



HAL
open science

Contrasting Maternal and Paternal Histories in the Linguistic Context of Burkina Faso

C. Barbieri, M. Whitten, K. Beyer, H. Schreiber, M. Li, Brigitte Pakendorf

► **To cite this version:**

C. Barbieri, M. Whitten, K. Beyer, H. Schreiber, M. Li, et al.. Contrasting Maternal and Paternal Histories in the Linguistic Context of Burkina Faso. *Molecular Biology and Evolution*, 2012, 29 (4), pp.1213-1223. 10.1093/molbev/msr291 . hal-01998849

HAL Id: hal-01998849

<https://hal.univ-lyon2.fr/hal-01998849>

Submitted on 16 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contrasting Maternal and Paternal Histories in the Linguistic Context of Burkina Faso

Chiara Barbieri,^{*1} Mark Whitten,¹ Klaus Beyer,² Henning Schreiber,³ Mingkun Li,⁴ and Brigitte Pakendorf¹

¹Max Planck Research Group on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

²Institut für Afrikanische Sprachwissenschaften, Johann Wolfgang Goethe-Universität, Frankfurt/Main, Germany

³Department of African Linguistics and Ethiopian Studies, University of Hamburg, Hamburg, Germany

⁴Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

***Corresponding author:** E-mail: chiara_barbieri@eva.mpg.de.

Associate editor: Sohini Ramachandran

Abstract

Burkina Faso is located in the heart of West Africa and is a representative of the local structured patterns of human variability. Here, different cultures and languages are found in a geographic contiguity, as a result of several waves of migration and the succession of long- and short-term empires. However, historical documentation for this area is only partial, focusing predominantly on the recent empires, and linguistic surveys lack the power to fully elucidate the social context of the contact-induced changes. In this paper, we report Y-chromosomal data and complete mtDNA genome sequences for ten populations from Burkina Faso whose languages belong to two very distantly related branches of the Niger–Congo phylum, the Gur and Mande language families. In addition, two further populations, the Mande-speaking Mandenka from Senegal and the Yoruba from Nigeria, were included for regional comparison. We focus on the different historical trajectories undergone by the maternal and paternal lineages. Our results reveal a striking structure in the paternal line, which matches the linguistic affiliation of the ethnolinguistic groups, in contrast to the near-complete homogeneity of the populations in the maternal line. However, while the ancient structure along the linguistic lines is apparent in the Y-chromosomal haplogroup affiliation, this has clearly been overlain by more recent migrations, as shown by significant correlations between the genetic distances based on Y chromosome short tandem repeats and geographic distances between the populations, as well as by the patterns of shared haplotypes. Using the complete mtDNA sequences, we are able to reconstruct population size variation in the past, showing a strong sign of expansion in the concomitance with the Holocene Climate Optimum approximately 12,000–10,000 years ago, which has been suggested as the cause of the spread of the Niger–Congo phylum in the area. However, subsequent climatic fluctuations do not appear to have had an impact on the demography of the inhabitants of West Africa, probably reflecting the adaptive advantages of cultural innovations, such as pastoralism and agriculture.

Key words: human, West Africa, Mande, Gur, mtDNA genomes, Y chromosome.

Introduction

Burkina Faso is situated in the center of sub-Saharan West Africa, sandwiched between the arid savannah of the Sahel in the north and the rainforests of the Sudan area in the south (Ouedba 1997). The area was sparsely populated until the end of the Pleistocene (Muzzolini 1993), around 12 kya. The following Holocene climate optimum approximately 10 kya represented a demographic turning point, favoring the spread of people and colonization of new territories (Alimen 1987; McIntosh and McIntosh 1988), with subsequent fluctuations in the climate resulting in north- and southward migrations depending on the sociocultural adaptations of the people (McIntosh and McIntosh 1988; Brooks 1989).

The major linguistic phylum present in sub-Saharan West Africa is Niger–Congo, which is the numerically largest as well as the most diverse and widespread phylum of the continent (Lewis 2009). While over all of Central Southern Africa, people speak only languages belonging

to the Bantu family, which is a low-level branch within Niger–Congo, in West Africa, a greater variety of Niger–Congo language families are found. Some authors have suggested a possible central West African origin and spread of the phylum, in concomitance with the Holocene climate, change roughly 10 kya (Dimmendaal 2008). This might have been enhanced by cultural and technical innovations like bow and arrow, poison for arrow tips, and dogs (Blench 2006); all improvements that are in accordance with a way of subsistence still based on hunting. Most of the languages spoken today in the area of Burkina Faso (63 out of 68, Lewis 2009) belong to one of two major families of the Niger–Congo phylum: either Mande or Gur (Voltaic). The first roughly covers the north and west, whereas the latter predominates in the south and east. The languages from these two major families are only distantly related: the Mande family, which is further divided into the West Mande and East Mande, is considered to be the

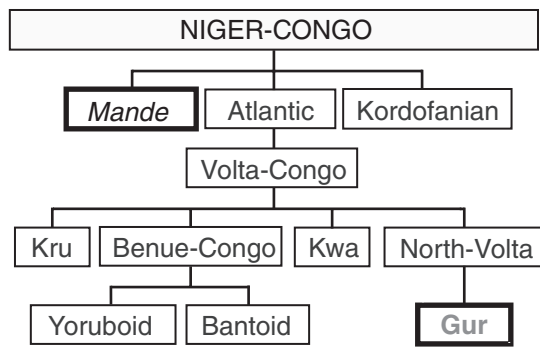


FIG. 1. Schematic tree of the Niger–Congo phylum showing the genealogical relationship between the Mande, Gur, and Benue–Congo families. (Modified from de Filippo C, Barbieri C, Whitten M, et al. By permission of Oxford University Press on behalf of the Society for Molecular Biology and Evolution.)

first split-off of the Niger–Congo phylum (Williamson 1989), and some authors even doubt its affiliation with the phylum itself (Mukarovsky 1966; Dimmendaal 2008). The Gur family, on the other hand, is one of the lower level divisions of the phylum, approximately at the same level as the Bantoid family to which the Bantu languages belong (Williamson and Blench 2000; see the schematic representation of the linguistic relationships in figure 1).

Previous studies of the West African populations have shown evidence for bidirectional east–west migrations following the Sahel Belt, as well as strong differentiation and lack of gene flow between sedentary agriculturalists and nomadic pastoralists (Černý et al. 2006, 2007, 2011; Pereira et al. 2010). However, the resolution of markers used in previous studies has generally been low, and only a small number of populations from Burkina Faso, such as the Mossi, Rimaibe, Fulani, and Tuareg, have been included in genetic studies focusing on the genetic variation in larger regions of Africa (Scozzari et al. 1997, 1999; Černý et al. 2006, 2011; Pereira et al. 2010).

This study is the first extended genetic survey conducted in Burkina Faso, with a focus on the central region of the country. The area of investigation is partially framed by the Black Volta river in the west, the White Volta in the east, the boundaries with Ghana in the south, and Mali in the north. In this territory, languages belonging to both the

Mande and the Gur families are spoken, and linguistic investigations show some sharing of features across family boundaries attributable to language contact (Beyer and Schreiber, Forthcoming). However, although it is obvious that the different ethnolinguistic groups in the area have been settled in close proximity for an extended period of time, the intensity and chronology of this contact situation is hard to judge based merely on linguistic evidence. Here, we investigate whether genetic variation correlates with linguistic affiliation to either of the two language families, and to what degree the contact discernible in the linguistic data is correlated with patterns of gene flow between populations speaking different languages. Using both whole mtDNA genome sequences as well as a substantial panel of Y-chromosomal markers, we compare the maternal and paternal histories of the populations under study. Furthermore, on a more fine-grained level, we compare the genetic variation in two linguistic islands: the Gur-speaking Pana in the northwest, who are surrounded by Mande-speaking populations, and the Mande-speaking Bisa in the southeast, who are isolated from their linguistic relatives by several Gur-speaking populations (cf. fig. 2). This permits us to elucidate whether linguistic or geographic proximity are better predictors of a genetic relationship.

