ARTICLE

Y-chromosome E haplogroups: their distribution and implication to the origin of Afro-Asiatic languages and pastoralism

Eyoab I Gebremeskel[1,2] and Muntaser E Ibrahim*[,1]

Archeological and paleontological evidences point to East Africa as the likely area of early evolution of modern humans. Genetic studies also indicate that populations from the region often contain, but not exclusively, representatives of the more basal clades of mitochondrial and Y-chromosome phylogenies. Most Y-chromosome haplogroup diversity in Africa, however, is present within macrohaplogroup E that seem to have appeared 21 000–32 000 YBP somewhere between the Red Sea and Lake Chad. The combined analysis of 17 bi-allelic markers in 1214 Y chromosomes together with cultural background of 49 populations displayed in various metrics: network, multidimensional scaling, principal component analysis and neighbor-joining plots, indicate a major contribution of East African populations to the foundation of the macrohaplogroup, suggesting a diversification that predates the appearance of some cultural traits and the subsequent expansion that is more associated with the cultural and linguistic diversity witnessed today. The proto-Afro-Asiatic group carrying the E-P2 mutation may have appeared at this point in time and subsequently gave rise to the different major population groups including current speakers of the Afro-Asiatic languages and pastoralist populations.
European Journal of Human Genetics (2014) **22,** 1387–1392; doi:10.1038/ejhg.2014.41; published online 26 March 2014

INTRODUCTION

Polymorphisms like bi-allelic mutations associated with the male-specific Y (MSY) chromosome portion, are important tools that proved essential in addressing aspects of human ancestry, migration episodes[1] and testing coalescence processes.[2] Interestingly, some bi-allelic markers of the Y chromosome not only have geographically defined distributions, but are also associated with certain facets of human culture like languages[3,4] and practice of pastoralism[5] all of which contribute to the phenomenon of genetic drift, probably the most single key element in shaping population genetic structures.[6] Intuitively, the high correlation between geographical distribution of some of the major E haplogroups and distribution of Afro-Asiatic languages, exemplary of established correlation between languages and genes as proposed by Cavalli-Sforza[7,8] prompted us to revisit such correlation in a multidisciplinary platform better suited to unravel hitherto untold chapters of human history. No better venue to put such approach into practice than the area of the Sahel and East Africa. The Sahel, which extends from the Atlantic to the Red Sea coast of Sudan and Eritrea and the Ethiopian highlands including fringes of the Sahara, has witnessed human population demographic events that were pivotal in prehistoric and historic periods of human history. Early occupation by *Homo sapiens* of the Red Sea coast of Eritrea,[9–13] and evidences of traces of earlier urban settlements in much of Eritrea[14–16] are some of the archaeological and paleontological evidences that suggest a major contribution of this area to prehistory and migration including the exodus of anatomically modern humans to Eurasia. Furthermore, in addition to the area being strikingly rich in genetic and linguistic diversity, it is one of the few remaining enclaves of traditional pastoralism, a dying human culture.[17]

Although suggestions has been made that East Africa is the likely place of origin of Y-chromosome haplogroups including the major E haplogroups, yet key questions on human origin and dispersal remain not fully addressed. One query, however, is whether the major macrohaplogroup E present almost in all continents and with particularly high frequency in East and North Africa in plethora of ancestral lineages, because of gene flow or an original early event of *in situ* evolution. Although a lot has been done to refine the E macrohaplogroup tree, sampling representative populations, like Eritreans, may still shed light on new dimensions of the history of populations bearing these mutations. Despite a single attempt to study Eritrean populations from the diaspora,[18] no systematic analysis has been done so far to address the genetic diversity of extant Eritrean populations pertinent to questions like the origin of the Afro-Asiatic languages and pastoralism in light of the distribution of E macrohaplogroup as a case study.

MATERIALS AND METHODS

Y-chromosome genotyping of bi-allelic markers

A total of 1214 Y chromosomes, positive for E haplogroups, were considered in the analysis. Out of an original sample of Eritrean males screened, 39 Y chromosomes (49%) turned to be positive for E markers and were included in the analysis. The language affiliation and present or past history of the populations analyzed are given in Supplementary Table S1. The culture is taken within the context of the current linguistic affiliation and information of past and present subsistence practices. The history of pastoralism is not restricted to cattle as it has been shown that livestock may change according to environment as the case with Baggara Arabs who were originally camel herders turned to cattle. Appropriate informed consent was obtained from all participants. DNA samples were obtained from buccal specimen using phosphate-buffered saline

