

Dartmouth College Dartmouth Digital Commons

Open Dartmouth: Faculty Open Access Articles

11-25-2013

X-Linked MTMR8 Diversity and Evolutionary History of Sub-Saharan Populations

Damian Labuda
Université de Montréal

Vania Yotova
Université de Montréal


Jean-François Lefebvre
Université de Montréal

Claudia Moreau
Université de Montréal

Gerd Utermann
Medical University of Innsbruck

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>

 Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

Recommended Citation

Labuda, Damian; Yotova, Vania; Lefebvre, Jean-François; Moreau, Claudia; Utermann, Gerd; and Williams, Scott M., "X-Linked MTMR8 Diversity and Evolutionary History of Sub-Saharan Populations" (2013). *Open Dartmouth: Faculty Open Access Articles*. 3549.

<https://digitalcommons.dartmouth.edu/facoa/3549>

This Article is brought to you for free and open access by Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Faculty Open Access Articles by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Authors

Damian Labuda, Vania Yotova, Jean-François Lefebvre, Claudia Moreau, Gerd Utermann, and Scott M. Williams

X-Linked *MTMR8* Diversity and Evolutionary History of Sub-Saharan Populations

Damian Labuda^{1,2*}, Vania Yotova¹, Jean-François Lefebvre¹, Claudia Moreau¹, Gerd Utermann³, Scott M. Williams⁴

1 Centre de Recherche, CHU Sainte-Justine, Université de Montréal, Montréal, Québec, Canada, **2** Département de Pédiatrie, Université de Montréal, Montréal, Québec, Canada, **3** Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria, **4** Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, United States of America

Abstract

The genetic diversity within an 11 kb segment of the *MTMR8* gene in a sample of 111 sub-Saharan and 49 non-African X chromosomes was investigated to assess the early evolutionary history of sub-Saharan Africans and the out-of-Africa expansion. The analyses revealed a complex genetic structure of the Africans that contributed to the emergence of modern humans. We observed partitioning of two thirds of old lineages among southern, west/central and east African populations indicating ancient population stratification predating the out of Africa migration. Age estimates of these lineages, older than coalescence times of uniparentally inherited markers, raise the question whether contemporary humans originated from a single population or as an amalgamation of different populations separated by years of independent evolution, thus suggesting a greater antiquity of our species than generally assumed. While the oldest sub-Saharan lineages, ~500 thousand years, are found among Khoe-San from southern-Africa, a distinct haplotype found among Biaka is likely due to admixture from an even older population. An East African population that gave rise to non-Africans underwent a selective sweep affecting the subcentromeric region where *MTMR8* is located. This and similar sweeps in four other regions of the X chromosome, documented in the literature, effectively reduced genetic diversity of non-African chromosomes and therefore may have exacerbated the effect of the demographic bottleneck usually ascribed to the out of Africa migration. Our data is suggestive, however, that a bottleneck, occurred in Africa before range expansion.

Citation: Labuda D, Yotova V, Lefebvre J-F, Moreau C, Utermann G, et al. (2013) X-Linked *MTMR8* Diversity and Evolutionary History of Sub-Saharan Populations. PLoS ONE 8(11): e80710. doi:10.1371/journal.pone.0080710

Editor: Francesc Calafell, Universitat Pompeu Fabra, Spain

Received: June 14, 2013; **Accepted:** October 7, 2013; **Published:** November 25, 2013

Copyright: © 2013 Labuda et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding provided by the National Science and Engineering Research Council of Canada. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: damian.labuda@umontreal.ca

Introduction

In light of recent data, human evolutionary history looks much more complex than what geneticists postulated only a decade ago [1]. For example, in addition to new evidence of archaic admixture outside Africa [2–4], the time scale of human evolution needs to be extended. The generation span of humans and their ancestors appears longer than was previously estimated, and an older age of independent human lineages that collectively contributed to the modern genome was documented [5–10]. Nevertheless, the greatest genetic diversity among human populations is observed in sub-Saharan Africa, which leads to revisiting the old question [11,12]:

Is greater genetic diversity in Africa due only to older and larger ancestral African population sizes as compared to those outside of Africa, or does it reflect the impact primarily of the out of Africa population bottleneck(s) in reducing the genetic diversity outside Africa? If at the time of out-of-Africa migration Africans and migrant populations did not differ in their genetic structure, but African populations stayed larger than non-Africans, greater African diversity would be expected to result from an accumulation of new low frequency local variants. In contrast, a restricted sampling of alleles due to a stringent out-of-Africa bottleneck [13]

would cause a relative paucity of the genetic diversity in non-Africans [14]. Both above scenarios turned out to be too simplistic. New findings have provided support for the idea that genetic diversity of non-Africans was additionally enriched by admixture with Neandertals [2,4] and Denisovans [3]. In Africa, low frequency derived alleles in the *dys44* segment are spread on different haplotypes, which implies substantial number of recombinations and gene conversions, and thus long evolutionary time since the corresponding mutations have occurred [10]. Such alleles are necessarily ancient and their limited occurrence suggested archaic admixture within Africa itself, supported by additional evidence [9,15–18]. Although extensive gene flow shaped the diversity of sub-Saharan Africans over various periods [19–22], traces of the ancestral subdivisions can still be recognized in the genetic record [6,20,23–26]; the homogenizing effect of gene flow that in general will partially conceal the record of the ancestral population structure, is expected to be less obvious in areas of low recombination and strong linkage disequilibrium [27]. Studying loci with such characteristics is therefore of great significance in terms of unravelling human population histories. In depth analyses of single autosomal and X-linked loci complement studies of uniparentally inherited mitogenomes and Y-chromosomes with their well resolved genealogies. However, the time depth of

autosomal or X-linked genealogies is roughly four and three-times deeper, respectively, than that of the genealogies of the uniparentally inherited loci. The “whole genome studies” amalgamate information from individual loci to provide a synthetic overall view of the human genome history. In contrast, single locus oriented studies, can reveal particular and sometimes unusually important historical events, such as that of archaic admixture within Africa, and/or that of Neandertal admixture outside Africa, as documented in our earlier studies focusing on specific DNA segments [4,6,10,28,29].

Here we studied an 11 kb DNA segment with low-recombination frequencies [30] located in the Xq11.1 genomic region. This segment is located within the myotubularin related protein 8 gene (*MTMR8*), in the centromeric portion of the X chromosome, earlier found to have reduced sequence diversity in non-African populations [31–33]. Our study showed that *MTMR8* diversity in sub-Saharan Africans can reveal much about the evolutionary history of these populations. Specifically, ancient population structure in the geographic partitioning reflects separate evolutionary histories of the southern, western/central and the eastern African clades. Our results are consistent with the enrichment of the African diversity through fragmentation of its populations over long periods of their history as well as by admixture among archaic populations. They also support the hypothesis of an early bottleneck in out of Africa migrants, prior to their range expansion outside of Africa.

Materials and Methods

DNA samples

A total of 160 X chromosomes were analyzed, 73 drawn from DNA samples of the Human Genome Diversity Panel of Centre d'Étude du Polymorphisme Humain, CEPH-HGDP [34]: Bantu speakers NE, Kenya (n = 10); Biaka Pygmy, Central African Republic (n = 13); San, Namibia (n = 7); Mandenka, Senegal (n = 13); M'Buti Pygmy, Democratic Republic of Congo (n = 13); Yoruba, Nigeria (n = 17). The remaining samples were from earlier studies (see [4,6,35,36]), and included Gabonese (n = 12); Khoen-San, collected in Smithdrift, South Africa (n = 11); Ethiopians (n = 15); Europeans (n = 13; 2 Bulgarians, 3 French, 2 French-Canadians, 2 Germans, 4 Italians), South East Asians (n = 14; 6 Chinese, 4 Vietnamese, 2 Filipinos, 2 Indians); Near East/North Africa (n = 12; 2 Iran, 4 Lebanon, 4 Egypt, 2 Morocco); Native Americans (n = 10; 6 Chipewyans, 4 Maya) (see table S1). They were obtained from the collaborating academic institutions or were collected by us, following a protocol approved by the Institutional Ethics Committee at the Centre Hospitalier Universitaire Sainte-Justine. Common chimpanzee DNA was extracted from a peripheral blood sample provided by Clément Lanthier from Granby Zoo (Québec), taken during the course of routine veterinary care. All Sub-Saharan African chromosomes (n = 111; 107 male and 2 female samples) and 7 non-African chromosomes (3 females and 1 male sample) were analyzed directly by resequencing; the remaining 42 non-African chromosomes were screened for additional polymorphisms by heteroduplex analysis followed by direct sequencing of the fragments of interest. Note that samples from Egypt and Morocco are counted among non-Africans. In addition, we used the genomic diversity data available online for HGDP and HapMap3 DNA populations [34,37–39].