Materials and Methods

Materials

Saliva samples were collected by M.W., H.S., and K.B. in two expeditions to Burkina Faso in March and October 2008, with the approval of the Ministry of Secondary and Higher Education and Scientific Research of Burkina Faso. After obtaining appropriate informed consent, a total of 352 samples from the following ten ethnolinguistic groups were collected: North Samo, South Samo, Marka, and Bisa, belonging to the Mande family, and Samoya, Pana, Mossi, Lyela, Nuna, and Kassena, belonging to Gur family (see, fig. 1 for a schematic representation of the linguistic relationship between the Mande and Gur families and fig. 2 for the geographic distribution of the sample populations). The ethnolinguistic affiliation of sample donors was ascertained through interviewing, with participants being asked about the place of birth and linguistic affiliation of their parents and grandparents. For both the Y-chromosome and the

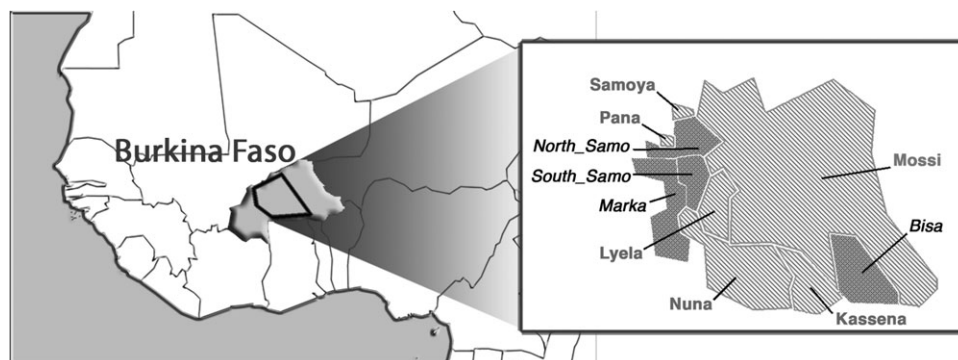


FIG. 2. Map of Western Africa showing the location of Burkina Faso and the ethnolinguistic groups included in the study. Mande populations are labeled in black italic font; Gur populations in gray bold font.

mtDNA analysis, related individuals and individuals of unclear or mixed ethnic affiliation were excluded. Because the previous studies of West African populations have as a rule only analyzed the mitochondrial hypervariable region and fewer Y-chromosomal single nucleotide polymorphisms (SNPs) and are thus not comparable to our data, we included only the Mandenka from Senegal and the Yoruba from Nigeria contained in the HGDP–CEPH panel (Human Genome Diversity Panel–Centre d’Étude du Polymorphisme Humain; Cann et al. 2002) for a regional comparison. The Mandenka belong to the Mande linguistic group (in particular, to the West Mande branch like the Marka from Burkina Faso) and are separated from our samples by a considerable geographic distance; the Yoruba, on the other hand, are linguistically more distantly related (speaking a language belonging to the Benue–Congo family of Niger–Congo) but live in relative proximity to other Gur speakers. Only unrelated individuals were included (Rosenberg 2006): 15 Mandenka and 12 Yoruba samples for the Y chromosome and 22 Mandenka and 22 Yoruba samples for the mtDNA analyses.

DNA Extraction and Y-Chromosome Analysis

DNA extraction from saliva, typing of Y-chromosome haplogroups via SnaPshot multiplex assays and typing of 12 Y-chromosome short tandem repeat (STR) loci with the Promega Y-Powerplex kit were described in de Filippo et al. (2011). After excluding individuals for whom DNA could not be successfully recovered, 362 individuals were typed for both SNP and STR data. In order to define sub-lineages of haplogroup B, the markers M181 (haplogroup B*), M150 (haplogroup B2a*), M152 (haplogroup B2a1a), and M112 (haplogroup B2b) were additionally typed in a separate multiplex SNaPshot assay. For details on the assay, see the [supplementary table 1, Supplementary Material](#) online.

mtDNA Sequencing

Genomic libraries were constructed by indexing and hybridization enrichment for mtDNA (Maricic et al. 2010) and subsequently sequenced on an Illumina GAIIx (Solexa) sequencer (for details of the sequencing protocol, see the [supplementary text, Supplementary Material](#) online). After excluding samples for which we could not get good quality sequence data, we recovered sequences from 335 unrelated individuals with maximum gaps of 0.2% and an average coverage of $100\times$ (with an average minimum of $17\times$ and an average maximum of $154\times$; the Mandenka and Yoruba from the CEPH panel were sequenced to an average of $900\times$). Gaps, sites of an uncertain base call (where a base was not present in $>70\%$ of the reads) and positions that had only $1\times$ coverage were considered as missing data; the missing data values for all samples were below 0.5%. The two poly-C regions (nucleotide positions 303–315 and 16183–16194), which cause severe sequencing errors, were deleted in all samples. All sequences were submitted to GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) and were given accession numbers JQ044791–JQ045125.

Data Analysis

For the Y-chromosomal and mtDNA data, analysis of molecular variance (AMOVA) and standard diversity indices for both haplogroups and STR haplotypes were computed using Arlequin ver. 3.11. For the STR analysis in Arlequin, a total of 14 samples with missing values for one or more loci were excluded. Nucleotide diversity and variance for the mtDNA sequence data in single populations were calculated in R with the function “nuc.div” of package *pegas* (Paradis 2010). Correspondence analysis (CA) of haplogroup frequencies in all populations was performed using the function “ca” from the R-package *ca* (Nenadic and Greenacre 2007). F_{st} , Φ_{st} , and R_{st} distances were calculated in Arlequin and plotted in R with a nonmetric multidimensional scaling (MDS) with function “isoMDS” from a package *MASS* (Venables and Ripley 2002). Y-chromosomal haplotype and mtDNA sequence sharing were estimated and plotted with in-house scripts for R.

Linguistic distance matrices were calculated on a preliminary word list of 46 items chosen to be historically informative, using unpublished data of Beyer and Kleinwillinghöfer for Gur, and data from Schreiber (2008; unpublished field materials), Creissels (2011), and Diallo (1988) for Mande. The linguistic distance between two languages was calculated in the standard lexicostatistic way as the proportion of shared cognates (i.e., words that are inherited in the respective languages from a common ancestor) subtracted from 1. Based on these distance matrices, neighbor-joining trees relating the Gur languages and Mande languages included in this study were constructed in R using the function “nj” of package *ape* (Paradis et al. 2004) and plotted with *FigTree* v1.3.1 (<http://tree.bio.ed.ac.uk>).

