

Y-Chromosome Variation Among Sudanese: Restricted Gene Flow, Concordance With Language, Geography, and History

Hisham Y. Hassan,¹ Peter A. Underhill,² Luca L. Cavalli-Sforza,² and Muntaser E. Ibrahim^{1*}

¹*Institute of Endemic Diseases, University of Khartoum, Sudan*

²*Department of Genetics, School of Medicine, Stanford University, CA*

KEY WORDS Sudan; Nile Valley; Y-chromosome; haplogroups

ABSTRACT We study the major levels of Y-chromosome haplogroup variation in 15 Sudanese populations by typing major Y-haplogroups in 445 unrelated males representing the three linguistic families in Sudan. Our analysis shows Sudanese populations fall into haplogroups A, B, E, F, I, J, K, and R in frequencies of 16.9, 7.9, 34.4, 3.1, 1.3, 22.5, 0.9, and 13% respectively. Haplogroups A, B, and E occur mainly in Nilo-Saharan speaking groups including Nilotics, Fur, Borgu, and Masalit; whereas haplogroups F, I, J, K, and R are more frequent among Afro-Asiatic speaking groups including Arabs, Beja, Copts, and Hausa, and Niger-Congo speakers from the Fulani ethnic group. Mantel tests reveal a strong correlation between genetic and linguistic struc-

tures ($r = 0.31$, $P = 0.007$), and a similar correlation between genetic and geographic distances ($r = 0.29$, $P = 0.025$) that appears after removing nomadic pastoralists of no known geographic locality from the analysis. The bulk of genetic diversity appears to be a consequence of recent migrations and demographic events mainly from Asia and Europe, evident in a higher migration rate for speakers of Afro-Asiatic as compared with the Nilo-Saharan family of languages, and a generally higher effective population size for the former. The data provide insights not only into the history of the Nile Valley, but also in part to the history of Africa and the area of the Sahel. *Am J Phys Anthropol* 137:316–323, 2008. © 2008 Wiley-Liss, Inc.

The sequential accumulation of genetic diversity along lines of common descent makes Y-chromosome a much more powerful tool for phylogeographic analysis than autosomes (Su et al., 1999; Jobling and Tyler-Smith, 2000; Semino et al., 2000; Underhill et al., 2000, 2001; Capelli et al., 2001; Hammer et al., 2001; Karafet et al., 2001). Archeological and genetic evidence based on mitochondrial, Y-chromosome, and autosomal DNA markers suggests that modern humans originated in Africa around 100,000–200,000 years ago (Cann et al., 1987; Grun and Stringer, 1991; Underhill et al., 2000), and that the Nile Valley may have provided a conducive environment for the first permanent settlements in Africa about 18,000 years ago (Phillipson, 1993), followed by the first adoption of agriculture in Africa by ancient Egyptians and Nubians around 10,000 BC or possibly earlier (Shillington, 1995). Climatic fluctuations would have played an important role in establishing the substrate of human habitation in the region. Widespread aridity occurred during the Last Glacial Maximum about 18 ky BP in North Africa followed by a Holocene lacustral phase (Street and Grove, 1976). Most if not all countries in the Nile Basin, including what is known today as the Sudan, are likely scenes for such pivotal human evolutionary events.

The languages spoken today in the Sudan belong to the three major African linguistic families: Nilo-Saharan, Afro-Asiatic, and Niger-Congo (Greenberg, 1963), encompassing an excess of 100 languages in total. Significant ethnic and cultural diversity exist, rendering the study of contemporary genetic diversity of human populations an interesting and appealing endeavor.

Several questions pertaining to the pattern of succession of the different groups/cultures in early Sudan have

been raised with the intention of acquiring clues into the history of the Nile Valley, state formation, and main demographic and migration events. Here, we seek insights into such history and attempt to understand Sudanese populations' structure from the male side of lineages, through the study of Y-chromosome variation of individuals representing some of the major ethnic groups in the country. These include groups known to have had an established history in what is today the Sudan, like Nuba and Nilotics, as well as groups that are known to have migrated relatively recently to Sudan (e.g., Hausa, Copts, and Meseria).

METHODS

Genotyping of Y-chromosome biallelic markers

A total of 445 unrelated male subjects belonging to 15 Sudanese populations (geographic locations are shown in Fig. 1) were analyzed. Appropriate informed consent was obtained from all participants. Sample sizes and linguistic affiliation for each population are reported in Table 1. DNA samples were obtained from blood or buccal specimen using phosphate buffer saline. DNA extraction was

*Correspondence to: Muntaser E. Ibrahim, Institute of Endemic Diseases, University of Khartoum, Qassr Street, Khartoum, Sudan. E-mail: mibrahim@iend.org

Received 8 October 2007; accepted 19 March 2008

DOI 10.1002/ajpa.20876

Published online 10 July 2008 in Wiley InterScience (www.interscience.wiley.com).

