Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Y chromosome and MT-COX2 variation in some Afro-Asiatic and Nilo-Saharan language families groups in Ethiopia

Solomon Balami Adugna
Salale University
Tura Safawo Jarso
tura.safawo@slu.edu.et

Salale University
Endashaw Bekele
Addis Ababa University
Muntaser Ibrahim
University of Khartoum

Research Article

Keywords: Y chromosome, MT-COX2, Haplogroup, PCR-RFLP, AS-PCR

Posted Date: May 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9696448/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Y chromosome and *MT-COX2* variation in some Afro-Asiatic and Nilo-Saharan language families groups in Ethiopia

Solomon Balemi,[1] Tura Safawo[1*], Endashaw Bekele,[2] and Muntaser Ibrahim[3]

Department of Biology, College of Natural Sciences, Salale University, P.O.Box 376, Fiche, Oromia, Ethiopia. E-mail: solbaa2000@yahoo.com
Department of Microbial, Cellular and Molecular Bioliogy, College of Natural Siences, Addis Ababa University, P.O.Box 1176, Addis Ababa, Ethiopia. E-mail: endashawbw@gmail.com
Departmnment of Molecular Bioliogy, Medical Campus, Khartoum University, Khortum, Sudan. E-mail: mibrahim@ied.org

Corrospoding author: Tura Safawo Jarso (turajarso21@gmail.com);Department of Biology, College of Natural Sciences, Salale University

Abstract: The patterns of Y chromosome and *MT-COX2* Variations were studied in five ethnic groups from Afro-Asiatic and Nilo-Saharan language families of Ethiopia. The Y chromosome macro-haplogroups (A, B, E and F) were genotyped in 120 unrelated adult males. The result shows haplogroup A-M13, B-M60, E-M78 and F-M89 were present with percentages of 44.04, 39.28, 86.11 and 21.42 % respectively, showing higher Y chromosome diversity than elsewhere in Africa. The ancient haplogroups A-M13 and B-M60 and the most widespread haplogroup E-M78 was present at high frequencies in Nilo-Saharan than Afro-Asiatic speaking groups. Intersetingly, haplogroup E-M78 subclade E1b1b1a1 was remarkably high. Based on *MT-COX2* sequence (*Genbank acc. no. MK955250-MK955274*), 15 substitutions (SNPs) were observed, of which the majority (80%) were Transitions. Haplotype variations analysis revealed a total of 13 haplotypes with 138 polymorphic sites. Evidence for the sign of past demographic expansion and high genetic diversity was noted. Phylogenetic analyses indicated that East Africans in general and Western Ethiopia in particular are at the root of the human evolutionary tree as a more ancestral lineages and the most probable location from which dissemination of human out of Africa has taken place, and suggest for further detailed studies on Ethiopian and African populations to contribute in filling the existing gaps in the *MT-COX2* database.

Key words: Y chromosome, *MT-COX2*, Haplogroup, PCR-RFLP, AS-PCR

1. INTRODUCTION

The studies of human population genetics in Ethiopia and the whole of Africa are crucial to understand the relationship of humans living outside Africa and the population genetic pattern within the African region. The characteristic of Y chromosomes and mtDNA makes them a much powerful tool for the evolutionary genetic studies of Ethiopian populations. Few studies made previously on the distribution of Y chromosomal haplogroup, and have shown that Ethiopian gene pools possess a considerable component variations, indicative of admixture with sub-Saharan African populations, Arabian and/or Near Eastern origin [1–5]. In other few studied groups, similar pattern of mtDNA variation was observed from Ethiopian populations [4,6] where both Sub-Saharan (L) and Eurasian mtDNA (M and N) haplogroups are observed. It was found that around one-half of Ethiopian (52.2%) mtDNA lineages belong to clades specific to sub-Saharan Africa, whereas the other half was divided between derived subclades of haplogroups M and N that are common outside Africa [6]. These finding gave an intermediate position of Ethiopian population, consistent with early admixture between sub-Saharan and Caucasian populations. This author have also indicated that both Semitic and Cushitic speaking populations of Ethiopia were close to each other and did not reveal significant differences by comparing samples of Amhara and Oromo peoples.

However, several of these previous studies are limited in a number of respects. They are mainly biased in terms of sampling appropriate and representative groups for genetic studies of Ethiopian populations. They included samples mainly from Semitic and to the lesser extent from the Cushitic populations without giving attention to Omotic and Nilo-Saharan speaking populations. In general, despite the great potential of studying human genetic variation in Ethiopia, which has crucial importance both regionally and globally, limited but encouraging progress has been made only recently. Plaster [7] and very recently [8] have conducted genetic studies with good ethnic coverage for populations in Ethiopia. However, studies involving Y-chromosome and mtDNA with appropriate sample size are very limited even today. Hence, the

present study was designed in order to contribute to the better understanding of the genetic structure of Ethiopian population groups through appropriate sample so that better understanding of the relationships between Ethiopian population groups and other African and non-African populations are worked out. To this end, analysis on sample of some Ethiopians population groups for a set of mtDNA and Y chromosome polymorphisms were made and compared with the data from other previous study on Ethiopian population, African and non-African populations.

2. MATERIALS AND METHODS

2.1. Ethical Approval and Declarations

Ethical approval for this study was obtained from the Salale University Institutional Research Ethics Review Committee (SlU-IRERC). Written informed consent was obtained from each participant who donated buccal cell samples for the study. Copies of the ethical approval letter and informed consent forms are included in the supplementary materials.

The research was conducted in accordance with internationally recognized ethical standards, including the World Medical Association's Declaration of Helsinki, as well as relevant national research ethics guidelines applicable in Ethiopia and the standard operating procedures and guideline of SlU-IRERC.

All participants were fully informed about the purpose, procedures, potential risks, and benefits of the study. Participation was voluntary, and participants were informed of their right to withdraw at any time without any consequences. Participant confidentiality and privacy were strictly maintained through anonymization and secure handling of both biological samples and associated data.

2.2. The Studied Populations and Sampling Strategies

The participants were recruited from Nuer ethnic group of Gambella region and Berta, Gumuz, Shinasha, and Mao ethnic groups of Benishangul-Gumuz region. They are located South West West, North west of Ethiopia along the Nile Valley (Figure 1). From evolutionary point of view,

these areas look key strategic region and are regarded as the corridors for the migration of human populations out of Africa. Also, there is an assumption that they are oldest from the perspective of human origin. These groups speak Nilotic and Omotic languages and thus they represent two major language families spoken in Ethiopia; Afro-Asiatic and Nilo-Saharan. 55 individuals from Nuer, 30 individuals from each of Berta and Gumuz, and 20 individuals from each of Shinasha and Mao, all are above the age of 18 up to 60 years old. These make a total of 155 samples representing five different ethnic group and two different geographic regions of Ethiopia.