Mantel and partial Mantel tests between geographic, genetic, and linguistic distances were performed in R (package *Vegan*, Oksanen et al. 2011), using two approaches: First, the correlation between genetic and geographic distances (based on Global Positioning System data from the individual sampling locations) was investigated for all pairs of villages that were represented with a sample size of five individuals or more for the Y-STRs (34 villages included) and the mtDNA full sequence data (27 villages). Second, we investigated the correlation between the genetic distances between pairs of populations (Y-chromosomal haplogroup, STR haplotype, and full mtDNA sequence data) and geographic distances between the centroids of sampling locations for each population, calculated with the function “centroid” of the package *geosphere* (Hijmans et al. 2010). In addition, in a partial Mantel test, we investigated the correlation between the linguistic distances and the genetic distances between populations for both mtDNA and Y chromosome, keeping geography constant. Due to the deep time depth separating the Gur and Mande language families, the vocabularies of Mande and Gur languages are too diverged to permit the secure identification of Niger–Congo cognates; this makes the calculation of linguistic distances between them unfeasible. Therefore, the partial

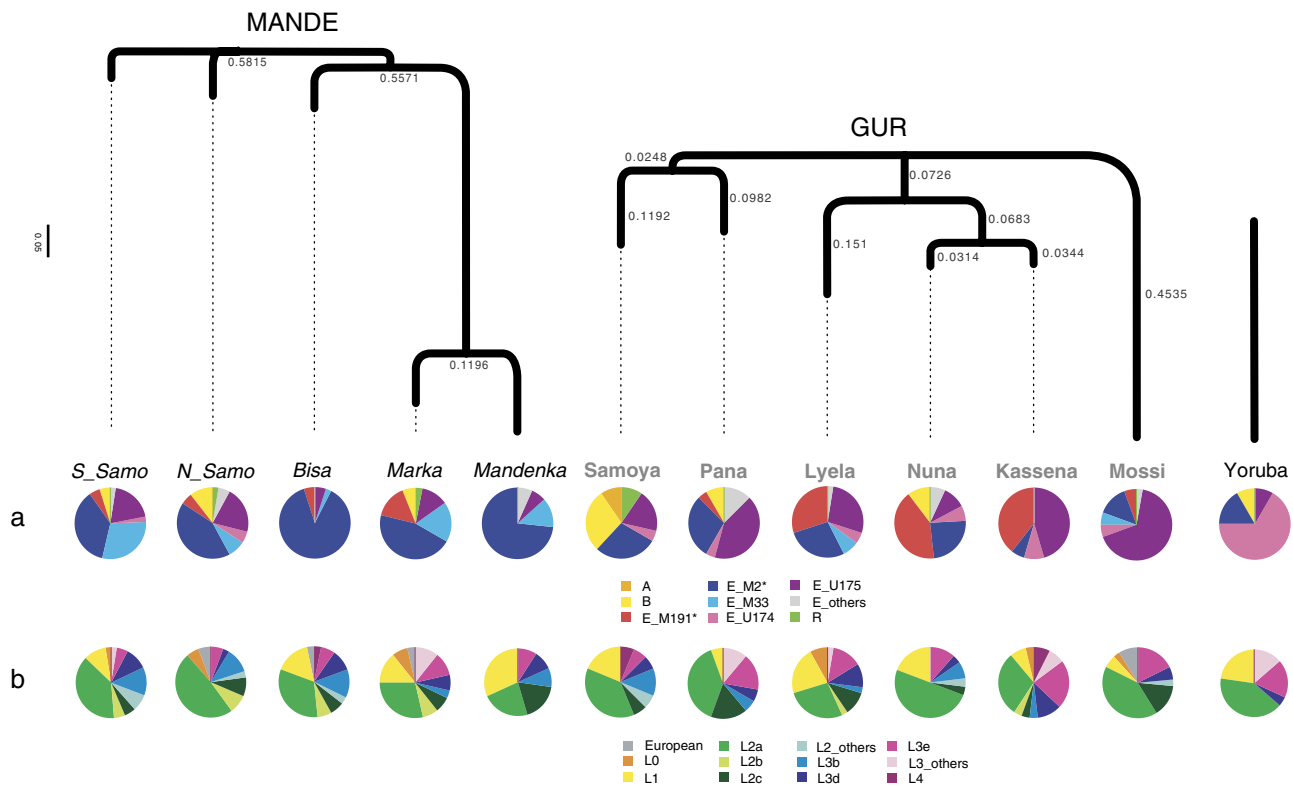


Fig. 3. Haplogroup composition of Mande and Gur populations as well as Yoruba, together with a Neighbour-Joining tree based on linguistic distances. (a) Y chromosome. (b) mtDNA. In the Y-chromosomal pie charts, haplogroups B-M150 and B-M181 were merged to B, and E-M35, E-M75 and E-M96 were merged to E_others; in the mtDNA pie charts, haplogroups L1b and L1c were merged to L1, L2d and L2e were merged to L2_others, L3f and L3h were merged to L3_others and U and H were merged to European.

Mantel analyses were performed separately for the Gur and Mande languages. The significance of the correlation was determined by 999 permutations.

Bayesian Skyline Plots (BSPs) were constructed for mtDNA sequence data (using only sequences belonging to the African haplogroups L1, L2, and L3) by means of BEAST (version 1.5.3), based on coding region sequences (position 577–16023, 15,447 bp). Prior to producing the plots, the best fitting substitution model was identified using Modeltest (version 3.7). A piecewise linear model was used with both a strict and relaxed clock and a rate of 1.691×10^{-8} for 50,000,000 generations, sampling every 1,000 steps (Atkinson et al. 2008). Because the relaxed clock model had higher likelihood scores, we report only these results.

For the AMOVA analysis based on geography and the haplotype sharing analysis, the Samoya, Pana, North Samo, South Samo, Marka, and Lyela were considered as “northern” populations, whereas the Mossi, Nuna, Kassena, and Bisa were considered “southern” populations, because three out of the four villages in which Mossi samples were collected lay in the south.

Results

Haplogroup Composition and Genetic Diversity

Figure 3 shows the frequencies of the major Y-chromosomal and mtDNA haplogroups in the 12 population groups included in the study. Overall, the populations from Burkina

Faso are characterized by high frequencies of Y-chromosome haplogroups E-M2*(xE-M191, xE-U175), E-M191*(xE-U174), and E-U175 (de Filippo et al. 2011); however, there is a noticeable distinction in Y-chromosomal haplogroup composition between the Mande- and the Gur-speaking populations (fig. 3a). The Mande groups, and especially the “linguistic island” Bisa as well as the geographically distant Mandenka, are characterized by particularly high frequencies of haplogroups E-M2*(xE-M191, xE-U175), whereas the Marka and South Samo, who live in close proximity to each other, have high frequencies of E-M33 (cf. supplementary table 2, Supplementary Material online). The Gur groups, in contrast, exhibit high frequencies of haplogroup E-U175, with the highest value of nearly 67% found in the Mossi, and have lower frequencies of E-M2*. The Lyela, Nuna, and Kassena from the southwest, who speak closely related languages, exhibit high frequencies (30–41%) of haplogroup E-M191*, which is very rare (<1%) in other populations of sub-Saharan Africa (de Filippo et al. 2011). The Yoruba, on the other hand, who speak a language belonging to the Benue–Congo branch of Niger–Congo, are characterized by high frequencies of haplogroup E-U174, which is absent or present in only low frequencies in the Mande-speaking populations, and present in the Gur-speaking populations at low to average frequencies.

This differentiation in Y-chromosomal haplogroup composition between the Mande and the Gur populations contrasts strikingly with the lack of such a distinction in the

Table 1. Diversity Values (DIV) for Y-chromosome and mtDNA Data, with Relative Standard Deviation (SD) in Parentheses.

