Examining models of modern human origins through the analysis of 43 fully sequenced human Y chromosomes

Shi Huang*

Center for Medical Genetics, Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, 110 Xiangya Road, Changsha, Hunan 410078, P.R. China

*Corresponding author: Shi Huang (shihuang1@gmail.com)

Short title: Testing human origin models with fully sequenced Y chromosomes

Keywords: Y chromosomes, maximum genetic diversity theory, neutral theory, infinite site model, Out of East Asia, Out of Africa,

Abstract

Over the past four decades, molecular studies have yielded two primary models for understanding the uniparental DNA phylogenetic trees of modern humans: the Out of Africa (OOA) and the Out of East Asia (OOEA) models. These models differ in their underlying assumptions, particularly in relation to early stem haplotypes, even though they share many haplotype relationships. Leveraging the wealth of new genetic variants unveiled through the comprehensive sequencing of 43 diverse human Y chromosomes, we here investigated the presence of shared variants among different haplotypes to determine which model better aligns with the genetic data. We validated our approach by confirming numerous well-established haplotype relationships that are consistent with both the OOA and OOEA models. Remarkably, our analysis revealed a compelling pattern: we were able to corroborate the existence of stem haplotypes specific to the OOEA model, but not those exclusive to the OOA model. For instance, we found that A0b and A1a shared the most variants with each other, aligning with the notion that both fall under the A00A1a stem haplotype of the OOEA model. So, it becomes evident that the genetic data obtained from the complete sequencing of the 43 newly analyzed human Y chromosomes lends robust support to the OOEA model as the more accurate representation of modern human origins.

Introduction:

The origins of modern humans remain a subject of ongoing and intensive research. Molecular analyses of contemporary human populations have generated two distinct models to explain the evolutionary history of modern humans: the "Out of Africa" (OOA) model and the "Out of East Asia" (OOEA) model. These models propose contrasting geographical locations as the point of origin for modern humans, with OOA suggesting Africa and OOEA suggesting East Asia. The OOA model gained prominence in the 1980s, primarily based on mitochondrial DNA (mtDNA) studies (1, 2). On the other hand, the OOEA model, also initially suggested by mtDNA studies in the 1980s (1), had been relatively overlooked until its recent independent rediscovery (3-5). Moreover, it has been revealed that all non-African Y chromosome haplotypes can be traced back to their origin in southern East Asia (6).

The foundation of rooting uniparental DNAs in Africa rests on two fundamental assumptions: the concept of a molecular clock and the prevalence of neutral mutations across the entire genome (1, 2, 7, 8). Within this neutral framework, the "infinite site assumption" holds particular significance. This assumption posits that mutations occur only once in the course of evolutionary history, thereby excluding the possibility of back mutations and convergent mutations. It serves as a critical factor in deducing the derived alleles that are pivotal in shaping the phylogenetic tree topology of the OOA model (7, 9). Nevertheless, the validity of the neutral assumption remains, at best, uncertain. Strong evidence suggests that mutation saturation and natural selection are more prevalent than initially believed (3, 10-15). In the Y chromosome phylogenetic tree of the OOA model, numerous incompatible alleles have emerged, showcasing recurrent mutations (16).

Genetic diversity patterns have continued to elude full explanation under the neutral theory (17), prompting the emergence of a novel framework in molecular evolution known as the "maximum genetic diversity theory" (12, 18, 19). Central to this new theory is the intuitive notion, initially demonstrated by Fisher nearly a century ago, that mutations of a given size are less likely to be favorable in complex than simple organisms (20, 21). The most recent supporting evidence for this concept lies in the observation that lower species tend to harbor a greater number of common variants, which, however, are often detrimental or non-tolerated in humans (22). In accordance

with this new theory, and bolstered by a wealth of strong evidence, genetic diversities or distances are primarily observed to be at optimal saturation levels, rather than continually scaling with time, as assumed by the molecular clock and neutral theory. The Y chromosome phylogenetic tree within the OOEA model is constructed based on this novel theoretical framework, setting it apart from the OOA model in three significant ways (see Fig. 1). Firstly, in OOEA, haplotypes are defined by haplotype-specific alleles, irrespective of whether they are derived or result from back mutations. Secondly, the rooting in the OOEA model is grounded in a straightforward premise: the original haplotype should be the one most commonly shared among individuals, as mutations leading to alternative haplotypes should be relatively rare occurrences (1). Finally, the haplotype representing the ancestral lineage is expected to exhibit a closer genetic relationship to ancient DNA samples from the time when modern humans initially appeared, in contrast to other contemporary haplotypes (3-5).

