Supplementary Materials for

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

## Dejian Yuan, Xiaoyun Lei, Yuanyuan Gui, Zuobin Zhu, Dapeng Wang, Jun Yu, and Shi Huang\*

## \*Corresponding author. E-mail: [huangshi@sklmg.edu.cn](mailto:huangshi@sklmg.edu.cn)

## This PDF file includes:

## Supplementary text

## Tables S1 to S14

## Figures S1 to S20

## References

SUPPLEMENTARY TEXT:

TABLE OF CONTENTS

S1. Genetic diversities in different types of SNPs in 1KG …………………………………… 3

S2. Conserved slow evolving sequences still have neutral variations…………………………3

S3. Known positively selected genes are fast evolving………………………………………....4

S4. Analysis of mtDNA haplotypes…………………………………………………………….…..4

S5. Admixture analysis…………………………………………………………………………….. 8

S6. References……………………………………………………………………………………….9

Supplementary Tables……………………………….…….......................................................10

Supplementary Figures……………………………….……......................................................13

Supplementary Information 1

Genetic diversity patterns in different types of SNPs in 1KG

In addition to SNP types examined in detail in the main text, we also examined the pairwise genetic distance (PGD) pattern of the splicing SNPs that are also expected to be functional (Supplementary Fig. S1 A-B). To examine whether the cut-off point in protein conservation in selecting the slow evolving proteins was appropriate, we verified that using nonsyn SNPs located within the next set of just slightly less conserved proteins (361 autosomal proteins with 800-1102 aa in length and with identity between human and monkey >99% but <100%) produced PGD patterns similar to the functional stop codon or splicing SNPs (Supplementary Fig. S1 C-D). We further showed that synonymous SNPs within the slow set of proteins gave patterns similar to the stop codon or splicing SNPs but unlike the nonsyn SNPs within the same set of slow evolving proteins (Supplementary Fig. S1 E-F).

Supplementary Information 2

Conserved slow evolving sequences still have neutral variations

As we aim to use slow evolving or highly conserved genes to do phylogenetic analysis, it is important to show that they still have neutral sequences that are not under natural selection. We determined whether they have fewer overlap sites by examining, of those different positions between human and monkey, how many are also different between any pair of the three species human, monkey and mouse ([*1*](#_ENREF_1)). As discussed before, such overlap sites would be rare if mutations occur in a neutral fashion following the infinite sites model of the Neutral theory rather than cluster at either positively selected sites or sites left untouched/neutral by negative selection. We examined the informative genes in the set of slow evolving proteins as described in Supplementary Table S3 that have orthologs in the species concerned and show less than 5 bp gaps in any pair of alignments. Of 31 informative proteins (Supplementary Table S5), the average overlap ratio is 9.6% (the number of overlap positions divided by the number of amino acid differences between human and monkey). In contrast, a randomly chosen set of 63 fast evolving genes (~93% identity between human and monkey), the average overlap ratio is 26.1% (2.73 fold greater than that of slow evolving genes, P < 0.01). These results confirm that highly conserved proteins have fewer overlap sites, and variants in them do follow the Neutral theory ([*2*](#_ENREF_2)).

Supplementary Information 3

Known positively selected genes are fast evolving

We next looked at whether known positively selected genes are more likely to be fast evolving. We did not find any of the slow genes present in the set of 56 positively selected genes identified by Nielson et al ([*3*](#_ENREF_3)) but found two (WDR7 and KIAA1429) in the set of 154 human genes identified by Bakewell et al, which does overlap with the Nielson et al set ([*4*](#_ENREF_4)). Among the list of ~14000 genes with informative identity scores between human and monkey, 24 and 87 were found in the 56 and 154 sets, respectively. These 111 genes have 88.6% identity between human and monkey with average length 601 aa. The genome average of ~14000 genes are 87.8% and 560 aa. Thus, positively selected genes have mutation rates similar to the genome average, which is much faster than the slow evolving set of genes identified here.

Supplementary Information 4

Phylogenetic tree of mtDNA haplotypes using slow SNPs and implications for admixtures

*5-1. Assigning the ancestral mtDNA type*

We used slow mtDNA SNPs (nonsyn and RNA SNPs) to redraw mtDNA haplotype tree based on the PhyloTree17. As described in the main text, the AMH ancestor had the R0 haplogroup. R0 here is essentially basal B4 if only slow SNPs are considered. The type 1 morph found in 1983 is 2-1-1-1-1 (Hpa1, BamH1, HaeII, Msp1,AvaII) and was considered the original type ([*6*](#_ENREF_6)). The minor alleles of these sites are rare and present in less than 30 sequences in the witoweb’s collection of 32059 full length mtDNAs. A few are relatively more common. Among these, Msp1 sites at 8112 and 8150 are lost in ~552 sequences in mitoweb, most of which are Bushman L0d. AvaII site at 13368 is lost in 1585 sequences in mitoweb and defines hg T. AvaII site at 12630 is lost in 167 sequences that includes many haplotypes including three B2b found in Amerindians and one B4g of East Asians. HaeII site 11002 is not present in type 1 but appears in other types and is present in ~300 sequences in the witoweb that include one B2 and many other haplotypes.

So, type 1 morph appears to be a collection of major alleles that are extremely common in worldwide populations and particularly so in Asians and Amerindians. It is expected that the original AMH should carry these major alleles because as the AMH population grew only a small fraction should carry alternative mutant alleles (mutations are rare events). Indeed, the 45K year old Ust’-Ishim’s mtDNA has fewer SNPs to begin with and the one he did have are all common alleles in today’s population (>67%) except one fast evolving site 16150. The existing conclusion on Ust’-Ishim is paradoxical, with its mtDNA carrying mostly common alleles while its autosomes mostly rare alleles, which is an artifact of inferring genealogy using fast evolving autosomal SNPs. The implication of our reasoning here is that the original AMH people were indeed the same as people today, consistent with the definition of modern. The out of Africa model however still requires much evolution from the original AMH as if they were not modern, resulting in their alleles being rare in today’s populations. But, that is precisely what one should expect for the archaic humans like Neanderthals.

It is also expected that present day people who are closest to the original AMH in terms of physiology and environments should also mostly carry the major alleles. Therefore, the population with the highest frequency of type 1 should be closest to the original AMH and their location should indicate the place of origin for AMH. This may be the only rationale left for building a mtDNA tree if radiation involved admixture with locale natives and if one can nullify the neutral interpretation of African high genetic diversity.

While the limited number of samples studied by Johnson et al (1983) found Amerindians to be all type 1, larger sample size in the mitoweb do show 4 Amerindians with B2 haplotype carrying non-type 1 morphs as shown above. Even if Amerindians do show the highest frequency of type 1, it remains uncertain whether the type 1 as defined by Johnson et al should be considered the best candidate for the original type (maybe a close one) as it is not a result of systematic evaluation of all common SNPs. Because most of the sites covered by the Johnson et al paper are fast evolving, the issue should be settled by a systemic new analysis using slow SNPs in mtDNA, which was what we have done here.

*5-2 Phylogenetic tree of mtDNA as defined by slow SNPs*

Many subtypes within R/N/M share alleles with Neanderthal or related humans, which are common in present day people known to show admixed features (N1b, N14, N21, O1a, P4, S, Y, N9b, I, N1, N5, W, N2a, J, K, T, N8, A, X2, Q, F1, R12’21, M5, M66, M67, M43, M63, M64, M10, M28, M29, M52, M77, M58, M7b, M80, D6, G, and L (Supplementary Fig. S4 and Table S12). Consistent with Africans allowing for more SNP variations in autosomes and Y, most L haplotypes other than L3 have longer branch length than non L haplogroups, especially long for L0 common in Bushman (Supplementary Table S11 and Fig. S4). Also as expected, haplotypes common in populations with admixed traits have longer branch length, including M1a1c, M2, M51, M23, M25, M32a, M67, D4d, N11b, N11a, N14, N1a, I, W, J, K1, R14, R23, P4, and T (Supplementary Fig. S4). Certain haplotypes in East Asians such as M20 and D4d also have long branch length, which is consistent with fossil findings of archaic human admixtures in East Asia. Also, for haplotypes sharing a common SNP, those common in groups with admixed traits have longer branches than those without, including M1a and M51 vs M20, M32a in Andamanese vs M56 in Indians, M73 in Philippines vs M79 in Chinese, and G1a in Japanese vs M12 in Indians (Fig. S4).

Consistent with a role for fast SNPs in adaptation, certain haplotypes representing very different human groups were found sharing fast SNPs. M4”67 of South Asians has many branches all sharing 12007 fast SNP common in L0, A2, and J1b (Fig. S4). M7a, b, c shared 2 fast SNPs but cover a wide range of people including Japanese, Chinese, Malays, and Laos. G and M12 share 1 fast SNP and G is common in Japanese but M12 is common in Indians.

Both 8071 and 10398 are slow sites under our definition but appear to be the fastest among the slow sites. They undergo frequent recurrent mutations. While most M have G allele and most N or R have A allele for both sites, some M (D, L0, L1) have 8701A, and some R/N (all B5, some B2d, B4c1a, H, and K, and most I and J) have 10398G. Some M (some L3e1a, L2, and C, and all D5) have 10398A. Also, 10398 is variable among archaic humans. Two Denisovans (FR695060 and FN673705) have A and one (KT780370) has G. It is also variable among apes with Chimpanzee as A and gorilla and orangutan as G. So, the existing tree claiming evolution direction from 10398A to G as if it were a one way street is contradicted by Denisovans and great apes and by the numerous recurrent changes among modern humans. The only sensible and consistent solution to the seemingly messy situation with these two sites is to consider their recurrent changes as adaptive both in terms of physiology and environment. To evolve from A at both 10398 and 8701 to G (that is from R to M) is consistent with the reality of R being far more common in ancient AMH DNAs. One can even observe R to M transitional haplotype such as B5 that has made the step to 10398G but not yet at 8701. Here, B5 is placed as a downstream sub-branch of R, consistent with R being the center ancestral type directly giving rise to many haplotypes.

Overall, the mtDNA tree here shows a star like expansion pattern from the original type. Most haplotypes within R and N branch derive directly from the original ancestral type. This is consistent with the expectation that following the original AMH ancestor in E. Asia, there should be many female offspring who migrated to many different areas and independently produced many different haplotypes. All these R and N related haplotypes are parallel to M. That M is a large haplogroup with two defining slow SNPs indicates a great population expansion some time after M had formed, likely as a result of AMH admixture with Neanderthal like humans, thus acquiring 8701 and 10398 mutations. During the population expansion, many females carrying M with only 8701 and 10398 mutations migrated to areas around South East Asia, and produced offspring and many different sub-haplotypes within M. The M carrying females were likely to be of low social economic status (SES) and vastly outnumber those females of high SES, and hence left more offspring, which may help making M a large haplogroup. It is important to note the dramatic difference between Y and mtDNA in terms of SES and popularity: Y haplotypes of ancient high SES individuals should be more popular today than those of low SES individuals, but mtDNA haplotypes should show the opposite pattern. An ancient male high SES individual may have mated with many females, most of which could only be expected to be of low SES individuals.