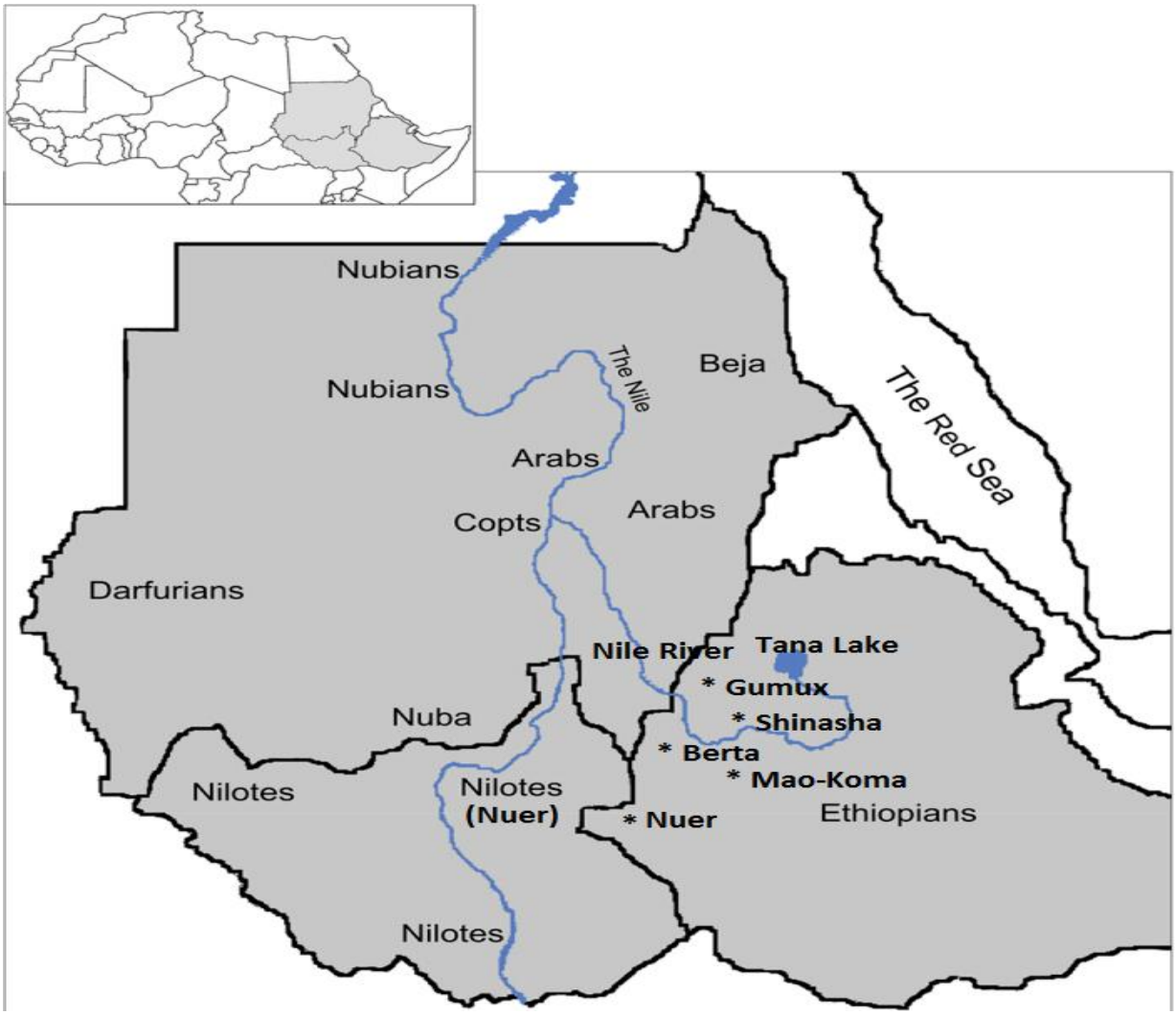


Figure1. A map of Ethiopia showing the location of study populations and the two neighbors Sudan (www. Idpiuk.org): The stars indicate the approximate locations of study sites. z

2.3. DNA Extraction

Buccal swabs samples were collected from all 155 individual volunteer participants. DNA extraction was performed according to Guanidine Chloride method as described by [9] with few modifications. Then agarose gel electrophoresis was performed to determine the presence or absence of genomic DNA in the extracted samples. The amount and quality of isolated DNA was measured using a Nano drop device (ND-1000 Spectrophotometer).

2.4. Y Chromosome Genotyping

The bi-allelic markers; M1 (YAP, *Alu* polymorphism), M13, M42, M60, M89 and M78 were genotyped in 124 unrelated males (out of total 155 participants) using PCR amplification. Gel electrophoresis was performed to determine the size of the amplified DNA product and categorized them as YAP + (insertion) or YAP- (deletion). Whereas YAP+ was further genotyped by allele specific PCR amplification to investigate the distribution of M78 binary sub clades, the YAP- were further screened for the other binary markers (M13, M42, M60, M89) using PCR-RFLP analysis. The markers were genotyped using oligonucleotide primers (Table 1).

Table1. Oligonucleotide primers used to generate the different Y chromosome haplogroups

PCR conditions			Polymorphism conditions		
Marker	Primer pairs (5' – 3')	Primer References	Enzymes	Ancestral alleles (bp)	Derived alleles (bp)
YAP	F:caggggaagataaagaaata R:actgctaaaaggggatggat	[10]	-	YAP- (150)	YAP+ (450)
M13	F:tcctaacctggtggtctttc R:tgagccatgattttatccaac	[11]	MboI (NEB)	G (156 + 77)	C (233)
M42	F:aaagcgagagattcaatccag	[12]	AluI (NEB)	A(340)	T(295+ 45)

	R:ttttagcaagttaagtcaccagc				
M60	F:gcactggcgttcatcatctg R:atgttcattatggttcaggagg	[12]	MboI (NEB)	GATC (241 + 147)	1bp-insertion GTATC (389)
M78	F1:cacttaacaaagatacttctttcc F2:cacttaacaaagatacttctttct R:attactttcctaggttctccca	[13]	-	C (319)	T (319)
M89	F:acagaaggatgctgctcagctt R:gcaactcaggcaaagtgagacat	(Akey, *et. al.,* 2001)	NlaIII (NEB)	C(20 + 67)	T(87)

2.5. Mitochondrial DNA Genotyping and Sequencing

After optimization*, MT-COX2* gene was amplified by conventional PCR using Maxime PCR PreMix Kit (i-Taq). *COXIIF* and *COXIIR* primers were used to obtain 623 bp fragment of the 684 bp region of the gene (Table 2 and Figure 2). Ten individuals from each ethnic group were randomly chosen for *MT-COX2* sequencing (total 50 samples). Of these the amplification was successful for 30 samples. Since some of these positive PCR products were found to contain multiple bands, they were sent for gel extraction sequencing at Macrogen, Inc. (Korea). Out of 30 sequences, only 25 were found to be high quality (*Genbank acc. no. MK955250-MK955274*). All the Mao's sequences were not good and were excluded. Only clean genomic region were included in the analysis after both end were trimmed. Both forward and reverse strands were used to check the correctness of the detected mutations. These sequences were combined with 269 global sequences obtained from public human mtDB database (http://www mitoma org) making a total of 294 sequences.

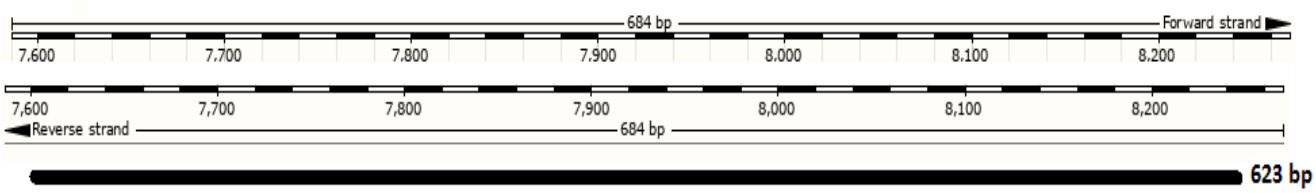


Figure 2. Cytochrome C Oxdase II location of mitochondrial Chromosome from 7586-8269 [14]

T**able 2. Primer pairs of mitochondrial** *MT-COX2* **genotyping**

Primers	Primer sequence	Region	Primer Reference
COXII F	5'-TAGGTCTACAAGACGCTACTT-3'	623 bp	Elhassan, *et. al*., 2004
COXII R	5'-AATTAATTCTAGGACGATGG-3'		

3.6. Data Analysis

Y chromosome phylogenetic tree was designed by hand according to updated version [15] of Y Chromosome Consortium nomenclature[16] (Y Chromosome Consortium, 2002). Principal Coordinate Analysis (PCA) was performed by employing PAST (paleontological statistics) algorithms, version 2.11 [17]. For *MT-COX2* Sequence analysis, the total 294 sequences were aligned using the program CLUSTALX version 1.81 (ClustalX, 1997)[18] on the sequence alignment editor Bioedit software [19] using Youruba refrence sequences [20]. Genetic distances were computed and were used for nucleotide sequence divergence between groups using MEGA5 software version 5.0 based on Tajima's Nei model [21,22]. Molecular Diversity and Neutrality Test were calculated using DNASP version 5.10[23] and ARLIQUIN 3.1 software [24]. The *MT-COX2* Phylogenetic Tree were performed using MEGA5[25] based on the model of [26].

3. RESULTS

3.1. Y chromosome Genotyping

The NRY region was successfully amplified as a 150 bp for YAP- and 450bp for YAP^{+} (Figure 3) using PCR amplification. All samples carrying YAP + were further typed for E-M78 sub clades revealing T/C SNPs and the sample PCR amplification results are indicated below (Figure

4). Also, all YAP- samples were genotyped for binary markers and the sample gel electrophoresis picture of the haplogroup (M42 and M60) are depicted below (Figure 5 and 6). Of 124 genotyped male, 4 individuals did not yield any amplified product for YAP+ and YAP- analysis.

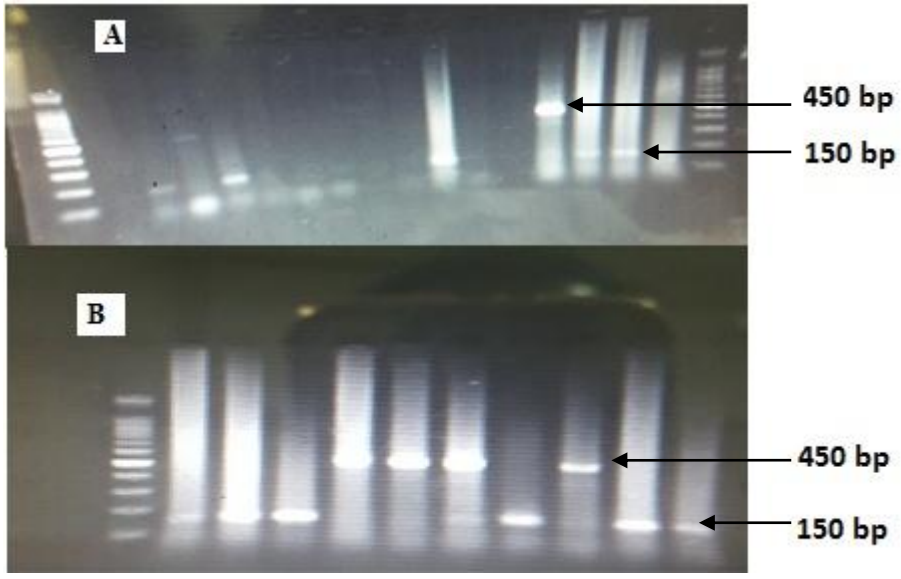


Figure 3. PCR Amplification of YAP locus to detect the +/- of *Alu* insertion. A) Bertha B) Nuer. 100bp DNA marker used. YAP+ is 450 bp and YaAP- is 150 bp. The second lane (A) and the last lane (B) are negative control. The products were run on 2% agarose gel.

