Ancient uniparental DNAs in distinguishing the competing theories of molecular evolution and modern human origins

Zhiyi Xia

Central South University

Hongyao Chen

Central South University

Ye Zhang

Central South University

Shi Huang (✉ shihuang1@gmail.com)

Central South University <https://orcid.org/0000-0003-2674-2830>

Research Article

Keywords: Y chromosomes, mtDNA, ancient DNAs, maximum genetic diversity, Out of East Asia, Out of Africa, infinite site model

Posted Date: November 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1087730/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Analyses of extant people have resulted in two models for the uniparental DNA phylogenetic trees of modern humans rooted in either Africa or East Asia. The Africa model is based on the neutral theory. The Asia model is reached from the maximum genetic diversity (MGD) theory. To test the two competing theories, we examined published data of ancient uniparental DNAs. Many ancient samples belonging to a terminal haplogroup were found to have mutated only in some, but not all, of the sites that define a more basal haplogroup. This pattern was found for the non-controversial haplogroups shared by the two competing models, and also for the haplogroups specific to the Asia model. Furthermore, many ancient samples that do not belong to some of the haplogroups of the Africa model nonetheless had mutations in them, which makes it impossible to unambiguously assign them to a haplogroup within the Africa model. Finally, uniparental DNAs of archaic humans were found to carry some modern alleles present in the first uniparental DNAs in the Asia model, indicating convergent evolution. Therefore, the data from ancient DNAs have verified the MGD theory and the actual existence of the haplogroups specific to the Asia model.

Introduction

The field of molecular evolution originally got started by the analyses of protein sequence alignments in the early 1960s. The most astonishing finding of those studies is the genetic equidistance phenomenon where, a simple species, e.g., fish, is approximately equidistant in amino acid identity to all the more complex species of land vertebrates such as reptiles, birds, pigs, and humans (Margoliash, 1963). This observation has been interpreted by the molecular clock hypothesis that all species have similar mutation rates (Kumar, 2005). The field then viewed this limited interpretation of an observed reality as the reality, and the neutral theory was next invented to try to explain the molecular clock “reality” (Kimura, 1968, 1969). The molecular clock and the neutral theory then served as the theoretical foundation for most of the phylogenetic results in the past 60 years, including the evolutionary trees of uniparental DNAs of modern humans with roots in Africa. A tacit assumption of this framework is that all observed genetic distances are still increasing with time and not yet at maximum saturation since mutations will always occur at new sites (not previously mutated) under this framework (infinite site model). Other assumptions include that most DNAs are selectively neutral and the number of neutral sites in a genome is infinite.

Despite the overwhelming amount of studies based on this neutral framework, most lack independent confirmations, and both self-inconsistencies and contradictions among different studies are abundant. After careful reviews of the vast literature in the phylogenetics field, many researchers have concluded that the neutral theory is an unsatisfying explanatory framework for evolutionary phenomena (Huang, 2016; Kern and Hahn, 2018; Leffler et al., 2012; Ohta and Gillespie, 1996). The molecular studies of human origins are all about interpreting the genetic diversity patterns of humans and yet the mystery of what determines genetic diversity remains to be fully solved (Huang, 2016; Leffler et al., 2012). The dogma that less conserved DNAs are less functional has been overturned (Kasinathan et al., 2020; Wang et al., 2020). There is a real possibility that both experimental and theoretical biology could in the near future demonstrate that most DNAs are not neutral (Dunham et al., 2012; Gates et al., 2021; Huang, 2016; Pouyet et al., 2018; Quinodoz et al., 2021; Tsuzuki et al., 2020). That would overthrow most of the results in the field, which might seem hard to believe, but is really not as most results are far from proven beyond any doubt or free from uncertain assumptions.

A competing alternative framework of molecular evolution, the maximum genetic diversity (MGD) theory published in 2008, was inspired by an independent rediscovery of the same genetic equidistance phenomenon (Huang, 2009, 2016). It turns out that the equidistance phenomenon is in fact a result of maximum saturation that is completely independent of time and mutation rate but is directly related to organismal complexity, and so the molecular clock (and in turn the neutral theory) interpretation of it could not be further from the truth (Hu et al., 2013; Huang, 2008, 2012). Complex species have a lower fraction of sites that can freely tolerate substitutions, and so the maximum dissimilarity between a complex species and a simpler one is mostly defined by the higher fraction of sites in the simpler species that can freely tolerate substitutions. The distinguishing feature of maximum distance is the enrichment of recurrent mutation sites, where independent mutations occur at the same sites in different lineages but lead to different amino acids (Huang, 2010; Wang et al., 2020). The MGD theory accepts the proven virtues of the neutral theory as a good description of truly neutral variants still at the linear phase of sequence divergence prior to maximum saturation, but differs from it in several fundamental aspects. 1. Most observed genetic distances or diversities are at maximum saturation. 2. Convergent mutations, back mutations, and recurrent mutation sites are common. 3. Most DNAs are functional or under selection. These notions of MGD are fully supported by known data (Huang, 2009, 2016). The differences between the two theories guarantee that

phylogenetic inferences based on the MGD theory would be vastly different from those based on the neutral theory. Indeed, with regard just to humans, the MGD theory has produced two very dramatic results. First, despite that chimpanzee is the closest to humans in raw DNA sequence similarity, it is only among the closest to humans in phylogeny (Huang, 2012). All great apes within the pongid clade are equally related to humans in phylogeny as commonly accepted before the molecular era. The closer similarity in DNA of chimpanzees to humans in 70% of the genome, compared to other great apes, comes from convergent evolution. So does the closer similarity of gorillas to humans in 30% of the genome or the closer similarity of orangutans to humans in 1% of the genome, which has been explained away by the unprovable idea of incomplete lineage sorting (Hobolth et al., 2011; Scally et al., 2012). Secondly, different major human populations have independently evolved for closer to 2 million years, and the roots of the uniparental DNAs of modern humans are in East Asia rather than in Africa (Yuan et al., 2017).

Two competing models of modern human origins termed “Multiregional” and “Recent Out-of-Africa” (OoA) have long been proposed (Brauer, 1982; Henn et al., 2018; Scerri et al., 2018; Stringer and Andrews, 1988; Wolpoff et al., 1984; Wu, 2004). The multiregional model considers extant people of any given region, e.g., East Asia, to be largely descended from ancient people living in the same region at ~200-2000 ky ago, such as Peking man, with some gene flow having occurred between populations in different regions (Wu, 2004). The model has support from fossils and cultural remains but molecular evidence has been lacking until recently (Yuan et al., 2017). Analyses based on the MGD theory suggest multiregional origins for autosomes but root both uniparental DNAs in East Asia (Yuan et al., 2017). The rooting of mtDNA tree in Asia independently confirms an earlier paper (Johnson et al., 1983), and has been verified by findings from ancient mtDNAs that show the earlier appearance of haplogroup R compared to N (Zhang and Huang, 2019a, b). It has also been found that all non-African Y haplogroups originated in southern East Asia, contradicting the serial founder model of OoA (Hallast et al., 2021).

The OoA model posits that modern humans originated in Africa and then migrated to Eurasia, largely replacing local archaic humans with limited genetic mixing (Brauer, 1982; Cann et al., 1987; Green et al., 2010; Henn et al., 2018; Scerri et al., 2018; Stringer and Andrews, 1988). The rooting of uniparental DNAs in Africa relies on the assumption of neutral mutations throughout the entire genome (Cann et al., 1987; Ke et al., 2001; Underhill et al., 2000). The infinite site assumption emerges from the neutral framework, which states that mutations appear once in the evolutionary history, and the related inference of derived alleles underlies the tree topology of the uniparental DNAs and the rooting of the Africa model (Ingman et al., 2000; Underhill et al., 2000). However, mutation saturation and natural selection are far more common than initially thought, which would invalidate the currently accepted inference of derived alleles (Chen et al., 2020; Huang, 2016; Lei et al., 2018; Teitz et al., 2018; Yuan et al., 2017; Zhu et al., 2015). The OoA model of Y lacks self-consistency as many haplogroups contain derived alleles that would define other haplogroups, which violates the method underlying the Africa model in the first place. There are numerous incompatible alleles in the Y tree of OoA that show recurrent mutations, violating the infinite site model (Poznik et al., 2016). The mtDNA tree of OoA also fails the self-consistency test and contains a large number of back mutations that violate the infinite site assumption required to build the OoA tree in the first place. These mutations can be found in the official mtDNA tree, phylotree (<http://www.phylotree.org/>), and are designated by an exclamation mark (!) in phylotree. The total number of these back mutations in phylotree is 1180 per our manual counting. Models are simplifications that are still consistent with reality. It is unacceptable to assume the infinite site model (no recurrent mutations) when recurrent mutations are common. The more correct or realistic model is that recurrent mutations are common, which is the position of the MGD theory.

In addition to being derived from a different evolutionary theory, the Out of East Asia (OoEA) model of uniparental DNAs differs from the OoA model in several fundamental ways (Figure 1 and Figure 2). First, haplogroups in OoEA are defined by alleles shared by all members within a haplogroup, regardless of their derived status. Secondly, mutations that are shared by the terminal branches and form the basal branches come from both common ancestry and convergent mutations in OoEA but only from common ancestry in OoA (Figure 3).

