Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA

Dejian Yuan[1], Xiaoyun Lei[1], Yuanyuan Gui[1], Mingrui Wang[1], Ye Zhang[1], Zuobin Zhu[2], Dapeng Wang[3], Jun Yu[3], and Shi Huang[1]*

[1]State Key Laboratory of Medical Genetics, School of Life Sciences, Xiangya Medical School, Central South University, 110 Xiangya Road, Changsha, Hunan 410078, P.R. China

[2]Department of Genetics, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China.

[3]CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, P.R. China

*Corresponding author: huangshi@sklmg.edu.cn, tel 86-731-84805342

Dejian Yuan, yuandejian@163.com

Xiaoyun Lei, leixiaoyun@sklmg.edu.cn

Yuanyuan Gui, guiyuanyuan@sklmg.edu.cn

Mingrui Wang, wangmingrui@sklmg.edu.cn

Ye Zhang, zhangye@sklmg.edu.cn

Zuobin Zhu, zhuzuobin2010@163.com

Dapeng Wang, wangdp123@gmail.com

Jun Yu, junyu@big.ac.cn

Key words: Multiregional; Out-of-Africa; neutral theory; maximum genetic diversity (MGD) hypothesis; Neanderthals; Denisovans; Heidelbergensis; Aboriginal Australians; Negritos

Abstract

The neutral theory has been used as a null model for interpreting nature and produced the Recent Out of Africa model of anatomically modern humans. Recent studies, however, have established that genetic diversities are mostly at maximum saturation levels maintained by selection, therefore challenging the explanatory power of the neutral theory and rendering the present molecular model of human origins untenable. Using improved methods and public data, we have revisited human evolution and found sharing of genetic variations among racial groups to be largely a result of parallel mutations rather than recent common ancestry and admixture as commonly assumed. We derived an age of 1.86-1.92 million years for the first split in modern human populations based on autosomal diversity data. We found evidence of modern Y and mtDNA originating in East Asia and dispersing via hybridization with archaic humans. Analyses of autosomes, Y and mtDNA all suggest that Denisovan and Neanderthal were archaic Africans with Eurasian admixtures and ancestors of South Asia Negritos and Aboriginal Australians. Verifying our model, we found more ancestry of Southern Chinese from Hunan in Africans relative to other East Asian groups examined. These results suggest multiregional evolution of autosomes and replacements of archaic Y and mtDNA by modern ones originating in East Asia, thereby leading to a coherent account of modern human origins.

Background

There are two competing models of modern human origins termed “Multiregional” and the recent “Out-of-Africa” hypothesis [1]. In the Multiregional model [2-4], recent human evolution is seen as the product of the early and middle Pleistocene radiation of *Homo erectus* from Africa. Thereafter, local differentiation led to the establishment of regional populations which evolved to produce anatomically modern humans (AMH) in different regions of the world, made of four major differentiated groups (Africans, Europeans, East Asians, and Aboriginal Australians). *Homo* has been a single species since the genus first appeared in the fossil record ~2.3-2.8 million years (myr) ago [2-4]. Support for this model is based on fossils and Paleolithic cultural remains but consistent molecular evidence has been lacking. While autosomal data have put a common ancestor of modern humans at ~1.5 myr ago, it is still far short of 2 myr [5]. In addition to regional continuity, the model further suggests hybridization among different groups [4]. Seeming difficulties here are the clear separation between modern and archaic mtDNAs and Y, the absence of archaic mtDNAs and Y in modern humans [6, 7], and the young age for the modern Y (~100 ky) and mtDNA (~200 ky) [8-10].

The single origin Out of Africa model assumes that there was a relatively recent common ancestral population for *Homo sapiens* which already showed most if not all of the anatomical features shared by present day people. This population originated in Africa ~200 ky ago, followed by an initiation of African regional differentiation, subsequent radiation from Africa, and final establishment of modern regional characteristics outside Africa [1, 10]. These modern Africans replaced the archaic *Homo* in Eurasia with limited genetic mixing [11-15]. Support for this model comes from the African location of the earliest fossils (~315 ky ago in Jebel Irhoud, Morocco) of mostly but not all AMH features [16, 17] and the molecular clock or neutral theory interpretation of the greater genetic

diversity in Africans [10]. Difficulties with this model include the discrepancy between autosomal and Y/mtDNA age, the Y haplotype A00 with age >300 ky [18], no evidence of first appearance of fully modern humans in Africa, seeming data for multiregionalism within Africa [19] and for multiple dispersals into Asia [20], fossils with AMH features of greater than 85 ky old (upto ~260 ky) in multiple Eurasia locations (the fully modern teeth in Daoxian Hunan, Longtanshan cave 1 Yunnan, Luna Cave Guangxi, Xuchang Henan, Bijie Guizhou, and Dali Shaanxi in China, Misliya and Shkul/Qafzeh in Israel, and Al Wusta-1 in Arabia) [21-28], and the generally weaker support from fossils and stone tools relative to the multiregional model. While the AMH fossils found outside Africa have been assumed to originate in Africa, an origin in Asia has not been excluded. In fact, in 1983, researchers have derived an mtDNA tree rooted in Asia [29]. Unfortunately, this model was overlooked without anyone ever explaining why the Asia model was less valid than the African Mitochondrial Eve model published 4 years later [10].

Most fatal to the Out of Africa model, however, is that the theoretical foundation for it, the molecular clock and neutral theory, is widely known to be incomplete or has yet to solve the century old riddle of what determines genetic diversity [30]. The neutral theory, while largely sound as a null model and a framework for pre-saturation evolutionary processes, has met with great difficulty as an explanatory framework for most molecular evolutionary phenomena [31-34] and as such should not have been so freely used to account for genetic diversity patterns. Obviously, inferring human origins by using genetic diversity data must wait until one has a complete understanding of what genetic diversity means. The standard for such an understanding should of course be a complete and coherent account of all known puzzles related to genetic diversity.

The unusually admixed features of the Aboriginal Australians have yet to be explained by any model [1]. A list of morphological features aimed at defining modern

humans would exclude both modern Aboriginal Australians and Neanderthals, indicating some shared traits between the two [35]. Also unexplained is the origin of Negritos in South Asia. Despite the obvious phenotypic similarities and close Y and mtDNA relationships, no special autosomal relationship has yet been found between Negritos and African pygmies or even among different Negrito groups in South Asia [36].

In recent years, a more complete molecular evolutionary theory, the maximum genetic distance or diversity (MGD) hypothesis, has been making steady progress in solving both evolutionary and contemporary biomedical problems [37-50]. The core concept of the MGD theory, maximum genetic diversity or distance no longer changing with time, is *a priori* expected and supported by numerous facts [37, 51, 52]. In contrast, the neutral theory and its infinite site model fail to take MGD into account and tacitly assume that nearly all observed genetic distances or diversities could still increase with time with no limit defined [53, 54]. The MGD theory has solved the two major puzzles of genetic diversity, the genetic equidistance phenomenon and the much narrower range of genetic diversity relative to the large variation in population size [30, 37]. The primary determinant of genetic diversity (or more precisely MGD) is species physiology [37, 55]. The genetic equidistance result of Margoliash in 1963 is in fact the first and best evidence for MGD rather than linear distance as mis-interpreted by the molecular clock and in turn the neutral theory [37, 39, 44, 45, 56-58]. Two contrasting patterns of the equidistance result have now been recognized, the maximum and the linear [45, 57]. The neutral theory explains only the linear pattern, which however represents only a minority of any genome today. The link between traits/diseases and the amount of SNPs shows an optimum genetic diversity level maintained by selection, thereby providing direct experimental disproof for the neutral assumption for common SNPs [38, 40, 46-48, 50, 59].

More direct functional data invalidating the neutral assumption have also been found [60, 61].

One simple method to determine whether any DNA fragment has reached MGD is to count the number of overlap sites (coincident substitutions) in a sequence alignment of three different species [44]. Such sites represent positions where mutations leading to different residues had occurred independently at the same position in at least two species, which would be a low probability event under the neutral theory or its infinite site assumption but common under the MGD theory [44]. The neutral theory is only valid for slow evolving genes yet to reach MGD, where its infinite site assumption holds and the number of overlap sites follows calculation from probability theory [44]. Unfortunately, however, nearly all existing phylogenetic results are from fast evolving sequences that were *assumed* to follow the infinite site model when they in fact do not as they have now been shown to be enriched with overlap sites [44].

Coincident substitutions at overlap sites do not contribute to genetic distances and make the relationship between distance and time hard if not impossible to model accurately. To overcome this, we have developed the “slow clock” method that only uses slow evolving DNAs with zero or few overlap sites. Contrary to naïve expectations based on the neutral theory, fixed and standing variations in slow evolving proteins are in fact enriched with conservative amino acid substitutions and hence more neutral and suitable for phylogenetic inferences [62]. The method has produced a separation time for the pongids and humans that is remarkably consistent with common sense and the original interpretation of fossil records and drastically different from the result of fast evolving DNAs [45]. Here we used the MGD theory and its related methods to revisit the evolution of modern humans. The unique value of the MGD theory in human origin studies is that it helps select the truly informative sequences that would follow the neutral theory. Once

149 such sequences are selected, the remaining methodologies would be mostly covered by
150 the neutral theory, and we fully grant the neutral theory to be valid for truly neutral
151 sequences still at the linear phase of accumulating variations. We just disagree with its
152 treatments of most genome sequences to be neutral and at the linear or near linear
153 phase of accumulating genetic diversity.
154

155 **Methods**

156 **Sequence download.** We downloaded ancient and modern human genome sequences
157 using publically available accession numbers. South Asian and Oceanian SNPs data
158 from Pugach et al (2013) were obtained from the authors [63]. The Hunan and Fujian
159 identity information of CHS sample of 1KG were obtained from the Coriell Institute
160 website.
161

162 **Selection of SNPs.** *Random selection of 255K SNPs as fast evolving SNPs.* We
163 selected 255K SNPs from 1KG data to represent the average variation of the genome
164 (Supplementary Table S1). We first generated a random number for each SNP on a given
165 chromosome followed by sorting the SNPs based on the random numbers, and then
166 selected the top ranked set of SNPs with the number of SNPs in the set proportional to
167 the size of the chromosome. SNPs from the slow set were removed. No consideration for
168 SNP frequency was applied. We used the downloaded VCF files of 1KG to generate the
169 genotyping data.
170

171 **Slow evolving SNPs.** The identification of slow evolving proteins and their associated
172 SNPs were as previously described [59]. Briefly, to obtain non-synonymous SNPs

173 located in the slowest evolving genes, we collected the whole genome protein data of
174 *Homo sapiens* (version 36.3) and M*acaca mulatta* (version 1) from the NCBI ftp site and
175 then compared the human protein to the monkey protein using local BLASTP program at
176 a cut-off of 1E-10. We only retained one human protein with multiple isoforms and chose
177 the monkey protein with the most significant E-value as the orthologous counterpart of
178 each human protein. The aligned proteins were ranked by percentage identities. Proteins
179 that show the highest identity between human and monkey were considered the slowest
180 evolving (including 423 genes > 304 amino acid in length with 100% identity and 178
181 genes > 1102 amino acid in length with 99% identity between monkey and human). We
182 downloaded the 1KG phase 3 data and assigned SNP categories using ANNOVAR. We
183 then picked out the nonsyn SNPs located in the slow evolving set of genes from the
184 downloaded VCF files of 1KG (Supplementary Table S2).
185
186 **Calling SNPs from genome sequences.** We used publically available software
187 SAMTOOLS, GATK, and VCFTOOLS to call SNPs from either downloaded BAM files or
188 BAM files we generated based on downloaded fastq data [64-66].
189
190 **Analysis of shared and unique SNPs.** Shared and unique SNPs were identified by
191 using downloaded allele frequency information from 1KG. Classification of SNPs into
192 different categories such as intergenic, intron etc were also according to downloaded
193 VCF files of 1KG. For some studies, we also calculated the alternative allele frequency of
194 a SNP in a randomly divided human group by using the PLINK software [67].
195
196 **Imputation.** Because commonly used SNPs chips for genome wide genotyping have
197 only a fraction of the slow SNPs defined here, we performed imputation to obtain more

198 coverage of the slow SNPs on the South Asian and Oceanian datasets of Pugach et al
199 (2013). We used the SHAPEIT2 software to do phasing for the SNP chip data [68] and
200 the IMPUTE2 software to impute based on 1KG data [69].
201
202 **Genetic distance calculation.** We used the custom software, dist, to calculate pairwise
203 genetic distance (PGD) or number of SNP mismatches from SNP data [59]. This software
204 is freely available at https://github.com/health1987/dist and has been described in detail
205 in previous publications [41, 70]. We obtained PGD for each of the 25 human groups in
206 the 1KG data and obtained average PGD per group for groups within each of the 5 major
207 continents as represented by the 1KG. We excluded highly admixed groups ASW, ACB,
208 CLM, and PUR in calculating the continental average.
209
210 **Principal component analysis (PCA).** PCA is commonly used to discover genetic
211 structures of a population. We utilized GCTA to analyze data in the PLINK binary PED
212 format to generate two files (*.eigenvec and *.eigenva). We drew PCA plot using
213 *.eigenvec file [67, 71]. One sample BEB_HG04131 was found on PC2-PC3 plot to be an
214 outlier and was hence excluded from the PC analysis and most distance calculations
215 presented here.
216
217 **Other methods.** Other common statistical methods used were Student's t test, chi
218 square test, and Fisher's exact test, 2 tailed.
219

Results

Contrast between fast and slow evolving DNAs in genetic diversity patterns

Different human groups are well known to share ~85% of common genetic variations [72]. However, sharing may not necessarily mean genetic exchanges or common ancestry as assumed by the field as saturation or parallel mutations could also explain it. These two explanations could be distinguished by asking whether the fractions of shared SNPs are similarly distributed in the fast versus the slow evolving sequences. Since the majority of human genomes are made of non-coding sequences and hence faster evolving relatively to coding sequences, we randomly selected from the 1000 genomes project phase 3 (1KG) data a set of 255K SNPs to represent the fast evolving SNPs or the average genome wide variation (Supplementary Table S1) [73]. To find the slow evolving SNPs, we first identified the slow evolving proteins by aligning human and Macaca proteomes and then selected only the non-synonymous (nonsyn) SNPs located in these proteins as previously described [59]. Proteins that show the highest identity between human and monkey were considered the slowest evolving, including 423 genes > 304 amino acid in length with 100% identity and 178 genes > 1102 amino acid in length with 99% identity between monkey and human. We downloaded 1KG data and obtained a list of ~15K nonsyn SNPs located in these slow evolving proteins as our slow set of SNPs (Supplementary Table S2 and S3).