Detection of sequence polymorphism

PCR-primers were designed to amplify 21 overlapping fragments covering an 11024 bp-long segment of *MTMR8* (NT011669), starting in intron 3 and ending in intron 5 (table

S2). Amplifications were carried out in 20 μ l using 5 ng DNA, 0.75 U of Platinum[®] Taq DNA polymerase (Invitrogen Canada Inc.) in the Platinum[®] PCR buffer containing 1.5 mM MgCl₂, 0.2 mM each dNTP, and 2 μ M each primer. Reaction started with 5 min incubation at 94°C followed by 35 cycles of 30 s at 94°C, 30 s at 55°C (except for fragment 9 at 61°C) and 30 s at 72°C, to end in 10 min incubation at 72°C. Both strands of the products were sequenced on ABI 3730 DNA sequencer. In addition to 118 chromosomes analyzed in this way we screened 21 female non-African samples (Asia, Europe, and Native America - 5 from each group, and 6 from Near East/North Africa) for the presence of polymorphisms by DNA temperature-melting heteroduplex-detecting analysis [40]. It was carried out in HR-1 High Resolution Melter, Idaho Technology, using LCGreen I Melting Dye in standard Roche LightCycler[®] glass capillary tubes. We used previously detected polymorphisms to create heterozygous samples serving as positive controls for heteroduplex detection in the analyzed DNA fragments. Two heteroduplexes were detected in non-African samples, adding two new haplotypes. By sequencing they were found to be due to a C>T transition in position 1890874 (mutation 32) and to a T>C polymorphism at position 1887715 (mutation 19), already known from resequencing of the African samples (table 1).

Haplotypes and network

The ancestral alleles of the *MTMR8* polymorphisms were inferred by comparison with the primate outgroup sequences of chimpanzee, orangutan and macaque. The human allele identical by state with a chimpanzee or at least two other outgroup sequences was considered ancestral. All distinct haplotypes could have been derived from the data without any ambiguity because all but two African samples represented hemizygous males and all females were either homozygous or when heterozygous it was only in a single haplotype position (table 1). The haplotype tree reduces to a “perfect phylogeny” tree after excluding derived allele A at the site 1889557 (mutation 25) in haplotype 5 (seen in two copies in Khoen-San), and derived allele A at the site 1890711 (mutation 31) in haplotype 2 (a single copy in M'Buti). These two particular polymorphisms are due to G>A transitions within hypermutable CpG-dinucleotides [41] thereby providing limited phylogenetic information. We assumed that haplotypes 2 and 5 were due to independent recurrent transitions (coded as mutation 25* and 31*, respectively) rather than resulting from an unlikely sequence of recombination events in this genomic region of particularly low recombination activity [30].

Statistical analysis

Coalescence analysis was carried out according to the method of Griffiths and Tavaré [42] using *genetree*, version 9.0, on the full data set (n = 160), on all Sub-Saharan Africans together (n = 111), their eight subpopulations, and Sub-Saharan Africans +1 (n = 112) by addition of haplotype 23, found in a single copy in Lebanon but possibly of African origin given its position in the network. Maximum likelihood estimates of Θ (i.e. Θ_{ML} ; see list of abbreviations in Material S1), the time to the most recent common ancestor (TMRCA) and the age of mutations were obtained conditional on the haplotype tree, assuming an infinite-sites model, random mating and constant population size or exponential growth. The number of iterations per run was sufficiently large for the results to remain constant over repeated runs differing only in the random seed number. Estimates of Θ_{ML} made with a model of exponential growth concurrently yielded an estimate of the growth rate β . Particularly, in this model the population size exponentially declines backward in time at rate β

Table 1. *MTMR8* segment haplotypes.

SNPs ID	Populations																			Counts																					
	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study	this study																						
Human Feb. 2009 (GRCh37/hg19) assembly	6355561	6355562	6356568	63565691	63565818	63566179	6356648	63566740	63566864	63567014	63567951	63568366	63568754	63568859	63569203	63569307	63569355	63569593	63569727	63569969	63569979	63570024	63571405	63571438	63571569	63571733	63571753	63571974	63572010	63572288	63572723	63572886	63573344	63573673	63573863	63575039	63575448	Total			
network mut. positions	1-2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	25*	26	27	28	29	30	31	31*	32	33	34	35	36	37			
Ancestral	G	A	C	C	G	G	C	A	C	A	C	A	T	C	A	A	T	A	A	A	G	G	A	G	G	G	G	C	A	A	G	G	C	G	C	A	T				
Neandertal	A	N		
Denisova	
H1	A	G	A	.	A	T	38	
H2	A	G	A	.	A	T	1	
H3	A	G	A	.	A	T	2	
H4	A	G	A	.	A	T	3	
H5	A	G	A	.	A	T	2	
H6	A	G	A	.	A	T	1	
H7	A	G	A	.	A	T	1	
H8	A	G	A	.	A	T	5	
H9	A	G	A	.	A	T	20	
H10	A	G	A	.	A	T	6	
H11	A	G	A	.	A	T	1	
H23	A	G	A	.	A	T	1	
H24	A	G	A	.	A	T	1	
H12	A	G	A	.	A	T	46	
H13	A	G	A	.	A	T	1	
H14	A	G	A	.	A	T	1	
H15	A	G	A	.	A	T	1	
H16	A	G	A	T	.	A	T	T	2		
H17	A	T	10		
H18	A	T	1		
H19	A	T	2		

Table 1. Cont.

SNPs ID	Counts	Populations
Human Feb. 2009 (GRCh37/hg19) assembly	Total	
	Non Africans	
	Yoruba	
	Mbuti	
	Mandenka	
	Khoe-San	
	Gabon	
	Eth	
	Bia	
	Bantu NE	
6355561 this study	63575448	G
6355562 this study	63575039	G
6355568 this study	63573863	G
63565691 this study	63573673	G
63565684 this study	63573344	G
63566740 this study	63572886	G
63566864 this study	63572723	G
63567014 this study	63572288	G
63567951 this study	63572010	G
6356859 this study	63571974	G
6356864 this study	63571753	A
63569203 this study	63571733	A
63569277 this study	63571569	A
63569355 this study	63571438	A
63569370 this study	63571405	A
63569393 this study	63570024	A
63569699 this study	63570024	A
63569727 this study	63569979	A
63569953 this study	63569969	A
63569979 this study	63569727	A
6357014 this study	63569593	G
6356864 this study	63569355	G
63566740 this study	63569307	G
63566864 this study	63569203	G
6356859 this study	63568859	G
63568754 this study	63568754	G
63568366 this study	63568366	G
63567951 this study	63567951	G
63567014 this study	63567014	G
63566864 this study	63566864	T
63566740 this study	63566740	T
635664109 this study	63566448	T
63566179 this study	63566179	T
63565818 this study	63565818	T
63565691 this study	63565691	T
63565568 this study	63565568	T
63565562 this study	63565562	T
63565561 this study	63565561	T
n		10 13 15 12 18 13 13 17 49 160

