

Genetic Structure of Pastoral and Farmer Populations in the African Sahel

Viktor Černý,^{*1} Luísa Pereira,^{2,3} Eliška Musilová,⁴ Martina Kujanová,⁴ Alžběta Vašíková,⁵ Paola Blasi,⁶ Luisa Garofalo,⁶ Pedro Soares,² Issa Diallo,⁷ Radim Brdička,⁵ and Andrea Novelletto⁶

¹Department of the Archaeology of Landscape and Archaeobiology, Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague, Czech Republic

²Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal

³Faculdade de Medicina da Universidade do Porto, Portugal

⁴Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic

⁵Department of Molecular Genetics, Institute of Hematology and Blood Transfusion, Prague, Czech Republic

⁶Department of Biology, University "Tor Vergata," Rome, Italy

⁷Département de Linguistique et Langues Nationales, Institut des Sciences des Sociétés, CNRST, Ouagadougou, Burkina Faso

***Corresponding author:** E-mail: cerny@arup.cas.cz.

Associate editor: Anne Stone

Abstract

Traditional pastoralists survive in few places in the world. They can still be encountered in the African Sahel, where annual alternations of dry and wet seasons force them to continual mobility. Little is known about the genetic structure of these populations. We present here the population distribution of 312 hypervariable segment I mitochondrial DNA (mtDNA) and 364 Y-short tandem repeat haplotypes in both farmer and pastoralist groups from the Lake Chad Basin and the West African Sahel. We show that the majority of pastoral populations (represented in the African Sahel by the Fulani nomads) fail to show significant departure from neutrality for mtDNA as evidenced by Fu's F_s statistics and exhibit lower levels of intrapopulation diversity measures for mtDNA when contrasted with farmers. These differences were not observed for the Y chromosome. Furthermore, analyses of molecular variance and population distributions of the mtDNA haplotypes show more heterogeneity in the sedentary groups than in the pastoralists. On the other hand, pastoralists retain a signature of a wide phylogenetic distance contributing to their male gene pool, whereas in at least some of the farmer populations, a founder effect and/or drift might have led to the presence of a single major lineage. Interestingly, these observations are in contrast with those recorded in Central Asia, where similar comparisons of farmer and pastoral groups have recently been carried out. We can conclude that in Africa, there have been no substantial mating exchanges between the Fulani pastoralists coming to the Lake Chad Basin from the West African Sahel and their farmer neighbors. At the same time, we suggest that the emergence of pastoralism might be an earlier and/or a demographically more important event than the introduction of sedentary agriculture, at least in this part of Africa.

Key words: African Sahel, mtDNA, Y chromosome, pastoralists, farmers.

Introduction

Although the contrasting patterns of genetic diversity between hunter gatherers and farmers have been examined in several studies (Excoffier and Schneider 1999; Pereira et al. 2001; Destro-Bisol et al. 2004; Pilkington et al. 2008; Verdu et al. 2009), we still know little regarding processes that have led to the present day genetic differentiation seen in farmers and pastoralists (Luca et al. 2008). Pastoralists' lifestyle conforms closely to the needs of domestic animals, and in this way, it resembles that of hunter gatherers, who are obliged to follow wild game. However, in today's age, there is no human subsistence activity being carried out in a social vacuum, and similarly to hunter-gatherers, pastoralists communicate, collaborate, and exchange their produce with farmers.

Extant studies analyzing social aspects of human evolutionary genetics deal mainly with the relationship of patrilocal versus matrilocality (Oota et al. 2001; Hamilton

et al. 2005). It has been shown that postmarital residence patterns significantly influence the genetic structure of a population. This is especially apparent in gender-specific genetic markers, such as mitochondrial DNA (mtDNA) and the nonrecombinant portion of the Y chromosome (NRY). Reduced Y-chromosomal diversity in some populations is generally explained by genetic drift caused by patrilocality (Marchani et al. 2008).

Insightful analysis of the genetic differences between farmers and pastoralists has come out of Central Asia (Chaix et al. 2007), where these groups coexist from some 6 thousand years ago (KYA). In contrast to their farmer neighbors, Central Asian pastoralists show a loss of NRY (measured by short tandem repeats or STRs) diversity that is not observed for mtDNA (measured by Hypervariable Segment I or HVS-I). According to the authors, the discrepancy can be explained by differences in social organization of these communities resulting in

genetic drift in pastoralist Y chromosomes. The study is an appealing example of how social structure can influence genetic structure in a short period of time (Balaresque and Jobling 2007).

In this article, we have tried to determine whether such differences between farmers and pastoralists are to be found elsewhere, looking concretely at the Sahel belt of Africa where they live in close proximity. The Sahel zone lies between the savannah zone in the south and Sahara desert in the north and represents one of the outstanding ecosystems of the planet. No agriculture is possible in the southern Sahara, though pastoralists still find some pasture for their animals during the short rainy season. Agriculture is, however, well sustainable in the more humid areas near the Savannah zone. The African Sahel, lying in between, thus forms a meeting ground for pastoralists and farmers.

Several forms of mutual relations and economical exchanges have been established between these two communities (Beauvillain 1977). Nomadic pastoralists cannot survive on their own resources alone, so women currently sell milk and dairy products to the farmers at the local markets and buy agricultural and craftwork products in return. The presence of the pastoralist's herds grazing on the farmer's fields after the harvest is also welcomed as a convenient form of natural fertilization.

Archaeologically, the origin of pastoral subsistence in Africa can be localized to the Eastern and Central Sahara (Lhote 1973; Wendorf et al. 1984; Dupuy 1999; Mohammed-Ali and Khabir 2003). Moreover, several rocky massifs in the Sahara contain a large number of prehistoric paintings, and engravings of domestic cattle that are dated later than the depictions of wild animals also found there and have led to the recognition of the presence of a Bovidian culture (Le Quellec 1997; Holl 2004; Smith 2005). Even though the exact chronology of such Saharan rock art is very contentious (Muzzolini 1992), it is commonly agreed that pastoralists were present here some 6 KYA. It has even been further suggested that some of the paintings found in the rocky shelters of the Tassili-n-Ajjer Mountains in Southeast Algeria were created by ancestors of the contemporary Fulani pastoralists because the same cattle cults and rites as depicted in the prehistoric paintings are still being performed by some Fulani groups today (Dupire 1962; Ba and Dieterlen 1966).

