Science
AAAS

Supplementary Materials for

Early Neolithic genomes from the eastern Fertile Crescent

Farnaz Broushaki, Mark G Thomas, Vivian Link, Saioa López, Lucy van Dorp, Karola Kirsanow, Zuzana Hofmanová, Yoan Diekmann, Lara M Cassidy, David Díez-del-Molino, Athanasios Kousathanas, Christian Sell, Harry K. Robson, Rui Martiniano, Jens Blöcher, Amelie Scheu, Susanne Kreutzer, Ruth Bollongino, Dean Bobo, Hossein Davudi, Olivia Munoz, Mathias Currat, Kamyar Abdi, Fereidoun Biglari, Oliver E. Craig, Daniel G Bradley, Stephen Shennan, Krishna Veeramah, Marjan Mashkour, Daniel Wegmann, Garrett Hellenthal, Joachim Burger

Correspondence to: Joachim Burger jburger@uni-mainz.de; Garrett Hellenthal ghellenthal@gmail.com; Daniel Wegmann daniel.wegmann@unifr.ch

This PDF file includes:

Supplementary Text
Figs. S1 to S52
Tables S1 to S37
Caption for Animation S1

Other Supplementary Materials for this manuscript includes the following:

Animation S1 (3D animation of PCA)

Supplementary Text

S1. Archaeological context

Marjan Mashkour, Olivia Munoz, Fereidoun Biglari, Farnaz Broushaki, Kamyar Abdi, Hossein Davudi & Stephen Shennan

Archaeological background of the Neolithic in Zagros and Iran

It has long been established that agriculture based on cereal crops and domestic animals began at the end of the last Ice Age in the Fertile Crescent, which stretches north from the southern Levant through eastern Anatolia and northern Mesopotamia then east into the Zagros Mountains on the border of Iran and Iraq. In recent years it has been established that these developments were widely dispersed, and that there was no single center (*20*), but the role of the Zagros has remained unclear. While it was at the forefront of research on the origins of agriculture during the 1960s and 70s, since the 1980s little work has been done.

The discovery of the Iranian Neolithic owes much to the impetus of Braidwood's investigations during the 1950s to 1960s in Iranian Kurdestan where he conducted excavations at Asiab and Sarab (*22*) (Fig. S1). Braidwood's pioneering research continued with the excavation of other highland sites such as Guran, Ganj Dareh and Tepe Abdul Hosein (*23-26*), which brought the earliest evidence for the emergence of farming communities in the central Zagros region, where agriculture was possible without irrigation on the eastern fringes of the Fertile Crescent (*25-29*). In parallel, Ali Kosh (*30*), Chogha Sefid (*31*) and Chogha Bonut (*32*), lowland sites in the foothills of the Zagros, were investigated to understand human interaction during the Neolithic in the Zagros. After the first wave of archaeological investigations, more recent work at Sheikh-e Abad and Jani (*2*) has produced the earliest evidence of Neolithic development in the Zagros. Sheikh-e Abad, where the earliest phase dates to c. 10,100 BCE (*2*), shows repeated occupation by mobile groups exploiting wild caprines, largely goats, with evidence of animal dung at Sheikh-e Abad and Jani appearing c. 8,200-7,800 BCE and probable animal pens c. 7,600 BCE. These developments are in keeping with evidence for goat management at the Zagros site of Ganj Dareh c. 7,900 BCE, and also for similar developments

further west in SE Anatolia (*33*). Further south, Chogha Golan and East Chia Sabz can be considered the earliest sites in the region where the processes of plant management and domestication began during the local PPNA in the 10th millennium BCE (*20*). At Chogha Golan archaeobotanical research (*1*) has provided evidence of pre-domestication cultivation of the wild ancestor of barley, including the presence of weeds of cultivation, at the end of the Younger Dryas, from c. 9,800 BCE, and thus contemporary with the earliest such sites further west as shown in Riehl *et al.*: Fig. 2 (*1*), while morphologically domesticated emmer wheat appears in significant quantities from c. 7,800 BCE. New investigations carried out in the Western and Eastern fringes of the Zagros, such as at the pre-pottery sites of Kelek-e Asad Morad (*34*) and Qaleh Rostam (*35*), show the continuing presence of Neolithic communities until the 6th millennium BCE with pre-pottery and pottery levels.

Archaeological investigations over the last decade clearly show the spread of the Neolithic towards the east through two distinct waves in the south and north of the central Iranian Plateau occupied by the Kavir and Lut deserts. The region of Fars in the southern Zagros and the region of Semnan in the northeast of Iran are promising areas for documenting this eastward spread of the Neolithic. The earliest evidence of pre-pottery settlement in the southern Zagros is found in Rahmat Abad and other sites in the vicinity. In the northeast of the Iranian plateau, the site of Tappeh Sang-e Chakhmaq (western mound) in the Shahroud area *(36, 37)* is the key site for evidencing the earliest pre-pottery occupation and the link to the Neolithization process in southern central Asia. The apparent later arrival of the Neolithic culture to the north and southeast of the Central Plateau and in the northwest of Iran may be a consequence of less dense archaeological investigations, although the antiquity of the central Zagros is hardly expected in these areas (*34, 38*).

The archaeological context of the analyzed sites

Three archaeological sites were chosen for this study because of their significance to understanding the origins of the Neolithic communities of the Iranian Plateau and their relation to Anatolia, and because the quality and preservation status of samples from these sites allowed us to extract the DNA and perform further examinations. Table S1 summarizes the main characteristics of our five presented individuals from the three sites.

Tepe Abdul Hosein

Tepe Abdul Hosein is an early Neolithic mound ~50 m in diameter located at an altitude of 1860 m in Luristan, i.e. in the central Zagros (*23, 39*), with features characteristic of the aceramic Neolithic in the region. The earliest occupation is represented by pits dug into the natural soil succeeded by shallow deposits with occupation surfaces and a gradual transition to well constructed floors and then substantial mud-brick architecture (*23*). Radiocarbon dates from the initial excavation, albeit with very large standard deviations, point to occupation in the late 9th and early 8th millennium BCE, confirmed by the dates on the skeleton in this study. Plant remains recovered were present from the earliest phase of occupation and include pistachios, almonds, and small quantities of barley (*Hordeum distichon*), emmer wheat (*Triticum dicoccum*) and lentils (*Lens culinare*), whose domestication status is unclear (*23, 40*).

Neolithic burials from Tepe Abdul Hosein

Nine burials were identified during the 1978 excavation campaign at Tepe Abdul Hosein (*23*), of which seven have been attributed to the Neolithic period, the remaining ones belonging to a phase of reuse occurring during the Islamic period. Among the individuals recovered from the Neolithic levels those sampled for DNA analyses (AH1, AH2, and AH4) were found in three different graves, respectively 13030, 19001, and 10035 located in the squares 12H, 11G and 20L (Fig. S2). AH1 yielded no collagen and failed to be dated. The ^{14}C radiocarbon datings for AH2 and AH4 are 8,205-7,756 cal. BCE and 8,204-7,755 cal. BCE, respectively (Table S2). All radiocarbon ages are calibrated using OxCal 4.2.

1. Burial n° 13030, skeleton #1 sample AH1

Burial context

Burial 13030 was recovered in square 12H, in the southeast quadrant of the mound, close to burial 13029. According to (*23*): 61, the skeletons recovered in this square represent "a family of three [that] was killed when burning debris fell on top of them". This author describes the burial 13030 as "a female skeleton sitting cross-legged on the floor with her head bent over an infant she held in her arms" (Fig. S3).

Anthropological analysis

This burial corresponds to an adult skeleton (skeleton #1), whose pelvic bone morphology indicates a female (*41*), and a perinatal individual (skeleton #2), whose age can be estimated between 38-40 weeks of gestation according to the long bone length (*42*). According to the length of the radius, the stature of the adult woman can be estimated between 1.54 and 1.63 m (*43*). Degenerative joint diseases were observed on the distal extremity of both first metacarpals, on the left ulna (proximal) and scapula (glenoïd cavity), while the other observed bones do not present such pathologies. This individual presented severe dental pathologies, including calculus deposits, severe tooth wear, carious lesions, periapical lesions, alveolar resorption and multiple ante-mortem tooth losses. Notably, all the teeth from the lower jaw have been lost ante-mortem. This indicates not only low oral hygiene, but also suggests that carbohydrates could have played an important role in the diet of this individual. Finally, an interproximal groove is visible on the left upper first premolar; it could result from the repeated use of a tooth pick but could also be related to the use of teeth to process fibers or tendons, as evidenced in prehistoric Pakistan (44, 45). Although the adult cranium is fragmented and incomplete, one may observe a deformation at the parietal-occipital junction, with a flattening of the left parietal that could be related to intentional shaping of the head, as observed on other individuals from Tepe Abdul Hosein (39).

2. Burial n° 19001, skeleton #1, sample AH2

Burial context

Burial 19001 was found in square 20L, at 20 cm below the summit of the mound, in its western margin (Fig. S2). According to a photograph from this square (Fig. S4), the burial contained the skeleton of an adult individual (skeleton #2) disposed in a squatting position with the right upper limb lying under the lower limbs, and left upper limb slightly flexed with the hand close to the left ankle. A long bone from a second adult individual (skeleton #3) is visible on the photograph, lying close to the left forearm of skeleton #2, but is not mentioned in the

description. A third individual (skeleton #1) is represented in the collection and was identified during the study, but its position and relation with Skeletons #2 and #3 is not mentioned by Pullar (*23-24*).

Anthropological analysis

a. Skeleton 19001 #1 belongs to a child aged 6.5-8 years old, according to teeth development and long bone length (*42, 46*).The observable deciduous upper and lower molars are severely worn suggesting frequent consumption of food with an abrasive consistency. Moreover, one carious lesion is recorded on the buccal face of the first left lower permanent molar, suggesting, as for the woman from burial 13030, a consumption of carbohydrates.

b. Skeleton 19001 #2 belongs to an adult individual. The whole skeleton is rather well represented although many bones are fragmented and incomplete. According to the morphology and metrics of the pelvic bones (*41, 47*) and to the morphological traits of the cranium, the individual was a female. Her stature is estimated between 1.57 and 1.65 m according to the length of the right femur (*43*). Several degenerative joint diseases were observed (vertebras; coxal bone: auricular surface and acetabulum; upper limbs; left and right carpal and metacarpal bones). Periosteal reactions were noticed on the diaphysis of the right humerus. The skull is incomplete and maxillary teeth are not represented. The mandibular teeth present several pathologies including severe alveolar resorption, calculus deposits, severe attrition, carious lesions, periapical lesions, and antemortem tooth losses (at least central incisors and 2nd right molar). The oblique wear observed on the right 1st molar and canine, associated with extreme wear of premolars, suggests a use of right teeth for extra-alimentary tasks, perhaps related to leather processing. The cranium presents a circumferential deformation which has been identified as an intentional shaping of the head (39) and it is now exposed, with the lower jaw, in the National Museum of Iran.

c. Skeleton 19001 #3 belongs to an adult individual, whose bones present a different patina (whiter in color), is represented by few metacarpal and phalanges, fragments of clavicle, scapula and lower limbs. The bones preserved do not allow a reliable sex assessment.

3. Burial n° 10035, sample AH4

Burial context

Skeleton 10035 was recovered in square 11G, in the southeast quadrant of the mound (Fig. S2). It was located in the northwest corner of the trench, the burial pit cutting the wall U (*23*). According to the drawings made during the excavation (Fig. S5), the burial has undergone post-depositional anthropogenic disturbance leading to the displacement and removal of several bones (left upper limb and bones from the right lower limb).