To test the amount of sharing, we examined the SNP frequency files from 1KG. For the three human groups, African (AFR), East Asian (ASN), and European (EUR), we considered a SNP as shared if it has frequency > 0 in more than one group and unique if it is present in only one group. We examined 3 different sets of SNPs, the slow set as defined above, syn SNPs in the slow genes as defined above (Supplementary Table S3), and the random set as defined above. The results showed a clear pattern of more sharing

in fast evolving SNPs (Table 1), indicating saturation level of genetic diversity, which further confirmed previous findings of higher genetic diversity in patients of complex diseases relative to normal matched controls [40, 48, 59]. That the observed sharing (24%) in fast SNPs was lower than the 85% for common SNPs was because we did not filter the SNPs by frequency and hence there were many private or low frequency SNPs in our set.

As fast evolving SNPs are enriched with common SNPs, which may in part explain more sharing as found above, we next studied only rare SNPs with the same allele frequency in a racial group for either fast or slow evolving variants. For each major racial group, we selected a set of SNPs that have the alternative allele appearing only twice in the group (either two heterozygous or one homozygous), as singleton SNPs may be sequencing errors and more common SNPs may be too few for informative studies. We divided the set into several subsets based on evolutionary rates, intergenic, intron, synonymus or missense changes in fast or slow evolving proteins with SNP numbers of a racial group in each subset ranging 281-1376757. Fast evolving proteins here include all those that do not belong to the slow group as defined above. We then examined the fractions of these SNPs that were present in other racial groups. The results showed higher shared fractions for SNPs in fast evolving DNAs except for sharing between AMR (American) and EUR (Fig. 1). As AMR is well known to be highly admixed with Europeans, the observed equal sharing in EUR of SNPs regardless of evolutionary rates in fact made sense and validated our approach. We further validated it by randomly dividing the EAS group into two subgroups EAS1 and EAS2. From the random 255K fast SNPs and the slow SNP set, we selected SNPs that appeared in EAS1 only twice and examined their sharing in EAS2 and other racial groups. The results showed equal sharing of fast and slow evolving SNPs in EAS2 but more sharing of fast evolving SNPs in other racial groups (Fig. 1F). Thus, these results invalidated the present notion of a very recent

common ancestor (<100 ky) and large amount of admixture as suggested by the Out of Africa model and support a more distant common ancestor and limited admixture as advocated by the multiregional model.

We next examined the genetic diversity levels within each of the 5 major human groups as sampled by 1KG, AFR, AMR (American), ASN, EUR, and SAS (South Asians), by calculating the average pairwise genetic distance (PGD) per group in different types of SNPs, including the slow set and the random set as defined above, and a stop codon gain/loss set (Fig. 2). In our analysis here, we have excluded 4 highly admixed groups, Americans of African Ancestry in SW USA (ASW), African Caribbeans in Barbados (ACB), Colombians from Medellin Colombia (CLM), and Puerto Ricans from Puerto Rico (PUR). Since certain deleterious SNPs may exist only in heterozygous (het) state rather than homozygous (hom) state, we calculated, in addition to total PGD contributed by both het and hom differences, also the hom PGD resulting from hom mismatches that should better represent neutral diversity. As shown in Fig. 2, hom PGD showed different pattern from total PGD only in the slow SNPs, with the hom PGD level of AFR below the average of five groups while that of AMR being the highest. Remarkably, the stop codon set showed similar pattern as the random set, with AFR having the largest PGD. This indicates functionality rather than neutrality for the random set since stop codon SNPs are definitely functional given its dramatic effect on protein structure [61] and since a neutral set is expected to be very different from the stop codon set. To verify the results of stop codon SNPs, we also found similar PGD pattern in a set of splicing site gain/loss SNPs that are also expected to be functional (Supplementary Information S1 and Fig. S1A-B). Overall, these results showed Europeans with the lowest diversity in stop codon and splicing SNPs and East Asians with the lowest diversity in the random set ($P<0.01$). Africans have the highest genetic diversity levels in all types of non-neutral SNPs

295 examined as well as the random set (P<0.01), thereby deeming the Out of Africa model
296 untenable as it is based on the now disapproved assumption of neutrality for a random
297 set of genome wide SNPs.

298 To confirm if we have made the appropriate cut-off in selecting the slow SNPs as our
299 phylogeny-informative set of neutral SNPs, we verified that the next set of just slightly
300 less conserved nonsyn SNPs (total number ~13.7K, Supplementary Table S4) within 361
301 autosomal proteins already behaved like the random set or the stop codon set (800-1102
302 aa in length with identity between human and monkey >99% but <100%) (Supplementary
303 Information S1, Fig. S1 C-D). Furthermore, syn SNPs within the slow set of proteins as
304 defined above (Supplementary Table S3) gave PGD patterns similar to the stop codon
305 SNPs but unlike the nonsyn SNPs within the same set of proteins (Supplementary
306 Information S1, Fig. S1 E-F). Finally, we confirmed that these slow evolving proteins still
307 have neutral nonsyn variations that are not under natural selection by showing that these
308 proteins have fewer overlap or recurrent mutation sites than relatively faster evolving
309 proteins (Supplementary Information S2 and Table S5), and that known positively
310 selected genes are faster evolving (Supplementary Information S3). Together, these
311 results suggest that only hom distance calculated from the slow nonsyn SNPs, hereafter
312 referred as the slow SNPs, can be informative to phylogenetic inferences. To view slow
313 nonsyn SNPs as neutral seems counter to expectations based on the neutral theory but
314 in fact makes sense under the MGD perspective as fast SNPs are at saturation levels that
315 are maintained by selection and play adaptive roles in response to fast changing
316 environments [62].

317 Using hom distance measured by slow SNPs, we found, as expected, Africans as the
318 outgroup to the other 4 groups as sampled in 1KG because the non-African groups are
319 closer to each other than to Africans (Supplementary Fig. S2A). Also as expected from *a*

priori reasoning but not from the existing model, Africans are closer to each other than to non-Africans. However, for the random set of 255K SNPs, total distance within Africans was similar to that between Africans and non-Africans, which is well known from previous studies and reflects saturation as we now realized from the MGD theory (Supplementary Fig. S2B). This result also established the maximum genetic equidistance phenomenon, previously known only at the inter-species level, at the intra-species level where groups with lower MGD are equidistant to the group with the highest MGD with the distance being equal to the MGD of the highest MGD group. The result independently confirms the difference between slow and fast SNPs and the fact that fast SNPs are at saturation level of genetic diversity.

We also reexamined the claim of an inverse correlation between intra-population genetic diversity and distance to Africa, purporting to support the Out of Africa model [74-76]. Clearly, such correlation was not observed for the slow SNPs where Americans have the highest diversity among all groups (Fig. 2). Neither was it found for the 255K fast SNPs where Americans were more diverse or have higher PGD and higher numbers of het alleles than East Asians; South Europeans (IBS and TSI) were more diverse than North Europeans (CEU and FIN); South Asians were more diverse than Europeans (Supplementary Fig. S3). Different selection pressures on the fast SNPs for different populations may explain these patterns. Ascertainment bias plus overlooking saturation may explain previous conclusions.

Divergence time between major human groups

To estimate the time of separation between major human groups, we determined the mutation rate of the slow evolving genes. We found 34 informative genes in the 178 slow evolving genes as defined above that showed gap-less alignment in any pair of

345 comparisons among humans, chimpanzees, orangutans, and monkeys (Supplementary
346 Table S6). Assuming gorilla and orangutan contributed similarly to their genetic distance
347 since their split 12 myr as inferred from the fossil records [77], we obtained a gorilla or
348 orangutan mutation rate of 0.000173 aa per myr per aa for the 34 genes (47628 aa).
349 Given a distance of 0.00385 aa per aa between human and orangutan and their
350 separation time of 17.6 myr [45], we used the formula 0.00385 = R_{human} x 17.6 + 0.000173
351 x 17.6 to obtain the human mutation rate as 4.57E-5 aa per myr per aa, which is 3.88
352 times slower than orangutan's. Given this mutation rate and the distance matrix (total
353 distance including both het and hom distances) as shown in Table 2 (only the largest
354 distance among groups are shown), we estimated the split time between ESN (Esen in
355 Nigeria) and GBR (British in England and Scotland) as 1.92 myr, consistent with the
356 known first migration out of Africa for the Homo species as shown by the fossil records.
357 The split between ESN and CHS (Southern Han Chinese) was similar or slightly shorter
358 at 1.86 myr and not significantly different from that between ESN and GBR. In fact, using
359 hom distance as measured by the slow SNPs which represent neutral distance better,
360 ESN is slightly closer to CHS (14.87) than to GBR (14.93). We only used the largest
361 distance between groups, which was between ESN and GBR, to calculate the time in
362 order to be more precise. Since admixture was common, shorter distances between
363 some pairs of groups may be a result of gene flow and hence not reflect true separation
364 time.
365
366 **Y chromosome phylogeny**
367 The existing Y phylogenetic tree depends on inferring derived alleles and in turn
368 requires the validity of the infinite site assumption, which means no maximum genetic
369 distance and no recurrent mutations. However, this assumption can be proven invalid

even just by the existing Y tree itself, since the tree shows numerous recurrent mutations that were simply ignored without valid reasons (Supplementary Table S7), especially for the early branches with some such as KxLT and HIJK contradicted by as many as 50% of all relevant SNPs [78]. That these self-contradictions mostly occurred for the early African branches such as BT and CT but rarely for the terminal Eurasian ones indicates the unrealistic nature of these early branches. Also, while haplotypes with few sequence variations from the ancestor of F, C, D, E, NO, KxLT, or K are routinely found in present day people, none could be found for these early branches. The branching pattern in Africans often involves one branch, such as A00, with few or no sub-branches while the other branch A0-T accounting for all of the remaining haplotypes on Earth, which is odd and against branching patterns known in experimental biology such as the embryonic differentiation into three layers with each layer giving rise to multiple cell types.

Given functionality for genome wide autosomal SNPs as discussed above, it is easily inferred that most SNPs in Y chr are also non-neutral. Evidence for extreme natural selection on Y is also known [79, 80]. We therefore redrew the Y tree based on shared alleles (rather than on derived alleles), which may mean common physiology more than common adaptations if physiology is the chief determinant of MGD. Using previously defined haplotypes for 1KG samples (Supplementary Table S8) and 58251 cleanly called SNPs (no individual with uncalled SNPs, Supplementary Table S9) [81], we found a major megahaplogroup ABCDE (Fig. 3). Megahaplotype F, defined as lacking any mutations that define other haplotypes, is the ancestor. All F-like or F* haplotypes sequenced so far are partial ABCDE carrying 4 (Lahu_HGDP01320), 13 (Malay_SSM072), or 14 (KHV_HG02040) of the 151 mutations that group ABCDE (Fig. 3) [81-83]. The F* haplotype is most common in East Asia, present in 5 of 7 (71.4%) Lahu males in Yunnan of South West China [84], 10-15% of Han and other minority Chinese,

and low percentages (<10%) in South Asians and French. Furthermore, the top 4 individuals among 1KG closest to the ~45 ky old Western Siberian Ust'-Ishim who carried NO haplotype and was expected to be most like the AMH ancestor were all East Asians with Asian haplotypes F and O (F2 in KHV_HG02040, O2 in CHB_NA18534, O3 in CHS_HG00559, O3 in KHV_HG02088), indicating least deviation from the ancestor for Asian haplotypes [14]. Those three O type East Asian individuals also were the closest to the three F* carrying individuals above. These results suggest the origin of F in East Asia with subsequent migration to other regions of the world (Supplementary Fig. S4).

In our tree, alleles previously used to define BT and CT now merely represent alleles associated with the original F ancestor. The AB grouping in our case makes more sense with phenotypes than the BT grouping since it groups African B with African A rather than with the CT group containing mostly Eurasians. The key feature of our tree is that every haplotype besides the original F is associated with haplotype specific SNPs and there are no inconsistent SNPs. Such self-consistency alone would qualify it as more correct than the self-inconsistent tree rooted in Africa.

A real haplotype should exist in a way that has only its own haplotype specific SNPs plus private SNPs. While it is more likely to be the case for terminal haplotypes, it is not impossible for ancestral haplotypes close to the root of the tree, which could be used to distinguish two different competing classifications on an ancestral haplotype. One of the major differences between our Out of Asia tree and the Out of Africa tree is the position of haplotype C, which belongs to ABCDE in the former and CT in the latter. The ABCDE haplotype is closer to the root in the Asia tree (among the first to branch out from the root) than CT is in the Africa tree (2nd to branch out from the root). Hence, relative to ABCDE, CT should have a higher chance to be like a more terminal haplotype. The number of CT defining SNPs is larger than ABCDE (264 vs 151, with additional 50 SNPs contradicting

CT), which should also make CT more like a terminal branch. In reality, however, one found the exact opposite. People with only ABCDE specific SNPs plus private SNPs have been found as in present day people Lahu_HGDP01320, Malay_SSM072, and KHV_HG02040 (Fig. 3). That these Y chr had only a portion of ABCDE defining SNPs is consistent with the dynamics of the ancient appearance of ABCDE. Also consistent with the ancestral status of ABCDE, the 38.7 ky old Kostenki14 had 83 among 84 informative SNPs supporting it as ABCDE and 88/92 as C, and the 30.6 ky Vestonice43 had 19/20 as ABCDE and 20/22 as C [85, 86]. In contrast, no one with only CT specific alleles plus private SNPs has been found to exist today or in the past. All ancient Europeans known to be CT and lacking alleles for downstream haplotypes were in fact missing informative sites for at least some haplotypes due to incomplete coverage in sequencing [85, 86]. These findings strongly support grouping C in ABCDE rather than in CT, thereby invaliding the Out of Africa tree.

mtDNA phylogeny

The existing mtDNA phylogenetic tree has exactly the same problems as the existing Y tree as discussed above. mtDNA has also been found to be under strong selection relative to more slowly evolving nuclear DNAs, consistent with the MGD theory [34, 87, 88]. Based on previously defined mtDNA haplotypes for 1KG (Supplementary Table S8)[81], we redrew the mtDNA tree using slow evolving SNPs, which alter amino acids or RNA sequences (Fig. 4A, Supplementary Information S4, Supplementary Fig. S5, Supplementary Table S10). Fast SNPs are more involved in adaptation to fast changing environments and should not be used whenever possible. Two lines of evidence suggest haplogroup R as the ancestor of all modern haplogroups. First, ancient humans are expected to be closer to the ancestor and the oldest AMH, Ust'-Ishim, carried the R*

haplotype [14]. Second, R0 is the least differentiated haplotype and closest to the ancient haplotype in Ust'-Ishim (Fig. 4B). That R0 is most common in Chinese among 1KG samples indicates origin of R in East Asia (Fig. 4B) and subsequent diversification in other regions of the world (Supplementary Fig. S6).