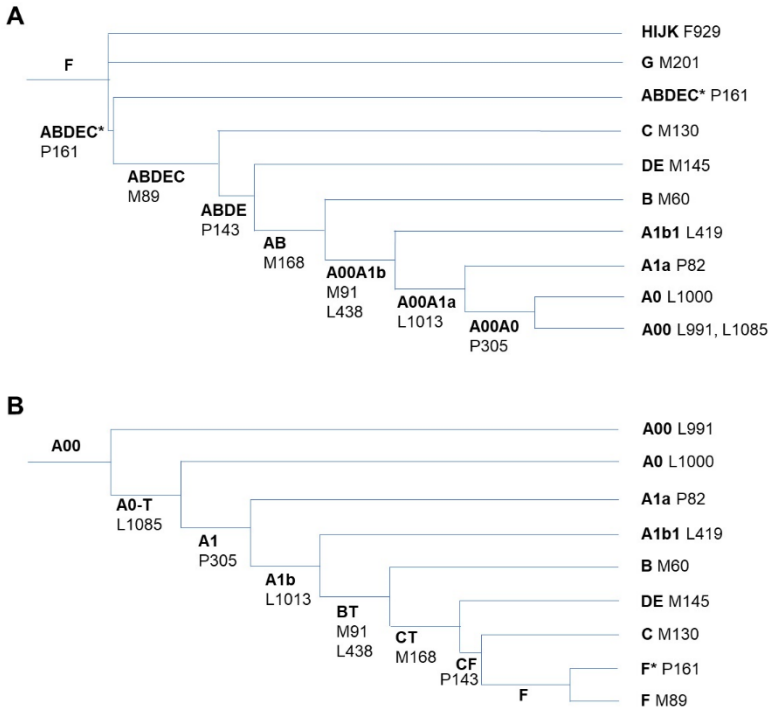

Fig. 1. Y chromosome phylogenetic trees of modern humans. A. The Out of East Asia model. B. The Out of Africa model. Only major branches and representative SNVs are shown with branch lengths not to scale. The haplotype topologies for F subtypes under the OOA model, including G, H, I, J, and K, are consistent and identical between the OOA model and the OOEA model. These haplotypes share the same genetic relationships and branching patterns in both models.

Both models, OOA and OOEA, rely on assumptions that carry inherent uncertainties, necessitating independent testing to establish their validity. Human Y chromosomes have until recently only been partially sequenced (53.8%) and the phylogenetic tree of Y chromosome is built by using single nucleotide variants (SNVs) located in the non-recombining Y region (NRY) or the male specific Y region (MSY). However, a groundbreaking development is the recent achievement of complete sequencing for 43 human Y chromosomes, representing a wide array of major human groups or haplotypes (23). This comprehensive sequencing has unveiled a plethora of SNVs and insertions or deletions (indels), predominantly found in the pseudoautosomal regions (PARs), which now offer a valuable resource for an independent examination of human origin models.

In our study, we scrutinized these variants to ascertain which model finds stronger support in the context of this new variant data. Our results indicate that numerous variants are shared across different haplotypes, providing an informative pattern that sheds light on the relationships among these haplotypes. Importantly, this analysis validates the Y chromosome phylogenetic tree of the OOEA model, offering compelling evidence in its favor.

Methods and Materials:

Datasets

The data utilized in this study were sourced from Supplementary Tables S22, S24, and S25, as published in the paper by Hallast et al. in 2023 (23). These tables provide comprehensive listings of large and small size indels (Table S22 and S24) and SNVs (Table S25), along with the occurrences of these variants across the 43 Y chromosomes/samples. In our analysis, we combined the information from Table S22 and S24, allowing us to investigate a total of 24,325 indels and 53,744 SNVs for our study.

Calculation of the shared variant fraction between two haplotypes

For each haplotype/sample, we first compiled a list of SNVs or indels that are present within that specific haplotype/sample. We then employed custom scripts designed for this purpose to tally the count of shared genetic variants between the given haplotype/sample and all the other haplotypes/samples among the 43 Y chromosomes

analyzed in our study. To quantify the extent of shared variants between two haplotypes, we computed the fraction of shared variants. This fraction was computed for each haplotype by dividing the number of shared variants by the total number of variants identified in that particular haplotype. To represent the overall fraction of shared variants between two haplotypes, we calculated the average of the fractions for both haplotypes in the pair. This approach allowed us to systematically assess the genetic relationships between the various haplotypes within our dataset.