The Fulani nomads are unambiguously the dominant pastoral group of the West African Sahel (Botte et al. 1999; Diallo and Schlee 2000). According to the historical sources available, their cultural expansion began in the mountains of Futa Djallon in West Africa (Newman 1995) in the 11th century AD and continued eastward through the inner Niger delta as far as the Lake Chad Basin, where their first archaeological traces can be dated to the 15th century (David 1971; Gauthier 1979). Some Fulani pastoralists have settled into sedentary lifestyles in the Sahel during the last 500 years but many remain faithful to their original nomadic lifestyle (Schultz 1984; Jabbar et al. 1995). We show that these populations differ significantly in several genetic aspects as well as their lifestyles.

Materials and Methods

The Subjects

The African Sahel is a place of sedentary farmers and nomadic pastoralists. Although the pastoralists are represented almost exclusively by the Fulani people, the farmers can be classified into various ethnolinguistic groups. We collected samples (buccal swabs) of several Fulani pastoral groups from five African countries—Chad, Cameroon, Niger, Burkina Faso, and Mali, and in total, we secured samples of 432 individuals of which only 186 had already been published previously (Černý et al. 2006). All these individuals identified themselves as Fulani (locally also called Peul, Fulbé, Fula, Wodabé, M'Bororo, etc.) speaking one of the “Fulfulde” dialects/languages belonging to the West Atlantic branch of the Niger-Congo linguistic family. As Fulani groups move frequently (from temporary camp to camp), we registered the approximate place of our encounter during their dry season wandering. For comparison purposes, we collected samples from their sedentary farmer neighbors as well. The farmer sample is not ethnically homogenous and represents different linguistic families—Afro-Asiatic (Buduma, Bulahay, Hide, Kotoko, Mafa, and Masa—all currently classified to the Chadic branch), Nilo-Saharan (Songhai, Kanuri, and Kanembou), and Niger-Congo (Fali). Of the 470 sedentary farmer samples analyzed here, 294 have been previously published (Černý et al. 2007). Thus, the total number of new samples presented in this study equals 422. The exact numbers of individuals analyzed for mtDNA and NRY in each population are shown in tables 1 and 2, respectively, geographic locations of the population samples are shown in figure 1. When performing the NRY analysis, the Fali data were excluded, and the Mafa and the Hide were merged because of small sample sizes and the close linguistic and cultural affinities between them.

Laboratory Methods

DNA extraction and polymerase chain reaction amplification of mtDNA HVS-I segment were undertaken as previously reported (Černý et al. 2004, 2007). Mutations were scored relative to the revised Cambridge reference sequence (Andrews et al. 1999), and numbers within the range 16030–16370 refer to the position of the mutation in that sequence. Three restriction fragment length polymorphisms outside this region (3592 *Hpa*I, 2349 *Mbo*I, and 16391 *Av*II) were also assayed in some cases, allowing confirmation of the phylogenetic assignments inferred by HVS-I. Eight Y chromosome STRs (DYS19, DYS388, DYS389_1, DYS389_2, DYS390, DYS391, DYS392, and DYS393) were analysed in 364 male samples of both sedentary and pastoral groups according to standard methods (Carracedo et al. 2001; Butler et al. 2002).

Data Analysis

To evaluate genetic structure at the intrapopulation level for both mtDNA and Y-STR, we have analysed gene (haplotype) diversity and mean number of pairwise differences

Table 1. mtDNA Genetic Diversity of Populations Practicing Different Lifestyles From the West African Sahel.

Pastoralists Code	Place	Country	<i>n</i>	<i>K</i>	Hs	<i>II</i>	Dii	<i>r</i>	<i>P(r)</i>	<i>D</i>	<i>P(D)</i>	Fs	<i>P(Fs)</i>	<i>R</i> ₂	<i>P(R</i> ₂ <i>)</i>
FDIA	Diafarabe	Mali	50	28	0.962	0.0192	6.55	0.0380	0.05	-1.0287	0.1500	-11.3336	0.0000*	0.0727	0.1332
FBAN	Banfora	Burkina Faso	50	21	0.945	0.0202	6.90	0.0446	0.03*	-0.1161	0.5210	-3.4083	0.1280	0.1043	0.5296
FTIN	Tindangou	Burkina Faso	50	19	0.905	0.0188	6.38	0.0519	0.05	-0.4501	0.3580	-2.4943	0.1940	0.0921	0.3520
FABA	Abala	Niger	26	13	0.932	0.0190	6.48	0.0170	0.61	-0.4316	0.3620	-1.1951	0.3090	0.1078	0.3534
FADE	Ader	Niger	50	15	0.907	0.0144	4.93	0.0499	0.23	-0.7930	0.2500	-1.3229	0.3400	0.0822	0.2454
FBAL	Balatungur	Niger	23	11	0.893	0.0100	6.22	0.0219	0.76	0.3454	0.6850	-0.3713	0.4480	0.1374	0.6602
FDIF	Diffa	Niger	32	23	0.980	0.0229	7.80	0.0084	0.80	-1.1637	0.1130	-8.9342	0.0070*	0.0773	0.0976
FZIN	Zinder	Niger	34	20	0.954	0.0223	7.60	0.0133	0.64	-0.8735	0.2040	-4.6555	0.0490	0.0832	0.1520
FTCH	Tcheboua	Cameroon	40	21	0.953	0.0207	7.07	0.0197	0.39	-0.8723	0.2040	-4.7844	0.0640	0.0826	0.1912
FBON	Bongor	Chad	50	27	0.934	0.0200	6.82	0.0190	0.28	-0.5110	0.3550	-9.4858	0.0100*	0.0899	0.3266
FLIN	Linia	Chad	27	17	0.952	0.0244	8.34	0.0205	0.39	-0.3052	0.4280	-3.1032	0.1320	0.1113	0.4212

