bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

On the peopling of the Americas: molecular evidence for the Paleoamerican and the Solutrean models

Dejian Yuan and Shi Huang*

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

*Corresponding author: Shi Huang, haungshi@sklmg.edu.cn

Abstract

Morphological and archaeological studies suggest that the Americas were first occupied by non-Mongoloids with Australo-Melanesian traits (the Paleoamerican model), which was subsequently followed by Southwest Europeans coming in along the pack ice of the North Atlantic Ocean (the Solutrean model) and by East Asians and Siberians arriving by way of the Bering Strait. Past DNA studies, however, have produced different accounts. With a better understanding of genetic diversity, we have now reinterpreted public DNA data. Consistent with our recent finding of a close relationship between South Pacific populations and Denisovans or Neanderthals who were archaic Africans with Eurasian admixtures, the ~9500 year old Kennewick Man skeleton with Australo-Melanesian affinity from North America was about equally related to Europeans and Africans, least related to East Asians among present-day people, and most related to the ~65000 year old Neanderthal Mezmaiskaya-1 from Adygea Russia among ancient Eurasian DNAs. The ~12700 year old Anzick-1 of the Clovis culture was most related to the ~18720 year old El Miron of the Magdalenian culture in Spain among ancient DNAs. Amerindian mtDNA haplotypes, unlike their Eurasian sister haplotypes, share informative SNPs with Australo-Melanesians, Africans, or Neanderthals. These results suggest a unifying account of informative findings on the settlement of the Americas.

Key words: Paleoamerican hypothesis, Solutrean hypothesis, Kennewick Man, Anzick-1, Neanderthals, Clovis culture, mtDNA haplotypes.

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction:

Morphological analyses of early skeletons in the Americas have suggested that characteristics of some Pleistocene and early Holocene skeletons are different from present-day Native Americans and fall within the variation of present-day indigenous people in South Pacific (Australians, Melanesians, Polynesians, and Negritos) and certain sub-Saharan African groups [1-3]. This is particularly so for the first South Americans, while the first North Americans seem to occupy an unresolved morphological position between modern South Pacific and European populations [4]. No resemblance was noted between the first Americans and either Northeast Asians or modern Native Americans. This has led to the Paleoamerican hypothesis that the initial pioneer population in the Americas had common ancestry with indigenous people in South Pacific which was largely replaced by populations with Northeast Asian affinities in the early Holocene but may have persisted in some locations in South America such as the extinct Pericúes and Fuego-Patagonians [5,6].

Archaeological studies have uncovered the Clovis culture, the oldest widespread archaeological complex defined in North America, dating from 13,000 to 12,600 years ago [7,8]. The culture is thought to be developed in North America from an ancestral technology [9]. Two competing hypotheses have been developed regarding the origins of the people who made Clovis tools. Based on the striking similarities between Solutrean and Clovis lithic technologies, the Solutrean hypothesis suggests that people of the Solutrean culture, 21,000 to 17,000 years ago, in Ice Age Europe migrated to North America along the pack ice of the North Atlantic Ocean during the Last Glacial Maximum [10]. Alternatively, the more conventional hypothesis suggests that people associated with the Clovis culture were from Asia by way of the Bering Strait and the similarities with Solutrean tools are thought to be coincidental [8].

However, neither the Paleoamerican nor the Solutrean model has found support from past molecular researches [11]. Ancient DNA studies on the ~9500 year old Kennewick Man skeleton found in the state of Washington in the United States, which was thought to be closely related to the Ainu and Polynesians on the basis of cranial morphology, have nonetheless grouped him with present-day Native Americans [12]. Moreover, populations such as the Fuego-Patagonians that were considered to be relicts of an early migration into the Americas and closely related to Australo-Melanesians are shown to be genetically related to contemporary Native Americans [13,14]. Genome sequencing of a ~24000 year old Mal'ta (MA-1) skeleton suggests that Native Americans are derived from a mixture of populations that are related to the Mal'ta lineage as well as one or more unknown East-Asian lineages [15]. The ~12700 year old Anzick-1 genome of the Clovis culture was also thought to be derived from MA-1 and directly ancestral to many contemporary Native Americans [16].

Why the dramatic differences between the molecular and non-molecular results? We have found the problem here to lie within the unrealistic assumptions of the present molecular methods. In fact, these assumptions have not even solved the long standing puzzle regarding the determinants of genetic diversity [17]. A new framework of molecular evolution, the maximum genetic diversity (MGD) hypothesis, has recently solved the puzzle and made it now possible for the first time to infer molecular models of human origins based on genetic diversity data [18,19]. It

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

is now known that genetic diversities are mostly at saturation level, which therefore renders most of the past molecular results invalid since those results were based on mistreating saturated phases of genetic distance as linear phases [20-23]. Only slow evolving nuclear sequences are still at linear phase and hence informative to divergence time. For the mtDNA genomes, the relatively slow evolving sequences within mtDNA are non-synonymous sites and RNA genes, which are related to physiology and allow phylogenetic inference based on shared physiology.

New results based on the MGD theory have indeed suggested a unifying account of the origin of modern humans [19]. The time for the first split in modern human autosomes was dated to be 1.91-1.96 million years ago, consistent with the multiregional hypothesis. Modern Y and mtDNA originated in East Asia and dispersed via hybridization with archaic humans. Analyses of autosomes, Y and mtDNA all suggest that Denisovan/Neanderthal like humans were archaic Africans with Eurasian admixtures who have transmitted their genomes mostly into the indigenous people in South Pacific. These new findings immediately make the Paleoamerican model highly plausible. To address this and related models, we here reanalyzed previously published DNA sequences based on the MGD framework.