Unlike Y, mtDNA diversification as defined by slow SNPs here is far more star like with multiple parallel haplotypes and few hierarchical structures (Fig. 4A, Supplementary Fig. S5), which is expected from the vast difference in the possible number of offspring between males and females. Many female individuals with R0 might each serve as an ancestor of a specific haplotype within R or N haplogroup, and R is not a sub-branch of N. M also directly derived from R0. L and numerous M subtypes shared a few defining SNPs.

To confirm M giving rise to L, we examined mtDNA distance between African (YRI: Yoruba in Ibadan, Nigeria) L and South Asian (BEB: Bengali from Bangladesh) M and found L3e to be the closest to M (Fig. 4C). Also, M of BEB or GIH is closer to L3e than M of CHS, indicating a more direct role for BEB or GIH in dispersing AMH mtDNA into Africa and a Southern route into Africa. Consistently, in autosome distance, BEB or GIH with M haplotype were closer to Africans than those with N (including R) haplotype (Fig. 4D), despite the fact that people with M had larger autosomal nucleotide diversity than those with N (PGD: M_BEB = 8.59, N_BEB = 7.9, M_GIH = 8.42, N_GIH = 8.36). M may be closer to the common ancestor of M and L since all L types except L3 are at least 2 fixed slow SNPs (769, 1018) away from the common ancestor. There might be a time when there were multiple M types with no L, and then one of the M types with mutations in 769 and 1018 sites became the common ancestor of all L types except L3.

Neanderthals and Denisovans

If major human groups have separated ~2 myr ago with region specific features developed not long after separation such as shovel shaped teeth in *H. erectus* from China (Yuanmou man and Peking man), Neanderthals and Denisovans with features more modern than *H. erectus* should be expected to belong to one of the modern groups today. However, previous studies have found Neanderthals to be outgroup to AMH and used D-statistics to show Neanderthal gene flow into non-Africans but oddly not Africans [11, 12]. The assumption of D-statistics is that all modern groups are equidistant to chimpanzees so that presence of derived alleles (different from chimpanzees) was due to gene flow from Neanderthal. If in fact Africans are closer to chimpanzees or carrying more ancestral alleles in general due to perhaps common adaptation to the Africa environment, the conclusion of gene flow into non-Africans would become invalid. We examined this by measuring genetic distance between 1KG and 10 previously sequenced chimpanzee genomes [89]. Using the random 255K SNPs set, we found closer hom distance between Africans and chimpanzees than between non-Africans and chimpanzees (Supplementary Fig. S7). As presence of Neanderthal derived alleles in a non-African are mostly in het state [14], which could be observed to be biased toward non-Africans only if Africans are in hom ancestral state, the fact of more hom ancestral alleles in Africans therefore deems invalid the previous finding of Neanderthal gene flow into non-Africans. Furthermore, as already noted above for Y and mtDNA trees, the finding of saturated level of genetic diversity makes the infinite site assumption invalid, which in turn makes the assignment of ancestral and derived alleles unrealistic. That the D statistics method may not be appropriate to detect Neanderthal introgression has also been independently found by others [90]. Thus, the relationship between Neanderthals/Denisovans and present day populations remains to be determined.

Making use of the published Neanderthal and Denisovan genomes [11, 12, 91], we calculated the genetic distance in slow SNPs between 1KG and Neanderthals (Altai, Vindija 33.16, 33.25, 33.26, and Mezmaiskaya1) or Denisovan3 (Fig. 5A). These ancient genomes showed closer distance to Africans except Vi33.25 to ASN and Vi33.26 to AMR. Denisovan3 was closer to Africans than Neanderthals were (Fig. 5A). In contrast to the Neanderthals and Denisovan3, their near contemporary AMH Ust'-Ishim from Western Siberia was closest to SAS (Fig. 5A) and grouped with SAS in PCA plots (4B-G). We also studied the more recently reported Neanderthal genomes of Vi33.19, Vi87, Les Z4, GoyetQ56, and Mezmaiskaya2 [6, 92] and found them to be also closest to AFR except Vi87 who was equally related to AFR and SAS (Supplementary Fig. S8). Using the slow SNPs but not the non-informative SNPs, we also found that two Neanderthals from the same location, Mezmaiskaya 1 and 2, but separated for ~20 ky were in fact closest to each other than either was to any other Neanderthals or ancient DNAs of modern humans found anywhere in the world (Supplementary Fig. S9), thus confirming regional continuity and invalidating the previous conclusion of Neanderthal population turnovers [6, 92]. These results suggest that Neanderthals and Denisovans were Africans who migrated into Eurasia and admixed with local non-Africans. The observations of an East Asian like Neanderthal (Vi33.25) in Europe at >45,000 years ago and of a South Asian like Western Siberian (Ust'-Ishim) from ~45,000 years ago indicates migration of Asians into Europe around the time of AMH origin in South East Asia.

Origins of Negritos and Aboriginal Australians

The Andamanese and the African pygmies seem obviously related in multiple aspects, including traits, Y relationship with the African megahaplogroup ABDE, and mtDNA haplotype M being closely related to African L. However, previous studies have

519 found Andamanese to be even more genetically distant to Africans than other Eurasians
520 [36]. Using the published genomes of 10 individuals from the Jarawa (JAR) and Onge
521 (ONG) populations in the Andaman Islands [36], we found that Andamanese are
522 relatively closer to Africans or have lower AFR/SAS(-BEB) distance ratio than other
523 nearby populations such as BEB, with ONG more so than JAR, consistent with the known
524 less admixture in ONG relative to JAR (Fig. 6A). PCA plots also showed Andamanese
525 closer to Africans than all five populations of SAS (Fig. 6B). Relative to the distance to
526 SAS, ONG showed smaller distance to Mbuti than to San or other Africans examined
527 except LWK (Fig. 6C). The Mbuti group here consists of 4 published genomes from the
528 Simons project [82] and the San group consists of 2 published genomes [93]. Given that
529 Andamanese were closer to Africans than other Indians were (Fig. 6A) but Mbuti pygmies
530 were not closer to Andamanese than some other Africans were, it can be inferred that
531 Andamanese came from Mbuti rather than the opposite.

532 The African affinity of Neanderthals prompted us to examine the distance between
533 Neanderthals (with relatively higher coverage genomes, Vi33.16 and Altai) and several
534 different Indian populations (ONG, JAR, BEB, and GIH) to see if ONG might have come
535 from Neanderthals or related humans. Relative to the distance to the ~4500 year old
536 African Mota [94], ONG was closer to Neanderthals Vi33.16 and Altai, as well as to
537 Ust'-Ishim who was known to have large amount of Neanderthal admixture, than other
538 Indians were (Fig. 6D). Also, if Andamanese came from Neanderthals, Neanderthals
539 should be closer to Mbuti than to San and other Africans, since Andamanese are closer
540 to Mbuti than to San (Fig. 6C). This was indeed the case for the Altai individual who had
541 high coverage genome for this analysis to be informative (Fig. 6E).

542 Since different Negrito groups in South Asia share similar traits, one expects them to
543 be genetically related. The new Y tree grouping C with ABDE further suggests a common

544 ancestry for different Negrito groups since the C haplotype is common in certain Negrito
545 groups in Philippines while D is common in some others such as Onge. We therefore
546 made use of a previously published SNPs genotyping data for a number of Oceanian
547 groups including the Negrito group Mamanwa and its neighboring group Manobo in
548 Philippines [63]. We measured the ONG/JAR distance ratio to look for the group that is
549 closest to ONG relative to its neighbor JAR and the Mamanwa/Manobo distance ratio to
550 look for the group closest to Mamanwa relative to its neighbor Manobo. Of the 13 groups
551 examined, Mamanwa showed the smallest ONG/JAR distance ratio besides ONG;
552 conversely, ONG showed the smallest Mamanwa/Manobo distance ratio besides
553 Mamanwa (Supplementary Fig. S10). These results suggest that the two Negrito groups
554 are more closely related to each other than either is to other groups as examined here.
555 We also examined the Aboriginal Australian (AUA) samples in the Pugach et al
556 (2013) dataset and a previously published ~100 year old AUA (AUA_100yr) who was
557 unlikely to have admixed with European colonizers [95]. These AUA samples showed
558 lower Mamanwa/Manobo ratio than other Oceanians (Supplementary Fig. S11). The AUA
559 samples from Pugach et al (2013) also showed lower AFR/ASN ratio than other
560 Oceanians, representing 68% of the average ratio for the Oceanians (excluding AUA and
561 NGH or New Guinea Highlanders). To examine if the African component of AUA had
562 come from Neanderthals, we calculated the Altai/ASN distance ratio of AUA and found it
563 to be 64% of the average ratio for the Oceanians in Pugach et al (2013) dataset, which
564 was significantly lower than the 68% found for AFR/ASN ratio, indicating closer
565 relationship of AUA to Altai than to AFR. These results showed similarity between AUA
566 and Negritos, indicating similar ancestry in Neanderthals and Denisovans.
567
568 **Testing the out of East Asia model**

We next tested certain obvious predictions of the out of East Asia model. First, the model predicts lower diversity in people directly associated with the original AMH and higher diversity in people resulting from admixture of AMH with archaic humans. We calculated the hom PGD in slow SNPs as well as het numbers for each of the 25 groups totaling 2534 individuals in 1KG. The lowest hom PGD level was found in LWK followed by slightly higher level in CHS (Supplementary Fig. S12A). However, LWK has significantly higher numbers of het than CHS (Supplementary Fig. S12B). As high level heterozygosity indicates high genetic diversity and would reduce hom distance, it is likely that CHS has lower genetic diversity than LWK. We further found that within CHS (made of 72 individuals from Hunan and 36 from Fujian), Hunan samples have lower hom PGD and het numbers than Fujian samples (Supplementary Fig. S12CD). These results indicate that CHS, in particular Hunan people, have lowest genetic diversity levels among the 25 groups in 1KG. Given that known admixed groups such as MXL and PUR showed the highest genetic diversity or PGD (Supplementary Fig. S12A), it may be inferred that CHS or Hunan people may have the least amount of admixture and hence represent the original AMH group, at least among the 25 groups sampled here. That Africans, as human ancestor from ~2 myr ago according to the multiregional model, did not show the highest genetic diversity level may seem unexpected but is in fact consistent with a key role for admixtures as claimed by the multiregional model as well as our out of East Asia model here. The original AMH group should have low admixture with archaic people in order for evolution into AMH to be possible since admixture may reverse AMH back to the archaic state.

Second, we would expect Southern East Asian groups to be closer to Africans. Although CHS represent samples collected from Southern China (Hunan and Fujian), while CHB (Han Chinese in Beijing) samples were from Northern China (Beijing), both in

fact contain Southern and Northern Chinese. We therefore made use of the Hunan versus Fujian samples in CHS, where Fujian people are known to be mostly migrants from Central North China during the West Jin, Tang, and Song dynasties. We calculated the distance of each group to Hunan or Fujian and obtained the Hunan/Fujian distance ratio of each group. Consistently, groups known to have more Northern Chinese admixtures, such as CHB, MXL (Mexican Ancestry from Los Angeles), PEL (Peruvians from Lima), JPT (Japanese in Tokyo), had higher Hunan/Fujian distance ratio than Southern groups such as CDX (Chinese Dai in Xishuangbanna) and KHV (Kinh in Ho Chi Minh City, Vietnam) (Fig. 7A). Of note, FIN is closest to Hunan people among EUR groups ($P<0.01$), suggesting that North Western migrations of Southern Chinese during the first wave of AMH dispersal from Hunan area may have contributed to the ancestry of FIN. Consistently, Western hunter-gatherers from the Paleolithic age also showed closer distance to Hunan (manuscript in preparation). All AFR groups showed lower Hunan/Fujian distance ratio than non-Africans with LWK in East Africa the lowest, consistent with migration of Southern Chinese into Africa and into the Horn of Africa first. That non-Africans had more Fujian admixtures is consistent with known migrations of Northern East Asians into both the West and the America in more recent times during the Neolithic and Bronze ages. We further found Hunan people to be relatively closer to Africans than other South East Asians such as CDX and KHV (Fig. 7B), indicating origin of AMH more likely in Hunan relative to other nearby regions.