The individual was apparently lying on its back, with the head towards the north. The right upper limb is extended along the thoracic grid with the hand lying on the pelvis; the absence of the left limb does not allow reconstruction of its position. The left lower limb is extended toward the north, while the left upper limb may have been flexed, according to the position of the right fibula and ankle bones.

Anthropological analysis

The skeleton belongs to an adult individual. Its preservation is poor, as many bones are missing or are fragmented. This prevented any reliable sex determination. The long bones of upper limb are rather gracile. No degenerative joint diseases have been observed on the bones present. The individual displays several oral pathologies, including carious lesions, periapical lesions, ante-mortem tooth losses. Tooth wear is moderate to severe, with premolars and molars more affected than anterior teeth, both on the upper and lower arcades. The cranium presents a circumferential deformation probably resulting from an intentional shaping of the head (*39*).

Wezmeh Cave

Wezmeh Cave is located in the Kermanshah region, 12 km southwest of the town of Islamabad-e Gharb and 3.5 km northeast of the village of Tajar-e Akbar, at an elevation of 1.430 m above sea level (Fig. S6). The cave was discovered in 1999 during an archeological survey and measures approximately 45 m^2 with a depth of 27 m (*48*). In 2001, following disturbance by illicit digging, a short field season was undertaken, during which a 3x3 m trench was excavated to the bedrock on the terrace immediately outside of the cave entrance, and six test pits at 2 m intervals were dug to bedrock within the cave. The surface collecting on the exterior slope and the excavations on the terrace and within the cave yielded an abundant faunal assemblage, particularly rich in carnivore remains and included several human bones. The osteological remains were collected in Late Pleistocene and Holocene deposits (*49*). It was concluded that the cave was a den occupied alternately by hyenas and bears.

Surprisingly, several fragmented human bones and teeth were discovered among the faunal remains, including one premolar whose size places it at the upper limits of Late Pleistocene human ranges of variation (*50*). Because of the exceptional nature of the assemblage, its size, taxonomic diversity, and especially the presence of extinct species and human bones, a program for absolute dating of targeted species was initiated. The animal teeth were aged using Uranium-series analyses by alpha spectrometry, and the human premolar was analyzed by non-destructive gamma spectrometry. The earliest and the latest dates obtained on animal bones are 70 kya and 11 kya, respectively. For the current study, a human metatarsal bone fragment has been analyzed and dated to 7,455-7,082 cal. BCE (Table S2). The ^{14}C date for the analyzed male individual shows a more recent date than the latest dates obtained from the animal assemblage. There had not been any prior indication of the presence of an early Neolithic settlement, although a Chalcolithic occupation by pastoral communities had been demonstrated by the presence of Early and Middle Chalcolithic sherds and other material found mixed with faunal and human remains. The analyzed bone is derived from an individual who may had been a prey of carnivores and brought to the cave, or belonged to a person who died in the cave when it was no longer used as a den; caves have traditionally been used by pastoral nomads as shelter for themselves and their herds. Although it is not possible to say if this individual was a forager or a herder, it is certain that Wezmeh cave was surrounded by early pastoral communities, such as Jani located about 15 km to the southeast of the cave and Ganj Dareh at about 80 km to the east- northeast of Wezmeh (*2, 26*).

Tepe Hasanlu

The Early Iron Age (1,250-800 BCE) in Iran belongs to a cultural interval representing extensive and complex long-distance trade *(51)* linking Iran to Central Asia in the East and to Mesopotamia and Eastern Anatolia in the West, even more than in the Late Bronze Age. Among the key Iron Age sites of Northwestern Iran, Tepe Hasanlu is of high importance due to its long-term occupation and well-defined stratigraphy.

Hasanlu is an imposing fortified citadel mound in western Azerbaijan, situated on the southern shore of Lake Urmia in the Solduz Valley. Tepe Hasanlu has a long history of excavations (*52, 53*) and was continuously occupied during almost ten different cultural periods from the late Neolithic during the second half of the 6th millennium BCE to the Ilkhanid period 13th century CE (*51, 54*). However, the Late Bronze Age (1,450-1,250 BCE), and Iron Age (1,250-550 BCE) were most extensively investigated and large collections of burials were uncovered for the two periods, particularly for Iron Age I and II (*55, 56*). Its pivotal position along trade routes through the Zagros Mountains made Hasanlu a crossroads between the northern Zagros and Mesopotamia and Anatolia. The significant role of Hasanlu as an important urban center and its socioeconomic role in the northwest of the Iranian Plateau during the Iron Age with other regions of the Near East is evidenced by the diversity and richness of the material culture of the site (*57-60*). The city of Hasanlu was destroyed around 800 BCE (*56, 61*).

For this study we analyzed a femur fragment from a male individual from a burial in Square / Operation F38. No specific anthropological work has been carried out in the past on this skeleton. The direct ^{14}C radiocarbon date (cal. 971-832 BCE) (Table S2) confirms the allocation of the burial to the Iron Age II or Hasanlu IVb period.

S2. Stable isotope analysis

Harry K. Robson & Oliver E. Craig

In total, we selected three humans and eight faunal samples (Fig. S7 and Table S3) for carbon and nitrogen stable isotope analysis. We extracted collagen from each bone sample and analyzed it with Elemental Analysis Isotope Ratio Mass Spectrometry (EA-IRMS).

We coarsely ground and demineralized (0.6 M HCl, 4°C) the bone samples (0.2-0.5 g). We rinsed the acid insoluble fractions three times with distilled water and gelatinized them (pH3 [0.001M] HCl, 80°C, 48 h). We ultrafiltered the supernatants containing the gelatin (30 kDa, Amicon® Ultra-4 Centrifugal Filter Units, Millipore, Billerica, MA, USA), froze them, and lyophilized them. We placed the collagen samples (0.8-1.2 mg) in tin capsules and analyzed them in duplicates or triplicates by EA-IRMS using a Sercon GSL analyzer coupled to a Sercon 20-22 Mass Spectrometer (Sercon, Crewe, UK) at the University of York. The results from the analysis are reported per mil (‰) relative to an international standard (V-PDB for $\delta^{13}C$ and AIR for $\delta^{15}N$). For all these instruments the analytical error was <0.2‰ (1σ), which we calculated from the repeated measurements of each sample and measurements of a bovine control from multiple extracts. The accuracy was <0.3‰ based on analysis of international standards (IAEA-600, IAEA -N2). Out of 11 samples prepared for analysis only four yielded collagen. These had %C, %N and C:N atomic ratios that were within the acceptable range (Table S3)(*62*).

S3. Sample preparation and sequencing

Farnaz Broushaki, Amelie Scheu & Susanne Kreutzer

Ancient DNA work on 26 prehistoric skeletal samples from the Zagros region of Iran was carried out in dedicated facilities of the Palaeogenetics Group at Johannes Gutenberg-University Mainz. Decontamination and sample preparation (drilling and milling) was performed as in (*4, 21*). For contamination monitoring, blank controls during milling *(21)* and during all further steps were processed as described below.

DNA extraction

Multiple DNA extractions per sample were carried out according to (*4, 21*). EDTA (10 - 14ml, 0.5M, pH 8; Ambion/Applied Biosystems, Life technologies, Darmstadt, Germany), N-lauroylsarcosine (250µl, 0.5%; Merck Millipore, Darmstadt, Germany) and proteinase K (30µl, 18U/µl; Roche, Mannheim, Germany) were used to lysate the bone powder. The volume of EDTA was modulated according to the quantity (300-500 mg) and density of bone.

The samples were then incubated on rocking shakers at 37° C until the bone powder was completely dissolved. The lysis was followed by a phenol/chloroform/isoamyl alcohol (25/24/1, Carl Roth, Germany) based extraction, washing and concentration to 200-250µl of extract using Amicon Ultra15 50kDa filters (Merck Millipore, Darmstadt, Germany).

Library Preparation

Next generation sequencing libraries were prepared according to (*4, 63*). Libraries were amplified with two indexing primers in three parallel PCRs using AmpliTaq Gold® DNA polymerase (Applied Biosystems) and varying cycle numbers (12-18). Library purification and quantification was performed according to (*4*).

MiSeq screening and quantitative real-time PCR

At least one library per sample and per extraction was screened for preservation of ancient DNA by shallow shotgun sequencing on an Illumina MiSeq sequencing machine (50bp, single end) at StarSEQ GmbH (Mainz, Germany) to 450k to 1.2 million reads. The endogenous DNA content was estimated according to (*4*). Molecule numbers were estimated by quantitative real-time PCR (qPCR) after Fill-In reaction during library preparation as described in (*4*).

Out of the 26 samples screened, four early Neolithic samples from Central Zagros (WC1, AH1, AH2, AH4) and one Iron Age sample from Northern Zagros (F38) showed an endogenous DNA content higher than a threshold of 10% or molecule numbers of at least 10^7 per µl of Fill-In product and were selected for deeper genome-wide sequencing on an Illumina HiSeq 2500 sequencing machine. Table S4, S5, and Figure S8 summarize qPCR and screening results.

In order to account for the low quantity of molecules in some libraries and to enlarge the complexity of the sequencing pool, additional libraries were produced and pooled for subsequent high coverage sequencing (see Table S7 for the pooling strategy).

Post mortem damage patterns and blank controls

Post mortem damage patterns were established for all screened libraries using MapDamage 2.0 (*69*). The C/T deamination rates at the 5'-end of reads are typical for ancient DNA. They range from 0.20 to 0.37 when averaged over the first base at the 5'-end of each read.

Molecule numbers of blank controls as estimated by qPCR range from 2.38×10^4 to 1.60×10^5. Values are 10^2-10^4 times lower than those of the corresponding prehistoric samples (Table S5, Fig. S8). MiSeq reads generated from the blank controls were used to calculate a contamination fraction C for each corresponding screened library of a sample according to the following formula (modified after (*70*)):

C = (MAP.B / (TR x Ng.B)) / (MAP.S / (TR x Ng.S)),
where
MAP.B = mapped reads of the blank controls
TR = total reads of the MiSeq run
Ng.B = ng of blank control in the MiSeq
MAP.S = mapped reads of the sample
Ng.S = ng of corresponding sample in the MiSeq.

The contamination fraction of the blank controls is on average 0.09 % (Table S6).

HiSeq sequencing

Deeper next generation shotgun sequencing was carried out on an Illumina HiSeq 2500 platform at the Institute for Molecular Genetics, Johannes Gutenberg-University Mainz, Germany. The sequencing and pooling strategy is summarized in Table S7. For WC1, two different pools were sequenced, each consisting of 8 and 10 libraries on 6 and 2 lanes, respectively (100bp, paired end). For AH1, AH2 and AH4, 5, 12, and 9 libraries, respectively, were combined to a single pool, which was sequenced on 8 lanes (two different runs, each run with 4 lanes; 100bp, single end). For F38, one pool consisting of 7 libraries on 2 lanes (100bp, paired end) was sequenced.

S4. Read processing and post mortem damage

Christian Sell, Jens Blöcher & Farnaz Broushaki

For demultiplexing, 0 to 2 mismatches were allowed per index read. 0 mismatches per index were only allowed in the case of samples sequenced in parallel on the same sequencing lane. Raw read pairs were trimmed and quality filtered and overlapping pairs were merged. For the merging of read pairs the ea-utils package was used (*71*). Only merged pairs were mapped against the human reference build GRCh37/hg19 with Burrows Wheeler Aligner aln (*72*) using default parameters. After the removal of duplicates the mapped reads were re-aligned using the GATK-Toolkit (*73*). Different read groups were assigned for each library and sequencing run. A summary of the results is shown in supplementary Table S8. MapDamage 2.0 (*69*) was used to visualize the average deamination patterns over all mapped reads per sample (see Fig. S9, Table S8). For a comprehensive description of the pipeline see (*4*).