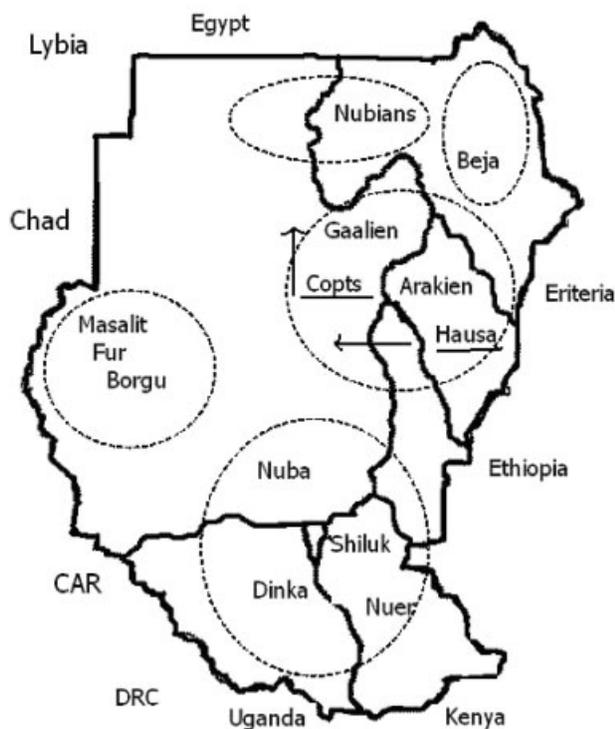


Fig. 1. Map of the Sudan. Showing the approximate locations of 15 populations typed for Y-chromosome markers in this study. The nomadic populations, Meseria, and Fulani were not shown in the map due to their wide distribution in west, east, and south-west of the country. Circles indicate the geographic regions. Arrows indicate directions of assigned location to population in Mantel Test, based on knowledge of history of migration within the past hundred years.

carried out according to Miller et al., (1988) with minor modifications. The biallelic variability at Y-chromosome-specific polymorphisms YAP, M2, M9, M11, M13, M23, M33, M40, M42, M51, M52, M60, M74, M78, M89, M170, M172, M173, M174, M175, M215, P25, and 12f2 (Y Chromosome Consortium [YCC], 2002) was used to generate male-specific haplotypes. To investigate the distribution of M78 binary subclades among the Sudanese population, all 114 Y-chromosomes carrying the M78 derived T allele were further genotyped for five binary markers-V12, V13, V22, V32, and V65-according to Cruciani et al. (2006, 2007).

Phylogenetic tree and principal component analysis

The Y-chromosome phylogenetic tree has been designed according to YCC nomenclature (YCC, 2002). Principal component analysis (PCA) was performed using PAST software (available online at <http://folk.uio.no/ohammer/past>). Published data of some African populations and Turks (Semino et al., 2002; Sanchez et al., 2005) were used alongside population data from this study to construct phylogenetic trees and PC plots.

Statistical analysis

Analysis of molecular variance (AMOVA) was performed to verify statistical differences between linguistic and geographic groups. Haplotype frequencies and mo-

lecular differences of Y-chromosome among haplogroups were taken into account. F_{ST} values were calculated based on the number of pairwise differences between Y-chromosome haplogroups. All calculations were performed using ARLIQUIN version 3.0 (Excoffier et al., 2005).

Mantel test

The correlation among genetic, linguistic, and geographic distances was assessed by the Mantel test (Mantel, 1967; Smouse et al., 1986), employing ARLEQUIN 3.0. This test is a statistical procedure that involves measuring the correlation coefficient between two matrices while holding a third one constant. Geographic distances between populations were calculated and converted to matrices using approximate latitude and longitude data. Fulani and Meseria were excluded from the analysis because they are widely spread nomadic populations. Linguistic distance matrices were calculated according to Excoffier et al. (1991) and Wood et al. (2005) after excluding the Nuba sample, because it encompasses different languages belonging to the Niger-Congo and Nilo-Saharan linguistic families.

F_{ST} and migration rate

Migration rate has been calculated according to the island model of migration and an $F_{ST} = 1/(1 + Nm)$ for haploid markers (Wright, 1951). Populations were categorized into four groups: Nilo-Saharan was divided into Nilotics and Non-Nilotics; Afro-Asiatic was divided into Arabs and Non-Arabs. The Niger-Congo samples were excluded from this analysis because of the small sample size.

RESULTS

Y-haplogroup diversity

Haplogroup frequencies in 15 Sudanese populations are given in Figure 2 following YCC nomenclature (2002). Haplogroups A-M13 and B-M60 are present at high frequencies in Nilo-Saharan groups except Nubians, with low frequencies in Afro-Asiatic groups although notable frequencies of B-M60 were found in Hausa (15.6%) and Copts (15.2%). Haplogroup E (four different haplotypes) accounts for the majority (34.4%) of the chromosome and is widespread in the Sudan. E-M78 represents 74.5% of haplogroup E, the highest frequencies observed in Masalit and Fur populations. E-M33 (5.2%) is largely confined to Fulani and Hausa, whereas E-M2 is restricted to Hausa. E-M215 was found to occur more in Nilo-Saharan rather than Afro-Asiatic speaking groups. In contrast, haplogroups F-M89, I-M170, J-12f2, and J-M172 were found to be more frequent in the Afro-Asiatic speaking groups. J-12f2 and J-M172 represents 94% and 6%, respectively, of haplogroup J with high frequencies among Nubians, Copts, and Arabs. Haplogroup K-M9 is restricted to Hausa and Gaalien with low frequencies and is absent in Nilo-Saharan and Niger-Congo. Haplogroup R-M173 appears to be the most frequent haplogroup in Fulani, and haplogroup R-P25 has the highest frequency in Hausa and Copts and is present at lower frequencies in north, east, and western Sudan. Haplogroups A-M51, A-M23, D-M174, H-M52, L-M11, O-M175, and P-M74 were completely absent from the populations analyzed.