[1]Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan; [2]Department of Biology, Eritrea Institute of Technology, Mai-Nefhi, Eritrea
*Correspondence: Professor ME Ibrahim, Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, PO Box 102, Qasser Street, Khartoum 11111, Sudan. Tel: +249 912576418; Fax: +249 9183793263; E-mail: mibrahim@iend.org

and DNA extraction was carried out according to Miller et al,[19] with minor modifications. The bi-allelic variability at Y-chromosome-specific polymorphisms E-M107, E-M123, E-M148, E-M191/P86, E-M200, E-M281, E-M329, E-M33, E-M34, E-M54, E-M81, E-P72, E-U175, E-V19, E-V32, E-V6 (Y Chromosome Consortium (YCC), 2008),[20] and E-M78[21] were used to generate MSY chromosome haplotypes. Primers for genotyping were selected according to Karafet et al[20] and Underhill et al[22] and the references herewith. Most of the genotyping was done at BGI Laboratory (Hong Kong, China). Published data of African, Asian and European populations[22–34] were used alongside population data from this study for comparative analysis.

Y-chromosome haplogroup tree

The Y-chromosome haplogroup tree has been constructed manually following YCC 2008 nomenclature[20] with some modifications.[35] The tree (Supplementary Figure S1) contains the E haplogroups of Eritrean populations from this study and those reported in the literature.[22–34] Genotyping results for E-V13, E-V12, E-V22 and E-V32 reported for Eritrean samples and elsewhere[23,27] were retracted to E-M78 haplogroup level. All the analyses in this study were done at the same resolution using the following 17 bi-allelic markers: E-M96, E-M33, E-P2, E-M2, E-M58, E-M191, E-M154, E-M329, E-M215, E-M35, E-M78, E-M81, E-M123, E-M34, E-V6, E-V16/E-M281 and E-M75.

Phylogenetic analysis

Median-joining network (Figure 1) was constructed using Network 4.6.1.1. (http://www.fluxus-engineering.com) program.[36] Effective mutation rate of $3.8–4.4 \times 10^{-4}$ mutations/nucleotide/generation[37] was used in estimating the divergence time of the populations using Network program.

A neighbor-joining (NJ) tree (Figure 2) was constructed[38] and evolutionary analysis was conducted using MEGA5 (Tamura et al[39]).

Genetic structure and population differentiations

Multi-dimensional scaling (MDS) and principal component analysis (PCA) were performed by using PAST (paleontological statistics) algorithms version 2.11 software (available online at http://folk.uio.no/ohammer/past)[40] based on FST values generated from Arlequin ver3.11 program.[41] Analysis of molecular variance (AMOVA) was performed to verify statistical differences between linguistic and geographic groups. Haplotype frequencies and molecular differences of Y chromosome among haplogroups were taken into account. FST values were calculated based on the number of pairwise differences between Y-chromosome haplogroups. All calculations were performed using Arlequin version 3.11 (Excoffier et al[41]) using the 17 bi-allelic data listed above. The correlation among genetic, linguistic and geographic distances was assessed by the Mantel test using FST matrices resulted from Arlequin analysis.