Population	Y-Chromosome Haplogroups				Y-STR Data		mtDNA Haplogroups		mtDNA Sequence Data			
	<i>n</i>	Haplogroup DIV (SD)	<i>n</i>	Haplotypes	Haplotype DIV (SD)	Average DIV Over Loci (SD)	<i>n</i>	Haplogroup DIV (SD)	<i>n</i>	Haplotypes	Haplotype DIV (SD)	Nucleotide DIV
South Samo	41	0.75 (0.04)	41	34	0.99 (0.01)	0.54 (0.29)	39	0.82 (0.05)	39	1.00 (0.006)	0.0018	0.00042
North Samo	38	0.78 (0.06)	34	32	1.00 (0.01)	0.54 (0.29)	35	0.75 (0.07)	32	1.00 (0.008)	0.0015	0.00036
Marka	33	0.74 (0.06)	33	31	0.99 (0.01)	0.57 (0.30)	28	0.89 (0.04)	27	1.00 (0.01)	0.0020	0.00049
Bisa	40	0.23 (0.09)	40	23	0.90 (0.04)	0.26 (0.16)	31	0.87 (0.04)	31	1.00 (0.008)	0.0021	0.00049
Mandenka	15	0.47 (0.15)	15	14	0.99 (0.03)	0.56 (0.31)	22	0.83 (0.05)	19	0.99 (0.017)	0.0018	0.00045
Samoya	21	0.86 (0.04)	15	14	0.99 (0.28)	0.60 (0.34)	16	0.86 (0.08)	15	0.99 (0.025)	0.0021	0.00052
Pana	24	0.79 (0.06)	23	18	0.96 (0.03)	0.45 (0.25)	18	0.82 (0.07)	18	1.00 (0.019)	0.0015	0.00038
Lyela	40	0.77 (0.03)	38	31	0.99 (0.01)	0.51 (0.27)	37	0.86 (0.03)	36	1.00 (0.007)	0.0021	0.00049
Nuna	29	0.77 (0.06)	29	24	0.9 (0.02)	0.53 (0.29)	26	0.73 (0.08)	26	1.00 (0.01)	0.0018	0.00043
Kassena	33	0.65 (0.05)	32	23	0.97 (0.02)	0.41 (0.23)	27	0.86 (0.04)	26	1.00 (0.01)	0.0018	0.00043
Mossi	36	0.54 (0.09)	36	29	0.98 (0.02)	0.39 (0.22)	34	0.78 (0.05)	33	1.00 (0.008)	0.0014	0.00033
Yoruba	12	0.56 (0.15)	12	12	1.00 (0.34)	0.48 (0.28)	22	0.78 (0.06)	22	1.00 (0.013)	0.0020	0.00049

NOTE.—*n* refers to the sample size considered in each analysis.

mtDNA (fig. 3b), with all groups having similarly high frequencies (average 36%) of the major haplogroup L2a (cf. supplementary table 2, Supplementary Material online); similarly, both Mande- and Gur-speaking groups as well as the Yoruba exhibit medium–high frequencies of haplogroups L1b, L2c, L3b, L3d, and L3e.

With respect to the genetic diversity values, it is notable that all groups from Burkina Faso have very high mitochondrial sequence diversity values of 1 or close to 1 (table 1). This contrasts with the Y-chromosomal diversity, where some groups stand out as having very low values: thus, the Bisa stand out as having the lowest Y-chromosomal haplogroup and haplotype diversities, as is expected given their haplogroup composition, with a single haplogroup, E-M2*, being present in extremely high frequency (~88%). Similarly, the Mossi exhibit a high frequency of a single haplogroup (E-U175) and have a correspondingly low Y-chromosomal haplogroup diversity (table 1). On the other hand, the Gur-speaking Samoya, who are settled in an area straddling the Mali-Burkina border, stand out amongst the populations analyzed here as having the highest Y-chromosomal haplogroup diversity in addition to the generally high mtDNA diversity, even though they are represented by the smallest sample size amongst the Burkina Faso populations. Their high Y-chromosomal diversity is reflected in their haplogroup composition, since they exhibit a number of haplogroups present at most in low frequencies in the other populations, such as A-M91, B-M150, B-M181, and R-M207 (fig. 3a; cf. supplementary table 2, Supplementary Material online).

Correspondence Analysis

A CA on the basis of haplogroup frequencies clearly depicts the differences between the Y-chromosomal and mtDNA data in separating populations speaking languages belonging to different language families (fig. 4a and b). In the plot based on Y-chromosomal haplogroup frequencies, the Mande-speaking and Gur-speaking populations as well as the Yoruba, who speak a Benue–Congo language, are

clearly distinct, with the exception of the Gur-speaking Samoya, who are an outlier (fig. 4a).

In contrast, the plot based on mtDNA haplogroup frequencies exhibits less structuring by linguistic affiliation. Populations speaking languages belonging to the different families cluster together, and the Mande-speaking Mandenka and Marka cluster more closely with the Gur-speaking Pana and Kassena, respectively, than with their linguistic relatives, the Mande-speaking populations (fig. 4b). An MDS plot based on pairwise Φ_{st} values between the populations underlines the complete lack of both linguistic and geographic structure in the mtDNA sequence data, with the exception of the separation of the North Samo as well as the geographically distant Mandenka from Senegal, whereas the corresponding MDS plot based on pairwise F_{st} values calculated from Y-chromosomal haplogroup frequencies shows the same language–family-based structure as the CA plot (supplementary fig. 1a and b, Supplementary Material online).

AMOVA Analyses

The differences between the Y chromosome and mtDNA with respect to the partitioning of genetic variation according to linguistic affiliation is strikingly evident in AMOVA analyses as well. As shown in table 2, the differences between the ten Burkina Faso populations included in our study account for close to 15% of the Y-chromosomal haplogroup variance. When the populations are grouped by linguistic affiliation, nearly 10% of the Y-chromosomal haplogroup variance is explained by differences between the linguistic groups, with approximately 9% of the variance explained by differences between the populations belonging to each of the language families (table 2). This structure, however, is not apparent in the Y-STR data, where the proportion of variance accounted for by linguistic affiliation is lower than the variance found in populations belonging to the same language family (approximately 7% vs. 13%). Interestingly, partitioning the groups into ‘northern’ and ‘southern’ populations leads to a contrasting view of the structure inherent in the Y-chromosomal haplogroup versus STR diversity: there is no difference between northern and southern

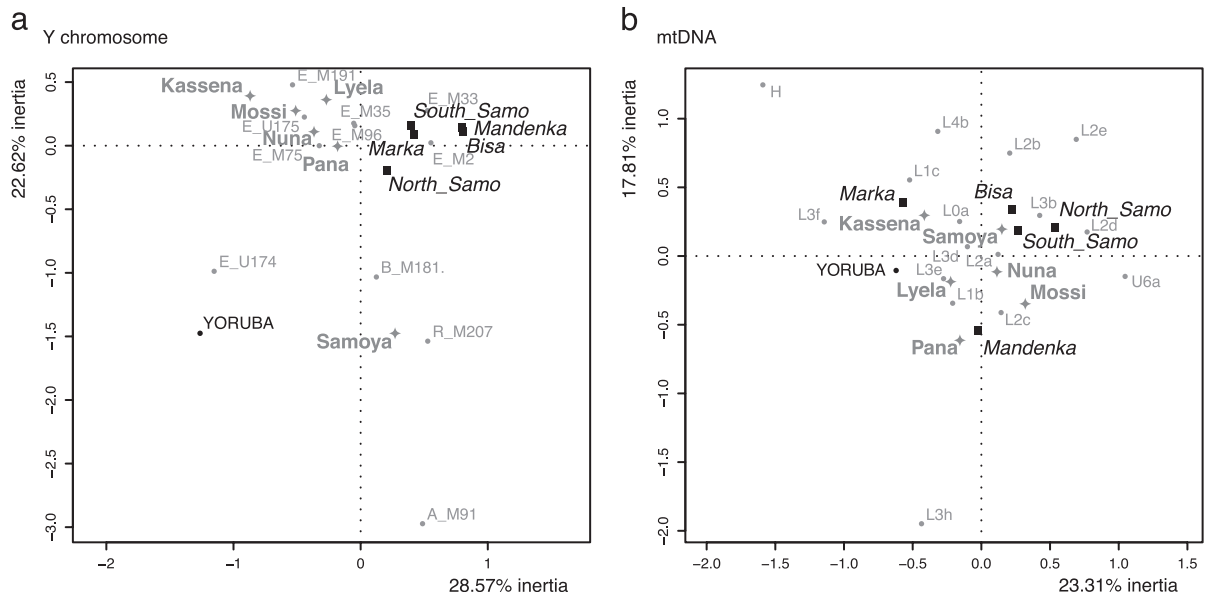


Fig. 4. CA plot based on haplogroup frequencies. Mande populations are indicated by squares and labeled in black italic font; Gur populations are indicated by diamonds and labeled in gray bold font. (a) Y chromosome. (b) mtDNA.

groups when basing the AMOVA on haplogroup variance, while some of the STR variance is accounted for by this geographical grouping (approximately 7% between-group variance vs. 13% within-group variance).