TABLE 1. Sample size and linguistic affiliation for populations analyzed

| Ethnic groups | | | Linguistic affiliation ^a | | |
|---------------|----------|----------------------------|-------------------------------------|----------------------------|-------------------------------|
| Group | Subgroup | Socioeconomical activities | N | Family | Level |
| Nilotic | Dinka | Pastoralists | 26 | Nilo-Saharan | Eastern Sudanic |
| | Shilluk | Agri-pastoralists | 15 | Nilo-Saharan | Eastern Sudanic |
| | Nuer | Pastoralists | 12 | Nilo-Saharan | Eastern Sudanic |
| Nubians | – | Agriculturists | 39 | Nilo-Saharan | Eastern Sudanic |
| Borgu | – | Agriculturists | 26 | Nilo-Saharan | Maban |
| Masalit | – | Agriculturists | 32 | Nilo-Saharan | Maban |
| Fur | – | Agriculturists | 32 | Nilo-Saharan | Fur |
| Fulani | – | Nomadic Pastoralists | 26 | Niger-Congo | Atlantic |
| Hausa | – | Agriculturists | 32 | Afro-Asiatic | Chadic |
| Nuba | – | Agri-pastoralists | 28 | Nilo-Saharan + Niger-Congo | Eastern Sudanic + Kordofanian |
| Copts | – | Agriculturists | 33 | Afro-Asiatic | Ancient Egyptian |
| Arabs | Gaalien | Agriculturists | 50 | Afro-Asiatic | Semitic |
| | Meseria | Nomadic Pastoralists | 28 | Afro-Asiatic | Semitic |
| | Arakien | Agriculturists | 24 | Afro-Asiatic | Semitic |
| Beja | – | Pastoralists | 42 | Afro-Asiatic | Cushitic |
| Total | | | 445 | | |

^a According to www.ethnologue.com

E-M78 subclades

The distribution of E-M78 subclades among Sudanese is shown in Table 2. Only two chromosomes fell under the paragroup E-M78*. E-V65 and E-V13 were completely absent in the samples analyzed, whereas the other subclades were relatively common. E-V12* accounts for 19.3% and is widely distributed among Sudanese. E-V32 (51.8%) is by far the most common subclades among Sudanese. It has the highest frequency among populations of western Sudan and Beja. E-V22 accounts for 27.2% and its highest frequency appears to be among Fulani, but it is also common in Nilo-Saharan speaking groups.

AMOVA

AMOVA results are shown in Table 3. When populations were grouped according to linguistic affiliation, most of the genetic variance (75.3%) was found within populations; a value that is similar to that obtained (75.9%) when the populations were grouped according to the geographic distribution. Variance among populations within the linguistic groups was 14.6%, which is higher than the variance among the geographic groups (11.3%). A notable amount of genetic variance (10.1%) was found among linguistic groups, which is lower than variance among geographic groups (12.8%).

Genetic structure and measures of population size

Data from this study alongside available data of African populations and Turkish populations were analyzed and displayed graphically in a PC plot (see Fig. 3) that portrays broad genetic affinities reflected in two main clusters of genetically closely related populations. The first cluster groups Nilo-Saharan speaking groups from the Sudan together with Oromo from Ethiopia. The second cluster encompasses Afro-Asiatic speaking populations from the Sudan as well as Nubians and Amhara from Ethiopia, Fulani, and Turks. Senegalese fell relatively distant from both clusters in the plot.

Mantel testing showed strong correlations between genetic and linguistic distances ($r = 0.31$, $P = 0.007$), and a similar correlation between genetic and geographic distances ($r = 0.29$, $P = 0.025$).

Tests of neutrality and migration rate

Table 4 shows the values of $\theta\pi$, θ_s , and Tajima's D . All D statistics were slightly positive in most cases, whereas they were slightly negative in the case of Arakien, Gaalien, Hausa, and Nubians. All P values were not significant. Table 5 shows the F_{ST} estimates and migration rate for the two linguistic groups. For Nm, Arabs showed the highest migration rate ($Nm = 5.0$), whereas the lowest population size apart from Fulani (excluded from analysis) was found to be in Nilotics ($Nm = 2.3$).