RESULTS

Y-chromosome haplogroup diversity

Bi-allelic frequencies of E haplogroups for the populations involved in the analysis are given in Supplementary Figure S1 following the recent nomenclature.[20]
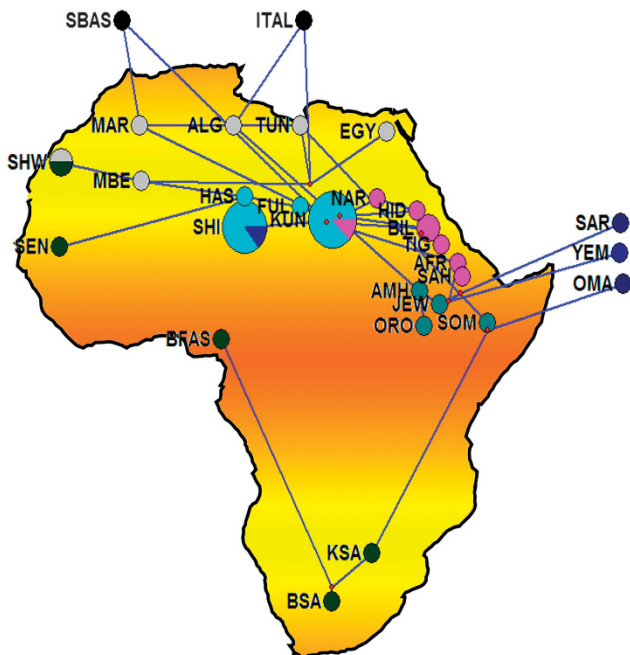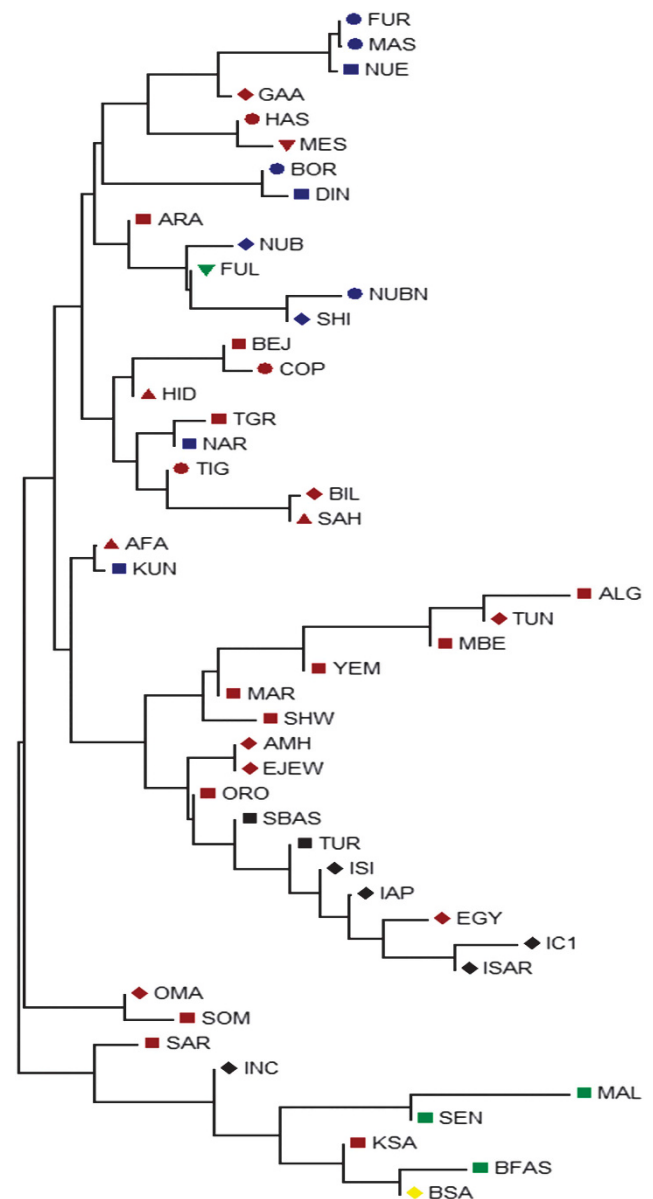

Figure 1 Median-joining (MJ) network. Network manipulated to fit the geography of the extant populations. MJ network was constructed using E haplogroup frequencies. Group represented by ITAL contains all the Italian samples pooled. Populations' descriptions are given in Supplementary Table S1.

Figure 2 NJ tree based on FST values generated from Arlequin 3.11. Population names are as given in Supplementary Table S1. Population life style: circle – agriculturalists; square – pastoralists; triangle – nomads; inverted triangle – nomadic pastoralists; diamond – agro-pastoralists. The populations are colored according to their language family: red – Afro-asiatic; blue – Nilo-Saharan; green – Niger-Kordofanian; yellow – Khoisan; black – Italic and Basque.

Phylogenetic analysis

The network analysis on the chromosomes carrying E haplogroups was robust enough with a main cluster near the root represented by Kunama (KUN) encompassing most of Eritreans and Sudanese populations, including Nilo-Saharan and Afro-Asiatic speakers suggesting that linguistic divergence is either a subsequent event to population divergence, language replacement or that the two linguistic families may have shared a common origin. The Southern African populations, which include the Khoisan and Bantu of South Africa populations, are shown to be divergent from the East African larger cluster through its connection to the Somali population. The network also suggests that dispersal of the haplogroup to Southern Africa may reflect the spread of pastoralism from North East Africa.[5] The Yemeni, Saudi Arabia and Oman populations on the other hand form a Near Eastern group. The link between the Yemeni and Omani populations with Afar and Saho populations from Eritrea could be attributed to the geographical proximity and possibly past genetic history. The Northern African populations tend to separate into two distinct groups: one containing Moroccan Arabs and Berbers and Saharawi, derived from the larger East African group and the other includes the Northern African populations of Algeria, Egypt and Tunisia, which forms a connection to both Europeans and Eritrean and Ethiopians hinting to recent genetic relationship between North and East African populations as is widely believed.[30]