The L hapolgroup can be divided into two classes L3 and non L3 (L0’1’5’2’4’6), which is consistent with the distribution pattern where all haplotypes within non-L3 are associated with Bushman and Pygmy or their related populations in Africa (Supplementary Fig. S4). All these L haplotypes have defining slow SNPs and the total number of these L branches are much few than those of non L haplotypes within M. This more hierarchical pattern indicates that there were only small number of female ancestors who migrated into Africa from Asia.

Supplementary Information 5

Admixture analysis

We used the common software, ADMIXTURE, to study ancestry relationships. ADMIXTURE implements a model-based estimation of ancestry in unrelated individuals, which may detect relatively recent admixture events in a population ([*7*](#_ENREF_7)). Admixture v1.3.0 was used to determine the ancestral population components. Input data were prepared using the same procedure as for the PCA. To explore ancestral populations, we used K=2–12. The cross validation (CV) method, implemented in ADMIXTURE, was used to estimate the best k value. All parameters were set to default. Despite the low number of SNPs in our dataset, the data was pruned for LD as ADMIXTURE generally assumes unlinked loci. PLINK was used to calculate an LD (r2) score for each pair of SNPs in a window of 200 SNPs, and one SNP from the pair was excluded if r2 > 0.4. This LD pruning removed 125 SNPs from a total of 15435 SNPs and the LD pruned dataset was used for ADMIXTURE analysis.

As any software must have assumptions which may or may not realistic, we tested whether the ADMIXTURE software can produce consistent results. We studied which group in the 1KG is most related to the Mbuti group here consisting of 4 genomes from SGDP. As shown in Supplementary Fig. S19, Mbuti was found related to Africans at K <7, and most related to LWK at K=7 or 8. But unexpectedly, at K=10 or 12, Mbuti was found most related to YRI. We also tested the affinity of the ancient Ust’-Ishim to specific groups in the 1KG but again failed to obtain consistent results (Supplementary Fig. S20). Depending on the K values selected, Ust’-Ishim could be mostly related to non-African (K =2), CHS (K=3 or 8), CEU (K=4 or 6), GIH (K=5, 9, or 11), or LWK (K=7). Therefore, the ADMIXTURE software failed to produce consistent results. Given such, it may be premature to use the software and we have instead used other methods with few uncertain assumptions as described in the main text.

Supplementary Information 6

References:

1. S. Huang, The overlap feature of the genetic equidistance result, a fundamental biological phenomenon overlooked for nearly half of a century. *Biological Theory* 5, 40 (2010).

2. S. Huang, Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing tarsiers. *Sci China Life Sci* 55, 709 (2012).

3. R. Nielsen *et al.*, A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* 3, e170 (Jun, 2005).

4. M. A. Bakewell, P. Shi, J. Zhang, More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc Natl Acad Sci U S A* 104, 7489 (May 1, 2007).

5. F. L. Mendez, G. D. Poznik, S. Castellano, C. D. Bustamante, The Divergence of Neandertal and Modern Human Y Chromosomes. *Am J Hum Genet* 98, 728 (Apr 07, 2016).

6. M. J. Johnson, D. C. Wallace, S. D. Ferris, M. C. Rattazzi, L. L. Cavalli-Sforza, Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* 19, 255 (1983).

7. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655 (Sep, 2009).

8. G. D. Poznik *et al.*, Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* 341, 562 (Aug 2, 2013).

9. I. Pugach, F. Delfin, E. Gunnarsdottir, M. Kayser, M. Stoneking, Genome-wide data substantiate Holocene gene flow from India to Australia. *Proc Natl Acad Sci U S A* 110, 1803 (Jan 29, 2013).

10. M. Rasmussen *et al.*, An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334, 94 (Oct 7, 2011).

## Supplementary Tables:

## Table S1. A randomly selected set of 255K SNPs from 1KG. See separate file 1.

## Table S2. Slow SNPs set: nonsyn SNPs in slow evolving proteins from 1KG, 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.

## See separate file 2.

## Table S3. List of Syn and Nonsyn SNPs in slow evolving proteins from 1KG.

## See separate file 3.

## Table S4. Nonsyn SNPs in 361 autosomal proteins with 800-1102 aa in length with identity >99% but <100% between human and monkey.

## See separate file 4.

## Table S5. Overlap ratio in slow and fast evolving proteins.

## See separate file 5.

## Table S6. 34 genes used for mutation rate calculation.

## See separate file 6.

## Table S7. Mutations inconsistent with the existing Y phylogeny as reported in Poznik et al. (2013) ([*8*](#_ENREF_8)). Also in separate file 7.

## .

## Table S8. Y and mtDNA haplotypes of the 1000 genomes samples.

## See separate file 8.

## Table S9. SNPs defining major haplogroups among 58251 cleanly called SNPs and their genotypes in the 1000 genomes.

## See separate file 9.

## Table S10. mtDNA SNPs in 1KG and four archaic humans with fast and slow status indicated.

## See separate file 10.

## Table S11. Number of slow mtDNA SNPs in haplotypes among 1KG.

## See separate file 11.

## Table S12. Sharing of archaic alleles in non-L haplotypes and L3 and L6a among 1KG.

## See separate file 12.

## Table S13. Rankings among 1KG mtDNA haplotypes in distance to archaic mtDNAs.

## See separate file 13.

## Table S14. Affinity of 1KG mtDNA haplotypes to European- or African-like archaic mtDNAs.

## See separate file 14.

## Supplementary Figures:

## 

## Fig. S1. Genetic diversity patterns as measured by different types of SNPs. Average pairwise genetic distances (PGD), either hom or total, within each of the five groups in the 1000 genomes are shown for splicing SNPs (A and B), nonsyn SNPs in 361 autosomal proteins (800-1102 aa in length with identity between human and monkey >99% but <100%) (C and D), and syn SNPs in slow evolving proteins (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) (E and F). 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.

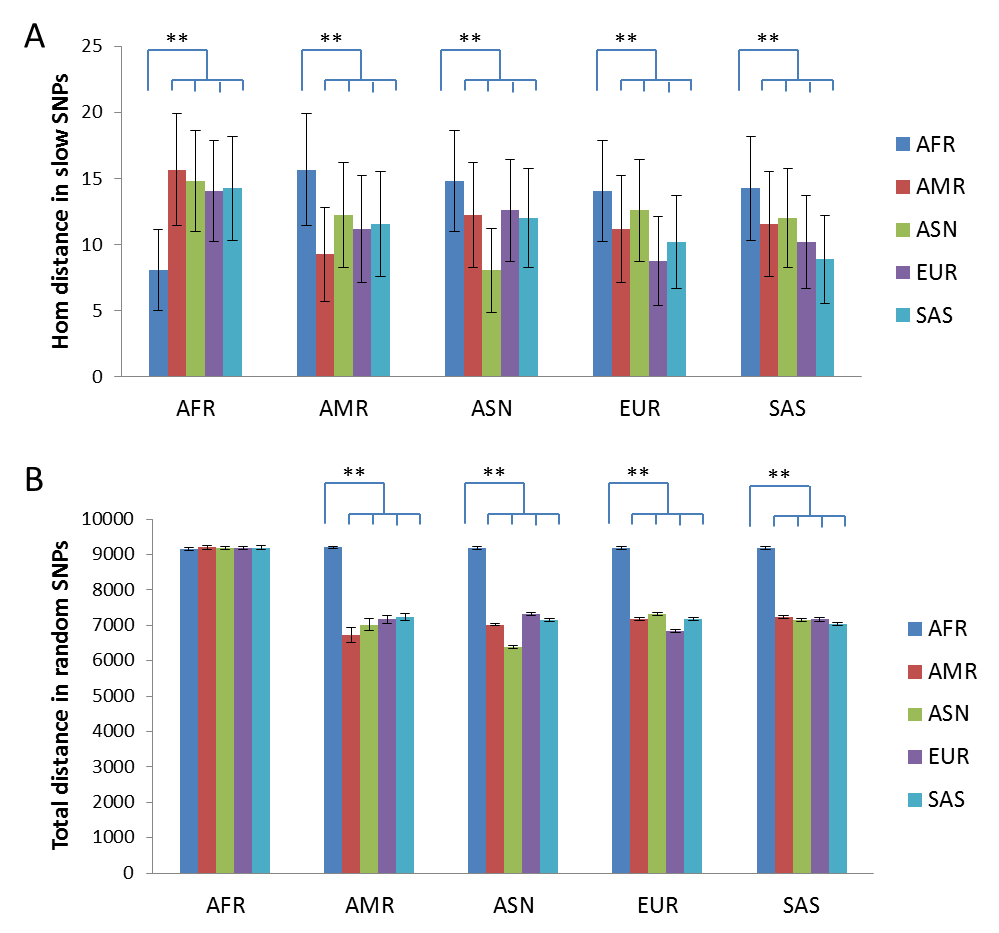


Figure S2. Genetic distance between human groups as measured by different types of SNPs. A. Genetic distance (Hom) measured by slow SNPs. B. Genetic distance (Total) measured by fast SNPs (255K randomly selected SNPs). Data are means with standard deviation. \*\*, P<0.01, t test, 2 tailed.