Calculation of the shared variant fraction between a stem haplotype and other haplotypes

The shared variant fraction between a stem haplotype and other haplotypes was calculated as follows, using the O stem haplotype as an example to illustrate our method. 1) Identification of SNVs or indels defining the O stem haplotype: SNVs or indels defining a stem haplotype, such as O, should be present in all terminal branch haplotypes. In the case of the O stem haplotype, there were one O1b sample and four O2a samples in the 43 Y chromosomes dataset. All SNVs or indels that were present in all five O samples were retrieved and designated as the O stem set. 2) Counting shared SNVs or indels: For a specific haplotype/sample, the number of shared SNVs or indels between that haplotype and the O stem set was counted. 3) Calculation of the shared fraction: The average shared fraction between the O stem haplotype and a particular haplotype (e.g., E) was determined by calculating the average of the two shared fractions, one for the O stem set and the other for that particular haplotype. The shared fraction for the O stem set was obtained by dividing the number of shared variants by the total number of variants in the O stem set. The shared fraction for the particular haplotype was computed by dividing the number of shared variants by the total number of variants in the particular haplotype. 4) Handling multiple samples: For haplotypes with multiple samples or terminal haplotypes represented in the 43 Y chromosomes dataset, such as haplotype E, the shared fraction between each individual E haplotype/sample and the O stem haplotype was independently calculated. These shared fractions for all E haplotypes/samples were then aggregated to calculate an average value. This average value was used to represent the shared fraction between the E haplotype as a whole and the O stem haplotype. This method allowed us to systematically assess and compare the shared genetic variants between the O stem haplotype and other haplotypes.

Results:

The genetic variants from the sequencing of 43 complete human Y chromosomes include 876 structural variants, 23,459 indels and 53,744 SNVs (23). Within this dataset, we identified numerous variants that serve to uniquely define the non-controversial haplotypes shared by both models. As an example, the NO stem haplotype is defined by 76 SNVs and 9 indels, while the O stem haplotype is delineated by 106 SNVs and 15 indels (Table 1). Furthermore, we pinpointed a range of variants that define the stem haplotypes exclusive to the OOEA model, including A00A1a, AB, ABDE, and ABDEC (Table 1 and Fig. 1A). An illustration of this is that A0b and A1a share 720 SNVs and 143 indels that are absent in other haplotypes/samples, supporting the A00A1a clade in the OOEA model. Interestingly, we observed that many variants are present in two or more haplotypes that do not share a most recent common ancestor (MRCA), indicating convergent mutations or recombination at the PARs. For example, G2a and H1a share 5 SNVs and 4 indels that are absent in other haplotypes/samples (Table 1). If these mutations had arisen in the MRCA of G and H, they would have been present in all H, I, J, and K haplotypes, as these 4 haplotypes share a more recent MRCA than with G.

Table 1. Number of haplotype-specific variants for each haplotype or combination of haplotypes. The listed haplotypes represent only a fraction of the complete set.

Haplotype	Number SNVs	Number indels
A00A1a	720	143
AB	324	60
ABDE	5	2
ABDEC	192	41
NO	76	9
O	106	15
P	8	0
R	1	0
A1a,B2b	4	27
G2a,H1a	5	4
G2a,J2	3	4
G2a,J1	3	3
G2a,O1b	1	3
B2b,G2a	5	8
B2b,H1a	19	13
C1a,G2a	3	4
C1a,H1a	24	11
N1a,H1a	3	4

Additionally, it's worth noting that many variants in one haplotype are also shared with other haplotypes (Table 2, see Supplementary Table S1 for an expanded version of Table 2 containing all 43 samples and Supplementary Table S2 for the same analysis of indels). For example, among 7746 SNVs present in A0b, only 3369 are unique to A0b and the rest is shared in various degrees with each of the other haplotypes (16-45%) represented by the 43 samples (Table 2 and Supplementary Table S1). The results indicate that these variants have likely spread due to recombination, given that most of them are situated within the PARs. This extensive sharing of variants among haplotypes further enhances our understanding of the genetic landscape of the human Y chromosomes.

Table 2. Fractions of shared SNVs among selected haplotypes or samples
The bold numbers represent the shared SNV fraction value for a pair of haplotypes/samples, which exceeds the value obtained when each of the pair is compared to the 41 other haplotypes/samples individually.

