bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The most recent common ancestor for Y chromosome lived about 3.67 million years ago

Tingting Sun[1], Qing Liu[1], Meiqi Shang[1], Kejian Wang[1*]

[1]State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou, China

[*]Correspondence: Kejian Wang, Email: wangkejian@caas.cn

Key words: mtDNA, Y-DNA, most recent common ancestor, modern *Homo sapiens*

Author's note: The manuscript has not been peer-reviewed. Any comments, suggestions and criticisms will be greatly appreciated.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The origin of human beings is one of the most important questions in science. A combination of numerous archaeological and genomic analyses has led to the widely accepted opinion that modern humans are the descendants of anatomically modern *Homo sapiens* that originated in Africa about 200 thousand years ago (KYA)[1,2]. In this study, we reanalysed the mitochondrial DNA and Y chromosome DNA of the 1000 Genomes Project[3], and found many minority-specific single-nucleotide polymorphisms. Using these polymorphisms, we recalculated the time taken for the evolution of modern humans. Analysis of mitochondrial DNA suggested that the most recent common female ancestor lived about 400 KYA and began to leave Africa about 180 KYA. In contrast, analysis of Y chromosome DNA revealed that the most recent common male ancestor lived about 3.67 million years ago (MYA) and began to migrate out of Africa about 2.05 MYA, a time which is consistent with the expansion time of *Homo erectus* identified by archaeological research. Based on the findings, we proposed a new migration routes and times of modern human, and speculated that anatomically modern *Homo sapiens* has been extensively interbred with local archaic human population during their dispersal across the globe.

The origin of humans is a topic of a large amount of research and discussion. Data from archaeological sites have led to the widely accepted view that humans originated in Africa, and have migrated out of Africa at least three times[4-7]. About 2–1.7 million years ago (MYA), *Homo erectus* spread to other continents, shortly after emerging in Africa[4,8,9]. The second expansion, the migration of ancestors of Neanderthals into Europe, occurred around 0.8–0.6 MYA, and the third wave was the dispersal of anatomically modern *Homo sapiens* about 130–90 thousand years ago (KYA)[10,11]. With the development of molecular anthropology, mitochondrial DNA (mtDNA)

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and Y chromosome DNA (Y-DNA), both of which are inherited in a haploid manner, have been extensively used to analyse the origins of modern humans[1,2,12,13]. Studies of mtDNA have shown that all human beings are descendants of a common female ancestor who lived about 200 KYA in Africa. Her descendants migrated out of Africa around 100 KYA[2,8,14]. Y-DNA can be used to trace the male linage. Early Y-DNA evidence suggested that the common male ancestor of modern humans lived in Africa about 150–60 KYA[15-17]. However, recent analyses of 1,244 Y chromosomes have led to speculation that the paternal ancestors lived about 190 KYA and left Africa about 100 KYA[1,8]. Combining archaeological and genetic studies, it is now widely accepted that anatomically modern humans left Africa about 130–90 KYA and then spread throughout the world, resulting in the global replacement and genetic extinction of pre-existing local archaic human populations[8,18].

Single-nucleotide polymorphisms (SNPs) are the most common type of polymorphism, involving variation of a single base pair. As the introduction of new SNPs occurs at a constant rate, common SNPs have been widely used to determine the time of common ancestor of different haplogroups over the course of human evolution[2,17]. However, until recently, only a very limited number of individuals had been sequenced and analysed. We speculated that each individual sequenced might contain a certain number of minority-specific variations, which will be classified as low-frequency SNPs (less than 1% or 5% of the population) and discarded in the study of population genetics and evolutionary biology[19,20]. To test this hypothesis, we reanalysed the highly accurate single nucleotide variants of 2,534 human individuals released by the 1000 Genomes Project[1,3]. We first analysed the SNPs present in mtDNA (about 14 kb, specific location: 489 bp–14,783 bp). There were 2,474 SNPs in the mtDNA of the 2,534 individuals. Of these SNPs, 2,037 (82%) were present in fewer than ten individuals (accounting for 0.36% of individuals analysed), and 926 (37%) SNPs were specific to a

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

single individual (Figure 1a). There were 48,226 SNPs found in Y-DNA (approximately 25 Mb, 2,762,077 bp–28,508,881 bp) from 1,233 male individuals. Of these SNPs, 40,576 (84%) were present in fewer than ten individuals (accounting for 0.73% of male individuals analysed), and 27,127 (56%) occurred in a single individual (Figure 1a). These results indicate the existence of a large number of minority-specific SNP sites, which are not likely to be analysed in population genetics.

