator of DigitalCommons@WayneState.

See next page for additional authors

Authors
Martina Čížková, Pavel Munclinger, Mame Yoro Diallo, Iva Kulichová, Mohammed Gassim Mokhtar, Alioune Deme, Luísa Pereira, and Viktor Černý

This article is available at DigitalCommons@WayneState: https://digitalcommons.wayne.edu/humbiol_preprints/129

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 mtDNA Sequences

Martina Čížková,[1] Pavel Munclinger,[2] Mame Yoro Diallo,[3] Iva Kulichová,[1] Mohammed G. Mokhtar,[4] Alioune Dème,[3] Luísa Pereira,[5,6,7] and Viktor Černý[8]*

[1]Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic.

[2]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic.

[3]Département d'Histoire, Faculté des Lettres et Sciences humaines, Université Cheikh Anta Diop de Dakar, Dakar, Senegal.

[4]Arabic Department, Faculty of Arts, University of Kordofan, Al-Ubayyid, Sudan.

[5]Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto, Portugal.

[6]Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal.

[7]Faculdade de Medicina da Universidade do Porto, Porto, Portugal.

[8]Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague, Czech Republic.

*Correspondence to: Viktor Černý, Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Letenska 4, 118 01 Prague, Czech Republic. E-mail: cerny@arup.cas.cz.

Short Title: Genetic Structure of the Western and Eastern African Sahel

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

KEY WORDS: FULANI, ARABS, MTDNA, AFRICAN PASTORALISM, SAHEL.

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Abstract The objective of this study is to provide deeper knowledge of the maternal genetic structure and demographic history of the human population of the dynamic Sahel/Savannah belt, the extensive region lying between the Sahara and tropical rainforests, spanning from the Atlantic Ocean to the Red Sea coast. The study aims to confirm or disconfirm archaeological and linguistic data indicating that the region's populations underwent diversification as a result of the spread of agropastoral food-producing subsistence lifestyles, over time dividing the region into separate areas of nomadic pastoralism on the one hand, and sedentary farming on the other. In order to perform both descriptive and coalescence analyses from the Sahel/Savannah belt's entire region, including western and eastern rather than just central populations studied previously, we generated a new mtDNA dataset not only having almost 2,000 samples (875 of which were newly collected); but also encompassing whole mtDNA D-loop segment rather than only the previously studied HVS-1. While comparing our analyses with previous results from the Lake Chad Basin (central Sahel/Savannah Belt) we revealed similar intra-population diversity measures (i.e., lower values of measures in pastoralists than farmers). However, the new dataset pointed to significant differences in mating strategies between western as compared to the eastern pastoralists: our results suggest higher gene flow between the Arabic pastoralists and neighboring farmers in the eastern than between the Fulani pastoralists and their sedentary neighbors in the western part of the Sahel/Savannah Belt. The findings are discussed in the light of archaeological and linguistic data, allowing us to postulate that the genetic differentiation of Fulani pastoralists from the common western African agropastoral gene pool occurred at around the same time as the arrival of the Arabic pastoralists to eastern Africa. However, it seems that while the process of divergence of the Fulani pastoralists in the west was accompanied by a loss of Fulani females to other populations, the Arab

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

pastoralists' immigration to the Sahel/Savannah belt conversely resulted in some gain of local females into this Arab population.

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

The African Sahel/Savannah belt (further referred to as the SSB) is a unique biome lying between the Sahara in the north and tropical rainforests to the south, and stretching from the Atlantic Ocean to the Red Sea coast. Populations living in the region can be classified into groups according to geography, language and subsistence lifestyle, including two distinct sympatric food-producing populations that cohabit across the whole length of the SSB: nomadic pastoralists and sedentary farmers. Lying in the middle of the SSB is the Lake Chad Basin, an area of remarkable natural and cultural diversity geographically dividing the whole region into western and eastern parts. While the pastoralists constitute rather large groups such as the Fulani in the western part, and Arabs in the east, there is also a large number of small geographically dispersed sedentary farmers with only a few more numerous groups such as the Kanembu, Kanuri, Hausa, and Mandinka, whose history is linked with great African empires (Connah 2001; Newman 1995). Last but not least, SSB populations can be classified according to linguistic affiliations as belonging to one of the Niger-Congo, Nilo-Saharan, or Afro-Asiatic language families (Blench 2006; Ehret 2002; Heine and Nurse 2000).

Important information on the population history of the region can also be gained from archaeology. Relicts of harpoons fashioned of bone and other items such as decorated wavy line or dotted wavy line ceramics and later-dated fish hooks made from shells or bones have been found in the area ranging from the middle Nile and the East African lakes to the western part of the Sahara. Such findings have led to the formulation of the theory of the Aquatic Civilization (Sutton 1974). According to this view, a common culture was spread by ancestors of today's Nilo-Saharans some 10,000 years ago, when the last African Humid Period caused the creation of the so-called "Green Sahara", which included water formations such as rivers, inland deltas, and large lakes such as Mega-Chad (Bouchette et al. 2017; Kuper and Kröpelin 2006; Skonieczny et al. 2015). Today, Nilo-Saharan populations extend

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

from the eastern to the central SSB; its westernmost population is found around the Inner Niger Delta and speaks Songhai.

After these humid conditions in the area began to dry up some 5,000 years ago, it is likely that the transformation to an agropastoral lifestyle was led by Afro-Asiatic peoples (Blench 1999; Blench 2006). They caused fragmentation of the regions' original continuum of Nilo-Saharan hunter-gatherers as some of their groups gradually also adopted the Afro-Asiatic example of food production strategies. Although associating archaeologically-evidenced ancient cultures with contemporary languages is a controversial, it is hard to overlook that Nilo-Saharan speakers still tend to systematically live near bodies of water, including the ancient northern shores of the Mega-Chad paleo-lake, and in locations where bone harpoons have been found (Drake et al. 2011). Further evidence shows that the original agropastoral subsistence pattern began gradually diverging some 4,000 year ago into two different but perhaps mutually dependent lifestyles: nomadic pastoralism based on the continuous movement (transhumance) of peoples and their domesticated animals for one, and sedentary farming for the other (Linseele 2013; MacIntosh 2005).

In western Africa, these cultural developments were accompanied by further diversification of the Niger-Congo languages that had already started to expand millennia earlier, at least some 8,000 years ago (Blench 2006). Today, pastoralism is practiced in the western part of the SSB by the Fulani, who speak a language belonging to this Niger-Congo family - this branch of languages expanded to West Africa in the Early Holocene, however, a long time before the diversification of the above-mentioned food producing lifestyles (Blench 2006). Archaeologically, the origin of the Fulani is believed to be the central Sahara area (Dupuy 1999). Some scholars have even suggested that certain central Saharan prehistoric paintings might have been created by Fulani ancestors (Ba and Dieterlen 1966); this last speculation, however, has been, highly criticized (Le Quellec 2004). Historical accounts

indicate that the Fulani expansion started around the 11th century CE in Fouta Djallon, Guinea, continuing eastward to the Lake Chad Basin, where it was impeded by Arabian pastoralists who had spread to the eastern part of the SSB only shortly earlier (Owens 1994). Today, only few Fulani groups can be found as far east as the Blue Nile area in Sudan (Delmet 2000). It should also be emphasized that many Fulani groups adopted farming practices in various parts of western Africa, while others (called *M'Bororo* or *Woɗaaɓe*) remained faithful to the nomadic lifestyle of their ancestors (Botte et al. 1999)

Pastoralism in the eastern part of the SSB is kept up today mainly by Arab groups of various names including the Shuwa, Baggara, and Abbala. It is believed that their ancestors started to expand from Arabia to Africa in the 7th century CE in what is known as the Islamization of Africa (Insoll 2003; MacMichael 1922), arriving to the Lake Chad Basin approximately five hundred years later (Levy and Holl 2002; Zeltner 2002) - hence, as mentioned above, slightly earlier than the Fulani. In addition, migrations of various tribes from the Arabian Peninsula to Africa continued until recently, a case in point being the Rashaayda Bedouins, whose ancestors reached the Red Sea coast in Sudan in the 1860s (Young 1996).

To-date, studies examining human SSB population genetics have collected samples mainly in its central part of the SSB (in other words, the Lake Chad Basin area). Based on uniparental (mitochondrial and the Y chromosome loci, henceforth referred to as mtDNA and NRY, respectively) and some biparental genetic systems, such research seems to indicate that the SSB constituted a "crossroads" of migrations to the region from both eastern and western Africa, and that contacts occurred predominantly among pastoralists (Cerezo et al. 2011; Černý et al. 2007; Triska et al. 2015). Analysis of uniparental genetic systems also showed that the majority of the Fulani lineages belong to the western African gene pool; a few of them, however, pointed to a northern African or even western Eurasian origin (Bučková et al.