It has been shown that different human populations share more alleles in fast evolving SNPs

than in slower evolving ones, which indicates convergent mutations in fast evolving DNAs (Yuan et al., 2017). Finally, the rooting in OoEA relies on the reasoning that the original haplotype should be the common type shared by most individuals, since mutations leading to alternative types should be rare events (Johnson et al., 1983; Yuan et al., 2017; Zhang and Huang, 2019a, b). This reasoning has been independently arrived at twice (Johnson et al., 1983; Yuan et al., 2017). The ancestor type should have many alleles different from the outgroup to qualify as a modern type. Thus, alleles representing the early basal branches in OoA are all

already present in the first uniparental ancestor in OoEA (Figure 4). Some of those alleles may revert back to archaic alleles to give rise to new haplogroups as modern humans migrated to new environments and admixed with archaic humans (Figure 4). Co-evolution alongside admixed autosomes may cause certain sites in modern uniparental DNAs to mutate back to archaic alleles, or cause certain sites in archaic uniparental DNAs to mutate to modern alleles. However, the rooting in OoA relies on similarity to outgroups and assuming constant mutation rates. Populations with the greatest genetic diversity are viewed as the earliest evolving. Also, among polymorphic sites, populations that have more shared alleles with the outgroups are considered as the closest to the root (Ingman et al., 2000; Underhill et al., 2000). Both of these reasonings are invalid. Genetic diversity today is at maximum saturation and is not related to time of evolution. Greater genetic diversity or long branch length such as Y haplogroup L00 and mtDNA haplogroup L0a in certain populations like the San people in Africa maybe a result of natural selection related to their hunter-gatherer lifestyle that selects for greater immunity in response to a more primitive lifestyle (Nemat-Gorgani et al., 2018). The reality of convergent mutations or back mutations suggests that one cannot use sharing of ancestral alleles with the outgroups among polymorphic sites of humans to define the oldest branch in a phylogenetic tree.

Phylogenetic trees are largely a speculation about history based on genetic variation patterns of present day samples. Nothing could be more powerful than the ancient DNAs in testing if these speculations are true or not. In particular, ancient DNAs could be used to test whether mutations in terminal branches do not occur until the basal branches have fully formed as expected from the neutral theory and OoA. However, OoEA and MGD expect some fractions of mutations in the shared basal branches to occur only after the terminal branches have already formed, which are due to convergent mutations in different terminal branches. Therefore, ancient samples belonging to a terminal haplogroup are expected to have mutated only in some, but not all, of the sites that define a more basal haplogroup (Figure 3). Furthermore, if this pattern is indeed the case, it can be further used to distinguish a true haplogroup from a fake one and hence a true tree from a fake tree: a fake basal haplogroup should not show such a partial mutation pattern or should be found fully mutated in ancient samples. For example, Y haplogroup B belongs to either AB basal haplogroup in OoEA or BT basal haplogroup in OoA. If ancient B samples show partial mutations in AB defining sites but always complete mutations in BT defining sites, then AB is real and BT is fake. As can be seen from the OoEA tree, the first original modern Y already carried BT alleles in all BT defining sites and later mutations in these sites would lead to formation of the basal haplogroup A00A1b in OoEA (Figure 4A). BT alleles are already fully formed in the first modern Y chromosome and are expected to be fully present in all B samples regardless if they are ancient or present day samples.

The biggest difference between the two competing models of uniparental DNA trees comes from how they handle the basal mega-haplogroups in the Africa model upstream of G in the Y tree (Figure 1) and upstream of L3 in the mtDNA tree (Figure 2). These mega-haplogroups are thought to result from new mutations since the original type in the Africa model. In contrast, the Asia model considers the alleles of mega-haplogroups like CT of Y and R of mtDNA to be the ancestor type carried by the first modern human individual, and it is the non-CT haplotypes such as AB or non-R haplotypes such as NML that have acquired new mutations. We here used ancient DNAs to test these two competing models of modern human origins. The results strongly verified the Asia model and invalidated the Africa model.

Materials And Methods

Ancient samples were selected for analyses here for no particular reasons other than age, sequence depth, and availability known to the authors. Whole genome sequencing data or in a few cases targeted capture array sequencing data of ancient DNAs in BAM files or fastq files or vcf files were downloaded from links provided by the previous publications (see Supplementary Table S1 for a list of the ancient samples and their references). Targeted capture array sequencing data were less informative as they only cover a limited range of Y chromosome sites but would not bias the results to favor one of the two competing origin models or would cover more informative sites in one model versus the other. Most of these ancient DNA sequences were available for download as Bam files. Some were in fastq format and BWA 0.7.10 was used to align the fastq format to the human genome GRCh37 using standard parameters (Li and Durbin, 2009). The mapped reads were then filtered and sorted using SAMtools 0.1.19 and Picardtools 1.107 (<http://picard.sourceforge.net>) (Li, 2011; Li et al., 2009). The Picardtools 1.107 was also used to remove duplicates. The genotypes of ancient DNAs were called under haploid setting using GATK-3.2-2 or Samtools software with mapping quality setting at Q=30 and base quality at 30 (Li et al., 2009; McKenna et al., 2010).

Haplogroups defining SNPs were identified using the ISOGG Y chromosome tree version 15.73 updated on August 5, 2020 (<https://isogg.org>) and the 1kGP dataset (<http://www.internationalgenome.org>). Briefly, we downloaded the ISOGG Y chromosome SNP index file and cleaned it up by removing duplicated positions. We also downloaded the Y chromosome SNPs data of the 1kGP, which had 60353 SNPs (Poznik et al., 2016). We merged the ISOGG sites with the 1kGP sites to produce haplotype assignments for SNPs in the 1kGP data. ISOGG only has SNP sites whose derived alleles (different from chimpanzee alleles) can define a haplotype. In the Out of East Asia tree, haplotypes are defined by shared alleles regardless if they are derived or not. Thus, many SNPs that define a haplotype in the OoEA model are ancestral and would not be found in the ISOGG list of SNPs. To find these, we made use of the 1kGP dataset that had the haplotype information of a SNP assigned based on the ISOGG data. 20257 of the 60353 SNPs in 1kGP data were assigned a haplotype in this way. We next counted the number of minor alleles of each SNPs in the 1kGP dataset among the 1233 male samples in 1kGP. If a SNP has the minor allele as ancestral allele that is present in a group of individuals that share a haplotype as defined by the derived alleles of some SNPs present in ISOGG, it was assigned either that same haplotype or a haplotype specific to the Asia model. The OoEA-specific haplogroups are defined by both the derived alleles of certain SNPs and the ancestral allele of some other SNPs whose derived alleles define a different haplotype specific to the Africa model. For example, 62 SNPs in 1kGP were assigned as haplotype F as defined by ISOGG, and their derived alleles are present in 849 samples belonging to F while their minor alleles are present in 384 samples that had A, B, C, D, or E haplotypes. In the OoEA model, the minor alleles of these SNPs would define the ABCDE mega-haplogroup. We found 93 SNPs that are not present in ISOGG as F defining but have minor alleles present in the same set of 384 samples. These 93 SNPs were thus assigned to the ABCDE mega-haplogroup, thereby bringing the total number of ABCDE-defining SNPs to 155 (62+93). In total, 21659 SNPs in the 1kGP dataset were assigned a haplotype name as found in the Asia model.

Chi-squared test in GraphPad Prism 6 was used to test if the number of unexpected mutation or absence of expected mutation in a mega-haplogroup is significantly different from the background error rate in calling the genotypes. The background error rate of calling was determined by calling the sites defining all the haplogroups from L to T (the fraction of unexpected calls or mutations among all informative sites called). If a sample belonging to a particular haplotype is found to have the defining alleles or mutations in an unrelated haplotype, such alleles/mutations were deemed unexpected. If the sample has the non-defining alleles in an unrelated haplotype, such alleles are deemed expected. If the sample has the defining alleles in the haplotype to which it belongs, the alleles are deemed expected. If on the other hand it has the non-defining alleles in the haplotype to which it belongs, the alleles are deemed un-expected. To select L to T haplogroups as controls is because most of the informative cases in this study involved the early basal haplogroups A to F.

Results

Partial mutations in both terminal and basal haplogroups in ancient Y chromosomes

To study whether ancient samples of modern humans should have mutated in only a fraction of the sites that define the present day haplogroups to which they belong, regardless whether they are the internal or terminal branches, we studied a total of 111 published ancient Y chromosome sequences of relatively high coverage (list of samples and references see Supplementary Table S1) for informative sites (60353 SNPs) as found in the 1000 genomes project (1kGP) or in the Y-DNA haplogroup tree from the International Society of Genetic Genealogy (ISOGG, <http://www.isogg.org>, version 15.73) (Poznik et al., 2016). These samples were selected for no particular reasons other than age (1000-45000 years ago), sequence depths, and availability known to the authors. For all of these samples we called genotypes using downloaded BAM files or BAM files made from downloaded fastq files.