			Group	ACB	GWD	LWK	YRI	GWD	JPT	MSL	BEB	CLM	PJL	FIN	CHS	CHS	MXL	TSI
ID	Haplotype	#Total SNV	#Unique	HG01890	HG02666	NA19384	NA19239	HG03248	NA18989	HG03579	HG03009	HG01258	HG02492	HG00358	HG00673	HG00621	NA19650	NA20509
HG01890	A0b	7746	3339		**0.45**	0.25	0.33	0.32	0.25	0.26	0.24	0.25	0.23	0.23	0.29	0.31	0.21	0.26
HG02666	A1a	6883	2589			0.30	0.36	0.40	0.27	0.29	0.24	0.27	0.29	0.26	0.30	0.31	0.24	0.28
NA19384	B2b	10952	2476				0.33	0.36	0.41	0.39	0.47	0.35	0.45	0.46	0.37	0.37	0.28	0.26
NA19239	E1a2a1a1a	5195	361					**0.64**	0.36	0.42	0.34	0.38	0.31	0.28	0.38	0.38	0.36	0.44
HG03248	E1a2b1a1a1	4809	322						0.36	0.39	0.33	0.36	0.38	0.34	0.33	0.35	0.34	0.31
NA18989	C1a	10060	2676							0.39	0.46	0.40	0.44	0.44	0.41	0.43	0.31	0.36
HG03579	G2a	6150	1162								0.43	0.48	0.39	0.40	0.36	0.36	0.38	0.40
HG03009	H1a	7356	1166									0.42	0.48	0.51	0.42	0.44	0.34	0.33
HG01258	J1	6834	1079										0.45	0.40	0.34	0.35	0.39	0.42
HG02492	J2	7362	960											0.50	0.40	0.41	0.31	0.28
HG00358	N1a	6228	816												0.42	0.42	0.28	0.28
HG00673	O2a2a1a1a	4477	512													**0.59**	0.26	0.38
HG00621	O2a2b1a1a1	4805	362														0.30	0.37
NA19650	R1b1a1b1a1a	2621	341															**0.60**
NA20509	R1b1a1b1a1a	2128	100															

Haplotypes that are closely related, sharing a MRCA, are expected to exhibit a higher degree of variant sharing among them. This shared genetic material may arise from common ancestry, convergent mutations, or recombination events occurring within the pseudoautosomal regions (PARs). Convergent mutations, for instance, may result from shared adaptations, while recombination at the PARs, involving the X chromosomes of the same or similar population groups, could lead to increased variant sharing among closely related Y haplotypes. To assess the proximity of Y haplotypes, one can examine the extent of shared variants among different samples. If, among the 43 haplotypes or samples, haplotype 1 shares the highest fraction of variants with haplotype 2, and vice versa, these two haplotypes are considered the closest to each other or are likely to belong to the same clade. For example, sample NA19239 demonstrated a 64% overlap

in SNVs with sample HG03248, a percentage higher than that seen between NA19239 and any other samples, or between HG03248 and any other samples (Table 2). It is worth noting that these two samples are indeed recognized as belonging to the same clade, E1a2a1a1a.