2013; Černý et al. 2006; Kulichová et al. 2017). Strikingly, genetic similarities among Fulani sub-populations sampled thousands of km apart demonstrated a common origin and/or homogenizing effect of the gene flow among their local demes (Černý et al. 2011; Čížková et al. 2017).

These studies also show that although some earlier gene flows from Arabia to Africa (and vice-versa) can be inferred by phylogeographic (Fernandes et al. 2015; Musilová et al. 2011) and population-based (Černý et al. 2016) studies, the bulk of the genetic input to the SSB occurred as a result of Arab immigration from the 7th century CE onwards. It seems that admixture during this time was relatively common; as far as the maternal gene pool is concerned, the highest admixture rate for an Arab group with sub-Saharan Africans appears to be in the Shuwa Arabs from north-eastern Nigeria and northern Cameroon (Černý et al. 2007). Further analyses of African Arabs such as the Aballa, Baggara, and Rashaayda yielded a negative correlation between the frequency of the "Arabian" lactase persistence –13,915*G allele and the frequency of sub-Saharan mtDNA, a phenomenon that can be explained by the progressive introduction of sub-Saharan females into the Arabian gene pool (Priehodová et al. 2017).

To expand on these previous results, in this study we further explore the genetic structure of the SSB by separately analyzing population differentiations and gene flows between nomadic pastoralists and sedentary farmers in western and eastern parts of the region. The main questions are whether the demographic reconstructions suggested by the archaeological and linguistic data summarized above are indeed reflected in the genetic make-up of contemporary populations, and whether the observations previously made in the central SSB can be reproduced in the western and eastern areas. Thus, we generated an extensive mtDNA D-loop dataset covering not only the Fulani pastoralists and sedentary farmers in the Lake Chad basin (as analyzed previously, Černý et al. 2011), but extending

across a much larger geographic area. Generally speaking, we discovered a low genetic structure, a finding suggesting a high migration rate and/or common genetic origin of the populations across the entire region. Moreover, our data provides the first evidence of differential mating strategies in the Fulani as compared to the Arabic pastoralists. In both cases, however, pastoralism can be considered a leading factor in population differentiation and/or admixture, remarkably shaping the genetic structure of the whole SSB area's population.

Material and Methods

Subjects and Laboratory Methods. In this study, we analyzed large mtDNA datasets containing 1,994 D-loop mtDNA sequences of nomadic pastoralists and sedentary farmers belonging to three linguistic groups across the SSB and Northeast Africa. Nomadic pastoralists in the dataset are represented by 10 populations, and sedentary farmers by 26 populations. Pastoralists in the western part of the SSB are represented both by Fulani groups still actively practicing the nomadic lifestyle (Ferlo and Ziniare), and others that have a pastoral history but have recently settled down and are now sedentary (Fouta Djallon and Halpularen). Pastoralists from the eastern part of the SSB are represented by both groups of indigenous African peoples such as the Daza, Beja, and Datog on one hand, and Arabs whose historically-documented origin can be traced to Arabian Peninsula on the other. Complete information about the populations used in this study is provided in Supplementary Table 1, with their geographic locations graphically visualized in Figure 1.

Sampling authorizations were secured from the respective state institutions of the countries involved, appropriate informed consent was procured before sample collection from all participants, and ethical approval for the project was obtained from the Charles University in Prague (no 2016/07). Samples were collected using the Oragene DNA Collection kit, and

DNA was extracted according to the protocol provided by the supplier (DNA Genotek). All newly sequenced samples (n = 875) of this dataset were sequenced using primer pairs P23 and P24 (as in Gonder et al. 2007). The mtDNA D-loop sequences were submitted to GenBank (accession numbers MH122955–MH123829). Total length is 1,227 bp (nps 15888–579). All obtained sequences were compared to the revised Cambridge Reference Sequence (rCRS) (Andrews et al. 1999), and mutations were scored.

For comparative purposes, we also included published datasets of 17 Sahelian populations (1,119 sequences) for which the whole D-loop region was available (see Supplementary Table 1). However, due to differences in the lengths of the sequences published in those studies compared to the ones generated in this study, their analyses were restricted to the segment 1,087 bp (nps 16030–579).

Data Analyses. To evaluate genetic diversity at the population level, we computed both the standard and molecular diversity indices, namely gene/haplotype (h) and nucleotide diversity (π). We estimated demographic events using Tajima's D and Fu's F_s tests of selective neutrality (Fu 1997; Tajima 1989), and we also tested goodness-of-fit indices, namely the sum of squared deviations (SSD) and Harpending's raggedness index (RI) comparing observed and expected mismatch distribution by applying both models of population expansion (demographic and spatial) and using a parametric bootstrap approach of 500 replicates (Excoffier et al. 2005; Excoffier and Schneider 1999; Harpending et al. 1993; Ray et al. 2003).

For inter-population comparisons, pairwise Reynolds genetic distances based on haplotype frequencies (Reynolds et al. 1983; Slatkin 1995) were calculated, and their significance was tested with the permutation procedure implemented in Arlequin software using 10,000 iterations (Excoffier and Lischer 2010). Reynolds genetic distances of mtDNA sequences were weighted using evolutionary distance between haplotypes (Φ_{ST} indices)

(Excoffier et al. 1992) and the Kimura-2P model with a Gamma correction of 0.4, a transition/transversion ratio of 10/1, while leaving indels out of consideration, as recommended in a similar study (Poloni et al. 2009).

The levels of genetic differentiation between or among different groups of populations (divided according to subsistence lifestyle, geographical location, and language) were assessed through analyses of molecular variance (AMOVA). We used a hierarchical framework and significance of the fixation indices was tested by 10,000 iterations of the permutation procedure implemented in Arlequin ver. 3.5.2.2 (Excoffier et al. 2005). We calculated also shared haplotypes between defined groups (based either on lifestyle and/or language). The coancestry coefficients of the values of Reynolds genetic distances were used to perform MDS (Multidimensional Scaling Analysis) in RStudio (Team 2016).

We also performed the Mantel test implemented in GenAlEx (Peakall and Smouse 2012; Peakall and Smouse 2006) with a setting of 10,000 permutations to investigate the significance of correlation coefficients between geographic and genetic distances. We further focused more deeply on relationships between genetic and geographic aspects, and conducted Spatial Analysis of Molecular Variance (SAMOVA), using SAMOVA 2.0 software, a program that defines the genetic structure of populations by a simulated annealing approach (Dupanloup et al. 2002).

Last but not least, estimation of population parameters was carried out by Migrate-n ver. 3.6.11 using a Bayesian MCMC (Markov chain Monte Carlo) inference model (Beerli 2009). The estimated parameters were Θ (which is $N_f\mu$, where N_f = effective female population size, f and μ = mtDNA mutation rate) and the migration rate M (which is m/μ, where m is the chance for a lineage to immigrate per generation, and μ is the mutation rate per site per generation; an M of 1 means that it is as likely for the sequence to migrate as it is

for a site on the sequence to mutate) with settings of 1,000,000 genealogies with every 5,000 recorded, and a burn-in of 10,000.

Results

Of the 1,994 D-loop sequences in our dataset of various African populations living across the SSB, we defined 1,352 (67.8%) unique haplotypes. Relatively high gene/haplotype diversity (h) values were observed in all populations (mean value = 0.988) with the exception of in two pastoralist groups – the Beja in Sudan (0.725) followed by the Fulani from Ferlo in Senegal (0.840). Both these populations also rank among the lowest values in nucleotide diversity (π) (Beja = 0.0099; Fulani Ferlo = 0.0095), as do several western sedentary populations (Gurunsi = 0.0096; South Samo = 0.0099; North Samo = 0.0094; Bisa = 0.0097). The complete results of intra-population diversity measures are presented in Supplementary Table 2. Plots of gene and nucleotide diversity can be seen in Supplementary Figure 1, clearly demonstrating that lower mean values (with higher confidence intervals/higher variance) of these indices are observed in the nomadic pastoralists. However, differences in average haplotype and nucleotide diversities among geographically specified subsistence modes were tested with pairwise Wilcoxon rank sum test and the only significant difference was found between nucleotide diversities of western sedentary and eastern nomadic people (Supplementary Table 3).

To search for a signal of demographic expansion, the data was analyzed using Tajima's D and Fu's F_s tests of selective neutrality. Although the Tajima's D test applied in our dataset yielded negative values that would indicate an excess of low frequency polymorphisms (or an excess of recent variants) – and thus an increase in population size and/or a selective sweep (Bamshad and Wooding 2003) – the vast majority of values was not significantly different from zero, meaning the null hypothesis cannot be rejected. Of the 39

groups studied, the analysis detected significant values in just two sedentary populations: the Somali and the Sandawe. When carrying out the more sensitive Fu's F_s test, on the other hand, all but one sedentary group show significant negative values (i.e., excepting the Nuba Koalib), while only half of the pastoral populations do so; this result is consistent with demographic expansion or genetic hitchhiking (Fu 1997).