The haplotype spans 11 Kb of the *MTMR8* gene, starting in intron 3 and ending in intron 5; the location of its polymorphic sites within the hg19 genome reference sequence are shown in the third line. New alleles appear on the background of ancestral (chimpanzee) alleles, which are also shared with Neanderthal and Denisova sequences, except for the polymorphic site 6 (highlighted in grey) where the derived allele is the same as that found in the Neanderthal genome. The polymorphic sites 25 and 31, involving CpG-dinucleotides, are assumed to have mutated twice, indicated by asterisks, to create separate haplotypes 5 and 2, respectively (both found among Khoe-San). doi:10.1371/journal.pone.0080710.t001

from a current size $N_{(t)}$ (or $1.5 N_{(t)}$ chromosomes), such that the size of the population at time t is $N_{(t)} = N_{(0)} e^{-\beta t}$ (note that $t = g/1.5N_{(0)}$, where g is the number of generations ago). Using *genetree* we jointly estimated Θ_{ML} and the growth parameter β . First, we explored different values of β at fixed Θ_{ML} (the starting values were those obtained under a constant population size scenario). Subsequently, for the selected β , we explored the likelihood density of Θ , to finally obtain the maximum likelihood estimates of both parameters after few rounds of such simulations. The ρ -statistics [43] that evaluates the average number of mutations till coalescence was calculated using Network 4.5.0.2 software [44]. This statistic [45,46] is equivalent to Thomson approach [47,48]. The substitution rate in the 11024 bp *MTMR8* segment was estimated at 1.49×10^{-4} per generation (or 1.35×10^{-8} per bp per generation) from the human-chimpanzee divergence of 0.0081 (± 0.0009) per bp, assuming a separation time of 7.5 million years (My) and an average generation time along both lineages of 25 years [7,49]. The corresponding autosomal rate can be evaluated by multiplying by $3(\alpha+1)/2(\alpha+2)$, where α is the ratio of male to female mutation rate [50,51].

ARLEQUIN software, v. 3.1 [52] was employed to compute different population statistics (haplotype count k and diversity G , count of segregating sites S and nucleotide diversity as well as different estimates of the scaled mutation parameter Θ , global and pairwise F_{ST}) and to carry out neutrality tests such as Tajima’s D [53] and Fu’s F_S [54], as well as tests according to Ewens and Watterson [55,56], Slatkin [57], and Chakraborty’s population amalgamation test [58]. For detailed description of the parameters and tests please refer to the Arlequin manual [52]. We also used DnaSP software v.5 [59] to estimate Θ_H and carry out the Fay & Wu H test [60].

Coalescence simulations

Simulations were performed using the msHot software [61], a modification of the ms program [62]. They were used to evaluate the effect of demography on the estimates of Θ_{ML} [42], Θ_S [63] and Θ_H [53] under a simple version of the standard neutral model at constant population size, at population growth, demographic bottleneck and at population subdivision.

Results

Table 1 presents sequence diversity of the *MTMR8* segment in a worldwide sample of 160 X chromosomes. Out of 24 haplotypes, 21 are observed in sub-Saharan Africa and only 4 are found outside of Africa (table 2). Furthermore, three of the non-African haplotypes are singletons. This dearth of diversity in non-Africans is also reflected in their very low nucleotide diversity ($\pi = 0.011 \times 10^{-3}$) and all other summary statistics (table 2 and table S3). In contrast, the nucleotide diversity of sub-Saharan Africans ($\pi = 0.46 \times 10^{-3}$) falls within the range observed at other X-linked segments [11,28,64,65]. Neutrality tests [53–56,60] are consistent with neutrality of *MTMR8* in sub-Saharan Africa (Table S3). Therefore, this locus is well suited to infer population history of sub-Saharan Africans. In non-African populations the situation is opposite, tests indicate a non-neutral evolution (Table S3) and the virtual absence of common sequence polymorphisms (table 1 and [32,33]) what renders *MTMR8* uninformative for population history inferences. Therefore, there was no reason to extend the sample size of non-Africans and we have focused our analysis on sub-Saharan populations.

The *MTMR8* haplotypes (table 1), considering haplotype 22 apart, form a simple network with the oldest branches separating Southern, Western/Central and Eastern African lineages (fig. 1).

Table 2. Sub-Saharan African and non-African Diversity and Neutrality Tests (all abbreviations are listed in Material S1).

	Sub Saharan Africa	Non-Africans	Total
Summary Statistics			
<i>n</i>	111	49	160
<i>k</i>	21	4	24
<i>S</i>	34	3	36
Θ_{π}	5.04	0.12	5.82
Θ_H	6.13	1.96	7.50
Θ_S	6.45	0.67	6.37
$\text{Hom} = (1-G)_{\text{obs}}$	0.16	0.88	0.19
Ewens-Watterson			
$\text{Hom} = (1-G)_{\text{exp}}$	0.12	0.52	0.12
Watterson's <i>p</i>	0.914	1.000	0.964
Slatkin's <i>p</i>	0.936	1.000	0.996
Chakraborty's			
<i>k</i> _{exp}	16.29	1.52	19.99
<i>P</i> (<i>k</i> or more haplotypes)	0.025	0.008	0.001
Tajima's			
<i>D</i>	-0.66	-1.70	-0.26
<i>p</i>	0.264	0.012	0.415
Fu's			
<i>k</i> _{exp}	16.29	1.52	19.99
<i>F</i> _S	-2.15	-4.26	-1.58
<i>p</i>	0.292	0.000	0.381
Fay & Wu's <i>H</i>			
<i>H</i>	-2.24	-1.84*	-3.08
Normalized <i>H</i>	-0.492	-2.51	-0.680
<i>p</i>	0.174	0.005	0.144

*the *H* statistics of **-9.44** ($p < 0.001$) is obtained for a combined population of non-Africans and Ethiopians ($\Theta_{\pi} = 1.24$; $\Theta_H = 10.6$. doi:10.1371/journal.pone.0080710.t002

The TMRCA was obtained by *genetree* [42] under a modest demographic growth in Africa ($\beta = 0.8$) with $\Theta_{ML} = 8.7$ ($N_{(0)ML} = 19,500$) (see Materials and Methods). Non-African samples were excluded from this analysis except for a single Lebanese chromosome carrying haplotype H23 because of its proximity to Africa and likely sub-Saharan origin (figs. 1 and 2). TMRCA was estimated at $29\,800 \pm 7\,000$ generations corresponding to 745 ± 175 thousand years (Ky) (fig. 2 and table 3). It falls within the time range of sequence divergence of Neandertals, Denisovans and modern humans [7,66] consistent with the derived allele A at the site 1884167 (mutation 6) being shared with Neandertals [2]. Interestingly, Denisova genome still carry the ancestral G, but the derived A shows up again in the newest high coverage Altai Neandertal sequence (<http://www.eva.mpg.de/neandertal/index.html>). In Biaka Pygmies from the Central African Republic we find a structurally distinct, ancient haplotype H22, whose lineage originates directly from the root of the tree (fig. 1). Given its characteristics we propose it as a plausible candidate to represent an archaic lineage within Sub-Saharan Africa [9,10] (see Discussion). The other very old haplotypes H19-H21 and H17 are principally Khoe-San, from southern Africa,