We first examined the non-controversial haplogroups that are shared by both the Africa and Asia models. We looked for ancient DNAs that showed mutations in some but not all of the sites that define a basal haplotype to which they belong. Of the 111 ancient samples, 32 were found to have partial mutations in a non-controversial basal haplogroup even though they all had mutations in a more terminal haplogroup (Table 1). The basal haplogroups showing partial mutations included B2, C, DE, D, E, G, IJ, I, J, NO, Q1b, and P. For example, I10873 had 57 of 59 informative sites mutated for the basal haplogroup B2-M182 and 2 out of 2 informative sites mutated for the more terminal branch B2b1, thus showing mutation in terminal haplogroup B2b1 prior to completion of mutation in the basal haplogroup B2. To exclude the possibility of sequencing or calling errors, we analyzed all samples for mutations in sites that define the haplogroups from L to T (L-T). I10873 showed 0 unexpected mutations in 2027 informative sites

defining L-T. Therefore, the 2 sites among the 59 informative sites for B2 that were found to be non-mutated is highly unlikely due to sequencing or calling errors ($P < 0.01$, 2/59 vs 0/2027). The results showed that ancient samples of a particular terminal haplogroup often had only partial mutations in the basal haplogroups to which they belong that are non-controversial or present in both the Africa and the Asia model.

Table 1

Mutation patterns of ancient Y chromosomes in non-controversial haplogroups. Haplogroup-defining SNPs were identified using ISOGG and the 1kGP dataset. The total combined number of sites from ISOGG and 1kGP was used in the analysis. The number of informative sites and the number of mutated sites are shown for the non-controversial haplogroups. * $P < 0.01$, chi-squared test, comparing the fraction of non-mutated sites in a haplogroup and the fraction of unexpected mutations in haplogroups L-T. If a sample belonged to an haplogroup within L-T, such as O, Q, or R, the counting of sites within L-T had excluded the relevant haplogroups to which the sample belong, as shown.

Sample	Year BP	HG	Unexp. L-T	HG partial mutation spectrum
I10873	3160	B2b1	0/2027	B 9/9, B2 57/59*, B2b1 2/2
BB7-240	45000	C1	9/1710	C 28/34*, C1 1/1
Kostenski14	36000	C1	0/3546	C 43/44*, C1 2/2
Schw72	7000	C1	7/794	C 14/16*, C1 1/1
I4089	5700	C1a	4/1717	C 34/35*, C1a 1/1
Vestonice16	30000	C1a2	0/1322	C 26/27*, C1 1/1, C1a 1/1, C1a2 1/14*
UKY001	14500	C2	7/1613	C 18/21*, C2 14/19*
AR14.5K	14500	C2	0/1020	C 19/20*, C2 12/17*
I13698	7350	C2	0/2498	C 61/62*, C2 43/49*
GLZ002	4500	C2	1/1389	C 18/20*, C2 12/15*
AR7K	7000	C2a	1/1041	C 23/23, C2 5/11*, C2a 1/1
S13886	4000	D1a2a3a	0/1647	D 5/6*, D1 22/22, D1a2a3a 7/7
I1985	3000	E1b1b1	1/2004	DE 10/10, E 33/34*, E1b1b1 43/43
I6900	3000	E1b1b1	2/1512	E 21/22*, E1b1b 4/5*, E1b1b1 24/24
LUK001	3900	E1b1b1	0/713	DE 1/2*, E 9/10*, E1b1b1 3/3
I3966	3800	E1b1b1	4/1398	DE 4/4, E 19/20*, E1b1b1 23/23
RISE423	3300	E1b1b1	7/996	DE 3/3, E 8/9*, E1b1b1 8/9*
TAF014	15000	E1b1b1a1	5/2135	DE 7/8*, E 39/41*, E1b1b1a1 15/19*
TAF013	15000	E1b1b1a1	8/2149	DE 11/11, E 44/45*, E1b1b1a1 17/23*
TAF011	15000	E1b1b1a1	6/2404	E1b1b1 51/52*, E1b1b1a1 16/22*
TAF010	15000	E1b1b1a1	7/1581	DE 1/2*, E1b1 11/12*, E1b1b1a1 6/12*
ZKO	10050	G2a	1/1061	G 12/13*, G2a2b 1/1
Pta08	5000	G2a	11/5081	G 154/157*, G2a 17/17
Iceman	5300	G2a	14/4736	G 86/98*, G2 19/22*, G2a 10/10
I0449	7000	I2a1a	0/2374	IJ 15/16*, I 36/36, I2 12/12
I0412	6249	I2a1a	1/2861	IJ 15/16*, I 48/51*, I2 13/13
KK1	9900	J2a	4/5659	J 69/70*, J2a 55/55
SM-SGDLM6	4000	O2a2b1a1a	2/2747 xNO/O	NO 11/12*, O 15/15
		Q1b1a1a1	3/4273 xP/Q	Q1b 37/39*, Q1b1 29/30*
Sumidouro5	10000			
AHUR2064	10000	Q1b1a1a1	1/4391 xP/Q	P 120/121*, Q1b1 32/32

Sample	Year BP	HG	Unexp. L-T	HG partial mutation spectrum
HAL001	4832	R1a1a1b	0/1524 xP/Q/R	P 50/51*, R 25/25, R1 23/23, R1a 9/9
NAU001	4800	R1a1a1b	0/2308 xP/Q/R	P 86/87*, R 39/39, R1 37/37, R1a 12/12

We next examined the ancient samples for mutations in the basal mega-haplogroups that are differently classified by the two competing models. We reasoned that if a basal mega-haplogroup is real it should show partial mutations in the haplogroup-defining sites. If it is not real, it would carry the complete set of haplogroup defining alleles.

The Asia model differs from the Africa model in the mega-haplogroups ancestral to haplogroup F (Figure 1 and 2). To test which model is true, we examined 76 ancient DNAs of the haplotype A, B, C, D, E, or F. The results showed that 36 of these had partial mutation patterns for the basal branches specific to the Asia model such as A (A00-A1b), AB, ABDE, and ABCDE (Table 2). In contrast, all of these 36 ancient DNAs carried complete set of haplotype defining alleles for the basal haplogroups specific to the Africa model, including A1, A1b, BT, CT, and CF. All ancient samples showed very low rate of unexpected mutations in haplogroups L-T, and so the observed absence of mutation in certain sites in the basal haplogroups in the Asia model is unlikely to be miscalls ($P < 0.01$, Table 2). For example, I10871 sample had haplogroup A00 and the following mutation patterns in the mega-haplogroups of the Asia model, 5 mutated among 5 informative sites in A00A0, 8 mutated among 10 sites in A00A1a indicating partial mutations, 146 among 146 in A00A1b, 98 among 100 in AB indicating partial mutations, 3 among 3 in ABDE, 65 among 67 in ABCDE indicating partial mutations. So, this sample showed partial mutations in A00A1a, AB, and ABCDE, and thus has served to validate the actual existence of these mega-haplogroups. It had very low rate of unexpected mutations (3/2705) in haplogroups L-T that were likely due to sequencing or calling errors, which was significantly lower than the rate of non-mutated sites in these mega-haplogroups ($P < 0.01$). Thus, the partial mutation patterns observed here were unlikely due to sequencing or calling errors. A partial list of sites that were non-mutated in certain ancient samples is presented in Table 3. To further confirm that these sites were not called by errors, we verified that the reads in the BAM files covering these sites were indeed aligned properly (representative screen shots for sample I10871 are shown in Figure 5).

Table 2

Mutation patterns of ancient Y chromosomes in the controversial haplogroups. For numbers in the cells, the first refers to the number of mutations in the OoEA-specific haplogroup, the second refers to the number of mutations in the OoA-specific haplogroup, and the third represents the number of informative sites. Numbers in bold highlight cases where only a fraction of the sites defining a basal haplogroup had been mutated. * $P < 0.01$, chi-squared test, comparing the fraction of non-mutated sites in a OoEA-specific haplogroup and the fraction of unexpected mutations in haplogroups L-T.