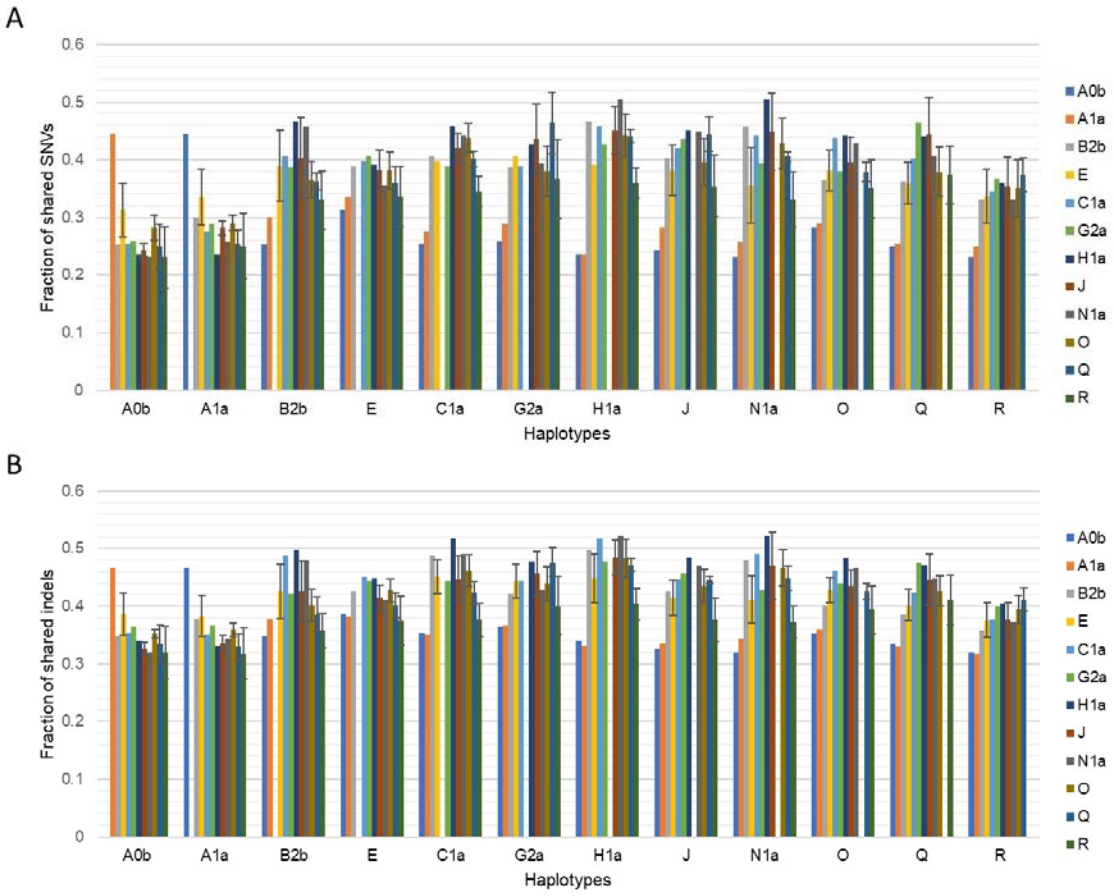

Fig. 2. Fraction of shared variants among individual terminal haplotypes. A. Fraction of shared SNVs. B. Fraction of shared indels. In the case of haplotypes like E, J, O, Q, and R that have more than one individual terminal haplotype/sample represented in the dataset of 43 chromosomes, the average value of all samples under a haplotype is shown with error bars representing the standard deviation of the data. The information used to generate the figures were extracted from Supplementary Table S1 and S2.

Utilizing the approach of quantifying variant sharing, we pinpointed 10 pairs of samples that emerged as each other's closest relatives based on the shared SNV fractions among the 43 samples (Table 2 and Supplementary Table S1). Remarkably, nine out of these 10 pairs are well-documented close relatives within the same clade, thereby

affirming the efficacy of this method in identifying closely related haplotypes. However, one particular pair, A0b and A1a, stands out as an anomaly, contrary to expectations under the OOA model, which suggests that A1a should be more closely related to haplotypes within the A1b clade, such as B, C, E, or F, among others (see Fig. 1B). In stark contrast, within the OOEA model (as shown in Fig. 1A), these two haplotypes fall under the same A00A1a clade and are more closely related to each other than to any other haplotypes. It's noteworthy that consistent results were obtained when applying the same approach to indels (Supplemental Table S2). What's particularly striking is the absence of any pair of unrelated samples or haplotypes demonstrating the pattern of sharing the most variants with each other. For instance, G2a/HG03579 shared the most SNVs (50.2%) with the unrelated Q1b/HG01928, while Q1b/HG01928 shared the most (57.5%) with the known related Q1b/HG01952, rather than with G2a/HG03579. Thus, our findings provide strong support for the A00A1a clade within the OOEA model, which encompasses A0b and A1a as terminal branches.