Third, as migration of AMH from Hunan via the Southern route to East Africa must cross the Indian subcontinent, one would expect closer relationship with Africans for groups within South Asia that are more related to Chinese relative to those more related to Europeans or more Southern relative to more Northern. Indeed, relative to Fujian people, the distance of different Indian groups to Africans follows exactly their direct

619 distance to Hunan people, as well as their direct distance to LWK, in the order of
620 increasing distance, BEB, GIH, ITU, STU, and PJL (Fig. 7BC). Also, Gujarati Indians
621 (GIH) in Western India is closer to Africans than Punjabi people from Northern Pakistan
622 (PJL) (Fig. 7B). Consistently, relative to PJL, both BEB and GIH are closer to Africans
623 with BEB closer than GIH (Fig. 7B). The observation of lower BEB/Fujian distance ratio
624 than Hunan/Fujian is consistent with Indians being in general closer to Africans than East
625 Asians (Fig. 7D) and being more recent ancestors to Africans than East Asians based on
626 the migration route of the out of East Asia model.
627 Fourth, we hypothesized that the branching process of Y may involve AMH
628 hybridization with archaic humans and subsequent adaptive co-evolution of Y and
629 admixed autosomes. As the first major split resulted in ABCDE, G, and HIJK haplogroups,
630 we tested whether the ABCDE megahaplogroup, whose sub-branches are mostly found
631 in Africans and South Asians or Oceanians with African like features, may have resulted
632 from admixture of F AMH with admixed archaic Africans such as Neanderthals who may
633 have migrated to South East Asia. We examined the Y chr sequences of three
634 Neanderthals [6, 7], and found all to share alleles at informative sites with haplotype A0, A,
635 AB and ABCDE but not with any non-ABCDE haplotypes (Table 3). These results further
636 confirmed the African affinity of Neanderthals as shown by autosome analyses above and
637 indicated that admixture of F AMH with Neanderthals may have resulted in African-like
638 descendants with ABCDE megahaplotype who largely preferred to live in the Southern
639 hemisphere. Consistently, East Asians (JPT) with D or C haplotype showed closer
640 autosomal distance to Andamanese (also with D haplotype) or African MSL (with E
641 haplotype) than those with O haplotype did (Supplementary Fig. S13).
642
643 Fifth, to similarly test whether mtDNA diversification from the original ancestor type to
644 more African type may involve AMH hybridization with archaic humans, we examined the

645 distance between archaic and modern mtDNAs in slow SNPs (Supplementary Table S10).
646 Although archaic mtDNAs were nearly equidistant as measured by mtDNA genome wide
647 SNPs to the modern group consisting of Europeans (CEU: Utah Residents with Northern
648 and Western European Ancestry), East Asians (CHS), and Africans (LWK: Luhya in
649 Webuye, Kenya), they were closer to Africans in SNPs found in archaic humans (sites
650 that differ between archaic mtDNA and the rCRS), indicating more sharing of archaic
651 alleles in Africans (Supplementary Fig. S14). This is likely due to independent adaptive
652 mutations since archaic mtDNAs are outgroups to modern mtDNAs as previous studies
653 have shown. We also confirmed it by showing that the average distance between archaic
654 and modern mtDNAs were larger than that within modern mtDNAs (Supplementary Fig.
655 S15A). The archaic mtDNAs are at least of two types, with Neanderthal Vi33.16 and Altai
656 belonging to one type or being close to each other than to other archaic mtDNAs while
657 Denisovan and Heidelbergensis [96] belonging to another type (Supplementary Fig.
658 S15B). Such results support the notion of multiple turnover events in mtDNA types in the
659 past ~2 myr of human evolution.

660 Our results here confirmed the first mtDNA tree ever built that placed the original
661 AMH type (type 1 morph) in East Asia [29]. Our original type here is defined by the major
662 alleles of 6 slow SNPs, 750, 1438, 2706, 8860, 14766, 15326 (Vi33.16 and Altai have all
663 6 except 2706, Denisovan has 14766 and 15326, and Heidelbergensis has 14766 and
664 1438). The earliest AMH Ust'-Ishim had the major alleles of these 6 SNPs and no
665 additional slow SNPs. Mutation at 14766 defines V and VH and further mutation at 2706
666 defines H (Supplementary Fig. S5A). All other haplotypes overwhelmingly carry the major
667 alleles of these 6 sites plus a few additional less common slow SNPs. We calculated the
668 number of slow SNPs in each haplotype present in the 1KG and found R0 of Chinese to

669 have the least amount among all non VH haplotypes, supporting R0 as the original type
670 (Supplementary Fig. S16 and Table S11).

671 We examined whether the amount of allele sharing with archaic mtDNAs supports
672 the above results linking archaic humans with South Asians or Oceanians. Using
673 PhyloTree17 and Mitoweb data, we identified all non-L haplotypes that share alleles with
674 archaic mtDNAs and calculated the number of shared alleles in each haplotype
675 (Supplementary Table S12, Supplementary Information S4). Among R and N haplotypes,
676 P4b1, R7a, J1b, and W3 were the most enriched with archaic alleles (4 alleles), which are
677 common in Oceanians or Arabians/Caucasians/South Asians. Among M haplotypes
678 (excluding L), G4, M17a, M27a, M76a, and M7b1a2a1a were the most enriched with
679 archaic alleles (5 alleles), which are common in native Japanese, South East Asians,
680 Papuans, and Indians. The M haplotype with the least amount of archaic alleles is D5 (1
681 allele), which is most common in Chinese.

682 We next calculated the distance of each modern haplotype in 1KG to archaic
683 mtDNAs as measured by either slow or fast SNPs found in archaic genomes, and
684 assigned each a distance ranking (Supplementary Fig. S17 and S18, Supplementary
685 Table S13). Haplotype L0 commonly found in San people was the closest to archaic
686 mtDNAs, consistent with the Neanderthal affinity with Y haplotype A. Haplotype G2 of
687 JPT and M5a of SAS were top ranked non-L haplotypes in distance to Heidelbergensis in
688 slow SNPs, consistent with the known admixed phenotypes of native Japanese and S.
689 Asians. Relative to G2 that is common in South Asians and Tibetans, G1 is common in
690 Russian far East and consistently closer to Altai in fast SNPs, indicating an adaptive role
691 for fast SNPs. Consistent with the expected routes of human entry into America, D1
692 common in Amerindians and Paleoamericans is closest to Altai in both slow and fast
693 SNPs among the 4 archaic mtDNAs. As expected, haplotypes common in African groups

694 such as U6 and L3e are closer in slow SNPs to more African type archaic mtDNAs of
695 Denisovan/Heidelbergensis, whereas those of Amerindians such as D1, D4 and A2 are
696 closer to Neanderthals (Supplementary Table S14). Such analyses further indicate
697 possible effects of archaic admixtures, with G2, I, T, and X2 affected by
698 Denisovan/Heidelbergensis, and J, K2, and W by Neanderthals.

699 We next merged 1KG data with the AUA specific P4b1 and the Andamanese
700 enriched M31a and found these haplotypes ranked among the top 13 in slow distance to
701 Heidelbergensis (but 84th in fast SNPs), just following G and M5a among non-L
702 haplotypes (Supplementary Fig. S19). Furthermore, they are uniquely much closer to
703 Heiderlbergensis/Denisovan than to Altai, consistent with being uniquely related to
704 Denisovan among living people today.

705 As M is defined by 10398G and 8701G, both present in archaic humans, it likely
706 resulted from admixture of R0 with archaic Africans. While the effect of archaic humans
707 can also be observed for some haplotypes within R and N, the M haplogroup may be the
708 most affected as indicated by its defining SNPs and the extensive sharing of alleles
709 between L (now within M) and archaic humans. Consistently, the ~40000 year old
710 Romanian Oase 1 had extensive Neanderthal admixture and carried an unusual N with
711 the 8701G allele, indicating clear Neanderthal effect on its mtDNA [13]. The Oase 1
712 mtDNA may be an intermediate in the transition from N/R to M. A modern N haplotype
713 with 8701G allele is N21 common in Malays.

714

715 **Discussion**

716 We have arrived at a new model of modern human origins based on a more complete
717 understanding of genetic diversity (Fig. 8). While the autosomes in our model are largely
718 consistent with the multiregional hypothesis, the mtDNA and Y have a single origin in

East Asia. We also identified Negritos and Aboriginal Australians as direct descendants of Neanderthals/Denisovans who were African migrants with Eurasian admixtures.

In highly conserved proteins, mutation rate may be inherently slower, as indicated by the lower mutation saturation relative to fast evolving proteins. The nonsyn SNPs in slow genes as defined here are neutral [62]. They are not deleterious and unlike the stop codon and splicing SNPs. They are also not under positive selection as positively selected genes tend to be fast evolving. The random set of SNPs, traditionally viewed as neutral, were here shown in fact to be similar to stop codon and splicing SNPs. Under the neutral theory, the slow and fast SNPs should produce the same phylogenetic tree topology. To the dramatic difference between slow and fast evolving DNAs as shown here, we cannot come up with a meaningful explanation using any known schemes other than the MGD theory [37].

Previous studies using genome wide SNPs (random or fast set) suggest a serial founder effect model resulting from the expansion of modern humans out of Africa, with intra-population diversity decreasing relative to their geographical distance from Africa and diversity between populations increasing with the geographical distances separating them [74-76]. However, such observations could be explained by ascertainment bias and different levels of saturation and selection pressures among groups. The finding here of more shared alleles among populations for fast versus slow SNPs suggests that the large fraction of shared alleles among populations is due to saturation and parallel mutations in fast SNPs, which has also been found in previous studies where higher sharing of alleles was found for microsatellite markers versus autosomal SNPs (microsatellites have higher mutation rates) [75].

We have shown that there are only three major human groups, Africans, East Asians, and Europeans/Indians. PCA plots using slow SNPs showed the first major division was

744 between Africans and others (PC1), the second between East Asians and others (PC2),
745 and third between South Asians and others (PC3). This supports the first split in humans
746 to be between Africans and others, the second split between East Asians and others, and
747 the third between South Asians and others. This indicates a Southern rather than
748 Northern route for the first wave of AMH migration out of East Asia. The oldest AMH
749 Ust'-Ishim clustered with South Asians. Also, the Y haplotype H of Indians diverged
750 before diversification of European haplotypes, which is consistent with our model as well
751 as the non-inhabitability for most parts of Europe during the Last Glacial Period.
752 Aboriginal Australians and the related Negritos, traditionally viewed as the fourth major
753 group, in fact consist of largely European/Indian and African genomes and their unique
754 traits might have come from admixture of incoming Neanderthals with local archaic
755 humans. Our calculation showed that the first major split of humans occurred 1.86-1.92
756 myr ago, well consistent with fossil evidence for the presence of *Homo* in Eurasia and the
757 multiregional model. The coexistence at ~1.76 myr ago in Africa of both Olduwan and
758 Acheulean technologies suggests the coexistence of multiple groups of humans
759 distinguished by separate stone-tool-making behaviors [97, 98]. The sudden appearance
760 of Acheulean technologies and pro-Neanderthals at ~0.5 myr ago in Europe (Sima de los
761 Huesos site of Atapuerca) can now be explained by a more recent out of Africa migration
762 by the ancestors of Mbuti people [99, 100].

763 Mitochondrial DNA (mtDNA) and the non-recombination region of Y chr (NRY) lack
764 recombination and provide records of history that are independent of autosomes. Most
765 SNPs in these DNAs can be proven to be under selection, e.g. certain SNPs or
766 haplotypes of mtDNA or Y chr are known to be related to human diseases or compatibility
767 with nuclear genomes [43, 101-105]. Sharing of alleles of mtDNA or Y chr should mean
768 similar selection, reflecting both environments and physiology or primarily physiology

769 when saturation has been reached. Sharing of physiology should be informative for a
770 phonetic approach of phylogeny. Coevolution of mtDNA, Y, and autosomes has been
771 found by many studies [43, 105-108], which may play a key role in the diversification into
772 multiple haplotypes during AMH radiation from its place of origin to other regions by
773 hybridization with archaic humans. People who have stayed relatively unchanged in
774 physiology and living environments from the ancestor would be expected to have few
775 deviations from the ancestor haplotype and their present day living place would indicate
776 place of origin for the ancestor. It is through such reasoning that we have come to place
777 the origin of modern Y and mtDNA in East Asia or Southern China. Our results showed
778 that groups with the same Y or mtDNA haplotypes are also closer in autosomes and traits.
779 The Y megahaplogroup ABCDE matches with the mtDNA megahaplogroup M. Such *a*
780 *priori* sensible results provide strong independent validation for our new phylogenetic
781 method.

782 Given that most SNPs in Y and mtDNA are not neutral, one cannot use the molecular
783 clock approach to determine the age of the haplotypes except for recent diversifications.
784 That the Y haplotype NO of the ~45,000 year old Ust'-Ishim differs from the putative
785 ancestor F by only ~27 SNPs whereas a present day haplotype could differ from the F
786 ancestor by as much as ~740 SNPs (Fig. 3) indicates that the ancestor F should not be
787 much older than ~45,000 years. This relatively young age is remarkably consistent with
788 the time point for the replacement of Neanderthals by AMH but appears to contradict the
789 oldest AMH fossils in Africa or in Hunan China [21]. However, nearly all AMH fossils older
790 than 40,000 years still have certain archaic features and independent evolution of
791 modern features has been noted to occur periodically over the past 950,000 years since
792 the time of *H. antecessor*[4, 109].

793 The novel concept here of modern replacing archaic versions of Y and mtDNA but
794 not autosomes is key to our model of out of East Asia. The lack of recombination in Y and
795 mtDNA makes this idea biologically inevitable. The fact that Heidelbergensis, Denisovans,
796 Neanderthals, and AMH all have distinct mtDNAs suggests that such replacements may
797 have taken place multiple times in the past. Modern examples consistent with the
798 replacement idea are the dominant presence of Asian Z mtDNA in the Saami people of
799 Northern Europe and the wide presence of Asian Y haplotype N in Finnish, who are
800 otherwise largely indistinguishable from Europeans in both autosomes and traits. Also
801 consistent is the finding of three super-grandfather Y haplotypes in China that are
802 relatively young in age (~5000-7000 years) but account for ~40% of Han Chinese males
803 today [110, 111]. Admixture of incoming Asian AMH with archaic humans in Europe or
804 Africa would lead to haplotype diversification in Y and mtDNA while still maintaining
805 regional specificity in autosomes and hence traits as traits are mostly determined by
806 autosomes. Therefore, the multiregional model is fully compatible with any single origin
807 model of mtDNA or Y chr and there is no real conflict between the timing of autosome
808 diversification and the much more recent appearance of the modern mtDNA and Y.
809 Interbreeding between modern and archaic humans may best explain certain human
810 fossils near or within the era of AMH that show both archaic and modern features, such
811 as *H. naledi* and the 'Red Deer Cave' people [112, 113].
812 The Out of Africa model requires uni-parental DNAs and autosomes to be both
813 modern to qualify as AMH but our model only requires uni-parental DNAs. The Y tree
814 topology shows features not consistent with the Out of Africa model. The Y tree shows
815 haplotype I splitting before R from NO and so I should carry less Asian autosomes than R
816 if the Out of Africa model is true but not if our model is. But, Europeans carrying I and R
817 are not known to be different in autosomes. Also, the Out of Africa model has divergence

818 of Europeans and Asians in the Middle East. However, the Y tree shows West Asia
819 haplotype J splitting after the differentiation of South Asia haplotype G, indicating
820 differentiation of Eurasia types following type F to be in East Asia rather than West Asia,
821 which is unexpected by the existing model. Recent ancient mtDNA studies found the
822 earliest modern R to be close to the earliest fossil date of AMH at 45 ky ago and 5000
823 years older than the earliest N, thus invalidating the existing tree and validating our Out of
824 East Asia tree here [114].