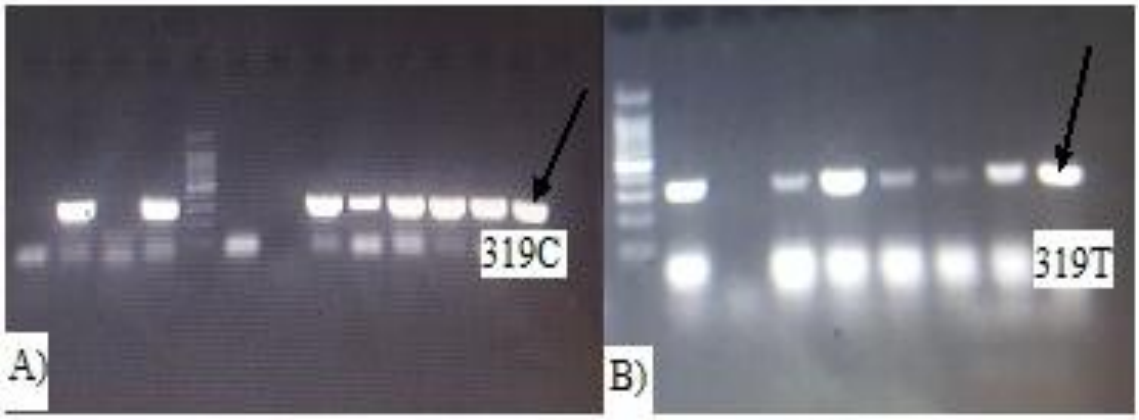


Figure 4. Sample PCR Amplification of YAP+ for M78. A) M78 carrying ancestral allele C B) M78 carrying derived allele T. 100bp marker used. The product were run on 2% gel.

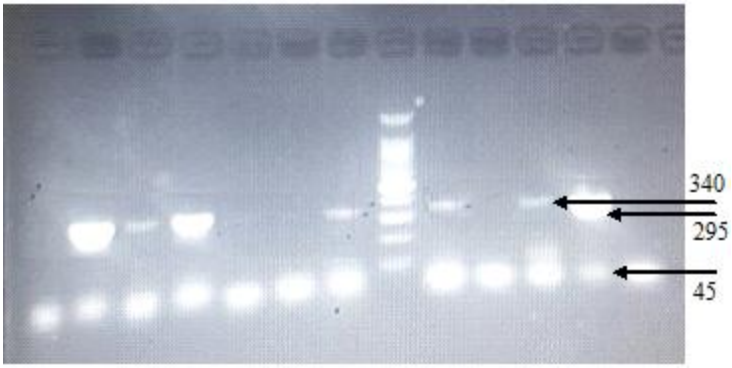


Figure 5. Sample RFLP Analysis of YAP- for M42. Lane 1, negative control, lane 8, 100bp marker.The product were run on 2% agarose gel.

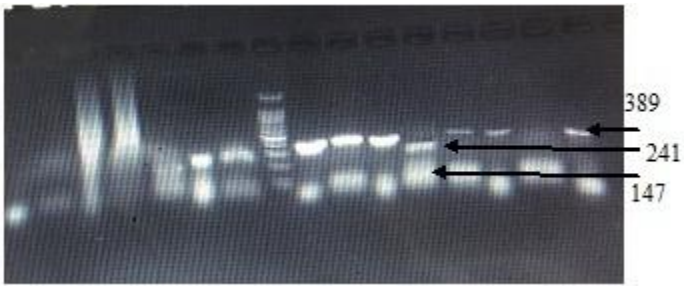


Figure 6. Sample RFLP Analysis of YAP- for M60. Lane one (from right to left) negative control, lane 8, 100bp marker. The product were run on 2% agarose gel.

Y chromosome Variation

All genotyped haplogroups (A, B, E and F) were observed with percentages of 44.04, 36.9, 86.11and 21.42 % respectively. Both A-M13 and B-M60 were present at high frequencies. Haplogroups E-M78 was by far the highest and they were found at high frequencies in Nilotic speaking than the Afro-Asiatic speaking groups. In the Nilo-Saharan speakers, haplogroup A-M13 was found at higher frequencies in Nuer (37.83%) followed by Gumuz (18.92%). With similar pattern, B-M60 was found in Nuer (36.36%) followed by Berta and Gumuz each with 21.21%. Moreover, E-M78 was present at the highest fequencies in Nuer (51.62%) and followed by Berta and Gumuz each with 19.35% and they all are Nilotics. Unexpectidely, haplogroup F-M89 were found at high frequencies in Berta (55.55%) and Gumuz (33.33%) but at low frequencies in Nuer and completely absent from the Afro-Asiatic speaking groups. Moreover, haplogroup B-M42 from which B-M60 desended was found to be present in all studied groups. Y chromosome haplogroup variation and distribution among the populations of the present study

and few neighbour populations is shown (Figure 7). Y chromosome haplogroup affinity and nomenclature were performed according to the Y chromosome Consortium rules and its updated version [15,27].

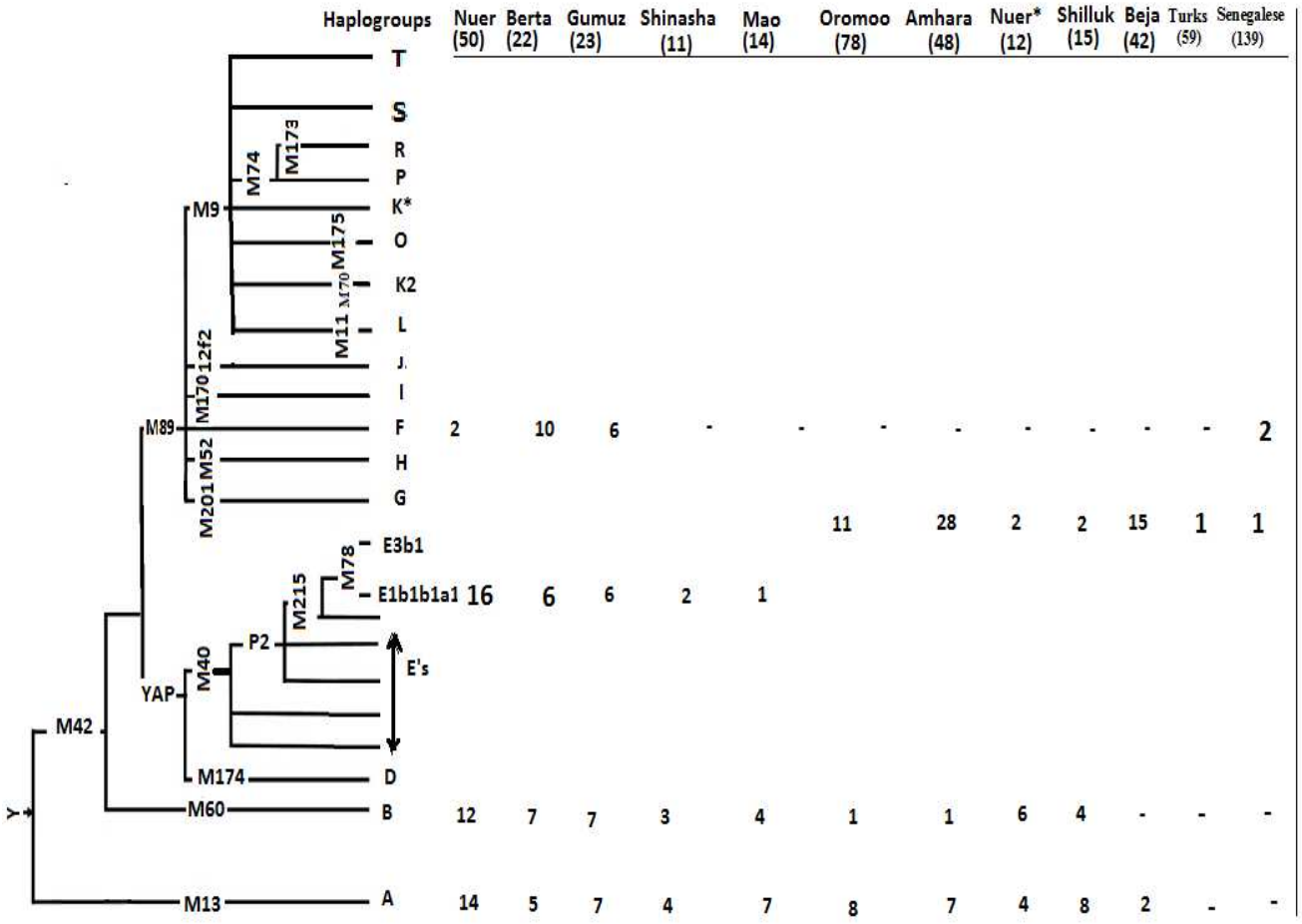


Figure 7. Phylogenetic distribution of the Y chromosome haplotypes and their frequencies in five Ethiopian populations in the present study, compared with three Sudanese [13]**, two Ethiopian ethnic groups (Amhara and Oromo) and Senegalese** [28] **and Turks** [29]**. Numbering of mutations and haplogroups nomenclature are according to** [27] **rules and updated by**[15]**. Nuer* is those belong to Sudan. The arrow shows the root of the tree.**

3.2. *MT-COX2* Genotyping

The sequence variants were visually counted in 25 Ethiopian sequences. Overall, fifteen different SNPs were identified of which, 12 (80%) were Transitions where as 3 (20%) were Transversions. Relatively the Nilo-Saharan groups exhibited highest number of synonymous

substitutions and lower number of nonsynonymous mutations than the Afro-Asiatic groups. The shared haplotypes and haplotypes frequencies among combined populations were calculated using Arlequin version 3.1[24] software. Thrteen (13) haplotypes with 138 polymorphic sites were observed among Ethiopian populations with haplotype diversity, Hd: 0.9233. But, fewer shared haplotype were observed among groups of the populations. Similarly, the global *MT-COX2* sequences was combined with the present and were analyzed using Arlequin to look at inter continental and inter populations diversity. In general, high diversity of SNPS and haplotypes than the world populations was observed, result tracing maternal ancestry back to Africa.