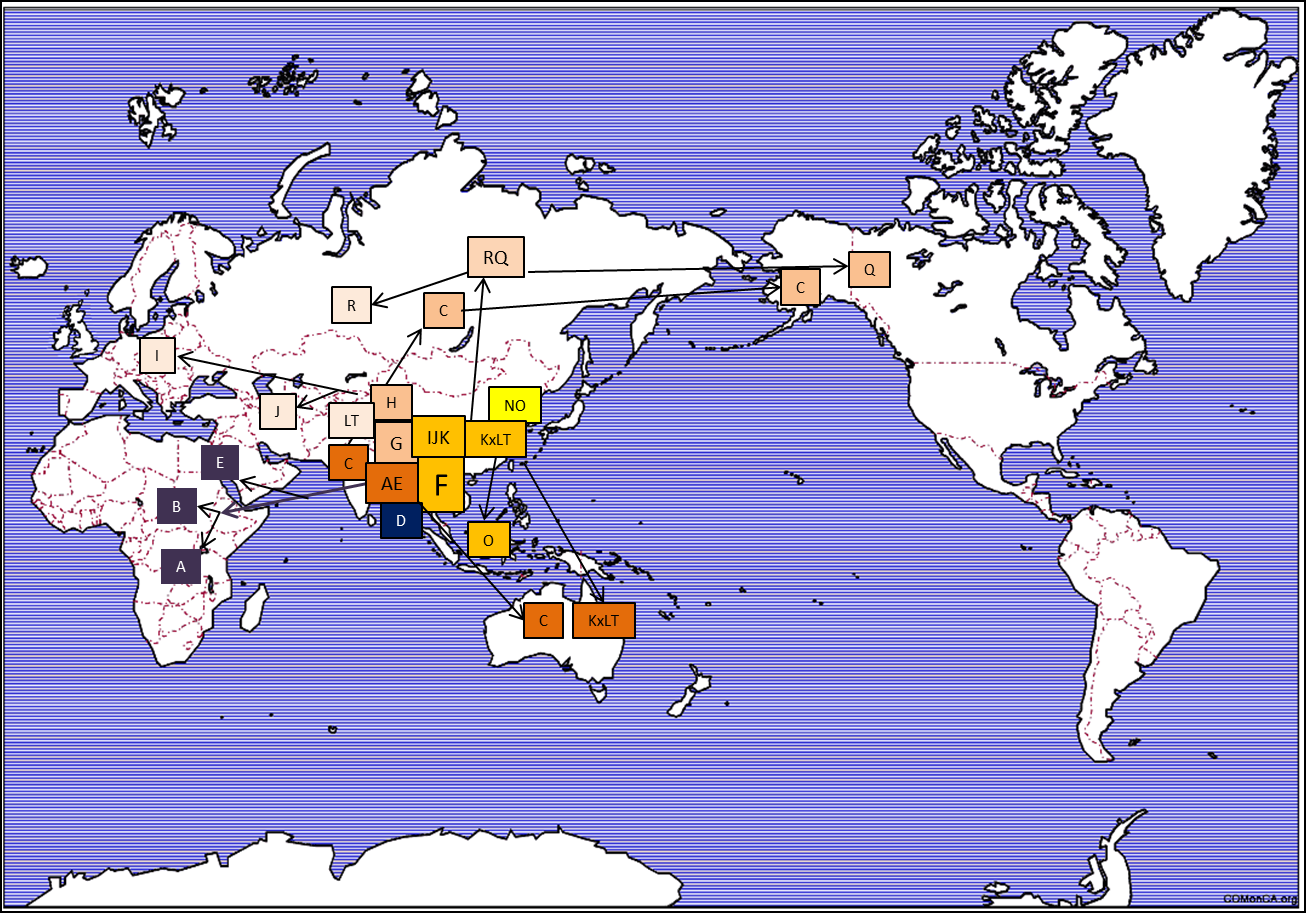
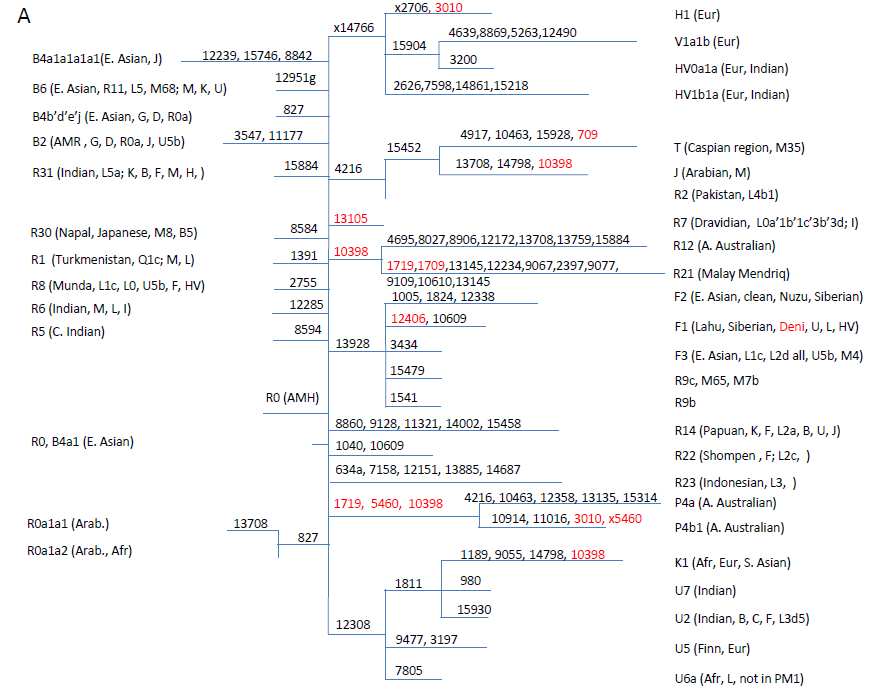
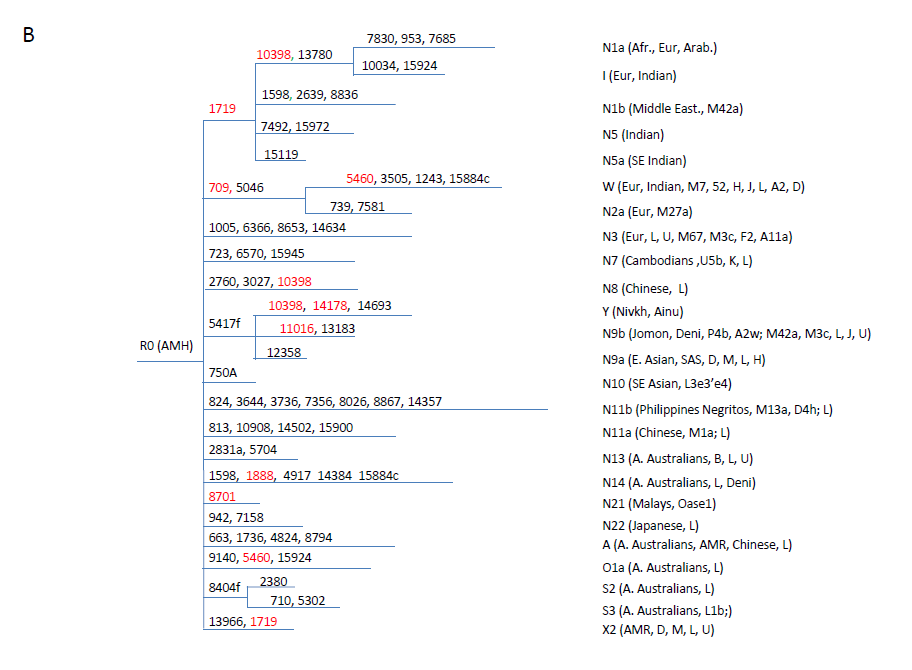
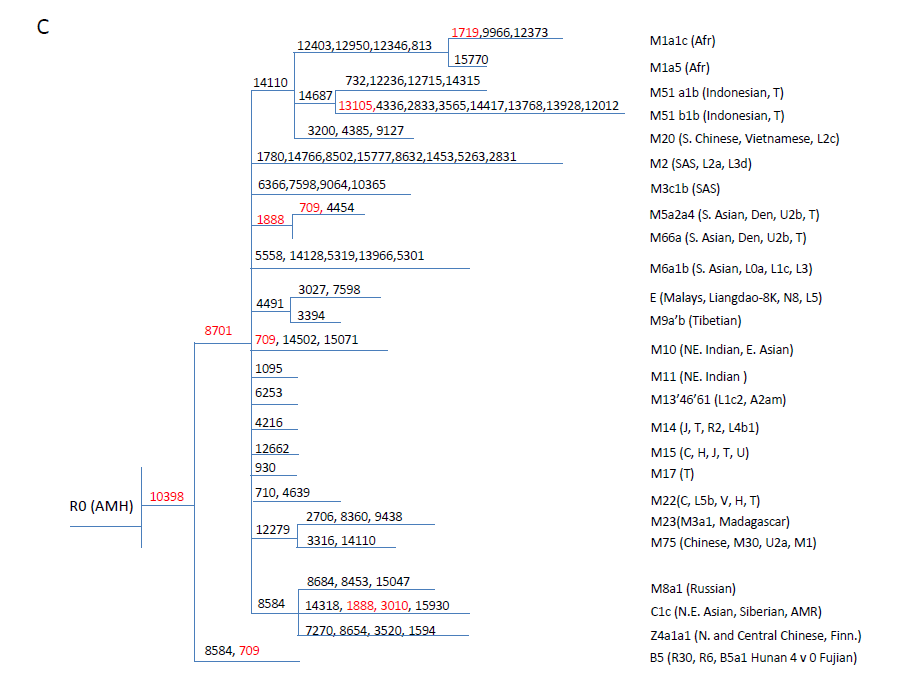
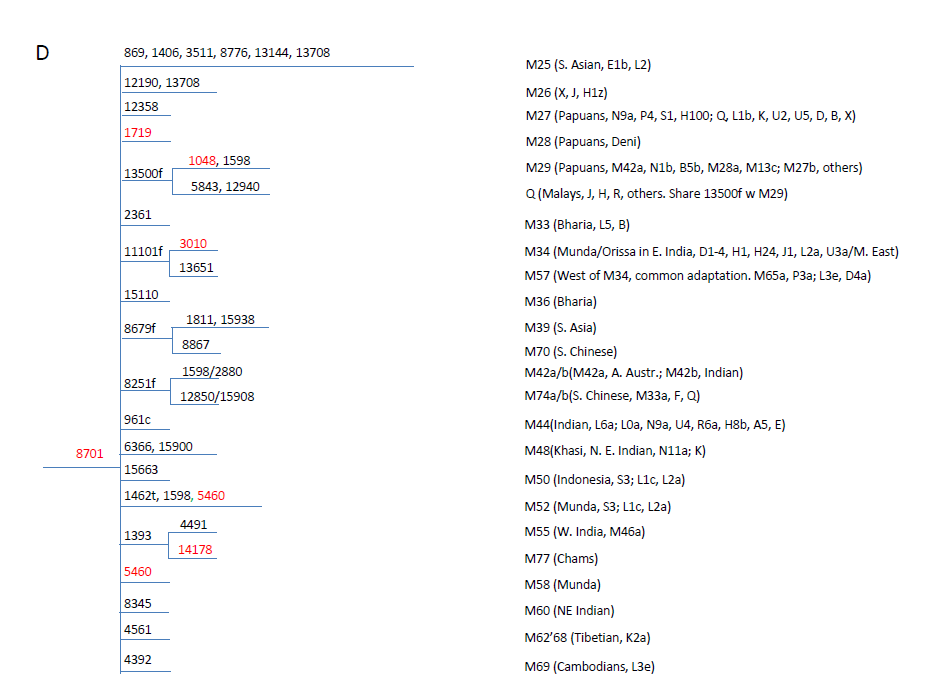
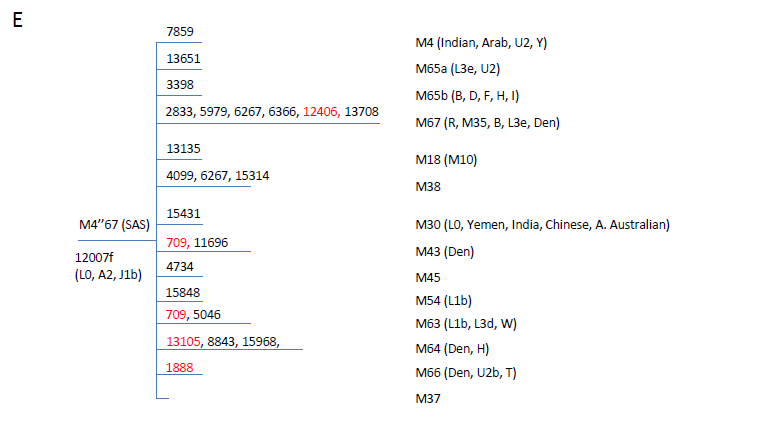
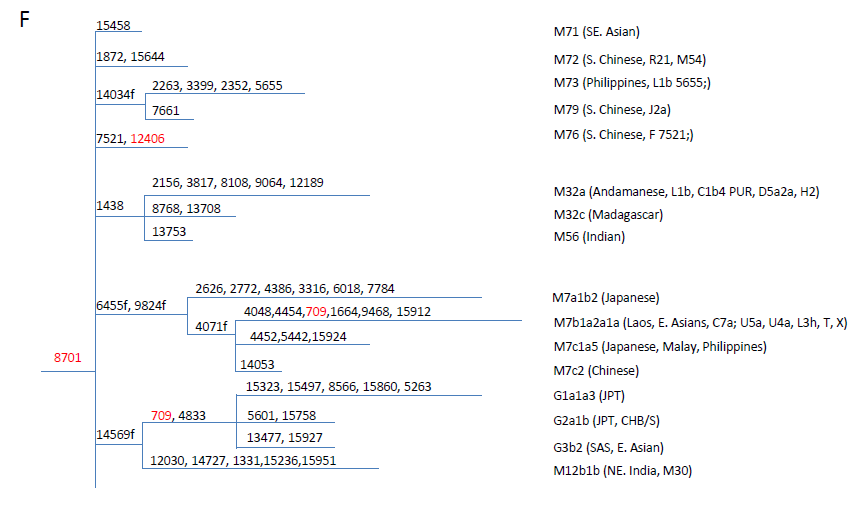