S5. Analysis of uniparental markers

Farnaz Broushaki, Rui Martiniano, Susanne Kreutzer & Christian Sell

<u>Mitochondrial haplogroups</u>

Mitochondrial haplogroups (mtDNA Hgs) were determined using Haplofind (*74*). The average coverage of the mitochondrial genomes ranges from 40 to 547x (Table S9). The four Neolithic individuals from Zagros represent mitochondrial haplogroups of the R clade and R2'JT subclade. WC1 is part of the J1d6 haplogroup. The highest frequency of J1d today is found in the Near East and the Caucasus but also in Mediterranean Europe (*75*). The J1d6 subclade is mainly found in Russia and Iran (76-78). Two samples from Tepe Abdul Hosein (AH1 and AH2) are part of the R2 branch, which is concentrated in southern Pakistan and India. However, it is also found in the Near East, the Caucasus, and the Arabian Peninsula (*79, 80*). One Siberian Pleistocene individual (*81*) is also assigned to the R lineage, although it falls at the root of the R haplogroup. The T2c haplogroup identified in AH4 is mostly found in the Near East and Mediterranean Europe. It is probably the first subclade that diverged from T2 with Near Eastern origins (*78*).

The Iron Age sample (F38) from Zagros is part of the N1a3a haplogroup of the N1a clade, which is observed at low frequency among modern Iranians. N1a is found mostly in the Arabian Peninsula and Northeast Africa, but also in Central Asia and Southern Siberia at lower frequencies. However, a few ancient samples such as the NE7 individual from the Neolithic period in Hungary (*70*) also represent an mtDNA HG derived from N1a. The N1a subclade appears in Europe together with the establishment of early farmer communities and has a relatively high frequency in early European farmers and also in Neolithic NW-Anatolia (*3, 17, 83, 84*).

<u>Y-chromosomal lineage determination</u>

We determined Y-chromosome haplogroups in ancient male samples using *clean_tree* (*85*). This software implements SAMtools (*86*) mpileup to call alleles at given genomic coordinates (2,710 SNPs), which we obtained from ISOGG 2013 (International Society of Genetic Geneology; "http://www.isogg.org/"). The number of SNPs identified in our samples are shown in Supplementary Table S10. It was not possible to determine the paternal lineage of AH2 due to insufficient number of reads. Derived alleles for F38 and WC1 can be found in Supplementary Table S11 and S12, respectively.

WC1 belongs to haplogroup G2b. G2 Y-chromosomal haplogroups are more frequent in Turkey, southern Caucasus and Iran (*87, 88*). Furthermore, some populations from Pakistan and Southern Asia such as Kalash, Brahui and Pashtun also have the G haplogroup at a higher frequency than other populations (*89, 90*). The very unique G2b subclade is rarely represented around the world. G2b is mostly found in Pashtuns (*91*) and among Europeans mainly in Ashkenazi Jews (*92*). However, it is also found at very low numbers in some other regions in the Near East (*87, 90, 93*). G2b lineages have not yet been identified in ancient samples. The dominating G-derived lineage in the European Neolithic so far is G2a (*3, 4*).

F38 belongs to sub-haplogroup R1b1a2a2-CTS1078/Z2103. This lineage can be included in the L23(xM412) clade, which is characterized by frequencies higher than 10% in the Caucasus, Turkey, Southeastern Europe, and Circum-Uralic populations, and is mostly found at very low frequencies in Western Europe. This pattern sharply contrasts with the distribution of M412-derived Y-chromosomes, which are very common in Western Europe but rare in the

East (*94*). In ancient DNA studies, derived alleles that define the R1b1a2a2 lineage were found in five Yamnaya individuals and in two samples from the Poltkava culture (*3, 17*). Although we observed two reads covering marker J2a1h1-M158, this is a C→T mutation and in addition, there are no upstream markers (ancestral alleles at J-M304, J2-M172 and J2a-M410) that would support including F38 in this lineage. In modern-day Iranian population, the most frequent Y-chromosome haplogroups are J2-M172 (22.5%) and J1-M267 (~10%), and the less frequent ones are R1a-M198 (0-25%) and R1b-M269 (*95*).

We were only able to identify the paternal lineage of two individuals (one Neolithic and one Iron Age sample) and therefore our findings should be interpreted with caution, as these might not be representative of the Y-chromosome haplogroups carried by populations of prehistoric Iran.

Mitochondrial and X-chromosome contamination estimates

We processed mitochondrial and X-chromosome reads and called variants as described in (*4*). In order to estimate contamination, X-chromosomal reads were extracted from male individuals and mitochondrial reads from all individuals and further handled as below. A summary of results is shown in Table S9.

We used the likelihood approach described in Fu *et al.* 2013 (*96*) in order to estimate the mitochondrial contamination rate in the prehistoric samples. The mitochondrial genomes show a range of $0.006 - 0.06\%$ of contaminating reads. In order to determine possible X-chromosome contamination in male samples, we used ANGSD (*97, 98*) as illustrated in (*4*). Both mitochondrial (0.006%-0.06%) and X-chromosomal (0.79%-1.14%) contamination estimates are low in all five samples (Table S9).

S6. Genotype calling and inference of local heterozygosity

Vivian Link, Athanasios Kousathanas, Christian Sell, Lara M. Cassidy & Daniel Wegmann

Methods

We produced both allele presence as well as diploid calls for all samples sequenced in this project following the genotyping pipeline we recently developed specifically for ancient DNA (*4*). This pipeline consist of the following steps performed on the mapped reads resulting from S4:

1. Splitting single end sequences by length
Due to the generally short fragments, most of the single-end sequences were shorter than the number of sequencing cycles performed. However, the length of some of the reads matched the number of cycles used. These two groups of reads are expected to have very different post mortem damage patterns at the 5' end in that only the short reads that reach the end of the DNA fragment show strongly elevated G->A transitions. We thus partitioned all single-end reads into two read groups by length (group 1: shorter than number of cycles; group 2: length matches number of cycles).

2. Merging read groups with too little reads
When we split the read groups of AH1, AH2 and AH4 by length, the group 2 read groups did not contain enough reads for our Base Quality Score Recalibration step (< 6 million reads). Thus, we merged all group 2 read groups. The same problem arose for the Kostenki genome, where we merged the read groups associated to the same library but different lanes and sequencing runs.

3. Estimating post mortem damage (PMD) patterns
We then inferred empirical post mortem damage patterns as described in (*9*) from the tabulated mismatches between the raw reads and the reference genome. These patterns were inferred independently for all read groups.

4. Base quality score recalibration
Next, we recalibrated base qualities using an approach similar in spirit to BQSR but that also takes post mortem damage patterns into account (*4*). When applying this approach, we used the raw quality scores, distance from the 3' and 5' end as well as the local read context as covariates. Following standard recommendations, we achieve this by focusing only on autosomes and the X chromosome and masking i) all sites known to be polymorphic in humans as predicted by 1000G, HapMap and dbSNP and ii) repetitive, telomeric and centromeric regions retrieved from the UCSC Table Browser (*112*) by using track *RepeatMasker* in group *Repeats* and track *Gap* in group *Mapping and Sequencing*.

5. Genotype calling
We finally called genotypes based on the recalibrated quality scores and taking post mortem damage patterns into account, as described in (*4*). Using this approach we generated three different sets of calls for each sample:

   a. Diploid MLE calls: we called diploid genotypes at each position in the genome based on Maximum Likelihood using the likelihood model for ancient DNA described in (*4*) This model is "reference free" in the sense that no external information on the present alleles is considered.

14

b. Diploid MLE calls with known alleles: To compare with existing SNP data sets, we also generated MLE calls at all positions in the data set described in (82, 99) using a biallelic model limited to the two alleles reported in those studies.

c. Allele Presence: Finally, we also generated allele presence calls following the approach in (4). These haploid calls report the most likely allele to be present at each genomic position.

Assessment of accuracy

To assess the accuracy of our pipeline to call diploid genotypes at an average coverage of 10x (matching the coverage of the sample WC1) we compared diploid MLE calls obtained from data of the high-coverage (42x) published genome Ust'-Ishim (*81*) and compared these to MLE calls obtained after downsampling this genome to 10x. The downsampled reads were pushed through the entire pipeline independently of the full data. Assuming the calls obtained from using all reads (genome-wide coverage of 42x) to be correct, we found genotypes called with an average coverage of 10x to be highly accurate. Specifically, 99.1% of all genotypes were called identically in both data sets. The accuracy of genotypes found to be heterozygous when using the full data was lower, but 72.3% were still called identical in both data sets. In addition, the vast majority (99.8%) of genotyping errors at heterozygous sites resulted in only one of the alleles being miscalled, mostly (71.7%) resulting in homozygous calls. Overall, only 0.06% of all calls differed in two alleles between these sets.

Comparisons with Existing Samples

In order to ensure bias-free comparisons with other previously reported samples, we obtained raw BAM files from several samples with sufficient coverage and processed them along the samples presented in this study using the same pipeline as described above. The samples for which genotypes were generated this way are:

1. LBK (*82*)
2. Loschbour (*82*)
3. NE1 (*70*)
4. BR2 (*70*)
5. Satsurblia (SATP) (*8*)
6. Kotias (KK1) (*8*)
7. Bichon (*8*)
8. Ust'-Ishim (*81*)
9. Mota (*10*)
10. Kostenki (*100*)
11. Kumtepe4 (*16*)
12. Kumtepe6 (*16*)
13. La Braña (*101*)

Inference of local heterozygosity

Since diversity estimates from called and filtered genotypes likely introduces biases, we inferred local heterozygosity in 1Mb windows for a large array of modern and ancient male samples following the "reference free" approach described in (*9*). To be more specific, we conducted the following steps:

1. We first prepared BAM files and inferred post mortem damage patterns (*9*) following steps 1-5 described above under "Genotype Calling".

2. We next used a robust "reference free" approach to recalibrate quality scores. This approach, described in detail in (*9*), takes post mortem damage into account and does not require any external information (neither knowledge of the reference nor of polymorphic positions) for recalibration. Instead, this approach relies on haploid sequences such as the X chromosome in males. Here we used the first 20Mb of the X chromosome for each individual.

3. We then applied the "reference free" approach to infer heterozygosity within a window, again as described in (*9*). This approach integrates over the uncertainty of genotypes while considering post mortem damage.

In this way, we inferred heterozygosity for five ancient male samples (WC1, BR2, KK1, Bichon and Mota) as well as three low coverage males from each of four 1000 genomes project populations for comparison. These were the British males HG00115, HG00116 and HG00117, the Tuscany males NA20509, NA20511, and NA20762, the Punjabi males HG02600, HG02660, and HG02681, and finally the Yoruban males NA18486, NA18519, and NA18522. For all samples, we excluded 1Mb windows closer than 5Mb to telomeres or centromeres as defined by the track *Gap* in group *Mapping and Sequencing* in the UCSC Table Browser.

Figure S11 shows diversity estimates obtained this way for WC1 in which particularly long runs of homozygosity (ROH) were detected (see below).

We used two different approaches to illustrate similarities between individuals in their distribution of diversity along the genome. First, we ran a PCA on the centered pairwise 1-Spearman correlations using the function *prcomp* in R (Fig. 4). Second, we used the R package *hclust* to cluster individuals based on Euclidean distances to reflect differences in level, and again 1-Spearman correlations to reflect differences in pattern (Fig. S10).

Runs of Homozygosity

We estimated the distribution of runs of homozygosity (ROH) in the genome for WC1 and six other high coverage ancient individuals (*4, 8, 70, 82*), as well as for 2527 modern individuals (*103-104*) following the same methods as outlined in (*4, 102*).