3.3. Genetic Distances, PCA and phylogeny

The Y chromosome and *MT-COX2* genetic distances (Table 3 and 4) of Ethiopian populations were compared with some other Ethiopian and neighboring Africans and other populations (Table 5 and 6). Also, the Y chromosome PCA of present Ethiopian and combined data with few neighbor Sudan was performed (Figure 8 A and B) based on F_{ST} genetic distances. Tests of neutrality were done for Y chromosome and *MT-COX2* sequences to search for demographic signs of population expansion. Moreover, Y chromosomes and *MT-COX2* phylogenetic relationships were observed. The Y chromosomes phylogeny shows populations which are generally grouped according to their geographic regions. *MT-COX2* sequences, phylogenetic trees: Neighbor-Joining (NJ) and Minimum Evolution (ME) respectively, showed similar topologies and branch length with few differences. The NJ tree of the 25 Ethiopian sequences is shown (Figure 9). Maximum Parsimony showed the evolutionary history for Ethiopian samples. When the present data are combined with some African and Asian populations, both NJ and ME phylogenetic tree of 166 individuals showed almost similar topology with a little bit difference from the maximum Parsimony (Figure 10).

Table 3. Y chromosome Pair wise Genetic Distances of Different five Ethiopian Populations. List of labels for population samples used in this table: 1: Nuer, 2: Berta, 3: Gumuz, 4: Shinasha and 5: Mao.

	1	2	3	4	5
[1]	0.000				

[2]	0.036	0.000			
[3]	0.021	0.049	0.000		
[4]	0.003	0.034	0.019	0.000	
[5]	0.133*	0.159*	0.143*	0.130*	0.000

*Significant at multiple tests adjusted P-Value

Markov chain length: 10,000 steps

Table 4. *MT-COX2* **Pair wise Genetic Distances of Different four Ethiopian Populations. List of labels for population samples used in this table: 1: Nuer, 2: Berta, 3: Gumuz, 4: Shinasha.**

	1	2	3	4
[1]	0.000			
[2]	0.099	0.000		
[3]	0.072	-0.107	0.000	
[4]	0.133	-0.034	-0.126	0.000

Matrix of significant F_{ST} P values; Significance Level=0.0500

Table 5. Y chromosome Pair wise Genetic Distances of Five Ethiopian populations compared with some other Ethiopian and Neighbor Africans. List of labels used in this table: 1: Nuer, 2: Berta, 3: Gumuz, 4: Shinasha, 5: Mao, 6: Oromo, 7: Amhara, 8: *Nuer, 9: Shilluk, 11: Beja**

[	1	2	3	4	5	6	7	8	9	10	11
[1]	0.0000										

[2] 0.0075 0.0000

[3] 0.0037 0.0075 0.0000

[4] 0.0019 0.0056 0.0019 0.0000

[5] 0.0037 0.0075 0.0000 0.0019 0.0000

[6] 0.0056 0.0093 0.0056 0.0037 0.0056 0.0000

[7] 0.0019 0.0056 0.0019 0.0000 0.0019 0.0037 0.0000

[8] 0.0037 0.0075 0.0037 0.0019 0.0037 0.0056 0.0019 0.0000

[9] 0.0037 0.0075 0.0037 0.0019 0.0037 0.0019 0.0019 0.0037 0.0000

[10] 0.0000 0.0075 0.0037 0.0019 0.0037 0.0056 0.0019 0.0037 0.0037 0.0000

[11] 0.0037 0.0075 0.0037 0.0019 0.0037 0.0019 0.0019 0.0037 0.0000 0.0037 0.0000

Nuer[*] belong to Sudanese group.

Table 6. *MT-COX2* Pairwise Genetic Distances of present Ethiopian populations compared with some other Ethiopian and Neighbor Africans and other populations. List of labels used in this table: 1: Hausa, 2: Nilotes, 3: Beja, 4: Nubian, 5: Nuba, 6: Ethiopian, 7: Eritrean, 8:Egypt, 9: Morocco, 10: Nigeria, 11:South Africa, 12:Tunisia, 13: San, 14: Pygymy, 15: Yoruba, 16: Saudi, 17: Yemen, 18: Israel, 19: Iraq, 20: Indian, 21: Chinese, 22: Japanese, 23: Ethiopian (present).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.00000																						
2	-0.85714	0.00000																					
3	-0.33333	-0.11702	0.00000																				
4	-0.27273	0.15625	0.22535	0.00000																			
5	-0.33333	0.05303	0.12281	0.18750	0.00000																		
6	0.00000	-0.21495	0.08046	0.11672	0.08046	0.00000																	
7	-1.00000	0.19631	0.42976	0.48454	0.45034	-0.32353	0.00000																
8	-0.82857	0.07527	0.14667	0.29139	0.26356	-0.24675	0.04245	0.00000															
9	1.00000	0.23529	0.50000	0.17647	0.36842	1.00000	0.90476	0.39048	0.00000														
10	0.00000	0.07143	0.33333	0.36364	0.33333	0.00000	-0.08434	-0.03337	1.00000	0.00000													
11	0.00000	-0.85714	-0.33333	-0.27273	-0.33333	0.00000	-1.00000	-0.82857	1.00000	0.00000	0.00000												
12	0.00000	-0.85714	-0.33333	-0.27273	-0.33333	0.00000	-1.00000	-0.82857	1.00000	0.00000	0.00000	0.00000											
13	0.00000	0.07143	0.33333	0.36364	0.33333	0.00000	-0.08434	-0.03337	1.00000	0.00000	0.00000	0.00000	0.00000										
14	0.00000	0.12641	0.33891	0.31653	0.30469	0.36842	0.47661	0.15556	0.75000	0.63415	0.00000	0.00000	0.63415	0.00000									
15	-0.80000	0.03150	0.13084	0.28532	0.24272	-0.20000	0.15523	-0.03226	0.55000	0.06250	-0.80000	-0.80000	0.06250	0.25000	0.00000								
16	1.00000	0.23304	0.46488	0.36652	0.37824	1.00000	0.80769	0.31915	1.00000	1.00000	1.00000	1.00000	1.00000	-0.20000	0.45455	0.00000							
17	-1.00000	0.00794	0.10714	0.21687	0.17808	-0.29032	0.11901	-0.00977	0.25000	0.00000	-1.00000	-1.00000	0.00000	0.09824	-0.02632	0.22830	0.00000						
18	0.16667	0.39474	0.50000	0.47826	0.47619	0.39394	0.66507	0.43637	0.54545	0.58333	0.16667	0.16667	0.58333	0.51163	0.48780	0.56044	0.25676	0.00000					
19	0.00000	-0.85714	-0.33333	-0.27273	-0.33333	0.00000	-1.00000	-0.82857	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.80000	1.00000	-1.00000	0.16667	0.00000			
20	0.00000	-0.21495	0.08046	0.11672	0.08046	0.00000	-0.32353	-0.24675	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.36842	-0.20000	1.00000	-0.29032	0.39394	0.00000	0.00000		
21	0.00000	0.07143	0.33333	0.36364	0.33333	0.00000	-0.08434	-0.03337	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.63415	0.06250	1.00000	0.00000	0.58333	0.00000	0.00000	0.00000	
22	-1.00000	-0.02644	0.20257	0.23286	0.20337	-0.20000	-0.02620	-0.04110	0.71429	0.18919	-1.00000	-1.00000	0.18919	-0.20000	0.01493	0.36842	-0.07989	0.45455	-1.00000	-0.20000	0.18919	0.00000	
23	0.00000	-0.21495	0.08046	0.11672	0.08046	0.00000	-0.32353	-0.24675	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.36842	-0.20000	1.00000	-0.29032	0.39394	0.00000	0.00000	0.00000	-0.20000	0.00000

Matrix of significant Fst P values; Significance Level=0.0500.