Farmers Code	Ethnic gr.	Country	<i>n</i>	<i>k</i>	Hs	<i>II</i>	Dii	<i>r</i>	<i>P(r)</i>	<i>D</i>	<i>P(D)</i>	Fs	<i>P(Fs)</i>	<i>R</i> ₂	<i>P(R</i> ₂ <i>)</i>
SONG	Songhai	Mali	31	25	0.981	0.0235	8.02	0.0242	0.06	-0.8026	0.2290	-12.7252	0.0000*	0.0874	0.1678
BUDU	Buduma	Niger	30	22	0.968	0.0219	7.46	0.0180	0.28	-1.0955	0.1430	-8.9027	0.0050*	0.0788	0.0856
KANU	Kanuri	Nigeria	59	48	0.991	0.0224	7.63	0.0121	0.79	-1.2964	0.0760	-24.9338	0.0000*	0.0615	0.0734
KANE	Kanembou	Chad	50	37	0.988	0.0258	8.79	0.0059	0.74	-1.1163	0.1380	-21.2632	0.0000*	0.0707	0.1128
BULA	Bulahay	Cameroon	51	30	0.967	0.0215	7.33	0.0164	0.16	-0.9148	0.1830	-12.2460	0.0000*	0.0774	0.1918
FALI	Fali	Cameroon	42	25	0.943	0.0236	8.04	0.0122	0.50	-1.2000	0.0870	-7.3971	0.0110*	0.0719	0.0914
HIDE	Hide	Cameroon	42	31	0.986	0.0266	9.08	0.0060	0.78	-0.9804	0.1610	-14.5501	0.0000*	0.0799	0.1630
KOTO	Kotoko	Cameroon	69	35	0.959	0.0223	7.59	0.0151	0.19	-1.0903	0.1170	-13.5743	0.0010*	0.0670	0.1344
MAFA	Mafa	Cameroon	55	35	0.978	0.0230	7.84	0.0088	0.86	-1.4615	0.0480*	-17.3338	0.0000*	0.0579	0.0394*
MASA	Masa	Cameroon	41	34	0.990	0.0225	7.66	0.0062	0.95	-0.8974	0.2120	-23.9985	0.0000*	0.0809	0.1688

NOTE.—*n*, sample size; *k*, number of different haplotypes; Hs, gene diversity; *II*, nucleotide diversity; Dii, mean number of pairwise differences; *r*, Harpending's (raggedness) index; *P(r)*, probability of Harpending's index; *D*, Tajima's test of selective neutrality; *P(D)*, probability of *D* statistics; Fs, Fu's test of selective neutrality; *P(Fs)*, probability of Fs statistics; *R*₂, *R*₂ statistic, *P(R*₂*)*, probability of *R*₂ statistics; and the numbers with asterisks are statistically significant at 5% level.

(in number of nucleotides for mtDNA and in number of repeats for Y-STRs) and compared the values obtained for these measures in pastoral and farmer populations using the Wilcoxon test. For mtDNA, we also calculated nucleotide diversity, Harpending's (raggedness) index, and its probability, as well as three tests of selective neutrality estimating past population growth, such as Tajima's *D* (Tajima 1989), Fu's *F*_s (Fu 1997), and Ramos-Onsins and Rozas's *R*₂

(Ramos-Onsins and Rozas 2002). All the parameters and tests except for *R*₂ (calculated using DnaSP software, version 5.10.01; Rozas and Rozas 1999) were performed using Arlequin software, version 3.5.1.2 (Excoffier et al. 2005).

To evaluate genetic structure at the interpopulation level, pairwise genetic distances based on haplotype frequencies (*F*_{ST}) or on haplotype frequencies and the squared difference between STR alleles (*R*_{ST}) were obtained with

Table 2. Y Chromosome STR Diversity of Populations Practicing Different Lifestyles From the West African Sahel.

Pastoralists Code	Place	Country	<i>n</i>	<i>k</i>	Hs	SE Hs	Dii
FDIA	Diafarabe	Mali	24	19	0.9710	0.0239	3.540
FBAN	Banfora	Burkina Faso	26	16	0.9292	0.0347	3.526
FTIN	Tindangou	Burkina Faso	12	7	0.8939	0.0627	2.970
FABA	Abala	Niger	21	10	0.8143	0.0809	3.605
FADE	Ader	Niger	21	9	0.7238	0.1011	3.414
FBAL	Balatungur	Niger	13	10	0.9487	0.0506	4.115
FDIF	Diffa	Niger	14	10	0.9560	0.0377	4.538
FZIN	Zinder	Niger	17	16	0.9926	0.0230	4.566
FTCH	Tcheboua	Cameroon	22	13	0.9221	0.0381	4.126
FBON	Bongor	Chad	25	15	0.9500	0.0237	3.710
FLIN	Linia	Chad	11	8	0.9455	0.0535	3.527

Farmers Code	Ethnic gr.	Country	<i>n</i>	<i>k</i>	Hs	SE Hs	Dii
SONG	Songhai	Mali	24	19	0.9819	0.0164	4.337
BUDU	Buduma	Niger	15	6	0.7714	0.0979	3.362
KANU	Kanuri	Nigeria	29	19	0.9360	0.0312	4.076
KANE	Kanembou	Chad	12	11	0.9848	0.0403	5.258
BULA	Bulahay	Cameroon	25	13	0.7767	0.0886	1.397
MAFA & HIDE	Mafa & Hide	Cameroon	10	4	0.7333	0.1005	1.222
KOTO	Kotoko	Cameroon	22	18	0.9827	0.0183	4.844
MASA	Masa	Cameroon	21	14	0.9143	0.0500	3.419

NOTE.—*n*, sample size; *k*, number of different haplotypes; Hs, gene diversity; SE Hs, standard error of gene diversity; and Dii, Mean number of pairwise differences.

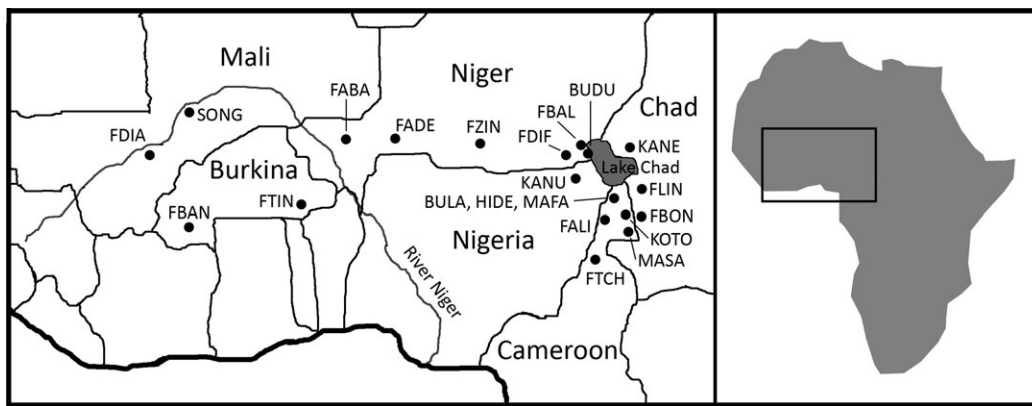


Fig. 1. Geographical distribution of the population samples in the West African Sahel. Only the countries from which the samples were secured are labeled on the map.