We used a total of 710,488 biallelic autosomal transversions to define ROH. These SNPs had been genotyped securely across all individuals. For the ancient samples, this involved filtering for a depth of coverage of 10 or above and a genotype quality of 30 or above. These steps insured ROH calls were directly comparable across individuals. We also required sites to have a minor allele frequency above 0.5%. ROH were identified using PLINK v1.90 with the specifications used in (*70*). We divided ROH in each individual into two categories, long ROH (>1.6 Mb) and short to intermediate ROH (<1.6 Mb), based on the classes defined in (*11*). We calculated the summed total length of ROH for both classes for each individual and plotted the resulting values against each other in Fig. 3.B. in the main text.

We also placed ROH for the seven ancient and 32 modern individuals in various size bins based on those defined in (*105*). We calculated the total length of ROH in each bin for each individual and plotted it in Fig. 3.A. in the main text.

S7. Principal component analysis

Yoan Diekmann, Mark Thomas

Methods

We generated the PCA plot shown in Fig. 2 using LASER version 2.02 (106). Animation S1 represents the same data in an interactive 3D format including the third principal component. We computed the coordinates in two steps, first for the modern and then for the ancient samples (datasets used here are described in the sections below).

1. Coordinates for modern samples:
We generated a reference space by standard PCA on genotype data of modern individuals and imputed missing entries by averaging the genotypes encoded as {0, 1, 2} over all individuals. We then projected the modern Zoroastrian samples onto this reference using TRACE from the LASER package (*106*).

2. Coordinates for ancient samples:
Second, we mapped ancient samples into the reference space. In this step, the input for LASER is sequence reads in the form of BAM files. Each sample was placed by first simulating sequencing data for each reference individual that matches the coverage pattern of the ancient sample. Then, a PCA was generated for the simulated reference data together with the ancient individual. Finally, Procrustes analysis of the latter PCA with respect to the reference from step 1 allowed us to project the sample into the reference space (*107*).

The advantages of this approach are twofold and especially relevant in the context of ancient DNA. First, the use of BAM files for the ancient samples circumvents the need for genotype calling, which can be challenging for ancient DNA due to contamination, post mortem damage and low coverage. However, the modern reference is still based on genotype data, which is abundant and of high quality for present-day populations. Second, Procrustes analysis allows for robust placement of samples despite low coverage, which is particularly valuable in the case of ancient DNA data.

Modern reference data

We used a genotyping dataset with samples from present-day Eurasian populations from a merged dataset published as part of (*99*) and (*108*) based on Illumina's genotyping platform (Illumina 550, 610, 660W arrays). We removed duplicate individuals (identified by $\pi = 1$ running PLINK version 1.9 (*109, 110*) with option '--genome') and a visual PCA outlier (sample Jordan444). After lifting the coordinates to human genome version hg19 (using PyLiftover (*111, 112*)) we obtained a total of 1,051 individuals and 510,811 autosomal polymorphic loci. The modern populations included in the data set are: Greek, Italian, South Italian, Sicilian, Tuscan, Spanish, French, Basque (South European); Bulgarian, Croatian Hungarian, German, Austrian, Irish, Orcadian, Welsh, Scottish, English, Ukrainian, Polish, Belorussian, Lithuanian, Norwegian, Finnish (East/Central/North European: Romanian); Armenian, Turkish, Georgian, Abhkasian, Adygei, North Ossetian, Balkar, Chechen, Kumyk, Kurd (West Asian); Palestinian, Druze, Lebanese, Syrian, Jordanian, Bedouin, Saudi (Near Eastern); Iranians, Iranian Zoroastrians (Iranian); Nogay, Tajik, Turkmen (West-Central Asian); Burusho, Kalash, Pathan, Sindhi, Balochi, Brahui, Makrani (Pakistani/Afghani); Kyrgyz, Hazara, Uygur, Uzbekistani (Central Asian); Bhunjia, Dhurwa, Mawasi, Nihali, Brahmin, Meena, Kshatriya, Chamar, Chenchu, Hakkipikki, Karnataka, Piramalaikallar, Dusadh, Kurumba, Velamas, Tamilnadu,

Kanjar, Muslim, Upcaste, Tharus, Lambadi, Kurmi, Meghawal, Bengali, Dharkar, Gond, Kol (Indian).

In addition, we newly present a set of Zoroastrian samples genotyped on the Affymetrix Human Origins chip.

Coloured areas highlighting the extent of modern populations in Fig. 2 are 95% contours of 2D kernel density estimates computed with the *ks* package in R.

Ancient data

All ancient samples shown in Fig. 2 are projected into the reference PCA space by LASER as explained above. Please see the Supplementary Information accompanying our recent publication (*4*) for details on the ancient samples, and Table S13 for number of sites used for projection and projection confidence.

S8. Shared drift paths and allele sharing amongst ancient and contemporary populations

Zuzana Hofmanová, Yoan Diekmann & Krishna Veeramah

<u>Introduction</u>

We used *f* statistics and *D* statistics via ADMIXTOOLS 1.1 package (as developed by (*6*)) in order to infer the population history of our ancient Iranian samples within the context of other modern and ancient Eurasian samples by determining relative levels of shared drift via correlations in allele frequencies (*f3* statistics) or differences in relative allele sharing (*D* statistics). In particular, we examined a) the relationships amongst the ancient Iranian samples generated in this study, b) to what extent these Iranian samples share ancestry with previously published pre-Neolithic populations and b) whether the ancient Iranian samples demonstrate evidence of being descendants of Neolithic and post-Neolithic populations.

<u>Material and Methods</u>

To maximize the number of populations (geographically and chronologically) to which we could compare our ancient Iranian samples, we limited the calculation of our *f* statistics and *D* statistics to SNPs found in the reference dataset from (*82*), while also including samples from (*3*), as well as other ancient DNA samples with whole genome data from literature (Table S14 for samples added to the reference data set). We ensured that duplicate samples and close relatives were not analyzed together by excluding one individual from each pair that had pi-hat higher than 0.2, resulting in total of 2,252 individuals including our five ancient Iranians.

In order to avoid bias introduced by differences in sequence quality and genotype calling amongst samples, we considered only pseudo-haploid calls. For reference samples from (*3, 82*), we chose a random allele from each heterozygous call, while Allele Presence calls with Phred-scaled Quality >= 30 (see Supplementary Section 6) were used for the samples we processed independently (Table S14) and for the sample from this study (Table S1). Additionally, conclusions were confirmed by analyzing a data set consisting of transversions only (as potential bias from post mortem damage would primarily affect transitions). Results with and without transitions were correlated with Spearman correlation (ϱ=0.80 for f-values and ϱ=0.81 for Z-scores) but it should be noted that power to identify significant *f* statistics and D statistics was substantially reduced when transitions were removed (resulting in only ~101K positions). We used Plink 1.9 *(109, 110)* for pi-hat calculation and data formatting and we used the programs from ADMIXTOOLS package (*6*) to calculate the *f* statistics (qp3Pop) and *D* statistics (qpDstat), with associated Z-scores estimated via a block jackknife re-sampling procedure using the software's default options.

<u>Shared drift with modern reference populations</u>

We used "outgroup" *f3* statistics as an exploratory statistical tool to establish relative relatedness of modern reference population to the ancient Iranian individuals. *f3* statistics of the form *f3*(C; A,B) usually test if the target population (C) is the result of admixture between two source populations (A, B) (or their close approximations). However, if we set the target population as an "outgroup" to A and B, we will determine a measure of the relative shared drift between A and B with respect to C. If A (or B) is set as a population to be studied, we can compare different B (or A) populations to determine their relative genetic similarity.

We applied this test of the form *f3(‡Khomani; Test, ancient_Iranian),* where *Test* was one of the modern populations in the dataset and the ‡Khomani San population was chosen as an

outgroup (it likely diverged from all other modern humans at least 100 kya and was therefore not likely to have experienced Eurasian admixture *(64, 113)*.

The four Neolithic Iranian samples WC1, AH1, AH2 and AH4 all demonstrate greatest genetic similarity to modern populations from an area spanning from Caucasus region to northern South Asia (Fig. S12-S15), with the highest values observed for two Iranian populations, Zoroastrians and Fars. While the Iron Age sample from Hasanlu also shows genetic similarity to some of the same populations as the Neolithic samples, Mediterranean and other European populations were also found amongst the 20 highest *f3* outgroup values (Fig. S16).

Shared drift with ancient populations

We also calculated "outgroup" *f3* statistics of the same form *f3(‡Khomani; Test, ancient_Iranian)* with *Test* being one of the ancient populations from the reference dataset (populations from dataset of *(3, 82)* and the samples shown in Table S14). Neolithic Iranians from Tepe Abdul Hosein demonstrated greatest genetic similarity with each other (Fig. S17-S20) and, somewhat surprisingly, with the Anatolian Chalcolithic sample Kumtepe6 *(16)*. The closeness among Tepe Abdul Hosein samples is consistent with their shared archaeological context (Supplementary section 1) and similar ^{14}C dating (Table S2), suggesting that all these individuals were likely sampled from the same population of early farmers from this area. The individual from Wezmeh Cave (WC1) did not originate from a context that would be informative on the subsistence strategy but had highly similar *f3* statistics results to AH1, AH2 and AH4. This, together with geographical and chronological proximity, suggests this individual also belonged to the general farming population in the region or was highly related to it. Therefore, whenever analyses require high base-pair coverage, WC1 was used to represent Early Neolithic Iranians (see Supplementary Sections 6 and 9). We obtained other relatively high outgroup *f3* statistic values for Neolithic Iranians in tests involving some of the Early European farmers, specifically Greek (Hofmanová *et al*. (4): Greek Early Neolithic and Greek Final Neolithic samples, Greek_EN and Greek_FN, respectively) and Hungarian *(3, 70)*, Bronze Age steppe and Hungarian populations *(3, 82)* and for Caucasus hunter-gatherers (CHG): KK1 and SATP from *(8)*. CHG in particular consistently demonstrate some of the higher *f3* statistic when compared to the Neolithic Iranians, generally showing significantly greater similarity to them than F38.

When F38 is itself utilized as the target population (*f3(‡Khomani; Test, F38)*), while it tends to also demonstrate similarity to the Neolithic Iranians as well Kumtepe6 and CHG, the "outgroup" *f3* statistic values are much more comparable to those for other European Neolithic samples (Fig. S21), in particular Greek Early (Rev5) and Final Neolithic samples from Hofmanová *et al*. 2016 (Klei10, Pal7) and CHG samples. Neolithic Iranians were abbreviated as *Neo_Iranian* and F38 was similarly labeled as *IA_Iranian*.

Genetic structure among ancient Iranian samples

Following the nomenclature of Patterson *et al.* (*6*), this statistics takes the form of D(W,X;Y,Z) and, if Z is set as an outgroup that has not experienced post divergence gene flow with W, X or Y, tests the consistency of the phylogeny (((W,X)Y)Z) by estimating relative derived allele sharing of W with Y versus X with Y. Significant deviations from 0 for this statistics suggest a violation of this tested tree and some form of unique shared ancestry (either because of gene flow or population structure in the ancestral W-X-Y population). Significantly positive *D* statistics will result because of shared ancestry between W and Y, and significantly

negative values will result because of shared ancestry between X and Y. In our implementation here ǂKhomani San is set as the Z or outgroup population, with W, X and Y being various combinations of modern or ancient Eurasian population.

These tests confirmed the inference of Neolithic Iranians being sampled from a single prehistoric Zagros population (see Table S15.1-2 for detailed results of *D* statistics). The results of tests involving F38 of the form *D(Neo_Iranian, Test, F38, ǂKhomani)* and *D(IA_Iranian, Test, Neo_Iranian, ǂKhomani)* were significantly positive for all modern populations. However, when we examined ancient groups we found evidence that F38 shared unique ancestry with Kumtepe6 compared to any Neolithic Iranian sample (Tables S15.4-S15.5), pointing to some level of potential gene flow between Iran and more westward populations after the Neolithic.