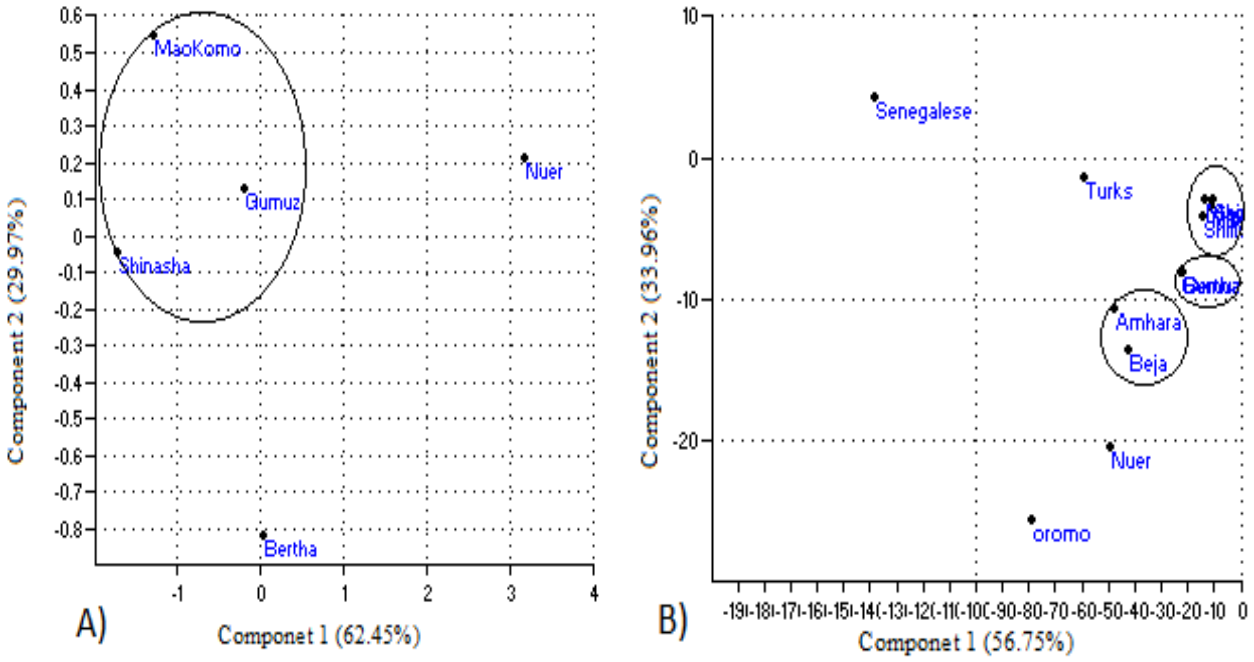


Figure 8. Principal Component Analysis based on Y chromosome haplogroups. A) Genetic affinities between populations in present study. B) Genetic affinities of Ethiopian populations of the present study compared with three Sudanese group [13]**, two Ethiopian groups (Amhara and Oromo) and Senegalese** [28]**and Turks** [29] **(A) and present and Circle shows the genetic affinities between populations.**

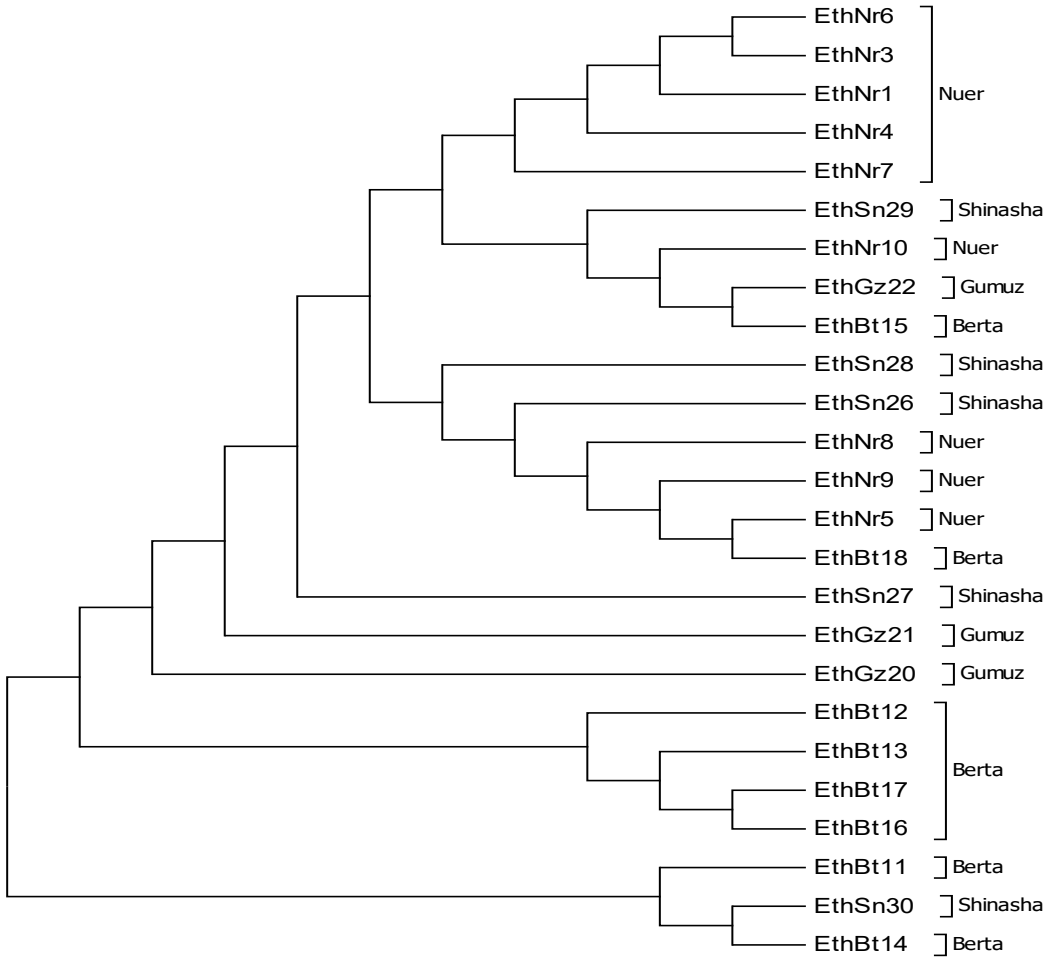


Figure 9. Evolutionary relationships using NJ Tree of 25 Ethiopian individuals based on the genetic distances of *MT-COX2* Sequences.

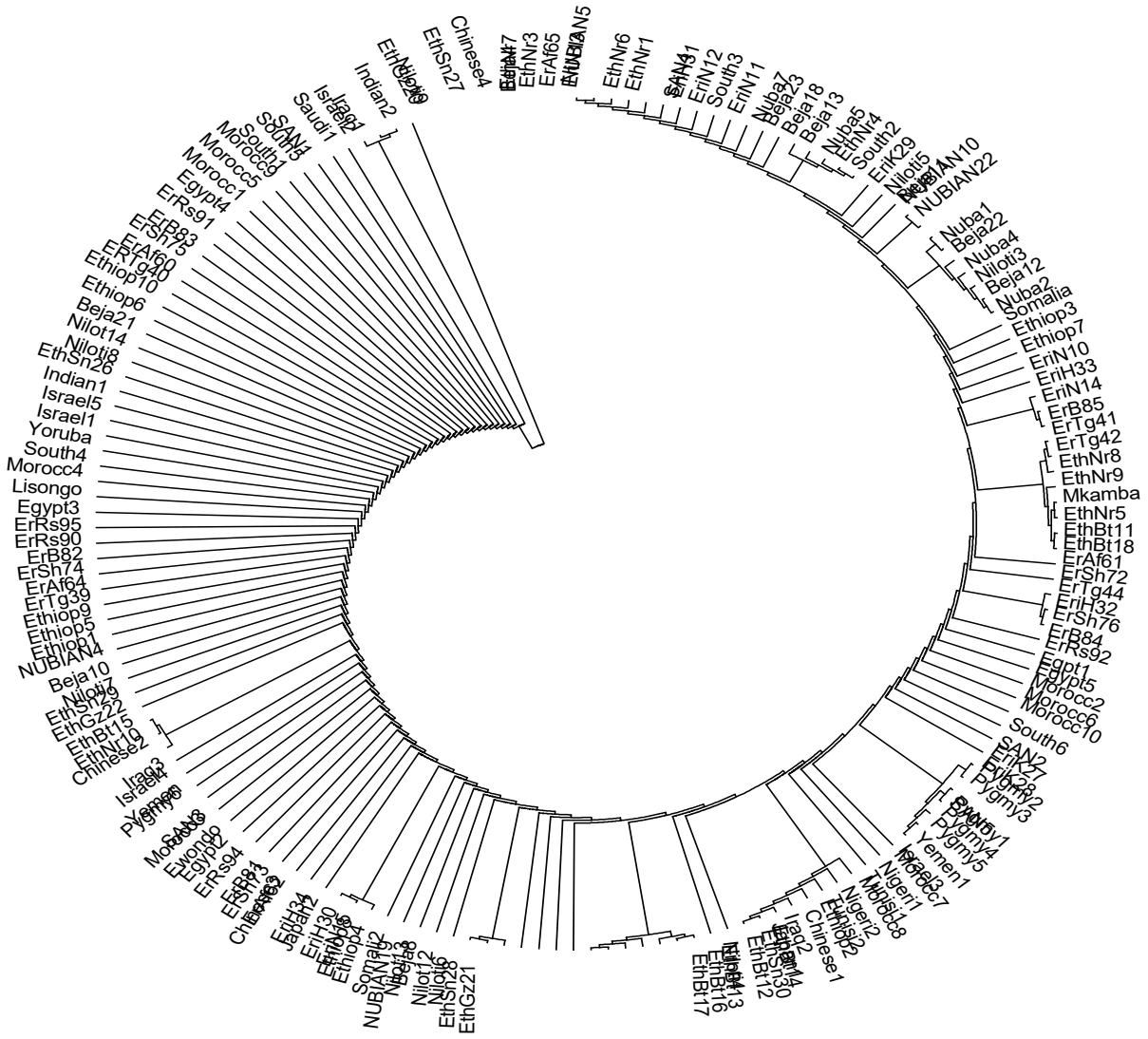


Figure 10. The evolutionary history using Maximum Parsimony Tree of 166 Africans and Asians populations based on the genetic distances of *MT-COX2* Sequences.