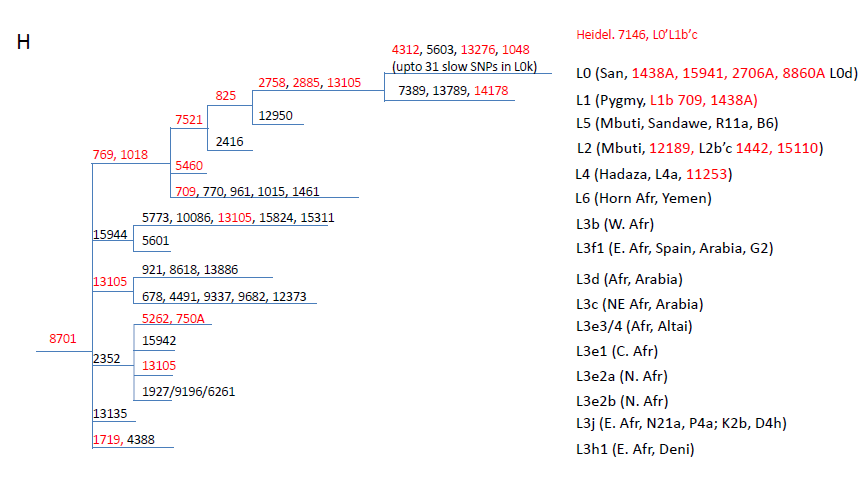
Fig. S3. Y chromosome migration map. The location and direction of migration of major Y haplotypes are shown on a world map. The map image was downloaded from Fotolia.





Supplementary Fig. S4. mtDNA haplotype tree as drawn by slow SNPs. Slow SNPs are shown as indicated and those shared with archaic mtDNAs are red colored. Certain fast SNPs such as 12007 are also shown as indicated by f (12007f). Haplotype defining SNPs that are identical to rCRS are indicated by “x”, such as x14766. Branch length is roughly to scale and branches that have no branch defining slow SNPs are indicated by very short lines about one third of the length of one SNP. Human groups associated with a haplotype and other haplotypes with recurrent SNPs are indicated in parenthesis. A. Tree for selected R haplotypes. B. Tree for selected N haplotypes. C. Tree for selected M and B haplotypes sharing SNP 10398. D. Tree for selected M haplotypes. E. Tree for selected M haplotypes M4”67. F. Tree for selected M and G haplotypes. G. Tree for selected D haplotypes. H. Tree for selected L haplotypes.

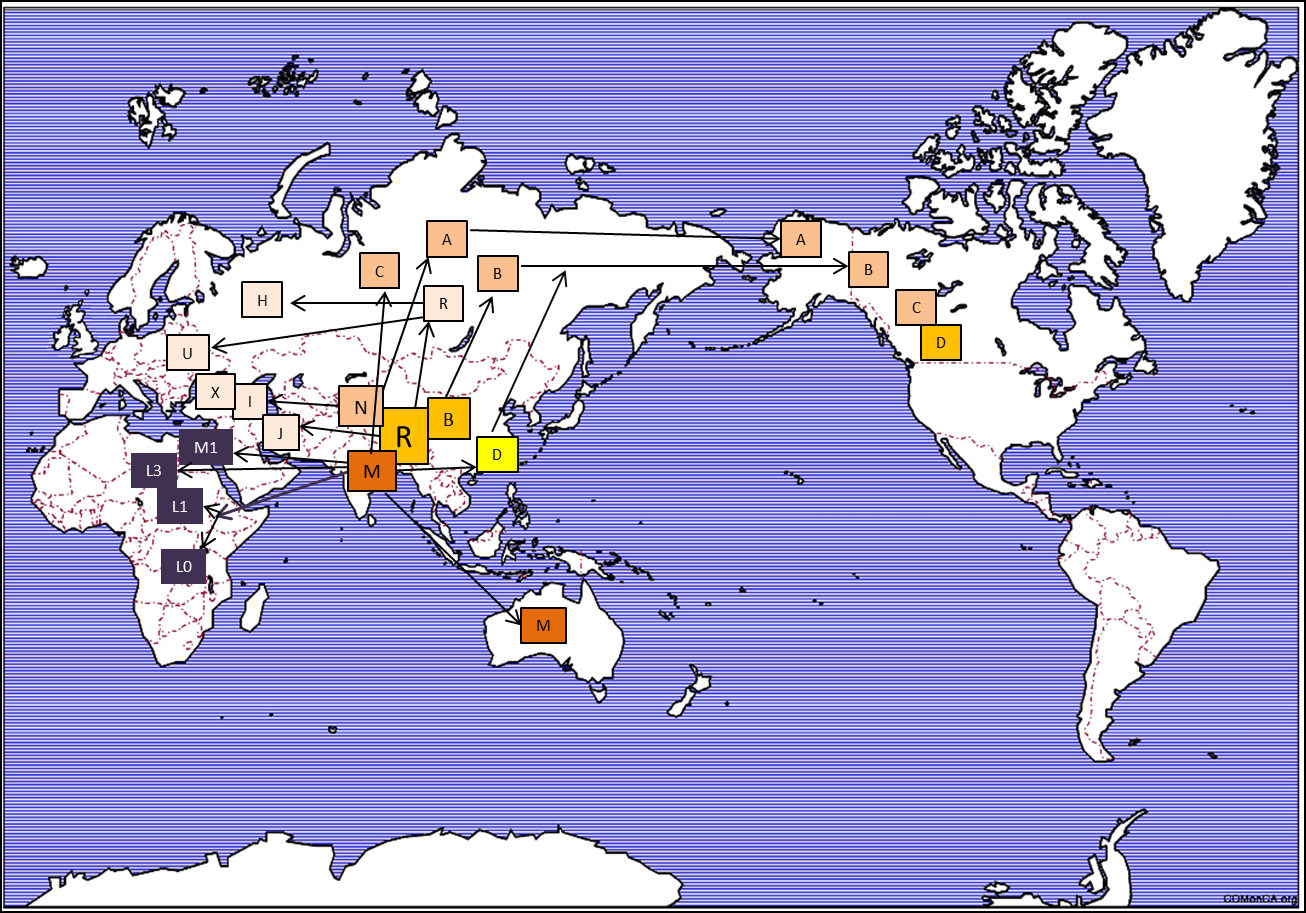


Fig. S5. mtDNA migration map. The location and direction of migration of major mtDNA haplotypes are shown on a world map. The map image was downloaded from Fotolia.

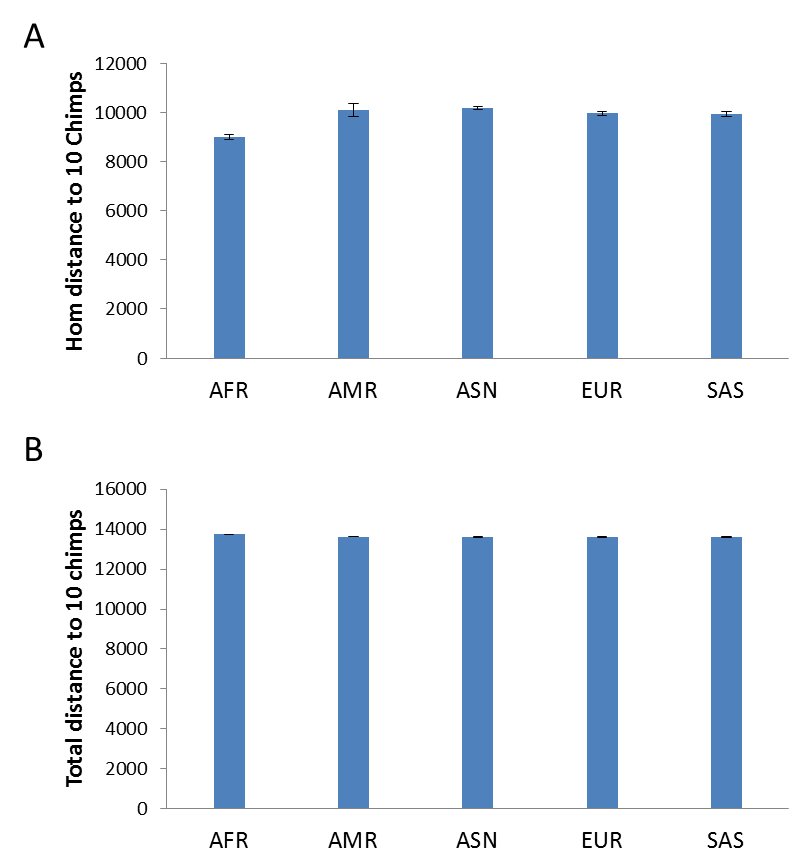


Fig. S6. Genetic distance between chimpanzees and different human groups in fast SNPs. Shown is the hom (A) or total distance (B) between 10 chimpanzees and the 5 human groups of the 1000 genomes as measured by the randomly selected set of 255K SNPs as defined in the main text. Data are means with standard deviation.