DISCUSSION

The PCA plot based on F_{ST} values of Sudanese Turks and African populations defines two main genetic episodes that feature striking concordance with linguistic and geographic variations. One cluster relates to populations who speak languages of the Nilo-Saharan family, the predominant linguistic family in the Sudan across the millennia. This cluster is defined by the predominance of the ancestral haplogroups A-M13 and B-M60, as well as the common and most widely distributed haplogroup (E-M78). The second grouping encompasses populations who are essentially speakers of languages belonging to the Afro-Asiatic family, with the exception of Nubians. The placement of the Oromo, who speak a language of the Afro-Asiatic family, in the first cluster is probably because of their possession of high frequencies of A-M13. Both A-M13 and B-M60 are haplogroups that are deeply rooted within the human Y-chromosome tree, and they are known to be common among populations in eastern Africa (Underhill et al., 2000; Semino et al., 2002).

Haplogroup E-M78, however, is more widely distributed and is thought to have an origin in eastern African. More recently, this haplogroup has been carefully dissected and was found to depict several well-established

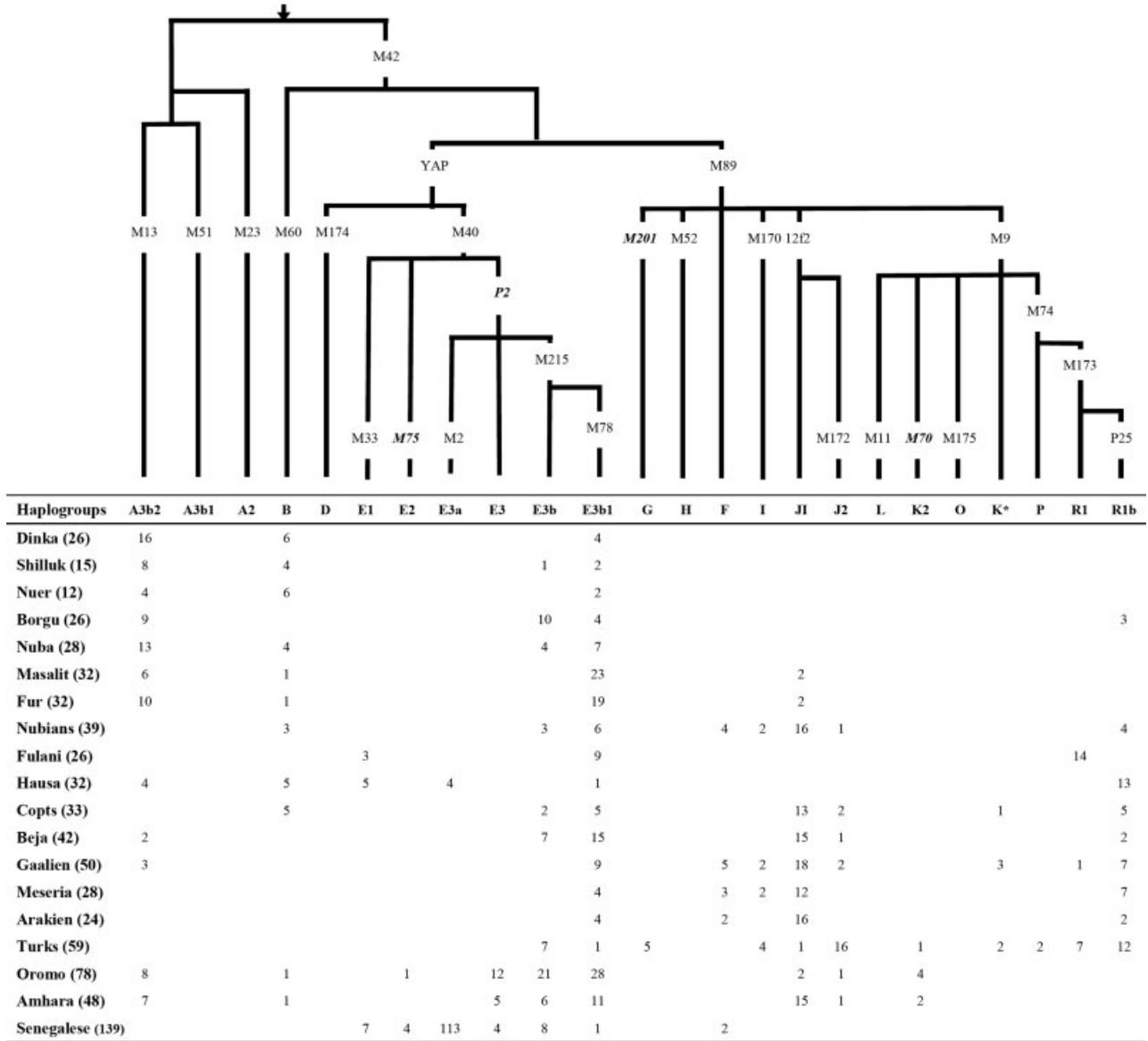


Fig. 2. Phylogenetic distribution of the Y-chromosome haplotypes and their frequencies in 15 Sudanese populations in the present study, compared with the frequencies in the Turks (Sanchez et al., 2005), two Ethiopian groups (Oromo and Amhara), and Senegalese (Semino et al., 2002). Numbering of mutations and haplogroups nomenclature are according to YCC (2002): those previously reported in the literature are shown in italics. The arrow shows the root of the tree.