representing 61% of their chromosomes. Their age can be estimated from the age of mutations marking the corresponding lineages (fig. 2 and table 3; note, however, that when a single branch carries two or more mutations, they are placed in arbitrary order; any of these mutations could be the oldest or the youngest since we cannot determine their relative arrival times, and by convention only the top mutation is considered to be the oldest). The H19-H21 and H17 lineages are separated from the root by mutation 6 timed by *genetree* at 658 ± 163 Ky. They are separated from all other haplotypes that share mutations 1–2 and 3, timed at 503 ± 91 Ky and 464 ± 86 Ky ago, respectively (fig. 2, table 3; note that the order of mutations that occur on the same branch is arbitrary). In turn, more than 75% of Central and Western African chromosomes carry structurally related haplotypes H1 through H11, with the two most frequent sub-Saharan haplotypes, H1 at 34% and H9 at 18%. They occur on two distinct branches H1 and H9, which are split between mutation 3 above and mutations 7, 14 and 21 below, indicating a time of divergence between 300 and 450 Ky ago (table 3). The third separate Eastern clade marked by the haplotype H12 is the most frequent worldwide. H12 and its derived haplotypes account for 73% of the Ethiopian chromosomes, one Kenyan (North East Bantu), one chromosome from Southern Africa (Khoe-San) and all non-African chromosomes. The age of the oldest mutation on this branch (arbitrarily assigned as mutation 7) is also estimated at about 350 Ky (table 3), suggesting divergence time similar to H1 and H9 lineages. The star-like form of H1 and H12 clades points to expansion of the populations carrying these haplotypes, well after the arrival of mutations at sites 25 and 19, i.e. after 151 ± 40 and 85 ± 31 Ky ago, respectively, according to *genetree* age estimates (table 3). Using ρ -statistics [43,45], equivalent to the approach described by Thomson et al. [47,48], we obtained values in general similar to those obtained by *genetree*. The exceptions were mutations 7 and 25 (table 3), “defining” branches leading to haplotypes H1 and H12. This can be explained by the fact that *genetree* takes into account full data [42], whereas ρ -statistics only considers the tree information from branches below the mutation in question [45,62]. Therefore, ρ -statistics estimates may be influenced by local “branch-specific” effects, such as demographic expansions (around H1 and H12) and large differences in branch lengths separating lineages (e.g. ages of mutations 25 and 7). In turn, the ρ -statistics estimate of the age of the mutation 7 is much older than that obtained by *genetree*, reflecting a relative excess of mutations observed along the H12 branch.

The geographic partition of distinct *MTMR8* lineages between southern, western/central and eastern African populations (fig. 1), with two thirds of the corresponding haplotypes remaining population and region specific, is reminiscent of an ancient stratification of sub-Saharan populations, captured in our data by the population amalgamation test [58] ($p = 0.025$; table 2). One third of these haplotypes are regionally shared, reflecting within Africa gene flow at different historical periods [19–22,67,68]. Non-standard population demography (i.e. other than non-subdivided constant size population), such as a demographic bottleneck, population subdivision with limited migration and/or population growth, are known to differentially affect estimates of the population mutation parameter Θ . Three distinct estimates of Θ differ in the *MTMR8* data from sub-Saharan Africans: $\Theta_{\pi} = 5.03$, from the average number of pairwise differences, $\Theta_S = 6.45$, from the number of segregating sites (table 2; table S3) and $\Theta_{ML} = 7.19$, a maximum likelihood estimate assuming constant population size model. By computer simulations we explored the effect of different demographic scenarios on the relative values of these estimates. Population bottlenecks reduced all three estimates of mutation rate

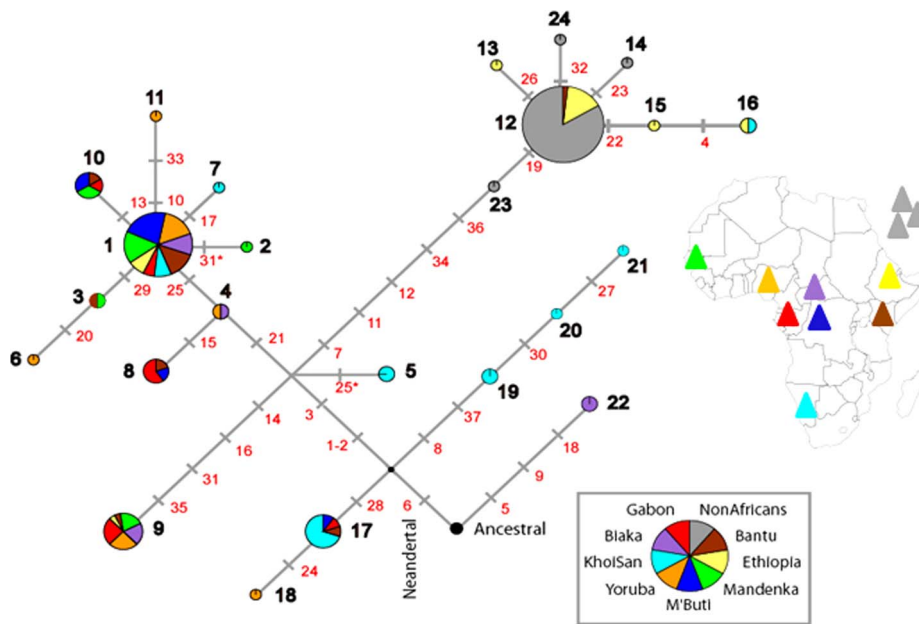


Figure 1. Network of *MTMR8* haplotypes. Haplotype frequencies are proportional to the surface of the circle (or to its single colored segment within a population group). Numbering of mutations and haplotypes is the same as in table 1. Asterisks indicate two mutations in the CpG-sites 25 and 31 that presumably represent independent substitutions leading to separate haplotypes 5 and 2, respectively (both found among Khoe-San). When a series of mutation occurs on a single branch their order of appearance is arbitrary as we cannot know which one was first or last based on the presented data.

doi:10.1371/journal.pone.0080710.g001

parameters similarly with respect to the input values. When populations are subdivided and the gene flow becomes increasingly restricted, down to the migration rate $4Nm = 0.04$, it is the Θ_{II} that increases the most followed by Θ_S , both characterized by a large variance between individual simulations, whereas Θ_{ML} increases only slightly. This is in contrast to what we observe here, i.e. $\Theta_{II} < \Theta_S < \Theta_{ML}$. However, the differential effect of population subdivision on the three estimates of Θ disappears, even if the populations were previously kept subdivided, after 500 generations of panmixia; in this scenario, all estimates tend to converge to the input value. Interestingly, a five-fold increase in population size 1500 generations ago has no marked effect on Θ_{II} , but it doubles Θ_S and triples Θ_{ML} . In a simulation experiment combining population subdivision with moderate gene flow ($4Nm = 2$) and population growth from 7 000 to 20 000 over the last 15 000 generations and assuming (input) Θ -value of 7.5, we reproduced the three observed Θ estimates: $\Theta_{II} = 5.39 < \Theta_S = 6.33 < \Theta_{ML} = 7.20$ (i.e. Θ_{ML} was estimated assuming constant population size). While the simulations do not provide true proof, they do show that both population subdivision and subsequent population growth can shape *MTMR8* diversity in ways consistent with the data, thereby offering a plausible explanation. A moderate increase in population size in Africa is suggested by the shape of the western/central and eastern branches of the network (fig. 1), by the coalescent analysis (fig. 2) and is also consistent with a slightly negative Tajima's *D* (table 2).

Discussion

Oldest tree branches belong to Khoe-San

Excluding H22, the oldest lineages H19–H21 and H17 belong to Khoe-San from South-Africa and Namibia, consistent with Khoe-San representing the oldest of all extant populations [68–73]. Our results are in line with earlier evidence from uniparentally

transmitted markers [24,74–77] and from microsatellite data [20]. Yet, while mtDNA and Y-chromosome studies estimate the oldest Khoe-San lineages at 140 Ky ago [24,74,78], our analysis finds the *MTMR8* Khoe-San lineages much older, originating before 300 Ky, possibly around 500 Ky ago and presumably even earlier. This is consistent with observations from a simple inspection of the data: the network structure, branch lengths as well as contrasting age estimates of the mutation 7 (fig. 2 and table 3). The estimates of Khoe-San population divergence, at around 150 Ky ago [71] (calibrated based on human chimpanzee separation at 7.6 My ago), are indeed expected to be much younger than the sequence divergence of the contributing lineages. On the other hand, the same authors approximated African-Eurasian divergence at 55–62 Ky ago, i.e. at almost half the 100 – 120 Ky ago estimated by Li and Durbin [72]. The latter dates are consistent with the archeological evidence of the presence of modern humans in the Near East at around 100 Ky ago [79–81] and raise the possibility that Khoe-San divergence estimated to have occurred ~ 2.5 times earlier [71] could thus be much older than 150 Ky, closer to the age of their lineages reported here (see however [67,82]).