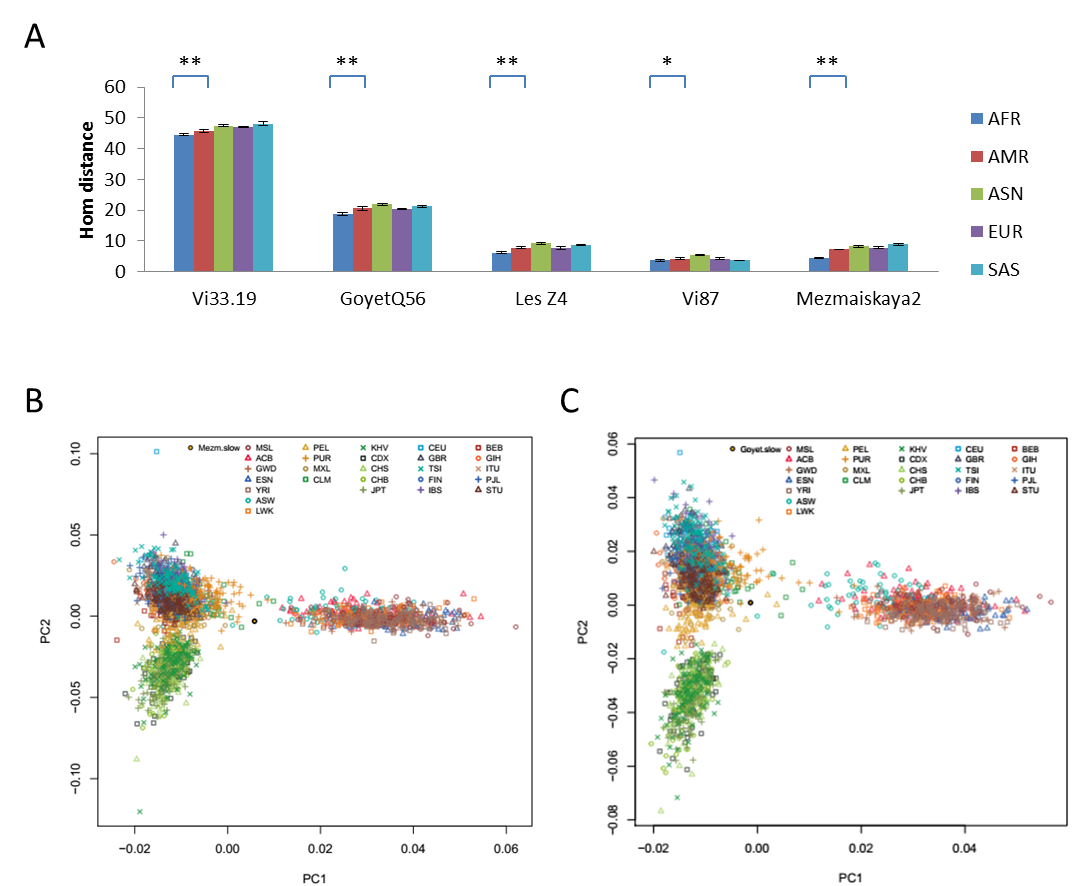


Fig. S7. Autosomal relationship between archaic and modern humans. (A) Shown are the genetic distances between the 5 groups of 1KG and Neanderthals. Data are means with standard deviation. \*\*, P<0.01, t test, 2 tailed. (B) and (C) Shown are PCA plot analyses for Mezmasikaya2 and GoyetQ56.

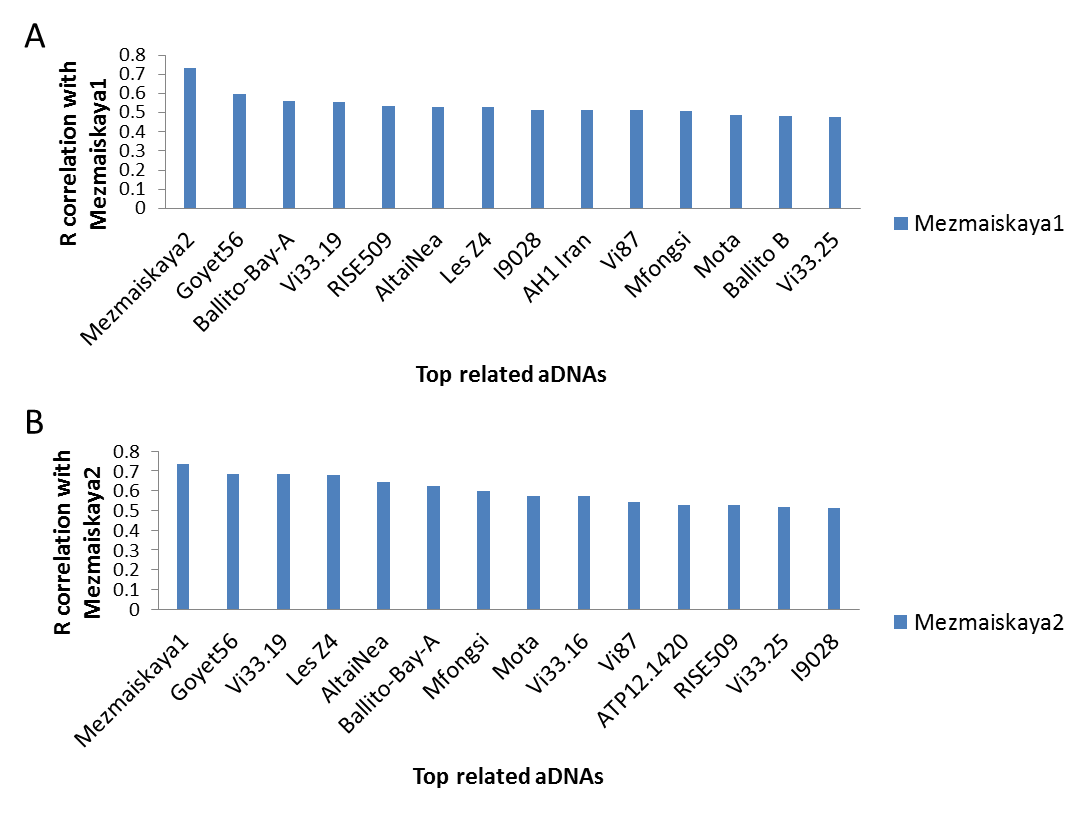


Fig. S8. Regional continuity of Neanderthals in Mezmaiskaya cave. Hom distances to 1KG samples in slow SNPs were calculated for each aDNA in a collection of 146 aDNAs including all Neanderthals sequenced so far. R correlation values between Mezmaiskaya1 or 2 with each aDNA in the 146 aDNA set were obtained using their distances to 510 EUR samples in 1KG. A. Shown are R values of aDNAs most related to Mezmaiskaya 1. B. Shown are R values of aDNAs most related to Mezmaiskaya 2.

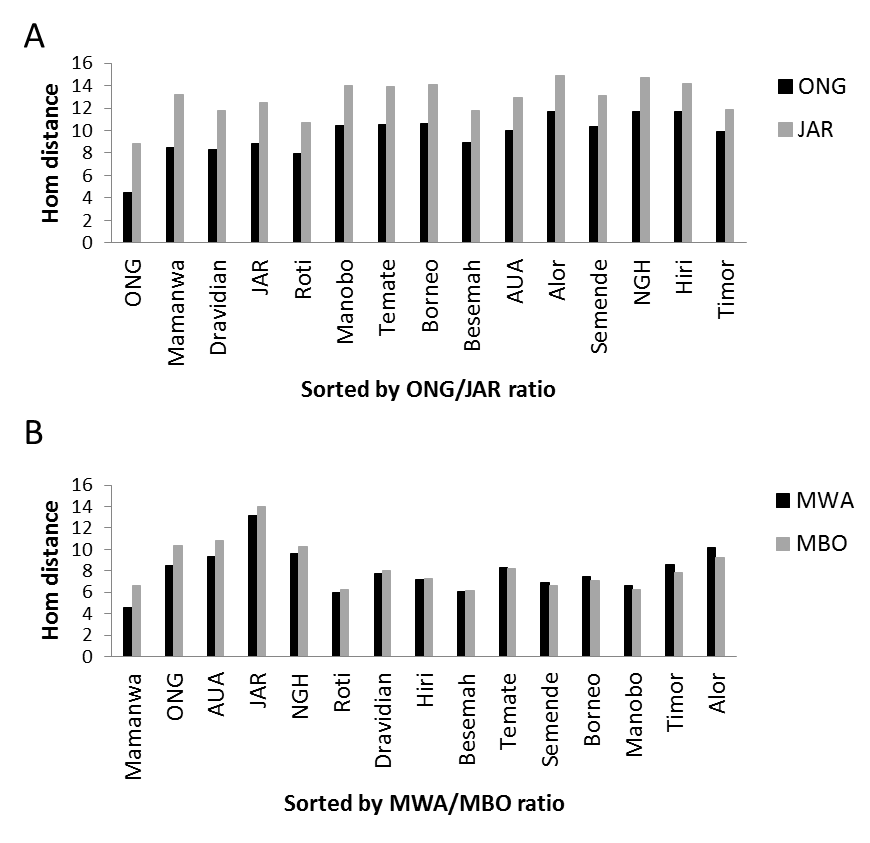


Fig. S9. Relationship between different Negrito groups in South Asia. A. Autosomal distances to ONG or JAR are shown for each of the Indian and South Asian island groups as sampled by Pugach et al (2013) ([*9*](#_ENREF_9)). The groups were ordered based ONG/JAR distance ratio from small to large. B. Autosomal distances to Mamanwa (MWA) or Manobo (MBO) are shown for ONG, JAR, and each of the Indian and South Asian island groups as sampled by Pugach et al (2013)([*9*](#_ENREF_9)). The groups were ordered from small to large based MWA/MBO distance ratio.