Changes in population size were also examined by means of pairwise distribution analyses. We calculated the Harpending's raggedness index (RI) and the sum of squared deviation (SSD) between the observed and expected mismatch for each of the populations using a spatial and demographic expansion model (Excoffier and Schneider 1999; Harpending et al. 1993). Results were not statistically significant, thus also pointing to both a spatial and demographic expansion model in all but two pastoralist groups having the highest overall (and significant) demographic indices, namely the Beja and Rashaayda. The results are listed in Supplementary Table 4.

Although the pairwise distribution analyses results do not seem to distinguish between populations of different subsistence strategies very well on first sight, the spatial expansion model also provided us an M parameter, which is a joint estimation of the effective population size (N_e) and migration rate (m). The first impression of the M parameter results is that M is higher on average in nomadic than in sedentary populations. On closer examination, we observed low M values not only in pastoralists, but also farmers such as the Daza (M = 17.18), Rashaayda (M = 11.06), Sandawe (M = 16.64), and Luhya (M = 12.32); it is important to note that extremely low values occurred only in the pastoralists such as the Beja (M = 1.54), Fulani from Ferlo (M = 3.18) and Fulani from Ziniare (M = 3.56) suggesting their genetic isolation and/or low N_e. From the inverse perspective, it seems that the Bisa is an outlier among western SSB sedentary populations. One clearly apparent distinction can be seen between contemporary and past pastoralists in the western part of the SSB, where the

past nomads show higher values of this measure. However, when testing the overall difference with Wilcoxon rank sum test, none of the average differences in lifestyles were significant. Complete results are shown in Supplementary Table 3 and 4.

Reynolds genetic distances did not reveal any essential population structure. Nonetheless, at first sight the plot displays the eastern pastoralists such as the Beja and Rashaayda as the most striking outliers, and indicates that eastern and western pastoralists are separated by a cluster of sedentary populations that communicate with eastern Africans via Arabian groups. Complete Reynolds genetic distance values along with the corresponding p-values are reported in Supplementary Table 5. A categorized and color-coded MDS plot with stress value 0.169 was also created in accordance with each population's lifestyle, language and region affiliations (Figure 2a-c). The result suggests that there seems to be a better correspondence between a population's region and its position in the genetic plots than with the other two criteria (language and lifestyle).

Levels of genetic differentiation among the studied populations were also assessed by an AMOVA within all groups, as well as using a hierarchical framework, based on regional grouping (western vs. eastern), language classification (Afro-Asiatic vs. Nilo-Saharan vs. Niger-Congo), subsistence pattern (nomadic vs. sedentary) and combinations thereof (see Table 1). Results show that in general, given the non-significant Φ_{CT} values, it can be stated that there is no significant differentiation between nomadic and sedentary populations. Results show, however, that the variance among nomadic populations (Φ_{ST} = 0.133 and 0.163) is almost twice higher than that among sedentary ones (Φ_{ST} = 0.073 and 0.085), which could indicate a higher level of drift or isolation among pastoralists. This overall difference between subsistence lifestyles was also confirmed when testing variance among subsistence groups regionally in the east, but examining the data showed negative and non-significant Φ_{CT} values indicating insignificant variation in genetic structure in the western SSB.

Moreover, the variance between eastern nomads and farmers is higher than that within either group, indicating strong differentiation by subsistence lifestyle in the eastern SSB. Besides lifestyle, we subjected the dataset to grouping by the language. We found the highest variance ($\Phi_{SC} = 0.109$) among populations within groups and no genetic structure between groups defined by these criteria jointly. Graphical representation of all the groupings and indices is provided in Figure 3.

Sequence and haplotype sharing between or among groups defined by subsistence strategy, linguistic classification and geographic region was also investigated (Table 2), resulting in opposite conclusions drawn based on sequence analyses: the data suggested higher gene flow in eastern populations. Results show that pastoralists and farmers share only 40 haplotypes, which comprises a higher percentage of sequences (individuals) in pastoralists than farmers (28% vs. 10%, respectively.) These 40 haplotypes are seen mainly in eastern pastoralists and farmers, but in some Fulani as well. We also examined the haplogroup classification of the shared haplotypes, and found a much higher contribution of non-sub-Saharan lineages among populations of eastern part of the SSB (e.g. R0a, M1, H2a1a, HV1, and T2) than in the west. The overview of all the haplotypes, haplogroups and their frequencies can be seen in Supplementary Table 6. Interesting evidence emerges when language groupings were considered in addition to region – although the Niger-Congo family contains the highest number of samples, it shares the lowest percentage of sequences with the other two groups. The highest percentage is shared between the Nilo-Saharan and Afro-Asiatic group (17%). As far as geographic distribution of populations is concerned, the groups (eastern and western) share fewer haplotypes (29) between each other than groups specified by subsistence strategy. Results of shared haplotypes are visualized in Supplementary Figure 2.

To explore the relationships between genetic and geographic distances we conducted a Mantel test and a SAMOVA analysis. The Mantel test (Figure 4) did not indicate any correlation in either group of pastoralists, but was positive and significant for farmers in both western ($R^2 = 0.321$, $p < 0.05$) and eastern ($R^2 = 0.070$, $p > 0.05$) regions, a finding which is compatible with a pattern of isolation by distance. The spatial analysis of molecular variance (SAMOVA) allowed us to define only one main group of populations and several single differentiated ones, which is in agreement with the MDS analysis similarly indicating the Rashaayda as representing the group most highly differentiated from the rest of the gene pool. After increasing the number of comparative groups (k) gradually from two to ten, results showed that after the Rashaayda, the most highly differentiated are the Beja, Sandawe, Burunge, Datog, Turu, Fulani from Ziniare, Fulani Halpularen and Somali (Table 3), revealing that the highest outliers are found in the eastern part of the SSB, and that of those outliers that were observed among western groups, all belong among the Fulani pastoralists.

Next, to estimate the extent and direction of gene flows among the groups of populations, we ran Migrate-n software coalescence analyses. Based on a Bayesian model with an expected variable migration rate (M) and constant index *theta* Θ, we found that the highest number of migrants has been exchanged between eastern farmers and pastoralists. These results, which highlight that the ratio of migrants in the western SSB is much lower than that in the east, go hand in hand with the notion of a higher differentiation of the pastoralists in the western SSB and/or higher pastoralist integration (mainly of Arabs) in its eastern part. The sedentary farmers' twice higher effective population size values, along with their present IBD model (Mantel test) population structure, are also consistent with the results of the selective neutrality tests indicating population expansion. A signal of migration between eastern and western pastoralists supports our results from haplotype sharing, but show higher gene flow from east to west than vice-versa. See Table 4 for complete results.

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Discussion

The aim of this study is to estimate the maternal genetic structure of the SSB, and try to understand it in light of archaeological and linguistic data. For this reason we based our results on a large mtDNA dataset providing higher molecular resolution than has been provided by similar studies to-date: instead of a short sequence of HVS-1 (currently about 350 bp), we generated a new dataset and gathered whole available D-loop segment encompassing 1,087 bp (nps 16030–579) in a sample size of almost 2,000 individuals representing 36 human populations distributed in two geographic regions across a large expanse of the SSB area (western and eastern), belonging to either of two lifestyles (nomadic pastoralist and sedentary farmer) and one of three linguistic groups (Nilo-Saharan, Afro-Asiatic, and Niger-Congo).

We provide a comprehensive comparison of diversity measures and distribution of genetic variation for several newly sampled populations from the perspective of two statistical approaches. The first, the frequency-based approach, either accepts or rejects the null hypothesis using permutations; the other, Bayesian analysis, involves probabilistic modeling based rather on prior assumptions and taking the maximum likelihood for granted. Due to fundamental differences in these methods, we observed an apparent contradiction in our results. All in all, we can confirm the previously observed contrast in the population dynamics between the two overall SSB subsistence strategies (Černý et al. 2006; Černý et al. 2007), with the generally lower values in nomadic pastoralists suggesting their genetic isolation. We have found that this observation is valid especially for the Beja and Rashaayda in the eastern part of the SSB, as well as for contemporary (but not really for now sedentary but historically active) Fulani pastoralists in the west. Therefore, in this regard, we can support the conclusions previously reached by Černý et al. (2011) using in the of

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

geographically (Lake Chad Basin) and ethnically (Fulani) restricted dataset suggesting a recent population bottleneck and a progressive loss of mtDNA diversity in pastoral populations due to genetic drift and asymmetric gene flow. Nevertheless, this finding has to be considered geographically specific to the SSB, as it has not been observed when studying pastoralists and farmers in other places such as Central Asia (Chaix et al. 2007.)