Since new SNPs are introduced into the genome at a constant rate[21], it can be inferred that each individual has a similar number of SNPs compared with the common ancestor. We further speculated that each individual also shares a similar number of SNP variations with a common out-group. To test the reliability of these minority-specific SNPs, we chose the genome of the chimpanzee as the common out-group because it was the closest living relatives of human beings and had similar mutation rate with human[22]. We found that each individual had a similar number of SNP differences with the chimpanzee analysed, with very limited variation in both mtDNA (541.36 ± 3.79) and in Y-DNA (2979.15 ± 26.41) (Figure 1b). This observation indicates that the minority-specific SNPs are reliable.

Next, we reanalysed mtDNA and Y-DNA using all SNPs, including the minority-specific SNPs, and calculated the time to the Most Recent Common Ancestor (MRCA) using a simple model (Figure 2). The principle underlying the model is as follows. The number of SNP differences between the out-group (O) and A/B individuals or groups (the SNP differences between each individual and outgroup are calculated and then take the average) is proportional to twice the divergence time (T_m) between the out-group and A/B. Similarly, the number of SNP differences between A and B is proportional to twice the differentiation time (T_n) between A and B. When the split time between the human and the out-group is known, the differentiation time between A and B can be determined. The time is calculated according to the following formula:

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

106 $$\frac{2T_n}{2T_m} = \frac{N_{AB}}{\frac{N_{AO} + N_{BO}}{2}}$$

107 To determine the time of major events affecting ancient humans, we first
108 constructed a neighbour-joining tree using the mtDNA of one chimpanzee
109 and all 2,534 individuals in the 1000 Genome Project[3]. The exact time of
110 divergence between humans and chimpanzees remains subject to debate, and
111 the most used time five million years was used in this study[15]. The
112 evolutionary tree suggested that the common mtDNA ancestor of modern
113 humans originated in Africa (Figure 3), which was in keeping with previous
114 studies[2,14]. In the polygenetic tree, Node 1 represents the MRCA of all
115 human individuals analysed. We calculated the time according to the above
116 formula and found that the MRCA of modern humans lived about 400 KYA
117 (Figure 3, Table 1 and Supplementary Table S1), which is about twice of the
118 time deduced from previous haplogroup analysis[2,14]. From Node 2, we
119 concluded that humans began to left Africa about 180 KYA, which is earlier
120 than the widely accepted 130–90 KYA inferred from early archaeological
121 and genomic studies[8,23], but approximately coincides with recent
122 archaeological findings of modern humans in the Misliya Cave in Israel,
123 dated to ~180 KYA, and the Apidima Cave in Greece, dated to around 210
124 KYA[24,25]. From Node 3, we inferred that the MRCA of all non-Africans also
125 lived about 180 KYA, which is also earlier than the 90 KYA deduced from
126 haplogroup analysis[2]. We also analysed the differentiation time of humans in
127 other branches. We found that the mtDNA of Europeans began to split from
128 that of West and South Asians (Node 12) at about 150 KYA, and their
129 MRCA lived around 70–40 KYA (Node 15 and 16), which is close to the 45
130 KYA discovered in archaeology[26,27]. We also found that the mtDNA of East
131 Asians split from that of West and South Asians (Node 10) around 130 KYA,
132 and their MRCA lived approximately 110–90 KYA (Node 7 and 11), which
133 is in agreement with the earliest known modern humans discovered by
134 archaeologists in southern China[28], which have been dated to around 120–80

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

KYA. We found that mtDNA of Americans diverged from that of East Asians (Node 4, 6, 8 and 13) around 110–70 KYA, and their MRCA lived approximately 60–40 KYA (Node 5, 9 and 14), which is much earlier than the widely accepted date for human arrival in America of about 15 KYA[29], but close to the 30 KYA deduced from a recently discovered Mexican cave[30,31].

We further constructed a neighbour-joining tree using the Y-DNA of 1,233 individuals and the chimpanzee (Figure 4). Using the same calculation, we found that the common patrilineal ancestors of modern humans (Node 1) arose in Africa about 3.67 MYA (Figure 4, Table 2 and Supplementary Table S2), which is nearly 20 times earlier than the 190 KYA estimated from recent haplogroup analysis[1], and even older than that deduced from the archaeological record of *Homo habilis*, which arose about 3 MYA[32]. From Node 2, we inferred that the Y-DNA of modern humans began to migrate out of Africa approximately 2.05 MYA, which coincides with the archaeological dispersal time (2–1.7 MYA) of *Homo erectus*[8,9]. We found that the MRCA of non-African lineages (Node 3) lived about 1.54 MYA, which is about 15 times earlier than the 76 KYA produced by previous haplogroup analysis[1]. It appeared that the Y-DNA of East Asians began to split from that of West and South Asians (Node 4) around 0.97 MYA, and their MRCA lived about 0.91 MYA (Node 5). According to our analysis, the Y-DNA of Europeans split from that of West and South Asians (Node 8) around 0.66 MYA, which approximately coincides with the arrival time (0.8–0.6 MYA) of ancestor of Neanderthals in Europe[9]. We also inferred that the American patrilineal ancestors separated from West and South Asians (Node 6) at about 0.43 MYA, and the MRCA of native Americans lived about 0.31 MYA (Node 7).