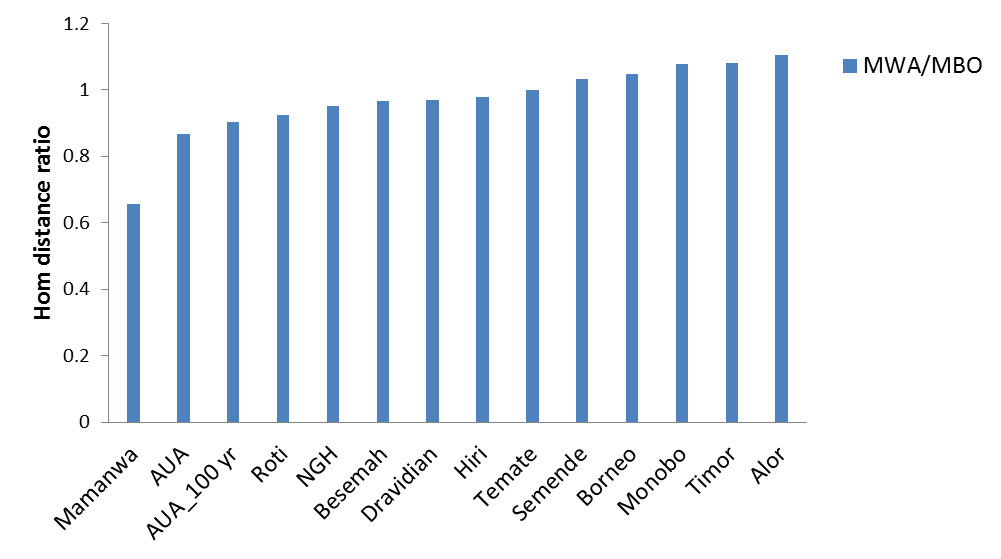


Fig. S10. Relationship between Aboriginal Australians (AUA) and Negritos. Distance to Mamanwa (MWA) or Manobo (MBO) are shown for a 100 year old AUA genome (AUA\_100 yr) as described by Rasmussen et al. (2011) ([*10*](#_ENREF_10)) and each of the Indian and South Asian island groups that also include AUA as sampled by Pugach et al (2013) ([*9*](#_ENREF_9)). The groups were ordered from small to large based MWA/MBO distance ratio.

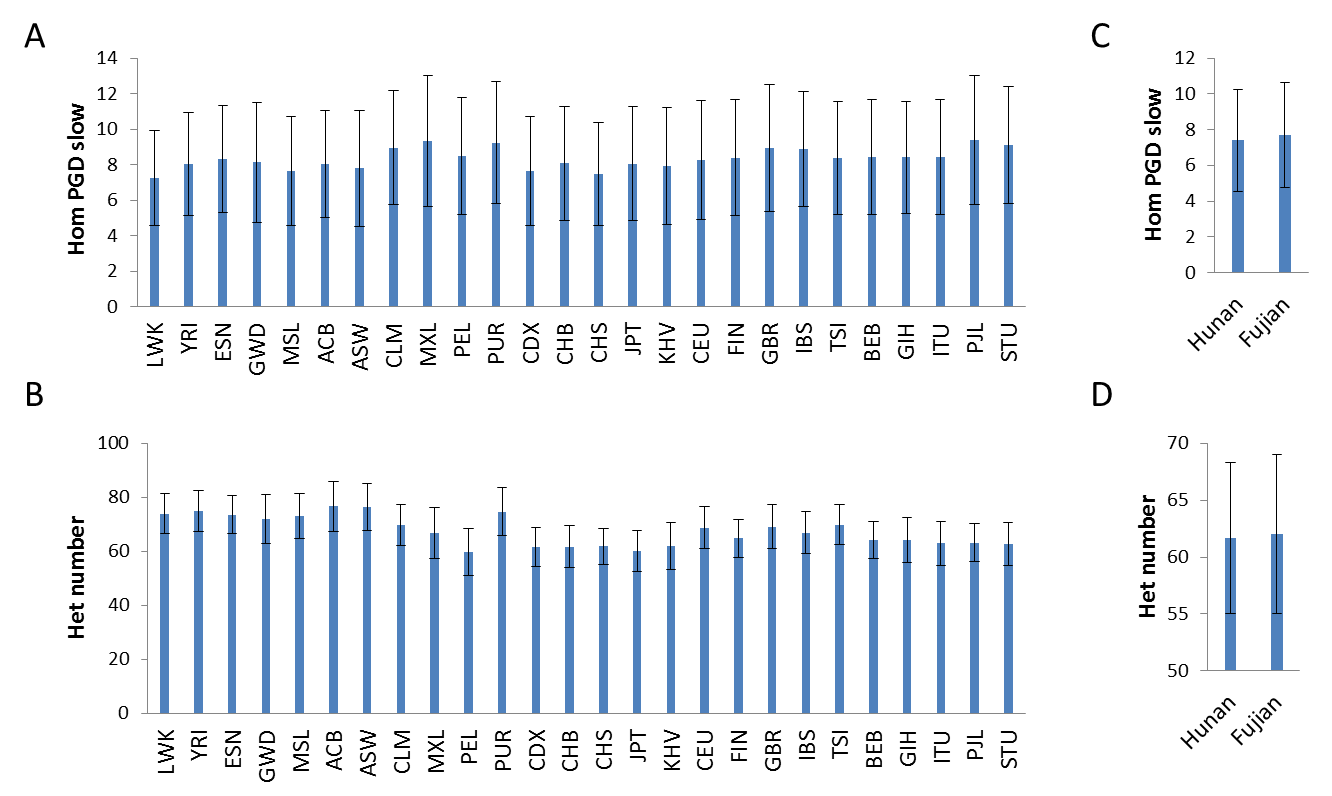


Fig. S11. Genetic diversity levels of different human groups in slow SNPs. A. Average pairwise genetic distance (Hom) in slow SNPs within each of the 25 groups of the 1000 genomes project. B. Average number of heterozygous sites in slow SNPs for each of the 25 groups. C. Average pairwise genetic distance (Hom) in slow SNPs within Hunan and Fujian group. D. Average number of heterozygous sites in slow SNPs for Hunan and Fujian group. Error bars indicate standard deviations. Data are means with standard deviation.

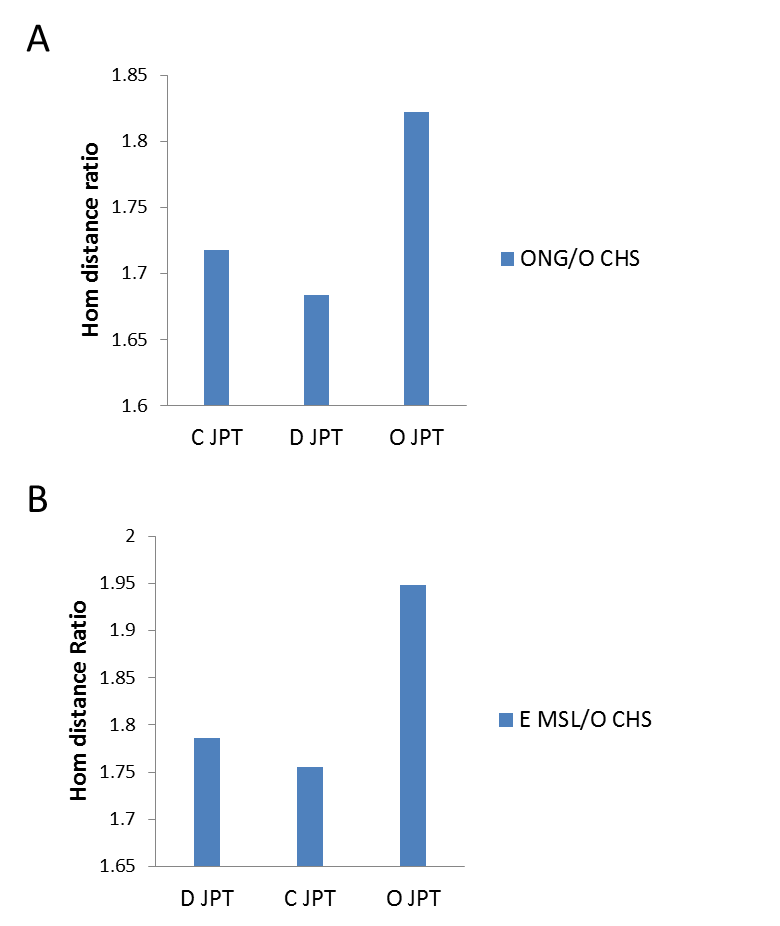


Fig. S12. Autosomal genetic distance in slow SNPs between Japanese and groups with African type Y haplotypes. A. Ratio of autosomal distance to Onge versus CHS with O haplotype for JPT individuals with C, D, or O haplotype. B. Ratio of autosomal distance to MSL with E haplotype versus CHS for Japanese individuals with C, D, or O haplotype.

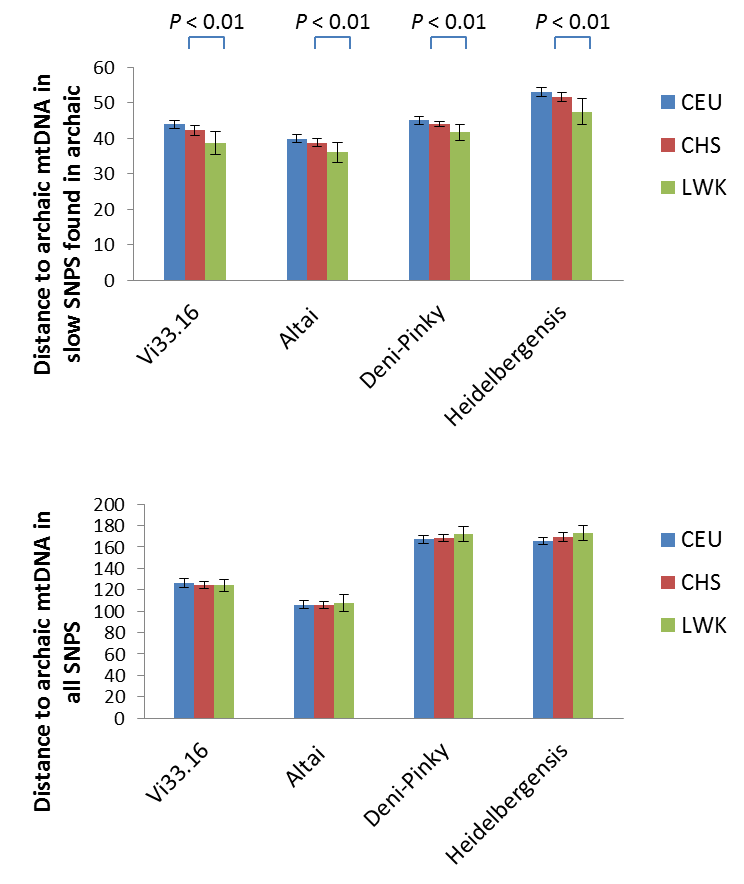


Fig. S13. Distance between archaic and modern mtDNAs. Shown are the average genetic distances between modern (CEU, CHS, and LWK from 1KG) and archaic mtDNAs as measured by using either slow SNPs found in archaic humans (A) or all SNPs (3860 SNPs) found in 1KG. Data are means with standard deviation. P values from t test are shown.