This contrasting pattern in the distribution of maternally inherited genetic variation between nomadic pastoralists and sedentary farmers can be observed in almost all the results. We can state that the Rashaayda and Beja are the two most diverging groups, which in terms of haplotype frequencies show (practically) no mutual sharing, nor sharing with the sedentary farmers inhabiting the eastern SSB; it is probable that this finding is a result of these two groups' relatively recent penetration to the region. On the other hand, other local pastoralists such as the Fulani, Daza, and Arabs do share some haplotypes with sedentary farmers, suggesting a relatively more frequent exchange of females. While it is true that the sharing of one or more haplotypes does not point directly to gene flow between groups, it can be assumed that very few non-sub-Saharan lineages were present before Arab immigration to Africa began in the 7th century, and that the newly incoming Arabs carried mostly Eurasian lineages. Therefore, the very occurrence of shared lineages between eastern pastoralists (mainly the Arabs) and farmers (considered indigenous peoples) implies that gene flow is the most likely cause. We found it noteworthy that the single most divergent group in our dataset, the Beja pastoralists, were mentioned by the Middle Age historian al-Maqrizi (Khutat) as having mixed with an Arabic tribe called Rabia during their migration to east Sudan; it is interesting that we did not, however, observe any non-sub-Saharan haplotype sharing between our Arab and Beja datasets: the only lineages they shared were sub-Saharan ones, suggesting gene flow into the Arabic population. After applying the migration model we discovered a very strong asymmetry in the gene flows between nomads and farmers in both

of the eastern and western SSB; while our analysis shows the asymmetry to be in the same direction from farmers to nomads in both regions, it is much more pronounced in the east, which again is consistent with previous results.

It can be objected that higher haplotype sharing seen in the eastern part of the SSB occurs for the non-sub-Saharan lineages, which have lower diversity and have been introduced to the area more recently. However, even if the Fulani's differentiation from the western SSB gene pool had occurred at a similar point in time as the Arab immigration, the main sub-Saharan lineages were already more diverse, making it more difficult to statistically discover sharing among them. Thus, it is possible that gene flow between populations of different subsistence systems occurs in both regions, with the flow in the eastern part of the SSB simply being more evident. It is obvious that gene flow is more common among Fulani sub-populations than between Fulani and sedentary groups, leading to prevalence of typical sub-Saharan lineages in this population. It is interesting to note that Eurasian haplogroup H1cb, previously interpreted as a sign of the Fulani's mid-Holocene contact with a population of Mediterranean origin (Kulichová et al. 2017) was also detected in six individuals in our dataset (Supplementary Table 6).

We suggest this pattern of genetic variation is the result of the long-term mating strategies maintained between these two subsistence groups co-habiting the SSB region for at least 4,000 years, maybe longer in its eastern part (Kuper and Riemer 2015). It has been shown that from the ecological point of view, pastoralism and farming are (at least in the northern Sahel, where droughts are currently recurrent) mutually exclusive livelihood alternatives – a combination of these practices are logistically and organizationally very demanding (Pedersen and Benjaminsen 2008). Indeed, the loss of genetic diversity and lower population dynamics in pastoralists is striking considering the fact that pastoralists are mostly

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

better nourished and suffer fewer infections in comparison with their neighboring farmers (Boström et al. 2012; Sheik-Mohamed and Velema 1999).

Our study suggests that pastoralism has had a different population history in the western compared to the eastern part of SSB. In the west, several lines of evidence point to separate cultural identity (Bonfiglioli 1988; Botte et al. 1999; Dupire 1970) and genetic diversification (Kulichová et al. 2017; Tishkoff et al. 2009; Triska et al. 2015) of the Fulani nomads relative to their sedentary neighbors. It has also been shown that a relatively high number of marriages within the patrilineal and patrilocal Fulani society is consanguineous (Hampshire and Smith 2001). Inter-ethnic marriages have not been studied so far across the SSB, and are considered rather as a rare or a recent phenomenon limited to urban areas (Blanc et al. 1990; Maiga et al. 2014). It can be suggested that the reasons for close intra-ethnic marriages are social and religious, and that their frequency in the past was probably not too different from what it is now.

The situation in the eastern part of the SSB is completely different; the African Arabs, unlike the Fulani, have not been differentiated from the common sub-Saharan gene pool, but arrived to Africa as new immigrants from the Arabian Peninsula some 1,300 years ago; this migration was described in several historical accounts (Zeltner 1980; Zeltner 2002). Contacts between the Arabic newcomers, whose population was probably not very numerous, and local peoples can also be attested to by linguistic evidence such as the several varieties of pidgin and creole tongues spoken in the eastern part of the SSB today, and by certain structural features of Kordofanian Baggara Arabic that are common to western Sudanic dialects (Manfredi 2012; Owens 1997). In fact, for the Baggara Arabs (unlike in the Fulani), inter-ethnic marriages with indigenous sub-Saharan populations is considered permissible, and appears to be increasing with time, as is suggested by genetic data (Bayoumi et al. 1985; Hassan et al. 2008; Priehodová et al. 2017).

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Acknowledgements

The study was financed by the Grant Agency of the Czech Republic (grant 13–37998S-P505). The authors are grateful to two anonymous reviewers for their comments on an earlier version of the manuscript.

Received 19 January 2018; revision accepted for publication 1 May 2018.

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Literature Cited

Andrews, R. M., I. Kubacka, P. F. Chinnery et al. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23:147.

Ba, A. H., and G. Dieterlen. 1966. Les fresques d'époque bovidienne du Tassili N'Ajjer et les traditions des Peul: hypothèses d'interprétation. *Journal de la Société des Africanistes* 36:141–157.

Bamshad, M., and S. P. Wooding. 2003. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* 4:99–111.

Barbieri, C., M. Whitten, K. Beyer et al. 2012. Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol. Biol. Evol.* 29:1,213–1,223.

Bayoumi, R. A., T. S. Taha, and N. Saha. 1985. A study of some genetic characteristics of the Fur and Baggara tribes of the Sudan. *Am. J. Phys. Anthropol.* 67:363–370.

Beerli, P. 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? In *Population Genetics for Animal Conservation*, G. Bertorelle, M. W. Bruford, H. C. Hauffe et al., eds. Cambridge: Cambridge University Press, 42–79.

Bendall, K. E., and B. C. Sykes. 1995. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am. J. Hum. Genet.* 57:248–256.

Blanc, M., A. Sanchez-Mazas, N. H. Van Blyenburgh et al. 1990. Interethnic genetic differentiation: GM polymorphism in eastern Senegal. *Am. J. Hum. Genet.* 46:383–392.

Blench, R. 1999. The westward wanderings of Cushitic pastoralists: Explorations in the prehistory of Central Africa. In *L'homme et l'animal dans le bassin du lac Tchad*, C. Baroin and J. Boutrais, eds. Paris: IRD, 39–80.

Blench, R. 2006. *Archaeology, Language and the African Past*. Lanham: Altamira Press.

Bonfiglioli, A. 1988. *Dudal: histoire de famille et histoire de troupeau chez un groupe de Wodaabe du Niger*. Cambridge: Cambridge University Press.

Boström, S., P. Giusti, C. Arama et al. 2012. Changes in the levels of cytokines, chemokines and malaria-specific antibodies in response to Plasmodium falciparum infection in children living in sympatry in Mali. *Malar. J.* 11:109.

Botte, R., J. Boutrais, and J. Schmitz. 1999. *Figures peules*. Paris: Karthala.

Bouchette, F., M. Schuster, J.-F. Ghienne et al. 2017. Hydrodynamics in Holocene Lake Mega-Chad. *Quat. Res.* 73:226–236.

Brandstätter, A., C. T. Peterson, J. A. Irwin et al. 2004. Mitochondrial DNA control region sequences from Nairobi (Kenya): Inferring phylogenetic parameters for the establishment of a forensic database. *Int. J. Legal Med.* 118:294–306.

Bučková, J., V. Černý, and A. Novelletto. 2013. Multiple and differentiated contributions to the male gene pool of pastoral and farmer populations of the African Sahel. *Am. J. Phys. Anthropol.* 151:10–21.

Cerezo, M., V. Černý, Á. Carracedo et al. 2011. New insights into the Lake Chad Basin population structure revealed by high-throughput genotyping of mitochondrial DNA coding SNPs. *PLoS One* 6:e18682.

Černý, V., M. Čížková, E. S. Poloni et al. 2016. Comprehensive view of the population history of Arabia as inferred by mtDNA variation. *Am. J. Phys. Anthropol.* 159:607–616.

Černý, V., M. Hájek, M. Bromová et al. 2006. MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum. Biol.* 78:9–27.

Černý, V., M. Hájek, R. Cmejla et al. 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. *Ann. Hum. Biol.* 31:554–569.