These results are consistent with the previous "outgroup" *f3* statistics suggesting different gene flow into Neolithic and Iron Age Iranians from other sources. In order to better characterize to what extent they differ with regard to their genetic similarity to the reference samples, we calculated *f4* statistics of the form *D(Neo_Iranian, IA_Iranian, Test, ǂKhomani)* with *Test* being one from other ancient populations. Unlike above we did not find a significant difference with regard to their relationship to Kumtepe6, though this may due to reduced power as a result of the low coverage of Kumtepe6. However, a number of significantly negative values (Table S20) were found, suggesting greater similarity of almost all other ancient populations, including *Steppe* populations (Yamnaya and pre-Yamnaya samples from (*3, 18*)) and EN Anatolian farmers (*Anatolian_Neolithic* combined from (*3, 4*)) to the Iron Age Iranian sample with only CHG and Mota showing greater similarity to the Neolithic Iranians (Table S15.6).

Origin of Iranian samples

In order to better understand the relationship of Neolithic Iranian to older hunter-gatherer populations, we calculated *D* statistics of the form *D(Neo_Iranian, pre-Neolithic, pre-Neolithic, ǂKhomani)* and *D(pre-Neolithic, pre-Neolithic, Neo_Iranian, ǂKhomani).* When calculating *D* statistics for *pre-Neolithic* populations we grouped samples, similarly to the source literature, as CHG (Caucasus hunter-gatherers; KK1, SATP from (*8*)), WHG (Western hunter-gatherers; LaBrana, Loschbour, Bichon, and other WHG samples identified as such in Mathieson *et al.* dataset (*3*)), EHG (Eastern hunter-gatherers; identified as such in Mathieson *et al.* dataset (*3*)) and SHG (Scandinavian hunter-gatherers; Motala samples from Lazaridis *et al.* (*82*), other samples identified as such in (*3*)), while other HG samples (Kostenki, Ust'-Ishim, MA1, AG2) were considered as representatives of distinct populations. However, we note that the ancient Iranians were analyzed on a per sample basis in order to avoid potential biases due to grouping during the first study they are analyzed in.

All combinations of the form *D(Neo_Iranian, pre-Neolithic, CHG, ǂKhomani)* and *D(CHG, pre-Neolithic, Neo_Iranian, ǂKhomani)* were significantly positive (Table S20) with high Z-scores (9.2<|Z|<35.3). This high amount of shared ancestry between CHG and Neolithic Iranians is consistent with previous results of "outgroup" *f3* statistic. However, when we directly tested if the CHG and Neolithic Iranians branched together with respect to other pre-Neolithic samples, we strikingly observed significantly negative *D* statistic in the form *D(Neo_Iranian, CHG, pre-Neolithic, ǂKhomani)* for WHG, SHG and EHG (Table S16.1). The results were not as high as the shared ancestry between CHG and Neolithic Iranians as |Z| < 8.4 and the actual values were also at least one order lower Additionally, we observed a decrease of Ust'-Ishim-like and Neanderthal ancestry (only significant for the Vindija genome) in

Neolithic Iranians (see Tables S16.2-4). This is unlikely to be due to a relative increase of Mota-like ancestry (see Tables S16.5-6).

Differences in genetic affinities of Near Eastern Early farmers

We recently demonstrated that Neolithic Aegean (Anatolian and Greek) populations are likely to be the source of early Neolithic farmers that spread into the rest of Europe (*3, 4*). Given that it has been proposed that farming ultimately developed even further east than the Aegean, we investigated whether Early Neolithic Iranian farmers (*Neo_Iranian*) may be the ultimate source for European Neolithic farmers. In particular, we examined this relationship within the context of Eurasian HG (hunter-gatherers), which would be an expected outgroup to Aegean and Iranian Neolithic farmers if all early farmers ultimately originated from one place and one population. The alternative hypothesis would be that Anatolian and Iranian farmers are derived from distinct hunter-gatherer populations (see the different models illustrated in Fig. S23).

When we calculated *D* statistics of the form *D(EN_farmer, Neo_Iranian, Euroasian HG, ǂKhomani)* where *EN_farmer* stands for Early Neolithic farmers except for Iranian Neolithic (farmer population names adopted from the reference literature), we generally observed significantly positive values, while *D(EN_farmer, Neo_Iranian, CHG, ǂKhomani)* estimates were significantly negative. This hints at different hunter-gatherer origins for Aegean and Iranian farmers, though distinct hunter-gatherer gene flow events into each farmer population after their divergence would also be compatible (Table S20 for all results, Table S17.1, Table S17.2, Table S17.3 and Table S15.3 for the early periods).

It seems likely that Neolithic Iranians and CHG may form a clade. However, when we estimated *D* statistics of the form *D(Neo_Iranian, HG, EN_farmer, ǂKhomani)*, we saw a pattern where CHG (and other HG) shared more drift with the external population, which in this case is the Anatolian farmers rather than European hunter-gatherers as in previous analysis in Table S18.1). This could be a result of gene flow between CHG and the ancestral Anatolian farmer/European hunter-gatherer population, which would appear to reinforce our earlier suggestion that CHG and Neolithic Iranians likely diverged quite far in the past (i.e. before the source population for localized neolithisation in the Aegean was established), though we cannot rule out independent gene flow events for both hunter-gatherers and Aegean farmers.

Only when we examine Ust'-Ishim, can we identify a hunter-gatherer population compared to which Anatolian and Iranian farmers share unique ancestry, implying that Ust'-Ishim could represent an outgroup to both farmer groups (Table S17.4).

When we perform the test *D(Anatolian_Neolithic, HG, Neo_Iranian, ǂKhomani)* (Table S17.5) we find that all European hunter-gatherers form a clade with Anatolian farmers with only a few rather sporadic significant values, except for CHG, which we believe is most closely related to Neolithic Iranians. Thus, our results again suggest that the Anatolian and Iranian farmer populations likely diverged from different hunter-gatherer populations, i.e. European HG and CHG, respectively. A less plausible alternative is that following a demic diffusion from Iran there was subsequently substantial levels of admixture with different European HG populations.

However, when we tested whether Anatolian Neolithic samples were a mixture of HG and Neolithic Iranian ancestry using a *f3* statistics test of the form *f3(Neo_Iranian, HG; Anatolia_Neolithic)* (with *inbred* option turned on), we obtained significantly positive results, allowing us to reject this hypothesis (Table S17.6). We conclude that Anatolian Neolithic

population is not a product of mixture of Neolithic Iranians and any of hunter-gatherer populations known from the literature.

Genetic affinities after the Neolithic

"Outgroup" *f3* statistics discussed above had intriguingly shown that Kumtepe6, a low-coverage Chalcolithic sample from Anatolia, demonstrated some of the highest levels of genetic similarity to Neolithic Iranians. In order to examine this in more detail and to identify other populations with genetic affinities to Neolithic Iranians, we estimated *D* statistic of the form *D(European_Early_farmer1, Neo_Iranians, European_Early_farmer2, ‡Khomani)* where *Early_farmer* groups were European early farmers from the reference dataset except for studied ancient Iranians. We found that Kumtepe6 (Table S18.1) shared unique ancestry with Neolithic Iranians, consistent with the previous observations. In addition, a similar pattern was also seen for the Greek Final Neolithic samples (Klei10, Pal7), which have been previously described as having contributed to Kumtepe6 ancestry (4), for certain combinations of populations.

Similarly, when we performed tests of the form *D(Early_farmer1, Early_farmer2, Neo_Iranian, ‡Khomani)* (Table S20) we again obtained high amounts of shared ancestry between Neolithic Iranians and Kumtepe6 as well as Greek Neolithic samples (both Early and Final Neolithic). The populations sharing the least ancestry with Neolithic Iranians are Remedello and Iceman, which are chronologically later than other Iberian, Hungarian, Anatolian and Central European farmers tested. For example *D(Remedello, LBK_EN, AH2/AH4/WC1, ‡Khomani)* is significantly negative and Remedello is thus the only population sharing significantly less ancestry with Iranians than LBK (actually the opposite is true for most population including early Neolithic samples from Hungary, Anatolia and Iberia).

Connections between ancient Iran and Yamnaya-like populations

Mathieson *et al*. 2015 (*3*) have previously suggested that pre-Yamnaya populations may possess Near Eastern ancestry. To examine if this ancestry may in fact be derived from Neolithic Iranians coming from the Fertile Crescent we estimated *D* statistic of the form *D(Steppe, EHG, Neo_Iranian, ‡Khomani)*. Significantly positive values were observed (Table S18.2, all in Table S20) for one comparison with Eneolithic Samara population and these values increased in later populations from this area, supporting this hypothesis, though we also note that no results were significant for Russian Early Bronze Age samples (Volga steppes).

However, when we examined whether these post-Neolithic ancient Steppe populations were closer to CHG or Neolithic Iranians (tested as *D(Neo_Iranian, CHG, Steppe, ‡Khomani)*,) we observed significant negative values for all but one Steppe population (Table S20). Therefore it appears likely that the Near Eastern drift observed in the Yamnaya and pre-Yamanya populations derives from CHG rather than Neolithic Iranians.

A direct test of admixture via *f3* statistics is not possible for one Iron Age sample but instead we tested if Neolithic Iranians can serve as a proxy for a source population of Yamnaya-like populations from Samara region. We tested *f3(Neo_Iranian, Test; Steppe)*, where *Test* is an ancient populations from literature (both pre- and post-Neolithic). For EHG and SHG as source populations, several comparisons showed (Table S18.7) significantly negative result (thus suggesting admixture). Also, one value (*f3(Neo-Iranian, Samara_Eneolithic; Poltavka)* showed that there was an increase of ancestry related to Neolithic Iranians in Poltavka (Middle Bronze age, post-Yamnaya culture) when compared to earlier periods in the same region.

Iron Age samples are very scarce in the datasets comprising Scythian Iron Age samples (*3*) and our sample F38. It was shown above that this sample shares high amount of ancestry

with Neolithic Iranians (Fig. S17-S21) but also has some specific patterns of ancestry shared with Kumtepe6 and Greek Final Neolithic and less of CHG-like and Mota-like influence (Table S15.4- S15.6). In order to investigate if we could pinpoint a source of this gene flow with F38, we tested $D(Early\_farmer1, F38, Early\_farmer2, \ddagger Khomani)$ and we found the most significant negative results for combinations involving Samarian and Yamnaya population as Early_farmer1 and Aegean and European Early farmer populations as Early_farmer2 (Table S18.3 as an example for Anatolian Neolithic as Early_farmer2). We therefore suggest that the source of the non-local gene flow in F38 is more likely to be related to the Anatolians than the "Samarian" (Yamnaya-like) populations. (see Table S18.5 for exceptions of this pattern detected via further D-statistic. A systematic exception was only ound for the Steppe Sintashta population that still shares less drift F38 than more westward populations, see Table 18.6).This result is relevant for the spread of Indo-European languages as we fail to prove a direct link between Yamnaya-like populations (a potential source) and the Iran area before the Iron age, whereas a link to Anatolian population seem to be stronger.

The exceptions to the pattern seen above are found in Table S18.4 and they concern rather low coverage samples Kumtepe6 and Greek Early Neolithic sample (Rev5) that was shown previously by "outgroup" *f3* statistic to be close to F38.

Ancestry of modern Iranian populations

Both Iron Age and Neolithic Iranians showed close affinities to modern populations as can be seen on Fig. S12-S16. We directly investigated if modern populations can be identified as mixture of Neolithic Iranians (for this test more than one sample per population is necessary) and other ancient populations in the dataset. Using the *inbred* option, we tested *f3*(*Neo_Iranian, Ancient; Test*), where *Test* was one of the modern populations in the dataset and *Ancient* was one of ancient DNA populations in the dataset.