Sample	Year BP	OoEA	A00A0	A00A1a	A00A1b	AB	ABDE	ABCDE	Unexp. mut L-T
		OoA	A1	A1b	BT	CT	CF	F	
I10871	7970	A00	5/0/5	8/2/10*	146/0/146	98/2/100*	3/0/3	65/2/67*	3/2705
BallitoBayA	1986	A1b1b2	0/6/6	0/22/22	72/5/77*	48/1/49*	1/0/1	25/0/25	6/3123
BallitoBayB	2149	A1b1b2	0/9/9	0/16/16	46/1/47*	27/1/28*	0/0/0	18/1/19*	2/2700
I9133	2000	A1b1b2a	0/6/6	0/19/19	69/0/69	53/1/54*	1/0/1	38/1/39*	4/3206
F6-620	43000	ABCDE	0/11/11	0/15/15	0/222/222	0/149/149	0/3/3	2/102/104*	3/2974
I2966	8100	B2b1	0/3/3	0/5/5	0/88/88	7/4/11*	1/0/1	8/0/8	0/868
ZMOJ	10050	C	0/1/1	0/3/3	0/80/80	0/59/59	0/0/0	4/2/6*	1/894
La368	7000	C	0/3/3	0/12/12	0/78/78	0/60/60	0/1/1	3/3/6*	0/1447
BB7-240	45000	C1	0/5/5	0/7/7	0/152/152	0/92/92	0/3/3	22/4/26*	12/1710
CC7-335	46000	C1	0/5/5	0/7/7	0/136/136	0/89/89	0/2/2	16/4/20*	16/1541
I2783	6000	C1	0/4/4	0/6/6	0/118/118	0/66/66	0/2/2	10/1/11*	1/1129
I4089	5700	C1a	0/4/4	0/8/8	0/144/144	0/102/102	0/3/3	25/1/26*	4/1717
Sunghir IV	36000	C1a2	0/11/11	0/35/35	0/264/264	0/187/187	0/4/4	83/1/84*	1/4973
Vestonice16		C1a2	0/4/4	0/8/8	0/105/105	0/73/73	0/2/2	20/1/21*	
	30000								0/1322
NE5	7000	C1a2	0/4/4	0/9/9	0/122/122	0/70/70	0/2/2	10/2/12*	2/1986
Kostenski14	36000	C1b	0/9/9	0/23/23	0/191/191	0/139/139	0/3/3	31/3/34*	0/3546
UKY001	14500	C2	0/6/6	0/7/7	0/135/135	0/89/89	0/3/3	14/6/20*	7/1613
DA357	6700	C2	0/2/2	0/6/6	0/63/63	0/39/39	0/0/0	1/1/2*	0/1086
I13698	7350	C2	0/6/6	0/10/10	0/195/195	0/126/126	0/2/2	44/1/45*	0/2498
S11696	7580	C2	0/6/6	0/7/7	0/164/164	0/118/118	0/2/2	41/1/42*	0/2143
S12977	4845	C2	0/4/4	0/10/10	0/167/167	0/106/106	0/2/2	36/1/37*	1/2099
GLZ001	4500	C2	0/1/1	0/4/4	0/114/114	0/75/75	0/1/1	9/1/10*	1/1225
AR19K	19000	C2	0/6/6	0/6/6	0/120/120	0/78/78	0/2/2	21/1/22*	0/1351
AR14.1K	14100	C2	0/4/4	0/2/2	0/82/82	0/62/62	0/2/2	16/1/17*	1/910
I12644	1000	E1	0/2/2	0/4/4	0/89/89	0/54/54	0/0/0	5/2/7*	1/881
ART015	5300	E1b	0/8/8	0/12/12	0/200/200	0/142/142	2/0/2	58/1/59*	2/2523
I3948	7860	E1b	0/2/2	0/8/8	0/140/140	0/90/90	1/0/1	31/1/32*	0/1687
Mota	4500	E1b1a1	0/11/11	0/38/38	0/301/301	0/211/211	4/0/4	120/2/122*	0/5588
ETM010	4650	E1b1b1	0/4/4	0/7/7	0/128/128	0/99/99	1/0/1	15/2/17*	12/1479

Sample	Year BP	OoEA	A00A0	A00A1a	A00A1b	AB	ABDE	ABCDE	Unexp. mut L-T
		OoA	A1	A1b	BT	CT	CF	F	
I1799	3000	E1b1b1	0/7/7	0/7/7	0/151/151	0/108/108	2/0/2	34/1/35*	1/1756
I4246	3910	E1b1b1a	0/5/5	0/4/4	0/107/107	0/67/67	2/0/2	15/1/16*	1/1214
TAF013	15000	E1b1b1a1	0/9/9	0/11/11	0/181/181	0/126/126	3/0/3	44/2/46*	8/2149
TAF011	15000	E1b1b1a1	0/9/9	0/14/14	0/206/206	0/140/140	2/0/2	51/2/53*	6/2404
TAF010	15000	E1b1b1a1	0/7/7	0/9/9	0/142/142	0/90/90	2/0/2	22/2/24*	7/1581
I1072	13000	E1b1b1b2	0/1/1	0/1/1	0/57/57	0/41/41	0/0/0	3/1/4*	1/686
IAM5	7100	E1b1b1	0/4/4	0/7/7	0/110/110	0/70/70	1/0/1	8/1/9*	0/1723

The sample F6-620 from the 43000 year old Barcho Karo Cave had 2 mutations out of 104 informative sites that define ABCDE and would be classified as partial ABCDE according to the Asia model. It could also be assigned as partial F (102 mutations out of 104 informative sites that define F) according to the Africa model. Given that the two other related samples from the same cave, BB7-240 and CC7-335, all belonged to C and ABCDE (Hajdinjak et al., 2021), it is more likely for F6-620 to be ABCDE rather than F.

We also examined 35 ancient DNAs carrying haplogroups that were under F, including G, H, I, K, P, Q, and R (see Supplementary Table S1). Only one sample, Iceman, was found to have partial mutations in the A1b haplogroup of the Africa model (32 mutations out of 33 informative sites). None of the other samples showed partial mutations for the mega-haplogroups specific to the Africa model to which they are supposed to belong, such as A1, A1b, BT, CT, CF, and F. Thus, the finding of few occurrence (1 among 35 samples examined) of partial mutation patterns in the OoA-specific mega-haplogroups upstream of G among ancient samples under F is significantly different from the common occurrence (36 among 76 samples examined) of partial mutations in ancient DNAs belonging to A to E (1/35 vs 36/76, $P < 0.01$). This cannot be explained by miscalls as that would require the miscalls to cluster in the A to E samples, which is unreasonable.

All together, we found that 59 out of 111 ancient samples examined (Table 1 and Table 2) showed partial mutations in a basal haplogroup despite carrying alleles defining a more terminal haplogroup. This is consistent with expectations based on the MGD theory or convergent mutations (Figure 3).

Table 3

List of sites not yet mutated in the OoEA-specific megahaplogroups but mutated in the wrong OoA-specific megahaplogroups. Bold bases indicate the alleles found in the ancient DNAs. Derived allele data is from ISOGG. n.a., not available.

Sample ID_hg	HG19 position	OoEA	OoA	OoEA	OoA	OoEA	OoA	OoEA	OoA	Derived	Cover.
		A00-A1a	A1b	A00-A1b	BT	AB	CT	ABCDE	F		
I10871_A00	19424283	T	G							G->T	8
	6826955	C	T							n.a.	7
	23030717					C	T			n.a.	1
	16779295					C	A			n.a.	2
	14149772							C	T	C->T	1
	16467111							C	T	n.a.	1
Ballito Bay A_A1a1b2	14123744			C	T					C->T	1
	16538095			C	T					C->T	1
	17966904			C	A					C->A	1
	18069115			G	A					G->A	1
	21248962			G	A					G->A	1
	9035554					C	T			n.a.	1
	7671530							G	A	G->A	4
Ballito Bay B_A1a1b2	15095089			C	T					C->T	1
	17630903					C	T			C->T	1
	17493513							C	T	C->T	1
I9133_A1a1b2a	7153375					C	T			C->T	1
	8572150							G	A	G->A	1
I2966_B2b	14074463					C	T			C->T	1
	17186912					C	T			C->T	1
	21267676					C	T			C->T	1
	8572150							G	A	G->A	2
Kostenski 14_C	21618856							C	T	C->T	1
	22108101							G	A	G->A	1
	23600700							C	T	n.a.	1
Sunghir IV_C	19375658							C	G	n.a.	1
Vestonice 16_C	7218079							G	A	n.a.	1
Mota_E	8553421							C	T	n.a.	1
	15553123							C	T	n.a.	9

Further invalidating the mega-haplogroups specific to the Africa model, many ancient DNAs showed unexpected mutations in certain OoA-specific mega-haplogroups to which they do not belong (Table 2). Partial mutation in the OoEA-specific megahaplogroups automatically means unexpected mutations in the wrong haplogroups specific to the OoA model. For example, I10871 sample had A00 haplogroup but was found to have mutations that would group it to certain OoA-specific mega-haplogroups to

which A00 does not belong, including 2 mutations that would group it with A1b, 2 mutations that would group it with CT, and 2 mutations that would group it with F (Table 2 and Figure 5). So, partial mutation of I10871-A00 in the OoEA-specific mega-haplogroup AB (98 mutation among 100 sites) automatically means unexpected mutations in the wrong OoA-specific mega-haplogroup CT (2 mutation among 100 sites). The finding of ancient A00 samples carrying CT alleles is highly unusual and unexpected. Under the OoA model, it would mean that there were certain sites that have mutated more than once (once in A00 branch and once in CT branch), which would violate the infinite site assumption that makes the construction of the Africa model possible in the first place. Also, this kind of widespread unexpected mutations in the wrong basal haplogroups in ancient DNAs makes the unambiguous assignment of haplogroups impossible. In contrast, all these ancient DNAs can be cleanly assigned to a basal haplogroup specific to the OoEA model. Taken together, the data strongly validated the OoEA model and invalidated the OoA model.

Partial mutations in both terminal and basal haplogroups in ancient mtDNAs

We also studied ancient mtDNAs to similarly examine whether the partial mutation pattern could be found in real basal haplogroups common to both the Asia and Africa models, and could only be found in basal haplogroups specific to the Asia model but not the Africa model. We used published sequences for most of these ancient mtDNAs (downloaded from NCBI). We also downloaded BAM files from published accession numbers of certain ancient sequences and called mtDNA sequences using stringent filters. Among 48 ancient mtDNAs examined (list of samples and references see Supplementary Table S1), six were found to show partial mutations in basal haplogroups, either real ones common to both origin models or the OoEA-specific basal haplogroups (Table 4). None of the ancient DNAs showed partial mutations in basal haplogroups specific to the Africa model. For example, the I2966 sample had L0k2 but lacked some mutations in the real basal haplogroups common to both origin models, including the absence of 1 mutation in L0a'b'f'g'k and the absence of 2 mutations in L0k (Table 4). I2966 also lacked some mutations in the OoEA-specific basal haplogroups, including the absence of 1 mutation in NML, of 1 mutation in L5'1'0, and of 1 mutation in L1'0 (Table 4). In contrast, none of the numerous ancient samples of haplogroup U examined here (Supplementary Table S1) showed partial mutations in the OoA-specific basal haplogroups to which U belongs, including L1'2'3'4'5'6, L2'3'4'5'6, L2'3'4'6, L3'4'6, L3'4, L3, N, and R.