The NJ tree, which was not rooted, on the other hand was quite robust in showing similar grouping to that of the network, MDS and PCA plots to imply a correlation with language and relevance to geography. With few exceptions, all populations carrying the haplogroup were either pastoralists or had recorded history of pastoralism. The populations that made exceptions, includes Hausa, Fur and Masalit, have strong agricultural practices, while the latter is thought to have recent history of mixed farming or foraging. The other exceptions are Copts from Egypt and Tigrigna from Eritrea, both with documented history of agricultural practices albeit historically part of larger communities with established pastoralist practices. The Nilo-Saharan speakers and Niger-Kordofanian were confined to the cluster from Sudan and Eritrea.

Genetic structure and population differentiation

The MDS and the PCA plots (Figure 3 and Supplementary Figure S2, respectively) generated from the E haplogroup frequency data portrayed similar pattern that complement the network result. Generally four main clusters can be identified from the MDS and PCA plots. In the MDS plot, one of the main clusters (grey shaded) constitutes almost all Eastern Africans including most Eritrean and Sudanese populations. The Saho and Afar populations of Eritrea tend to cluster with the Near Eastern or Arabian populations (brown shaded). The West and Southern African populations (blue shaded) form the third cluster, while North African populations forming the fourth cluster (green shaded). Interestingly, populations from Egypt, Tunisia and Ethiopia (Ethiopian Jews) assumed an intermediate position between the East African and Near Eastern clusters. The PCA (Supplementary Figure S2) also gave the same result clustering the majority of East Africans (grey shaded) in the first component and North Africans (brown shaded) separated from Middle East populations (blue shaded) in the second component. The first two components account 83% of the variations observed.
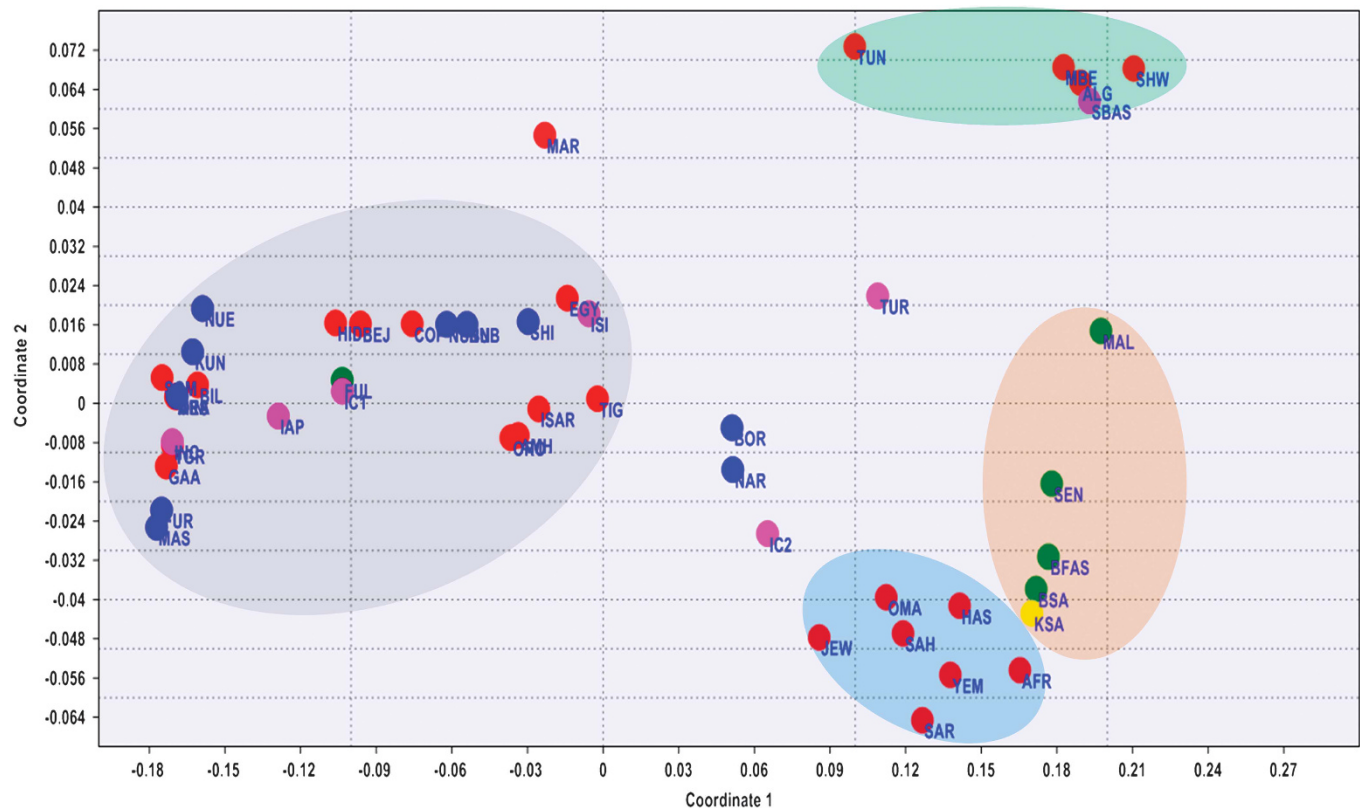