Results:

Kennewick Man and the Paleoamerican model

We examined the relationship of Kennewick Man with present day human groups in the 1000 genomes project (1kGP) [24]. In genetic distance by informative SNPs (homozygous mismatches in slow SNPs) as defined previously [19], Kennewick Man was closest to South American Peruvians from Lima Peru (PEL), most distant to Asians (closest to Hunan people within Asians), and about equally related to Europeans and Africans (Fig. 1A, and Supplementary Fig. S1A for PCA plot). In contrast, two other ancient DNAs from North America, Anzick-1 and the 4000 year old Eskimo Saqqaq [25], were all more related to Europeans than to Africans (Fig. 1A, Supplementary Fig. S1B for PCA plot). Anzick-1 was closest to Hunan and Southern Han Chinese (CHS) but Saqqaq was closest to Japanese in Tokyo (JPT), Han Chinese in Beijing (CHB), and Fujian population among East Asians, indicating at least two different migrations of East Asians with the ancient one by South East Asians, consistent with modern human origin in South East Asia as recently discovered [19]. These results indicate significant African ancestry in Kennewick Man, more so than East Asian ancestry, consistent with his Australo-Melanesian and African traits. Interestingly, relative to PEL and Mexican Ancestry from Los Angeles (MXL), Kennewick Man was not closer to Puerto Ricans from Puerto Rico (PUR) and Colombians from Medellin (CLM) who are known to have recent African admixtures, suggesting very different African ancestry between Kennewick Man and PUR/CLM (see below).

We next studied the relationship of Kennewick Man with ancient DNAs of Neanderthals and modern humans. Because different ancient samples were sequenced at different coverages, it is unrealistic to use SNP mismatches to infer relationships as there would be few shared SNPs among different samples. As an alternative, we calculated the correlation coefficient R of two

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

genomes in their distance to the 1kGP samples, assuming that different random sampling of a fraction of the whole set of ~15K slow SNPs are roughly equivalent in representing the whole set. We first verified this method by showing that the ~6000 year old Iceman farmer from Italy [26] was more correlated with the ~7000 year old Stuttgart farmer from Germany [27] than with the ~45000 year old Siberian Ust'-Ishim [28] in their distance to all 1kGP samples (Supplemental Fig. S2A-B. See Supplementary Table S1 for distance to 1kGP for 39 aDNAs). We also confirmed that a randomly picked CHS individual from 1kGP was more correlated in average R values with the CHS group of individuals than with the Toscani population in Italia (TSI) regardless whether the distance used in the correlation involved just one of the five major groups in 1kGP or all 1kGP samples (Supplementary Fig. S2C).

Using the correlation approach, we found Kennewick Man to be most correlated with the ~65000 year old Neanderthal Mezmaiskaya-1 from Adygea Russia [29] among all ancient DNAs examined in their distance to all 1kGP samples (Fig. 1B). In distance to only MXL and PEL samples, Mezmaiskaya-1 was among the top 3 most correlated with Kennewick Man, following just Anzick-1 and MA-1 (Supplementary Fig. S3). On a principal component analysis (PCA) plot, Mezmaiskaya-1 was located between Europeans and Africans (Supplementary Fig. S4), similar to Kennewick Man. We have previously shown that Mezmaiskaya-1, Altai [29], and Vi33.16 have more African than non-African genomes [19]. These 3 Neanderthals all showed far more correlation to Kennewick Man than to Anzick-1 and Saqqaq, similar to the ~4500 year old African Mota [30] but unlike the other Neanderthals Vi33.25 and Vi33.26 who have been shown previously to be closer to East Asians and Native Americans, respectively [19]. All other aDNAs of modern humans examined here showed more correlation with Anzick-1 and Saqqaq than with Kennewick Man. While the Denisovan genome [31] also has been shown previously to be more related to Africans than to non-Africans [19], it showed little correlation with Kennewick Man, consistent with its affinity to Australo-Melenesians [19]. These results suggest specific Neanderthal-associated African ancestry in the Kennewick Man genome.

We next asked whether some PEL individuals in 1kGP are like Kennewick Man in having more ancestry from Neanderthal type Africans. Based on correlation with Mezmaiskaya-1 in distance to 1kGP, we selected two PEL samples with one being top ranked and the other bottom ranked among PEL samples, and measured their genetic distance to the 5 major groups in 1kGP. These two PEL samples were about similarly related to non-African groups but the one most related to Mezmaiskaya-1, PEL_HG02006, was much closer to Africans than the other, PEL_HG01927, and more related to Africans than to East Asians (Fig. 2A). This differential relatedness to Africans could not be observed when fast SNPs, a random set of 250K SNPs representing genome average as described previously [19], were used to measure distance, consistent with previous failure to detect African admixtures using what are now realized to be adaptive SNPs (Fig. 2B).

To further confirm the above results that the African ancestry in the PEL samples may not have come from recent African admixture as in the case of PUR and CLM, we examined the genetic distances to 1kGP for all American samples that showed close relationship with Africans as indicated by PCA plot, including 2 PEL, 4 PUR, and 4 CLM samples (Supplementary Fig. S5). Unlike the two PEL genomes that were far closer to American (AMR) than to non-AMR groups

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and similarly related to AFR and ASN, the other samples all failed to show significant relationships with AMR relative to non-AMR samples while were all closer to AFR than to ASN (Fig. 2C). These results confirm that the African ancestry in the two PEL individuals is different from those in the PUR and CLM individuals.

The now extinct South American Fuego-Patagonians are known to show Paleoamerican cranial traits. We next tested whether these Fuego-Patagonians genomes [13] may be more related to Kennewick Man among the three ancient Amerindians. Relative to North Amerindian Eskimos and Aleutians, Kennewick Man was the most related to Puego-Patagonians and Saqqaq the least related, consistent with expectations (Fig. 3A). We also examined ancient and present-day Europeans and found that ancient hunter gatherers from Georgia or Caucasus (CHG, ~9700 year old Kotias and ~13300 year old Satsurblia) were more related to Fuego-Patagonians than to Eskimos/Aleutians [32,33], whereas the opposite was found for ancient Western hunter gatherers, farmers, Eastern hunter gatherers, and present day Europeans (Fig. 3A and Supplementary Fig. S6). Consistently, CHG was more related to Neanderthal Altai and Mezmaiskaya-1 among all ancient DNAs examined (Fig. 3B). While the Spanish farmer CB13 [34] showed slightly higher correlation with Mezmaiskaya-1 than the CHG Kotias did, it was most related to Ust'-Ishim whereas Kotias clearly had more Neanderthal genomes since its top 2 related aDNAs were both Neanderthals.