Population bottleneck and sub-Saharan population structure preceding out-of-Africa migration

The intriguing feature of the *MTMR8* tree is the presence of the long lineage of haplotype H12 shared by most Ethiopians, one Kenyan, one Khoe-San and all non-Africans (figs. 1 and 2). Such patterns have been seen before. In the tree derived from *PDHAI* there is a long branch separating the bulk of sub-Saharan-African haplotypes from non-Africans and a small subset of Africans [65]. As in the case of *MTMR8*, this suggests that non-Africans emerged from a subset of Africans living in relative isolation for a substantial length of time consistent with the idea that the out-of-Africa bottleneck started in Africa before the exodus [6,10]. Note that what we observe is not a local *MTMR8* effect. However, the effect

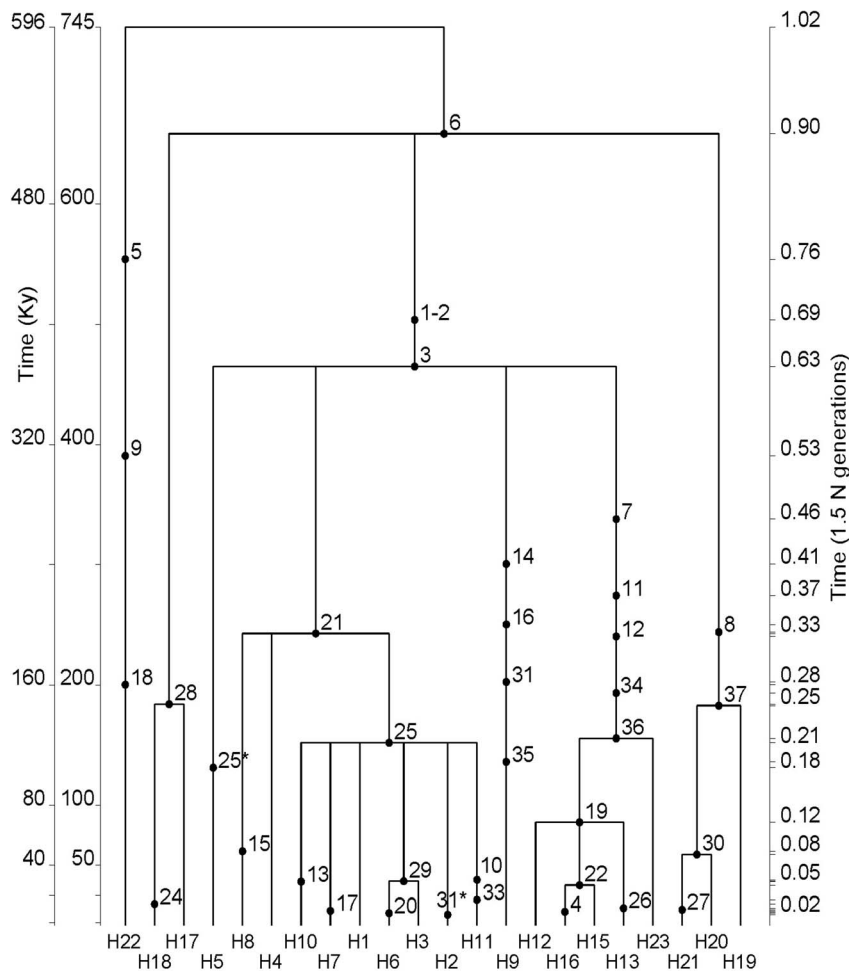


Figure 2. Coalescent analysis of the MTMR8 tree in Sub-Saharan Africa. The time scale in thousands of years is calculated using 7.5 My (internal left scale) or 6 My (external left scale) of sequence divergence between human and chimpanzee lineages. Numbering of mutations and haplotypes is the same as in fig. 1 and table 1. Note that a Levantine chromosome carrying haplotype 23 was included in this analysis. doi:10.1371/journal.pone.0080710.g002

Table 3. Genetree and ρ -statistics time estimates of mutations marking MTMR8 segment history (figs. 1 and 2) based on 111 sub-Saharan African chromosomes and a Lebanese haplotype 23 chromosome.

Method	Time estimates in Ky (\pm S.D.)	
	<i>genetree</i>	ρ -statistics
TMCA	745 \pm 180	931 \pm 290
mut 6	656 \pm 145	784 \pm 249
mut 3	464 \pm 85	532 \pm 163
mut 7	336 \pm 82	899 \pm 373
mut 14	299 \pm 85	503 \pm 291
mut 21	243 \pm 64	200 \pm 147
mut 25	151 \pm 40	47 \pm 24
mut 36	155 \pm 55	228 \pm 162
mut 19	85 \pm 31	78 \pm 48

doi:10.1371/journal.pone.0080710.t003

of this bottleneck may be enhanced in this locus due to a selection sweep that seems to have affected 5.4 Mb of the subcentromeric region, including *MTMR8* and 13 other genes [32,33]. Indeed, our results of the neutrality tests (table 2) are significant in non-Africans and thus consistent with a selection sweep hypothesis. The outcome of the H test of Fay & Wu [60] is of special interest, with $H = -1.84$ ($p = 0.005$) for non-Africans, which increased to $H = -9.44$ ($p < 0.001$) when non-Africans and Ethiopians were analyzed (fig. 2 and table 3). Outside sub-Saharan Africa we find 46 H12 chromosomes, two differing by single mutations in Europe (H14 and H24) and one chromosome from Lebanon, H23. In Africans the H12 derived lineages are also characterized by low diversity, nine H12 and four H12-derived chromosomes, but the haplotype diversity of H12 branch haplotypes in Africa is significantly higher than all other continental samples taken together ($p = 0.03$ Fisher exact test). And this, in spite of the fact that our Ethiopian sample partly represents non-African chromosomes due to effects of the Eurasian gene-flow on Ethiopian diversity, thus diluting its African component [20,35,83,84]. In Africans we find a haplotype H16 derived from H12 by two mutations. Specifically, H16 is found both in Ethiopians and Khoe-San, two populations known to share deep paternal lineages [67,68,85]. Our resulting hypothesis is that a population

bottleneck preceded the out of Africa migration and subsequent range expansion, because, at that time, sub-Saharan Africans were stratified and different populations evolved in isolation for an extended period of time. This is plausible, especially through part of the Middle and Late Pleistocene [86], when climatic conditions were conducive or even forced geographic isolation [87–89]. Interestingly, all non-African descendants, for both mtDNA and Y-chromosome uniparentally transmitted lines, each share one and the same close African ancestor [90]. This alone provides evidence of a period of isolation in a small single founding group, during which all other founding lineages were lost by genetic drift [82]. If so, the geographic partitioning of genetic diversity we observe should be essentially due to the partitioning of old variants, rather than to the accumulation of novel, population-specific variation, which would reflect recent population history. In *MTMR8*, about two thirds of the haplotypes from each of the three geographic poles (south, west/central or east) represent distinct old lineages (table 1, fig. 1). A similar situation was observed at other X-linked loci and uniparentally inherited markers [6,10,23,24,75,82], in spite of recent gene flow affecting several loci to a different extent. To confirm that our observations are not due to a stochastic fluke, limited to a short 11 Kb DNA fragment, we also analysed haplotypes of a 380 Kb segment spanning *MTMR8* locus using the data of HapMap3 populations. The resulting haplotypes' network (fig. S1) is consistent with our earlier findings. It shows a tripartite split between two separate African clusters (only two, in the absence of Central and South-African populations in HapMap3 collection) and one non-African as in the figure 1. Li and Durbin [72] observed an increase in the effective population size in Africa between 200 and 60 Ky ago, which they interpreted as the effect of population fragmentation with reduced migration. Our simulation experiments, while unable to prove this scenario of a complex African population structure before range expansion outside Africa, support its plausibility.