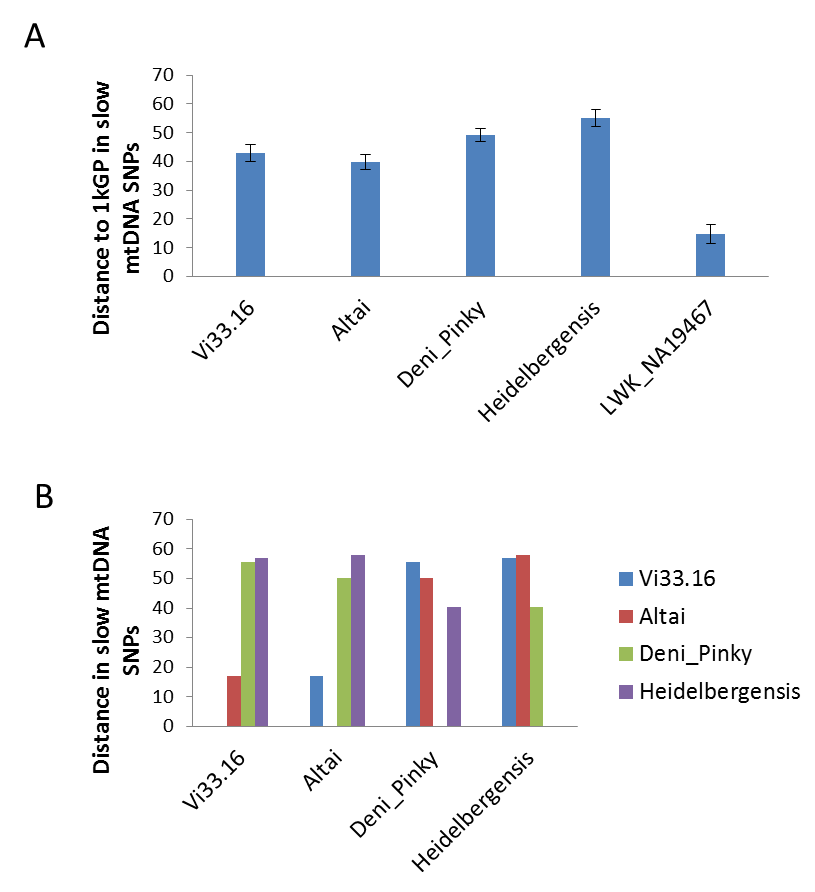
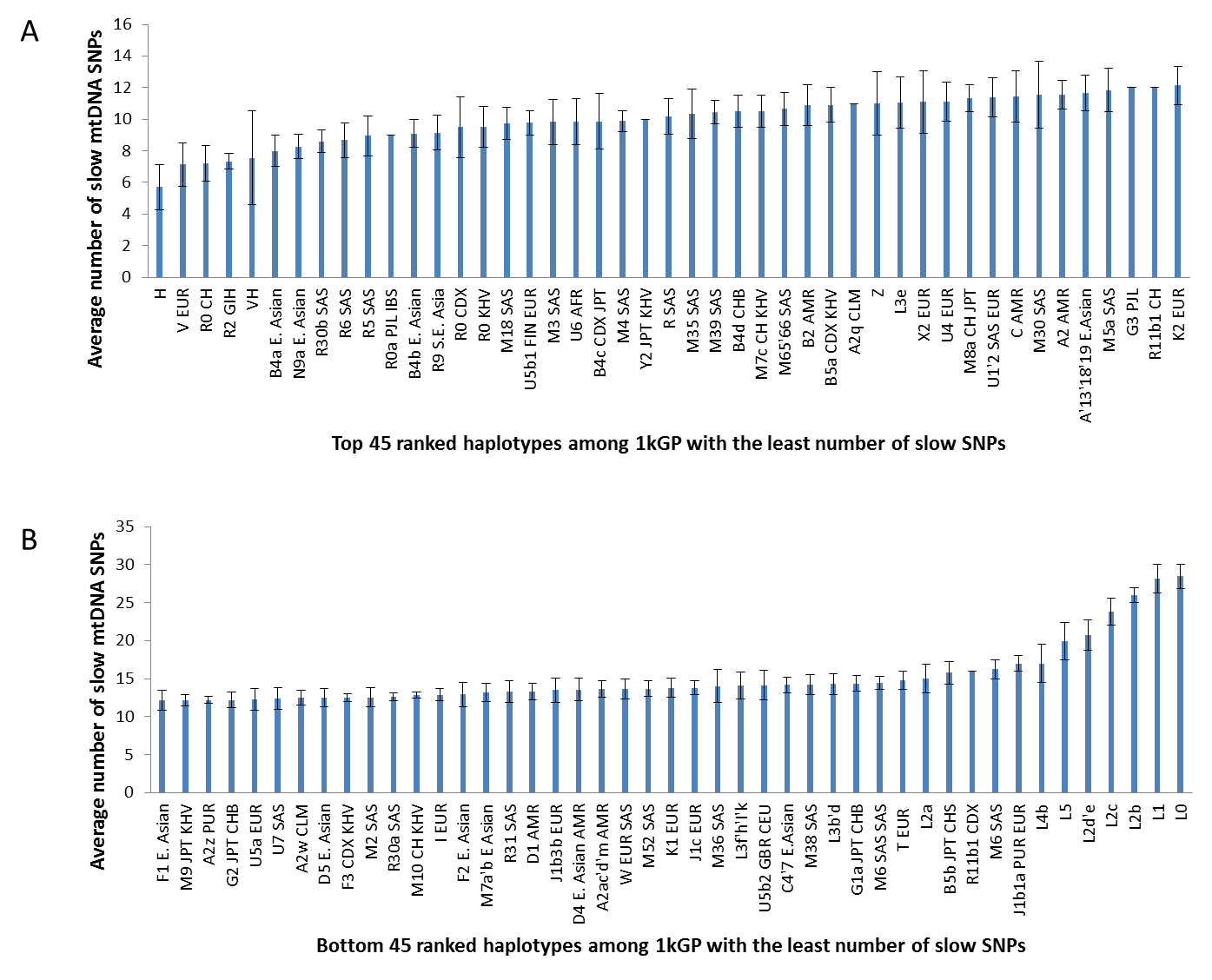
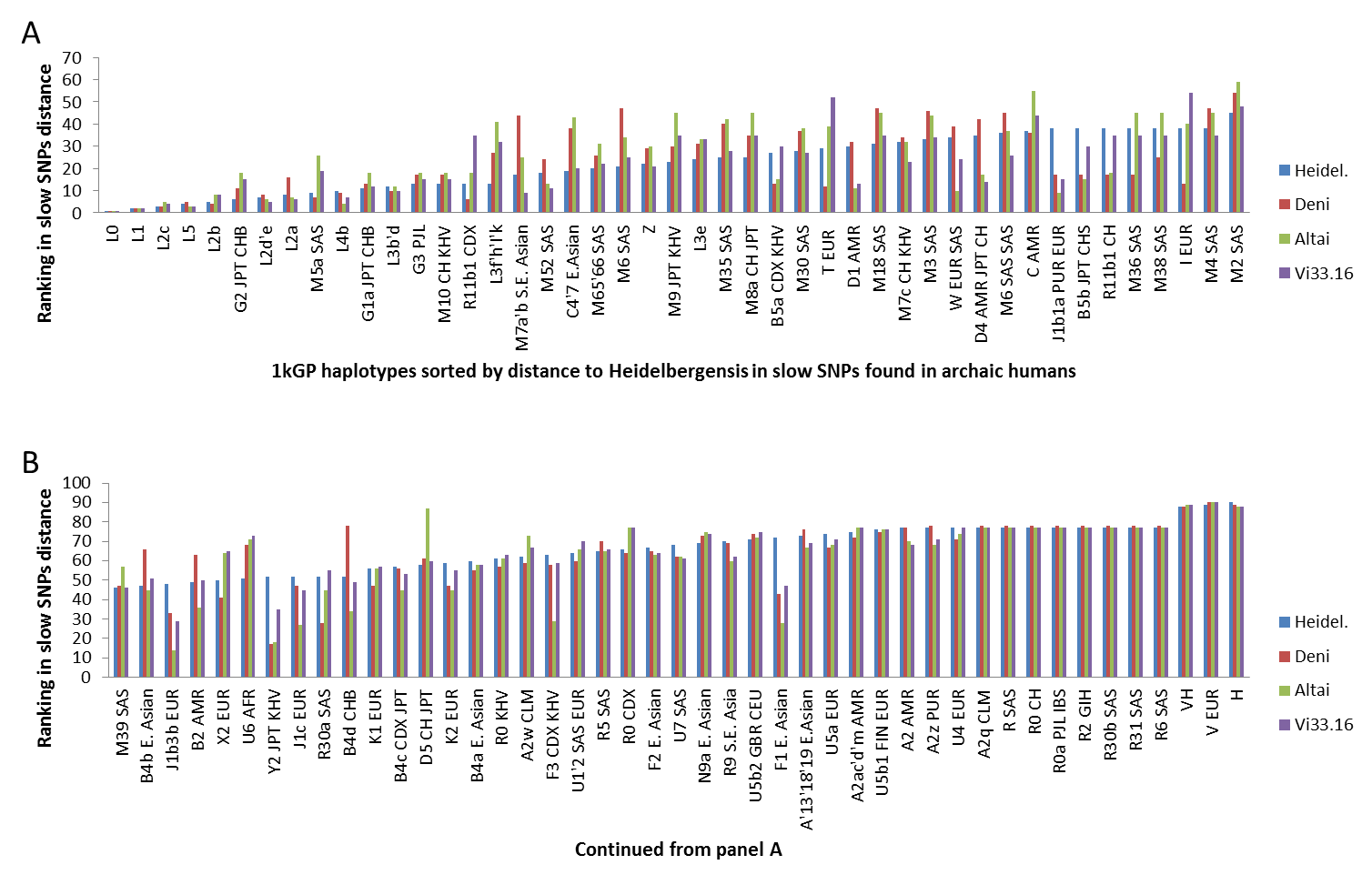