Černý, V., L. Pereira, E. Musilová et al. 2011. Genetic structure of pastoral and farmer populations in the African Sahel. *Mol. Biol. Evol.* 28:2,491–2,500.

Černý, V., A. Salas, M. Hájek et al. 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. *Ann. Hum. Genet.* 71:433–452.

Chaix, R., L. Quintana-Murci, T. Hegay et al. 2007. From social to genetic structures in central Asia. *Curr. Biol.* 17:43–48.

Čížková, M., Z. Hofmanová, M. G. Mokhtar et al. 2017. Alu insertion polymorphisms in the African Sahel and the origin of Fulani pastoralists. *Ann. Hum. Biol.* 44:537–545.

Connah, G. 2001. *African Civilizations: An Archaeological Perspective.* Cambridge: Cambridge University Press.

Delmet, C. 2000. *Les Peuls nomades au Soudan*. In *L'ethnicité peule dans des contextes nouveaux,* Y. Diallo and G. Schlee, eds. Paris: Karthala, 191–206.

Drake, N. A., R. M. Blench, S. J. Armitage et al. 2011. Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proc. Natl. Acad. Sci. U. S. A.* 108:458–462.

Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11:2,571–2,581.

Dupire, M. 1970. *Organisation sociale des Peul: étude d'ethnographie comparée*. Paris: Plon.

Dupuy, C. 1999. Les apports de l'achéologie et de l'ethnologie à la connaisance de l'histoire ancienne des Peuls. In *Figures peules*, R. Botte, J. Boutrais, and J. Schmitz, eds. Paris: Karthala, 53–72.

Ehret, C. 2002. *Civilizations of Africa: A History to 1800*. Virginia: University Press of Virginia.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1:47–50.

Excoffier, L., and H. E. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10:564–567.

Excoffier, L., and S. Schneider. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc. Natl. Acad. Sci. U. S. A.* 96:10,597–10,602.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.

Fernandes, V., P. Triska, J. B. Pereira et al. 2015. Genetic stratigraphy of key demographic events in Arabia. *PLoS One* 10:e0118625.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925.

Gonder, M. K., H. M. Mortensen, F. A. Reed et al. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* 24:757–768.

Guillot, G., F. Mortier, and A. Estoup. 2005. Geneland: A computer package for landscape genetics. *Mol. Ecol. Notes* 5:712–715.

Hampshire, K. R., and M. T. Smith. 2001. Consanguineous marriage among the Fulani. *Hum. Biol.* 73:597–603.

Harpending, H. C., S. T. Sherry, A. G. Rogers et al. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.* 34:483–496.

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Hassan, H. Y., P. A. Underhill, L. L. Cavalli-Sforza et al. 2008. Y-chromosome variation among Sudanese: Restricted gene flow, concordance with language, geography, and history. *Am. J. Phys. Anthropol.* 137:316–323.

Heine, B., and D. Nurse. 2000. *African Languages: An Introduction*. Cambridge: Cambridge University Press.

Insoll, T. 2003. *The Archaeology of Islam in Sub-Saharan Africa (Cambridge World Archaeology).* Cambridge: Cambridge University Press.

Kulichová, I., V. Fernandes, A. Deme, et al. 2017. Internal diversification of non-Sub-Saharan haplogroups in Sahelian populations and the spread of pastoralism beyond the Sahara. *Am. J. Phys. Anthropol.* 164:424–434.

Kuper, R., and S. Kröpelin. 2006. Climate-controlled Holocene occupation in the Sahara: Motor of Africa's evolution. *Science* 313:803–807.

Kuper, R., and H. Riemer. 2015. *Herders before Pastoralism: Prehistoric Prelude in the Eastern Sahara*. New York: Berghan Books.

Le Quellec, J.-L. 2004. *Arts rupestres et mythologies en Afrique*. Paris: Flammarion.

Levy, T. E., and A. F. C. Holl. 2002. Migrations, ethnogenesis, and settlement dynamics: Israelites in Iron Age Canaan and Shuwa-Arabs in the Chad Basin. *J. Anthropol. Archaeol.* 21:83–118.

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

MacIntosh, R. J. 2005. *Ancient Middle Niger: Urbanism and the Self-Organizing Landscape.* Cambridge: Cambridge University Press.

MacMichael, H. 1922. *A History of the Arabs in the Sudan and Some Account of the People Who Preceded Them and of the Tribes Inhabiting Darfur*. Cambridge: Cambridge University Press.

Maiga, B., A. Dolo, S. Campino et al. 2014. Glucose-6-phosphate dehydrogenase polymorphisms and susceptibility to mild malaria in Dogon and Fulani, Mali. *Malar. J.* 13:270.

Manfredi, S. 2012. Dialect mixing and dialect levelling in Kordofanian Baggara Arabic (Western Sudan). In *Dynamiques langagières en Arabophonies: variations, contacts, migrations et créations artistiques: hommage offert à Dominique Caubet par ses élèves et collègues,* A. Barontini, ed. Universidad de Zaragoza Área de Estudios Árabes e Islámicos, 141–162.

Musilová, E., V. Fernandes, N. M. Silva et al. 2011. Population history of the Red Sea—genetic exchanges between the Arabian Peninsula and East Africa signaled in the mitochondrial DNA HV1 haplogroup. *Am. J. Phys. Anthropol.* 145:592–598.

Newman, J. L. 1995. *The Peopling of Africa: A Geographic Interpretation*. New Haven, Connecticut: Yale University Press.

Oota, H., W. Settheetham-Ishida, D. Tiwawech et al. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat. Genet.* 29:20–21.

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

Owens, J. 1997. Arabic-based pidgins and creoles. In *Contact Languages: A Wider Perspective,* S. Thomason, ed. Amsterdam: John Benjamins, 125–172.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2,537–2,539.

Pedersen, J., and T. A. Benjaminsen. 2008. One leg or two? Food security and pastoralism in the Northern Sahel. *Hum. Ecol.* 36:43–57.

Poloni, E. S., Y. Naciri, R. Bucho et al. 2009. Genetic evidence for complexity in ethnic differentiation and history in East Africa. *Ann. Hum. Genet.* 73:582–600.

Priehodová, E., F. Austerlitz, M. Čížková et al. 2017. The historical spread of Arabian Pastoralists to the eastern African Sahel evidenced by the lactase persistence -13,915*G allele and mitochondrial DNA. *Am. J. Hum. Biol.* 29:e22950.

Ray, N., M. Currat, and L. Excoffier. 2003. Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* 20:76–86.

Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* 105:767–779.

RStudio Team. 2016. RStudio: Integrated Development for R. Boston: RStudio, Inc. http://www.rstudio.com/.

Sheik-Mohamed, A., and J. P. Velema. 1999. Where health care has no access: The nomadic populations of sub-Saharan Africa. *Trop. Med. Int. Health* 4:695–707.

Skonieczny, C., P. Paillou, A. Bory et al. 2015. African humid periods triggered the reactivation of a large river system in Western Sahara. *Nat. Commun.* 6:8,751.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.

Soares, P., F. Alshamali, J. B. Pereira et al. 2012. The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* 29:915–927.

Sutton, J. E. G. 1974. The aquatic civilization of Middle Africa. *J. Afr. Hist.* 15:527–546.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* 526:68.

Tishkoff, S. A., F. A. Reed, F. R. Friedlaender et al. 2009. The genetic structure and history of Africans and African Americans. *Science* 324:1,035–1,044.

Triska, P., P. Soares, E. Patin et al. 2015. Extensive admixture and selective pressure across the Sahel Belt. *Genome Biol. Evol.* 7:3,484–3,495.

Young, W. 1996. *The Rashaayda Bedouin: Arab Pastoralists of Eastern Sudan*. Fort Worth, Texas: Harcourt Brace College Publishers.

Zeltner, J. 1980. *Pages d'histoire du Kanem, pays tchadien*. Paris: l'Harmattan.