Arlequin. When partitioning the total variation, the R_{ST} index takes into account the squared number of repeat differences between two haplotypes. This is a relevant measure inasmuch it is linearly related to the expected coalescence time of the alleles compared (Slatkin 1995). Although we always calculated both indexes in Y-STR analyses, we preferentially present R_{ST} data to render the effect of the phylogenetic divergence between haplotypes and not simply their difference in state. This choice was supported by a significant ($P = 0.05$ for the global statistics and $P = 0.02$ for the average of pairwise statistics) contribution of allele sizes to genetic differentiation, as determined using the method of Hardy (Hardy et al. 2003). Multidimensional scaling (MDS) of F_{ST} and R_{ST} genetic distances was obtained by means of Statistica software version 5.5 (StatSoft, Inc.). Geographic distances were calculated based on latitude and longitude data using the online *gpsvisualizer* program. The Mantel test was used to assay the significance of the correlation between genetic and geographic distances with 10,000 permutations.

The absolute number of migrants exchanged between population pairs (M) was estimated for both mtDNA and Y-STR by transforming the matrix of pairwise F_{ST} (without interallele distances) according to the formula $M = (1 - F_{ST}) / (2 \times F_{ST})$. The values were rendered by using increasing line thickness for three classes of values.

The sharing of haplotypes between the pastoral and sedentary groups was calculated by Arlequin and visualized by the Network program version 4.516 (Bandelt et al. 1995). For mtDNA, the reduced median algorithm was used, with a reducing factor of two, followed by the median joining algorithm to resolve intermediate nodes. The fastest sites (16093, 16129, 16189, 16311, and 16362) were downweighted to 5 from an overall value of 10 for all the remaining sites (Soares et al. 2009). For Y-STRs, the median joining algorithm with the >1 criterion was used, and each locus was given a weight inversely proportional to its repeat variance.

Results

mtDNA

The pastoral groups exhibit consistently lower levels of intrapopulation diversity measures including gene diversity,

nucleotide diversity, and mean number of pairwise differences (table 1). We have examined these differences using the Wilcoxon test and confirmed their statistical significance (two-sided P value for nomads vs. farmers = 0.002 for gene diversity, 0.005 for nucleotide diversity, and 0.006 for pairwise differences). Excepting the pastoral group from Banfora, no other population in our sample shows a statistically significant Harpending's (raggedness) index.

Departure from the neutral model of populations evolving at constant size in mutation-drift equilibrium with no selection was tested by employing D , R_2 , and F_s statistics. With one exception (the Mafa), all Tajima's D and R_2 tests show no departure from neutrality expectations. However, F_u 's F_s statistics showed higher discriminatory power, revealing a significant departure from neutrality, compatible with population expansion in all the farmer populations analyzed in this study. In the pastoralists (table 1), however, significant departure from neutrality was restricted to only 3 of 11 groups.

The correlations between genetic (F_{ST}) and geographic distances in both farmer and pastoral populations are shown in figure 2a and b, respectively. No statistically significant correlation is observed for the pastoralists ($R^2 = 0.0474$; $P = 0.891$), but it is positive and statistically significant for the sedentary farmers ($R^2 = 0.4848$; $P = 0.006$).

Mutual genetic distances among all populations displayed by MDS analysis (fig. 3a) localized all pastoral groups in the left part of the graph along with only one "intrusive" farmer population (the Songhai). All other farmers are situated in the right part of the graph with similar distances for at least the Lake Chad Basin populations.

The analysis of molecular variance results showed that when grouping populations into pastoralists and farmers, 92.35% of the variation is observed within populations, 1.84% between populations/within groups, and 5.81% between groups, thus confirming a greater heterogeneity between than within groups. The same is also confirmed by the F_{CT} value of 0.058 (variation between groups) compared with the F_{SC} of 0.019 (variation between populations within groups). Individually, farmers displayed a higher F_{ST} (0.024) than the nomads (0.014), testifying to the greater heterogeneity of the former.

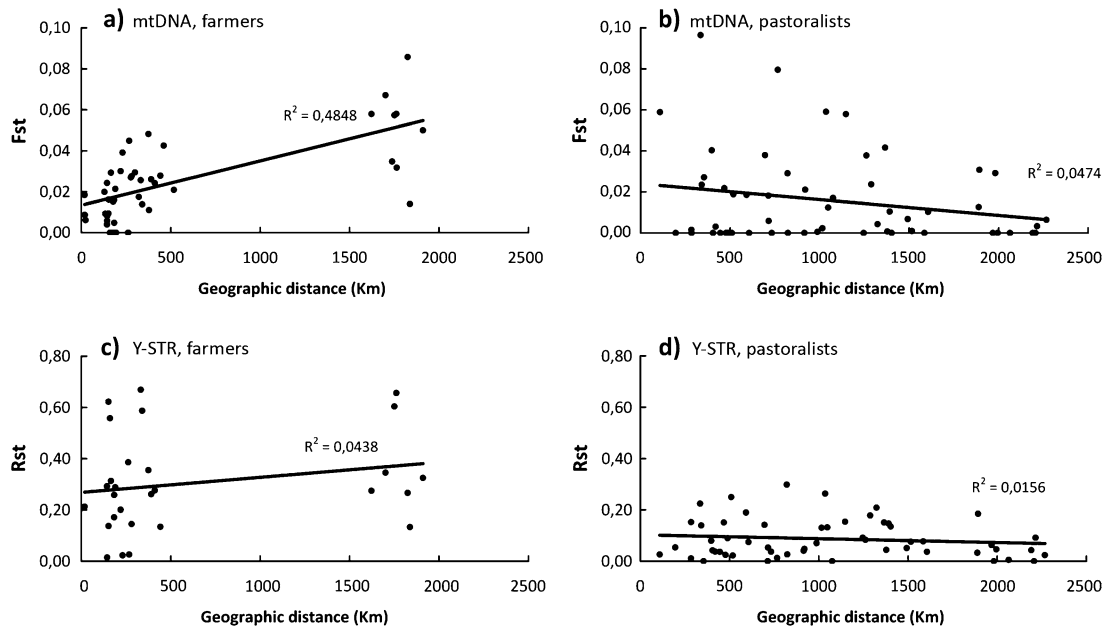


Fig. 2. Plots of geographic and genetic (F_{ST} for mtDNA and R_{ST} for Y-STR) distances for farmer (a, c) and pastoral (b, d) groups.

As genetic affinity can be maintained by gene flow between populations, we provide [figure 4A](#), where the estimated number of migrants between pairs of population is obtained by transforming the F_{ST} matrix. From the results, intensive gene flow in mtDNA can be inferred between each pair of pastoral populations and between some pastoralist groups and the Songhai. Strong gene flow is also apparent among several sedentary groups, especially the Masa, but some of them such as the Buduma and Kanembou appear very isolated in this regard.