subclades with defined geographical clustering (Cruciani et al., 2006, 2007). Although this haplogroup is common to most Sudanese populations, it has exceptionally high frequency among populations like those of western Sudan (particularly Darfur) and the Beja in eastern Sudan. The analysis of M78 subclades among Sudanese suggests that two subclades, E-V12 and E-V22, which are very common in northern African (Cruciani et al., 2007), might have been brought to Sudan from North Africa after the progressive desertification of the Sahara around 6,000–8,000 years ago. Sudden climate change might have forced several Neolithic cultures/people to shift northwards to the Mediterranean and southwards to the Sahel and Nile Valley (Dutour et al., 1988; Rando et al., 1998). E-V32 is the most frequent subclade among Sudanese. The Masalit possesses by far the highest frequency of the E-M78 and of the E-V32 haplogroup, sug-

gesting either a recent bottleneck in the population or a proximity to the origin of the haplogroup. Both E-V13, which is believed to originate in western Asia with its low frequency in North Africa, and E-V65 of North African origin (Cruciani et al., 2007), were not found among Sudanese.

Although the PC plot places the Beja and Amhara from Ethiopia in one sub-cluster based on shared frequencies of the haplogroup J1, the distribution of M78 subclades (Table 2) indicates that the Beja are perhaps related as well to the Oromo on the basis of the considerable frequencies of E-V32 among Oromo in comparison to Amhara (Cruciani et al., 2007). These findings affirm the historical contact between Ethiopia and eastern Sudan (Hassan, 1968, 1973; Passarino et al., 1998), and the fact that these populations speak languages of the Afro-Asiatic family tree reinforces the strong correlation

TABLE 2. Frequencies (N) of the Y-chromosome M78 subclades in 15 Sudanese populations

| Populations | M78 | M78* | V12* | V32 | V13 | V22 | V65 |
|-------------|-----|---------|-----------|-----------|-----|----------|-----|
| 1. Dinka | 4 | – | 1 | – | – | 3 | – |
| 2. Shilluk | 2 | – | – | – | – | 2 | – |
| 3. Nuer | 2 | – | 2 | – | – | – | – |
| 4. Borgu | 4 | – | – | 3 | – | 1 | – |
| 5. Nuba | 7 | 1 | 1 | 3 | – | 2 | – |
| 6. Masalit | 23 | 1 | – | 17 | – | 5 | – |
| 7. Fur | 19 | – | – | 13 | – | 6 | – |
| 8. Nubians | 6 | – | 5 | 1 | – | – | – |
| 9. Fulani | 9 | – | – | 1 | – | 8 | – |
| 10. Hausa | 1 | – | – | 1 | – | – | – |
| 11. Copts | 5 | – | 5 | – | – | – | – |
| 12. Beja | 15 | – | 2 | 13 | – | – | – |
| 13. Gaalien | 9 | – | 3 | 3 | – | 3 | – |
| 14. Meseria | 4 | – | 1 | 3 | – | – | – |
| 15. Arakien | 4 | – | 2 | 1 | – | 1 | – |
| Total N (%) | 114 | 2 (1.8) | 22 (19.3) | 59 (51.8) | – | 31(27.2) | – |

TABLE 3. Analysis of molecular variance (AMOVA)

| Groups | No. of groups | Within populations | | Among populations within groups | | Among groups | |
|-------------------|---------------|--------------------|----------|---------------------------------|----------|--------------|----------|
| | | Variance (%) | F_{ST} | Variance (%) | F_{SC} | Variance (%) | F_{CT} |
| Linguistic groups | 3 | 75.3 | 0.25 | 14.6 | 0.16 | 10.1 | 0.11 |
| Geographic groups | 5 | 75.9 | 0.24 | 11.3 | 0.13 | 12.8 | 0.13 |