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

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Table 1. Analyses of the Molecular Variance (AMOVA)

Grouping	N° of Groups	N° of Pop	Hierarchical AMOVA: Proportion of variation (Fixation indices)		
			Among Groups (ΦCT)	Among Populations Within Group (ΦSC)	Among Populations (ΦST)
ALL	1	36			0.08881*
nomadic	1	10			0.13298*
sedentary	1	26			0.07340*
western	1	19			0.03549*
eastern	1	17			0.08437*
Nilo-Saharan	1	4			0.06998*
Afro-Asiatic	1	7			0.09561*
Niger-Congo	1	23			0.05607*
subsistence (nomadic vs. sedentary)	2	36	0.00253 n.s.	0.08791*	0.09023*
geography (western vs. eastern)	2	36	0.05265*	0.06260*	0.11196*
language (Nilo-Saharan vs. Afro-Asiatic vs. Niger-Congo)	3	34	0.03633*	0.06632*	0.10024*
nomadic (western vs. eastern)	2	10	0.07557**	0.09444*	0.16287*
sedentary (western vs. eastern)	2	26	0.02475**	0.06164*	0.08487*
western (nomadic vs. sedentary)	2	19	0.00000 n.s.	0.08440*	0.08433*
eastern (nomadic vs. sedentary)	2	17	0.03697*	0.02219*	0.05834*
subsistence + language	6	34	0.00000 n.s.	0.10936*	0.09980*

Population in groups: subsistence - groupings to see in Supplementary Table 1; geography (Western: FZR, BED, FFE, GRS, GUR, MOS, SSR, FFD, HAL, SSA, NSA, NUN, MAR, LYE, BIS, GAM, MEN, YOR, ESA; Eastern: ARA, BEJ, NUB, ABG, RAS, SOM, KEN, BUR, DAT, SAN, TUR, KAW, ABA, DAZ, DNG, MAB, LUH) ; same division in groups "nomadic (western vs. eastern)", "sedentary (western vs. eastern)", "western (nomadic vs. sedentary)" and "eastern (nomadic vs. sedentary)" but just within subsistence and geographic groups (to see in Supplemetary Table S1.); language (Afro-Asiatic: ABA, ARA, BEJ, DNG, ABG, RAS, SOM; Nilo-Saharan: DAZ, MAB, NUB, DAT; Niger-Congo: BED, FFE, FFD, FZR, HAL, GRS, GUR, MOS, SSR, KAW, SSA, NSA, NUN, MAR, LYE, BIS, BUR, TUR, YOR, ESA, GAM, MEN, LUH); subsistence + language division is made into 6 groups according to lifestyle and language group

* P-value < 0,001

** P-value < 0,05

n.s. = non-significant

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Table 2. Haplotype and mtDNA Sequences Sharing of Groupings Based on Subsistence Strategy, Linguistic Affiliation and Geographic Region

SEQUENCE AND HAPLOTYPE SHARING

Subsistence	Sequences[a]	Haplotypes
sedentary	1502	1096
nomadic	492	299
sedentary-nomadic	147-138 (10%-28%)	40

Language	Sequences[a]	Haplotypes
Afro-Asiatic (AA)	448	324
Nilo-Saharan (NS)	175	140
Niger-Congo (NC)	1371	939
AA-NS	59-30 (13%-17%)	20
NS-NC	22-63 (13%-5%)	14
AA-NC	31-85 (7%-6%)	21

Geography	Sequences[a]	Haplotypes
eastern	1014	713
western	980	674
eastern-western	91-112 (9%-11%)	29

[a] Number before dash is the number of sequences shared between subsistence strategy or language on the first place and second place, after the dash is the opposite (shared haplotypes between the second one and the first one). Classification into linguistic groups to see in Supplementary Table 1

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Table 3. Spatial Analysis of Molecular Variance (SAMOVA) of Sahelian Populations

N° of groups	Populations									F_{CT}	F_{SC}
2	RAS	ABA, ARA, BED, BEJ, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, FZR, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, BUR, DAT, SAN, TUR, YOR, ESA, GAM, MEN, LUH								0.11521	0.08256
3	RAS	SAN	ABA, ARA, BED, BEJ, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, FZR, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, BUR, DAT, TUR, YOR, ESA, GAM, MEN, LUH							0.10804	0.07422
4	RAS	SAN	BEJ	ABA, ARA, BED, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, FZR, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, BUR, DAT, TUR, YOR, ESA, GAM, MEN, LUH						0.10420	0.07000
5	RAS	SAN	BEJ	BUR	ABA, ARA, BED, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, FZR, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, DAT, TUR, YOR, ESA, GAM, MEN, LUH						0.10277	0.06662
6	RAS	SAN	BEJ	BUR	DAT	ABA, ARA, BED, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, FZR, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, TUR, YOR, ESA, GAM, MEN, LUH					0.10316	0.06283
7	RAS	SAN	BEJ	BUR	DAT	TUR	ABA, ARA, BED, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, FZR, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, YOR, ESA, GAM, MEN, LUH					0.10027	0.06102
8	RAS	SAN	BEJ	BUR	DAT	TUR	FZR	ABA, ARA, BED, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, YOR, ESA, GAM, MEN, LUH				0.09726	0.05729
9	RAS	SAN	BEJ	BUR	DAT	TUR	FZR	HAL	ABA, ARA, BED, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, YOR, ESA, GAM, MEN, LUH			0.09510	0.05431

10	RAS	SAN	BEJ	BUR	DAT	TUR	FZR	HAL	SOM	ABA, ARA, BED, DAZ, DNG, FFE, GRS, GUR, MAB, MOS, NUB, SSR, ABG, FFD, KAW, HAL, SSA, NSA, NUN, MAR, LYE, BIS, SOM, KEN, YOR, ESA, GAM, MEN, LUH	0.09377	0.04425

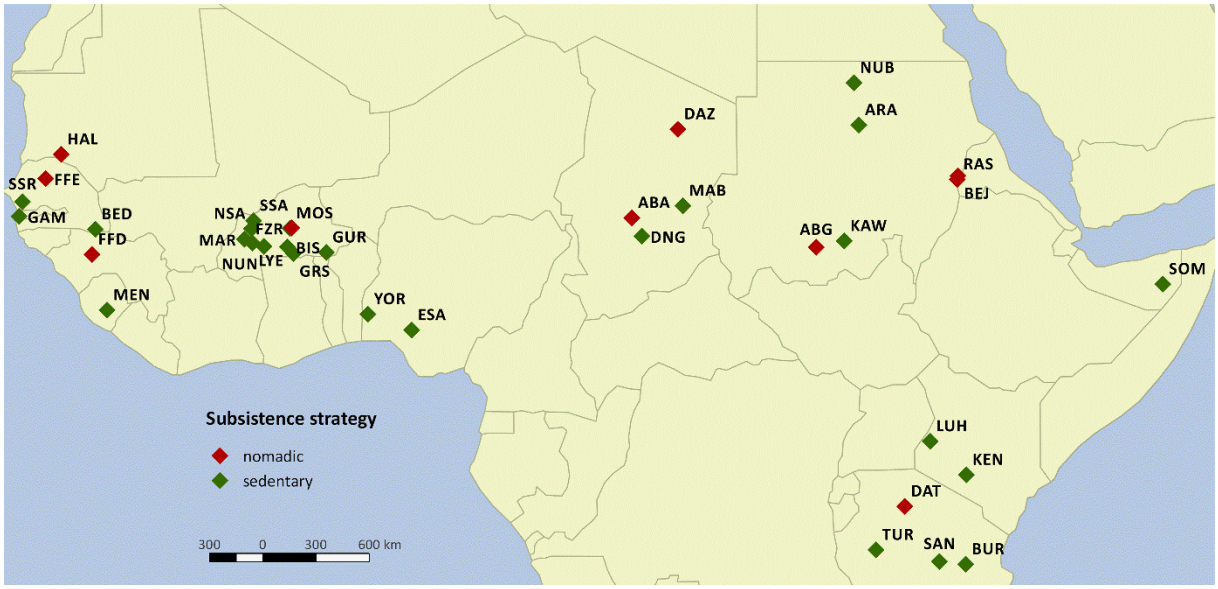
Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Table 4. Values of Scaled Effective Population Size and Immigration Rate Based on Bayesian Coalescence

Groups		Θ	M	Direction (from-to)	Number of migrants (Θ*M)
Within region					
east	nomadic (n)	0,01111	771,9	n-s	9
	sedentary (s)	0,02362	860,0	s-n	20
west	nomadic (n)	0,00507	883,2	n-s	4
	sedentary (s)	0,01692	906,0	s-n	15
Within subsistence strategy					
nomadic	east (e)	0,01111	303,0	e-w	3
	west (w)	0,00507	544,1	w-e	3
sedentary	east (e)	0,02362	864,9	e-w	20
	west (w)	0,01692	807,0	w-e	14