Figure 3 MDS plot based on the *FST* values generated from Arlequin 3.11 and using *Rho* similarity measures and with stress value of 0.07101. Populations' descriptions are given in Supplementary Table S1.

As indicated in the AMOVA summary (Table 1), when Eritrean populations were grouped according to geographic location, most of the genetic variance (82.11%) was found within populations; a value that is similar to that obtained (82.44%) when the populations were grouped according to their linguistic affiliation. Variance among populations within the linguistic groups was 14.71%, which is slightly higher than the variance among the geographic groups (13.17%). The genetic variance among the linguistic and geographical groups was 2.85% and 4.71%, respectively. AMOVA analysis was also carried out to see the variation of populations from this study and from published works in relation to their linguistic and geographic affiliation. For this purpose, populations were grouped as Afro-Asiatic, Nilo-Saharan and Niger-Kordofanian with the exclusion of Middle East and Europeans populations in both cases. Most of the genetic variance (52.66%) was found to be within populations. The genetic variance among groups and populations within groups were 18.73% and 28.66%, respectively. The AMOVA result after grouping the population into North, South, West and East Africans was different from grouping the populations according to their linguistic affiliation. The results were 25.89% variance among groups, 19.63% among populations within groups and 54.48% within populations. Mantel test showed no correlation between geographical isolation and linguistic affiliation of the populations and their genetic distance.

DISCUSSION

The Sahel, which extends between the Atlantic coast of Africa and the Red Sea plateau, represents one of the least sampled areas and populations in the domain of human genetics. The position of Eritrea adjacent to the Red Sea coast provides opportunities for insights regarding human migrations within and beyond the African landscape.

Worth noting in the current data set is the absence of differentiation of Eritrean populations along their geographical and linguistic affiliation, which may be a reflection of their admixture[42] or a common founding population with subsequent drift. Sharing of derived alleles for E and other more deep Y-chromosome lineages (unpublished data) of Eritreans with other populations from the region renders this part of East Africa a likely scene for some of the earliest demographic episodes within as well as subsequent expansion off the continent; a scenario that seems to corroborate paleontological, archeological and genetic evidences.[9,43–45]

The network cluster associated with the Eritrean Nilo-Saharan Kunama (Figure 1) may represent an expansion event following the out-of-Africa migration,[31,46] possibly close to the origin of the ancestral Y-chromosome clades.[47–49] The expansion, carrying the diversified E-P2 mutation, may be responsible for the migration of male populations to different parts of the continent and henceforth

Table 1 Summary of AMOVA analysis

Source of variations	Category	% Variation	Fixation indices (P-value)
Groups	Geography	18.73	FSC: 0.35259 (0.00000)
	Linguistic	25.89	FSC: 0.26488 (0.00000)
Populations within groups	Geography	28.66	FST: 0.47383 (0.00000)
	Linguistic	19.63	FST: 0.45523 (0.00000)
Among populations	Geography	52.62	FCT: 0.18728 (0.00000)
	Linguistic	54.48	FCT: 0.25873 (0.00000)

The grouping of populations in their linguistic and geographical location was done as discussed in the text.

the rise and spread of the bearers of the macrohaplogroup.[50] These type of population movements, or demic expansions, driven by climatic change and/or spread of pastoralism and to some extent agriculture,[51–54] are not uncommon in human history. This scenario is more substantiated by the refining of the E-P2 (Trombetta et al[35]) and its two basal clades E-M2 and E-M329, which are believed to be prevalent exclusively in Western Africa and Eastern Africa, respectively.