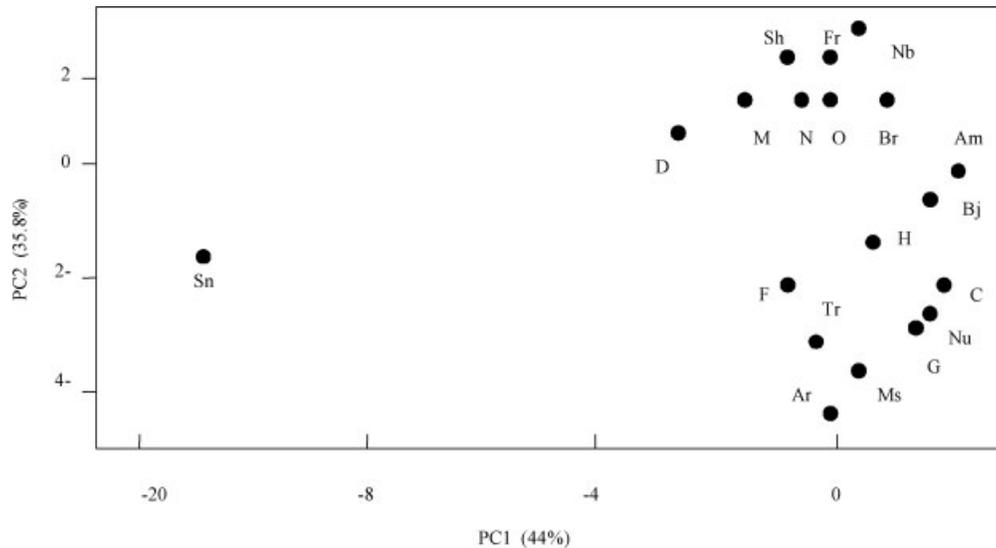


Fig. 3. PC plot of the Y-chromosome of Sudanese populations from the present study, compared with Turks (Sanchez et al., 2005), two Ethiopian groups (Oromo and Amhara), and Senegalese (Semino et al., 2002). Codes are as follows: **Am**, Amhara; **Ar**, Arakien; **Bj**, Beja; **Br**, Borgu; **C**, Copts; **D**, Dinka; **F**, Fulani; **Fr**, Fur; **G**, Gaalien; **H**, Hausa; **Mt**, Masalit; **Ms**, Meseria; **N**, Nuer; **Nb**, Nuba; **Nu**, Nubians; **Or**, Oromo; **Sh**, Shilluk; **Sn**, Senegalese; **Tr**, Turks.

between linguistic and genetic diversity (Cavalli-Sforza, 1997).

The Mantel test highlights the pronounced divide and correlation across linguistic and geographic lines in our data set. When populations are grouped according to linguistic affiliation, the proportion of among-groups variance ($F_{CT} = 0.11$) is similar to that when populations are grouped according to geographic location ($F_{CT} = 0.13$), which further indicates that Y-chromosome variation is significantly partitioned among both geographic and linguistic groups. The correlation with geography

became obvious when strictly nomadic pastoralists, who are widely spread in the country, were removed from the analysis. These include Meseria and Fulani, tribes that typically traversed the Sahel and who have managed to settle only recently in the Sudan.

It seems that gene flow is not only recent (Holocene onward) but also largely of focal nature. Most speakers of Nilo-Saharan languages, the major linguistic family spoken in the country, show very little evidence of gene flow and demonstrate low migration rate, with exception of the Nubians, who appear to have sustained consider-

TABLE 4. Measures of genetic diversity estimated from Y-chromosome data

| Linguistic groups | Populations | N | Hp | θ_K | θ_S (SD) | θ_π (SD) | D (P) |
|-------------------|-------------|----|----|------------|-----------------|-------------------|----------------|
| Afro-Asiatic | Arakien | 24 | 4 | 1.105 | 1.875 (0.901) | 1.812 (1.205) | -0.105 (>0.10) |
| | Beja | 42 | 6 | 1.679 | 2.324 (0.965) | 2.670 (1.612) | 0.440 (>0.10) |
| | Copts | 33 | 7 | 2.424 | 2.710 (1.124) | 2.955 (1.764) | 0.284 (>0.10) |
| | Gaalien | 50 | 9 | 2.941 | 2.679 (1.045) | 2.670 (1.605) | -0.010 (>0.10) |
| | Hausa | 32 | 6 | 0.976 | 1.986 (0.898) | 1.831 (1.202) | -0.235 (>0.10) |
| | Meseria | 28 | 5 | 1.502 | 2.056 (0.940) | 2.384 (1.488) | 0.492 (>0.10) |
| Nilo-Saharan (NS) | Borgu | 26 | 4 | 1.066 | 2.096 (0.965) | 2.858 (1.732) | 1.143 (>0.10) |
| | Dinka | 26 | 3 | 0.645 | 1.572 (0.785) | 2.265 (1.432) | 1.303 (>0.10) |
| | Fur | 32 | 4 | 0.976 | 1.986 (0.898) | 2.250 (1.414) | 0.397 (>0.10) |
| | Masalit | 32 | 4 | 1.905 | 2.731 (1.137) | 3.494 (2.034) | 0.886 (>0.10) |
| | Nuer | 12 | 3 | 0.934 | 1.987 (1.067) | 2.667 (1.719) | 1.305 (>0.10) |
| | Nubians | 39 | 8 | 2.763 | 2.838 (1.133) | 2.729 (1.644) | -0.119 (>0.10) |
| | Shilluk | 15 | 4 | 1.428 | 2.153 (1.085) | 2.629 (1.667) | 0.796 (>0.10) |
| | Fulani | 26 | 3 | 0.645 | 1.310 (0.693) | 2.234 (1.417) | 1.997 (>0.10) |
| NS+NC | Nuba | 28 | 4 | 1.032 | 1.799 (0.854) | 2.677 (1.636) | 1.467 (>0.10) |

Note: N = sample size; Hp = number of different haplogroups observed; SD = standard deviation; D = Tajima's D; P = P value for D.