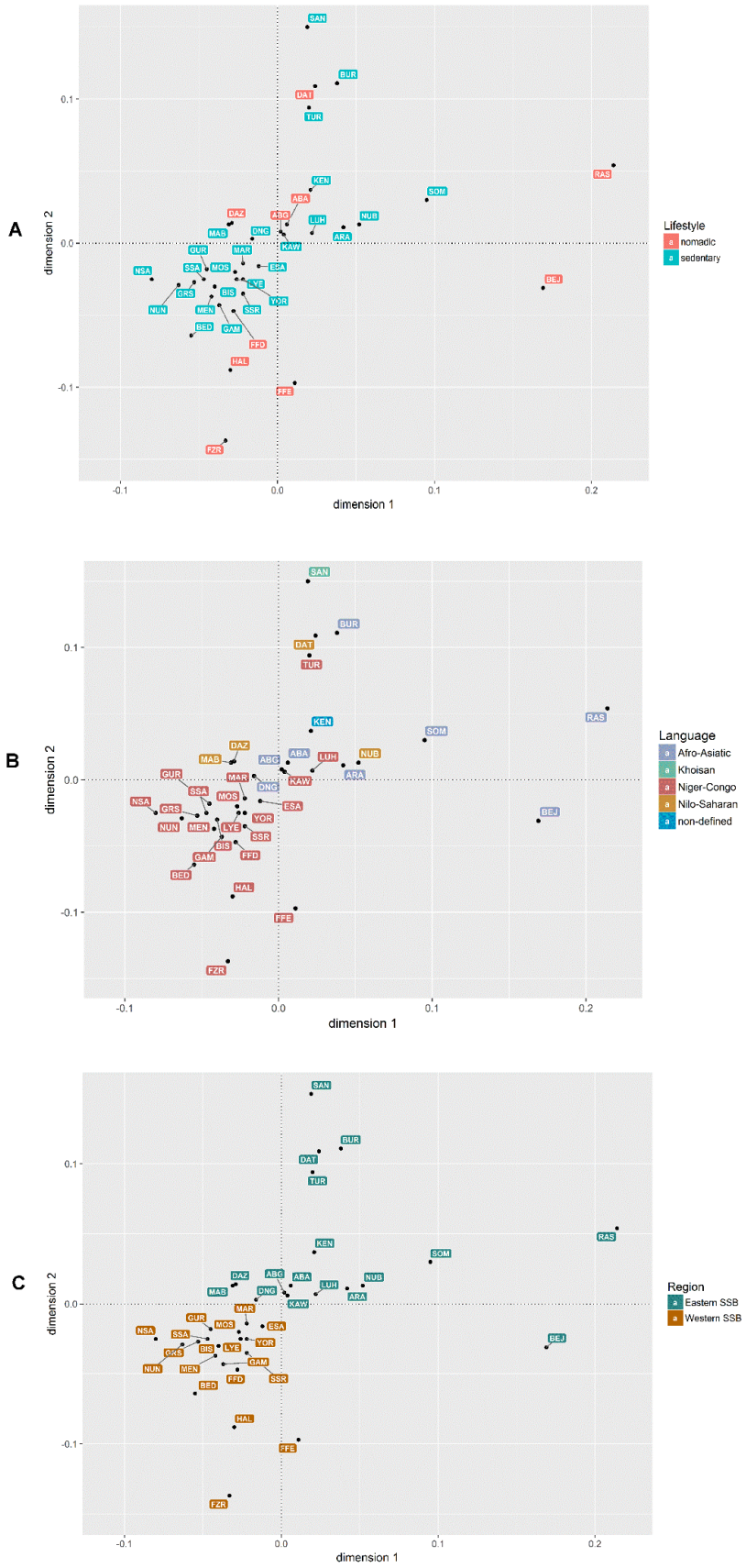
M = mutation-scaled effective immigration rate; Θ = mutation-scaled effective population size (in mtDNA it is Ne*mutation rate per site per generation)

Figure 1.



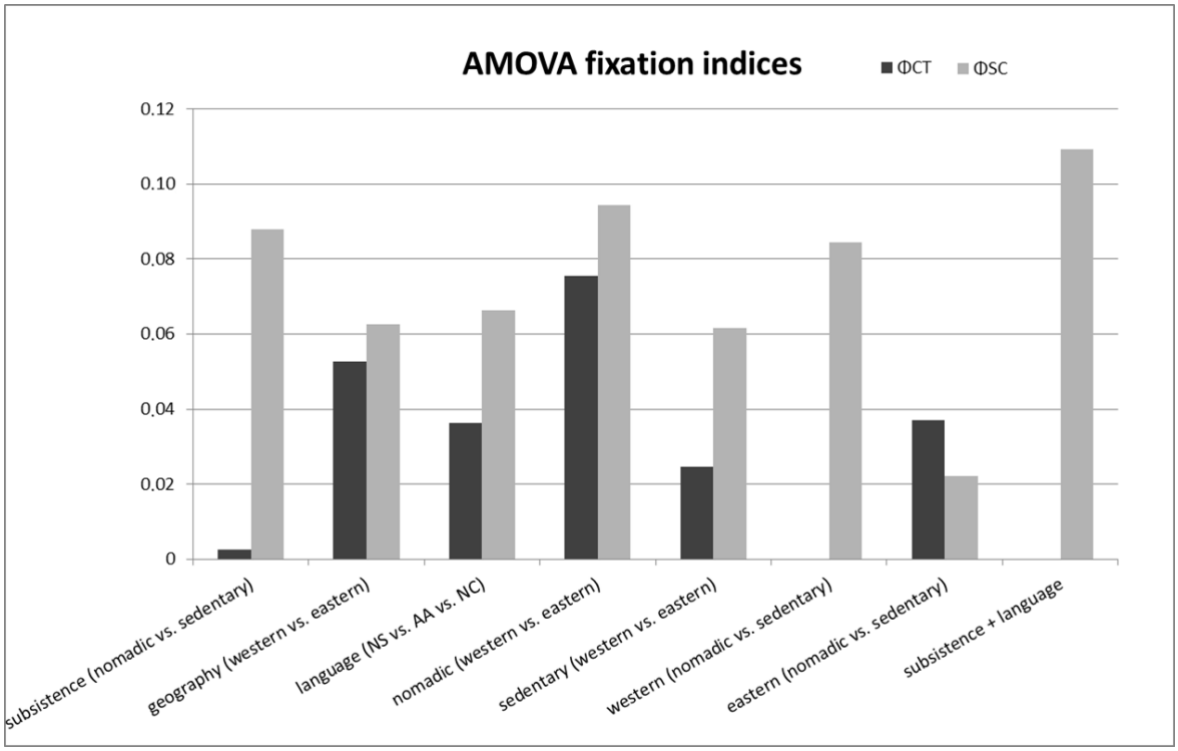
Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Figure 2.



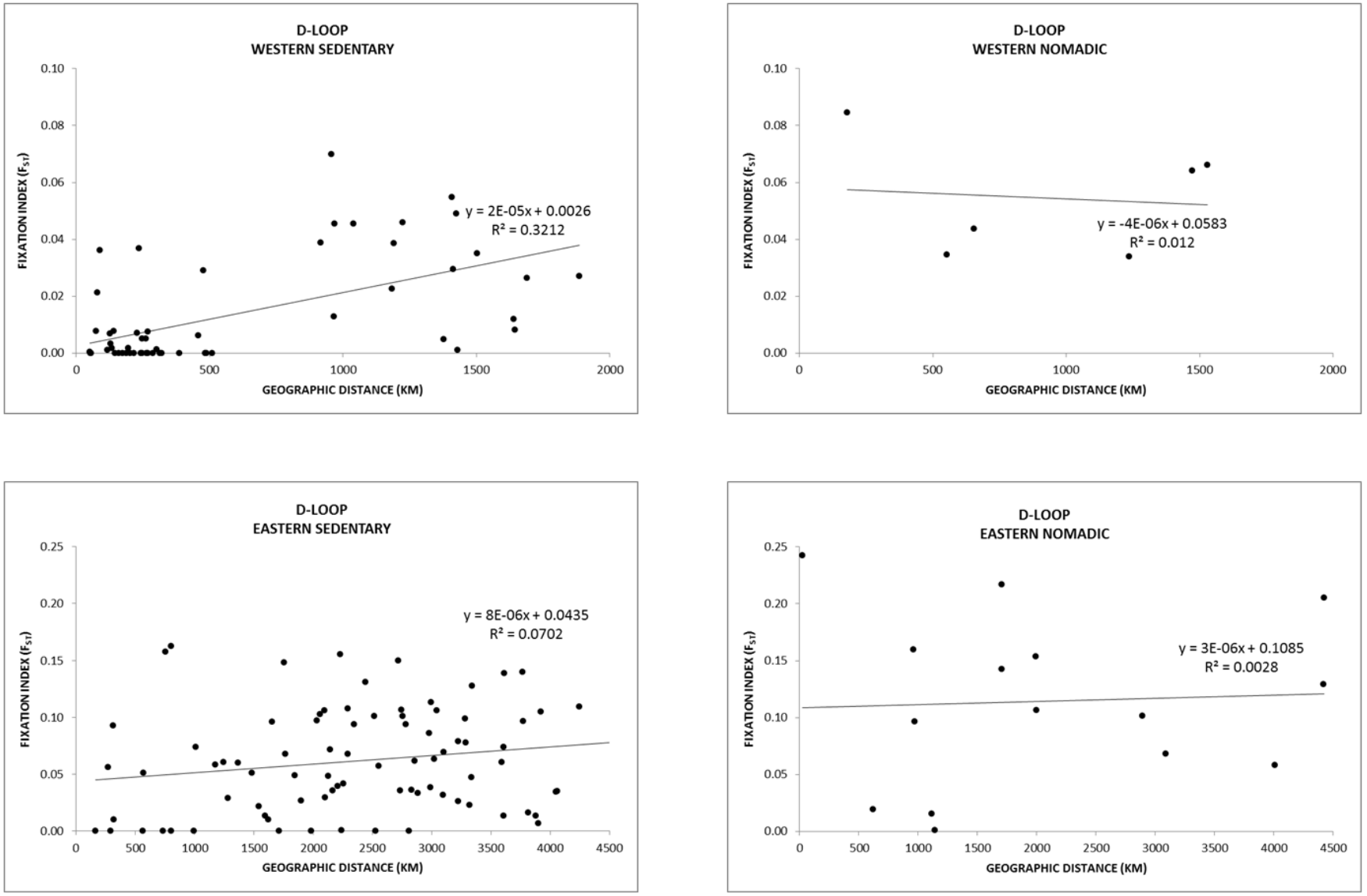
Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Figure 3.