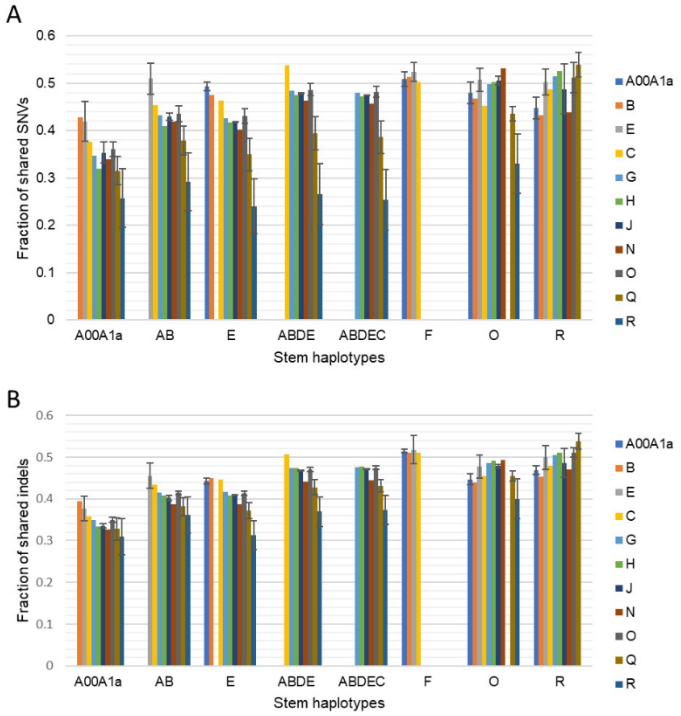

Fig. 3. Fraction of shared variants between a stem haplotype in the OOEA model and other individual terminal haplotypes in the 43 human Y chromosomes. A. Fraction of shared SNVs. B. Fraction of shared indels. For haplotypes such as A, E, J, O, Q, and R that have more than one individual terminal haplotype/sample represented in the 43 chromosomes, the average value of all samples under a haplotype is shown with error bars representing the standard deviation.

Under the framework of the OOA model, the expected topological relationship between A1a and A0b is akin to that between B and A1b1, E and B, or C and E (as depicted in Fig. 1B). Thus, if the OOA model anticipates that A1a should exhibit the highest degree of variant sharing with A0b, it would logically imply that E should share the most variants with B, or that C should share the most variants with E. To put this prediction to the test, we conducted a thorough analysis of the sharing patterns among all individual haplotypes and samples, as presented in Fig. 2. The results showed a significant disparity from the expected outcomes under the OOA model. Specifically, E was found to share a greater number of variants with G, C, or H than with B, and C shared more variants with H, N, and O than with E. Consequently, the close relationship observed between A0b and A1a is fundamentally at odds with the predictions of the OOA model.

The approach of inferring close relationships through the level of shared variants offers a fresh perspective on the evaluation of the two competing models of human origins. To begin, we established the validity of this approach by scrutinizing whether it could confirm non-controversial relationships among the stem haplotypes. In this analysis, we focused on SNVs or indels that are especially relevant for defining a stem branch or haplotype. These variants are present in all individual terminal haplotypes belonging to the respective stem haplotype. Such stem-haplotype defining variants can provide a more informative insight into the relationship between a stem haplotype and another haplotype compared to the variants found in a terminal haplotype. Our findings confirmed the reliability of this approach, as the O stem haplotype shared the most SNVs or indels with the N haplotype, aligning with the well-documented NO stem haplotype. Moreover, the R stem haplotype exhibited the highest degree of shared SNVs or indels with the Q haplotype, consistent with both falling under the well-established P stem haplotype (see Fig. 3). It is important to note that instances where an individual terminal haplotype, such as O1b/NA18534, did not emerge as the one sharing the most variants with the most O-related N haplotype might be attributed to individual-specific variations that are not representative of the broader O stem haplotype (refer to Table 2 and Supplementary Table S1-S2).

Next, we turned our attention to examining whether the stem haplotypes exclusive to the OOEA model exhibited the anticipated relationships. As depicted in Fig. 3, the stem haplotype A00A1a of the OOEA model shared the most variants with B2b, a finding that

aligns with the presence of the AB stem haplotype. The AB stem haplotype in the OOEA model exhibited the highest degree of shared variants with E, while the stem haplotype E shared the most with A and B, all consistent with the existence of the ABDE stem haplotype. Notably, the ABDE stem haplotype shared the most variants with C, reinforcing the presence of the ABDEC stem haplotype. Additionally, F was found to share variants nearly equally with A, B, E, and C, which is in line with expectations derived from the OOEA model. These results lend substantial support to the actual existence of all the stem haplotypes unique to the OOEA model. By contrast, the predictions derived from the OOA model, wherein E should be closer to C than B or A and F should be closer to C than E, do not align with the observed relationships depicted in Fig. 3. Consequently, these findings present a clear case against the OOA model.

All the stem haplotypes within the OOEA model that are common in African populations, including A00A1a, AB, ABDE, and ABDEC, are anticipated to exhibit equal genetic relatedness to all F subtypes. The observations presented in Fig. 3 corroborate this expectation, with the exception of R and Q. This discrepancy can likely be attributed to the fact that the reference genome used for variant calling (GRCh38 Y) corresponds to the R haplotype, which would inherently result in a reduced number of variants called for R and its closely related Q haplotype, as demonstrated in Table 2 and Supplementary Table S1-S2 where the total number of variants called for each haplotype or sample are listed. Interestingly, the O stem haplotype was found to share slightly more variants with these African stem haplotypes compared to the other F subtypes, such as G, H, and J (Fig. 3). This observation appears to defy the expected pattern based on the haplotype topologies, as O is generally considered more closely related to R compared to G, H, and J. Consequently, one might anticipate lower genetic sharing with African stem haplotypes for O, similar to what was observed for R.