4. DISCUSSION

4.1. Y Chromosome Binary Markers in Ethiopia

Genotyping of YAP marker revealed YAP+ and YAP- polymorphism individuals with the percentages of 30 and 70 respectively. These show that the frequency of the YAP^{+} was moderately high in the studied populations. The previous general studies of YAP+ and YAP- polymorphism have reported higher frequencies of the YAP+ within the African populations [30,31] in agreement with the present study, suggesting that this marker would be an African

ancestry. The YAP^{+} in Africa falls in to haplogroups E. Moreover, very high frequency of the YAP+ was reported to have appeared in African, Japan and Tibetans population followed by the Western Eurasians populations [30].

Whereas, all YAP+ was further typed for E-M78 to obtain T/C SNPs, all the YAP- individuals were analyzed for A-M13, B-M60 and F-M89. The presences of these haplogroups were bserved among Ethiopian populations with higher frequencies of 44.04, 39.28, 86.11, 21.42 % respectively. E-M78 haplogroup was found to be the most frequent followed by haplogroups A-M13 and B-M60, which were present at higher frequencies and are exclusive to the African continent. In agreement with this, Gomes, *et. al*. [32] have reported that the majority of Y chromosomes in sub-Saharan Africa carry haplogroup A and B with a frequency of approximately 33% each. This author reported that Nilotic speaking groups contain high frequencies of these haplogroups than Afro-Asiatic, which is consistent with the present study.

The frequency of haplogroup A-M13 observed here support the previous studies. Within the Nilo-Saharan speaking groups, it was found at higher frequencies in Nuer (37.84%) followed by Gumuz (18.92%), but moderate in Afro-Asiatic. In support of this, A-M13 was reported to be prevalent in Nilo-Saharan populations [13,33]. The high to moderate frequencies of M13 among both Nilo-Saharan and Afro-Asiatic populations in East Africa, as well as the presence of closely related and shared haplotypes observed in world mtDNA data suggest the extensive contact and gene flow over a long period of time. Haplogroup A-M13 was the most common in the present sample with a frequency of 44.04%, greater than that already found in Nilotic groups from Sudan (30%) [13]. Likewise, small frequencies were found in several Afro-Asiatic speaking populations from East Africa [13,33]. This patterns of frequencies were also supported by autosomal data. As described by [34], whereas East Africa contains most of the A-M13 Y chromosomes, central and West Africans have low diversity. These together with the potentially bi-directional nature of migration along the Sahel increase the potential for the haplogroup to have arisen anywhere along the belt. The lack of comparative data for the sub clades of haplogroup A-M13 makes the analysis of these haplogroups difficult in Africa if not impossible.

Haplogroup B-M60 was found at the highest frequencies in samples of Nilo-Saharan speaking groups than the Afro-Asiatic speakers. It was found in Nuer with frequency of 36.36% followed

by Berta and Gumuz each with frequency of 21.21%. Similar study of the Y chromosome of populations in Sudan revealed higher frequencies of haplogroup B-M60 in Southern Sudanese (30%), Hausa (16%) and Nuba (14%) [13], which are close to the result of the present study. According to another studies, haplogroup B-M60B has been found in approximately 2.5% and 6% of Sudanese males and Malagasy males of Madagascar respectively[35,36] (Underhill, *et. al*., 2000 and Hurles, *et. al*., 2005). Like A-M13, the higher frequency of haplogroup B-M60 suggests the African ancestry of these haplogroups. It also support the report that haplogroup B-M60 is the oldest and diverse human Y haplogroups in African populations.

Haplogroup E1b1b1a (E-M78), previously called E3b1 (according to YCC (2002)) was found at highest frequency in Nuer with frequency of 51.62%. This result is in agreement with the original thought that E-M78 and its sub clades have originated in North East Africa (Cruciani, *et. al.,* 2007). It's distribution in studied Ethiopian group is highest than the high frequency reported in Europe, middle east and East Africa. Gebremeskel and Ibrahim,[37] also reported that majority of Y-chromosome haplogroup diversity in Africa is found within this haplogrup. They also suggest that it probably to have originated 21,000 – 32,000 YBP somewhere between the Red Sea and Lake Chad. This may suggest the Ethiopian origin of this subclades although further detail anlaysis with more representative sample are required for confirmation. It has also higher frequency in other Nilo-saharan speaking groups, Bertha and Gumuz both having frequency of 19.35%. Interestingly, the frequency of the E-M78 observed in Ethiopian Nilo-Saharan populations is higher than the percentages reported in the other African populations [13]. This may indicate that these populations are experiencing recent gene flow from other populations [38] or alternatively represent a genetically distinct Ethiopian populations, though these also require further extensive and in depth investigation in order to confirm. In this study, the analysis of E-M78 among Ethiopian samples confirmed the presences of deep terminal branch E1b1b1a and thus somewhat refined the phylogeny of haplogroup E-M78 in Ethiopia.

Haplogroup F-M89 were unexpectidely found in Nilo-Saharan speaking groups with high frequencies as in Berta (55.55%) and Gumuz (33.33%), but relatively at low frequencies in Nuer (11.11%), and completely absent from the Afro-Asiatic speaking groups (Figure 7). From personal communication, elders from the two groups have claimed their Arab ancestry from their present numerious Arabs like traditions and Islam religion. Thus, this mutation might have

disseminated to Out of Africa mainly to Asia, through the ancient Arabs, where it is currently found at highest frequencies. The F-M89 haplogroup individual was M89+ and thus considered to be in haplogroup F. Unfortunately, it was not tested for the presence of mutations that define the haplogoups descended from F (sub clades of F-M89) in order to confirm the present findings. The high frequency of F-M89 observed in the present samples may suggest that Haplogroup F-M89 appear to have originated in Africa, from where it was disseminated to outside of African continent early past during prehistory. To shed light on the hypothetical "Back to Africa" migration, Y-chromosome haplogroups (J, R1b and T) associated with this hypothesis should be genotyped. Also its high frequency may be due to small samples sizes and thus sampling effects of studied population. Y chromosome haplogroup variation and distribution among the five populations of the present study and few neighbour populations are illustrated (Figure 7).

4.2. Genetic Structure and Patterns of Distributions

The PCA plot based on F_{ST} values of the present samples showed that populations were grouped according to their geographic regions. The Nuer and Berta are genetically differentiated from each other and also from other studied groups. The Nuer and Berta might have differentiated from one another and from the rest populations because of the presence of high frequency of haplogroups E-M78 sub clades in Nuer and high frequency of Hamplogroup F-M89 in Berta. PCA plot based on F_{ST} values of combined data with few data from other Ethiopians, Sudanese, Turks and Senegalese populations, showed three main genetic clusters that was grouped according to linguistic and geographic variations except for Beja and Amhara (Figure 8). The first cluster includes populations who speak languages of the Nilo-Saharan family (Berta and Gumuz), one of linguistic family in Ethiopia. The second cluster includes populations (Beja and Amhara) who are essentially speakers of different languages belonging to the Afro-Asiatic family. The third cluster encompasses the Sudanese groups. These clusters are defined by the predominance of the ancient haplogroups A-M13, B-M60, and E-M78. 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 East Africa [12,28].The placement of the Ethiopian Nuer close to the Oromo group who speak language that belong to Afro-Asiatic family clearly revealed the importance of language relationship by revealing the presence of high frequencies of A-M13 in both groups and their geographic proximity. Although they are not close to each

other, Amhara and Beja are clustered together may be because of the presence of shared ancient haplogroup A-M13 and B-M60. The distribution of E- M78 sub clades indicates that the Beja are perhaps related as well to the Oromo on the basis of the considerable frequencies of E-M78 among Oromo in comparison to Amhara [39]. These in addition making the Cushitic Oromo more ancestral and also indicate the presence of the historical contact between Ethiopia and eastern Sudan and showed the strong correlation between linguistic and genetic diversity as these populations speak languages of the Afro-Asiatic family [4,13,40].

The AMOVA analysis highlights the population differentiation and shows association of linguistic with genetics in the data set. The analysis revealed that 85.2% of the genetic variation occurred within populations, which might suggest a low and less stable regional genetic structure in the studied population. The overall F_{ST} estimated (F_{ST} = 0.084) values within populations showed strong population differentiation. When the populations are grouped according to linguistic and geographic families, almost similar overall FST was observed. This may indicates that Y chromosome variation is partitioned among both geographic and linguistic groups, showing the absence of gene flow supporting the PCA plots.

4.3. The *MT-COX2* Variation in Ethiopian Populations

This study is the first to sequence and analyze the ***MT-COX2*** coding region among Ethiopian populations. The ***MT-COX2*** sequences was alligned against the Yoruba and fifteen SNPs were detected. There were a total of 3 non synonymous and 12 synonymous polymorphisms yielding an NS/S ratio of 0.25. This result is in agreement with the work of Elson, *et. al.* [41] which indicates that negative selection operated on the coding region of the human mtDNA during evolution. Small numbers of synonymus than non synonymous mutations observed is in agreement with ancestral sequence evolution concept that point out 'older' sequence changes in the human mtDNA coding sequences have relatively fewer transversion substitution than a set of 'younger' sequence changes.