We next used selected SNPs data from the Human Genome Diversity Project (HGDP) samples and South Pacific individuals as reported by Pugach et al (2013) [35] to further confirm Kennewick Man affinity with South Americans and indigenous people in South Pacific. We calculated the ratio of R score (in MXL_PEL distance) of Kennewick Man vs Anzick-1 as well as of Saqqaq or PEL_HG02006 or PEL_HG01927 vs Anzick-1. Relative to Saqqaq and the two PEL individuals, Kennewick Man was more related to the Negrito group Mamanwa from the Philippines than to its non-Negrito neighbor Monobo, more to Amerindian groups than to others in HGDP, more to South American Surui than to North American Mayan, and more to African Yoruba than to Russian (Fig. 3C). In R scores, Kennewick Man was more related to Aboriginal Australians (AUA), Melanesian, and Mamanwa than to Han Chinese.

Anzick-1 and the Solutrean model

We examined the Solutrean hypothesis by testing whether Anzick-1 may be specifically related to the ~18720 year old El Miron from the Magdalenian culture in Spain that was immediately preceded by the Solutrean [36]. In distance to EUR samples, El Miron was the only ancient DNA among all examined that showed positive correlation only with Anzick-1 but not with the other two ancient Amerindians Kennewick Man and Saqqaq (Fig. 4A). The correlation of other European aDNAs with Anzick-1 likely indicates a general European element but not special ancestry relationship as they were also similarly correlated with the other two ancient Amerindians. The ~7300 year old farmer CB13 from Spain was the most correlated with Anzick-1 among European aDNAs, consistent with a special connection between El Miron and Anzick-1 and local genetic continuity.

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

As the number of informative slow SNPs available for El Miron (582) was relatively small due to low coverage, we verified that they were good enough to show highest R value with La Brana-1 among 25 ancient DNA genomes as might be expected given that both were from Spain (Fig. 4B) [37]. Furthermore, among 230 ancient European genomes from 27 populations [38], two in the top 3 correlated with El Miron were from Iberia. Further confirming the validity of our approach, we found as expected that the ~24000 year old MA-1 belonging to the Gravettian culture in South Central Siberia was most related to the ~30000 year old Vestonice-16 (with 666 slow SNPs) of the same culture from South Moravia of Czech Republic and vice versa among all aDNAs of similar ages from Europe (Supplementary Fig. S7). Unlike previous findings of MA-1 being closest to contemporary Amerindians [15], our results showed MA-1 to be closest to contemporary EUR followed by highly admixed CLM and PUR groups, which is more consistent with expectation based on the location of MA-1 in Siberia (Supplementary Fig. S8).

We next asked whether the American group of 1kGP may contain individuals specifically related to El Miron. Of all 352 MXL, PEL, CLM, and PUR samples, we found one MXL_NA19764 that correlated with El Miron and La Brana-1 but not with the ~14000 year old Villabruna from Italy (898 slow SNPs) and the ~30000 year old Vestonice-16 (Figure 5A). Interestingly, a significant fraction of the samples (0.28 MXL, 0.14 PEL, 0.13 CLM, 0.11 PUR) showed negative correlation with El Miron while only a few samples did so for 26 other ancient European genomes examined (Figure 5A, Supplementary Table S2). Among those MXL samples with positive correlation with El Miron, the average R value in El Miron correlation was similar to that for the Vestonice-16 correlation (~0.15 in both cases), indicating that a generally weak relationship with El Miron cannot explain the negative correlation with MXL. We further found that MXL samples with negative correlation to El Miron were closer to present-day Africans (Fig. 5B) and Kennewick Man (Fig. 5C) but more distant to El Miron and La Brana-1 with El Miron far more distant than La Brann-1 (Fig. 5C). This pattern suggests that El Miron related people, upon landing in North America, had admixed with African like Amerindians who had settled earlier, therefore resulting in offspring with less El Miron and more African genomes.

This negative correlation of El Miron with a significant fraction of Native Americans was unique among the 26 ancient European genomes examined (Supplementary Table S2), suggesting that it was El Miron-related population rather than any other group of Europeans that went into America and hence had their genomes partially replaced in their descendants due to admixture with African like ancestors of Kennewick Man. Amerindians in the North (MXL) had more individuals with negative correlation with El Miron than those in the South (PEL), consistent with where the Clovis culture was located. We further verified the special correlation between El Miron and Anzick-1 by using their distance to AMR samples (MXL and PEL) in correlation studies (Supplementary Fig. S9). These results strongly suggest that El Miron-related people were among the direct ancestors of Anzick-1. Given the location of El Miron and the associated Magdalenian culture (immediately preceded by the Solutrean), the results here are most compatible with the Solutrean model.

mtDNA haplotypes

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Our previous study suggests adaptive evolution of mtDNAs with nuclear genomes in admixed people, with indigenous people in South Pacific sharing more informative (slow evolving) SNPs with Denisovans and Neanderthals [19]. Using the Mitoweb database, we examined slow SNPs in mtDNAs to confirm the above autosomal results. Kennewick Man carried the X2a haplotype that is only found in Amerindians and different from other X haplotypes common in Eurasians. The absence of X2 in Asia has been considered as evidence for the Solutrean hypothesis. We found four X2 associated slow SNPs (1719, 12397, 13966, 14502), among which only 14502 is X2a specific. While 1719 is present in Denisovans, some Australo-Melanesians, and others, 12397 and 13966 are not found in South Pacific. However, the X2a specific 14502 is found in P7 (1/1) and S2 (2/5) that are common in Australo-Melanesians. It is also frequently found in M10a (32/32) and R8b (6/21) common in South Asians, N11a (4/5) common in Chinese, and H42a (10/18) common in the Basques. These results suggests admixture between Solutrean migrants carrying X2 and the first Paleoamericans or ancestors of Kennewick Man who shared ancestry with people in South Pacific and Denisovans/Neanderthals.