We present significant results for Anatolian, Steppe and EHG/WHG/CHG populations in Table S19 (other results in Table S20 are showing similar patterns for related populations). As can be seen, the modern samples Near Eastern and Indian repeatedly showed significant negative values for this test. Interestingly, Iranian Zoroastrians, that had high "outgroup" *f3* statistic values with Neolithic Iranians (Fig. S12-S16), were not a result of mixture of Neolithic Iranians with known ancient populations, whereas most modern Iranian populations in the dataset tested significant for this test.

Neanderthal ancestry

All statistics presented in this section are computed with 'qpF4ratio' from the ADMIXTOOLS package (*6*). We implement the f4-ratio scheme described by Fu *et al.*, 2016 (*14*), in Section 3 (pp. 21) of the SM, however, we use reference population data from Lazaridis *et al.*, 2014 (*82*) instead of full genome sequences from the Simons Genome Diversity Project Dataset (*104*). Therefore, all our results are based on a maximum of around 500K SNPs, which explains the differences we observe compared to the original values for samples published in Fu *et al.* that we re-analyse here.

The Neanderthal ancestry proportions are given in Table S21. Samples from three different sources are included (specified in column 'dataset'): 'Fu', 'Lazaridis', and 'PAT_gt'. As already mentioned above, we re-analyzed the samples published in Fu *et al.* 2016 (labeled 'Fu'), excluding low coverage samples as in Figure 2 of Fu *et al.*. This is 'pseudo-haploid' data

24

resulting from a random allele calling procedure, except for Loschbour, Ust_'Ishim and Stuttgart, which are diploid. Contemporary populations are from Lazaridis *et al.*, 2014 (*82*) (labeled 'Lazaridis'), and consist of diploid data. Finally, samples labeled 'PAT_gt' are samples we called with our genotype caller described in Supplementary section S6.

We see the same trend of reduction in Neanderthal ancestry with time as observed by Fu *et al.* 2016 (Fig. S22). We obtain good agreement of inferred ancestry proportions between samples that are both in the original Fu *et al.* dataset and were re-called with our genotype caller, including the 'pseudo-haploid' Fu *et al.* samples without heterozygote sites. We infer very low levels of Neanderthal ancestry for WC1. However, WC1 does not behave as an outlier as temporally close samples are inferred to have similar proportions.

S9. Comparing haplotype sharing patterns amongst samples using a mixture model

Lucy van Dorp, Saioa López & Garrett Hellenthal

Introduction

The aim of this section is to use a novel statistical mixture model to represent the DNA from modern and ancient groups or individuals as mixtures of that from other sampled groups or individuals. These mixture patterns identify which sampled groups are most related to one another genetically, reflecting shared common ancestry relative to other groups due to e.g. admixture or other historical processes. There are two major advantages of our mixture model approach over commonly used techniques to measure genetic structure, such as F_{ST} (*114*) or f statistics (*6*). The first is that under our approach each group/individual can be described as a mixture of the DNA of other groups/individuals, in contrast to being compared to only one group or a few groups at a time. The second is that our approach here uses haplotype (or phase) information, which can considerably increase power to infer demographic events and population structure (*68, 108, 115*) while being less susceptible to biases that result from SNP ascertainment schemes (*66, 67*).

Methods

<u>Description of dataset analyzed</u>

We merged our new highest coverage Neolithic sample from Iran (WC1) with two other existing merged datasets:

a. "Lazaridis": contains genomes of modern individuals from (*82*), plus additional individuals described below, all genotyped on the Affymetrix Human Origins chip (final dataset after procedure described below which includes aDNA samples: 2317 individuals from 230 labeled groups and 525,796 SNPs).

b. "Busby": contains genomes of all modern individuals from (*99, 108*) and additional modern Armenian groups from (*116*), all genotyped on Illumina Infinium series chips (final dataset after procedure described below which includes aDNA samples: 2576 individuals from 174 labelled groups and 386,642 SNPs).

Throughout we use the labels from (*82, 99, 116*) to refer to the groups in these datasets. For each of the Lazaridis and Busby merges, in addition to WC1, we also merged in aDNA samples that were called as described in this paper. This included aDNA from ancient hominid groups Altai Neanderthal (ALT) and Denisova (DEN), as well as a Mesolithic hunter-gatherer individual sampled from Luxembourg ("Luxembourg_Mesolithic" or Loschbour), a Neolithic individual sampled from Germany ("Stuttgart-LBK380" or LBK), a Neolithic individual sampled from Anatolia (Bar8) (*4*), a Neolithic hunter-gatherer sampled from west Georgia (KK1) (*8*), a Neolithic individual sampled from Hungary (NE1) (*70*), a 4,500 year old genome sampled from Ethiopia (Mota) (*10*) and a 45,000 year old genome sampled from western Siberia (Ust'-Ishim) (*81*).

For the Lazaridis merge, we also added novel modern samples from two groups from Iran that were genotyped on the same Affymetrix Human Origins chip; groups that we label based on their self-identification of religion or ethnicity. This gave us two new modern groups for the

Lazaridis merge beyond those included in (*82*): Iran_Zoroastrian (29 individuals) and Iran_Fars (17 individuals).

When generating each merge, we extracted only overlapping positions with a quality score >10 and a read depth >2 from each of the ancient individuals, and excluded any SNPs across the modern samples with genotype call rates <95% using PLINK v1.9 (*109*). For the Lazaridis merge, we removed 18 individuals (Href, Chimp, Gorilla, Orang, Macaque, Marmoset, Denisova_light, Vindija_light, Mez1, Otzi, Saqqaq, MA1, AG2, Skoglunk_HG, Skoglund_farmer, Motala_merge, Motala12 and Labrana) from the original dataset presented in (*82*) that either did not have heterozygous calls or were not of interest in this study, plus the Jewish populations due to their migration and varying admixture history. For the Busby merge, we lifted the coordinates of this merge to human genome version hg19 (using PyLiftover (*111, 112*) and additionally (i) removed one individual from each pair of related individuals with a PLINK inferred IBD coefficient PI_HAT >0.25 and (ii) excluded SNPs across the modern samples with a minor allele frequency less than <5% using PLINK v1.9 (*109*).

In total, for the Lazaridis merge we analyzed 2317 individuals at 525,796 SNPs. For the Busby merge, we analyzed 2576 individuals and 386,642 SNPs. For each merge, haplotype phasing was performed using SHAPEIT (*117*) with default parameters and incorporating the build 37 genetic map combined across populations available at:

https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html.

Inferring "painting profiles"

We followed the chromosome painting approach described in (*118*) to compare haplotype patterns of a "recipient" chromosome to those from a set of "donor" chromosomes. In particular, we used the program CHROMOPAINTER (*118*) to infer along the recipient's genome which donor they are most closely related to ancestrally relative to all other donors.

For both the Lazaridis and Busby merges, we used all modern groups as donors when painting (i.e. as recipients) each modern and ancient individual, where "ancient" individuals refer to: WC1, KK1, NE1, Loschbour, LBK, Mota, Ust'-Ishim, Bar8 and "modern" individuals refer to all other groups in the merged dataset (including the ancient hominid groups ALT and DEN). Under this painting approach any recipient individual cannot use itself as a donor. For this reason, when constructing our "painting profiles" we used a "leave-one-out" approach, analogous to the one described in (*4*) and (*108*). In particular, if each donor group {1,...,D} contains {n_1,...,n_D} individuals, respectively, with $N = \sum_{d=1}^{D} n_d$ total donor individuals, we fix the set of donors to contain n_d-1 individuals from each of the *D* groups (i.e. giving *N-D* donor individuals in total). For example, under our described painting analysis each Iranian individual can only use $n_{Iranian}$-1 other Iranian individuals as donors, and therefore we fix every other recipient individual to use only $n_{Iranian}$-1 Iranian individuals as donors. Exceptions to this "leave-one-out" rule are the two ancient hominid genomes ALT and DEN, for which we have only a single sample of each; for this reason these genomes are used as donors for every group and otherwise are not used as contributors when inferring final "proportions of ancestry" as described below.

When using CHROMOPAINTER, we initially estimated the mutation/emission ("-M") and switch rate ("-n") parameters using 10 steps of Expectation-Maximisation (E-M) algorithm ("i.e. -i 10 -in –iM"), starting with default values and running on a subset of individuals (every 10 in our merged datasets) for a subset of chromosomes (1,4,15,22). We averaged inferred

values of each parameter across these chromosomes, weighting the average by number of SNPs, and then averaged across individuals. We then fixed these values (i.e. using "-M" and "-n") and ran on all chromosomes and all individuals. We otherwise used default values in CHROMOPAINTER. For the Lazaridis merge, we used {-M =0.000703, -n = 223.3433}. For the Busby merge, we used {-M =0.00059, -n = 236.5327}.

For each recipient r, we define f_d^r to be the total proportion of genome-wide DNA for which individual r is inferred to be most closely related to (i.e. is painted by) a donor chromosome from group d (i.e. based on the "chunklengths.out" output from CHROMOPAINTER) (*118*). We note that r here can refer to a single individual (e.g. for ancient samples) or can represent an average across individuals with the same group label (e.g. for modern groups). Here we let each distinct population label represent a different donor group d, leading to D=222 total donor groups for the Lazaridis merge and D=166 total donor groups for the Busby merge.

Allele presence merge and inferring "allele frequency profiles"

In order to test the robustness of our findings, we performed a separate analysis that did not use our diploid calls, following the procedure described in (4). In particular, we performed a merge with the Busby dataset based on a different "allele presence" calling technique comprising the ancient samples WC1, LBK, NE1 and Bar8 that treated each aDNA sample as a haploid genome. As in (*4*) let H_1^i, ..., H_L^i be the observed data for a sampled haploid chromosome i at SNPs 1,...,L, where H_l^i is the probability the given SNP is of allele type a_l at SNP l (here a_l is arbitrarily chosen to be one of the two possible allele types at SNP l according to the Busby merge). Here each modern individual (plus ALT, DEN) has two haploid chromosomes with $H_l^i \in \{0,1\}$ for all non-missing data. In contrast, each aDNA sample (i.e. WC1, NE1, LBK, Bar8) is represented as a single haploid chromosome with $0 \leq H_l^i \leq 1$ based on the posterior probability of being a particular allele type according to our calling algorithm, subject to the modifications below.

We first removed SNPs where for any new aDNA sample:

- Coverage < 2
- $Q_i > 0.001$, where Q_i is equal to 1 minus the maximum posterior probability at SNP i across the four possible allele types (A,G,C,T)
- The allele type with maximum posterior probability does not match the two possible allele types at SNP i in the Busby merged dataset

For SNPs passing this criteria, if Q_i=0 we set H_l^i=1.0 or H_l^i=0.0, depending on which allele matching that in the Busby merge had the highest posterior probability. For SNPs with Q_i>0, we checked whether the allele type with the second highest posterior probability matched the other possible allele type in the Busby merge. If so, we set H_l^i= 1.0 - Q_i (respectively H_l^i=Q_i, depending on which allele matching the Busby merge had the highest posterior probability); otherwise we set H_l^i=1.0 (respectively H_l^i=0.0). Due to these strict criteria, note that H_l^i will nearly always be 0 or 1, or extremely close to these values, and can thus be thought of as such throughout the following.