TABLE 5. F_{ST} and Nm estimates in two linguistic groups in Sudanese populations

| Linguistic groups | Subgroups | F_{ST} | Nm | Tajima's D (P) |
|-------------------|-------------|----------|-----|----------------|
| Afro-Asiatic | Arabs | 0.166 | 5.0 | -0.498 (>0.1) |
| | Non-Arabs | 0.171 | 4.9 | -0.119 (>0.1) |
| Nilo-Saharan | Nilotic | 0.300 | 2.3 | 0.366 (>0.1) |
| | Non-Nilotic | 0.250 | 3.0 | 1.559 (>0.1) |

^a F_{ST} is a measure of interpopulation variability, whereas Nm is the effective number of migrants.

able gene flow from Asia and Europe together with the Beja. Both Beja and Nubians lie at entering ports of the Sudan; the Beja in the Red Sea area where past and recent settlements of both Turks and Arabs are evident, and Nubians occupy a strip along the Nile bordering south of Egypt, where successive waves of migration and conquest of the Sudan have passed over the millennia (MacMichael, 1967; Hassan, 1968). This is attested by the remarkable presence of the J-M172 chromosome known to be quite frequent in Turkey and the Levant, as well as other Eurasian haplogroups, including haplogroup J-12f2 (Al-Zahery, 2003; Giacomo et al., 2004) and the bondage and genetic continuum of the Nubians with their kin in southern Egypt is indicated by comparable frequencies of E-V12 the predominant M78 subclade among southern Egyptians (Cruciani et al., 2007).

The group that displayed the highest population size in fact was the Gaalien from central Sudan. This group occupies a trading crossroad that extends back to the ancient Kingdom of Meroe. The Gaalien exhibits a Y-profile that gives insight into past and recent migrations to the Sudan. Interestingly, they still maintain low frequencies of haplogroup A-M13 and E-M78, which suggests older rooting and relates them to other neighboring populations. Considerable frequencies of Eurasian haplogroups including J-12f2 are also present, consistent with a more recent Arabic oral tradition and descent.

Among other groups with a relatively large population size are the Hausa and Copts. Hausa display elevated frequencies of the haplogroup R-P25, which is considered as an evidence for back migration from Asia to Sub-Saharan Africa (Cruciani et al., 2002), although a recent study questions the reliability of this marker being used

in singularity (Adams et al., 2006). Other groups with varying frequencies of this haplogroup, like the Borgu and Nubians, appear to have acquired it from Afro-Asiatic speaking groups through gene flow. We have recently shown that this haplogroup is strongly associated with the sickle cell gene. Both markers might have co-introgressed during the past 300 years to eastern Sahel (Bereir et al., 2007).

The relatively high-effective population size of the Copts is unlikely to have been influenced by their recent history in the Sudan. The current communities are known to be largely the product of recent migrations from Egypt over the past two centuries. The Copt samples displayed a most interesting Y-profile, enough (as much as that of Gaalien in Sudan) to suggest that they actually represent a living record of the peopling of Egypt. The significant frequency of B-M60 in this group might be a relic of a history of colonization of southern Egypt probably by Nilotics in the early state formation, something that conforms both to recorded history and to Egyptian mythology.

The Fulani, who possess the lowest population size in this study, have an interesting genetic structure, effectively consisting of two haplogroups or founding lineages. One of the lineages is R-M173 (53.8%), and its sheer frequency suggests either a recent migration of this group to Africa and/or a restricted gene flow due to linguistic or cultural barriers. The high frequency of subclade E-V22, which is believed to be northeast African (Cruciani et al., 2007) and haplogroup R-M173, suggests an amalgamation of two populations/cultures that took place sometime in the past in eastern or central Africa. This is also evident from the frequency of the "T" allele of the lactase persistence gene that is uniquely present in considerable frequencies among the Fulani (Mulcare et al., 2004). Interestingly, Fulani language is classified in the Niger-Congo family of languages, which is more prevalent in West Africa and among Bantu speakers, yet their Y-chromosomes show very little evidence of West African genetic affiliation.

It seems, however, that the effective size of the pastoralists and nomadic pastoralists is generally much smaller than groups of sedentary agriculturalists life style. This is intriguing in the sense that one would expect nomadic tribes to be more able to admix, spread, and receive genes than their sedentary counterparts.

However, this data might point to the fact that population size in human history is largely affected by culture, including the formation of states, rather than the population mobility. Both Hausa and Copts descend from longstanding cultures of city states and empires that have historically expanded, drawing into their spin other groups and populations.

For all studied populations, P values of Tajima's D (Table 4) were not significant and, therefore, the hypothesis of expansion could be rejected at the $P = 0.05$ level. This may be due to the recent nature of the expansion, consistent with N_m that supports a larger effective size migration rate and expansion of Afro-Asiatic speaking groups who are known to have migrated recently to the country.

Accordingly, we suggest that regional variation in Y-chromosome sequences in Sudan is likely to have been shaped by human migrations, some of which occurred in the recent past. For example, the high effective population size of Afro-Asiatic males could be explained by a recent higher migration rate; diversity among populations of Afro-Asiatic versus Nilo-Saharan speaking groups and the low male migration rate in Nilo-Saharan may be due to the in situ evolution of Nilo-Saharan in East Africa.

CONCLUSION AND FUTURE WORK

The strong concordances between the language and genetics ($P = 0.007$), and geography and genetics ($P = 0.025$), suggest that language, geography, and cultural traits may have played a significant role in the genetic structure of Sudanese populations, in a country with diverse linguistic and cultural traits. Although most of the Y-chromosome markers genotyped define the deep ancestry of the phylogeny, our interpretations of recent causative events are plausible given the strong linguistic correlations and the concordance with history. Such perspectives, however, should be tested by employing more recently derived markers within the major haplogroups to further explore some of the pertinent issues addressed in this manuscript.