825 The ~45,000-year-old AMH Ust'-Ishim from Siberia was previously found to have left
826 no descendants among present populations and to be more related to East Asians than
827 to Europeans/Indians [14]. However, our results showed this individual as Indians. This
828 discrepancy is to be expected. It has been routinely found as surprising in previous
829 studies on ancient DNAs that there is no genetic continuity between ancient and present
830 day people. Such unexpected anomalies can now be understood as artifacts of using
831 non-informative or fast SNPs which turned over quickly to be adaptive. We have verified
832 this for most ancient European DNAs that while non-informative SNPs placed them as
833 outliers, the slow SNPs as defined here all placed them as indistinguishable from present
834 day Europeans.

835 Our finding of Neanderthals and Denisovans as primarily Africans with Eurasian
836 admixture is well supported by fossil data indicating *H. heidelbergensis*, present in both
837 Africa and Europe, as ancestors of Neanderthals. The taurodont teeth are common in
838 Neanderthals, Heidelbergensis and certain South African fossils [115]. The occipital
839 bunning of Neanderthals are also common in modern Africans [116]. Neanderthals are
840 known to share multiple traits with Europeans such as the prominent shape and size of
841 the nose [35, 117], which supports our finding that Neanderthals were often closer to
842 Europeans than to E. Asians (8/10 examined here). Thus, traits and genotypes are

843 coupled after all as they should, unlike the strange and surprising conclusion of more
844 Neanderthal alleles in East Asians than in Europeans as analyzed by using
845 non-informative fast SNPs [11, 12]. Our result that Denisovan is nearly equally related to
846 East Asians and Europeans (slightly more related to East Asians) is consistent with
847 where Denisovan was found. Seemingly unexpectedly, certain Neanderthals found in
848 Europe was most closely related to Asians (Vi33.25) or Americans (Vi33.26). However,
849 this would be expected if Africans associated with the Neanderthal exit had entered Asia
850 or South Asia via the Northern route from Siberia or possibly a Southern route. The
851 general lack of Neanderthal fossils in this Southern route may reflect the relatively small
852 effort so far invested in this region (with only few *Homo* fossil finds like Narmada from
853 ~200 kya who is broadly classified as *H. heidelbergensis*). Indeed several fossils in China
854 show Neanderthal features such as the inner-ear formation in the ~100 ky old *Xujiayao*
855 and *Xuchang Man* [4, 118-120]. Certain mysterious Southern China fossils such as the 11
856 -15.5 ky old 'Red Deer Cave' people with hybrid features of modern and archaic humans
857 may also be candidates for Asian relatives of Neanderthals, especially considering their
858 taurodont teeth [112]. Early modern human fossils with typical Mongoloid features in
859 South West China (Liujiang, Ziyang, Lijiang, and Chuandong) also have weak occipital
860 buns commonly found in Neanderthals [4, 119, 121]. Thus, although Neanderthals were
861 mostly found in Europe and Middle East, they likely also made their way to North East
862 Asia (Denisovan and Teshik-Tash) and South East Asia [122].
863 Fossils or traits indicating AMH migration from East Asia into Africa or Europe have
864 been noted before. First, native Africans such as Khoisans are well known to have certain
865 East Asian features such as shoveling teeth, epicanthic fold, and lighter skins. Mbuti
866 pygmies look very much like the Andamanese. The much lower frequency of shoveling
867 teeth in African fossils and Khoisan relative to ancient and modern Chinese suggests that

this type of teeth could only originate in China with its African presence due to migration. The type of shoveling teeth found in Neanderthals and Pleistocene *Homo* from Atapuerca-Sima de los Huesos may either be a different type from that of Asians and Africans or come from early disposal of *Homo* from Asia to Europe [123, 124]. Second, a combination of three features has been noted to be region-specific to China fossils with lower frequency also found in North Africa: a non-depressed nasal root, non-projecting perpendicularly oriented nasal bones and facial flatness [125]. Third, Dali man of China (~260 kya) had lower upper facial index and flat nasomolar angle, but these two modern features only first appeared in Europe in Cro Magnons (Xinzhi Wu, personal communication).

The appearance of modern humans should be accompanied by new technologies just as the knife type stone tools were associated with the first appearance of the genus *Homo*. A technology just one step more advanced than stone tools is pottery making. Consistent with our model, the earliest pottery making intended for practical usage was found in Hunan and the neighboring Jiangxi in Southern China at 18,000-20,000 years ago [126, 127]. While future investigations could extend the time even earlier, one should not expect a new technology to appear simultaneously with the first appearance of AMH since it would take time for the first modern humans to grow into a large enough population to be able to invent new cultures. It is also remarkable to note that the next new invention after pottery, rice or agriculture, also likely came from Hunan [128]. Hunan is also the site of the earliest fully modern human fossils [21]. Placing AMH origin in China is also in line with the observation that the best argument for regional continuity has been built using data from China [4]. The observation here that different modern Chinese people could have independent genetic lineages separated by hundreds of thousands of years is consistent with the morphological observation that *H. erectus* and *H. sapiens* in

Northern China are not identical to those in Southern China [129]. Among all East Asians examined here, the genomes of Hunan people were found most enriched in Africans. Therefore, our model of modern human origins in East Asia, in particular Hunan Province in China, provides a satisfying account of all relevant data.

The study here shows different genetic diversity levels in different human groups depending on different types of SNPs. Europeans show the lowest genetic diversity level in stop codon and splicing SNPs while Africans the highest, which has also been found in a recent study [130]. However, East Asians show the lowest genetic diversity in genome average and hence in non-coding sequences. Thus, different populations encounter different selective pressures, the precise nature of which would require future research. Already, however, some tentative hints emerge on the genetic basis of certain complex traits that are commonly thought to be culturally shaped. The difference in selective pressure on non-coding or regulatory regions versus proteins or parts is reminiscent of the thinking style difference between the East and West in philosophy and medicine, i.e., the holistic versus the analytical [131]. The high genetic diversity of Africans may confer adaptive advantages as in high diversity in immunity [132].

Conclusions:

The MGD theory provides a more complete understanding of the long standing puzzle of what determines genetic diversity, which makes inferring human origins from genetic diversity patterns realistically possible. By better identification of phylogenetically informative genes and constraining neutral theory application to these genes, we provide strong molecular evidence for multiregional evolution of autosomes and for East Asia origin of modern Y and mtDNA. Further work utilizing the MGD theory is ongoing and may yield more surprising and yet satisfying results in human evolution.

Abbreviations:

AMH: anatomically modern humans

MGD: maximum genetic diversity

SNP: single nucleotide polymorphisms

AUA: Aboriginal Australian

PGD: pairwise genetic distance

PCA: principal component analysis

Myr: million years

AFR: African

ASN: East Asian

EUR: European

SAS: South Asian

ESN: Esen in Nigeria

GBR: British in England and Scotland

CHS: Southern Han Chinese

CHB: Han Chinese in Beijing

JPT: Japanese in Tokyo

BEB: Bengali from Bangladesh

YRI: Yoruba in Ibadan, Nigeria

CEU: Utah Residents with Northern and Western European Ancestry

LWK: Luhya in Webuye, Kenya

Declarations:

Acknowledgements:

We thank Shuhua Xu, Xitong Lu, Denghui Luo, Jie Liang and Xiaohua Tan for technical assistance. We thank Xinzhi Wu for sharing unpublished work. We are grateful to Mark Stoneking, Irina Pugach, David Reich, Joseph Pickrell, Arti Tandon, Sarah Tishkoff, Joseph Lachance, Philip Johnson, and Brenna Henn for sharing DNA datasets. Although not all analyses of these datasets were presented here, they have been helpful to make this work possible. We thank Xinzhi Wu, German Dziebel, and Joseph Daniels for critical reading of the manuscript. We thank Feng Gao, Xing Gao, Yamei Hou, Wu Liu, Erik Trinkaus, Lingxia Zhao, and Changqing Zeng for valuable discussions.

Funding:

Supported by the National Natural Science Foundation of China grant 81171880, the National Basic Research Program of China grant 2011CB51001, and the Furong Scholars program (S. H.).

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and material:

The datasets generated and analyzed for this study are available in the Additional files.

968 **Author contributions:**

969 SH and DY designed the study. DW and JY identified the slow evolving proteins. DY,
970 XL, YG, MW, YZ, ZZ, and SH performed data analyses. SH and DY wrote the manuscript
971 and all authors made comments.

972

973 **Competing Interests:**

974 The authors declare that they have no competing interests that might be perceived to
975 influence the results and/or discussion reported in this paper.

976 **References:**

977 [1] Stringer CB, Andrews P. Genetic and fossil evidence for the origin of modern
978 humans. Science 1988;239:1263-8.
979 [2] Thorne AG, Wolpoff MH. Regional continuity in Australasian Pleistocene hominid
980 evolution. Am J Phys Anthropol 1981;55:337-49.
981 [3] Wolpoff MH, Wu XZ, Thorne AG. Modern homo sapiens origins: a general theory
982 of hominid evolution involving the fossil evidence from east Asia. New York: Alan R. Liss,
983 1984.
984 [4] Wu X. On the origin of modern humans in China. Quaternary International
985 2004;117:131-40.
986 [5] Blum MG, Jakobsson M. Deep divergences of human gene trees and models of
987 human origins. Mol Biol Evol 2011;28:889-98.
988 [6] Hajdinjak M, Fu Q, Hubner A, Petr M, Mafessoni F, Grote S, et al. Reconstructing
989 the genetic history of late Neanderthals. Nature 2018;555:652-6.
990 [7] Mendez FL, Poznik GD, Castellano S, Bustamante CD. The Divergence of
991 Neandertal and Modern Human Y Chromosomes. Am J Hum Genet 2016;98:728-34.
992 [8] Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common
993 ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad
994 Sci U S A 2000;97:7360-5.
995 [9] Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. Global patterns
996 of human mitochondrial DNA and Y-chromosome structure are not influenced by higher
997 migration rates of females versus males. Nat Genet 2004;36:1122-5.
998 [10] Cann RL, Stoneking AC, Wilson AC. Mitochondrial DNA and human evolution.
999 Nature 1987;325:31-6.
1000 [11] Green RE, Krause J, et., al. A draft sequence of the Neandertal Genome.
1001 Science 2010;328:710-22.
1002 [12] Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A
1003 High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science 2012.
1004 [13] Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An
1005 early modern human from Romania with a recent Neanderthal ancestor. Nature
1006 2015;524:216-9.

[14] Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 2014;514:445-9.

[15] Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. Science 2014;343:1017-21.

[16] Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature 2017;546:289-92.

[17] White TD, Asfaw B, DeGusta D, Gilbert H, Richards G, Suwa G, et al. Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 2003;423:742-7.

[18] Mendez FL, Krahn T, Schrack B, Krahn AM, Veeramah KR, Woerner AE, et al. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. Am J Hum Genet 2013;92:454-9.

[19] Scerri EML, Thomas MG, Manica A, Gunz P, Stock JT, Stringer C, et al. Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? Trends Ecol Evol 2018;33:582-94.

[20] Bae CJ, Douka K, Petraglia MD. On the origin of modern humans: Asian perspectives. Science 2017;358.

[21] Liu W, Martinon-Torres M, Cai YJ, Xing S, Tong HW, Pei SW, et al. The earliest unequivocally modern humans in southern China. Nature 2015;526:696-9.

[22] Li ZY, Wu XJ, Zhou LP, Liu W, Gao X, Nian XM, et al. Late Pleistocene archaic human crania from Xuchang, China. Science 2017;355:969-72.

[23] Zhao L, Zhang L, Du B, Nian XM, Zheng Y, Zhang Z, et al. New discovery of human fossils and associated mammal faunas in Bijie Guizhou. Acta Anthropologica Sinica 2016;35:24-35.

[24] Athreya S, Wu X. A multivariate assessment of the Dali hominin cranium from China: Morphological affinities and implications for Pleistocene evolution in East Asia. Am J Phys Anthropol 2017;164:679-701.

[25] Hershkovitz I, Weber GW, Quam R, Duval M, Grun R, Kinsley L, et al. The earliest modern humans outside Africa. Science 2018;359:456-9.

[26] Groucutt HS, Grun R, Zalmout IAS, Drake NA, Armitage SJ, Candy I, et al. Homo sapiens in Arabia by 85,000 years ago. Nat Ecol Evol 2018;2:800-9.

[27] Curnoe D, Ji X, Shajin H, Tacon PSC, Li Y. Dental remains from Longtanshan cave 1 (Yunnan, China), and the initial presence of anatomically modern humans in East Asia. Quaternary International 2016;400:180-6.

[28] Bae CJ, Wang W, Zhao J, Huang S, Tian F, Shen G. Modern human teeth from Late Pleistocene Luna Cave (Guangxi, China). Quaternary International 2014;354:169-83.

[29] 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;19:255-71.

[30] 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:e1001388.

[31] Hahn MW. Toward a selection theory of molecular evolution. Evolution 2008;62:255-65.

[32] Kreitman M. The neutral theory is dead. Long live the neutral theory. Bioessays 1996;18:678-83; discussion 83.

[33] Ohta T, Gillespie JH. Development of Neutral and Nearly Neutral Theories. Theor Popul Biol 1996;49:128-42.

[34] Ballard JW, Kreitman M. Is mitochondrial DNA a strictly neutral marker? Trends Ecol Evol 1995;10:485-8.

[35] Wolpoff MH, Caspari R. Race and Human Evolution: A Fatal Attraction. New York: Simon & Schuster, 2007.

[36] Mondal M, Casals F, Xu T, Dall'Olio GM, Pybus M, Netea MG, et al. Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. Nat Genet 2016;48:1066-70.