To infer an "allele frequency profile", we followed the "unlinked" approach described in (*118*) (e.g. the "unlinked coancestry matrix") to compare the alleles of a recipient haploid chromosome to that of a set of donor haploid chromosomes, while accounting for uncertainty

in the calls. Again, as in (4), let $H_1^i, \ldots, H_L^i$ be the observed data for a sampled haploid chromosome i at SNPs 1,...,L, where H_l^i is the probability the given SNP is of allele type a_l at SNP l, as described above. For recipient chromosome i, let $X_l^i(k)$ be the score assigned to donor chromosome $k \in [1, \ldots, K]$ at SNP l, with K the total number of donor chromosomes and:

$$X_l^i(k) = \frac{H_l^i H_l^k + (1.0 - H_l^i)(1.0 - H_l^k)}{\sum_{j=1}^{K} H_l^i H_l^j + (1.0 - H_l^i)(1.0 - H_l^j)} .$$

We then calculate the total genome-wide allele matching score for recipient i and donor k as $X^i(k) = \sum_{l=1}^{L} X_l^i(k)$. We can then sum $X^i(k)$ across donor chromosomes k that belong to the same donor group d. Similarly, for recipient individuals with two haploid chromosomes, we can sum $X^i(k)$ across these two haploid chromosomes to get a final vector of scores for that recipient individual, or we can sum $X^i(k)$ across all i from a given recipient group to get a final vector of scores for that group.

As noted above, we partition our K donor chromosomes into D donor groups (here each of these D groups refers to a distinct population label). We define the "allele frequency profile" for recipient r as $f^r \equiv \{f_1^r, \ldots, f_D^r\}$, with:

$$f_d^r = \frac{\sum_{k=1}^{K} 1_{[k,d]} X^r(k)}{\sum_{j=1}^{D} [\sum_{k=1}^{K} 1_{[k,j]} X^r(k)]} ,$$

where $1_{[k,d]} = 1$ if donor chromosome k is assigned to donor group d and 0 otherwise. Note that $\sum_{d=1}^{D} f_d^r = 1.0$. As noted above, r can represent a single haploid chromosome, a single individual, or all haploid chromosomes from a common group.

We performed this approach using the same set of donors and recipients as described for the inference of "painting profiles", e.g. using all modern groups as donors when calculating the allele frequency profile for each of WC1, NE1, LBK, Bar8, with "modern" groups referring to all modern groups in the Busby merge and the ancient hominid groups ALT, DEN as before. Results under this analysis are provided in Fig. S38.

To assess the effect of different sequencing depth in the ancient samples, we also randomly sub-sampled SNPs across NE1, an aDNA sample with a high average coverage >20, to exactly match the coverage distribution of WC1, which has an average coverage of 10.4. This sub-sampling and coverage matching strategy led to 95,161 SNPs, with analysis results provided in Fig. S38d.

Measuring differences in inferred painting profiles (TVD)

In order to quantify differences in the inferred painting (or allele frequency) profiles between each of the ancient samples, we apply a total variation distance (TVD) measure as in (68, 115). As before, let f_d^r be the genome-wide proportion of DNA that a recipient individual (or group) r copies from each of donor individuals or groups $d \in [1, \ldots, D]$. To compare the painting (or allele frequency) profiles of two recipient individuals (or groups) X and Y - in this case each pairwise combination of the ancient samples (Ust'-Ishim, KK1, Loschbour, WC1, Bar8, NE1, LBK, Mota), we calculate TVD_{XY} as follows:

$$TVD_{XY} = 0.5 \sum_{d=1}^{D} |f_d^X - f_d^Y|.$$

For the *TVD* results when comparing aDNA samples to only African groups as depicted in Fig. S31ab and S38f, instead of donor *groups* we summed differences across all d ∈ [1,...,D] donor *individuals*, using *D*=125 African individuals for the Lazaridis merge and *D*=64 African individuals for the Busby merge. Similarly we used this approach for Fig. S31cd, though with the addition of a single Sardinian donor individual.

<u>Inferring final "proportions of ancestry" based on the "painting profiles"</u>

Our inferred "painting profiles" suffer some limitations. For example, *a priori* a donor group *d* with a disproportionately large number of sampled individuals may have relatively higher proportion of matching haplotype patterns (i.e. relatively large f_d^r) across all recipient groups $r \in [1,..., R]$, potentially leading to a biased interpretation of results. To cope with this, we use additional mixture modeling described in this section to "clean" the raw CHROMOPAINTER inference. In particular, we can use this technique to compare a group's painting profiles to that of any other groups we include in the mixture. In contrast to the mixture model techniques described in (*4, 68, 108, 115*), we use a novel statistical method that gave qualitatively more intuitive results when applied to modern groups.

Let $l^r \equiv \{l_1^r, ..., l_D^r\}$ be the observed CHROMOPAINTER painting for recipient individual (or group) *r*, with l_d^r the centimorgan (cM) length of genome-wide DNA that individual (group) *r* paints (or copies) from donor group $d \in [1,..., D]$. Here for any *r*, $\sum_{d=1}^{D} l_d^r = C$, where *C* is equal to the total genome length of DNA (in cM), and note that $f_d^r \equiv \frac{l_d^r}{C}$. We have analogous painting profiles l^j for all other recipient groups $j \neq r \in [1,..., R]$, which we can convert to f^j. To measure the relative amount of drift (or "self-copying") in group *r* in analysis (III) described below, we introduce a *D*-vector f^{r*} with $f_r^{r*} = f_r^r$ and all other entries 0.

We assume that:

$$Pr(l^r | l^1, ..., l^S, C, \beta^r) = Multinomial \left(C; \sum_{s=1}^{S}[\beta_s^r f_1^s], ..., \sum_{s=1}^{S}[\beta_s^r f_D^s] \right),$$

where $\beta^r \equiv \{\beta_1^r, ..., \beta_S^r\}$ are the mixture coefficients we aim to infer and *s* = 1, ..., *S* represents a set of "surrogate" groups used to describe the ancestry of group *r*. Specifically, the set of surrogates can contain all other *R* - 1 recipient groups, or it may contain any subset of these other *R* - 1 recipient groups. For analysis (III) described below, i.e. when allowing self-copying, this set of surrogates includes f^{r*} defined above. We explore several combinations of surrogates below.

We take a Bayesian approach to inferring β^r, further assuming the following:

$$Pr(\beta^r | \lambda) = Dirichlet (\lambda, ..., \lambda),$$
$$Pr(\lambda) = Uniform(0,1).$$

For each recipient *r*, we wish to sample the mixing coefficients $\{\beta_1^r, ..., \beta_S^r\}$ based on their posterior probabilities conditional on $l \equiv \{l^r, l^1, ... l^S\}$. We do so using the following Markov Chain Monte Carlo (MCMC) technique. We start with an initial value of $\lambda(0) = 0.5$ and sample

30

our initial values of $\beta^r(0) \equiv \{\beta_1^r(0), \dots, \beta_S^r(0)\}$ from the prior distribution Dirichlet $(\lambda(0), \dots, \lambda(0))$. Then for $m = 1, \dots, M$:

A. Update $\beta^r(m) \equiv \{\beta_1^r(m), \dots, \beta_S^r(m)\}$ using a Metropolis-Hastings (M-H) step:

    (i) Randomly choose two surrogates s_1, s_2.

    (ii) Randomly sample $X \sim$ Unif $(0, 0.1)$.

    (iii) Set $\beta_{s1}^r(m) = \beta_{s1}^r(m-1) + X$.

    (iv) Set $\beta_{s2}^r(m) = \beta_{s2}^r(m-1) - X$.

    (v) For all $s \neq \{s1, s2\} \in [1, \dots, S]$, set $\beta_s^r(m) = \beta_s^r(m-1)$.

    (vi) For numerical stability, if $\beta_{s1}^r(m) > 1 - 1e^{-7}$, set $\beta_{s1}^r(m) = 1 - 1e^{-7}$. Similarly, if $\beta_{s2}^r(m) < 1e^{-7}$, set $\beta_{s2}^r(m) = 1e^{-7}$. Re-scale so that $\sum_{s=1}^{S} \beta_s^r(m) = 1.0$.

    (vii) Accept $\beta^r(m)$ with probability min(α, 1.0), where:

$$\alpha = \frac{Pr(l^r | l^1, \dots, l^S, C, \beta^r(m)) Pr(\beta^r(m) | \lambda(m-1))}{Pr(l^r | l^1, \dots, l^S, C, \beta^r(m-1)) Pr(\beta^r(m-1) | \lambda(m-1))}.$$

B. Update $\lambda(m)$ using a M-H step:

    (i) Propose a new $\lambda(m)$ from a Normal $(\lambda(m-1),$ sd $= 0.2)$

    (ii) Automatically reject if $\lambda(m) \notin [0,1]$

    (iii) Otherwise accept $\lambda(m)$ with probability min(α, 1.0), where:

$$\alpha = \frac{Pr(\beta^r(m) | \lambda(m))}{Pr(\beta^r(m) | \lambda(m-1))}.$$

For large *M*, this algorithm is guaranteed to converge to the true posterior distribution of the β^r's (e.g. (*119*)). In practice, for all results presented here we use M=20,000, sampling every 1,000[th] iteration after an initial "burn-in" of 10,000 iterations. Also, for each recipient individual (or group) *r*, we combined results across ten independent runs of the above procedure, weighting our estimates and standard errors by the posterior probabilities of the samples (using 190 total samples). We refer to the final estimates of $\{\beta_1^r, \dots, \beta_S^r\}$, weighted by posterior values, as our inferred proportions of ancestry for group *r* conditional on this set of *S* surrogates. This procedure differs from the mixture model procedure described in (*4, 68, 108, 115*) in that it assumes that l^r is multinomial distributed and solves for β^r using a Bayesian approach rather than a non-negative least squares optimization.

We perform the following four mixture model analyses (though here "modern" groups exclude ALT, DEN), which differ in the set of surrogates used:

    (I)   "all moderns" – form each ancient and modern genome using all modern groups as surrogates

(II) "all moderns + ancients" – form each ancient and modern genome using all modern+ancient groups as surrogates

(II-) "moderns - excluding neighbours + ancients" – form each ancient and modern genome using ancient groups and modern groups excluding those populations from within the same country as surrogates. For the Lazaridis merge this was performed separately on populations from Iran (Iranian, Iran_Zoroastrian, Iran_Fars), India (Kharia, Lodhi, Mala, Tiwari, Vishwabrahmin, GujaratiA_GIH, GujaratiB_GIH, GujaratiC_GIH, GujaratiD_GIH), Pakistan (Balochi, Brahui, Burusho, Hazara, Kalash, Makrani, Pathan, Punjabi_Lahore_PJL, Sindhi) and Georgia (Abkhasian, Georgian_Megrels). For the Busby merge this was performed separately on populations from Armenia (Armenian, Chambarak, Dprabak, Gavar, LebArmenian, Martuni, Yegvard, Yerevan), India (Indian, Indianjew, Tamilnadu, Bengali, Bhunjia, Brahmin, Chamar, Chenchu, Dharkar, Dhurwa, Dusadh, Gond, Hakkipikki, Kanjar, Kol, Kshatriya, Kurmi, Kurumba, Lambadi, Mawasi, Meena, Meghawal, Muslim, Naga, Nihali, Piramalaikallar, Tharus, Upcaste, Velamas, Malayan) and Pakistan (Balochi, Brahui, Burusho, Hazara, Kalash, Makrani, Pathan, Sindhi).

(III) "ancients + Yoruba + Han" – form each ancient and modern genome using all other ancient genomes, plus the modern Yoruba and Han as surrogates.

In each case, a group cannot use itself as a surrogate or else it would match itself exactly. For analysis (I) we were interested in how modern and ancient groups relate ancestrally to different modern groups. For analyses (II), (II-) and (III), we were interested in how modern and ancient groups relate ancestrally to different sets of modern and/or ancient groups. For analyses (I), (II) and (II-) we use a slightly alternative version similar that used in (*108*). In particular when inferring coefficients for each group r using the MCMC algorithm defined above, we set $l_r^r = 0$, $l_r^s = 0$ for all $s \in [1, ..., S]$ to mitigate any effects of self-copying in group r. Therefore, self-copying is only considered under analysis (III). Also for analysis (III), we included the modern Yoruba and Han groups as surrogates, since our ancient samples contain no good proxies for sub-Saharan Africa and East Asia, each of which may have contributed recent ancestry (e.g. through admixture) to our recipient groups. Finally, for all analyses we excluded Bar8 as a surrogate, as this sample had the lowest coverage (7.21) out of all of our aDNA samples (median coverage range of remaining aDNA samples was 19, with a range of 10.4-42).