High value of haplotype diversity were observed and is almost similar to a previous report based on the same investigated species from populations in Sudan using the same ***MT-COX2*** region as marker [42]. This could be attributed to the complex and variable nature of studied population with deep evolutionary history in the Western Ethiopia, a high mutation rate in the study fragment and the large population size of the species. However, no shared haplotype were

observed among all population groups in this study. Thus, the analysis of ***MT-COX2*** haplotype distribution showed the absence of single shared haplotypes to all studied individuals. This is a pattern frequently attributed to populations that have undergone past demographic expansion as described by [43,44]. In addition the absence of common haplotype indicates that there is no gene flow [45]. In comparisions to the continental *MT-**COX2*** sequences, the diversity within Ethiopians by far is greater than among the Ethiopian and the global mtDB sequences. In the combined ***MT-COX2*** sequences of Ethiopian and mtDB global sequences all haplotype was shared by Africans, except two that are only shared by Asians and Australians showing Africans are the genetic ancestor of the other population. By comparing the present and combined data, it is clear that the present fewer sequences have contributed 19.69% of the total haplotyoes resolved. This result is consistent with the report of [42]. In general, this result clearly indicated that high diversity of SNPs and haplotypes are present in Ethiopian samples than the global samples trace the matrilineal ancestry back to Ethiopia. This pattern of diversity support the idea that mtDNA in human populations of African origin are the most diverse in the world and that ancient human first existed and stayed in Africa for several thousands of years.

Tests of neutrality were done for Y chromosome and *MT-COX2* sequences to search for demographic signs of population expansion. Tajima's D test based on the Y chromosome haplogroups was statistically significant positive values for all populations but varies among populations, indicating the less possibility that these positive values are due to recent population demographic expansions. Instead the positive values could signify balancing selection. Likewise, based on the *MT-COX2* data, Tajima's D test (also Fu's *Fs)* analysis were performed to examine the historical demographic expansions of populations, which may play an important role in determining the patterns of genetic variability. Tajima's D test based on the *MT-COX2* region sequences was significantly negative for three populations (Nuer, Berta and Shinasha) and zero for Gumuz population (non significant). Non-significant values in Fu and Li's D* test statistics (-3.64183, $P < 0.02$) and Fu and Li's F* test statistic (-3.84989, $P < 0.02$) combined with a significant Fu's Fs statistic (2.372) have been interpreted by Fu [46] as evidence of the possible presence of background selection, thus supporting indications of demographic expansions. Similarly, the measure of Fu's Fs statistics revealed values significantly different from zero in all analyzed individuals, suggesting a past expansion in the population studied. It was reported that Fu's *FS* test is the most powerful tests for detecting population growth. The resulting negative

values for Tajima' D and positive values for Fu's *Fs*-tests indicate a complex demographic history of Ethiopian populations, which might corresponds to bottleneck and subsequent population growth. Moreover, study conducted by Elson, *et. al.,*[41] that assessed selection in the mitochondrial DNA (mtDNA) coding-region sequences support the present study. These authors provided evidence that selection has been a significant force during the evolution of the human mitochondrial genome and indicated evidence for both gene-specific and lineage-specific selection. Unfortunately, there are only small data on present populations, and more extensive studies are needed. At the present, based on the present results, a simple model of isolation by languages can also explain the population genetic structure of populations within region since population genetic differentiation (estimated as F_{ST}) does appear to be related to linguistic distances separating populations.

To further see the Ethiopian genetic affinities and their relationship to neighbor Sudan, PCA was performed. The Y chromosome PCA of the present Ethiopian and combined with few data from neighbor Sudan were performed based on F_{ST} genetic distances (Figure 8 A and B). 92.42% of the total variation was contributed by the first PCA. This PCA showed the general genetic affinities between populations. Broadly three groups of genetically close populations were observed. This includes Mao, Shinasha and Gumuz. It shows that the Berta and Nuer are not only genetically distinct from each other, but also from the other groups. The second PCA run was used to analyze the combined data of the present study and other available data from Ethiopians, Sudanese, Senegalese and Turks. The PCA showed general genetic similarities between some populations. Three main groups of genetically closely related populations were observed. The first groups are Beja and Amhara, the second Berta and Gumuz and the third Shilluk, Shinasha, Mao and Nuer* (Sudan group). This show that the Y chromosome lineage distributions seem to cluster geographically, and for some lineages, based on linguistic classification. The correlation between genetic distances with language is stronger than it is with geography classification. Additionally, correlation between genetic variation and geography/language is stronger for Y chromosome lineage data than for *mt*DNA lineages.

Also, the evolutionary relationships among Ethiopian and global *MT-COX2* sequences was seen. Most of the Ethiopian mtDNA types cluster together on the common lineage. In all cases, tree topologies were supported, basal branching pattern, with one deep branch lead generally to

Ethiopians, followed by less deep branches of Africans, and non-Africans with intercalation of few Ethiopian sequences. In a variety of phylogenetic studies aimed at constructing human mtDNA trees, this deepest clade has always lead to sub-Saharan African individuals, and with the complete absence of Ethiopian sequences in the analysis, it was impossible to draw conclusion about the root of the tree encompassing the most ancient and diverse individuals. In the present study, mtDNA tree, for the first time, has been reconstructed, and found to be consistent with the number of phylogenetic trees constructed by complete mtDNA sequences with the exception of the deepest clade constituent.

Ethics approval and consent to participate

This study involving human participants was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments. Ethical approval for the study titled "Y chromosome and MT-COX2 variation in some Afro-Asiatic and Nilo-Saharan language families groups in Ethiopia" was granted by the Salale University Institutional Research Ethics Review Committee (SIU-IRERC) under reference number SIU-IRERC-ET01/23, dated 27 January 2023. The research was approved for a one-year implementation period, subject to adherence to established ethical principles. All participants provided written informed consent prior to their involvement in the study.

Consent for publication

Not applicable

Declaration of Funding

The authors declare no competing interests. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interest:

No

Acknowledgments

The authors would like to thank Salale University, Ethiopia, for institutional support, and Addis Ababa University for providing laboratory facilities for this work.

Availability of data and materials:

The data supporting this study are provided within the manuscript and its supplementary information file. The sequences have been deposited in GenBank under accession numbers MK955250–MK955274/(https://www.ncbi.nlm.nih.gov/nuccore/MK955274.1/).(MT-COX2 sequences reported in this paper) . Additional software and database resources used to acquire or analyze the data are available at: DnaSP (https://www.ub.edu/dnasp/) and the Human Mitochondrial Genome Database (mtDB) via MitoMap (https://www.mitomap.org)

Author's contribution: Solomon Balami contributed to conceptualization, data analysis and interpretation, and writing of the original draft. Tura Safawo Jarso contributed to data analysis and interpretation, and reviewing and editing the manuscript. Endeshaw Bekele conceptualization and guiding of all the research activities, Muntaser Ibrahim Laboratory data analysis and facilitation

References

[1] L.L. Cavalli-Sforza, The history and geography of human genes., Princeton University Press., 1994.

[2] F. Cruciani, R. La Fratta, P. Santolamazza, D. Sellitto, R. Pascone, P. Moral, E. Watson, V. Guida, E.B. Colomb, B. Zaharova, J. Lavinha, G. Vona, R. Aman, F. Calì, N. Akar, M. Richards, A. Torroni, A. Novelletto, R. Scozzari, Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out of Africa, Am. J. Hum. Genet. 74 (2004) 1014–1022. doi:10.1086/386294.

[3] J.R. Luis, D.J. Rowold, M. Regueiro, B. Caeiro, C. Cinnioğlu, C. Roseman, P.A. Underhill, L.L. Cavalli-Sforza, R.J. Herrera, The Levant versus the Horn of Africa: Evidence for Bidirectional Corridors of Human Migrations, Am. J. Hum. Genet. 74 (2004) 532–544. doi:10.1086/382286.

[4] G. Passarino, O. Semino, L. Quintana-Murci, L. Excoffier, M. Hammer, A.S. Santachiara-Benerecetti, Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms, Am. J. Hum. Genet. 62 (1998) 420–434. doi:10.1086/301702.

[5] M.G. Thomas, T. Parfitt, D.A. Weiss, K. Skorecki, J.F. Wilson, M. Le Roux, N. Bradman, D.B. Goldstein, Y chromosomes traveling south: The cohen modal haplotype and the origins of the Lemba - The “black Jews of southern Africa,” Am. J. Hum. Genet. 66 (2000) 674–686. doi:10.1086/302749.

[6] T. Kivisild, M. Reidla, E. Metspalu, A. Rosa, A. Brehm, E. Pennarun, J. Parik, T. Geberhiwot, E. Usanga, R. Villems, Ethiopian mitochondrial DNA heritage: Tracking

gene flow across and around the gate of tears, Am. J. Hum. Genet. 75 (2004) 752–770. doi:10.1086/425161.

[7] C.A. Plaster, Variation in Y chromosome, mitochondrial DNA and labels of identity in Ethiopia., 2011.

[8] G. Lopez, S. and Hellenthal, Exploring the genetics and ancestry of peoples of Ethiopia. ., 2017.