We also found other Amerindian specific mtDNA haplotypes to share more slow SNPs with Africans, Europeans, and Neanderthals relative to their sister haplotypes in East Asia. Of five A haplotype specific slow SNPs (663, 1736, 4824, 8027, and 8794), only 8027 is A2 specific and also present in L1c of Africans and some R0 of East Asians while the other 4 are not common in non A types, indicating impact of Africans in the formation of the Amerindian specific A2 haplotype. For the Amerindian B2 type, there are two B2 specific slow SNPs that are commonly found in Europeans with 11177 in U5b (21/592) and 3547 in HV1b (15/30), consistent with effect of admixture with certain European groups. Haplotypes C1, C4c, and D1 are highly specific to Paleoamericans and Native Americans but do not have defining slow SNPs different from those of other C and D types. However, they shared more Neanderthal mtDNA alleles than other C or D subtypes common in East Asians [19]. These results on mtDNAs provide independent evidence confirming the autosomal links between Paleoamericans and Australo-Melanesians, Africans, Europeans, or Neanderthals.

Discussions:

The DNA results here are remarkably consistent with previous models based on morphological and archaeological findings. By linking Kennewick Man with the Neanderthal Mezmaiskaya-1 and contemporary Europeans and Africans, the puzzle of his craniofacial affinity with Australo-Melanesians now has an intellectually satisfying answer. The specific genetic relationship between Anzick-1 and El Miron provides strong evidence for the Solutrean model.

We have previously shown that among present-day people Australo-Melanesians have the most ancestry from Denisovans who were archaic Africans with Eurasian admixtures [19]. The results here indicate close relationship of Kennewick Man with Neanderthals but not Denisovans, therefore largely eliminating the possibility of a direct migration route into America via the Pacific Ocean for Australo-Melanesians in South Pacific. The Y chromosome haplotype C is common in indigenous people in South Pacific, Asia, Siberia, and North America, and belongs to the newly found megahaplogroup ABCDE that is closely related to Neanderthals and Denisovans [19], which

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

is consistent with Native Americans sharing paternal ancestry with Neanderthals and Australo-Melanesians. Therefore, data on all three informative DNAs (autosomes, mtDNA, and Y) consistently suggest that the first South Americans had archaic African or Neanderthal ancestry.

The timing of first entry into America by humans remains to be estimated by future DNA studies. Several human sites in America, including the Topper site in South Carolina United States, the Burnham site in Oklahoma United States, sites in Brazil and Chile, suggest that humans were in the New world as early as 30,000 years ago to perhaps 60,000 [38]. This old age for the first settlement is consistent with the findings here that ancient Amerindians had more ancestry from Southern Chinese (Hunan) relative to Northern Chinese (Fujian), the recent finding of modern humans originating in South East Asia, and the African or Australo-Melanesian like genome in Paleoamericans or Native Americans that could only have come from Neanderthals [19]. Also compatible with results here is the recent discovery of a ~130000 year old human site in North America [39], indicating that admixture of Neanderthals and modern humans might have happened in America, which appears to be the more likely model given the absence of mtDNA C1, C4c and D1 haplotypes in North East Asia and Siberia that shared more informative alleles with Neanderthals than their respective sister haplotypes in North East Asia and Siberia [19]. If Neanderthals were present in North East Asia and Siberia, there seems *a priori* no reason that they could not find their way through the Bering Strait.

We found the ~24000 year old MA-1 of Siberia to be the most correlated with Kennewick Man among European aDNAs in their distance to Amerindians (Supplementary Fig. S3). This suggests that the ancestry of Kennewick Man may include Siberians, which is consistent with intuitive expectations. MA-1 has previously been found to be most related to Amerindians rather than Europeans [15]. But our findings here that MA-1 was most related to contemporary Europeans is more in line with the common sense that local genetic continuity should be the rule rather than the exception. Indeed, morphological studies in general consistently find local continuity in traits and cultures. MA-1 also shows a special relationship with Vestonice-16, both belonging to the same culture Gravettian. Past studies however failed to see that relationship because they have used fast evolving SNPs that turned over fast in response to fast changing environments [36].

Our results suggest that South West Europeans of ~15000-20000 years ago such as El Miron had a special genetic connection with the Clovis people Anzick-1 in North America, more so than East Europeans and Siberians of that age. Consistent with leaving Europe in ancient times, the genomes of these people are largely absent in present-day Europeans, which is unlike all other ancient European genomes examined here. Such results suggest that the part of ancestry in Anzick-1 that is shared with El Miron could only have come from Europeans who had come to North America by way of the Atlantic Ocean [10]. Therefore, people of the Clovis culture appear to be the progeny of Europeans and the first South Americans. This explains the morphological finding that the first North Americans show traits in between modern Australo-Melanesians and Europeans [1-3]. Our results also showed a fraction of present-day Native Americans to share more ancestry with Kennewick Man or El Miron. Some North Native Americans showed replacement of El Miron genomes with Kennewick Man related African genomes, consistent with the first North Americans occupying an unresolved morphological position between modern

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

South Pacific and European populations [4]. These results could not be found when using adaptive fast changing DNA sequences because present-day people living in the same location may share adaptive sequences despite different ancestry.

It must be emphasized that any result must be held as uncertain at best until it has been verified by at least one other independent approach unless that result is a logical deduction from self-evident axioms. Previous molecular results based on unrealistic assumptions have routinely contradicted trait evidence and common sense. In striking contrast, the results from our new molecular method, both here and in a previous work [19], have consistently found unity between molecules and traits. We expect this intellectually satisfying pattern to continue to hold in future studies of other evolutionary puzzles.

Acknowledgements:

Supported by the National Natural Science Foundation of China grant 81171880 and the National Basic Research Program of China grant 2011CB51001.

Author contributions: DY and SH designed the study, performed data analyses, and wrote the manuscript.

Methods and Materials:

Sequence download We downloaded ancient and modern human genome sequences using publically available accession numbers. South Asian and Oceanian SNPs data from Pugach et al (2013) were obtained from the authors [35].

Selection of SNPs The identification of slow and fast evolving proteins and their associated SNPs were as previously described [19].

Calling SNPs from genome sequences We used publically available software SAMTOOLS, GATK, and VCFTOOLS to call SNPs from either downloaded BAM files or BAM files we generated based on downloaded fastq data [40-42].

Imputation We performed imputation to obtain more coverage of the slow SNPs on the Human Genome Diversity Project (HGDP) dataset and the South Asian and Oceanian dataset of Pugach et al (2013) [35]. We used the SHAPEIT2 software to do phasing for the SNP chip data [43] and the IMPUTE2 software to impute based on 1kGP data [44].