Nevertheless, it is important to note that historical inferences from studies of the X-linked loci may differ from these based on autosomal record. The effective population size of the X-chromosomes is smaller than that of the autosomes and X-chromosomes spend only one third of the time in males and two third in females, where they recombine. Male and female demography may differ, due to different migration patterns of males and females, to patrilocality or matrilocality, to polygamy, and other phenomena changing the female-to-male ratio. As a result, X-chromosomes diversity, as compared to autosomal loci, cannot be simply accounted for by a three-to-four ratio between X-chromosomes and autosomes population sizes [51,91,92]. Interestingly, during the out-of-Africa bottleneck, the X-chromosome diversity appears to have been disproportionately reduced relative to the rest of the nuclear genome [93]. Whether this was due to natural selection or to demographic effects as revealed here, it shows how important studies of the X-chromosome diversity (often excluded from genome-wide diversity analyses) are to understand human evolutionary history.

Archaic admixture

The Biaka haplotype H22 does not share any derived alleles with the remaining haplotypes and is absent outside Africa. Only this haplotype has the alternative allele at our mutation 6 (fig. 1 and table 1) thus driving the *MTMR8* TMRCA back in time prior to the divergence of human and Neandertal lineages. Similar, structurally distinct haplotypes, representing African-only lineages were observed by others. For example, haplotype P in the *CMAH* locus on chromosome 6p21.32 is seen in two copies in Biaka Pygmies out of 132 analyzed chromosomes. It carries 16 derived

alleles that are not shared with all other haplotypes carrying a different subset of 39 derived sites. Out of the 56 *CMAH* polymorphisms only one derived allele is shared between P and another haplotype C3 found in the same Biaka population [16]. In the left portion of the *CD209* locus with its 57 segregating sites, 17 derived alleles are exclusive to three rare African haplotypes while the remaining 40 alleles occur on other chromosomes. These three distinct haplotypes are found in San from Namibia (2 copies), in Bantu speakers from Gabon and from South Africa (5), in Yoruba's from Nigeria (3) and Mandenka from Senegal (2) [94]. In turn, in a DNA segment from Xp21.1 haplotype A carries again 6 derived alleles that are exclusive to this haplotype found in only two copies in Mbuti Pygmies and absent on other haplotypes found on the majority of chromosomes [15]. In *CMAH*, *CDC209* and Xp21.1 segment gene trees these rare ancient haplotypes have driven the corresponding TMRCA's above 2 My ago, artificially elongating tree branches due to the presence of derived alleles not shared with all remaining worldwide haplotypes. Other examples of similarly distinct African haplotypes were also found in the *4qMB179*, *13qMG107* and *18qMB60* regions on chromosomes 4q, 13q and 18q, respectively [9]. In the light of our results and examples above, our earlier interpretation of the greater genetic diversity in Africa being partly due to the introgression from an archaic population [6,10,36] is gaining more support recently [15,17,18,95]. There is the possibility that these haplotypes are not archaic and have been kept intact for such great amounts of time simply due to the stochastic process. Interestingly, based on HGDP polymorphisms [38] H22 extends over 800 Kb before "joining" new alleles common to other sub-Saharan samples (HT16 in Table S5). It is noteworthy that H19 haplotype (HT17 in Table S5) does not differ from Neandertal and Denisova haplotype over the whole length of the extended haplotype (1971 Kb), confirming its greater antiquity and also consistent with its age estimates.

Uncertainty in time estimates

Age estimates play an important role in our analysis by providing a temporal framework that allows the integration of results of other genetic studies as well as evolutionary events inferred from the genetic analysis with the paleontological, archaeological and palaeoclimatic context. The ρ -statistic used here is equivalent to the one in Thompson *et al.* (2000) which is considered as a good estimator [48]. Yet, simulations indicated that, albeit infrequently, it has a tendency to underestimate the true value [46,48]. Uncertainty also comes from genetic estimates such as the substitution rates and the generation times. In present-day hunter-gatherer societies generation time is estimated to be approximately 32 and 26 years for males and females, respectively [96,97], which leads to an average generation interval of ~ 28 years for chromosome X. Here we used 25 years per generation as a phylogenetic average [7] and the same 25 years generation span to convert time into years in the *MTMR8* tree (table 3). As a consequence of using the same generation estimate (~ 25 years) in both calculations, we end up with a homogenous clock with respect to substitutions per year along the entire length of human and chimpanzee branches. This clock not only ignores any changes in generation time but also the possibility of a relative rate slowdown on the human branch [98]. While using 25 rather than 28 years per generation may be justified because of the long evolutionary depth of the human chromosome X lineages [99], it means that our times in fig. 2 and table 3 are possibly underestimates. This effect will be even more pronounced if our phylogenetic substitution rate were overestimated [7,100], as suggested by recent determinations of the substitution rate directly

in human pedigrees [101-103]. On the other hand, there are good reasons to believe that these recent pedigree estimates may be too low [104]. Our substitution rate, recalculated from its autosomal equivalent assuming $\alpha = 3.3$ [102,104] is $\sim 1.65 \times 10^{-8}$ per bp per generation. This figure is very close to the estimates of 1.3×10^{-8} and 1.8×10^{-8} from human Mendelian disease frequencies by Lynch [105] and Kondrashov [106], respectively, and to a recent estimate of $1.4 - 2.3 \times 10^{-8}$, obtained from a different approach that avoids phylogenetic calibration [104]. Nonetheless we recognize that our ages are likely to be underestimates. On the other hand, our rate was calibrated assuming human-chimpanzee sequence divergence of 7.5 My [7,49,107]. Using 6 My instead, as in many earlier studies, would increase our rate estimate by 20% and lower our time estimates by the same factor (see internal left side scale in fig. 2 and table S4). Importantly, such a shift of the time scale would not affect our results and conclusions that point to a longer evolutionary time frame for human evolution, considering a possibility that contemporary humans originated by amalgamation of lineages from different populations that were separated by years of independent evolution.

Conclusions

The analysis of the *MTMR8* segment diversity and its population tree provided new insights into the evolutionary history of sub-Saharan Africans and demographic events preceding the out of Africa migration. Consistent with earlier studies we found the oldest *MTMR8* lineages among Khoe-San from South-Africa. Worth noting a unique Biaka haplotype branched off directly from the root of the *MTMR8* haplotype tree and may represent a trace of archaic admixture within Africa. Inferred ancient population stratification and the age of the separately evolving lineages deduced from *MTMR8* evolution may imply an older dating of our species, older in fact than the earliest fossils of anatomically modern humans. Additionally, our data indicate that at the time of out of Africa migration sub-Saharan Africans were subdivided for a substantial amount of time. The *MTMR8* segment carries signature of a selection sweep that most likely started prior to the exodus. Such selection events could additionally potentiate the effect of the out-of-Africa demographic bottleneck by reducing the genetic diversity of non-African X chromosomes. Because of its characteristics this subcentromeric region of the X-chromosomes appears particularly interesting to study sub-Saharan populations' history and warrants more detailed analysis using full sequence data.

Web Resources

Arlequin software, <http://cmpg.unibe.ch/software/arlequin3/>; Network software, <http://www.fluxus-engineering.com/>; Program *genetree*, <http://www.stats.ox.ac.uk/~griff/software.html>; Program *msHot*, <http://home.uchicago.edu/~rhudson1/source/mksamples.html>; HGDP data, <http://www.cephb.fr/en/hgdp/>; HapMap3 data, <http://hapmap.ncbi.nlm.nih.gov/>.