Previous genomic evidence from mtDNA and Y-DNA suggested that the paternal and maternal ancestors of modern humans both lived within 200 KYA, and migrated out of Africa about 100 KYA[1,2,8]. This time is roughly

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

consistent with the archaeologically determined time (130–90 KYA) of anatomically modern *Homo sapiens*[10,33], but much later than that of archaic human populations, such as *Homo erectus* and archaic *Homo sapiens*[4,34]. Thus, it is now widely accepted that modern humans are the offspring of anatomically modern *Homo sapiens*, while pre-existing local archaic human populations became extinct with the arrival of modern *Homo sapiens*[35]. In this study, analysis of the mtDNA showed that the time of the female MRCA approximately corresponds with the rise of anatomically modern *Homo sapiens*, implying that the matrilineal ancestors of modern humans are likely to be anatomically modern *Homo sapiens*. However, analysis of the Y-DNA implied that the time of the male MRCA is even earlier than the time of *Homo habilis.* In addition, the time left Africa of male MRCA is consistent with that of *Homo erectus*. Based on the findings, we proposed a new migration routes and times of modern human (Figure 5). Furthermore, we speculated that anatomically modern *Homo sapiens* has been extensively interbred with local archaic human population during their dispersal across the globe. All modern humans inherited mtDNA from female anatomically modern *Homo sapiens* and Y-DNA from male archaic human population in different locations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Read mapping and SNP calling for out-group

The sequencing of 24 chromosomes and the mitochondrial genome—chromosomes 2A and 2B merged into chromosome 2—of the chimpanzee (GCF_002880755.1_Clint_PTRv2_genomic.fna https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/880/755/GCF_002880755.1_Clint_PTRv2/GCF_002880755.1_Clint_PTRv2_genomic.fna.gz) was simulated using ART software with the parameters '–l 150', '–f 60', '–m 500', '–s 10', '-p', '-na' and '–ef'[36]. The high quality paired-end reads were aligned to the hs37d5 reference genome using BWA (v0.7.17) with the parameters: 'mem -t 16 -k 32–M'[37]. PCR duplicates were removed using Picard tools (http://broadinstitute.github.io/picard/). We performed SNP calling using a HaplotypeCaller approach as implemented in the package GATK[38]. A Perl programme was used to filter potentially false SNPs using the following criteria: a homozygous genotype; supporting reads for the reference or alternative allele greater than 30; and the ratio of them was greater than 3.

Population genetic analysis

Single-nucleotide polymorphism genotyping data available from the HapMap3 Project (Altshuler et al. 2010) was downloaded from the NCBI (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/)[3]. The linkage disequilibrium (LD r^2) of all SNPs in the mitochondrial genome within 50,000 kb were calculated using the Haploview software[39]. For the Y chromosome, 200 SNPs were randomly selected from the SNP genotyping dataset to calculate the LD within 50,000 kb using Haploview. The final SNP dataset was formed by removing unmatched SNP positions between the chimpanzee and humans in the LD block. The IBS distance matrix of the mitochondrial genome and the Y-DNA was computed using PLINK on the SNP set which was located in the LD block[40]. The same distance matrix was used to construct a phylogenetic tree by the neighbour-joining method,

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

implementing in PHYLIP and visualised in R[41,42].

Analysis of SNP differences between different groups

Individual-specific SNPs and the number of SNP differences between each individual and chimpanzee were calculated using a Perl script. When counting the number of SNP differences, genotypes designated '-', were not counted. For branches with more than one individual, the average number of SNP differences over all individuals was calculated.

Author Contributions

KW conceived, designed and supervised the study, interpreted the data and wrote the paper; TS conceived the study, conducted most of the analysis and wrote the paper; QL and MS discussed the results and verified the data. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional files

Supplementary Table S1: Individuals of two branches from each differentiation node of mtDNA

Supplementary Table S2: Individuals of two branches from each differentiation node of Y-DNA

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1 Poznik, G. D. *et al.* Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet* **48**, 593-599, doi:10.1038/ng.3559 (2016).

2 Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. & Tishkoff, S. A. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* **24**, 757-768, doi:10.1093/molbev/msl209 (2007).

3 Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58, doi:10.1038/nature09298 (2010).