We extended our analysis to assess the stem haplotypes exclusive to the OOA model, including BT, CT, and CF. According to the OOA model's predictions, BT should exhibit the closest relationship to A1a, CT should be most closely related to B, and CF should have the strongest affinity with E. However, our findings deviate from these expectations. Our results indicated that BT shares a similar degree of relationship with both A0b and A1a, CT is similarly related to B2b, A1a, and A0b, and CF exhibits similarities in relationships with E, B2b, A1a, and A0b (Fig. 4). These results raise questions about the

validity of the stem haplotypes unique to the OOA model, suggesting that they may not correspond to actual genetic relationships.

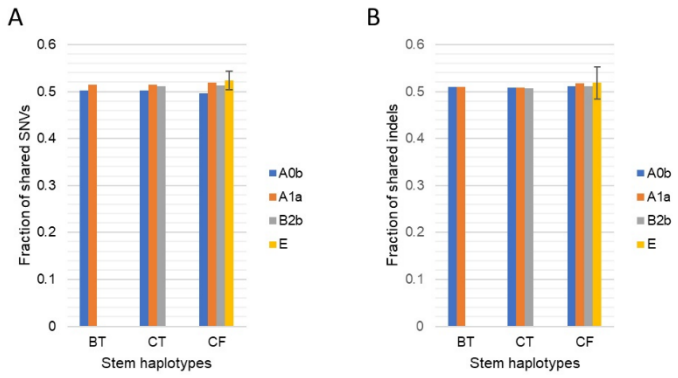

Fig. 4. Fraction of shared variants between a stem haplotype in the OOA model and other individual terminal haplotypes in the 43 human Y chromosomes. A. Fraction of shared SNVs. B. Fraction of shared indels. For haplotypes such as E that have more than one individual terminal haplotype/sample represented in the 43 chromosomes, the average value of all samples under a stem haplotype is shown with error bars representing the standard deviation.

Discussions:

In our study, we observed that the stem haplotypes exclusive to the OOEA model consistently exhibited the anticipated sharing pattern, thereby supporting the model's validity. However, by contrast, a majority of the stem haplotypes specific to the OOA model failed to manifest the expected sharing pattern, raising questions about the model's accuracy.

Within the OOEA model, a well-defined hierarchy is observed, where A0b and A1a fall within the A00A1a clade, A00A1a resides within the AB clade, AB is within ABDE clade, E is encompassed within the ABDE, ABDE is nested within the ABDEC clade, and C finds its place within the ABDEC clade. Remarkably, the anticipated variant sharing patterns are in complete alignment with the actual observations we have presented in our study. These findings offer compelling support for the OOEA model's representation of the genetic relationships among these haplotypes.

Our study presents three observations that pose significant challenges to the OOA model. First, in contrast to the OOA model's predictions, we found that A1a exhibited the closest relationship with A0b. However, this observation was not paralleled by the

expected proximity of E to B, or C to E as prescribed by the OOA model, if one were to consider the possibility of this observation being compatible with the OOA model. Second, according to the OOA model, the E stem haplotype should be more closely related to C than to B, and the F stem haplotype should have a closer genetic affinity with C than with E. Our study did not align with these expectations. Lastly, the stem haplotypes unique to the OOA model, namely BT, CT, and CF, all failed to exhibit the anticipated sharing patterns, even though the same methodology could confirm the expected closer relationship between R and Q or between O and N. These observations collectively cast doubt on the OOA model's ability to explain the genetic relationships we observed in our study.

The observation that the O stem haplotype, prevalent in East Asian populations, shares more genetic variants with African stem haplotypes compared to other F subtypes, such as G, H, and J, which are more commonly found in West and South Asian populations (Fig. 3), is in line with the hypothesis that African stem haplotypes may have been influenced to a greater extent by East Asian populations. This observation supports the idea of a westward migration, as proposed in the OOEA model, rather than an eastward migration as suggested by the OOA model. According to the OOEA model, the original modern human population is believed to have resided in East Asia and carried the F stem haplotype. This population is thought to have later given rise to a distinct population characterized by the ABDEC stem haplotype. Therefore, based on the OOEA model, East Asians would have been the only modern population present during that ancient period with the potential to influence the development and genetic characteristics of the ABDEC stem haplotype and the ABDEC population. By contrast, under the OOA model, one would expect the haplotypes from West and South Asia to exhibit a closer genetic relationship with African populations due to their geographical proximity. Thus, our findings further challenge the OOA model and reinforce the validity of the OOEA model.