[37] Huang S. New thoughts on an old riddle: What determines genetic diversity within and between species? Genomics 2016;108:3-10.

[38] Zhu Z, Lu Q, Wang J, Huang S. Collective effects of common SNPs in foraging decisions in Caenorhabditis elegans and an integrative method of identification of candidate genes. Sci. Rep. 2015:doi:10.1038/srep16904.

[39] Luo D, Huang S. The genetic equidistance phenomenon at the proteomic level. Genomics 2016;108:25-30.

[40] 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:e0133421.

[41] Zhu Z, Man X, Xia M, Huang Y, Yuan D, Huang S. Collective effects of SNPs on transgenerational inheritance in Caenorhabditis elegans and budding yeast. Genomics 2015;106:23-9.

[42] Biswas K, Chakraborty S, Podder S, Ghosh TC. Insights into the dN/dS ratio heterogeneity between brain specific genes and widely expressed genes in species of different complexity. Genomics 2016;108:11-7.

[43] Zhu Z, Lu Q, Zeng F, Wang J, Huang S. Compatibility between mitochondrial and nuclear genomes correlates wtih quantitative trait of lifespan in Caenorhabditis elegans. Sci. Rep. 2015:doi:10.1038/srep17303.

[44] Huang S. The overlap feature of the genetic equidistance result, a fundamental biological phenomenon overlooked for nearly half of a century. Biological Theory 2010;5:40-52.

[45] Huang S. Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing tarsiers. Sci China Life Sci 2012;55:709-25.

[46] Lei X, Yuan J, Zhu Z, Huang S. Collective effects of common SNPs and risk prediction in lung cancer. Heredity 2018:doi:10.1038/s41437-018-0063-4.

[47] Gui Y, Lei X, Huang S. Collective effects of common SNPs and genetic risk prediction in type 1 diabetes. Clin Genet 2017;93:1069-74.

[48] Yuan D, Zhu Z, Tan X, Liang J, Zeng C, Zhang J, et al. Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms. Sci China Life Sci 2014;57:876-88.

[49] Lei X, Huang S. Enrichment of minor allele of SNPs and genetic prediction of type 2 diabetes risk in British population. PLoS ONE 2017;12:e0187644.

[50] He P, Lei X, Yuan D, Zhu Z, Huang S. Accumulation of minor alleles and risk prediction in schizophrenia. Sci Rep 2017;7:11661.

[51] Huang S. Histone methylation and the initiation of cancer, Cancer Epigenetics. New York: CRC Press, 2008.

[52] Huang S. Inverse relationship between genetic diversity and epigenetic complexity. Nature Precedings 2008:doi.org/10.1038/npre.2009.1751.2.

[53] Kimura M. Evolutionary rate at the molecular level. Nature 1968;217:624-6.

[54] King JL, Jukes TH. Non-Darwinian evolution. Science 1969;164:788-98.

[55] Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 2014;515:261-3.

[56] Margoliash E. Primary structure and evolution of cytochrome c. Proc. Natl. Acad. Sci. 1963;50:672-9.

[57] Hu T, Long M, Yuan D, Zhu Z, Huang Y, Huang S. The genetic equidistance result, misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. Sci China Life Sci 2013;56:254-61.

[58] Huang S. The genetic equidistance result of molecular evolution is independent of mutation rates. J. Comp. Sci. Syst. Biol. 2008;1:092-102.

[59] Yuan D, Zhu Z, Tan X, Liang J, Zeng C, Zhang J, et al. Minor alleles of common SNPs quantitatively affect traits/diseases and are under both positive and negative selection. arXiv:1209.2911 2012.

[60] Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.

[61] Prieto-Godino LL, Rytz R, Bargeton B, Abuin L, Arguello JR, Peraro MD, et al. Olfactory receptor pseudo-pseudogenes. Nature 2016;539:93-7.

[62] Wang M, Wang D, Yu J, Huang S. Enrichment in conservative amino acid changes among fixed and standing missense variations in slow evolving genes. bioRxiv 2019:doi.org/10.1101/644666.

[63] Pugach I, Delfin F, Gunnarsdottir E, Kayser M, Stoneking M. Genome-wide data substantiate Holocene gene flow from India to Australia. Proc Natl Acad Sci U S A 2013;110:1803-8.

[64] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.

[65] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156-8.

[66] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

[67] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.

[68] Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 2013;10:5-6.

[69] Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.

[70] Zhu Z, Lu X, Yuan D, Huang S. Close genetic relationships between a spousal pair with autism-affected children and high minor allele content in cases in autism-associated SNPs. Genomics 2016:10.1016/j.ygeno.2016.12.001.

[71] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76-82.

[72] Lewontin R. The apportionment of human diversity. Evolutionary Biology 1972;6:391-8.

[73] Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature 2015;526:68-74.

[74] Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic distance in

1154 human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci U S
1155 A 2005;102:15942-7.
1156 [75] Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al.
1157 Worldwide human relationships inferred from genome-wide patterns of variation. Science
1158 2008;319:1100-4.
1159 [76] Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, et al. A
1160 worldwide survey of haplotype variation and linkage disequilibrium in the human genome.
1161 Nat Genet 2006;38:1251-60.
1162 [77] Suwa G, Kono RT, Katoh S, Asfaw B, Beyene Y. A new species of great ape
1163 from the late Miocene epoch in Ethiopia. Nature 2007;448:921-4.
1164 [78] Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, et al.
1165 Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males
1166 versus females. Science 2013;341:562-5.
1167 [79] Wilson Sayres MA, Lohmueller KE, Nielsen R. Natural selection reduced
1168 diversity on human y chromosomes. PLoS Genet 2014;10:e1004064.
1169 [80] Teitz LS, Pyntikova T, Skaletsky H, Page DC. Selection Has Countered High
1170 Mutability to Preserve the Ancestral Copy Number of Y Chromosome Amplicons in
1171 Diverse Human Lineages. Am J Hum Genet 2018;103:261-75.
1172 [81] Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, et al.
1173 Punctuated bursts in human male demography inferred from 1,244 worldwide
1174 Y-chromosome sequences. Nat Genet 2016;48:593-9.
1175 [82] Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons
1176 Genome Diversity Project: 300 genomes from 142 diverse populations. Nature
1177 2016;538:201-6.
1178 [83] Karmin M, Saag L, Vicente M, Wilson Sayres MA, Jarve M, Talas UG, et al. A
1179 recent bottleneck of Y chromosome diversity coincides with a global change in culture.
1180 Genome Res 2015;25:459-66.
1181 [84] Black ML, Wise CA, Wang W, Bittles AH. Combining genetics and population
1182 history in the study of ethnic diversity in the People's Republic of China. Hum Biol
1183 2006;78:277-93.
1184 [85] Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic
1185 history of Ice Age Europe. Nature 2016;534:200-5.
1186 [86] Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, et al.
1187 Genomic insights into the origin of farming in the ancient Near East. Nature
1188 2016;536:419-24.
1189 [87] Teske PR, Golla TR, Sandoval-Castillo J, Emami-Khoyi A, van der Lingen CD,
1190 von der Heyden S, et al. Mitochondrial DNA is unsuitable to test for isolation by distance.
1191 Sci Rep 2018;8:8448.
1192 [88] Towarnicki SG, Ballard JWO. Mitotype Interacts With Diet to Influence Longevity,
1193 Fitness, and Mitochondrial Functions in Adult Female Drosophila. Front Genet
1194 2018;9:593.
1195 [89] de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J,
1196 et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science
1197 2016;354:477-81.
1198 [90] Amos W. Testing an alternative explanation for relatively greater base-sharing
1199 between Neanderthals and non-African humans. bioRxiv 2017:doi:
1200 https://doi.org/10.1101/133306.
1201 [91] Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The
1202 complete genome sequence of a Neanderthal from the Altai Mountains. Nature
1203 2014;505:43-9.

[92] Prufer K, de Filippo C, Grote S, Mafessoni F, Korlevic P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 2017;358:655-8.

[93] Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, et al. Complete Khoisan and Bantu genomes from southern Africa. Nature 2010;463:943-7.

[94] Gallego Llorente M, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science 2015;350:820-2.

[95] Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 2011;334:94-8.

[96] Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga JL, et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature 2014;505:403-6.

[97] Lepre CJ, Roche H, Kent DV, Harmand S, Quinn RL, Brugal J-P, et al. An earlier origin for the Acheulian. Nature 2011;477:82-5.

[98] Asfaw B, Gilbert WH, Beyene Y, Hart WK, Renne PR, Gabriel GW, et al. Remains of Homo erectus from Bouri, middle Awash, Ethiopia. Nature 2002;416:317-20.

[99] Bischoff JL, Williams RW, Rosenbauer RJ, Aramburu A, Arsuaga JL, Garcia N, et al. High-resolution U-series dates from the Sima de los Huesos hominids yields 600+/-kyrs: implications for the evolution of the early Neanderthal lineage. J. Archaeol. Sci. 2007;34:763-70.

[100] Lycett SJ. Understanding ancient hominin dispersals using artefactual data: a phylogeographic analysis of Acheulean handaxes. PLoS ONE 2009;4:e7404.

[101] Shoffner JM, Brown MD, Torroni A, Lott MT, Cabell MF, Mirra SS, et al. Mitochondrial DNA variants observed in Alzheimer disease and Parkinson disease patients. Genomics 1993;17:171-84.

[102] van der Walt JM, Nicodemus KK, Martin ER, Scott WK, Nance MA, Watts RL, et al. Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. Am J Hum Genet 2003;72:804-11.

[103] Picard M, Wallace DC, Burelle Y. The rise of mitochondria in medicine. Mitochondrion 2016;30:105-16.

[104] Charchar FJ, Bloomer LD, Barnes TA, Cowley MJ, Nelson CP, Wang Y, et al. Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. Lancet 2012;379:915-22.

[105] Sloan DB, Havird JC, Sharbrough J. The On-Again, Off-Again Relationship between Mitochondrial Genomes and Species Boundaries. Mol Ecol 2016.

[106] Osada N, Akashi H. Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. Mol Biol Evol 2012;29:337-46.

[107] Gemmell NJ, Sin FY. Mitochondrial mutations may drive Y chromosome evolution. Bioessays 2002;24:275-9.

[108] Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol Evol 2004;19:645-53.

[109] Bermudez de Castro JM, Arsuaga JL, Carbonell E, Rosas A, Martinez I, Mosquera M. A hominid from the lower Pleistocene of Atapuerca, Spain: possible ancestor to Neandertals and modern humans. Science 1997;276:1392-5.

[110] Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, et al. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. PLoS ONE 2014;9:e105691.

[111] He P, Chen N, Hu Z, Zhu Z, Xia K, Huang S. Neolithic super-grandfather Y haplotypes, their related surnames, and autism spectrum disorder. bioRxiv 2017;https://doi.org/10.1101/077222.

[112] Curnoe D, Xueping J, Herries AI, Kanning B, Tacon PS, Zhende B, et al. Human remains from the Pleistocene-Holocene transition of southwest China suggest a complex evolutionary history for East Asians. PLoS ONE 2012;7:e31918.

[113] Berger LR, Hawks J, de Ruiter DJ, Churchill SE, Schmid P, Delezene LK, et al. Homo naledi, a new species of the genus Homo from the Dinaledi Chamber, South Africa. Elife 2015;4.

[114] 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.

[115] Shaw JC. Taurodont Teeth in South African Races. J Anat 1928;62:476-98 1.

[116] Liu W, Mbua E, Wu X, Zhang Y. Comparisons of cranial features between Chinese and Afrcian holocene humans and their implications. Acta Anthropologica Sinica 2003;22:89-104.

[117] Thorne AG, Wolpoff MH. The multiregional evolution of humans. Sci Am 1992;266:76-9, 82-3.

[118] Wu XJ, Crevecoeur I, Liu W, Xing S, Trinkaus E. Temporal labyrinths of eastern Eurasian Pleistocene humans. Proc Natl Acad Sci U S A 2014;111:10509-13.

[119] Wu X. Comparative study of early Homo sapiens from China and Europe. Acta Anthropologica Sinica 1988;7:287-93.

[120] Li ZY, Wu XJ, Zhou LP, Liu W, Gao X, Nian XM, et al. Late Pleistocene archaic human crania from Xuchang, China. Science 2017;355:969-72.

[121] Wu X, Poirier FE. Human evolution in China: a metric description of the fossils and a review of the sites. Oxford: Oxford University Press, 1995.

[122] Gunz P, Bulygina E. The Mousterian child from Teshik-Tash is a Neanderthal: a geometric morphometric study of the frontal bone. Am J Phys Anthropol 2012;149:365-79.

[123] Martinon-Torres M, Bermudez de Castro JM, Gomez-Robles A, Arsuaga JL, Carbonell E, Lordkipanidze D, et al. Dental evidence on the hominin dispersals during the Pleistocene. Proc Natl Acad Sci U S A 2007;104:13279-82.

[124] Wolpoff MH. Human Evolution. New York: McGraw-Hill, Inc, 1996.

[125] Brauer G, Stringer C. Models, polarization, and perspectives on modern human origins. New York: Aldine de Gruyter, 1997.

[126] Boaretto E, Wu X, Yuan J, Bar-Yosef O, Chu V, Pan Y, et al. Radiocarbon dating of charcoal and bone collagen associated with early pottery at Yuchanyan Cave, Hunan Province, China. Proc Natl Acad Sci U S A 2009;106:9595-600.

[127] Wu X, Zhang C, Goldberg P, Cohen D, Pan Y, Arpin T, et al. Early pottery at 20,000 years ago in Xianrendong Cave, China. Science 2012;336:1696-700.

[128] Zhang W, Yuan J. A preliminary study of ancient excavated rice from Yuchanyan site, Dao County, Hunan Province, P.R.China. . Acta Agronomica Sinica 1998;24:416-20.

[129] Wu R, Wu X. Paleolithic Sites In China. Shanghai Scientific and Technological Education Publishing House, 1999.

[130] Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-91.

[131] Nisbett RE, Peng K, Choi I, Norenzayan A. Culture and systems of thought: holistic versus analytic cognition. Psychol Rev 2001;108:291-310.