Furthermore, the absence of a mutation in a basal haplogroup in the OoEA model automatically means the presence of a mutation in a wrong basal haplogroup specific to the OoA model. For example, I2966-L0k2 lacked the mutation 16223T in the haplogroup NML but instead had the 16223A allele that would classify it to the haplogroup R (Table 4). Therefore, if the OoA model is true, it would be impossible to unambiguously assign certain ancient mtDNAs to a specific terminal haplogroup. Such difficulty however is non-existent in the OoEA model. Thus, the partial mutation patterns of ancient mtDNAs observed here strongly validated the mtDNA tree rooted in Asia and invalidated the one rooted in Africa.

Table 4

List of non-mutated sites in ancient mtDNAs in the non-controversial haplogroups and in the OoEA-specific mega-haplogroups. SNPs represent mutations in the OoEA model and mutations are given in the format [modern base] [position number][archaic base], e.g. "A189G". Modern alleles in these sites are those that were carried by the first original modern mtDNA in the OoEA model. Archaic alleles are those that were carried by the archaic humans prior to any admixture with modern humans or prior to any convergent evolutions (see Figure 4B).

SNP_OoEA	rCRS	Haplogroup		I13981	BBayB	I9133	I4422	I4421	I2966
				2700BP	2000BP	2000BP	5290BP	5400BP	8100BP
		OoA	OoEA	L0a	L0d2a1	L0d1b2b	L0k1	L0k2	L0k2
A189G	A	L0a'b'f'g'k	L0a'b'f'g'k				A		
T16172C	T	L0a'b'f'g'k	L0a'b'f'g'k						T
T14374C	T	L0k	L0k						T
A16166C	A	L0k	L0k						A
C16223T	C	R	NML						C
A4104G	A	L3'4'6	L2'5'1'0			A			
C16187T	C	L2'3'4'6'	L5'1'0		C				C
T2885C	T	L2-L6	L1'0						T
A13276G	A	L1-L6	L0	A					
A9347G	A	L0	L0					A	

Modern alleles in archaic uniparental DNAs

As modern humans migrated to new places and admixed with local archaic humans, coevolution alongside admixed autosomes may have caused certain sites in modern uniparental DNAs to mutate back to archaic alleles, or certain sites in archaic uniparental DNAs to mutate into modern alleles. Such coevolution is a fundamental part of the Asia model but is not acknowledged by the Africa model given its neutral and infinite site assumptions which rule out coevolution. We here examined four previously published (two Denisovans and two Neanderthals) Y chromosome high coverage sequencing datasets to see whether these archaic Y chromosomes may carry modern alleles in sites where they are not supposed to according to the Africa model (Petr et al., 2020).

Table 5

Modern alleles in archaic Y chromosomes. The total combined number of sites from ISOGG and 1kGP was used in the analysis. For numbers in the cells, the first refers to the number of mutations in the OoEA-specific haplogroup, the second refers to the number of mutations in the OoA-specific haplogroup, and the third represents the number of informative sites. Numbers in bold highlight cases where only a fraction of the sites defining a basal haplogroup had been mutated. * $P < 0.01$, chi-squared test, comparing the fraction of non-mutated sites in a OoEA-specific haplogroup and the fraction of unexpected mutations in haplogroups L-T. The number of modern alleles (alleles present in the first modern Y under OoEA model) among all informative sites are shown under modern allele column.

OoEA	A0000A000	A00	A00A0	A00A1a	A00A1b	AB	ABDE	ABCDE	Unexp. L-T
OoA	A00T	A0T	A1	A1b	BT	CT	CF	F	
Denisova4	14/5/19*	4/3/7*	3/0/3	5/3/8*	47/9/56*	28/2/30*	1/0/1	18/5/23*	6/2722
Denisova8	51/8/59*	10/5/15*	8/0/8	15/0/15	112/4/116*	79/1/80	1/0/1	50/3/53*	19/3394
Mezmaiskaya2	64/5/69*	12/3/15*	7/0/7	18/2/20*	149/0/149	98/0/98	4/0/4	73/0/73	9/3895
Spy 94a	35/0/35	2/0/2	2/0/2	3/1/4*	11/3/14*	12/1/13*	0/0/0	12/3/15*	10/1696

We merged the genotypes of the archaic Y chromosomes with the Y-DNA haplogroup tree from the International Society of Genetic Genealogy (ISOGG, <http://www.isogg.org>, version 15.73) and also the list of Y chromosome SNPs from the 1000 genomes project (Poznik et al., 2016). We counted the number of modern alleles among informative sites for the major mega-haplogroups of modern humans, as well as the unexpected mutations in L-T haplogroups (Table 5). Archaic Y chromosomes were found to carry a high fraction of modern alleles in sites that, according to the Asia model, differentiate the original modern Y (not yet affected by

admixture) from the original archaic Y (Table 5), including A0000A000 (or A00T in the Africa model. A0000 is Denisovan haplotype and A000 is Neanderthal haplotype), A00A1a (A1b), A00A1b(BT), AB (CT), and ABCDE (F). For example, for sites defining A00T, the Asia model has the original modern Y chromosome carrying A00T alleles while the original archaic Y carrying non A00T alleles or A0000A000 alleles (Figure 4A). Coevolution with admixed modern autosomes would result in mutation to modern alleles (A00T alleles) in individuals carrying the archaic Y chromosomes A0000A000 alleles. Thus, among A00T defining sites, the fraction of modern A00T alleles was 5/19, 8/59, and 5/69 for Denisova 4, Denisova 8, and Mezmaiskaya 2, respectively, which were all significantly higher than the fraction of unexpected mutations among sites defining L-T haplogroups ($P < 0.01$, Table 5).

Table 6

Modern alleles among archaic Y chromosomes in sites defining haplogroup A0-T. The sites listed here define A0-T in the OoA tree but A00 in the OoEA tree. Modern alleles in these sites are those that were carried by the first original modern Y chromosome in the OoEA model or those carried by haplogroups A0-T (A0 to T). Archaic alleles are those that were carried by the archaic humans prior to any admixture with modern humans or prior to any convergent evolutions (see Figure 4A). n.i., not informative.

SNP name	HG	HG19	Mutation	Denisova 4	Denisova 8	Mezmaisk. 2	Spy 94a
L1101	A0-T	6859819	T->C	archaic	archaic	archaic	n.i.
L1105	A0-T	7590048	C->T	archaic	archaic	archaic	n.i.
L1130	A0-T	16661010	T->G	archaic	archaic	archaic	n.i.
L1155	A0-T	22191266	G->C	n.i.	archaic	archaic	n.i.
L1121	A0-T	14496448	G->A	n.i.	n.i.	archaic	n.i.
L1135	A0-T	18147303	C->A	n.i.	archaic	archaic	n.i.
L1142	A0-T	21288562	C->T	archaic	archaic	n.i.	archaic
L1120	A0-T	14496439	G->T	modern	modern	archaic	n.i.
L1136	A0-T	18914286	A->G	modern	modern	modern	n.i.
L1143	A0-T	21593345	A->G	modern	modern	n.i.	n.i.
L1150	A0-T	21903971	A->G	n.i.	n.i.	archaic	n.i.
L1127	A0-T	16281810	A->G	n.i.	n.i.	archaic	archaic
L1098	A0-T	6847637	C->A	n.i.	archaic	archaic	n.i.
L1123	A0-T	15425676	C->T	n.i.	archaic	archaic	n.i.
L1125	A0-T	15467768	A->G	n.i.	archaic	archaic	n.i.
L1137	A0-T	19047091	C->T	n.i.	archaic	n.i.	n.i.
L1129	A0-T	16596846	T->C	n.i.	modern	modern	n.i.
L1132	A0-T	16718811	A->G	n.i.	modern	modern	n.i.

For sites defining A0T, the Asia model has the original modern Y chromosome carrying A0T alleles while the original archaic Y carrying A00 or non A0T alleles (Figure 4A). Coevolution with admixed modern autosomes would result in mutation to modern alleles (A0T alleles) in individuals carrying the archaic Y chromosomes. Indeed, among A0T defining sites, the fraction of modern A0T alleles was 3/7, 5/15, and 3/15, for Denisova 4, Denisova 8, and Mezmaiskaya 2, respectively (Table 5 and Table 6), which were all significantly higher than the fraction of unexpected mutations among sites defining L-T haplogroups ($P < 0.01$, Table 5).

For the most part, the mutated sites were shared among the four samples (Table 6), for example, the SNP L1136 was shared among Denisova 4, Denisova 8, and Mezmaiskaya 2 (not informative in Spy 94a). There were also mutations that differentiate the Denisovans from Neanderthals, for example, L1120 site was mutated to modern alleles in the two Denisovans but not in the Neanderthal Mezmaiskaya 2 (not informative in Spy 94a) (Table 6). This pattern is consistent with expectation based on shared natural selection or coevolution.