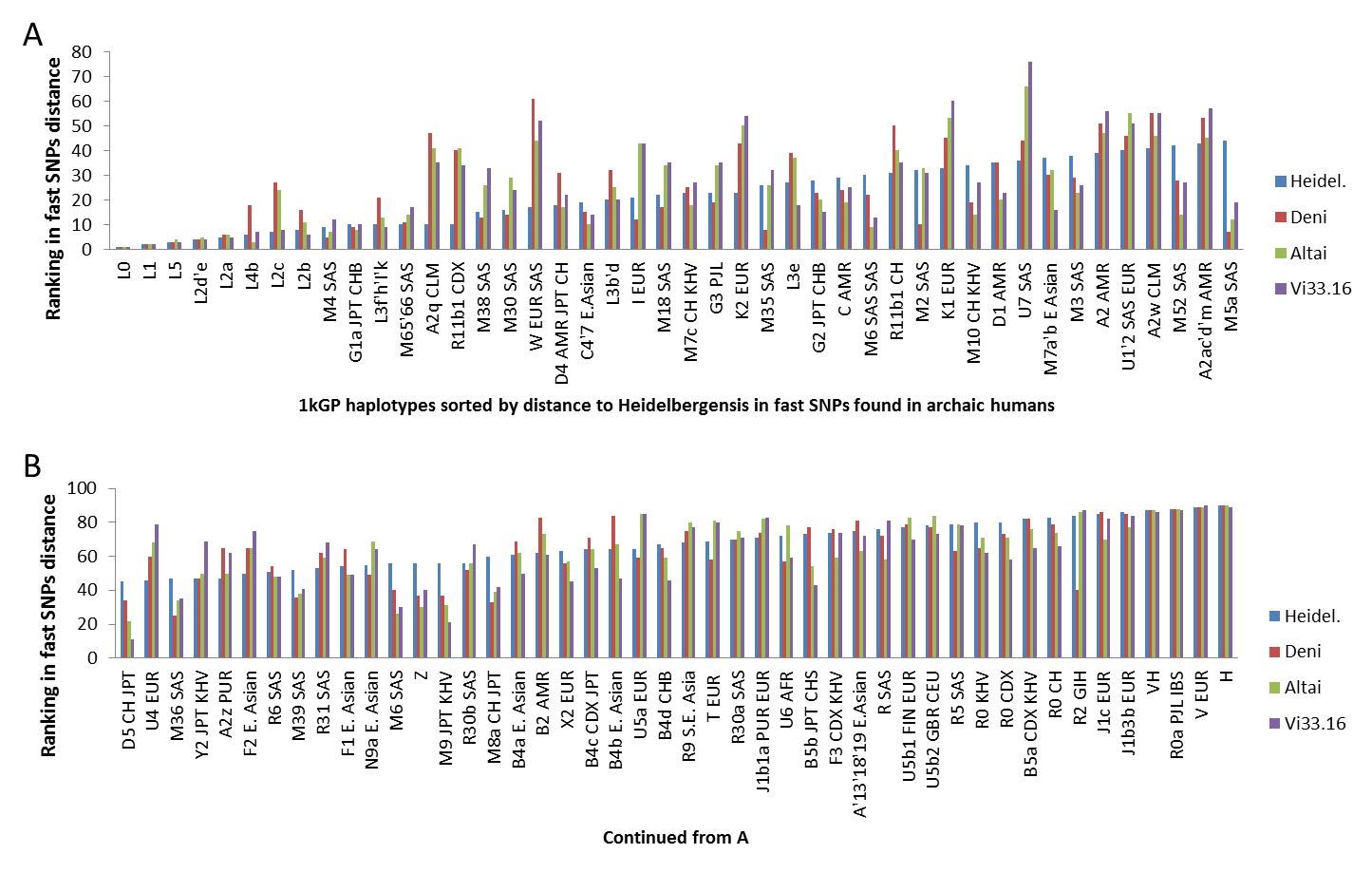
Fig. S14. Archaic mtDNAs as outgroup to modern mtDNAs. A. Genetic distance in slow mtDNA SNPs between modern humans (including all in 1KG) and a randomly selected modern human with the most diverged L0a haplotype (LWK\_NA19467), or four archaic mtDNAs as shown. Data are means with standard deviation. B. Genetic distance in slow mtDNA SNPs among four archaic mtDNAs.

****

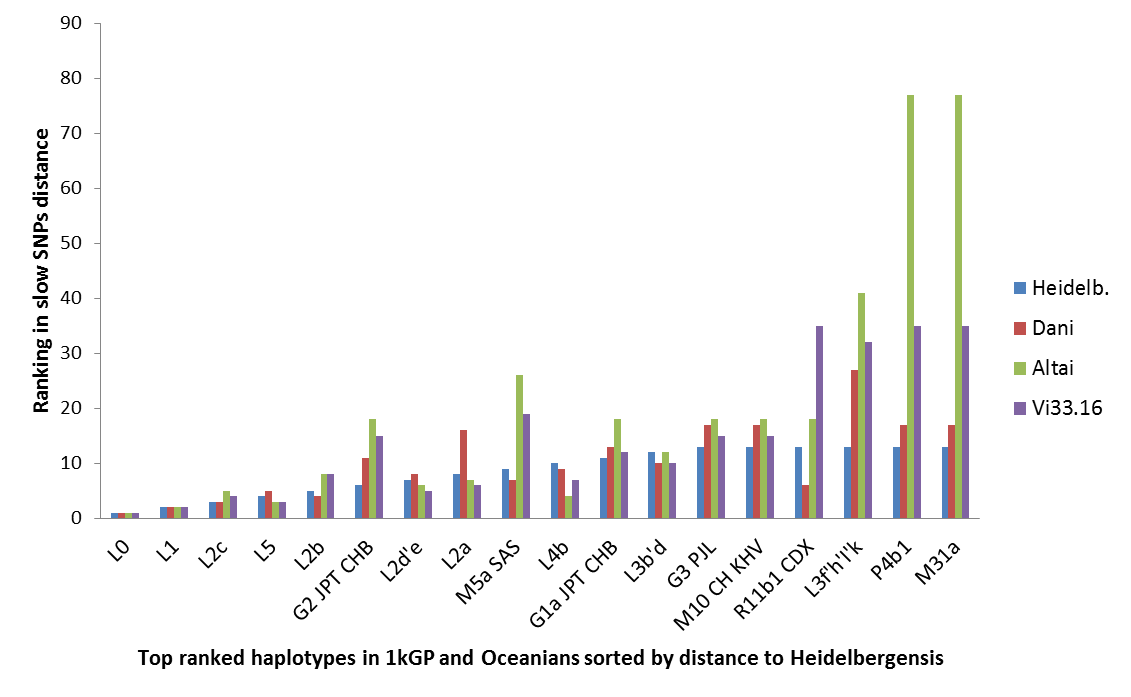
**Fig. S15. Number of slow mtDNA SNPs in major mtDNA haplotypes.** **A.** Top half ranked haplotypes with the least amount of slow SNPs. **B.** Bottom half ranked haplotypes.

****

**Fig. S16. Distance ranking of 1KG mtDNA haplotypes in distance to archaic mtDNAs measured by archaic specific slow SNPs. A.** Top half ranked haplotypes with the smallest distance. **B.** Bottom half ranked haplotypes.

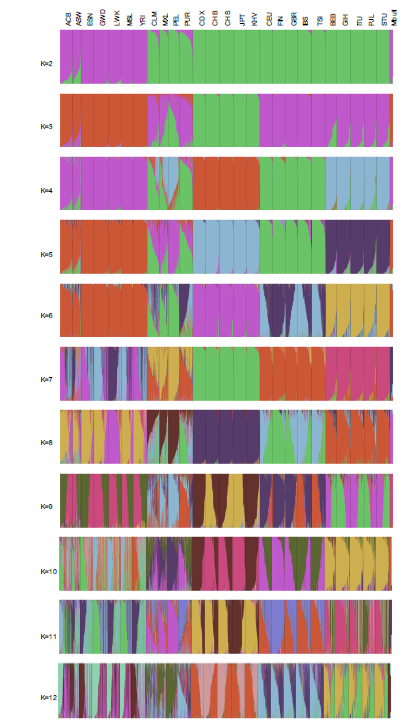
****

**Fig. S17. Distance ranking of 1KG mtDNA haplotypes in distance to archaic mtDNAs measured by archaic specific fast SNPs. A.** Top half ranked haplotypes with the smallest distance. **B.** Bottom half ranked haplotypes.

****

**Fig. S18. Top ranked 1KG mtDNA haplotypes and two Oceanian specific haplotypes in distance to archaic mtDNAs.** Distances were measured by archaic specific slow SNPs. Top 13 ranked 1KG haplotypes and two Oceanian specific haplotypes (P4b1 and M31a) sorted by distance to Heidelbergensis. Ranking ratio of Altai vs Heidelbergensis/Denisovan: 5.1 for P4b1 or M31a, 2.2 for G2, and 3.2 for M5a.

A



B

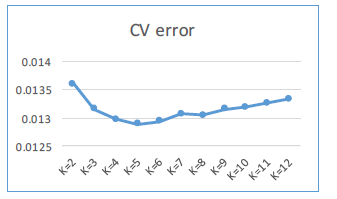
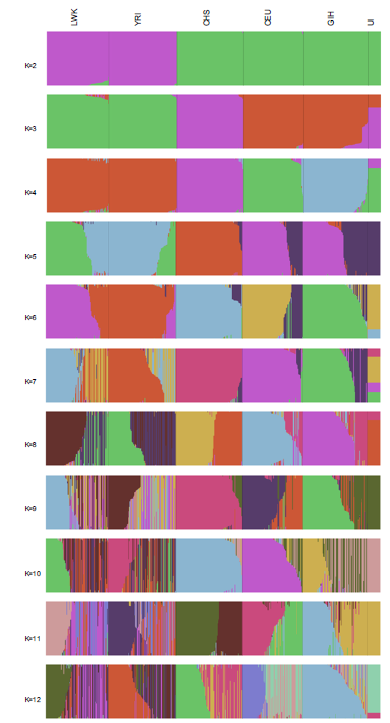


Fig. S19. ADMIXTURE plots for K=2 to 12 on Mbuti and 1KG. ADMIXTURE analysis was done using LD pruned slow SNPs data on Mbuti (4 individuals) and the 1KG. Plots for K=2 to 12 are shown in A, and values of cross validation (CV) are shown in B.

A



B

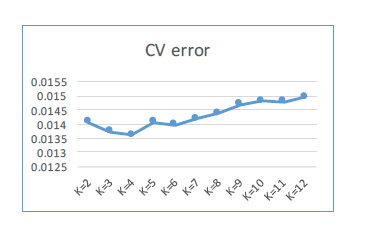


Fig. S20. ADMIXTURE plots for K=2 to 12 on Ust’-Ishim and selected groups from 1KG. ADMIXTURE analysis was done using LD pruned slow SNPs data on Ust’-Ishim (UI) and 5 selected groups from the 1KG. Plots for K=2 to 12 are shown in A, and values of cross validation (CV) are shown in B.