Analysis of the mtDNA haplotype distribution revealed another interesting point. Although the majority of the haplotypes harboured in pastoral groups is shared, the farmer groups bear rather differing haplotypes. In [supplementary figure S1](#), [Supplementary Material](#) online, 432 “pastoral” sequences (in gray) are clustered into a much lower number of haplotypes than the 470 “farmer” sequences (in white). One important issue is that each of the most

frequent haplotypes in the network is shared between the groups. The nine most common haplotypes in the network are each found in more than 15 individuals and are shared in 7 of 11 pastoral groups. In total, these haplotypes correspond to about 70% of the mtDNA gene pool in pastoralists as a whole. The mtDNA haplotypes and how they are shared among the analyzed populations are shown in the [supplementary table S1](#), [Supplementary Material](#) online.

Y-STRs

The parameters of intrapopulation diversity are reported in [table 2](#). Four populations stand out for having a lower level of gene diversity (<0.80) attributable to the presence of a predominant haplotype (see [supplementary table S2](#), [Supplementary Material](#) online). Even though there is no significant difference in gene diversity between the two groups of different lifestyle (Wilcoxon test $P = 0.804$), three of the four instances of low diversity occur in the farmer

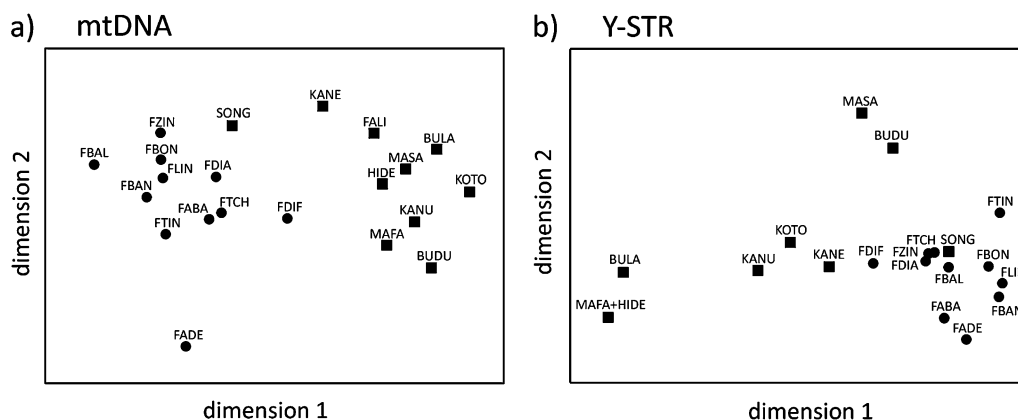


Fig. 3. MDS plot based on genetic distances (F_{ST} for mtDNA and R_{ST} for Y-STR) for the populations studied in the present article; the stress value is 0.083 for mtDNA and 0.132 for Y-STR. Farmers are marked by squares, pastoralists by circles.

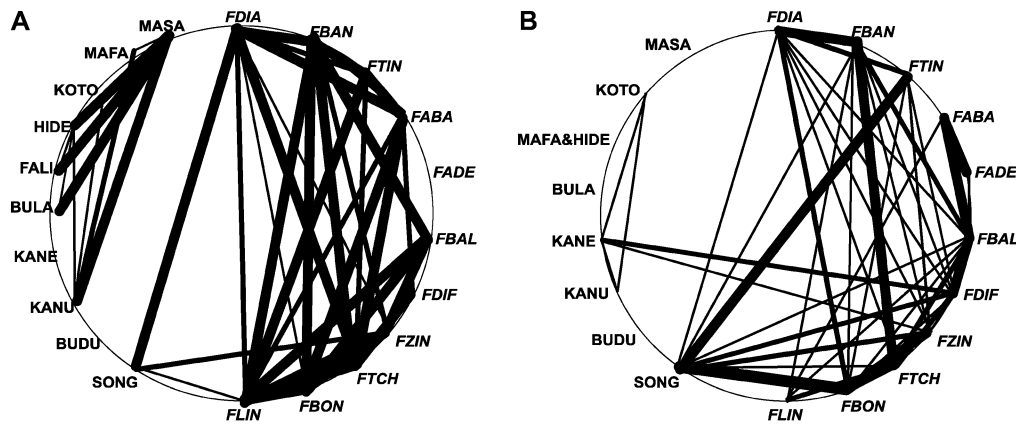


Fig. 4. Graphical representation of pairwise population estimates of migration, as obtained from the transformed F_{ST} matrices for mtDNA (A) and Y-STRs (B). Thin, medium, and thick lines represent values of N_m (absolute numbers of migrants) > 50 , > 100 and infinite ($F_{ST} = 0$) for mtDNA, and > 10 , > 50 , and infinite for Y-STRs, respectively. Codes for pastoralists are in italics.

populations. Haplotype diversity is not paralleled by the mean number of pairwise differences, which is 3.414 and 3.362 in the FADE and the BUDU, respectively, whereas dropping to 1.397 and 1.222 in the BULA and the MAFA and HIDE, respectively. This result becomes clearer when considering the phylogenetic relationships between haplotypes contributing to gene diversity (see below). Overall, the mean number of pairwise differences does not differ significantly between pastoralists and farmers (Wilcoxon test $P = 0.441$).

The relationship between genetic (R_{ST}) and geographic distances in both farmer and pastoral populations are shown in figure 2c and d, respectively. In contrast to mtDNA, no statistically significant correlation was observed. R_{ST} values (y axis) are generally lower among the pastoralists. Six of the pairwise R_{ST} values are greater than 0.5 in farmers and are observed both between populations living at short distances (left side of the plot) and as far as $\sim 1,800$ km apart. Of the six pairwise comparisons producing such high values, three involved one population with low diversity and three involved two populations with low diversity. In all cases, no more than a single haplotype was shared between the populations compared.

A summary representation of pairwise R_{ST} as obtained by MDS is displayed in figure 3b. Although the pastoral populations form a tight cluster on the bottom right of the plot, the farmers are much more dispersed, lying from the top centre to the bottom left. Compared with the F_{ST} plot (not shown here), the R_{ST} plot was better able to distinguish a subset of the farmers, such as Kanuri, Kotoko, Kanembou, and one group of Fulani pastoralists from Diffa. In agreement with the mtDNA results, the Songhai are the only farmers genetically close to the pastoralists.

In the Y-STR migration matrix (fig. 4b), the intensive gene flow among pastoralists mirrors that seen with mtDNA. Conversely, only a minimal exchange among farmers was apparent, with the Kanembou displaying consistent male-driven gene flow.