In stark contrast to the structure evident in the Y-chromosomal data, an AMOVA on the basis of the mtDNA (table 3) shows a complete lack of structure, with 100% of the variance being present within populations and no differences discernible between the individual populations. This holds for AMOVA analyses based both on the haplogroup composition and on Φ_{st} values between complete sequences; none of the variance is explained by either linguistic grouping or the division into “northern” and “southern” populations.

The same contrast between structure along linguistic affiliation reflected in the Y-chromosomal data and a complete lack of structure in the mtDNA is evident from AMOVA analyses including the Mandenka and Yoruba (cf. supplementary table 3, Supplementary Material online), with a slightly higher level of diversity in the mtDNA se-

quence data driven by the Mandenka, who are also separate in an MDS plot based on Φ_{st} (cf. supplementary fig. 1b, Supplementary Material online).

Correlation between Linguistic Affiliation and Genetic Relationships

As can be seen in figure 3, the Y-chromosomal haplogroup affiliation appears to fit the linguistic branching pattern within the Gur and the Mande family quite well, whereas the mtDNA haplogroup composition does not reflect this branching at all. For instance, the Lyela, Kassena, and Nuna, who are linguistically closely related, all have high frequencies of the otherwise rare haplogroup E-M191*. To better explore this possible correlation, a partial Mantel test was performed between linguistic distances and Y-chromosomal F_{st} and mtDNA Φ_{st} distances conditioned on geographic distance, with the Gur and Mande populations kept separate due to their high linguistic divergence. In contrast to the structuring by linguistic affiliation seen in the CA and AMOVA analyses, the partial Mantel test showed no significant correlations

Table 2. Analysis of Molecular Variance (AMOVA) for Y chromosome Data.

	n	Percentage of Variation		
		Among Groups	Among Pops/Within Groups	Within Pops
Haplogroup composition				
All 10 populations	1		14.59**	85.41**
Gur versus Mande	2	9.42**	8.76**	81.82**
North versus South	2	-1.81	15.69**	86.12**
STR data				
All 10 populations	1		17.57**	82.43**
32 Villages	1		20.84**	79.16**
Gur versus Mande	2	6.69	13.2**	80.11**
North versus South	2	7.09*	13.12**	79.79**

NOTE.—Asterisks indicate P values $<0.05^*$ and $<0.01^{**}$.

Table 3. Analysis of Molecular Variance (AMOVA) for mtDNA Data.

	n	Percentage of Variation		
		Among Groups	Among Pops/Within Groups	Within Pops
Haplogroup composition				
All 10 populations	1		-0.29	100.29**
Gur versus Mande	2	0.44	-0.53	100.09**
North versus South	2	-0.3	-0.12	100.42**
Sequence data				
All 10 populations	1		0.16	99.84
27 Villages	1		0.65	99.35
Gur versus Mande	2	0.12	0.009	99.871
North versus South	2	-0.27	0.3	99.97

NOTE.—Asterisks indicate P values $<0.05^*$ and $<0.01^{**}$.

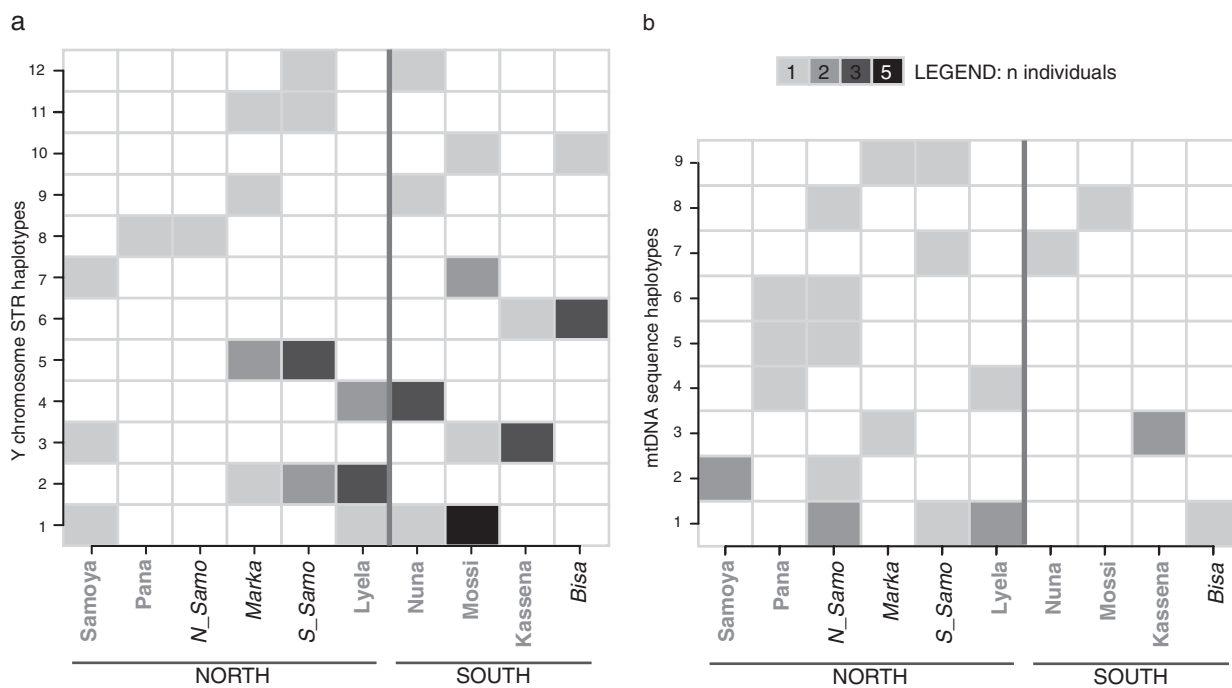


Fig. 5. Heat plot of haplotype sharing between populations in Burkina Faso. The populations are ordered in a roughly geographical order. Mande populations are labeled in black italic font; Gur populations are labeled in gray bold font. Only the haplotypes shared between at least two populations are displayed, and they are numbered and ordered on the vertical axis from the most frequent to the less frequent ones. (a) Y-STR haplotypes. (b) mtDNA sequence types.

between the Mande languages and their Y-chromosomal distances ($r = -0.558$, $P = 0.73$) and only a weak correlation between the Gur languages and their Y-chromosomal distances ($r = 0.556$, $P = 0.06$). As expected, the mtDNA distances do not correlate with the linguistic distances between the languages (Mande: mtDNA: $r = -0.07$, $P = 0.38$; Gur: $r = 0.21$, $P = 0.33$).

Correlation between Geographic and Genetic Distance and Haplotype Sharing

A Mantel test of the correlation between geographic and genetic distances demonstrates the impact of geography on Y-chromosomal variation in contrast to mtDNA variation, both at a fine-scaled level considering the distances between villages as well as at a broader scale considering distances between populations. At the village scale, the correlation between geographic distances and Φ_{st} distances based on mtDNA sequences is not significant ($r = -0.008$, $P = 0.52$), whereas it is significant for the comparison of R_{st} distances between Y-STR haplotypes and the geographic distances between villages ($r = 0.24$, $P = 0.003$). At the level of populations, the correlation is significant at both the level of Y-chromosomal haplogroups and Y-STR haplotypes ($r = 0.402$, $P = 0.019$ and $r = 0.345$, $P = 0.032$, respectively), but is again not significant for the mtDNA sequence data ($r = -0.0641$, $P = 0.60$).

An analysis of Y-STR haplotype sharing (fig. 5a) illustrates this effect of geographic proximity on Y-chromosomal variation, since in general, haplotypes are shared between geographically neighboring populations, such as the Marka

and South Samo, Lyela, and Nuna, or the Mossi and the Kassena and Bisa. A striking exception is represented by the Samoya, who are settled in the northernmost area of the sampled region and who share haplotypes not with their immediate neighbors, but with the more southerly groups Lyela, Nuna, Mossi, and Kassena. In the mtDNA, on the other hand, while sharing is more pronounced between neighboring populations, this involves only the northern populations, who share with each other and with southern populations; there is no haplotype sharing between southern populations only (fig. 5b).