Supporting Information

Figure S1 Network of *MTMR8* extended haplotypes in HapMap3 populations [39]. The analyzed region extends over 380 Kb (between sequence positions 63312040 to 63693104) and includes 40 SNPs. A total of 53 haplotypes were observed in 1180

References

- Alves I, Sramkova Hanulova A, Foll M, Excoffier L (2012) Genomic data reveal a complex making of humans. *PLoS Genet* 8: e1002837.

X-chromosomes. Africans: LWK, MKK, YRI, and non-Africans: CEU, TSI, JPT, GIH, CHD. Haplotype frequencies are proportional to the surface of the circle (or to its single colored segment within a population group). (JPG)

Material S1 Supplementary Material. (DOCX)

Table S1 Samples and their Geographic/Ethnic Origin. (DOCX)

Table S2 PCR Primer Sequences. (DOCX)

Table S3 Summary Statistics of Populations Samples. Abbreviations: n – number of chromosomes; S – number of segregating sites (SNPs); k – number of haplotypes; G – gene (haplotype) diversity; $(1-G)$ – haplotype homozygosity; Θ – estimator of population mutation rate $4N\mu$ (N – effective population size; μ – mutation rate per DNA segment per generation); $\Theta\pi$ – estimate from nucleotide diversity [108]; Θ_S – estimate from the number of segregating sites [63]; Θ_H – estimate from frequency of the derived alleles [60]; Θ_{ML} – maximum likelihood estimate by *genetree* [42]; Θ_k – estimate from the number of haplotypes [55]; Θ_G – estimator from haplotype diversity [109,110] (DOCX)

Table S4 Time estimates (Ky \pm S.D.) marking the *MTMR8* segment history (figs. 1 and 2) based on phylogenetic calibration of the mutation rate using human-chimpanzee divergence of 6 My. (DOCX)

Table S5 Extended haplotypes of a subset of our Sub-Saharan samples from HGDP project that were used in our study (Table S1) and for which genome-wide genotypes were available at <http://www.cephb.fr/en/hgdp/> [38]. Two positions (A>G8 and T>C36 of the haplotype) whose derived alleles are shared with Neandertals and Denisovans are separated by 1585 Kb, whereas the distance between the same A>G8 and the leftmost C>T24 in the Biaka HT16 (H22 in Fig.1) is 812 Kb. The polymorphism C>T13, with derived T highlighted in red corresponds to the mutation 16 in the network in Fig.1 of the main text. The correspondence between the extended haplotypes below and the Fig.1 network haplotypes are on the right. (DOCX)

Acknowledgments

We are grateful to Gil McVean who interested us in this genomic region, to Malgorzata Labuda who selected the DNA segment we analysed and to Luis Barreiro for his critical comments. We are also indebted to all individuals who consented to participate in genetic studies and donated their biological material, to all colleagues who participated in sample collection and also to those who generously shared DNA samples from their collections.

Author Contributions

Conceived and designed the experiments: DL. Performed the experiments: VY. Analyzed the data: VY JFL CM. Contributed reagents/materials/analysis tools: GU SMW. Wrote the paper: DL. Contributed to writing the manuscript: VY JFL CM GU SMW.

- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328: 710–722.

3. Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, et al. (2011) Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *Am J Hum Genet* 89: 516–528.
4. Yotova V, Lefebvre JF, Moreau C, Gbela E, Hovhannesyan K, et al. (2011) An X-linked haplotype of Neandertal origin is present among all non-African populations. *Mol Biol Evol* 28: 1957–1962.
5. Hawks J (2012) Longer time scale for human evolution. *Proc Natl Acad Sci U S A* 109: 15531–15532.
6. Yotova V, Lefebvre JF, Kohany O, Jurka J, Michalski R, et al. (2007) Tracing genetic history of modern humans using X-chromosome lineages. *Hum Genet* 122: 431–443.
7. Langergraber KE, Prufer K, Rowney C, Boesch C, Crockford C, et al. (2012) Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci U S A* 109: 15716–15721.
8. Scally A, Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* 13: 745–753.
9. Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD (2011) Genetic evidence for archaic admixture in Africa. *Proc Natl Acad Sci U S A* 108: 15123–15128.
10. Labuda D, Zietkiewicz E, Yotova V (2000) Archaic lineages in the history of modern humans. *Genetics* 156: 799–808.
11. Zietkiewicz E, Yotova V, Jamik M, Korab-Laskowska M, Kidd KK, et al. (1998) Genetic structure of the ancestral population of modern humans. *J Mol Evol* 47: 146–155.
12. Relethford JH, Jorde LB (1999) Genetic evidence for larger African population size during recent human evolution. *Am J Phys Anthropol* 108: 251–260.
13. Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, et al. (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins [see comments]. *Science* 271: 1380–1387.
14. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci U S A* 102: 15942–15947.
15. Garrigan D, Mobasher Z, Kingan SB, Wilder JA, Hammer MF (2005) Deep haplotype divergence and long-range linkage disequilibrium at xp21.1 provide evidence that humans descend from a structured ancestral population. *Genetics* 170: 1849–1856.
16. Hayakawa T, Aki I, Varki A, Satta Y, Takahata N (2006) Fixation of the human-specific CMP-N-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. *Genetics* 172: 1139–1146.
17. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, et al. (2012) Evolutionary History and Adaptation from High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers. *Cell* 150: 457–469.
18. Wall JD, Lohmueller KE, Plagnol V (2009) Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol Biol Evol* 26: 1823–1827.
19. Atkinson QD, Gray RD, Drummond AJ (2009) Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc Biol Sci* 276: 367–373.
20. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, et al. (2009) The genetic structure and history of Africans and African Americans. *Science* 324: 1035–1044.
21. Watson E, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61: 691–704.
22. Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, et al. (2004) Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 74: 1014–1022.
23. Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, et al. (2002) The making of the African mtDNA landscape. *Am J Hum Genet* 71: 1082–1111.
24. Behar DM, Villemers R, Soodyall H, Blue-Smith J, Pereira L, et al. (2008) The dawn of human matrilineal diversity. *Am J Hum Genet* 82: 1130–1140.
25. Shimada MK, Panchapakesan K, Tishkoff SA, Nato AQ Jr, Hey J (2007) Divergent haplotypes and human history as revealed in a worldwide survey of X-linked DNA sequence variation. *Mol Biol Evol* 24: 687–698.
26. Schlebusch CM, Lombard M, Soodyall H (2013) MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. *BMC Evol Biol* 13: 56.
27. Keinan A, Reich D (2010) Human population differentiation is strongly correlated with local recombination rate. *PLoS Genet* 6: e1000886.
28. Jaruzelska J, Zietkiewicz E, Batzer M, Cole DE, Moisan JP, et al. (1999) Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics* 152: 1091–1101.
29. Zietkiewicz E, Yotova V, Jamik M, Korab-Laskowska M, Kidd KK, et al. (1997) Nuclear DNA diversity in worldwide distributed human populations. *Gene* 205: 161–171.
30. Mahtani MM, Willard HF (1998) Physical and genetic mapping of the human X chromosome centromere: repression of recombination. *Genome Res* 8: 100–110.
31. Nachman MW, D'Agostino SL, Tillquist CR, Mobasher Z, Hammer MF (2004) Nucleotide variation at *Msn* and *Alas2*, two genes flanking the centromere of the X chromosome in humans. *Genetics* 167: 423–437.
32. Lambert CA, Connelly CF, Madeoy J, Qiu R, Olson MV, et al. (2010) Highly punctuated patterns of population structure on the X chromosome and implications for African evolutionary history. *Am J Hum Genet* 86: 34–44.
33. Casto AM, Li JZ, Absher D, Myers R, Ramachandran S, et al. (2010) Characterization of X-Linked SNP genotypic variation in globally-distributed human populations. *Genome Biol* 11: R10.
34. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, et al. (2002) A human genome diversity cell line panel. *Science* 296: 261–262.
35. Lovell A, Moreau C, Yotova V, Xiao F, Bourgeois S, et al. (2005) Ethiopia: between Sub-Saharan Africa and western Eurasia. *Ann Hum Genet* 69: 275–287.
36. Zietkiewicz E, Yotova V, Gehl D, Wambach T, Arrieta I, et al. (2003) Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern humans' diversity. *Am J Hum Genet* 73: 994–1015.
37. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851–861.
38. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319: 1100–1104.
39. Altschuler DM, Gibbs RA, Peltonen L, Demitzakis E, Schaffner SF, et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52–58.
40. Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem* 49: 853–860.
41. Labuda D, Striker G (1989) Sequence conservation in Alu evolution. *Nucleic Acids Res* 17: 2477–2491.
42. Griffiths RC, Tavare S (1994) Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci* 344: 403–410.
43. Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59: 935–945.
44. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.
45. Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67: 718–726.
46. Cox MP (2008) Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol* 80: 335–357.
47. Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A* 97: 7360–7365.
48. Hudson RR (2007) The variance of coalescent time estimates from DNA sequences. *J Mol Evol* 64: 702–705.
49. Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life. *Mol Biol Evol* 24: 26–53.
50. Labuda D, Lefebvre JF, Roy-Gagnon MH (2010) Response to Lohmueller et al.: Sex-Averaged Recombination and Mutation Rates on the X Chromosome: A Comment on Labuda et al. *Am J Hum Genet* 86: 980–981.
51. Labuda D, Lefebvre JF, Nadeau P, Roy-Gagnon MH (2010) Female-to-male breeding ratio in modern humans—an analysis based on historical recombinations. *Am J Hum Genet* 86: 353–363 and *AJHG* 386:978–981.
52. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.
53. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
54. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
55. Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3: 87–112.
56. Watterson GA (1978) The Homozygosity Test of Neutrality. *Genetics* 88: 405–417.
57. Slatkin M (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genet Res* 64: 71–74.
58. Chakraborty R (1990) Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. *Am J Hum Genet* 47: 87–94.
59. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
60. Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.
61. Hellenthal G, Stephens M (2007) msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics* 23: 520–521.
62. Hudson RR (1990) Gene Genealogies and the Coalescent Process. *Oxford Surveys in Evolutionary Biology* 7: 1–44.
63. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7: 256–276.
64. Kitano T, Schwarz C, Nickel B, Paabo S (2003) Gene diversity patterns at 10 X-chromosomal loci in humans and chimpanzees. *Mol Biol Evol* 20: 1281–1289.