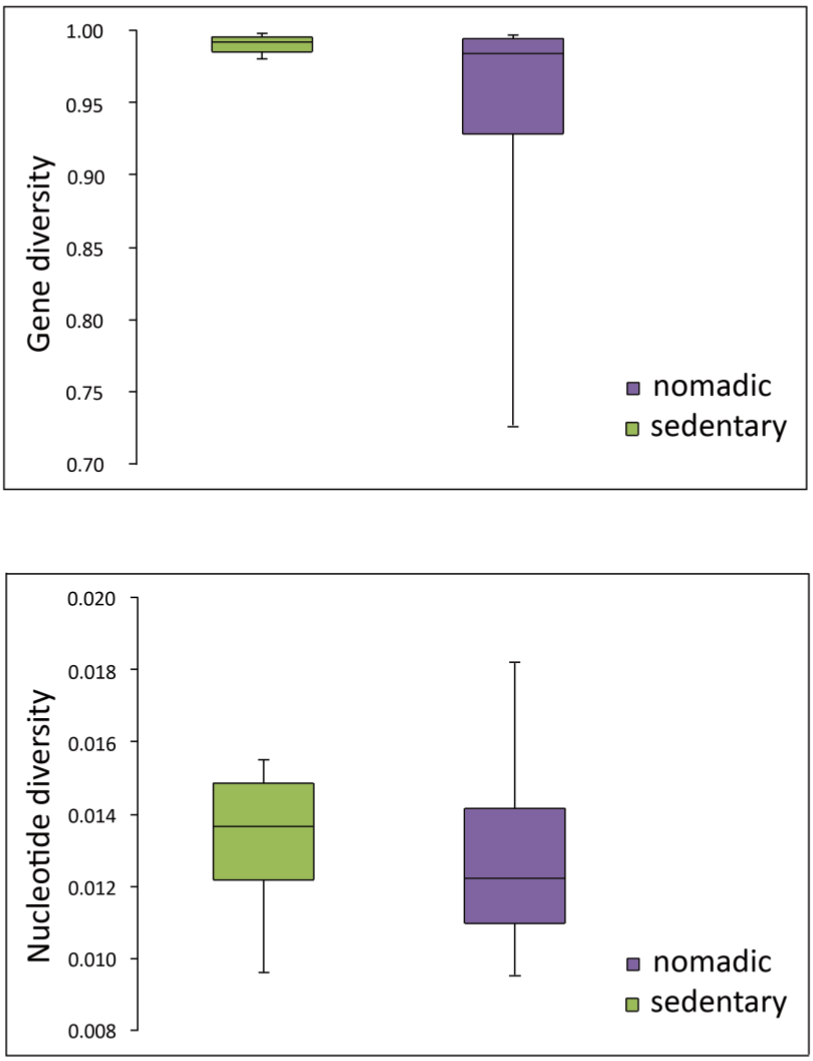
Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Figure 4.



Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Supplementary Figure S1.



Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

Supplementary Figure S2.

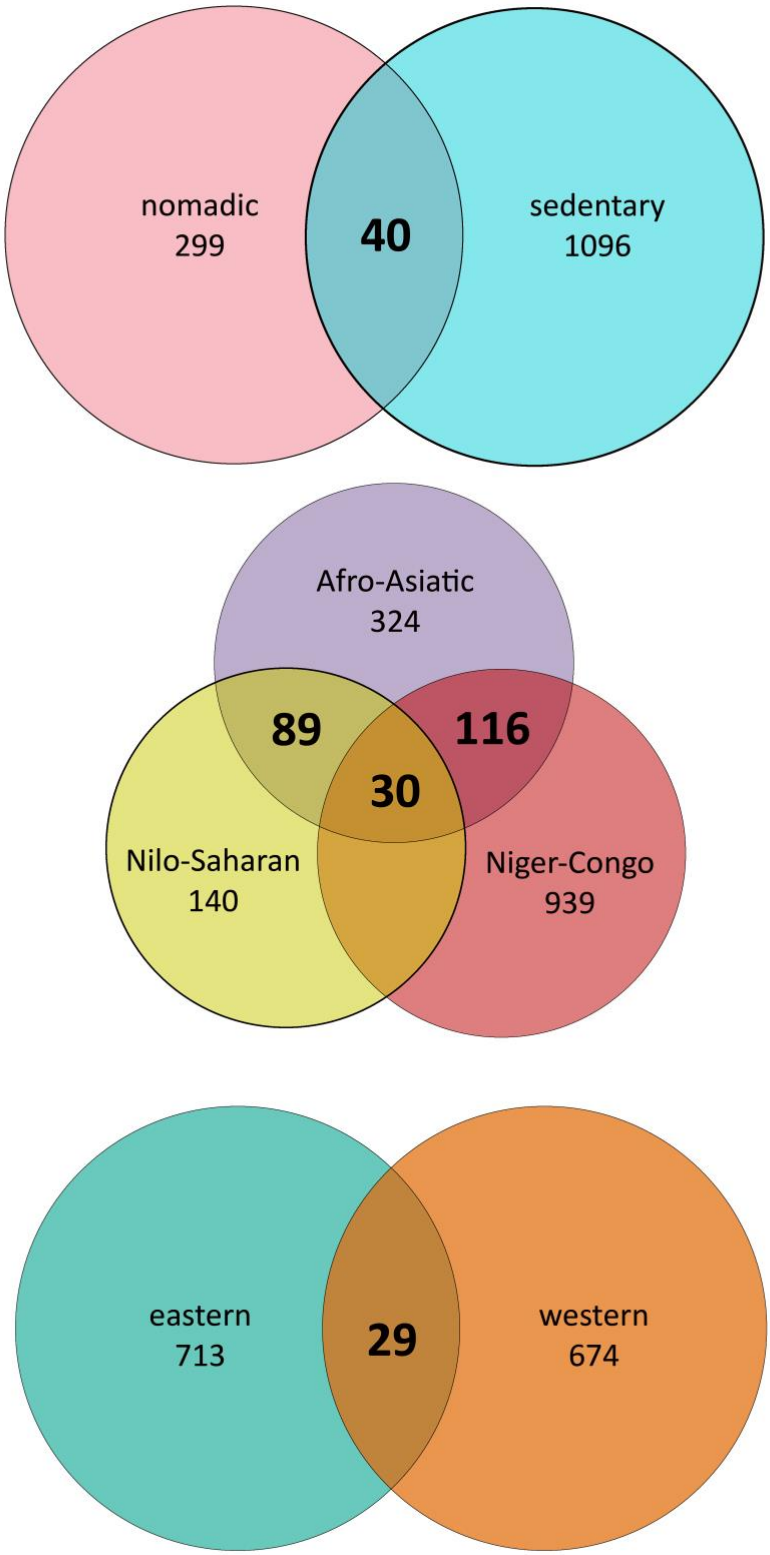


Figure captions

Figure 1. Geographic distribution of the samples included in this study.

Figure 2. MDS plot of pairwise Reynolds genetic distances between 36 populations of the SSB dataset, reproduced three times, color-coded according to (A) lifestyle, (B) linguistic affiliation and (C) geographical region. The stress value is 0.169.

Figure 3. AMOVA fixation indices for defined groups of populations as specified in Table 1.

Figure 4. Plots of Mantel test results comparing fixation indices and geographical distances of both subsistence strategies in both regions.

Supplementary Figure S1. Boxplots of gene and nucleotide diversity.

Supplementary Figure S2. Haplotype sharing between populations of different subsistence strategies (A), linguistic affiliations (B) and geographic region (C).

Pre-print version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.