Bayesian Skyline Plots

Using coding region sequences, we constructed BSPs to infer prehistoric demographic events that may have had an impact on present-day genetic variability in Burkina Faso. The analysis for all sequences shows a stable effective population size that rises steeply at about 9–14 kya, with a median of approximately 13 kya (fig. 6). This clear signal of a strong population expansion is consistently present in BSPs generated over all five runs performed with the relaxed clock model (data not shown).

Discussion

The Linguistic Context of Genetic Structure

As becomes evident from the various analyses described above, the populations of West Africa included in this study have undergone strikingly different historical trajectories in their maternal and paternal lineages. While in the maternal line there is absolutely no discernible structure

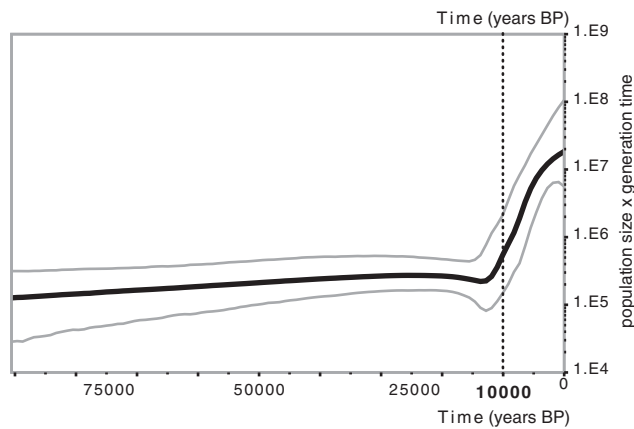


Fig. 6. Bayesian Skyline Plot of mtDNA coding regions for all individuals from Burkina Faso belonging to African haplogroup L, performed with a relaxed clock model.

amongst the populations, irrespective of which language family they are affiliated with and where they are settled, in the paternal line there is a clear correlation between the Y-chromosomal haplogroup composition and affiliation with one of the three major language families Mande, Gur, and Benue–Congo (cf. tables 2 and 3, fig. 4). This indicates a strong impact of patrilocality and exogamy on the genetic variability in this area, with women marrying into foreign ethnolinguistic communities irrespective of the geographic distance from their family homes and their linguistic affiliation, while men remain in their native village amongst members of their speech community. Currently, all the populations included in our sample are indeed strictly patrilocal, thus fitting this explanation, and the deep structure discernible in the Y chromosome provides an indication that this pattern of residence has been stable in this area over millennia. However, as shown by the AMOVA results based on Y-STR haplotype variation (tables 2 and 3) as well as by the haplotype sharing analysis and the correlation between geographic and genetic distance, this ancient structure along linguistic lines evident in the Y-chromosomal haplogroup composition has been overlaid by relatively more recent male migration between neighboring communities.

Interestingly, although all the populations included in this study form a homogenous group with respect to mtDNA sequence variation (cf. table 3), there are only nine haplotypes shared between populations (cf. fig. 5b), indicating a surprising lack of recent gene flow between the populations. While known relatives up to the grandparental generation as well as individuals of mixed ethnic affiliation were excluded by the sampling strategy, this finding is nevertheless unexpected. It indicates that the female migrations that led to the homogenization of the mtDNA gene pool are an old and stable feature of this area, where interethnic marriages were a means to strengthen political village alliances. Thus, alliances in defense against slave raids were sustained by the circulation of marriageable women across ethnolinguistic boundaries (Hubbell 2001). It is arguably this maternal gene flow that may have mediated

the current multilingualism and convergence of languages irrespective of their affiliation (Beyer and Schreiber, forthcoming). Interestingly, the analysis of haplotype sharing (fig. 5b) provides evidence for closer maternal ties among the ethnolinguistic groups in the north than among those in the south, which would be expected to lead to closer linguistic ties and concomitantly a higher degree of convergence among the languages spoken in the north; this last point requires further linguistic investigation.

While the Y-chromosomal haplogroup composition of the populations is structured at the level of the language families included here (Mande vs. Gur vs. Benue–Congo), the linguistic distinction between West Mande (the Marka and the Mandenka) and East Mande (the North and South Samo and the Bisa) is not reflected in the Y-chromosomal haplogroup variation (fig. 3). In the CA plot (fig. 4a) the Marka cluster closely with their geographic neighbors, while the Mandenka cluster very closely with the southernmost Mande-speaking group, the Bisa. This lack of congruence with linguistic subgrouping is confirmed by the partial Mantel test, which does not show any correlation between the linguistic and Y-chromosomal distances for the Mande family, and an AMOVA analysis using only the Mande-speaking populations, where grouping the populations by linguistic subgroup does not account for any variance in the data (cf. supplementary table 3, Supplementary Material online). Language shift from speakers of Gur languages to Marka may explain the divergence between the two West Mande populations; in this context, it is interesting to note that the Marka have relatively high frequencies of Y-chromosomal haplogroup E-M191*, which is found at a high frequency in their neighbors. The possibility of language shift is strengthened by the fact that historically the territory of Pana speakers (whose language belongs to the Gur family) extended much further south than it does today, whereas it is nowadays occupied by Marka and/or Samo people (cf. Beyer 2001). While close contact between Marka and Pana is attested by the large amount of borrowed Marka vocabulary in Pana (Beyer 2006), detailed linguistic analyses are required to assess the impact of possible language shift from Pana to Marka.

In contrast to the lack of substructure amongst the Mande-speaking groups, the presence in high frequencies (30–40%) of the rare Y-chromosomal haplogroup E-M191* (found at a frequency of 1.5% in Zambian East Bantu, but absent in other sub-Saharan African populations; de Filippo et al. 2011) in the three southwestern Gur-speaking populations Lyela, Nuna, and Kassena indicates a possible shared paternal history of these populations within the Gur-speaking groups. Even though the sharing of a paraphyletic haplogroup such as E-M191* is of course not proof of shared ancestry, it is noteworthy that these groups are united by sharing the M191 mutation in high frequency without the additional derived mutation (U-174) which is generally found on the background of M191 in the populations of Burkina Faso and large parts of sub-Saharan Africa (de Filippo et al. 2011). Further analysis of Y chromosomes belonging to the E-M191* haplogroup

are needed, however, to conclusively resolve the question of possible shared ancestry of these populations. Although this possible shared history does not emerge in the CA plots, it is in good accordance with the homogeneity of these groups with respect to cultural and linguistic features (Manessy 1979; fig. 3), and provides evidence for possible genetic parallels to the cultural and linguistic ties.

Nevertheless, while there is a clear correlation of genetic variation with linguistic affiliation for the Y-chromosomal data, there is also a distinct signal of geographic structure overlying the linguistic structure (cf. fig. 5a). This geographic structure is confirmed by a Mantel test of the correlation between geographic and genetic distances, which is not significant for the mtDNA sequences ($r = -0.008$, $P = 0.52$ for the comparison between villages and $r = -0.064$, $P = 0.60$ for the comparison between populations), but is significant for the comparison of Rst distances between Y-STR haplotypes and geographic distances between villages ($r = 0.24$, $P = 0.003$) as well as for the comparison between geographic distances between populations and genetic distances based on Y-chromosomal haplogroups and STR haplotypes ($r = 0.402$, $P = 0.019$, and $r = 0.345$, $P = 0.032$, respectively). This is indicative of more recent paternal gene flow taking place between neighboring populations. This relatively more recent paternal gene flow is also reflected by the difference in the AMOVA analysis based on the division between northern and southern groups: although this geographic split is not reflected at all in the Y-chromosomal haplogroup variation, it does emerge in the Y-STR variation (tables 2 and 3).