[132] Nemat-Gorgani N, Hilton HG, Henn BM, Lin M, Gignoux CR, Myrick JW, et al. Different Selected Mechanisms Attenuated the Inhibitory Interaction of KIR2DL1 with C2(+) HLA-C in Two Indigenous Human Populations in Southern Africa. J Immunol 2018;200:2640-55.

Tables:

Table 1. Sharing of different types of SNPs among three groups. Shared SNPs are present in more than one group and unique SNPs are present in only one group. Shown are fractions of each type of SNPs. SNPs that are not found in any of the three groups (AFR, ASN, and EUR) are grouped as no variations (No var.).

	Shared	**Unique**	**No var.**	**#SNPs**
Nonsyn slow	0.05	0.66	0.29	15422
Syn slow	0.11	0.64	0.24	16591
Random set	0.24	0.52	0.24	254489

Table 2. Time of divergence between human populations. The separation time and average pairwise genetic distance (total distance including both het and hom distances) between human populations (ESN, GBR, CHS) in 9578 slow evolving autosome SNPs located in the 178 genes (>99% and <100% identity between human and Macca) with total length 291083 aa. The human mutation rate was estimated as 4.57E-5 aa/myr/aa x 291083 aa = 13.32 aa/myr.

	Myr (total # of aa mismatches)		
Groups	ESN	GBR	CHS
ESN	1.77 (47.21)	1.92 (51.03)	1.86 (49.62)
GBR		1.53 (40.65)	1.61 (42.8)
CHS			1.4 (37.19)

1331
1332
1333 **Table 3. Sharing of Y chr alleles between Neanderthals and modern humans**
1334

	Mezmaiskaya2			Spy94			El Sidron		
Haplotypes	**#SNPs**	**#Shared**	**Fraction**	**#SNPs**	**#Shared**	**Fraction**	**#SNPs**	**#Shared**	**Fraction**
A	4	2	0.5	1	1	1	4	3	0.75
A0	3	3	1				3	2	0.67
AB	2	1	0.5				1	1	1
ABCDE	2	2	1	2	2	1	2	2	1
B	2	0	0	1	0	0			
C	4	0	0	2	0	0			
D2	8	0	0	3	0	0			
E	14	0	0	7	0	0	1	1	1
G	3	0	0	4	0	0	1	0	0
H1	8	0	0	7	0	0			
I1	3	0	0	2	0	0			
IJ	2	0	0						
J	2	0	0						
L	2	0	0	2	0	0			
LT	3	0	0						
N	4	0	0	2	0	0			
O	12	0	0	5	0	0			
Q				1	0	0			
QR	3	0	0				1	0	0
R1a	12	0	0	2	0	0	1	0	0
T	9	0	0	6	0	0			

1335
1336

1337

Figure legends:

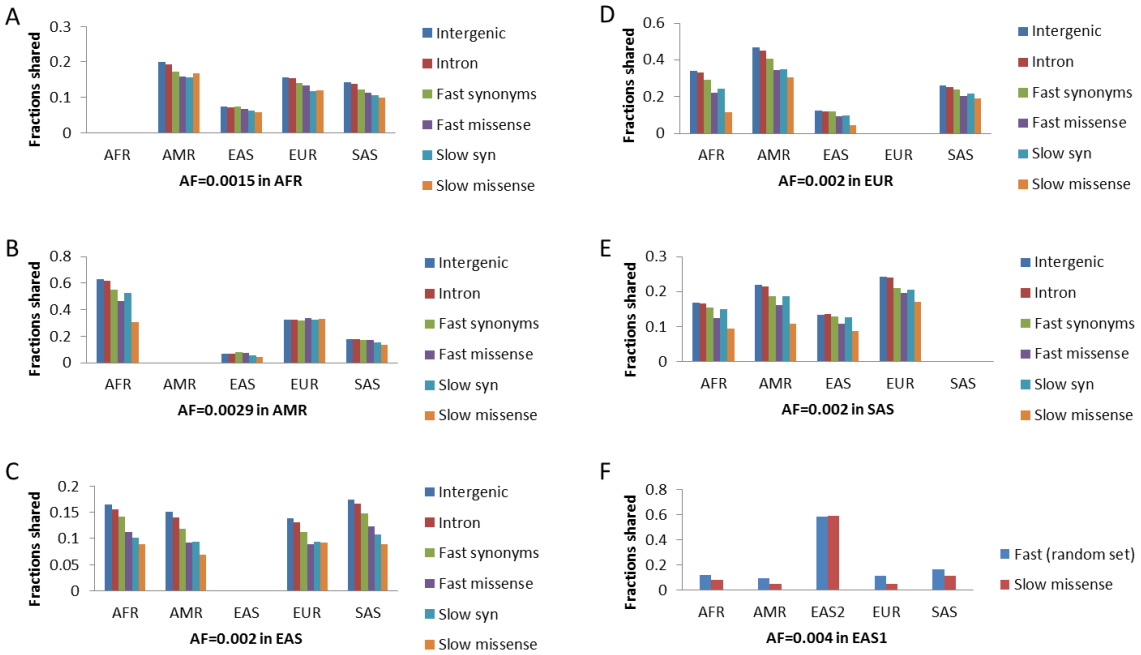

Figure 1. Sharing of SNPs among different racial groups. The fractions of shared SNPs in each racial group are shown for SNPs with alternative allele appearing only twice in AFR (A), AMR (B), EAS (C), EUR (D), SAS (E), or EAS1. EAS1 and EAS2 were two randomly divided subgroups of EAS. These SNPs were classified based on evolutionary rates as indicated.

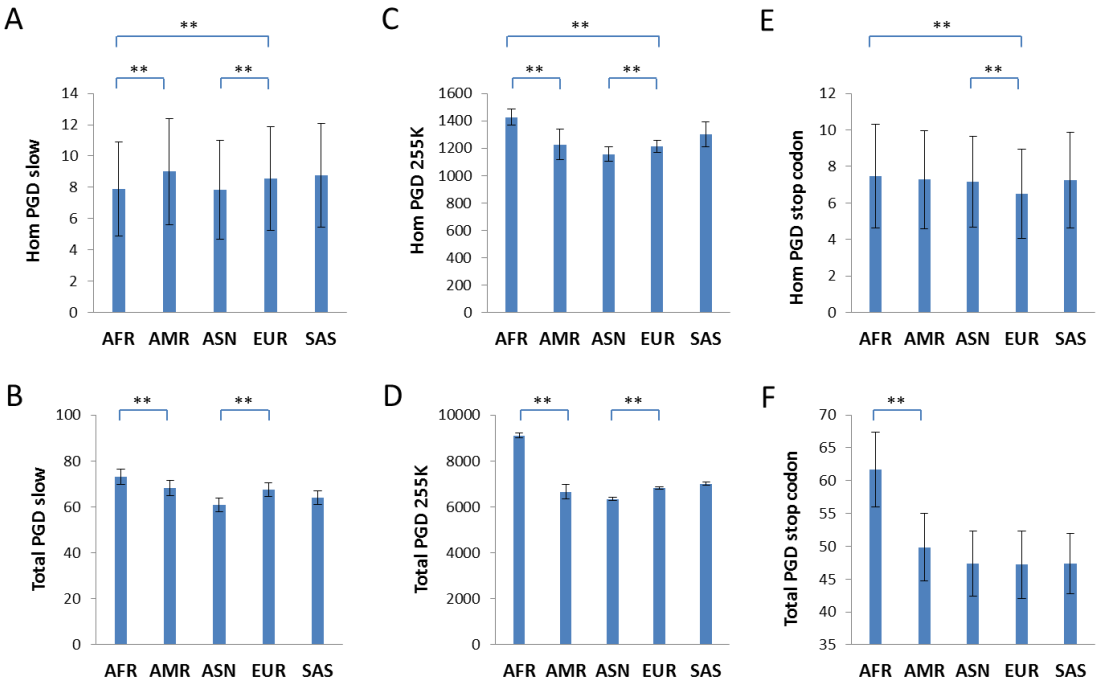

Figure 2. Pairwise genetic distance as measured by different types of SNPs.

Pairwise genetic distance (PGD), either by homozygous mismatches (Hom) or by both homozygous and heterozygous mismatches (Total), as measured by three different types of SNPs is shown for each of the 5 major human groups in the 1KG. Known heavily admixed groups such as ASW and ACB in the African group or CLM and PUR in the American group were excluded in the analysis. Data are means with standard deviation. **, P <0.01, t test, 2 tailed.

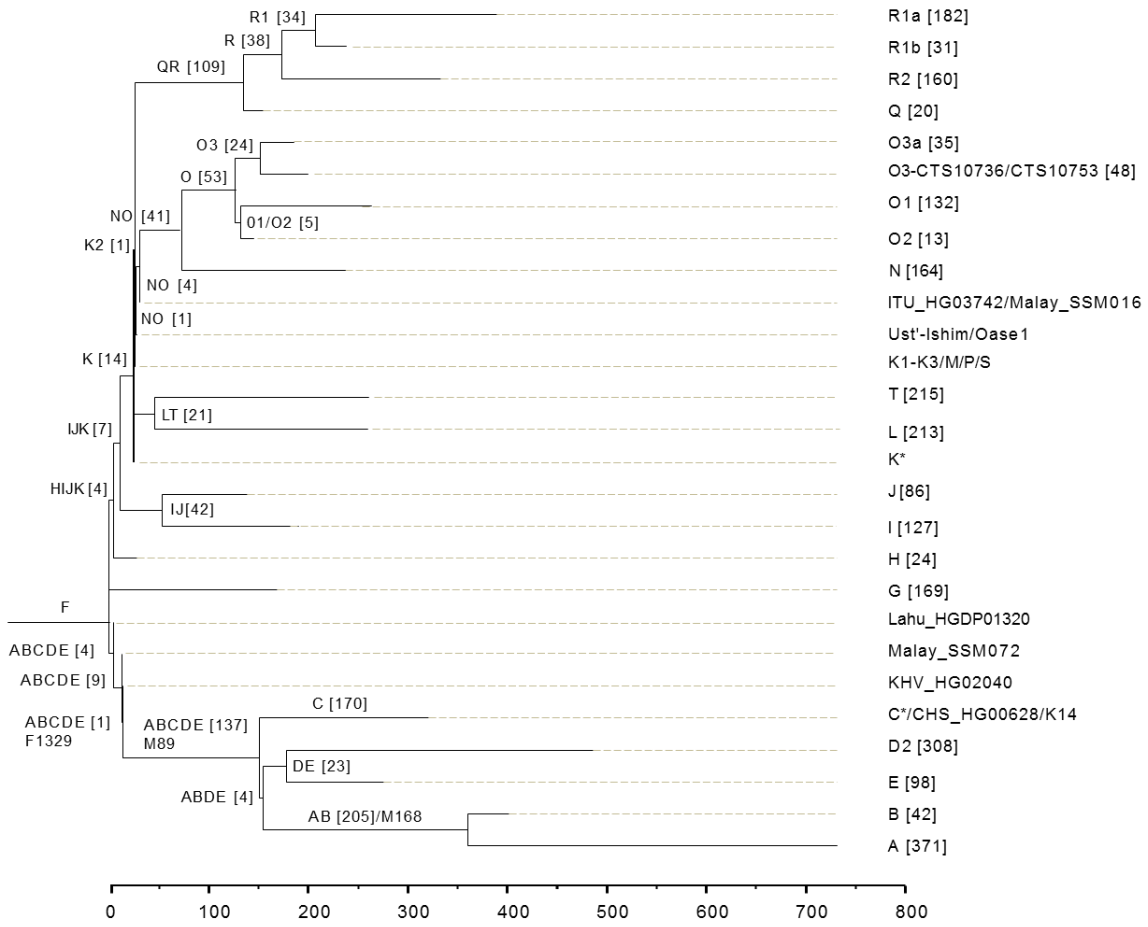

Figure 3. Y chromosome phylogeny. Branch lengths are drawn proportional to the number of SNPs. Only major haplogroups are shown with defining SNPs indicated for some. Numbers in parenthesis indicate the number of SNPs defining a haplogroup among the 58251 cleanly called SNPs in the 1KG. Individuals with few changes from an ancestor haplotype are also listed as shown.

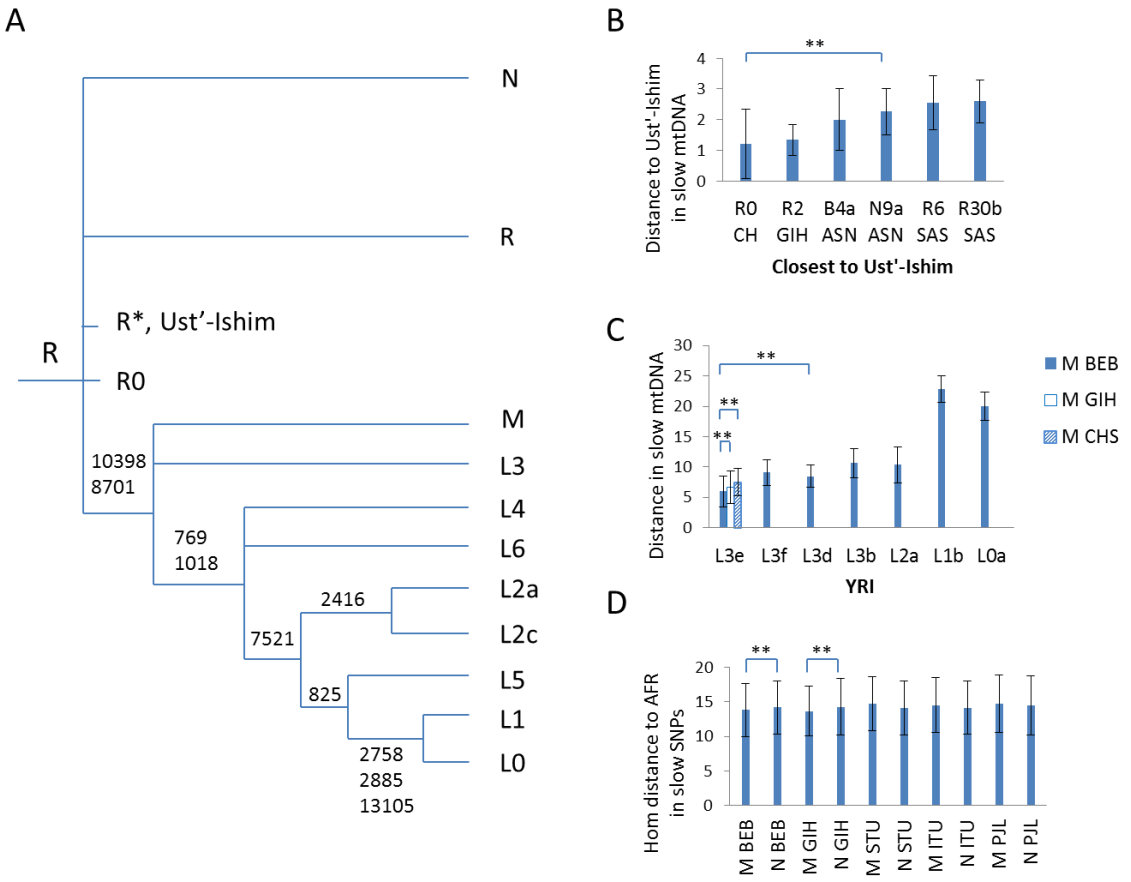

Figure 4. mtDNA phylogeny. (**a**) The mtDNA tree was drawn using slow evolving SNPs as indicated with the common ancestor haplotype defined as being closest to the ~45,000 year old Ust'-Ishim. Only major branches are shown and no slow SNPs could be found to separate N and R. (**b**) Genetic distance in slow mtDNA SNPs to Ust'-Ishim mtDNA for haplotypes in 1KG. Only the closest few are shown. (**c**) Genetic distance in slow mtDNA SNPs to the M haplotype in BEB, GIH, or CHS for different L haplotypes in the YRI group.(**d**) Genetic distance in slow autosomal SNPs to individuals of South Asian BEB (or GIH, STU, ITU, PJL) carrying either the M or N haplotype. Data are means with standard deviation. **, P<0.01, t test, 2 tailed.