We next examined several archaic mtDNAs to examine the presence of modern alleles in archaic mtDNAs. As shown in Figure 4B, the original modern mtDNA in the OoEA model carried modern alleles in most of the major basal mega-haplogroups that are different from the alleles carried by archaic humans. But the original modern mtDNA in the OoA model carried archaic alleles in most of the major basal mega-haplogroups that are the same as the alleles carried by archaic humans. For example, sites defining ML haplogroup in OoEA had modern alleles in the original modern mtDNA that are different from archaic alleles carried by archaic humans, but the same sites define N haplogroup in OoA and have archaic alleles in the original modern mtDNA. Evolution from the first modern mtDNA into an African haplogroup such as L0 involved many back mutations to archaic alleles (Figure 4B). Under the OoEA model, archaic mtDNAs had many modern alleles, i.e., alleles carried by the first modern mtDNA (Table 7). This sharing of modern alleles could be due to convergent evolution, both dependent or independent of admixture between modern and archaic humans. For archaic humans contemporaneous with modern humans such as Neanderthals and Denisovans, convergent evolution could be related to admixture. In contrast, convergent evolution is forbidden in the assumptions required to build the OoA tree, and it is therefore difficult for the OoA model to account for the presence of modern alleles in archaic mtDNAs.

Table 7

Modern alleles in archaic mtDNAs. Sites defining the basal mega-haplogroups are listed here. SNPs represent mutations in the OoEA model and mutations are given in the format [modern base][position number][archaic base], e.g. "G263A". Modern alleles in these sites are those that were carried by the first original modern mtDNA in the OoEA model. Archaic alleles are those that were carried by the archaic humans prior to any admixture with modern humans or prior to any convergent evolutions (see Figure 4B). Archaic alleles also are shared between archaic humans and OoEA-specific haplogroups, e.g., 16223T is common to both Vindi.33.16 and NML haplogroup (formation of NML is a result of back mutation to archaic alleles).

SNPs	Mega-haplogroup		Vindi.33.16	Mezmais.1	Deniso. 8	DeniPinky
	OoA	OoEA	38000BP	50000BP	120000BP	50000BP
G263A	L0	L0	modern	modern	modern	modern
C1048T	L0	L0	modern	modern	archaic	archaic
C3516a	L0	L0	modern	modern	modern	modern
T5442C	L0	L0	modern	modern	modern	modern
T6185C	L0	L0	modern	modern	modern	modern
C9042T	L0	L0	modern	modern	modern	modern
A9347G	L0	L0	modern	modern	modern	modern
G10589A	L0	L0	modern	modern	modern	modern
G12007A	L0	L0	modern	modern	archaic	archaic
A12720G	L0	L0	modern	modern	modern	modern
T146C	L1-L6	L0	archaic	modern	archaic	modern
T182C	L1-L6	L0	modern	modern	modern	modern
C4312T	L1-L6	L0	archaic	archaic	archaic	archaic
C10664T	L1-L6	L0	archaic	archaic	modern	modern
T10915C	L1-L6	L0	archaic	archaic	archaic	archaic
G11914A	L1-L6	L0	archaic	archaic	archaic	archaic
A13276G	L1-L6	L0	archaic	archaic	archaic	archaic
A16230G	L1-L6	L0	archaic	archaic	modern	modern
T152C	L2-L6	L1'0	archaic	modern	archaic	archaic
G2758A	L2-L6	L1'0	modern	archaic	archaic	archaic
T2885C	L2-L6	L1'0	archaic	archaic	archaic	archaic
A7146G	L2-L6	L1'0	modern	archaic	archaic	archaic
C8468T	L2-L6	L1'0	archaic	archaic	modern	modern
T195C	L2'3'4'6'	L5'1'0	modern	modern	archaic	modern
G247A	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	modern
T825A	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	archaic
C8655T	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	archaic
G10688A	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	archaic
T10810C	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	archaic
A13105G	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	archaic

SNPs	Mega-haplogroup		Vindi.33.16	Mezmais.1	Deniso. 8	DeniPinky
	OoA	OoEA	38000BP	50000BP	120000BP	50000BP
C13506T	L2'3'4'6'	L5'1'0	modern	modern	archaic	archaic
A15301G!	L2'3'4'6'	L5'1'0	modern	modern	archaic	archaic
G16129A	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	archaic
C16187T	L2'3'4'6'	L5'1'0	modern	modern	modern	modern
T16189C	L2'3'4'6'	L5'1'0	archaic	archaic	archaic	modern
A4104G	L3'4'6	L2'5'1'0	archaic	archaic	archaic	archaic
G7521A	L3'4'6	L2'5'1'0	archaic	archaic	archaic	archaic
C182T	L3'4	L6'2'5'1'0	modern	modern	modern	modern
C3594T	L3'4	L6'2'5'1'0	archaic	archaic	archaic	archaic
C7256T	L3'4	L6'2'5'1'0	archaic	archaic	archaic	archaic
C13650T	L3'4	L6'2'5'1'0	modern	modern	archaic	archaic
C16278T	L3'4	L6'2'5'1'0	archaic	archaic	archaic	archaic
G769A	L3	L4'6'2'5'1'0	archaic	archaic	modern	modern
G1018A	L3	L4'6'2'5'1'0	archaic	archaic	archaic	archaic
T16311C	L3	L4'6'2'5'1'0	archaic	archaic	archaic	archaic
A8701G	N	ML	archaic	archaic	archaic	archaic
T9540C	N	ML	archaic	archaic	archaic	archaic
A10398G	N	ML	archaic	archaic	archaic	archaic
T10873C	N	ML	archaic	archaic	archaic	archaic
G15301A	N	ML	archaic	archaic	modern	modern
C12705T	R	NML	archaic	archaic	archaic	archaic
C16223T	R	NML	archaic	archaic	modern	modern

Discussion

The mutation pattern in the ancient Y chromosomes and mtDNAs as revealed here confirms the expectation from the MGD theory that ancient samples belonging to a terminal haplogroup should mutate in only a fraction of the sites that define a haplogroup they belong to, regardless if it is a terminal or basal haplogroup. Such a pattern is not expected from the popular neutral framework for phylogenetic inferences. Our study here has systematically examined a large number of ancient samples for both uniparental DNAs, and strongly established the partial mutation pattern as a real phenomenon of ancient uniparental DNAs.

Sequencing ancient DNAs are known to have more errors. Three observations in the case of Y chr indicate that such errors cannot explain our findings. First, as a negative control for background error rate in our calling method, we determined for each ancient Y sample the rate of calling an unexpected allele among all informative sites that define the haplogroups from L to T. We showed in all cases that this rate was significantly lower than the rate of calling the absence of mutations that establish the partial mutation pattern as reported here. Secondly, partial mutation patterns in the controversial Y haplogroups almost all happened to the ancient samples belonging to the non-F haplogroups. Nearly all randomly sampled ancient DNAs belonging to haplogroup F failed to show partial mutation patterns in the OoA-specific haplogroups (A0-T, A1, A1b, BT, CT, and CF), which is significantly different from the high incidence of partial mutations in the OoEA-specific haplogroups (ABCDE, ABDE, AB, A00A1b, A00A1a, A00A0) among randomly selected non-F samples (1/35 vs 36/76, $P < 0.01$). It is unreasonable for miscalls to cluster within non-F samples. Finally, among the

76 non-F samples, 75 (except I10871-A00) are informative for at least one basal haplogroup among the OoA-specific haplogroups in terms having partial mutations and yet all showed complete set of haplotype-defining alleles as fully expected by the OoEA model, which is inconsistent with a significant miscall rate. It is unlikely for miscalls to only happen to OoEA specific haplogroups for these 75 non-F samples (35 out of 75 samples had partial mutations in OoEA haplogroups) while not happen to any OoA specific haplogroups (0 out of 75 informative samples had partial mutations in OoA haplogroups).

Ancient DNAs are supposed to be more informative than extant DNAs with regard to past events and could thus serve as the best evidence to either verify or invalidate any phylogenetic trees that are built by using extant DNAs. It is therefore surprising that the field has yet to use the now abundant ancient DNAs to verify the standard model of modern human origins, the OoA model. Is it because the model cannot pass the ancient DNA test? Our study here represents the first such test. Several observations here strongly invalidated the OoA model and its associated neutral theory and supported the competing OoEA model and its associated MGD theory. First, the partial mutation pattern is common in the non-controversial haplogroups, either terminal or basal. For an ancient terminal haplogroup to have partial mutations in a basal haplogroup could only be allowed by the MGD theory but not by the neutral theory (Figure 3). The completion of mutations in all the sites defining a basal haplogroup in present day samples must entail convergent mutations from different terminal haplogroups. Just like a real tree, both the stem and the leaf branches grow together. Convergent mutations are to be expected in uniparental DNAs given their widespread presence in fast evolving autosomal DNAs that have reached mutation saturation (Huang, 2010; Wang et al., 2020; Yuan et al., 2017). Sharing of alleles among different human populations has been found to happen mostly in fast evolving SNPs but less common in slowly evolving ones, which is consistent with convergent evolution rather than admixture or recent common ancestry (Yuan et al., 2017). Convergent mutations have also been found to be very common in long term evolutionary experiments on microorganisms (Johnson et al., 2021; Katz et al., 2021). As mutations are stochastic, one does not expect all ancient samples to show partial mutation patterns in the basal haplogroups. Nonetheless, we observed a high rate of occurrence of partial mutations in the non-controversial Y chr basal haplogroups among the 111 ancient samples studied (32 out of 111, Table 1).