When the populations were partitioned according to their lifestyle (supplementary table S3, Supplementary Material online, first three lines, left) the index F_{CT} mea-

suring the contribution of the distinction between groups to the total variance was 0.131. Residual heterogeneity among populations within groups produced an F_{SC} of 0.138. The R statistics, which also weights the number of repeats separating alleles, produced an increase of both fixation indexes (0.208; supplementary table S3, Supplementary Material online, first three lines, right). This denotes that variation in STR numbers between alleles contributes to interpopulation and intergroup differences, again stressing the relevance of evolutionary diversification between haplotypes found in different populations. However, heterogeneity between populations was not distributed equally in both groups. In particular (supplementary table S3, Supplementary Material online, central and bottom lines, left) F_{ST} was much lower among pastoral (0.049) than sedentary (0.241) populations. Furthermore, consideration of interallele size differences resulted in a nearly doubled fixation index among nomadic populations (0.084), but only a 34% increase among sedentary populations (0.322).

Further insights into the composition of the male gene pool of the populations examined here were gained by considering the phylogenetic relationships between Y-STR haplotypes. Among the eight-loci haplotypes studied in the 364 males, 100 were private, that is, observed in only one individual. Among the remaining 264 individuals (73% of total), 54 different nonprivate haplotypes were observed. The network connecting these 54 shared haplotypes is shown in supplementary figure S2, Supplementary Material online, with each plot displaying which haplotype is found in the corresponding population. The overall network consists of three clusters, connected by median vectors, and positioned in the top, right, and bottom portions of the graph, respectively (right-bottom panel in supplementary fig. S2, Supplementary Material online). There is great variation in the occurrence of haplotypes belonging to each cluster in the two lifestyle populations. In most cases, haplotypes of the second and third cluster are represented among pastoralists (11 plots at the top of supplementary fig. S2, Supplementary Material online) with sporadic

representatives of the first cluster. Conversely, among the farmers, haplotypes of the first cluster are very common and are the sole ones represented in the Bulahay and in the pooled Mafa–Hide group. This accounts for the low levels of both haplotype diversity and mean number of pairwise differences in these samples. Conversely, in cases where modal- and low-frequency haplotypes belong to different clusters, the latter ones contribute disproportionately to mean pairwise differences. Among farmers, the Songhai represents the only group lacking these haplotypes, in line with their general similarity with the Fulani pastoralists.

Discussion

When comparing our results with those from Central Asia (Chaix et al. 2007), a rather different pattern is observed. The Y chromosome gene pool of the African pastoralists is more diverse than that of farmers. We argue that in spite of the small number of contemporary pastoralists in the Sahel, the shift to pastoralism in Africa might have involved a much larger ancestral population than the shift to farming. This is in clear contrast with Central Asian pastoralists where the male effective population size is smaller in contrast to their farmer neighbors (Chaix et al. 2007; Segurel et al. 2008). Substantial gene flow between the constantly mobile pastoralist groups maintained their initial Y chromosome genetic diversity and did not disturb their linguistic integrity: only one dominant language (“Fulfulde”) characterizes these peoples at the present time. However, the mtDNA gene pool of the Fulani pastoralists is composed of a much lower number of haplotypes (even though shared through large distances) than the farmers and might reflect a recent female population bottleneck, similarly as in hunter gatherers (Excoffier and Schneider 2000). Conversely, among the patrilineal and patrilocal farmers, where many different languages are represented, a stronger male-specific drift and a broader repertoire of mtDNA haplotypes denote possible recruitment of females across linguistic barriers in a distance-dependent manner.

Unfortunately, almost nothing is known about matrimonial exchanges between the pastoralists and farmers from the West African Sahel. Ethnographic studies report only cases within a specific family Fulani clan (Stenning 1959; Bonfiglioli 1988) or deal with matters of economy (Beauvillain 1977). Our study shows that the sharing of neither mtDNA nor NRY haplotypes is very common between these groups (supplementary figs. S1 and S2, Supplementary Material online), and so mating exchanges must have been rather limited. One exception is the Songhai population from Mali. With regards to the fact that the Fulani nomads might recently (in the 11th century AD) have spread to Lake Chad from West Africa, the relationship between these groups can be explained by the harboring of common ancestral haplotypes. Unfortunately, farmer population samples from the area between Lake Chad Basin and Middle Niger Delta were not available to us. Some mtDNA data sets exists in the literature (Watson et al. 1997), but we relied only on the samples secured directly in the field using our collection criteria.

Considering the large extent of mtDNA haplotype sharing between the pastoral groups, two scenarios emerge: a first in which these groups approximate a single population on the maternal side of the gene pool (indicating extremely large gene flow between groups throughout the African Sahel), and a second where the current repertoire of mtDNA haplotypes represents a pool of common ancestral haplotypes. This latter case would imply that genetic drift was not very strong in the pastoral groups. In fact, both scenarios can explain the observed results. According to the archaeobotanical evidence, farming in the West African Sahel is a later phenomenon than pastoralism (Neumann 2003), and important gene flow within the ancestral population of the contemporary pastoralists might thus have been interrupted only quite recently. On the other hand, among the farmers (speaking various languages), divergence of their pool(s) can have been achieved earlier and more easily, maintaining a lower gene flow among their subpopulations.

The Y chromosome gene pool of the pastoralists mainly consists of two distantly related haplotype sets. Their differentiation in terms of repeat numbers suggests an antiquity predating herding practices. This composition can result either from a dual origin or from a large ancestral effective size in a stationary population. In the latter case, the two long branches of the coalescent tree would have survived until today, but the lack of many Fulani-specific haplotypes would necessitate the postulation of a continuous exchange with other external populations.

Based on comparison with works reporting joint STR and single nucleotide polymorphism data in similar African populations (Goncalves et al. 2005; Rosa et al. 2007; Tishkoff et al. 2007; Berniell-Lee et al. 2009; Cruciani et al. 2010a) and also including the three modal haplotypes, a large part of the corresponding clusters can be tentatively affiliated with haplogroups R1b-V88, E3b-M35, and E3a-M2, respectively. Maintenance of haplotypes of the second and third clusters in all pastoral groups is compatible with high levels of gene flow and/or a common origin. The presence of a few haplotypes of the first cluster in nomads is very interesting. In fact, these too are deeply differentiated and are similar to those found in high frequency in populations that acquired them after herding practices were already established in the Sahel (Cruciani et al. 2010a). This component of the male-specific pool thus might testify to the occasional recruitment of males into the pastoralist population. The Y-STR pattern among farmers is radically different, their gene pools being dominated by haplotypes of the first cluster. The populations are differentiated by the frequencies that various haplotypes have reached within each of them (large F_{ST}), irrespectively of the geographical distance separating them. This pattern is in agreement with genetic drift acting within the male component of each population, accompanied by scarce male-driven gene flow.