Genetic distance calculation We used the custom software, dist**,** to calculate pairwise genetic distance (PGD) or number of SNP mismatches from SNP data as described previously [20].

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

PC analysis We utilized GCTA to analyze data in the PLINK binary PED format to generate two files (*.eigenvec and *.eigenva). We drew PCA plot using *.eigenvec file [45,46]. One sample BEB_HG04131 was found on PC2-PC3 plot to be an outlier and was hence excluded from the PC analysis and most distance calculations presented here.

Correlations between different individuals To test the relatedness of two individuals, we calculated the distance of each individual to the 2534 individuals in 1000 genomes phase 3 data. We then obtained Spearman or Pearson correlation coefficient R of these two individuals in their distance to 1kGP or one of the 5 groups using Graphpad Prism 6 software. For between population correlations, the average R value of all individuals was calculated from all pairwise correlations. Spearman and Pearson analyses largely gave similar results and we have therefore mostly presented Spearman results.

Other methods Other common statistical methods used were Student's t test, chi square test, and Fisher's exact test, 2 tailed.

References:

1 Neves, W. A., Prous, A., Gonzalez-Jose, R., Kipnis, R. & Powell, J. Early Holocene human skeletal remains from Santana do Riacho, Brazil: implications for the settlement of the New World. *J Hum Evol* **45**, 19-42 (2003).

2 Jantz, R. L. & Owsley, D. W. Variation among early North American crania. *Am J Phys Anthropol* **114**, 146-155, doi:10.1002/1096-8644(200102)114:2<146::AID-AJPA1014>3.0.CO;2-E (2001).

3 von Cramon-Taubadel, N., Strauss, A. & Hubbe, M. Evolutionary population history of early Paleoamerican cranial morphology. *Science advances* **3**, e1602289, doi:10.1126/sciadv.1602289 (2017).

4 Steele, D. G. & Powell, J. F. Peopling of the Americas: paleobiological evidence. *Human biology* **64**, 303-336 (1992).

5 Neves, W. A., Hubbe, M. & Correal, G. Human skeletal remains from Sabana de Bogota, Colombia: a case of Paleoamerican morphology late survival in South America? *Am J Phys Anthropol* **133**, 1080-1098, doi:10.1002/ajpa.20637 (2007).

6 Gonzalez-Jose, R. *et al.* Craniometric evidence for Palaeoamerican survival in Baja California. *Nature* **425**, 62-65, doi:10.1038/nature01816 (2003).

7 Waters, M. R. & Stafford, T. W., Jr. Redefining the age of Clovis: implications for the peopling of the Americas. *Science* **315**, 1122-1126, doi:10.1126/science.1137166 (2007).

8 Goebel, T., Waters, M. R. & O'Rourke, D. H. The late Pleistocene dispersal of modern humans in the Americas. *Science* **319**, 1497-1502, doi:10.1126/science.1153569 (2008).

9 Meltzer, D. J. *First Peoples in a New World*. (University of California Press, 2009).

10 Stanford, D. J. & Bradley, B. A. *Across Atlantic Ice: the Origin of America's Clovis Culture*. (University of California Press, 2012).

11 Nielsen, R. *et al.* Tracing the peopling of the world through genomics. *Nature* **541**, 302-310, doi:10.1038/nature21347 (2017).

12 Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455-458, doi:10.1038/nature14625 (2015).

13 Raghavan, M. *et al.* POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884, doi:10.1126/science.aab3884 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

14 Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104-108, doi:10.1038/nature14895 (2015).

15 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91, doi:10.1038/nature12736 (2014).

16 Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225-229, doi:10.1038/nature13025 (2014).

17 Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol* **10**, e1001388, doi:10.1371/journal.pbio.1001388 (2012).

18 Huang, S. New thoughts on an old riddle: What determines genetic diversity within and between species? *Genomics* **108**, 3-10, doi:10.1016/j.ygeno.2016.01.008 (2016).

19 Yuan, D. *et al.* Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA. *bioRxiv*, doi: https://doi.org/10.1101/106864 (2017).

20 Yuan, D. *et al.* Minor alleles of common SNPs quantitatively affect traits/diseases and are under both positive and negative selection. *arXiv:1209.2911* (2012).

21 Yuan, D. *et al.* Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms. *Sci China Life Sci* **57**, 876-888, doi:10.1007/s11427-014-4704-4 (2014).

22 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* **10**, e0133421, doi:10.1371/journal.pone.0133421 (2015).

23 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.*, doi:10.1038/srep16904 (2015).

24 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

25 Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762, doi:10.1038/nature08835 (2010).

26 Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature communications* **3**, 698, doi:10.1038/ncomms1701 (2012).

27 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).

28 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, doi:10.1038/nature13810 (2014).

29 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

30 Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820-822, doi:10.1126/science.aad2879 (2015).

31 Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, doi:10.1126/science.1224344 (2012).

32 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications* **6**, 8912, doi:10.1038/ncomms9912 (2015).

33 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206, doi:10.1038/nature18964 (2016).

34 Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol Biol Evol* **32**, 3132-3142, doi:10.1093/molbev/msv181 (2015).

35 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* **110**, 1803-1808, doi:10.1073/pnas.1211927110 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

36 Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205, doi:10.1038/nature17993 (2016).
37 Sanchez-Quinto, F. *et al.* Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr Biol* **22**, 1494-1499, doi:10.1016/j.cub.2012.06.005 (2012).
38 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, doi:10.1038/nature16152 (2015).
39 Holen, S. R. *et al.* A 130,000-year-old archaeological site in southern California, USA. *Nature* **544**, 479-483 (2017).
40 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
41 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158, doi:10.1093/bioinformatics/btr330 (2011).
42 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
43 Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6, doi:10.1038/nmeth.2307 (2013).
44 Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529, doi:10.1371/journal.pgen.1000529 (2009).
45 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).
46 Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure Legends:

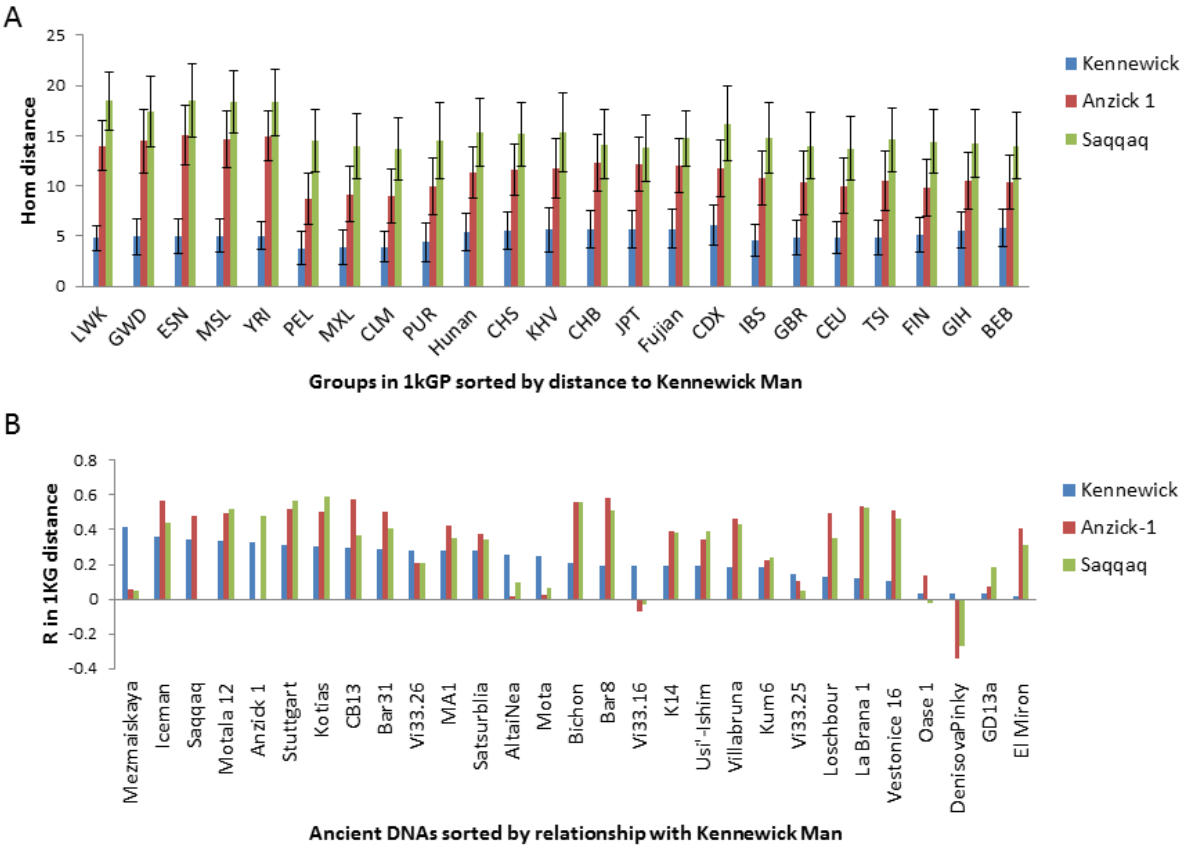

Fig. 1. Relationships between ancient Amerindians and 1kGP samples or ancient humans. **A.** Genetic distance between three ancient Amerindians and 1kGP groups. Average distance and standard deviations are shown. **B.** Correlation between ancient Amerindian DNAs and other ancient DNAs in their distance to all samples in 1kGP.

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

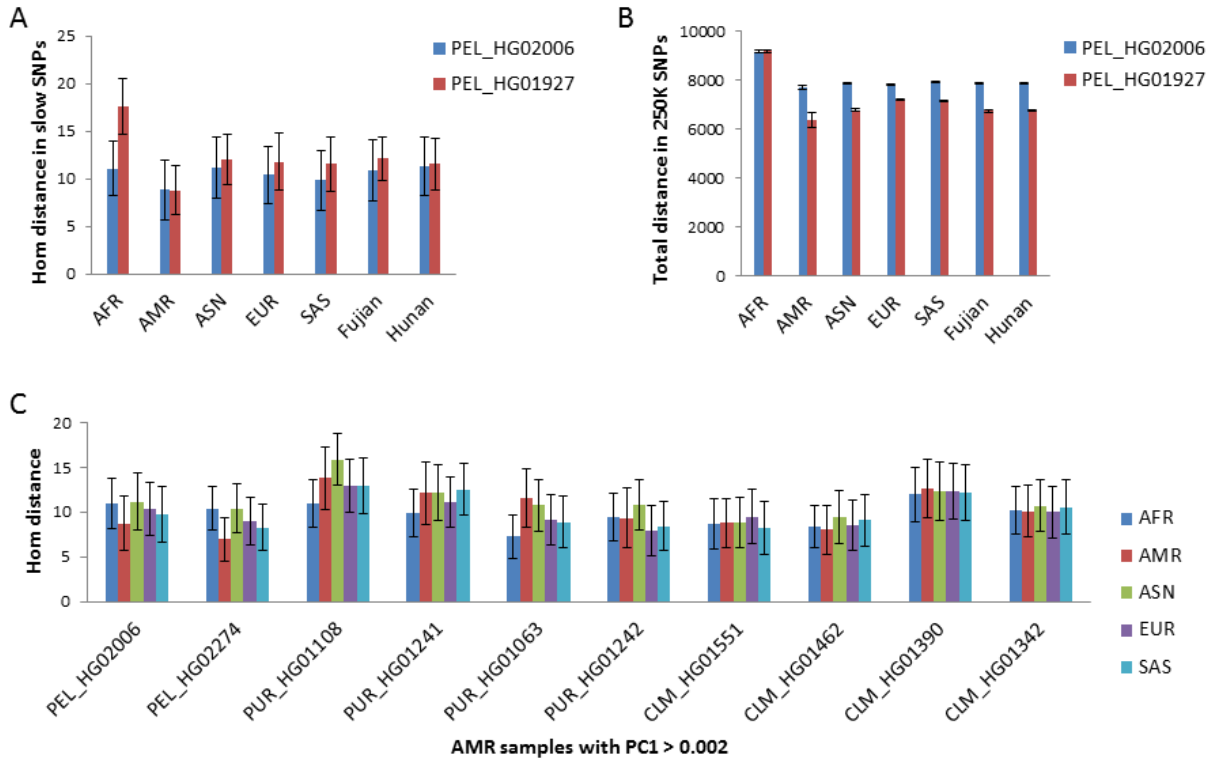

Fig. 2. Amerindians with significant African ancestry. Distance in slow SNPs (A) or fast SNPs (B) to 1kGP groups for two PEL samples of different ancestry. **C.** Distance in slow SNPs to 1kGP groups for AMR samples with significant African ancestry or with PC1 > 0.002 in a PCA type analysis. Average distance and standard deviations are shown.