65. Harris EE, Hey J (1999) X chromosome evidence for ancient human histories. *Proc Natl Acad Sci U S A* 96: 3320–3324.
66. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, et al. (2012) A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338: 222–226.
67. Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, et al. (2012) Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science* 338 (6105): 374–369.
68. Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, et al. (2012) The genetic prehistory of southern Africa. *Nat Commun* 3: 1143.
69. Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, et al. (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci U S A* 108: 5154–5162.
70. Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, et al. (2012) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Mol Biol Evol* 29: 617–630.
71. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43: 1031–1034.
72. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475: 493–496.
73. Schlebusch CM, Soodyall H (2012) Extensive Population Structure in San, Khoe, and Mixed Ancestry Populations from Southern Africa Revealed by 44 Short 5-SNP Haplotypes. *Hum Biol* 84: 695–724.
74. Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA (2007) Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 24: 757–768.
75. de Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdottir ED, et al. (2011) Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol Biol Evol* 28: 1255–1269.
76. Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, et al. (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.
77. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, et al. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26: 358–361.
78. Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, et al. (2011) A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *Am J Hum Genet* 88: 814–818.
79. Petraglia MD, Haslam M, Fuller DQ, Boivin N, Clarkson C (2010) Out of Africa: new hypotheses and evidence for the dispersal of *Homo sapiens* along the Indian Ocean rim. *Ann Hum Biol* 37: 288–311.
80. Mellars P (2006) Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313: 796–800.
81. Shen G, Wu X, Wang Q, Tu H, Feng YX, et al. (2013) Mass spectrometric U-series dating of Huanglong Cave in Hubei Province, central China: Evidence for early presence of modern humans in eastern Asia. *J Hum Evol* 65: 162–167.
82. Oppenheimer S (2012) Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos Trans R Soc Lond B Biol Sci* 367: 770–784.
83. Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, et al. (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet* 75: 752–770.
84. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, et al. (2012) Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am J Hum Genet* 91: 83–96.
85. Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA (2002) Ethiopians and Khoisans share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 70: 265–268.
86. Gunz P, Bookstein FL, Mitteroecker P, Stadlmayr A, Seidler H, et al. (2009) Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proc Natl Acad Sci U S A* 106: 6094–6098.
87. Scholz CA, Johnson TC, Cohen AS, King JW, Peck JA, et al. (2007) East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proc Natl Acad Sci U S A* 104: 16416–16421.
88. Cohen AS, Stone JR, Beuning KR, Park LE, Reinthal PN, et al. (2007) Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. *Proc Natl Acad Sci U S A* 104: 16422–16427.
89. Eriksson A, Manica A (2012) Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc Natl Acad Sci U S A* 109: 13956–13960.
90. Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41: 539–564.
91. Heyer E, Chaix R, Pavard S, Austerlitz F (2012) Sex-specific demographic behaviours that shape human genomic variation. *Mol Ecol* 21: 597–612.
92. Emery LS, Felsenstein J, Akey JM (2010) Estimators of the human effective sex ratio detect sex biases on different timescales. *Am J Hum Genet* 87: 848–856.
93. Keinan A, Mullikin JC, Patterson N, Reich D (2009) Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat Genet* 41: 66–70.
94. Barreiro LB, Patin E, Neyrolles O, Cann HM, Gicquel B, et al. (2005) The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. *Am J Hum Genet* 77: 869–886.
95. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, et al. (2003) Traces of human migrations in *Helicobacter pylori* populations. *Science* 299: 1582–1585.
96. Matsumura S, Forster P (2008) Generation time and effective population size in Polar Eskimos. *Proc Biol Sci* 275: 1501–1508.
97. Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 128: 415–423.
98. Elango N, Thomas JW, Yi SV (2006) Variable molecular clocks in hominoids. *Proc Natl Acad Sci U S A* 103: 1370–1375.
99. Blum MG, Jakobsson M (2011) Deep divergences of human gene trees and models of human origins. *Mol Biol Evol* 28: 889–898.
100. Sjödin P, Sjöstrand AE, Jakobsson M, Blum MG (2012) Resequencing Data Provide No Evidence for a Human Bottleneck in Africa during the Penultimate Glacial Period. *Mol Biol Evol* 29: 1851–1860.
101. Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, et al. (2011) Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43: 712–714.
102. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, et al. (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488: 471–475.
103. Roach JC, Glusman G, Hubley R, Montsaroff SZ, Holloway AK, et al. (2011) Chromosomal haplotypes by genetic phasing of human families. *Am J Hum Genet* 89: 382–397.
104. Sun JX, Helgason A, Masson G, Ebenesersdottir SS, Li H, et al. (2012) A direct characterization of human mutation based on microsatellites. *Nat Genet* 44: 1161–1165.
105. Lynch M (2010) Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A* 107: 961–968.
106. Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 21: 12–27.
107. Scally A, Duthieil JY, Hillier LW, Jordan GE, Goodhead I, et al. (2012) Insights into hominid evolution from the gorilla genome sequence. *Nature* 483: 169–175.
108. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.
109. Zouros E (1979) Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics* 92: 623–646.
110. Chakraborty R, Weiss KM (1991) Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. *Am J Phys Anthropol* 86: 497–506.