The West African Sahel pastoralists and Lake Chad Basin farmers are well separated by the MDS analyses for both mtDNA and Y-STR. The Songhai living in the Middle Niger Delta are currently classified to the Nilo-Saharan linguistic

family that is spread and diversified mainly in East and Central Africa. However, according to some linguists, the attribution of the Songhai to this group is very contentious apart from the fact that the whole Nilo-Saharan family is very heterogeneous (Bender 2000). We explored the F_{ST} relationship between the Songhai (both our and Songhai sample of Watson et al. (1996)) and 121 African populations in our mtDNA data set (data not shown) and found nonsignificant pairwise differences preferentially with West African populations such as the Guineans, Wolof, and Serer and also with some groups of the central Sahel such as the Hausa, Yoruba, Fulani, Mandara, Ouldeme, Kanuri, and even some Tuareg groups. By contrast, the comparisons of Songhai with Northern, Eastern, and Southern African populations always revealed significant differences. It thus seems that the genetic similarity of the Fulani and Songhai can be better explained by their common ancestry in West or West central Africa.

Our results, together with the West African influences of Fulani pastoralists, support the general idea of the African Sahel acting as a bidirectional corridor along its horizontal axis (Bereir et al. 2007; Černý et al. 2007; Hassan et al. 2008; Tishkoff et al. 2009). Peoples such as the Mafa, Hide, and Bulahay can be taken as one ethnic group speaking slightly different dialects (David et al. 1991). These groups, living in the Mandara Mountains together with the Masa and Kotoko who live along the Shari River, belong to the Chadic branch of the Afro-Asiatic family. Indeed, they account for a major proportion of interpopulation differentiation, as far as Y-STR variability is concerned (supplementary table S3, Supplementary Material online), indicating a parallel genetic and cultural peculiarity of this cluster of populations. The reconstruction of the original homeland of Chadic speakers based on genetic evidence is disputed (Cruciani et al. 2010a; Lancaster 2010). An East African origin is favored for their similarity with some Cushitic groups in Somalia and by the phylogeny of L3f3 mtDNA haplogroup (Černý et al. 2009), largely represented in our series. Perhaps, the recent founder effect together with subsequent demographic isolation in the Mandara Mountains can help in explaining the reduced genetic diversity of these “Montagnard” (the Mafa, Hide, and Bulahay) groups. However, Cruciani et al. (Cruciani et al. 2002, 2010a, 2010b) described the NRY lineage R1b-V88 as associated with the spread of Chadic languages within Africa. This lineage was also found in North Africa and thus favors a northern route to the Lake Chad Basin. The most common NRY STR haplotypes found in our sedentary populations share the 15-11-13 motif at DYS19-DYS391-DYS393 with haplotypes found on this lineage. It is worth noting that this phylogeographic uncertainty replicates that based on linguistic grounds: the “trans-Saharan” versus “inter-Saharan” hypotheses (Blench 2006).

Supplementary Material

Supplementary tables S1–S3 and figures S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The project was supported by the Grant Agency of the Czech Republic (Grant no. 206/08/1587), the Portuguese Fundação para a Ciência e a Tecnologia (project—PTDC/ANT/66275/2006; post-doc grant—SFRH/BPD/64233/2009), and the Italian Ministry of University (MIUR-PRIN2007). Instituto de Patologia e Imunologia Molecular da Universidade do Porto is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT, the Portuguese Foundation for Science and Technology.

References

- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 23:147.
- Ba AH, Dieterlen G. 1966. Les fresques d'époque bovidienne du Tassili-n-Ajjer et les traditions des Peuls. *Journal de la Société des Africanistes.* 36:141–157.
- Balaresque P, Jobling MA. 2007. Human populations: houses for spouses. *Curr Biol.* 17:R14–R16.
- Bandelt H-J, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753.
- Beauvillain A. 1977. Les Peuls du Dallol Bosso. Centre National de Recherche Scientifiques et Humaines. Niamey: Edition Études Nigériennes, No. 42.
- Bender LM. 2000. Nilo-Saharan. In: Heine B, Nurse D, editors. African languages: an introduction. Cambridge: Cambridge University Press. p. 44–74.
- Bereir RE, Hassan HY, Salih NA, Underhill PA, Cavalli-Sforza LL, Hussain AA, Kwiatkowski D, Ibrahim ME. 2007. Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel. *Eur J Hum Genet.* 15:1183–1185.
- Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mougouia-Daouda P, van der Veen L, Hombert JM, Quintana-Murci L, Comas D. 2009. Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol.* 26:1581–1589.
- Blench R. 2006. Archaeology, languages, and the African past. Lanham (MD): Alta Mira Press.
- Bonfiglioli AM. 1988. Duda: histoire de famille et histoire de troupeau chez un groupe de Wodaabe du Niger. Paris: Editions de la Maison des sciences de l'homme.
- Botte R, Boutrais J, Schmitz J. 1999. Figures peules. Paris (France): Karthala.
- Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. 2002. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int.* 129:10–24.
- Carracedo A, Beckmann A, Bengs A, et al. (28 co-authors). 2001. Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format. *Forensic Sci Int.* 119:28–41.
- Černý V, Fernandes V, Costa MD, Hájek M, Mulligan CJ, Pereira L. 2009. Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evol Biol.* 9:63.
- Černý V, Hájek M, Bromová M, Čmejla R, Diallo I, Brdička R. 2006. MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum Biol.* 78:9–27.
- Černý V, Hájek M, Čmejla R, Brůžek J, Brdička R. 2004. mtDNA sequences of Chadic-speaking populations from northern