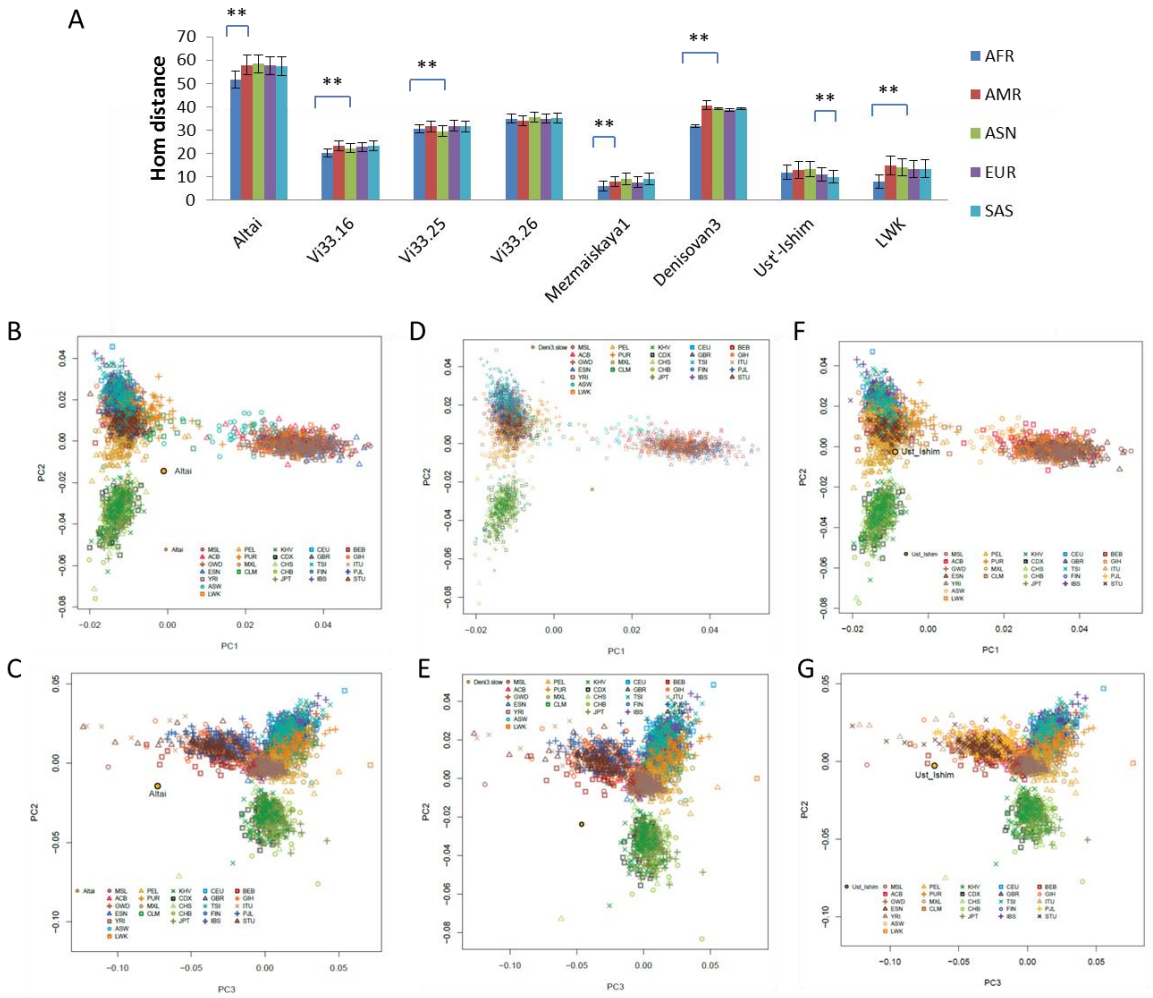

Figure 5. Autosomal relationship between archaic and modern humans. (**a**) Shown are the genetic distances between the 5 groups of 1KG and Neanderthals, Denisovan3, Ust'-Ishim, or the modern African group LWK. Data are means with standard deviation. (**b-g**) PCA plot analyses (b, d, f for PC1-PC2; c, e, g for PC3-PC2) for Altai, Denisovan3, or Ust'-Ishim merged with 1KG. **, P<0.01, t test, 2 tailed.

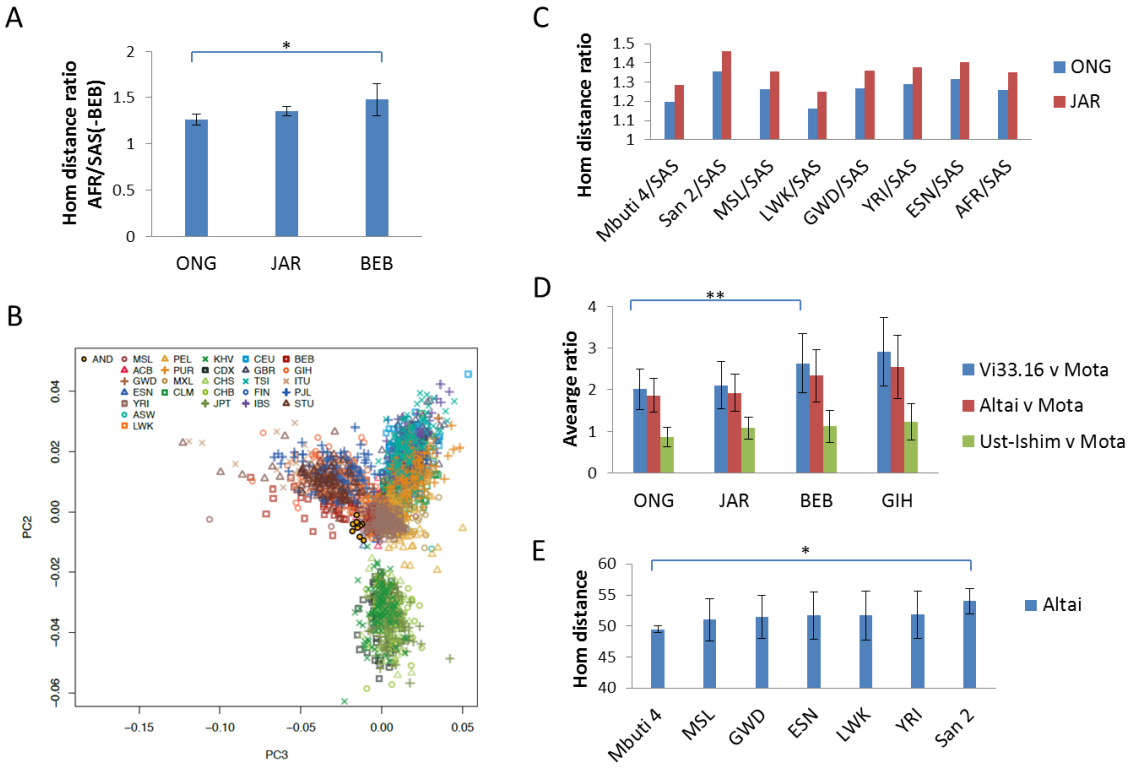

Figure 6. Origin of Negritos.(**a**) Shown are the ratios of ONG, JAR, or BEB autosomal distance to AFR versus SAS(-BEB). SAS (-BEB) excluded the BEB group from SAS groups. (**b**) PCA plot (PC3-PC2) analysis of 10 Andamanese and 1KG using slow autosomal SNPs. (**c**) Shown are the ratios of ONG or JAR autosomal distance to African groups versus SAS. (**d**) Hom distance ratio of ancient humans versus the Mota African for four South Asian groups (ONG, BEB, GIH, JAR). (**e**) Autosomal distance between Altai and various African groups. Data are means with standard deviation. **, P<0.01, t test or chi-squared test, 2 tailed.

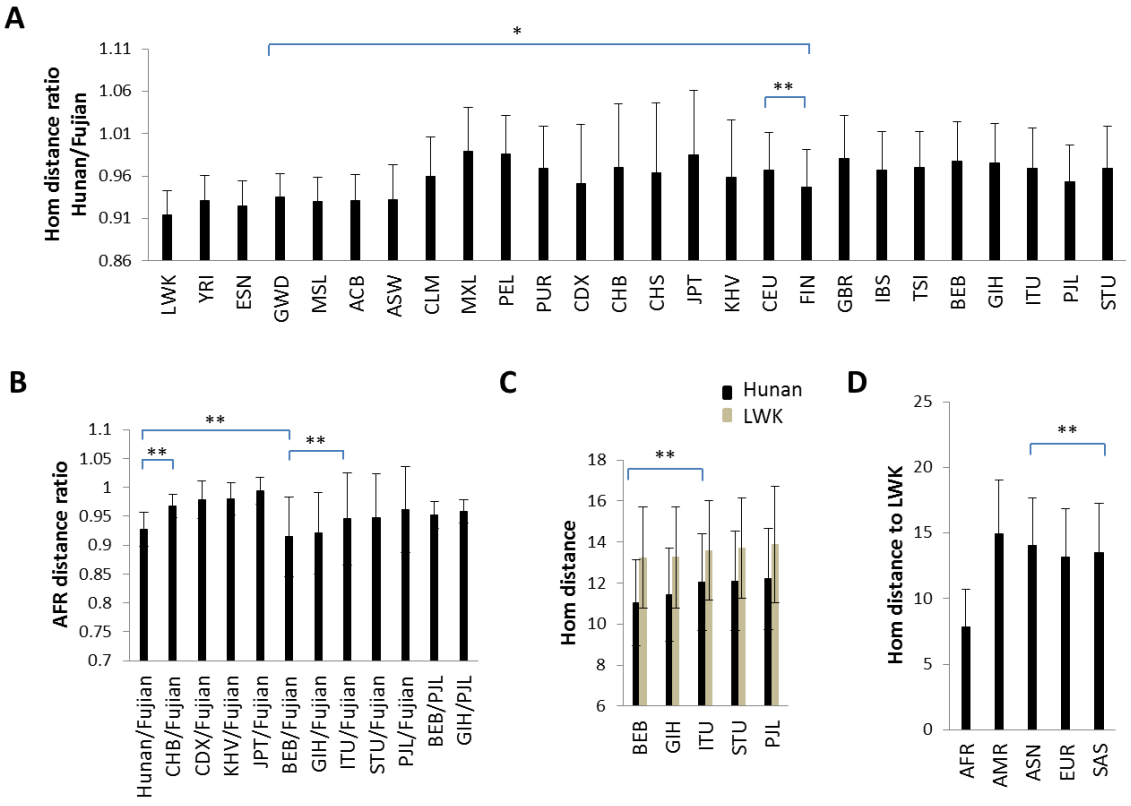

Figure 7. Hunan ancestry in Africans. (**a**) Ratios of autosomal distance to Hunan versus Fujian for each of the 25 groups in 1KG. (**b**) Ratios of autosomal distance to Hunan (or other East Asian and South Asian groups in 1KG) versus Fujian. (**c**) Autosomal distance to Hunan or LWK for various South Asian groups. (**d**) Autosomal distance to LWK for the 5 groups in 1KG. **, P<0.01, *, P<0.05, t test or chi-squared test, 2 tailed. Standard deviations are shown.

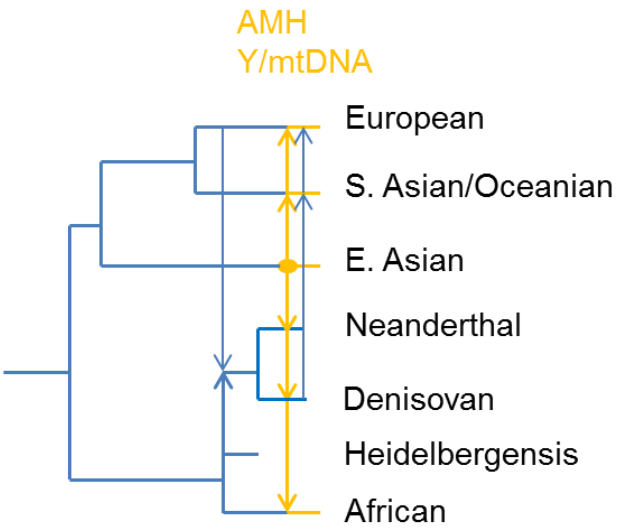

Figure 8. Model of human evolution. A schematic tree showing the phylogenetic relationship of major human groups, including Africans, East Asians, South Asians/Oceanians, Europeans, Heidelbergensis, Neanderthals, and Denisovans.