Secondly, the partial mutation pattern is commonly found also in the OoEA-specific mega-haplogroups but not found in the OoA-specific mega-haplogroups. For Y chr, for example, 35 out of 75 non-F samples had partial mutations in OoEA haplogroups but none of these had partial mutations in OoA specific haplogroups. Also, nearly none of the 35 F samples examined showed partial mutations in the OoA specific Y haplogroups. Such a pattern is fully expected if the OoEA model is true and if the OoA model is false. If the partial mutation pattern is a real phenomenon, which appears to be the case, it would apply to any haplogroup. Thus, the absence of that pattern for the OoA-specific haplogroups strongly suggests that the OoA model is unrealistic. Ancient DNAs are either non-informative in distinguishing the two competing models (have all the expected alleles from either models) or informative in falsifying the OoA model and validating the Asia model. We did not find any informative cases in the ancient DNA record that would support OoA.

Thirdly, our analyses showed that the absence of a mutation in a site defining an OoEA-specific haplogroup automatically means the presence of a mutation in a site that defines an OoA specific haplogroup to which however the concerned sample does not belong. A haplogroup, be it ancient or present, should not carry mutations that define haplogroups to which they do not belong. This expectation is only met by the Asia model but not the Africa model. For example, I10871-A00 Y chromosome carried alleles that would assign it to A1b, CT, and F of the Africa model, to which it does not belong. I2966-L0k2 mtDNA carried alleles that would assign it to L2'3'4'6' and L2-L6 of the Africa model, to which it does not belong. It is however not surprising for ancient samples to fail to be cleanly assigned to a haplogroup within the OoA model if the model is simply incorrect.

We discuss in more detail the implications of our finding using Mota-E1b1a1 Y chromosome as an example with regard to the ABCDE mega-haplogroup under the OoEA tree. Mota had partial mutations in the E1b1a1-M291 terminal haplogroup, mutating only 14 of 178 informative sites. Mota had two SNPs that would assign it to F. They are among a set of 155 SNPs that can divide all present day haplotypes into two mega-groups, F (F to T) and ABCDE (A to E). Mota was informative for 122 of these 155 SNPs. For the two alleles of each of these SNPs, A1 and A2, if one mega-group carries the A1 allele, the other mega-group would carry the A2 allele. Depending on assuming to which mega-group the first modern male belonged, one gets either the OoA model or the OoEA model. In the OoA tree, the first original modern male is supposed to carry the alleles shared by all present day ABCDE samples in these 155 SNPs (Figure 4A). In the subsequent diversification of haplotypes under OoA, these ABCDE-defining alleles were mutated to form the F haplogroup or become F-defining alleles. It is therefore clear that all samples of the ABCDE mega-group, regardless if

they are ancient or present, should carry the complete set of ABCDE-defining alleles in all these 155 SNPs, since the first modern male individual who was the ancestor to all ancient and present day male individuals was supposed to already have all of these alleles under the OoA model. Also, ancient F samples are expected to show partial mutations, having most of the ABCDE-defining alleles (but not all) changed into F-defining alleles. So, the finding of an ancient E (Mota) carrying two F-defining alleles would invalidate the OoA model or its assumption that the first modern male carried ABCDE-defining alleles in these 155 SNPs. It indicates that Mota-E1b was on its way to become F, mutating 2 of the 122 informative SNPs into F-defining alleles, which is absurd. If the Mota E lineage were to survive to today, it would have to have these two SNPs mutated back to ABCDE-defining alleles. But back mutations are forbidden by the assumption of infinite site (meaning no back mutations) that is required to build the OoA tree in the first place. Furthermore, all of the ancient F-samples (28 of 28) studied here carried complete set of F-defining alleles in these 155 SNPs, therefore showing no partial mutation patterns, which is unexpected if F was formed by mutating ABCDE alleles carried by the first Y. The two unexpected F-alleles in Mota cannot be explained by miscalls, since we found as a control that there was not a single unexpected call for Mota among 5588 informative sites that define L to T haplotypes (2/122 vs 0/5588, $P < 0.0001$).

In contrast, in the OoEA tree, the first modern male is supposed to carry the alleles shared by all present day F samples in these 155 SNPs (Figure 4A). In the subsequent diversification of haplotypes under the OoEA model, these F-defining alleles were mutated to form the ABCDE mega-haplogroup or become ABCDE-defining alleles. It therefore follows that all F samples, regardless if they are ancient or present, should carry the complete set of F-defining alleles in all these 155 SNPs, since the first modern male already had all of them under the OoEA tree. Indeed, we found that all ancient F (F to T) samples (28 of 28 studied) carried the complete set of F-defining alleles in all these 155 SNPs, which is in stark contrast to the above mentioned observation that many ancient A to E samples did not carry the complete set of ABCDE-defining alleles in these 155 SNPs. Many of these ABCDE samples (34 out of 76 studied, Table 2), including Mota, had partial mutations in these 155 SNPs, changing most but not all of the F-defining alleles into ABCDE-defining alleles, which is to be expected if the ABCDE mega-haplogroup is real and was formed by mutations from the first Y who carried the F haplotype or the complete set of F-defining alleles in these 155 SNPs.

Finally, we found that certain alleles in modern uniparental DNAs are also present in archaic humans. Sharing of alleles between archaic and modern humans in most cases is more likely a result of coevolution due to common adaptation to a shared environment or physiology, or natural selection. As convergent evolution is common under the MGD theory but forbidden by the neutral theory, the presence of modern alleles in archaic humans is evidence for the MGD theory. Only the OoEA model can explain such presence by postulating coevolution of admixed autosomes and uniparental DNAs (Figure 4).

We here used ancient DNAs to test the two different models of the human uniparental DNA phylogenetic trees and their respective underlying theory of evolution. The results confirm the actual existence of the haplogroups specific to the Asia model and support the MGD theory that underlies the model. As might be expected already from the failure of the OoA model to be self-consistent, ancient uniparental DNAs strongly invalidated the OoA model and the neutral theory underlying it. The great value of having two competing models is to make it unnecessary to make up ad hoc epicycles in order to make a model escape falsification by contradicting data. If one model is supported by new data while the other is contradicted by it, it is straightforward to recognize the correct model.

Declarations

Acknowledgments:

Supported by the National Natural Science Foundation of China grant 81171880 (S.H.).

Additional Information:

Competing Interests:

The authors declare that they have no competing interests.

Author contributions:

Z.X., H.C., and Y.Z. performed data analysis. S.H. devised the project, analyzed the data, and wrote the manuscript. All authors edited the manuscript.

References

- Brauer, G. (1982). Current problems and research on the origin of Homo sapiens in Africa. *Humanbiol Budapest* 9, 69–78.
- Cann, R.L., Stoneking, A.C., and Wilson, A.C. (1987). Mitochondrial DNA and human evolution. *Nature* 325, 31–36.
- Chen, H., Lei, X., Yuan, D., and Huang, S. (2020). The relationship between the minor allele content and Alzheimer's disease. *Genomics* 112, 2426–2432.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fietze, S., Harrow, J., Kaul, R., *et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Gates, A.J., Gysi, D.M., Kellis, M., and Barabasi, A.L. (2021). A wealth of discovery built on the Human Genome Project - by the numbers. *Nature* 590, 212–215.
- Green, R.E., Krause, J., *et al.* (2010). A draft sequence of the Neandertal Genome. *Science* 328, 710–722.
- Hajdinjak, M., Mafessoni, F., Skov, L., Vernot, B., Hubner, A., Fu, Q., Essel, E., Nagel, S., Nickel, B., Richter, J., *et al.* (2021). Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. *Nature* 592, 253–257.
- Hallast, P., Agdzhoyan, A., Balanovsky, O., Xue, Y., and Tyler-Smith, C. (2021). A Southeast Asian origin for present-day non-African human Y chromosomes. *Human genetics* 140, 299–307.
- Henn, B.M., Steele, T.E., and Weaver, T.D. (2018). Clarifying distinct models of modern human origins in Africa. *Curr Opin Genet Dev* 53, 148–156.
- Hobolth, A., Dutheil, J.Y., Hawks, J., Schierup, M.H., and Mailund, T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res* 21, 349–356.
- Hu, T., Long, M., Yuan, D., Zhu, Z., Huang, Y., and Huang, S. (2013). The genetic equidistance result, misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. *Sci China Life Sci* 56, 254–261.
- Huang, S. (2008). The genetic equidistance result of molecular evolution is independent of mutation rates. *J Comp Sci Syst Biol* 1, 092–102.
- Huang, S. (2009). Inverse relationship between genetic diversity and epigenetic complexity. *Nature Precedings* doi.org/10.1038/npre.2009.1751.1032.
- Huang, S. (2010). The overlap feature of the genetic equidistance result, a fundamental biological phenomenon overlooked for nearly half of a century. *Biological Theory* 5, 40–52.
- Huang, S. (2012). Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing tarsiers. *Sci China Life Sci* 55, 709–725.
- Huang, S. (2016). New thoughts on an old riddle: What determines genetic diversity within and between species? *Genomics* 108, 3–10.
- Ingman, M., Kaessmann, H., Paabo, S., and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708–713.
- Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C., and Cavalli-Sforza, L.L. (1983). Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* 19, 255–271.