[9] T. Sambrook, J., Fritsch, E.F. and Maniatis, Molecular cloning: A laboratory manual ., 2nd Editio, Cold Spring Harbor Laboratory, New York, 1990.

[10] M.F. Hammer, S. Horai, Y chromosomal dna variation and the peopling of Japan, Jpn. J. Hum. Genet. 41 (1996) 79.

[11] P.A. Underhill, L. Jin, A.A. Lin, S. Qasim Mehdi, T. Jenkins, D. Vollrath, R.W. Davis, L.L. Cavalli-Sforza, P.J. Oefner, Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography, Genome Res. 7 (1997) 996–1005. doi:10.1101/gr.7.10.996.

[12] J. Shen, P., Wang, F., Underhill, P. A., Franco, C., Yang, W. H., Roxas, A., Sung, R., Lin, A. A., Hyman, R. W., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L., Oefner, P. J., & Chu, Population genetic implications from DNA polymorphism in, in: Proc. Natl. Acad. Sci., 2000: pp. 4225–4230.

[13] H.Y. Hassan, P.A. Underhill, L.L. Cavalli-Sforza, M.E. Ibrahim, Y-chromosome variation among sudanese: Restricted gene flow, concordance with language, geography, and history, Am. J. Phys. Anthropol. 137 (2008) 316–323. doi:10.1002/ajpa.20876.

[14] Ensembl., Ensembl Homo sapiens, release 91. European Bioinformatics Institute., (2017). https://www.ensembl.org.

[15] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res. 18 (2008) 830–838. doi:10.1101/gr.7172008.

[16] Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups., Genome Res,. 12: (n.d.) 339–348.

[17] Ø. Hammer, D.A.T. Harper, P.D. Ryan, Past: Paleontological statistics software package for education and data analysis, Palaeontol. Electron. 4 (2001).

[18] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882. doi:10.1093/nar/25.24.4876.

[19] T.A. Hall, BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT., Nucleic Acids Symp. Ser. 41 (1999) 95–98.

[20] M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and the origin of modem humans, Nature. 408 (2000) 708–713. doi:10.1038/35047064.

[21] M. Tajima, F., & Nei, Estimation of evolutionary distance between nucleotide sequences, J. Mol. Evol. 10 (1993) 512–526. 329.

[22] B.B. Aggarwal, Y. Takada, S. Shishodia, A.M. Gutierrez, O. V. Oommen, H. Ichikawa, Y. Baba, A. Kumar, Nuclear transcription factor NF-kappa B: Role in biology and medicine, Indian J. Exp. Biol. 42 (2004) 341–353.

[23] DNA, P0. Software for comprehensive analysis of Data.olymorphism, DnaSP version 5.1, (2009).

[24] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): An integrated software package for population genetics data analysis, Evol. Bioinforma. 1 (2005) 47–50. doi:10.1177/117693430500100003.

[25] S. Tamura, K Nei, M., & Kumar, MEGA4: Molecular evolutionary genetics analysis (MEGA) software version., Dudley, J., 4.0., Mol. Biol. Evol. 24 (2007) 1596–1599.

[26] K. Tamura, M. Nei, Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees, Mol. Biol. Evol. 10 (1993) 512–526. doi:10.1093/oxfordjournals.molbev.a040023.

[27] M. Hammer, A nomenclature system for the tree of human Y-Chromosomal binary haplogroups, Genome Res. 12 (2002) 339–348. doi:10.1101/gr.217602.

[28] O. Semino, A.S. Santachiara-Benerecetti, F. Falaschi, L.L. Cavalli-Sforza, P.A. Underhill, Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny, Am. J. Hum. Genet. 70 (2002) 265–268. doi:10.1086/338306.

[29] J.J. Sanchez, C. Børsting, N. Morling, Typing of Y chromosome SNPs with multiplex PCR methods., Methods Mol. Biol. 297 (2005) 209–228. doi:10.1385/1-59259-867-6:209.

[30] M.F. Hammer, A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies, Mol. Biol. Evol. 11 (1994) 749–761. doi:10.1093/oxfordjournals.molbev.a040155.

[31] S. Agrawal, F. Khan, A. Pandey, M. Tripathi, R.J. Herrera, YAP, signature of an African-Middle Eastern migration into Northern India, Curr. Sci. 88 (2005) 1977–1980.

[32] V. Gomes, P. Sánchez-Diz, A. Amorim, Á. Carracedo, L. Gusmão, Digging deeper into East African human y chromosome lineages, Hum. Genet. 127 (2010) 603–613. doi:10.1007/s00439-010-0808-5.

[33] E.T. Wood, D.A. Stover, C. Ehret, G. Destro-Bisol, G. Spedini, H. McLeod, L. Louie, M. Bamshad, B.I. Strassmann, H. Soodyall, M.F. Hammer, Contrasting patterns of Y chromosome and mtDNA variation in Africa: Evidence for sex-biased demographic processes, Eur. J. Hum. Genet. 13 (2005) 867–876. doi:10.1038/sj.ejhg.5201408.

[34] S.A. Tishkoff, F.A. Reed, F.R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J.B. Hirbo, A.A. Awomoyi, J.M. Bodo, O. Doumbo, M. Ibrahim, A.T. Juma, M.J. Kotze, G. Lema, J.H. Moore, H. Mortensen, T.B. Nyambo, S.A. Omar, K. Powell, G.S. Pretorius, M.W. Smith, M.A. Thera, C. Wambebe, J.L. Weber, S.M. Williams, The genetic structure and history of Africans and African Americans, Science (80-.). 324 (2009) 1035–1044. doi:10.1126/science.1172257.

[35] P.A. Underhill, P. Shen, A.A. Lin, L. Jin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonné-Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J.R. Kidd, S.Q.

Mehdi, M.T. Seielstad, R.S. Wells, A. Piazza, R.W. Davis, M.W. Feldman, L.L. Cavalli-Sforza, P.J. Oefner, Y chromosome sequence variation and the history of human populations, Nat. Genet. 26 (2000) 358–361. doi:10.1038/81685.

[36] M.E. Hurles, B.C. Sykes, M.A. Jobling, P. Forster, The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages, Am. J. Hum. Genet. 76 (2005) 894–901. doi:10.1086/430051.

[37] E.I. Gebremeskel, M.E. Ibrahim, Y-chromosome E haplogroups: their distribution and implication to the origin of Afro-Asiatic languages and pastoralism, Eur. J. Hum. Genet. 22 (2014) 1387–1392. doi:10.1038/ejhg.2014.41.

[38] F. Cruciani, P. Santolamazza, P. Shen, V. Macaulay, P. Moral, A. Olckers, D. Modiano, S. Holmes, G. Destro-Bisol, V. Coia, D.C. Wallace, P.J. Oefner, A. Torroni, L.L. Cavalli-Sforza, R. Scozzari, P.A. Underhill, 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 (2002) 1197–1214. doi:10.1086/340257.

[39] F. Cruciani, R. La Fratta, B. Trombetta, P. Santolamazza, D. Sellitto, E.B. Colomb, J.M. Dugoujon, F. Crivellaro, T. Benincasa, R. Pascone, P. Moral, E. Watson, B. Melegh, G. Barbujani, S. Fuselli, G. Vona, B. Zagradisnik, G. Assum, R. Brdicka, A.I. Kozlov, G.D. Efremov, A. Coppa, A. Novelletto, R. Scozzari, 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 (2007) 1300–1311. doi:10.1093/molbev/msm049.

[40] L.L. Cavalli-Sforza, Genes, peoples, and languages, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 7719–7724. doi:10.1073/pnas.94.15.7719.

[41] J.L. Elson, D.M. Turnbull, N. Howell, Comparative Genomics and the Evolution of Human Mitochondrial DNA: Assessing the Effects of Selection, Am. J. Hum. Genet. 74 (2004) 229–238. doi:10.1086/381505.

[42] N. Elhassan, E.I. Gebremeskel, M.A. Elnour, D. Isabirye, J. Okello, A. Hussien, D. Kwiatksowski, J. Hirbo, S. Tishkoff, M.E. Ibrahim, The episode of genetic drift defining

the migration of humans out of africa is derived from a large east african population size, PLoS One. 9 (2014). doi:10.1371/journal.pone.0097674.

[43] M. Slatkin, R.R. Hudson, Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations, Genetics. 129 (1991) 555–562. doi:10.1093/genetics/129.2.555.

[44] A.R. Rogers, H. Harpending, Population growth makes waves in the distribution of pairwise genetic differences, Mol. Biol. Evol. 9 (1992) 552–569. doi:10.1093/oxfordjournals.molbev.a040727.

[45] M. Krings, A.E.H. Salem, K. Bauer, H. Geisert, A.K. Malek, L. Chaix, C. Simon, D. Welsby, A. Di Rienzo, G. Utermann, A. Sajantila, S. Pääbo, M. Stoneking, mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration?, Am. J. Hum. Genet. 64 (1999) 1166–1176. doi:10.1086/302314.

[46] Y.X. Fu, Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection, Genetics. 147 (1997) 915–925. doi:10.1093/genetics/147.2.915.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplemetry2Figure4YAP78nuerandberthadocx.docx
- supplemetry3Figure5and6YAPtypingwithM42M60andM89.docx
- supplementry1figure3YAPis450bpandYaAPis150bpnuerandberta.docx