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

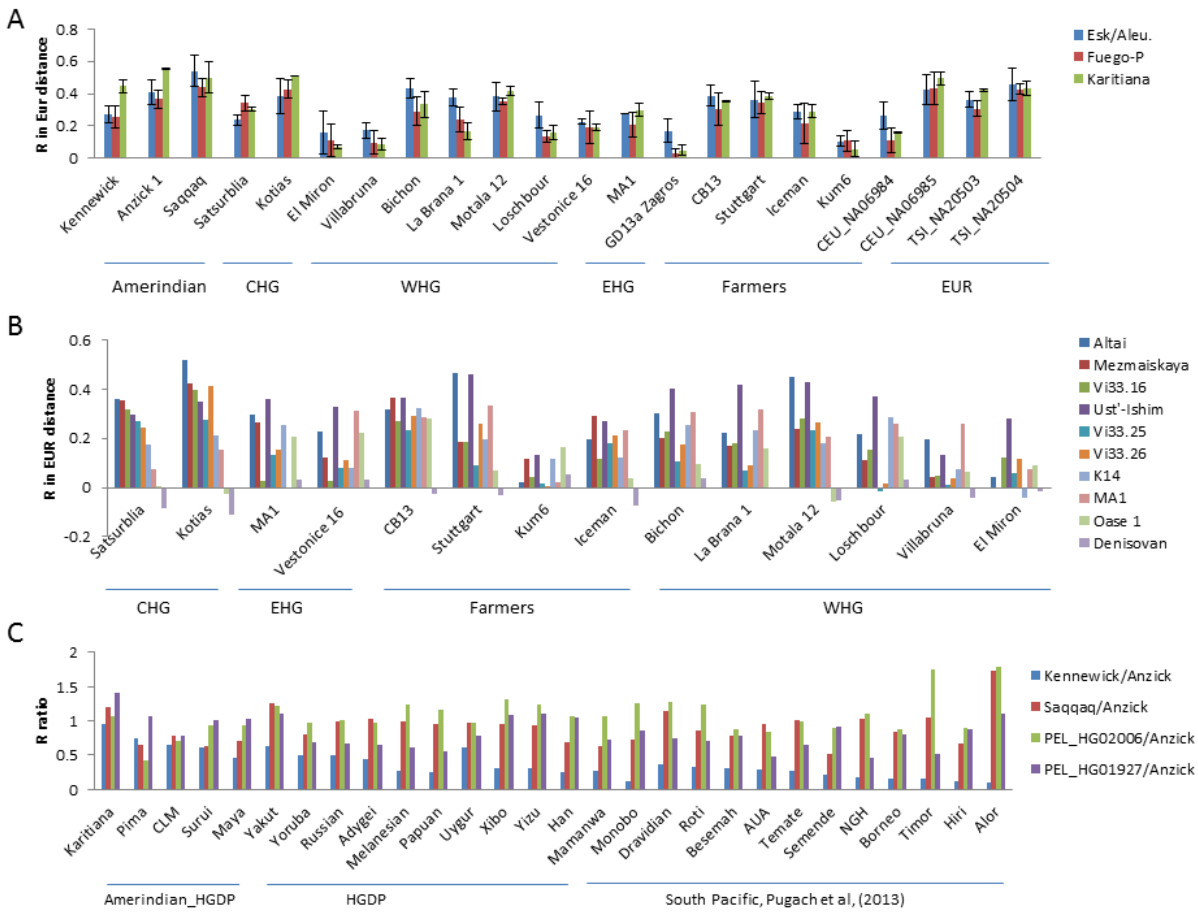

Fig. 3. Relationship with Amerindians and indigenous people in South Pacific for ancient and living genomes. A. Correlation with North and South Amerindians for ancient and living DNAs. Average R score of Spearman correlation in distance to EUR samples and standard deviations are shown. **B**. Correlation with Neanderthal and selected ancient DNAs for selected European ancient DNAs. **C.** Correlation with selected groups in HGDP and groups in South Pacific as reported by Pugach et al, (2013) [35].