4 Spoor, F. *et al.* Implications of new early Homo fossils from Ileret, east of Lake Turkana, Kenya. *Nature* **448**, 688-691, doi:10.1038/nature05986 (2007).

5 McDougall, I., Brown, F. H. & Fleagle, J. G. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* **433**, 733-736, doi:10.1038/nature03258 (2005).

6 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

7 Hublin, J. J. *et al.* New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature* **546**, 289-292, doi:10.1038/nature22336 (2017).

8 Templeton, A. Out of Africa again and again. *Nature* **416**, 45-51, doi:10.1038/416045a (2002).

9 Delson, E. An early modern human outside Africa. *Nature* **571**, 487-488 (2019).

10 Groucutt, H. S. *et al.* Homo sapiens in Arabia by 85,000 years ago. *Nat Ecol Evol* **2**, 800-809, doi:10.1038/s41559-018-0518-2 (2018).

11 Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L. & Skoglund, P. Origins of modern human ancestry. *Nature* **590**, 229-237, doi:10.1038/s41586-021-03244-5 (2021).

12 Schuster, S. C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463**, 943-947, doi:10.1038/nature08795 (2010).

13 Poznik, G. D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562-565, doi:10.1126/science.1237619 (2013).

14 Cann, R. L., Stoneking, M. & Wilson, A. C. Mitochondrial DNA and human evolution. *Nature* **325**, 31-36, doi:10.1038/325031a0 (1987).

15 Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman, M. W. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A* **97**, 7360-7365, doi:10.1073/pnas.97.13.7360 (2000).

16 Wei, W. *et al.* A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res* **23**, 388-395, doi:10.1101/gr.143198.112 (2013).

17 Cruciani, F. *et al.* A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *Am J Hum Genet* **88**, 814-818, doi:10.1016/j.ajhg.2011.05.002 (2011).

18 Stoneking, M. & Soodyall, H. Human evolution and the mitochondrial genome. *Current opinion in genetics & development* **6**, 731-736, doi:10.1016/s0959-437x(96)80028-1 (1996).

19 Zhang, C., Dong, S. S., Xu, J. Y., He, W. M. & Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786-1788, doi:10.1093/bioinformatics/bty875 (2019).

20 Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587 (2003).

21 Bromham, L. & Penny, D. The modern molecular clock. *Nat Rev Genet* **4**, 216-224,

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

287 doi:10.1038/nrg1020 (2003).
288 22 Venn, O. *et al.* Nonhuman genetics. Strong male bias drives germline mutation in
289 chimpanzees. *Science* **344**, 1272-1275, doi:10.1126/science.344.6189.1272 (2014).
290 23 Reyes-Centeno, H. *et al.* Genomic and cranial phenotype data support multiple
291 modern human dispersals from Africa and a southern route into Asia. *Proc Natl Acad*
292 *Sci U S A* **111**, 7248-7253, doi:10.1073/pnas.1323666111 (2014).
293 24 Hershkovitz, I. *et al.* The earliest modern humans outside Africa. *Science* **359**,
294 456-459, doi:10.1126/science.aap8369 (2018).
295 25 Harvati, K. *et al.* Apidima Cave fossils provide earliest evidence of Homo sapiens in
296 Eurasia. *Nature* **571**, 500-504, doi:10.1038/s41586-019-1376-z (2019).
297 26 Hublin, J. J. *et al.* Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave,
298 Bulgaria. *Nature* **581**, 299-302, doi:10.1038/s41586-020-2259-z (2020).
299 27 Fewlass, H. *et al.* A (14)C chronology for the Middle to Upper Palaeolithic transition
300 at Bacho Kiro Cave, Bulgaria. *Nat Ecol Evol* **4**, 794-801,
301 doi:10.1038/s41559-020-1136-3 (2020).
302 28 Liu, W. *et al.* The earliest unequivocally modern humans in southern China. *Nature*
303 **526**, 696-699, doi:10.1038/nature15696 (2015).
304 29 Goebel, T., Waters, M. R. & O'Rourke, D. H. The late Pleistocene dispersal of modern
305 humans in the Americas. *Science* **319**, 1497-1502, doi:10.1126/science.1153569
306 (2008).
307 30 Clark, P. U. *et al.* The Last Glacial Maximum. *Science* **325**, 710-714,
308 doi:10.1126/science.1172873 (2009).
309 31 Ardelean, C. F. *et al.* Evidence of human occupation in Mexico around the Last Glacial
310 Maximum. *Nature* **584**, 87-92, doi:10.1038/s41586-020-2509-0 (2020).
311 32 Leakey, L. S., Tobias, P. V. & Napier, J. R. A NEW SPECIES OF THE GENUS HOMO FROM
312 OLDUVAI GORGE. *Nature* **202**, 7-9, doi:10.1038/202007a0 (1964).
313 33 Grün, R. *et al.* U-series and ESR analyses of bones and teeth relating to the human
314 burials from Skhul. *J Hum Evol* **49**, 316-334, doi:10.1016/j.jhevol.2005.04.006 (2005).
315 34 Carbonell, E. *et al.* The first hominin of Europe. *Nature* **452**, 465-469,
316 doi:10.1038/nature06815 (2008).
317 35 Ayala, F. J. Cloning humans? Biological, ethical, and social considerations. *Proc Natl*
318 *Acad Sci U S A* **112**, 8879-8886, doi:10.1073/pnas.1501798112 (2015).
319 36 Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read
320 simulator. *Bioinformatics* **28**, 593-594, doi:10.1093/bioinformatics/btr708 (2012).
321 37 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
322 transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
323 38 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
324 analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303,
325 doi:10.1101/gr.107524.110 (2010).
326 39 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of
327 LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
328 40 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and
329 population-based linkage analyses. *The American journal of human genetics* **81**,
330 559-575 (2007).
331 41 Felsenstein, J. Phylogeny inference package. *Department of Genetics, University of*
332 *Washington. Seatle* (1993).
333 42 Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an R package for
334 visualization and annotation of phylogenetic trees with their covariates and other
335 associated data. *Methods in Ecology and Evolution* **8**, 28-36 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