- Cameroon suggest their affinities with eastern Africa. *Ann Hum Biol.* 31:554–569.
- 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. *Ann Hum Genet.* 71:433–452.
- Chaix R, Quintana-Murci L, Hegay T, Hammer MF, Mobasher Z, Austerlitz F, Heyer E. 2007. From social to genetic structures in central Asia. *Curr Biol.* 17:43–48.
- Cruciani F, Santolamazza P, Shen P, et al. (16 co-authors) 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet.* 70:1197–1214.
- Cruciani F, Trombetta B, Sellitto D, Massaia A, Destro-Bisol G, Watson E, Beraud Colomb E, Dugoujon J-M, Moral P, Scozzari R. 2010a. Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur J Hum Genet.* 18:800–807.
- Cruciani F, Trombetta B, Sellitto D, Massaia A, Destro-Bisol G, Watson E, Beraud Colomb E, Dugoujon JM, Moral P, Scozzari R. 2010b. Reply to Lancaster. *Eur J Hum Genet.* 18:1186–1187.
- David N. 1971. The Fulani compound and the archaeologist. *World Archaeol.* 3:111–131.
- David N, Gavua KB, MacEachern AS, Sterner JA. 1991. Ethnicity and material culture in North Cameroon. *Can J Archaeol.* 15:171–177.
- Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, Tofanelli S, Spedini G, Capelli C. 2004. Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol Biol Evol.* 21:1673–1682.
- Diallo Y, Schlee G. 2000. L'ethnité peule dans des contextes nouveaux: la dynamique des frontières. Paris (France): Karthala. p. 255 p., [258] p. of plates.
- Dupire M. 1962. Peuls nomades: Étude descriptive des Wodaabe du Sahel nigérien. Paris (France): Karthala.
- Dupuy C. 1999. Les apports de l'archéologie et de l'ethnologie à la connaissance de l'histoire ancienne des Peuls. In: Botte R, Boutrais J, Schmitz J, editors. Figures peules. Paris (France): Karthala. p. 53–72.
- Excoffier L, Schneider S. 1999. Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. *Proc Natl Acad Sci U S A.* 96:10597–10602.
- Excoffier L, Schneider S. 2000. The demography of human populations inferred from patterns of mitochondrial DNA diversity. In: Renfrew C, Boyle C, editors. Archaeogenetics: DNA and the population prehistory of Europe. Cambridge: McDonald Institute for Archaeological Research. p. 101–108.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1:47–50.
- Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925.
- Gauthier J-G. 1979. Archéologie du pays Fali, Nord Cameroun. Paris (France): Editions du CNRS.
- Goncalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, Underhill PA, Kivisild T, Brehm A. 2005. Y-chromosome lineages from Portugal, Madeira and Acores record elements of Sephardim and Berber ancestry. *Ann Hum Genet.* 69:443–454.
- Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilineal populations than in matrilineal populations. *Proc Natl Acad Sci U S A.* 102:7476–7480.
- Hardy OJ, Charbonnel N, Freville H, Heuertz M. 2003. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* 163:1467–1482.
- Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME. 2008. Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history. *Am J Phys Anthropol.* 137:316–323.
- Holl A. 2004. Saharan rock art: archaeology of Tassilian pastoralist iconography. Walnut Creek (CA): AltaMira Press.
- Jabbar MA, Reynolds L, Francis PA. 1995. Sedentarisation of cattle farmers in the derived savannah region of south-west Nigeria: results of a survey. *Trop Anim Health Prod.* 27:55–64.
- Lancaster A. 2010. Chadic languages and Y haplogroups. *Eur J Hum Genet.* 18:1185–1187.
- Le Quellec J-L. 1997. Art rupestre et préhistoire au Sahara. Paris (France): Payot.
- Lhote H. 1973. A la découverte des fresques du Tassili. Paris (France): Arthaud.
- Luca F, Bubba G, Basile M, Brdicka R, Michalodimitrakis E, Rickards O, Vershubsky G, Quintana-Murci L, Kozlov AI, Novelletto A. 2008. Multiple advantageous amino acid variants in the NAT2 gene in human populations. *PLoS One.* 3:e3136.
- Marchani EE, Watkins WS, Bulayeva K, Harpending HC, Jorde LB. 2008. Culture creates genetic structure in the Caucasus: autosomal, mitochondrial, and Y-chromosomal variation in Daghestan. *BMC Genet.* 9:47.
- Mohammed-Ali AS, Khabir A-RM. 2003. The wavy line and dotted wavy line pottery in the prehistory of the central Nile and the Sahara-Sahel belt. *Afr Archaeol Rev.* 20:25–58.
- Muzzolini A. 1992. Le Bovidien dans l'art rupestre Saharien, un Réexamen critique. *L'Anthropologie.* 96:737–758.
- Neumann K. 2003. The late emergence of agriculture in sub-Saharan Africa: archaeobotanical evidence and ecological considerations. In: Neumann K, Butler A, Kahlheber S, editors. Food, fuel and fields. Progress in African archaeobotany. Koln (Germany): Heinrich-Barth-Institute. p. 71–92.
- Newman JL. 1995. The peopling of Africa: a geographic interpretation. New Haven (CT): Yale University Press.
- Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat Genet.* 29:20–21.
- Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet.* 65:439–458.
- Pillington MM, Wilder JA, Mendez FL, et al. (13 co-authors) 2008. Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. *Mol Biol Evol.* 25:517–525.
- Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol.* 19:2092–2100.
- Rosa A, Ornelas C, Jobling MA, Brehm A, Villems R. 2007. Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC Evol Biol.* 7:124.
- Rozas J, Rozas R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175.
- Schultz E. 1984. From Pagan to Pullo: ethnic identity change in Northern Cameroon. *Africa: J Int Afri Instit.* 54:46–64.
- Segurel L, Martinez-Cruz B, Quintana-Murci L, et al. (11 co-authors) 2008. Sex-specific genetic structure and social organization in Central Asia: insights from a multi-locus study. *PLoS Genet.* 4:e1000200.
- Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.
- Smith AB. 2005. African herders: emergence of pastoral traditions. Walnut Creek (CA): AltaMira Press.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for

- purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet.* 84:740–759.
- Stenning DJ. 1959. Savannah nomads; a study of the Wodaabe pastoral Fulani of Western Bornu Province, Northern Region, Nigeria. London: Published for the International African Institute by Oxford University Press.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tishkoff SA, Gonder MK, Henn BM, et al. (12 co-authors) 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol.* 24:2180–2195.
- Tishkoff SA, Reed FA, Friedlaender FR, et al. (25 co-authors) 2009. The genetic structure and history of Africans and African Americans. *Science.* 324:1035–1044.
- Verdu P, Austerlitz F, Estoup A, et al. (14 co-authors) 2009. Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Curr Biol.* 19:312–318.
- Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Paabo S. 1996. mtDNA sequence diversity in Africa. *Am J Hum Genet.* 59:437–444.
- Watson E, Forster P, Richards M, Bandelt HJ. 1997. Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet.* 61:691–704.
- Wendorf F, Schild R, Close AE. 1984. Cattle-keepers of the eastern Sahara: the neolithic of Bir Kiseiba. Dallas (TX): Dept. of Anthropology [and] Institute for the Study of Earth and Man, Southern Methodist University.