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

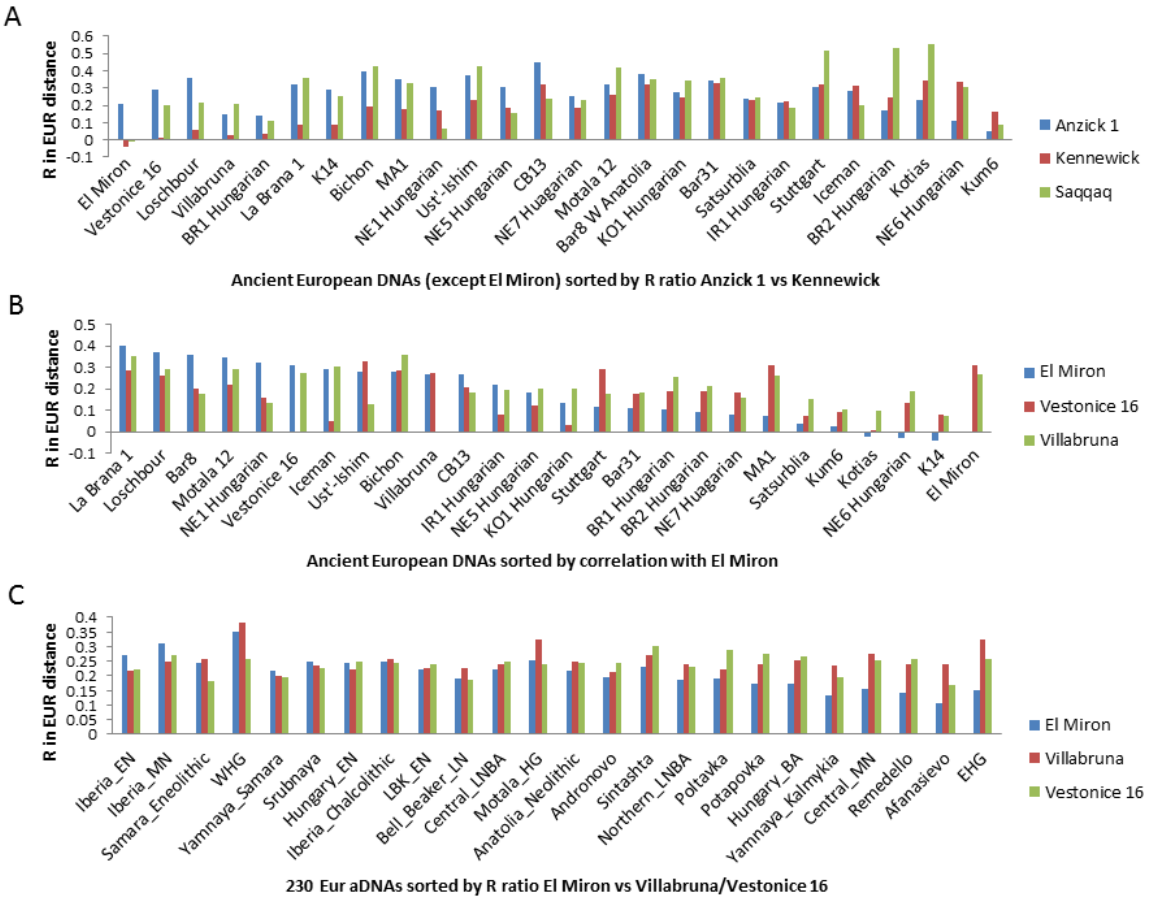

Fig. 4. Correlation of El Miron with ancient Amerindians and Europeans in distance to EUR samples in 1kGP. A. Correlation between ancient Europeans and Amerindians. **B.** Correlation of El Miron with ancient Europeans. **C.** Correlation of El Miron with 230 ancient European DNAs from Mathieson et al (2015) [38].

bioRxiv preprint doi: https://doi.org/10.1101/130989; this version posted May 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

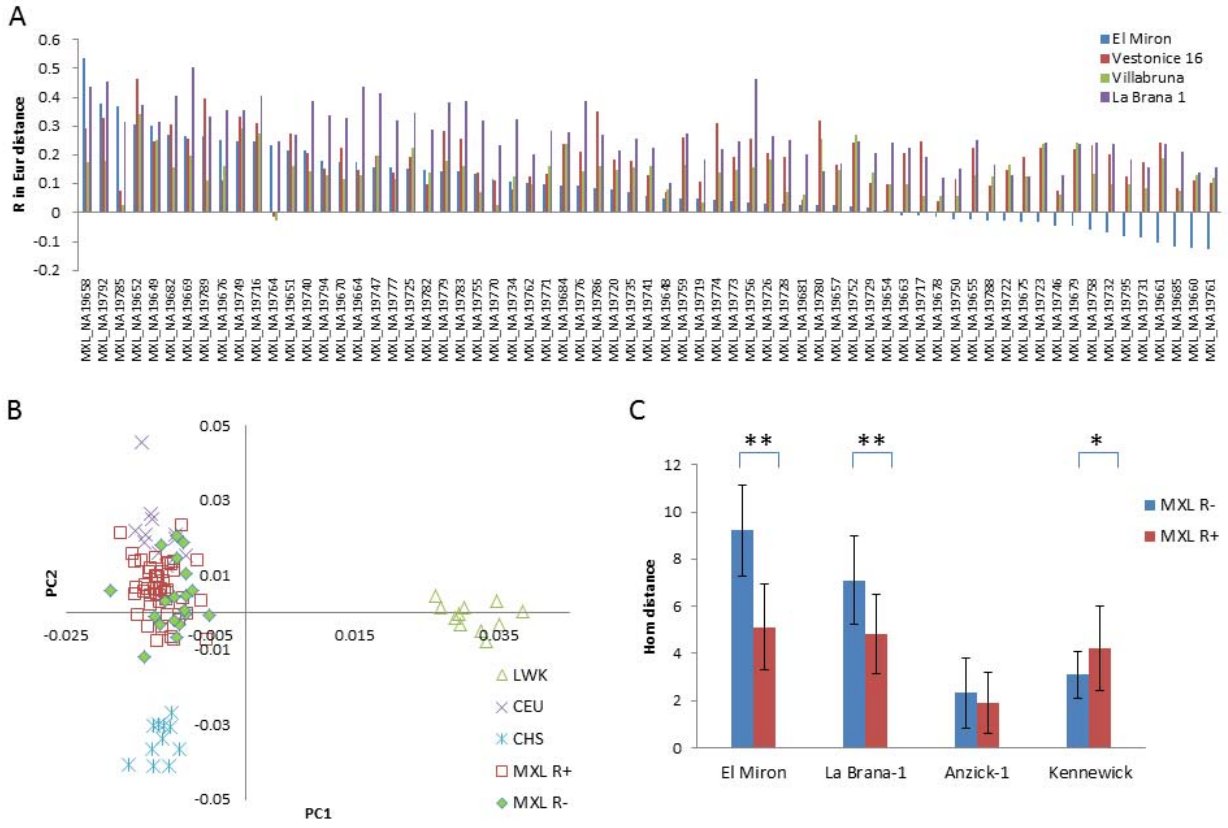

Fig. 5. Correlation of El Miron with MXL samples in 1kGP. A. Correlation of ancient DNAs with MXL individuals in distance to EUR samples. **B**. PCA plots of MXL samples with positive (MXL R+) or negative (MXL R-) correlation with El Miron. **C.** Genetic distance between ancient DNAs and MXL R+ or MXL R- samples. Standard deviations are shown, **, $P < 0.01$; *, $P < 0.05$.