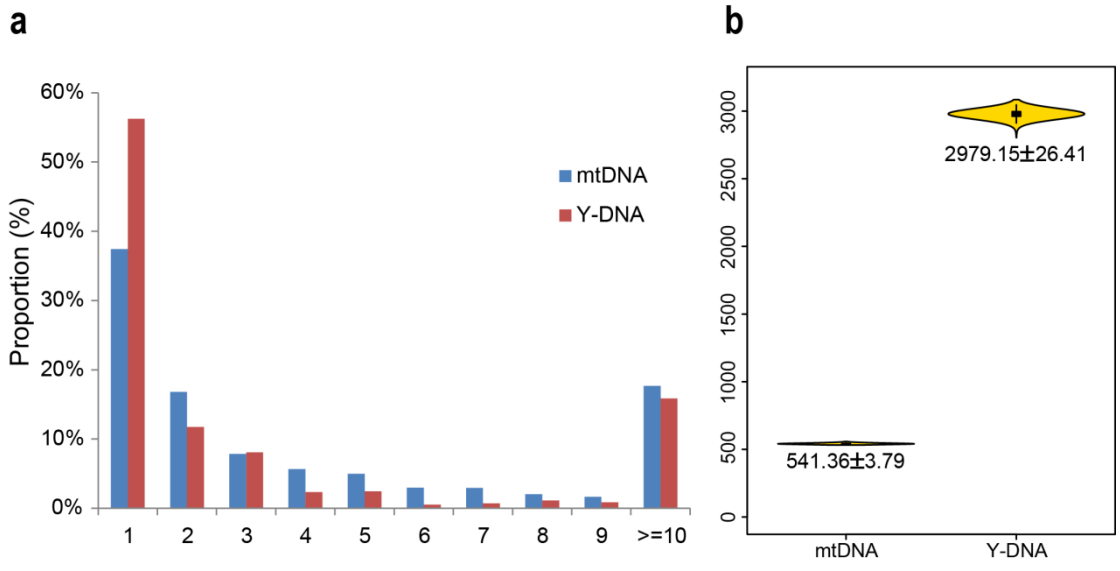

Figure 1: Analysis of the SNPs specificity and differences in mtDNA and Y-DNA

a. Proportions of specific SNPs in mtDNA and Y-DNA. The X-axis represents the number of individuals sharing the same minority-specific SNP. The Y-axis represents the proportion of SNPs. **b. SNP differences between all individuals and one chimpanzee.** Black outlines are violin plots. Black rectangles indicate interquartile ranges, and black lines indicate the 95% confidence interval. Numbers represent mean ±SD.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