Interestingly, this ancestral cluster includes populations like Fulani who has previously shown to display Eastern African ancestry, common history with the Hausa who are the furthest Afro-Asiatic speakers to the west in the Sahel, with a large effective size and complex genetic background.[23] The Fulani who currently speak a language classified as Niger-Kordofanian may have lost their original tongue to associated sedentary group similar to other cattle herders in Africa a common tendency among pastoralists. Clearly cultural trends exemplified by populations, like Hausa or Massalit, the latter who have neither strong tradition in agriculture nor animal husbandry, were established subsequent to the initial differentiation of haplogroup E. For example, the early clusters within the network also include Nilo-Saharan speakers like Kunama of Eritrea and Nilotic of Sudan who are ardent nomadic pastoralists but speak a language of non-Afro-Asiatic background the predominant linguistic family within the macrohaplogroup.

The subclades of the network some of which are associated with the practice of pastoralism are most likely to have taken place in the Sahara, among an early population that spoke ancestral language common to both Nilo-Saharan and Afro-Asiatic speakers, although it is yet to be determined whether pastoralism was an original culture to Nilo-Saharan speakers, a cultural acquisition or vice versa; and an interesting notion to entertain in the light of the proposition that pastoralism may be quite an antiquated event in human history.[17] Pushing the dates of the event associated with the origin and spread of pastoralism to a proposed 12 000–22 000 YBP, as suggested by the network dating, will solve the matter spontaneously as the language differences would not have appeared by then and an original pastoralist ancestral group with a common culture and language[50] is a plausible scenario to entertain. Such dates will accommodate both the Semitic/pastoralism-associated expansion and the introduction of *Bos taurus* to Europe from North East Africa or Middle East.[55] The network result put North African populations like the Saharawi, Morocco Berbers and Arabs in a separate cluster. Given the proposed origin of Maghreb ancestors[56–59] in North Africa, our network dating suggested a divergence of North Western African populations from Eastern African as early as 32 000 YBP, which is close to the estimated dates to the origin of E-P2 macrohaplogroup.[30,60] It can be further inferred that the high frequency of E-M81 in North Africa and its association to the Berber-speaking populations[25,30,32,60,61] may have occurred after the splitting of that early group, leading to local differentiation and flow of some markers as far as Southern Europe.[30,60,62]

A branching in the network may once again represent an episode of human migration that carried the haplogroup E-M35 and its subhaplogroups farther to the western coast of the Red Sea to Yemen, Oman and Saudi Arabia and concurrently down to Southern Africa as part of a more recent Y chromosome motivated out of Africa migration episode.

The PCA and MDS display similar interesting grouping of the Afar and Saho populations of Eritrea with their Near Eastern Arabian populations to conjure up on the genetic relationship of the two sides of the Red Sea. The arrival of the E-M35 and derived subclades, for

example, E-M123/E-M34, to Arabia appears to be strongly linked to expansion into East Africa, North Africa, Europe, Southern Africa, an event that is likely related to pastoralism, hastened by its advent and amenable for analysis and dating using approaches similar to what was proposed for the co-migration of Y chromosome and disease traits.[63]

The presence of archeological[10–13] and agro-pastoral[9,14,16] evidences from this side of the Red Sea and the history of migration of animals across the Red Sea,[64] however, calls for more molecular dissection of common haplogroups shared by these coastal populations. As suggested by others, this may give clues not only to the origin of E-M123, J-M267, K-M70, but also to the origin of Semitic languages.[65,66] Indeed the trail of such historical movements are detectable by molecular signatures of markers like Y chromosome giving insights into episodes of even more regional nature, for example, the high frequency of E-V32 in Eritrea, in concordance to oral history, supports the historical ties between North East Africa (Egypt) and East Africa including Eritrea, Sudan, Ethiopia and Somalia.