Geographic and Genetic Outliers

The Gur-speaking Pana settled in the north of the study area and the Mande-speaking Bisa settled to the south of the area are separated from their linguistic relatives by neighbors speaking unrelated languages. This separation might be expected to lead to a greater influx of individuals from geographically close but linguistically foreign communities. However, as can be seen in the CA and MDS plots based on Y-chromosomal data (cf. fig. 4a and supplementary fig. 1a, Supplementary Material online), both the Bisa and the Pana group with their linguistic neighbors, the Mande- and Gur-speaking populations, respectively, whereas in the mtDNA analyses, where the lack of linguistic structure is striking, both populations are located at a distance from their geographic neighbors (cf. fig. 4b and supplementary fig. 1b, Supplementary Material online). The Pana are quite isolated, sharing only one Y-STR and two mtDNA haplotypes with their closest geographic neighbors, the North Samo, indicating some gene flow between them, as well as one mtDNA haplotype shared with the Lyela. The Bisa share only one mtDNA haplotype with the more northerly populations North and South Samo as well as Lyela, but three Y-STR haplotypes with their geographic neighbors: two with the Kassena and one with the Mossi. The very low Y-chromosomal haplogroup and haplotype diversity in the Bisa (cf. table 1), where 14 out of 40 men share the same Y-STR haplotype, is indicative of a strong bottleneck or founder event in the history of this

group. Because the mtDNA diversity in the Bisa is as high as that in the other populations (table 1), the signal of this bottleneck or founder event appears to have been erased in the maternal line through intermarriage with neighboring groups—for which we, however, do not find any evidence in the genetic data. This fact might be due to the intermarriage having taken place with populations to the south who are not included here; interactions with more southerly populations from Ghana indeed emerge from dialectal data for Bisa (Keuthmann et al. 2004). On the other hand, the low Y-chromosomal diversity coupled with high mtDNA diversity appears indicative of a high degree of polygyny in the Bisa, similar to what was found for the highlands of West New Guinea (Kayser et al. 2003); however, there is no evidence for this in the historical or ethnographic data. On the contrary, the Bisa preferred sororal polygyny, where one man married several sisters (Pégard 1965), a type of polygyny that should not lead to an increase of mtDNA diversity. The much lower Y-chromosomal than mtDNA diversity might therefore, perhaps, be explained by the strict endogamy of the Bisa, where men were only allowed to settle if they belonged to allied Bisa clans, coupled with long-term military conflicts with the Mossi that might have diminished the paternal gene pool (Pégard 1965).

The Mossi, too, stand out amongst the populations included in this study by their low Y-chromosomal haplogroup diversity but relatively high mtDNA diversity. It is known from historical data (Fage 1964; Izard 1984) that the founders of the Mossi kingdoms immigrated into central Burkina Faso in the 15th century CE, where they established a succession of empires with military dominance of the surrounding populations from the 16th century. The genetic data indicate that this process was accomplished by an actual demic diffusion, with a later population expansion leading to an increase in Y-STR haplotype diversity, and incorporation of women from resident populations increasing mtDNA diversity.

The Gur-speaking Samoya are another group that stand out in our investigation. Linguistically included in the Northern Gurunsi subgroup, they retain connections with groups in Mali, like the Dogon (Beyer forthcoming; Tiendrebéogo 1983). Genetically, although they are indistinguishable from the other populations with respect to mtDNA (fig. 4b, supplementary fig. 1b, Supplementary Material online), they are a clear outlier in the Y-chromosomal analyses (cf. fig. 4a). First, they are the only group that shares Y-STR haplotypes not with their geographic neighbors but with populations located at a distance, such as the Lyela, Nuna, Kassena, and Mossi (fig. 5a). These haplotypes might have been spread via the Mossi, where they are found in higher frequency, and who were the socially dominant ethnolinguistic group in the area for several centuries. Second, they harbor several otherwise rare haplogroups with concomitant high Y-chromosomal haplogroup diversity. This appears to be indicative of gene flow in the paternal line, presumably with populations residing outside of the area of our investigation. Such evidence for paternal gene

flow in the Samoya is somewhat surprising for a patrilineal community. An alternative explanation might be that populations of different origin previously residing in the area were incorporated through language and cultural shift to an ultimately Samoya ethnolinguistic identity. Because the people of West Africa appear to be quite homogenous with respect to their mtDNA composition, as seen in [figure 4b](#) and [table 3](#), such a language shift would only be detectable in the Y chromosomes.

Impact of Climate Change on West African Demography

Approximately 10–12 kya, a major climatic change took place in West Africa, with a change from the hyperarid conditions during the Last Glacial Maximum toward more humid conditions, which led to a northward shift of both the equatorial belt as well as the Saharan desert, allowing for a major spread of the savannah and Sahel regions (McIntosh and McIntosh 1988). This made more resources available and thus opened up new territories for the expansion of humans: this putative demographic phase has been associated with the first spread of the Niger–Congo phylum (Dimmendaal 2008). Interestingly, the mtDNA sequence data variability for the Burkina Faso populations, as analyzed with Bayesian Skyline Plots, shows a demographic scenario in line with this interpretation: a clear sign of a very strong demographic expansion starting around 9–14 kya and continuing until the present. This high effective population size of the maternal gene pool of Burkina Faso is also evident in the high levels of sequence diversity of the groups included here ([table 1](#)). While the major climate change at the beginning of the Holocene arguably resulted in the strong demographic expansion detectable in the Bayesian Skyline Plots, later periods of climate deterioration with increased desertification, such as those that occurred 8–7 kya or 3 kya (McIntosh and McIntosh 1988), did not result in detectable reductions of population size. This lack of impact of later periods of climate change on the effective population size might be accounted for by migrations to avoid inhospitable regions as well as cultural innovations allowing adaptations to the changing climate. Thus, the domestication of cattle and ovicaprids as well as the domestication of plants, which would have enabled further population expansions, fall within such periods of climate deterioration and have been interpreted as an adaptive response to these difficult conditions (McIntosh and McIntosh 1988; Brooks 1989).

In summary, our investigation of both Y-chromosomal and mtDNA variation in populations from a fairly small region of Burkina Faso reveals a stark contrast between the maternal and paternal prehistory of these groups. Although the populations are entirely homogenous in their mtDNAs, the Y-chromosomal data show a striking correlation with linguistic affiliation. That this pattern, which is probably the result of the social practice of exogamic patrilocality, is not restricted to central Burkina Faso is demonstrated by the inclusion of the Mandenka from Senegal and the Yoruba from Nigeria. However, to obtain a more complete picture

of the population dynamics of Western Africa, the inclusion of groups from other areas would be necessary. Given the apparent ties of some of the populations studied here with populations to the north and south, including samples from Mali and Ghana would help elucidate the history of the area, and allow a better understanding of the past demographic dynamics in spite of the lack of historical documentation and fragmentation of the archaeological record. In addition, analysis of genome-wide polymorphisms might reveal more fine-grained structure within the individual populations studied here.

Supplementary Material

Supplementary text, tables 1–3, and figure 1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This study focuses on the prehistory of populations as reflected by mtDNA and Y-chromosomal variation, which can illuminate only certain aspects of overall population history. It does not aim at ascribing ethnicity to individual groups, nor does it intend to evaluate the self-identification of such groups. We sincerely thank all the sample donors for their contribution to our investigation, Djiri Lawono, Tjama Dieudonné, Seremé Yako, Kondo Umar Seremé, Seremé Oumar, Seremé Lanko, Pepin Ilboudo, Antoine Tière, Abdoul Karim Bandaogo, Anselme Ky, Andrea Reikat, Dr. Vincent Sedogo, and Leon Jander for generous support and assistance in the field, Cesare de Filippo for useful R scripts and suggestions, Martin Kircher for help with processing raw sequencing data, and all the members of the research group on Human Population History at the Max Planck Institute for Evolutionary Anthropology for helpful discussion. This study was supported by the Max Planck Society; C.B. was supported by a grant from the Deutsche Forschungsgemeinschaft (to B. P.); and the linguistic research was supported by funds from the Deutsche Forschungsgemeinschaft (to H.S. and K.B.).