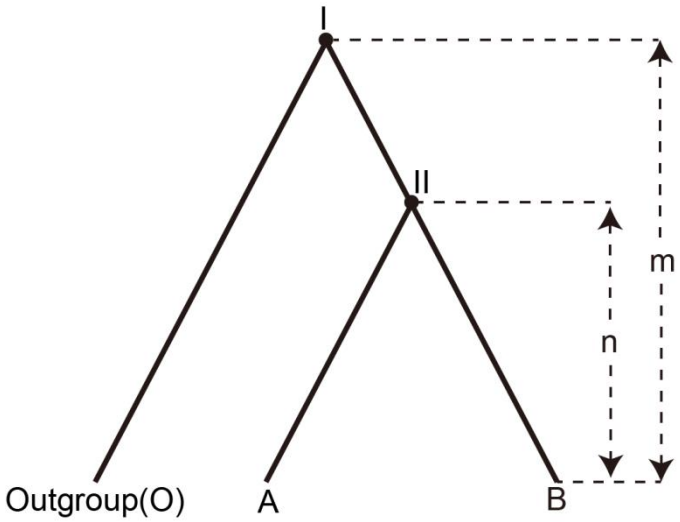

Figure 2: Time calculation model used in this study

A and B represent two different individuals or groups. Node I represents the Most Recent Common Ancestor (MRCA) of Out-group (O) and A/B. The corresponding time is represented as m. Node II represents the MRCA of A and B, and the corresponding time is indicated as n.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

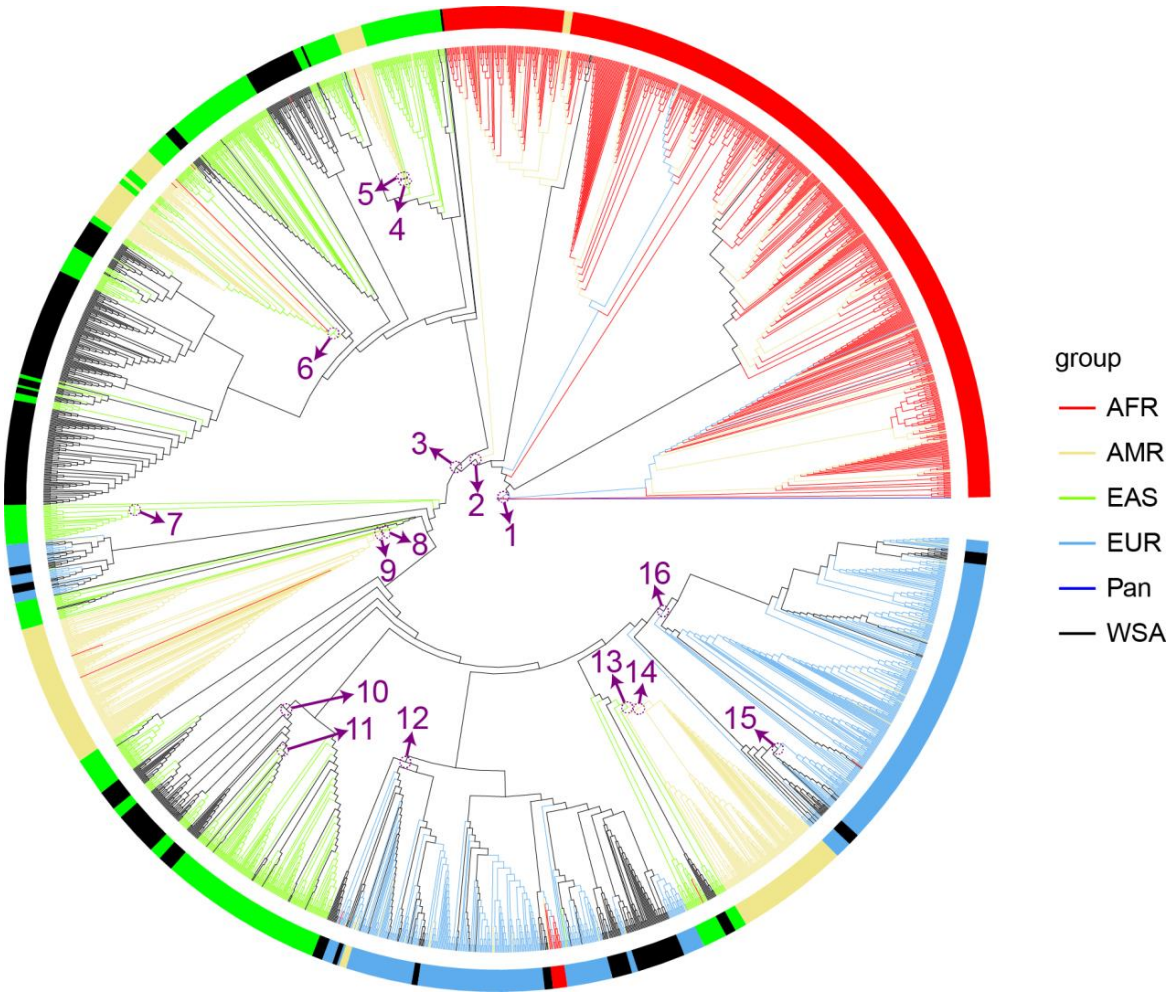

Figure 3: Evolutionary tree based on mtDNA

Populations from Africa (AFR), America (AMR), East Asia (EAS), Europe (EUR), West and South Asia (WSA) were used to construct an evolutionary tree based on mtDNA. Dashed circles and corresponding numbers represent the differentiation nodes between the two groups.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