In the past, we have employed ancient DNA analysis to provide independent validation of the OOEA model, as detailed in our previous work (24). The results presented in our current study offer yet another strong confirmation of the OOEA model's accuracy. Future molecular tests to further assess these models may concentrate on obtaining more comprehensive sequencing data from additional ancient and contemporary human

Y chromosomes, with particular emphasis on populations from Africa and East Asia. Expanding the scope of investigations to encompass a wider array of genetic data will contribute to a more comprehensive and refined understanding of human origins.

Acknowledgements:
This research was conducted without the direct support of any specific research grants.

Competing interests:
The author declares no competing interests.

References:

1. Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol. 1983;19(3-4):255-71.
2. Cann RL, Stoneking AC, Wilson AC. Mitochondrial DNA and human evolution. Nature. 1987;325:31-6.
3. Yuan D, Lei X, Gui Y, Wang M, Zhang Y, Zhu Z, et al. Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA. bioRxiv. 2019:doi.org/10.1101/101410
4. Zhang Y, Huang S. The Out of East Asia theory of modern human origins supported by recent ancient mtDNA findings. Acta Anthropologica Sinica. 2019;38(4):491-8.
5. Zhang Y, Huang S. The Out of East Asia model versus the African Eve model of modern human origins in light of ancient mtDNA findings. bioRxiv. 2019:doi/10.1101/546234.
6. Hallast P, Agdzhoyan A, Balanovsky O, Xue Y, Tyler-Smith C. A Southeast Asian origin for present-day non-African human Y chromosomes. Human genetics. 2021;140(2):299-307.
7. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, et al. Y chromosome sequence variation and the history of human populations. Nat Genet. 2000;26(3):358-61.
8. Ke Y, Su B, Song X, Lu D, Chen L, Li H, et al. African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. Science. 2001;292(5519):1151-3.

9. Ingman M, Kaessmann H, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. Nature. 2000;408:708-13.
10. Kreitman M. The neutral theory is dead. Long live the neutral theory. Bioessays. 1996;18(8):678-83; discussion 83.
11. Kern AD, Hahn MW. The Neutral Theory in Light of Natural Selection. Mol Biol Evol. 2018;35(6):1366-71.
12. Huang S. New thoughts on an old riddle: What determines genetic diversity within and between species? Genomics. 2016;108(1):3-10.
13. Teitz LS, Pyntikova T, Skaletsky H, Page DC. Selection Has Countered High Mutability to Preserve the Ancestral Copy Number of Y Chromosome Amplicons in Diverse Human Lineages. Am J Hum Genet. 2018;103(2):261-75.
14. Zhu Z, Yuan D, Luo D, Lu X, Huang S. Enrichment of Minor Alleles of Common SNPs and Improved Risk Prediction for Parkinson's Disease. PLoS ONE. 2015;10(7):e0133421.
15. Wang M, Huang S. The collective effects of genetic variants and complex traits. J Hum Genet. 2022(https://doi.org/10.1038/s10038-022-01105-1).
16. Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat Genet. 2016;48(6):593-9.
17. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, et al. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 2012;10(9):e1001388.
18. Bickel D. Testing Hypotheses of Molecular Evolution. In: Phylogenetic Trees and Molecular Evolution.: Springer, Cham.; 2022.
19. 黄石, 朱作斌. 系统生物学与进化理论. 朱作斌, 张潇, 王亮, editors. 南京: 东南大学出版社; 2022.
20. Fisher RA. The genetical theory of natural selection. Oxford, U.K.: Oxford University Press; 1930.
21. Orr HA. Adaptation and the cost of complexity. Evolution. 2000;54(1):13-20.
22. Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, et al. The landscape of tolerated genetic variation in humans and primates. Science. 2023;380(6648):eabn8153.

23. Hallast P, Ebert P, Loftus M, Yilmaz F, Audano PA, Logsdon GA, et al. Assembly of 43 human Y chromosomes reveals extensive complexity and variation. Nature. 2023;621(7978):355-64.
24. Xia Z, Chen H, Zhang Y, Huang S. Ancient uniparental DNAs in distinguishing the competing theories of molecular evolution and modern human origins. Research Square. 2021:DOI:10.21203/rs.3.rs-1087730/v1.