References

- Alimen H. 1987. Evolution du climat et des civilisations depuis 40 000 ans du nord au sud du Sahara Occidental (premières conceptions confrontées aux données récentes). *Bulletin de l'Association française pour l'étude du quaternaire* 24(4):215–227.
- Atkinson QD, Gray RD, Drummond AJ. 2008. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol Biol Evol*. 25(2):468.
- Beyer K. 2001. Das Pana, eine vom Aussterben bedrohte Sprache. In: Egbert H, Jöckel A, Kühme W, Vierke U., editors. *Afrika im Wandel. Beiträge zur Abschlußkonferenz des Bayreuther Graduiertenkolleg "Interkulturelle Beziehungen in Afrika 1990-1999."* Berlin (Germany): Verlag für Wissenschaft und Forschung. p. 25–36.
- Beyer K. 2006. Das Pana im Netzwerk arealer Beziehungen: das Lexikon. In: Ibrizimow D, Winkelmann K, editors. *Zwischen Bantu und BurkinaFestschrift für Gudrun Mieke zum 65. Geburtstag.* Köln (Germany): Rüdiger Köppe Verlag. p. 9–22.

- Beyer K, Schreiber H. Forthcoming. Intermingling speech groups: morpho-syntactic outcomes of language contact in a linguistic area in Burkina Faso (West Africa). In: Lèglise I, Chamoreau C, editors. *The interplay of variation and change in contact settings—morphosyntactic studies*. London: Benjamins.
- Blench R. 2006. Archaeology, language, and the African past. Lanham (MD): AltaMira Pr.
- Brooks GE. 1989. Ecological perspectives on Mande population movements, commercial networks, and settlement patterns from the Atlantic wet phase (ca. 5500–2500 B.C.) to the present. *Hist Afr*. 16:23–40.
- Cann HM, de Toma C, Cazes L, et al. (41 co-authors). 2002. A human genome diversity cell line panel. *Science* 296(5566):261.
- Černý V, Hajek 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(1):9–27.
- Černý V, Salas A, Hajek 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(4):433–452.
- Černý V, Pereira L, Musilová E, et al. (11 co-authors). 2011. Genetic structure of pastoral and farmer populations in the African Sahel. *Mol Biol Evol*. 28:2491–2500.
- Creissels D. 2011. Lexique mandinka-français. Revised version [cited 2011 Jul 12]. Available from: http://www.deniscreissels.fr/public/Creissels-lexique_mandinka_2011.pdf.
- Diallo M. 1988. *Éléments de systématique et de dialectologie du marka-kan (Burkina-Faso)* [dissertation]. [Grenoble (France)]: Université Stendhal (Grenoble 3).
- Dimmendaal GJ. 2008. Language ecology and linguistic diversity on the African continent. *Lang Linguist Compass* 2(5):840–858.
- Fage JD. 1964. Reflections on the early history of the Mossi-Dagomba group of states. In: Vansina J, editor. *The historian in tropical Africa*. London: Oxford. p. 177–189.
- de Filippo C, Barbieri C, Whitten M, et al. (13 co-authors). 2011. Y-chromosomal variation in Sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol Biol Evol*. 28(3):1255.
- Hijmans RJ, Williams E, Vennes C. 2010. The “geosphere” package (Version 1.2–4) [cited 2011 July 12]. Available from: <http://cran.r-project.org/web/packages/geosphere/index.html>.
- Hubbell A. 2001. A view of the slave trade from the margin: Souroudougou in the late nineteenth-century slave trade of the Niger bend. *J Afr Hist*. 42(1):25–47.
- Izard M. 1984. Les peuples et les royaumes de la boucle du Niger et du bassin des Volta du Xlle au XVle siècle. In: Niane DT, editor. *Histoire générale de l’Afrique*, t.IV. Paris: Éditions Unesco, p. 237–264.
- Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet*. 72(2):281–302.
- Keuthmann K, Schreiber H, Vossen R. 2004. Die westafrikanische Savanne—eine Zeitreise durch 20 000 Jahre: 2.4. weil nicht sein kann, was nicht sein darf. Zum Problem von Sprachinseln in der Westafrikanischen Savanne am Beispiel des Bisa in Burkina Faso”. In: Albert K-D, Lohr D, Neumann K, editors. *Mensch und Natur in Westafrika. Ergebnisse aus dem Sonderforschungsberreich 268. Kulturentwicklung und Sprachgeschichte im Naturraum Westafrikanische Savanne*. Weinheim (Germany): Wileyvch. p. 9139–9167.
- Lewis MP, Summer Institute of Linguistics. 2009. *Ethnologue: Languages of the world*. 16th ed. Dallas (TX): SIL International.
- Manessy G. 1979. Contribution à la classification généalogique des langues voltaïques. Vol. 37. Paris: Peeters Pub & Booksellers
- Maricic T, Whitten M, Pääbo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11):e14004.
- McIntosh SK, McIntosh RJ. 1988. From stone to metal: new perspectives on the later prehistory of West Africa. *J World Prehist*. 2:89–133.
- Mukarovsky HG. 1966. Zur Stellung der Mandesprachen. *Anthropos* 61:679–688.
- Muzzolini A. 1993. The emergence of a food-producing economy in the Sahara. In: Shaw T, Sinclair P, Andah B, Okpoko A, editors. *The archaeology of Africa: food, metals and towns*. London: Routledge. p. 227–239.
- Nenadic O, Greenacre M. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Softw*. 20:1–13.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, O’Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2011. *vegan: community ecology package*. R package version 1.17-8 [cited 2011 Mar 15]. Available from: <http://CRAN.R-project.org/package=vegan>.
- Ouadba JM. 1997. *Development of national monograph on the biological diversity of Burkina Faso: data gathering, ecological considerations*. Ouagadougou: Minist. Envir. et de l’Eau.
- Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Paradis E. 2010. pegas: an R package for population genetics with an integrated—modular approach. *Bioinformatics* 26:419–420.
- Pégard O. 1965. Structure et relations sociales en pays Bisa (Haute-Volta). *Cahiers d’Études Africaines* 18:161–247.
- Pereira L, Černý V, Cerezo M, Silva NM, Hájek M, Vašíková A, Kujanová M, Brdička R, Salas A. 2010. Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. *Eur J Human Genet*. 18(8):915–923.
- Rosenberg NA. 2006. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet*. 70(6):841–847.
- Scozzari R, Cruciani F, Malaspina P, et al. (21 co-authors). 1997. Differential structuring of human populations for homologous X and Y microsatellite loci. *Am J Hum Genet*. 61(3):719–733.
- Scozzari R, Cruciani F, Santolamazza P, et al. (17 co-authors). 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet*. 65(3):829–846.
- Schreiber H. 2008. *Eine historische Phonologie der Niger-Volta-Sprachen*. Köln (Germany): Rüdiger Köppe.
- Tiendrebéogo G. 1983. *Langues et groupes ethniques de Haute Volta*. Abidjan (Côte d’Ivoire): Institut de Linguistique Appliqué, Université d’Abidjan/Nationale de Côte d’Ivoire.
- Venables WN, Ripley BD. 2002. *Modern applied statistics with S*. New York: Springer. p. 495.
- Williamson K. 1989. Niger-Congo overview. In: Bendor-Samuel JT, Rhonda LH, editors. *The Niger–Congo languages—a classification and description of Africa’s largest language family*. Lanham (MD): University Press of America. p. 3–45.
- Williamson K, Blench R. 2000. Niger-Congo. In: Heine B, Nurse D, editors. *African languages: an introduction*. Cambridge (UK): Cambridge University Press. p. 11–42.