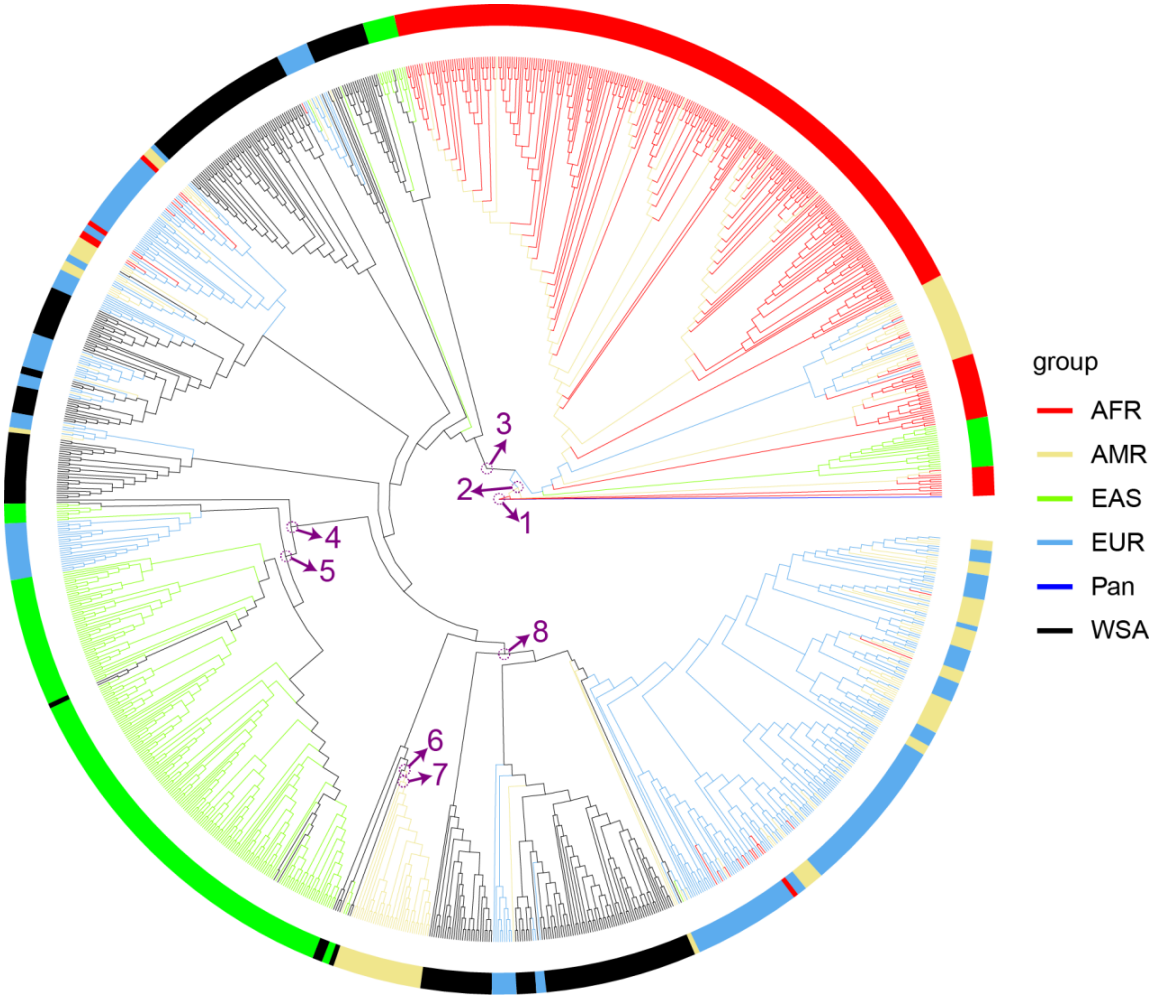

Figure 4: Evolutionary tree based on Y-DNA

Populations from Africa (AFR), America (AMR), East Asia (EAS), Europe (EUR), West and South Asia (WSA) were used to construct an evolutionary tree based on Y-DNA. Dashed circles and corresponding numbers represent the differentiation nodes between the two groups.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