CONCLUSION AND FUTURE WORK

Although most of the data sets in our study define the deep ancestry of the phylogeny, they still shed some information to our interpretations of recent phenomena such as the current genetic diversity of the E haplogroup in an implication to the origin and spread of Afro-Asiatic languages and to the history of pastoralism.[67] Such perspectives, however, should be tested by using more recently derived markers[5,47] within the major haplogroups to explain the archeological findings and the historical and current demography of the region. Moreover, more comparative genetic analysis between the two sides of the Red Sea, specially emphasizing on E-M123/E-M34 or E-M78 haplogroups, will not only refine the route of exit of *H. sapiens sapiens* from East Africa but also the genealogies of Afro-Asiatic languages in the region.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Underhill PA, Jin L, Lin AA *et al*: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; **10**: 996–1005.
2 Wei W, Ayub Q, Chen Y *et al*: A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res* 2013; **23**: 388–395.
3 Cavalli-Sforza LL: Genes, peoples, and languages. *Proc Natl Acad Sci USA* 1997; **94**: 7719–7724.
4 Cruciani F, Trombetta B, Sellitto D *et al*: Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur J Hum Genet* 2010; **18**: 800–807.
5 Henn BM, Gignoux C, Lin AA *et al*: Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proc Natl Acad Sci USA* 2008; **105**: 10693–10698.
6 Chiaroni J, Underhill PA, Cavalli-Sforza LL: Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci USA* 2009; **106**: 20174–20179.
7 Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J: Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci USA* 1988; **85**: 6002–6006.
8 Cavalli-Sforza LL, Feldman MW: The application of molecular genetic approaches to the study of human evolution. *Nat Genet Suppl* 2003; **33**: 266–275.
9 Walter RC, Buffler RT, Bruggemann JH *et al*: Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* 2000; **405**: 65–69.
10 Beyin A: The Bab al Mandab vs the Nile-Levant: an appraisal of the two dispersal routes for early modern humans out of Africa. *Afr Archaeol Rev* 2006; **23**: 5–30.
11 Beyin A, Shea JJ: Reconnaissance of prehistoric sites on the Red Sea coast of Eritrea, NE Africa. *J Field Archaeol* 2007; **32**: 1–16.
12 Mayer DEB-Y, Beyin A: Late stone age shell middens on the Red Sea coast of Eritrea. *J Isl Coast Archaeol* 2009; **4**: 108–124.
13 Beyin A: A surface middle stone age assemblage from the Red Sea coast of Eritrea: implications for upper Pleistocene human dispersals out of Africa. *Quat Int* 2013; **300**: 195–212.
14 Schmidt P: Urban precursors in the Horn: early 1st-millennium BC communities in Eritrea. *Antiquity-Oxford* 2001; **75**: 849–859.
15 Schmidt PR: Variability in Eritrea and the archaeology of the Northern Horn during the first millennium BC: subsistence, ritual, and gold production. *Afr Archaeol Rev* 2010; **26**: 305–325.
16 Curtis MC: Relating the ancient Ona culture to the wider Northern Horn: discerning patterns and problems in the archaeology of the first millennium BC. *Afr Archaeol Rev* 2009; **26**: 327–350.
17 Černý V, Pereira L, Musilová E *et al*: A genetic structure of pastoral and farmer populations in the African Sahel. *Genome Evol* 2011; **28**: 2491–2500.
18 Kivisild T, Reidla M, Metspalu E *et al*: Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet* 2004; **75**: 752–770.
19 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleic cells. *Nucleic Acids Res* 1988; **16**: 1215–1217.
20 Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 2008; **18**: 830–838.
21 Y Chromosome Consortium (YCC). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; **12**: 339–348.
22 Underhill PA, Passarino G, Lin AA *et al*: The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001; **65**: 43–62.
23 Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME: Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history. *Am J Phys Anthropol* 2008; **137**: 316–323.
24 Semino O, Santachiara-benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PPA: Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 2002; **70**: 265–268.
25 Cruciani F, Santolamazza P, Shen P *et al*: A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 2002; **70**: 1197–1214.
26 Cadenas AM, Zhivotovsky La, Cavalli-Sforza LL, Underhill Pa, Herrera RJ: Y-chromosome diversity characterizes the Gulf of Oman. *Eur J Hum Genet* 2008; **16**: 374–386.
27 Abu-Amero K, Hellani A, Gonzalez AM, Larruga JM, Cabrera VM, Underhill PA: Saudi Arabian Y-chromosome diversity and its relationship with nearby regions. *BMC Genet* 2009; **10**: 59.
28 Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinniog C, Roseman C: The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 2004; **74**: 532–544.
29 Sanchez JJ, Hallenberg C, Børsting C, Hernandez A, Morling N: High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. *Eur J Hum Genet* 2005; **13**: 856–866.
30 Semino O, Magri C, Benuzzi G *et al*: Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* 2004; **74**: 1023–1034.
31 Underhill PA, Shen P, Lin AA *et al*: Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000; **26**: 358–361.
32 Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA: High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between Northwestern Africa and the Iberian peninsula. *Am J Hum Genet* 2001; **68**: 1019–1029.
33 Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL: Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 2004; **114**: 127–148.
34 Scozzari R, Cruciani F, Malaspina P *et al*: Differential structuring of human populations for homologous X and Y microsatellite loci. *Am J Hum Genet* 1997; **61**: 719–733.
35 Trombetta B, Cruciani F, Sellitto D, Scozzari RA: New topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. *PLoS One* 2011; **6**: e16073.
36 Bandelt HJ, Forster P, Röhl a: Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; **16**: 37–48.
37 Repping S, Daalen S, van, Brown L *et al*: High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet* 2006; **38**: 463–467.
38 Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees'. *Mol Biol Evol* 1987; **4**: 406–425.
39 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods research resource. *Mol Biol Evol* 2011; **28**: 2731–2739.
40 Hammer Ø, Harper DAT, Ryan PD: PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* 2001; **4**: 1–9.