- Johnson, M.S., Gopalakrishnan, S., Goyal, J., Dillingham, M.E., Bakerlee, C.W., Humphrey, P.T., Jagdish, T., Jerison, E.R., Kosheleva, K., Lawrence, K.R., *et al.* (2021). Phenotypic and molecular evolution across 10,000 generations in laboratory budding yeast populations. *eLife* *10*.
- Kasinathan, B., Colmenares III, S., McConnell, H., Young, J.M., Karpen, G.H., and Malik, H.S. (2020). Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in *Drosophila*. *eLife* *9*, e63368. <https://doi.org/63310.67554/eLife.63368>.
- Katz, S., Avrani, S., Yavneh, M., Hilau, S., Gross, J., and Hersberg, R. (2021). Dynamics of adaptation during three years of evolution under long-term stationary phase. *Mol Biol Evol.*
- Ke, Y., Su, B., Song, X., Lu, D., Chen, L., Li, H., Qi, C., Marzuki, S., Deka, R., Underhill, P., *et al.* (2001). African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* *292*, 1151–1153.
- Kern, A.D., and Hahn, M.W. (2018). The Neutral Theory in Light of Natural Selection. *Mol Biol Evol* *35*, 1366–1371.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* *217*, 624–626.
- Kimura, M. (1969). The rate of molecular evolution considered from the standpoint of population genetics. *Proc Natl Acad Sci U S A* *63*, 1181–1188.
- Kumar, S. (2005). Molecular clocks: four decades of evolution. *Nat Rev Genet* *6*, 654–662.
- Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Segurel, L., Venkat, A., Andolfatto, P., and Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol* *10*, e1001388.
- Lei, X., Yuan, J., Zhu, Z., and Huang, S. (2018). Collective effects of common SNPs and risk prediction in lung cancer. *Heredity*, doi:10.1038/s41437-41018-40063-41434.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* *27*, 2987–2993.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* *25*, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Margoliash, E. (1963). Primary structure and evolution of cytochrome c. *Proc Natl Acad Sci* *50*, 672–679.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* *20*, 1297–1303.
- Nemat-Gorgani, N., Hilton, H.G., Henn, B.M., Lin, M., Gignoux, C.R., Myrick, J.W., Werely, C.J., Granka, J.M., Moller, M., Hoal, E.G., *et al.* (2018). Different Selected Mechanisms Attenuated the Inhibitory Interaction of KIR2DL1 with C2(+) HLA-C in Two Indigenous Human Populations in Southern Africa. *J Immunol* *200*, 2640–2655.
- Ohta, T., and Gillespie, J.H. (1996). Development of Neutral and Nearly Neutral Theories. *Theoretical population biology* *49*, 128–142.
- Petr, M., Hajdinjak, M., Fu, Q., Essel, E., Rougier, H., Crevecoeur, I., Semal, P., Golovanova, L.V., Doronichev, V.B., Lalueza-Fox, C., *et al.* (2020). The evolutionary history of Neandertal and Denisovan Y chromosomes. *Science* *369*, 1653–1656.
- Pouyet, F., Asechbacher, S., Thiery, A., and Excoffier, L. (2018). Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife* *7*, e36317. <https://doi.org/36310.37554/elif.36317>.

- Poznik, G.D., Xue, Y., Mendez, F.L., Willems, T.F., Massaia, A., Wilson Sayres, M.A., Ayub, Q., McCarthy, S.A., Narechania, A., Kashin, S., *et al.* (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet* *48*, 593–599.
- Quinodoz, S.A., Jachowicz, J.W., Bhat, P., Ollikainen, N., Banerjee, A.K., Goronzy, I.N., Blanco, M.R., Chovanec, P., Chow, A., Markaki, Y., *et al.* (2021). RNA promotes the formation of spatial compartments in the nucleus. *Cell* *184*, 5775-5790 e5730.
- Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., and al., e. (2012). Insights into hominid evolution from the gorilla genome sequence. *Nature* *483*, 169-175.
- Scerri, E.M.L., Thomas, M.G., Manica, A., Gunz, P., Stock, J.T., Stringer, C., Grove, M., Groucutt, H.S., Timmermann, A., Rightmire, G.P., *et al.* (2018). Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? *Trends Ecol Evol* *33*, 582–594.
- Stringer, C.B., and Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science* *239*, 1263–1268.
- Teitz, L.S., Pyntikova, T., Skaletsky, H., and Page, D.C. (2018). Selection Has Countered High Mutability to Preserve the Ancestral Copy Number of Y Chromosome Amplicons in Diverse Human Lineages. *Am J Hum Genet* *103*, 261–275.
- Tsuzuki, M., Sethuraman, S., Coke, A.N., Roth, M.H., Boyle, A.P., and Wierzbicki, A.T. (2020). Broad noncoding transcription suggests genome surveillance by RNA polymerase V. *Proc Natl Acad Sci U S A* <https://doi.org/10.1073/pnas.2014419117>.
- Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., *et al.* (2000). Y chromosome sequence variation and the history of human populations. *Nat Genet* *26*, 358–361.
- Wang, M., Wang, D., Yu, J., and Huang, S. (2020). Enrichment in conservative amino acid changes among fixed and standing missense variations in slowly evolving proteins. *PeerJ* *8*, e9983 <https://doi.org/9910.7717/peerj.9983>.
- Wolpoff, M.H., Wu, X.Z., and Thorne, A.G. (1984). *Modern homo sapiens origins: a general theory of hominid evolution involving the fossil evidence from east Asia* (New York: Alan R. Liss).
- Wu, X. (2004). On the origin of modern humans in China. *Quaternary International* *117*, 131–140.
- Yuan, D., Lei, X., Gui, Y., Wang, M., Zhang, Y., Zhu, Z., Wang, D., Yu, J., and Huang, S. (2017). Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA. *bioRxiv*, doi.org/10.1101/101410
- Zhang, Y., and Huang, S. (2019a). The Out of East Asia model versus the African Eve model of modern human origins in light of ancient mtDNA findings. *bioRxiv*, [doi/10.1101/546234](https://doi.org/10.1101/546234).
- Zhang, Y., and Huang, S. (2019b). The Out of East Asia theory of modern human origins supported by recent ancient mtDNA findings. *Acta Anthropologica Sinica* *38*, 491–498.
- Zhu, Z., Yuan, D., Luo, D., Lu, X., and Huang, S. (2015). Enrichment of Minor Alleles of Common SNPs and Improved Risk Prediction for Parkinson's Disease. *PLoS ONE* *10*, e0133421.

Figures

Figure 1

Y chromosome phylogenetic trees of modern humans. Only major branches and representative SNPs are shown with branch lengths not to scale. The tree topology was built without making use of any ancient DNAs. A. The Out of East Asia model. B. The Out of Africa model.

Figure 2

mtDNA phylogenetic trees of modern humans. Only major branches are shown with branch lengths not to scale. Branch-defining SNPs are from Phylotree.org mtDNA tree Build 17 (Feb 17, 2016 version). A. The Out of East Asia model. B. The Out of Africa model from Phylotree.org mtDNA tree Build 17. Back mutations in the OoA model are designated by an exclamation mark (!).

Model based on the MGD theory

Model based on the neutral theory

Figure 3

Expected pattern of mutations in terminal and basal haplogroups of uniparental DNAs from the MGD theory or the neutral theory. Before basal haplogroup C diverged into terminal ones C1 and C2, no sites in C1 and C2 have mutated and only some sites in C have mutated under the MGD theory. After C1 and C2 diverged, C1 (or C2) had mutations at some sites defining C1 (or C2). In ancient samples, formation of terminal haplogroups C1 and C2 can take place before complete mutation in sites defining the basal haplogroup C under the MGD theory but cannot under the neutral theory. Convergent mutations are supposed to be common under the MGD theory but non-existent under the neutral theory and would result in complete mutations in sites defining the basal haplogroup C in present day C1 or C2 haplogroups in the model based on the MGD theory.

Figure 4

Schematics of the allele patterns in uniparental DNAs in the Africa model versus the Asia model. Sequences are represented by horizontal lines and sites defining representative haplogroups are indicated by rectangular boxes (length not to scale). A. Y chromosome. A0000 represents Denisovan haplotype and A000 represents Neanderthal haplotype. The term “A00 to T” represents all individual haplotypes that are defined by haplotype-specific mutations not present in archaic humans. In the case of A0 and G, the small black box in the “A00 to T” box indicates A0- or G-specific mutations, respectively. A0 is a result of admixture while G is not. B. mtDNA. The term “L0 to Z” represents all individual haplotypes that are defined by haplotype-specific mutations not present in archaic humans. In the case of L0 and U, the small black box in the “L0 to Z” box indicates L0- or G-specific mutations, respectively. L0 is a result of admixture while G is not.

A. 6826955, A1b=T, A00-A1a=C

B. 19424283, A1b=G, A00-A1a=C

C. 16779295, CT=A, AB=C

D. 23030717, CT=T, AB=C

E. 14149772, F=T, ABCDE=C

F. 16467111, F=T, ABCDE=C

Figure 5

Screen shots of aligned reads covering the sites not yet mutated in the OoEA-specific haplogroups belonging to the ancient sample I10871. Files of aligned sequencing reads (BAM files) of sample I10871-A00 were downloaded from published source and screen shots of the aligned reads covering the sites not yet mutated are shown. For each of the 6 sites shown here, HG19 positions, allele for the OoA-specific haplogroup, and allele for the OoEA-specific haplogroup are also indicated. The non-mutated site is marked by a red star.

Supplementary Files

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

- [SupplementaryTableS1samplelist.xlsx](#)