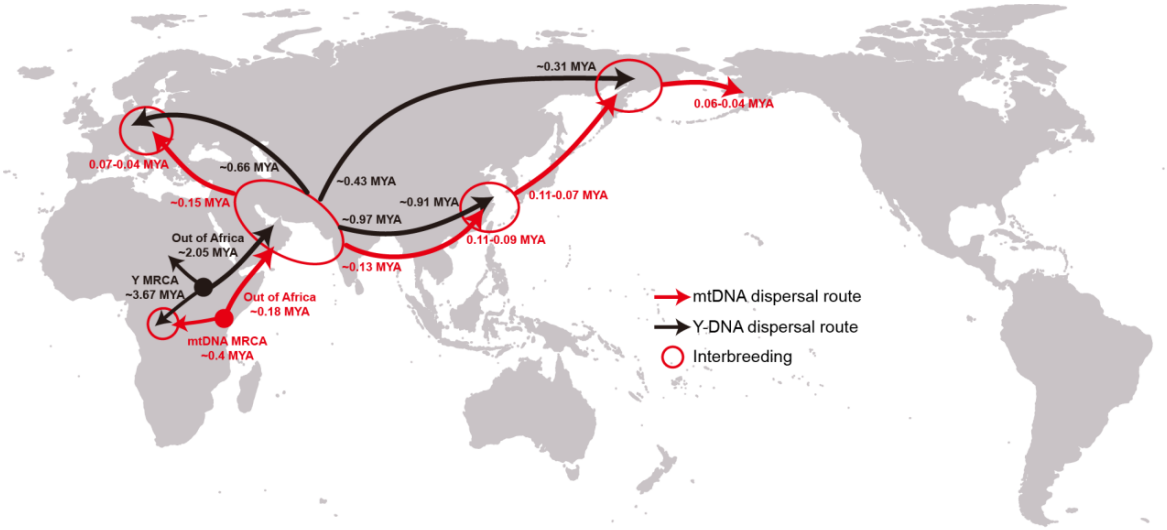

Figure 5: Migration routes and times of prehistoric human populations around the world, based on evidence from mtDNA and Y-DNA

Red and black arrows indicate the putative migration routes of mtDNA and Y-DNA, respectively. The numbers at the end of each arrow indicate the time when the populations of the two regions began to diverge. The numbers on the arrows represent the time of the common ancestor of the arriving region's population. Time unit: million years ago (MYA).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: A proposed migratory timeline for humans migrating across the world based on evidence from mtDNA

Events	Node	N_{AO}	N_{BO}	N_{AB}	Corresponding time (MYA)
Common ancestor of all modern humans	1	541.35	542.60	43.06	0.40
Human dispersal out of Africa	2	540.68	542.38	19.23	0.18
Common ancestor of all non-Africans	3	539.78	542.57	19.65	0.18
Americans split from East Asians	4	543.26	541.00	7.87	0.07
	6	543.11	545.35	10.93	0.10
	8	541.14	543.00	9.27	0.09
	13	539.35	537.00	11.64	0.11
Common ancestor of Americans	5	543.23	544.00	6.91	0.06
	9	541.13	542.00	4.29	0.04
	14	540.00	539.06	5.69	0.05
Common ancestor of East Asians	7	537.85	539.00	10.07	0.09
	11	540.81	542.13	11.83	0.11
East Asians split from West and South Asians	10	541.31	540.33	13.53	0.13
Europeans split from West and South Asians	12	542.58	537.22	15.70	0.15
Common ancestor of Europeans	15	535.56	536.53	4.44	0.04
	16	536.23	539.50	7.24	0.07

N_{AO}: Mean number of SNPs differences between branch A and Outgroup (chimpanzee).

N_{BO}: Mean number of SNPs differences between branch B and Outgroup (chimpanzee).

N_{AB}: Mean number of SNPs differences between branch A and branch B.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438302; this version posted April 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

388 **Table 2: A proposed migratory time for prehistoric human migration**
389 **across the world based on evidence from Y-DNA**

390

Events	Node	N_{AO}	N_{BO}	N_{AB}	Corresponding time (MYA)
Common ancestor of all modern humans	1	3000.50	2979.11	2197.19	3.67
Human dispersal out of Africa	2	2979.27	2954.75	1216.75	2.05
Common ancestor of all non-Africans	3	2972.42	2996.88	918.40	1.54
East Asians split from West and South Asians	4	2970.18	2951.00	575.35	0.97
Common ancestor of East Asians	5	2969.86	2972.22	538.49	0.91
Americans split from West and South Asians	6	2951.19	2958.00	255.26	0.43
Common ancestor of Americans	7	2950.74	2959.00	185.65	0.31
Europeans split from West and South Asians	8	2965.26	2984.72	391.14	0.66

391

392 N_{AO}: Mean number of SNPs differences between branch A and Outgroup
393 (chimpanzee).

394 N_{BO}: Mean number of SNPs differences between branch B and Outgroup
395 (chimpanzee).

396 N_{AB}: Mean number of SNPs differences between branch A and branch B.