ACKNOWLEDGMENTS

We express our gratitude to all donors for providing DNA samples and to the people who contributed to the sample collection for their helpful collaboration, which made this study possible. Our gratitude extends to Professor Himla Soodyall of WITS and Dr. Khider Abdelkarem for their comments and helpful discussion.

LITERATURE CITED

- Adams SM, King TE, Bosch E, Jobling MA. 2006. The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion. *Forensic Sci Int* 159:14–20.
- Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS. 2003. Y-chromosome and mtDNA polymorphisms in Iraq, a cross road of the early human dispersal and of post-Neolithic migrations. *Mol Phyl Evol* 28:458–472.
- 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.
- Cann RL, Stonking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.
- Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill PA, Pascali VL, Ko TM, Goldstein DB. 2001. A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet* 68:432–443.
- Cavalli-Sforza LL. 1997. Genes, peoples, and languages. *Proc Natl Acad Sci USA* 94:7719–7724.
- Cruciani F, Fratta RL, Torroni A, Underhill PA, Scozzari R. 2006. Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. *Hum Mutat* 27:831–832.
- Cruciani F, Fratta RL, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon JM, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradsnik B, Assum G, Brdicka R, Kozlov A, Efremov GD, Coppa A, Novelletto A, Scozzari R. 2007. Tracing Past Human Male Movements in Northern/Eastern Africa and Western Eurasia: New Clues from Y-chromosomal Haplogroups E-M78 and J-M12. *Mol Biol Evol* 24:1300–1311.
- Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olekers A, Modiano D, Destro-Bisol G, Holmes S, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA. 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.
- Dutour O, Vernet R, Aumassip G. 1988. Le peuplement préhistorique du Sahara. In: Aumassip G, et al., editors. *Milieux, hommes et techniques du Sahara préU historique*. Paris: *Problèmes [mes actuels]*. p 39–52.
- Excoffier L, Harding RM, Sokal RR, Pellegrini B, Sanchez-Mazas A. 1991. Spatial differentiation of Rh and Gm haplotype frequencies in Sub-Saharan Africa and its relation to linguistic affinities. *Hum Biol* 63:276–307.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinformatics Online* 1:47–50.
- Giacomo F, Luca F, Popa LO, Akar N, Agnagou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrilu L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarino V, Mammi C, Michalodimitrakis EN, Paoli G, Pappa KI, Pedicini G, Terrenato L, Tofanelli A, Mapaspina P, Novelletto A. 2004. Y chromosome haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* 115:357–374.
- Greenberg JH. 1963. *The languages of Africa*. Bloomington: Indiana University Press.
- Grun R, Stringer CB. 1991. Electron spin resonance dating and the evolution of modern humans. *Archaeometry* 33:153–199.
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18:1189–1203.
- Hassan YF. 1968. *Sudan in Africa*. Khartoum: Khartoum University Press.
- Hassan YF. 1973. *The Arabs and the Sudan*. Khartoum: Khartoum University Press.
- Jobling MA, Tyler-Smith C. 2000. New uses for new haplotypes the human Y chromosome, disease and selection. *Trends Genet* 16:356–362.
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69:615–628.
- MacMichael HA. 1967. *A History of the Arabs in the Sudan*, Vol. 2. Cambridge: Cambridge University Press.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–222.
- Miller SA, Dykes DD, Plesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215.
- Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekgn A, Swallow DM, Bradman N, Thomas MG. 2004. The T allele of a single-nucleotide polymorphism 13.9 kb upstream

- of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am J Hum Genet* 74:1102–1110.
- Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS. 1998. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 62:420–434.
- Phillipson DW. 1993. *African archaeology*. Cambridge: Cambridge University Press.
- Rando JC, Pinto F, Gonz ales AM, Hern andez M, Larruga JM, Cabrera VM, Bandelt H-J. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* 62:531–550.
- Sanchez J, Hallenbergh C, B orsting C, Hernandez A, Morling N. 2005. High frequencies of Y chromosome lineages characterized by E3b1. DYS19-11, DYS392-12 in Somali males. *Eur J Hum Genet* 13:856–866.
- Semino O, Passarino G, Oefner P, Lin A, Arbuzova S, Beckman L, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti A, Cavalli-Sforza L, Underhill P. 2000. The genetic legacy of paleolithic homo sapiens in extant Europeans: A Y chromosome perspective. *Science* 290:1155–1159.
- Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA. 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 70:265–268.
- Shillington K. 1995. *History of Africa*. New York: St. Martin's Press. p 16–17.
- Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst Zool* 35:627–632.
- Street FA, Grove AT. 1976. Environmental and climatic implications of Late Quaternary lake-level fluctuations in Africa. *Nature* 261:385–390.
- Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L. 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet* 65:1718–1724.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazo'n Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43–62.
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim ME, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361.
- Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, Santachiara AS, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: Evidence for sex-biased demographic processes. *Eur J Hum Genet* 13:867–876.
- Wright S. 1951. The genetical structure of populations. *Ann Eugen* 15:323–354.
- Y Chromosome Consortium (YCC). 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–334.