1392

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

42 Pollera A: *The Native Peoples of Eritrea; Translated by Linda Lappin*. New Jersey: Red Sea Press, 2002.

43 Oppenheimer S: Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos Trans R Soc Lond B Biol Sci* 2012; **367**: 770–784.

44 McEvoy B, Powell J, Goddard M: Human population dispersal out of Africa estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res* 2011; **21**: 821–829.

45 Fernandes V, Alshamali F, Alves M *et al*: The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *Am J Hum Genet* 2012; **90**: 347–355.

46 Jobling M, Tyler-Smith C: The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 2003; **4**: 598–612.

47 Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, Scozzari R: A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *Am J Hum Genet* 2011; **88**: 814–818.

48 Scozzari R, Massaia A, Atanasio ED *et al*: Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree. *PLoS One* 2012; **7**: e49170.

49 Mendez FL, Krahn T, Schrack B *et al*: An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am J Hum Genet* 2013; **92**: 454–459.

50 Lancaster AY: Haplogroups, archaeological cultures and language families: a review of the possibility of multidisciplinary comparisons using the case of E-M35. *J Genet Geneal* 2009; **5**: 35–65.

51 Mellars P: Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci USA* 2006; **103**: 9381–9387.

52 Zheng H-X, Yan S, Qin Z-D, Jin L: MtDNA analysis of global populations support that major population expansions began before Neolithic time. *Sci Rep* 2012; **2**: 745.

53 Gupta AK: Origin of agriculture and domestication of plants and animals linked to early Holocene climate amelioration. *Curr Sci* 2004; **87**: 54–59.

54 Diamond J, Bellwood P: Farmers and their languages: the first expansions. *Science* 2003; **300**: 597–603.

55 Pellecchia M, Negrini R, Colli L *et al*: The mystery of Etruscan origins: novel clues from Bos taurus mitochondrial DNA. *Proc R Soc London B* 2007; **274**: 1175–1179.

56 Ennafaa H, Cabrera VM, Abu-amero KK *et al*: Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet* 2009; **10**: 8.

57 Fadhlaoui-Zid K, Rodríguez-Botigué L, Naoui N, Benammar-Elgaaied A, Calafell F, Comas D: Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. *Am J Phys Anthropol* 2011; **145**: 107–117.

58 Henn BM, Botigué LR, Gravel S *et al*: Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet* 2012; **8**: e1002397.

59 Fadhaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar El gaaied A: Mitochondrial DNA heterogeneity in Tunisian berbers. *Ann Hum Genet* 2004; **68**: 222–233.

60 Cruciani F, Fratta R, La, Santolamazza P *et al*: Analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 2004; **74**: 1014–1022.

61 Arredi B, Poloni ES, Paracchini S *et al*: A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 2004; **75**: 338–345.

62 Scozzari R, Cruciani F, Pangrazio A *et al*: Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 2001; **62**: 871–884.

63 Bereir RE, Hassan HY, Salih NA: Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel. *Eur J Hum Genet* 2007; **15**: 1183–1185.

64 Wildman DE, Bergman TJ, al-Aghbari A *et al*: Mitochondrial evidence for the origin of hamadryas baboons. *Mol Phylogenet Evol* 2004; **32**: 287–296.

65 Cruciani F, Fratta R, La, Trombetta B *et al*: 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* 2007; **24**: 1300–1311.

66 Kitchen A, Ehret C, Assefa S, Mulligan CJ: Bayesian phylogenetic analysis of Semitic languages identifies an early bronze age origin of Semitic in the Near East. *Proc R Soc London B* 2009; **276**: 2703–2710.

67 Batini C, Ferri G, Destro-bisol G *et al*: Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Mol Biol Evol* 2011; **28**: 2603–2613.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)