Results

Inferred proportions of ancestry for aDNA samples

Our inferred proportions of ancestry are provided for all Neolithic ancient samples (i.e. WC1, KK1, LBK, NE1, LOS, Mota, Ust'-Ishim) for each of analyses (I) (Table S22) and (II) for the Lazaridis and Busby merges in Fig. S24-S25 and for analysis (III) in Fig. S28-S29. Estimates for all analyses (I)-(III) along with standard errors, are provided in Tables S24 and S25 for Lazaridis and Busby merges respectively.

Inferred proportions of ancestry for modern samples

Our inferred proportions of ancestry are provided for the modern groups for analyses (II) (Fig. S26), (II-) (Fig. S27) and (III) (Fig. S28-S29) (Table S23). Estimates for all modern groups

for all analyses ((I),(II),(III)), along with standard errors, are provided in Tables S24 and S25 for Lazaridis and Busby merges respectively.

Differences in inferred painting profiles for aDNA samples

To quantify differences in inferred painting profiles, we calculated TVD for each pairing of the aDNA samples Ust'-Ishim, KK1, Loschbour, WC1, Bar8, NE1, LBK and Mota under analysis (I). The matrix of TVD differences is represented as a symmetrical heatmap for both the Busby and Lazaridis merges (Fig. S30), and analogously for differences in inferred allele frequency profiles in Fig. S38e. Differences in each component (i.e. donor group) of the inferred painting profiles for pairwise combinations of the ancient samples (WC1, Bar8, LBK, NE1) are displayed graphically in Fig. S32-S37, with analogous differences in inferred allele frequency profile displayed in Fig. S38a-d. The key for the population labels for these plots is provided in Tables S24 and S25 for Lazaridis and Busby merges, respectively. We also report the TVD of pairwise differences when aDNA samples (WC1, Bar8, LBK, NE1) are painted in relation to African donor individuals with no reported evidence of recent West Eurasian admixture, either not including (Fig. S31ab, S38f) or including (Fig. S31cd) an additional Sardinian donor individual.

Patterns of ancestry among ancient and modern groups

In general, results are very consistent across the Lazaridis and Busby merges, despite the uses of different (though overlapping) individuals (e.g. Fig. S24-S25). Expanding upon points made in the main text, we note the following observations about our analysis of the ancient samples Bar8, WC1, KK1, LBK, Loschbour, NE1, Mota, Ust'-Ishim.

1. When inferring proportions of ancestry for any given ancient sample r using all modern samples as surrogates (analysis (I)), only modern groups geographically near to where r was sampled typically give $\beta_s^r > 0$, with the vast majority of the >160 modern groups s not contributing at all across both dataset merges (Fig. S24-S25). This encouragingly suggests that our aDNA calls appear relatively stable. A good example is the Caucasus hunter-gatherer KK1, which receives large contributions from modern groups in Georgia, Armenia and neighbouring groups with only small contributions from elsewhere (Fig. S24).

2. As another example, under analysis (I) the new Iranian Neolithic genome WC1 looks genetically most similar to groups from modern-day Iran (Iran_Fars, Iranian), Pakistan (Pathan, Sindhi, Makrani, Balochi), Armenia (Armenian_LebArmenian) and India (Meghawal) (Fig. S24). This is in strong contrast to the Neolithic genomes from Anatolia (Bar8), Germany (LBK) and Hungary (NE1) (Fig. S24, S32-S37) who receive the highest contributions from modern-day groups in southwest Europe as reported previously (*4, 16, 82*), notably Sardinians, Tuscans and other Spanish, Italian and Greek populations (Table S22).

3. As expected, the non-Iranian Neolithic samples LBK and NE1 contribute to each other under analysis (II) and also contribute substantially to the Neolithic Anatolian sample Bar8 (Fig. S24).

4. When comparing ancient groups to only aDNA samples plus modern day Yoruba and Han (i.e. analysis (III)), the Iranian Neolithic WC1 matches substantially more to the Caucasus hunter-gatherer KK1 than to any other ancient sample or the modern Yoruba and Han, with an

additional contribution from the 45 kya Siberian Ust'-Ishim. In contrast, as expected the non-Iranian Neolithic samples LBK and NE1 contribute substantially to each other (Fig. S28-S29).

5. Under analysis (III), the Caucasus hunter-gatherer KK1 matches substantially to WC1, mirroring the latter's large contribution from KK1, with additional contributions from NE1 and LBK. The other hunter-gatherer genome Loschbour from Luxembourg matches almost entirely to NE1 under this analysis, likely a reflection of the lack of any good surrogates to this sample's ancestors (Fig. S28-S29).

6. As expected, under analysis (III) the 4,500-year-old Ethiopian Mota is most similar to modern-day Yoruba in this analysis, with smaller contributions from Ust'-Ishim and WC1, which perhaps reflects the lack of any close ancient surrogates to Mota in this analysis. The 45 kya Siberian genome Ust'-Ishim is most similar to WC1 in this analysis, with additional contributions from geographically and temporally disparate groups: the modern-day Han, Loschbour and Mota, which also likely reflects the lack of close ancient surrogates for Ust'-Ishim in this analysis (Fig. S28-S29).

7. When comparing how inferred painting profiles of the aDNA samples differ in their patterns of haplotype sharing with modern groups, we show that there are greater differences between pairwise comparisons of WC1 and any other Neolithic genome than amongst any pairwise comparisons of the European and Anatolian Neolithic genomes LBK, NE1 and Bar8 (e.g. Fig. S30, Fig. S32-S34 compared to Fig. S35-S37). Consistent with Fig. S24, the majority of these differences correspond to WC1 relating more to modern groups from Pakistan, India, Iran and Armenia compared to LBK, NE1 and Bar8 and less to modern populations from Europe (most notably those from Greece, Spain and Sardinia). While haplotype-based analyses can be influenced by phasing errors, these observations are consistent if we use an alternative approach that compares allele frequencies of individual SNPs among our Neolithic and modern samples while also matching for sequencing depth (Fig. S38).

8. When comparing haplotype (or allele frequency) sharing patterns to only African groups that show no clear evidence of recent admixture from West Eurasians, the Neolithic aDNA samples (WC1, LBK, NE1, Bar8) look genetically similar to each other (Fig. S31 and Fig. S38f), with only subtle differences between samples. For example, when instead comparing each aDNA sample to the same African individuals while adding only a single Sardinian individual, substantial differences between WC1 and the European/Anatolian aDNA samples become apparent (Fig. S31cd). These observations are consistent with a scenario where differences between WC1 and the European/Anatolian Neolithics (e.g. as seen when comparing to modern African and non-African groups in Fig. S30-S31) arose after a single dispersal out-of-Africa that gave rise to the ancestors of these Neolithic samples.

For our analyses of the modern groups, we note the following observations:

1. As expected, most modern groups receive little or no contributions from aDNA samples when including both modern populations and ancient samples as surrogates under analysis (II) (Fig. S26). A notable exception to this is a sizeable contribution from LBK and NE1 to Sardinians. LBK and NE1 also make smaller contributions to Italian, Spanish and Greek populations. Amongst others Armenians, Georgians, Adygei, Abkhazians, Tajik, Balkar and Kumyk also receive small contributions from KK1 sampled nearby. Furthermore, the new Iranian Neolithic sample WC1 makes small contributions to modern groups sampled from Pakistan (in particular the Pathan and Kalash) and Iran (Iranian and Iran_Fars) and also to Turkmen, Tajik and Yemen (Table S23).

2. Excluding neighbouring modern groups as surrogates (analysis (II-)) typically increases the contribution from ancient samples, demonstrating a component that is otherwise masked by geographically proximal modern surrogates (Fig. S27). This most notably increases WC1 contributions to substantial levels in the Kalash, Balochi, Makrani, Pathan, Brahui and Sindhi from modern-day Pakistan, modern Iranians, Iran_Fars and Iran Zoroastrians and several Armenian groups: Chambarak, Dprabak, Gavar, Martuni, Yegvard and Yerevan (Table **S23**). These Armenian groups each also receive contributions from KK1.

3. When comparing modern groups to only aDNA samples, plus modern day Yoruba and Han (i.e. analysis (III)), the patterns in inferred proportions of ancestry vary substantially, often consistent with geography. For example, modern Caucasus groups from Georgia and Armenia match genetically to the Caucasus hunter-gatherer KK1, mirroring how KK1 matches genetically to these groups as described above (Fig. S28-S29). A notable exception is the Kalmyk who match almost entirely to modern day Han, likely a reflection of their East Asian origins (120) and highlighting how results should be interpreted in the context of included surrogate groups.

4. Similarly, under analysis (III) the modern populations that receive the highest contributions from the Iranian Neolithic sample WC1 were sampled from nearby countries (Iran, Iraq, Syria, India and Pakistan - notably the Brahui, Balochi, Sindhi and Makrani) (Fig. S28-S29). Among these modern groups, the majority from India have notable additional contributions from the 45 kya Siberian genome Ust'-Ishim, while groups from Iran (Iranian, Iranian_Fars, Iran_Zoroastrian) instead have additional contributions from LBK and to a lesser extent NE1 and Mota.

5. In Western Europe, under analysis (III) groups are more genetically similar to the Neolithic sample from Germany LBK, the Neolithic hunter-gatherer from Luxembourg Loschbour and the Neolithic hunter-gatherer from Georgia KK1 (Fig. S28-S29).

6. Under analysis (III), as expected, groups sampled from Africa primarily match the Ethiopian aDNA sample Mota and/or the modern Yoruba, with smaller contributions from NE1, LBK and WC1 in east African groups, the latter perhaps reflecting recent admixture (121) (Fig. S28-S29).

7. Not surprisingly, under analysis (III) many of the groups sampled from central and East Asia match to the modern-day Han, though with contributions from Loschbour in some groups (e.g. Ket, Koryak, Selkup) (Fig. S29).

8. Strikingly, under analysis (III) Papuans, Melanesians and the Onge receive high contributions from Ust'-Ishim, in stark contrast to other modern groups e.g. from south and east Asia and perhaps reflecting these groups' complicated origins (*122-124*). Several native groups from South America (Colombian, Karitiana, Pima, Surui) also receive notable contributions from Ust'-Ishim. However, many of these groups appear genetically differentiated from the prediction based on the inference from analysis (III) (i.e. their levels of "self-copying" are high as depicted by smaller pies in Fig. S29), perhaps suggesting either a substantial amount of population-specific drift in these groups and/or that none of the ancient genomes and modern groups included as surrogates in analysis (III) reflect well the DNA of these groups.

S10. Population continuity

David Díez-del-Molino & Mark G. Thomas

A continuous population with no admixture or replacement over an extended period of time is an unlikely model for any human population. However, it does serve as a useful null-hypothesis when investigating the relationships between ancient and modern genomes. In this context, we used the population continuity method described in (*4*) to explore the relationships between WC1 and a panel of selected modern populations. Because WC1 has a higher coverage (~10x) than the ancient genomes analyzed in (*4*) and we were thus able to use more high-confidence biallelic calls, we implemented a modification of the population continuity test that allowed for comparison between both ancient and modern diploid genotypes.

For each combination of an ancient genome and a sample of modern genomes we performed the steps described in (*4*) with minor modifications. Briefly, we considered only positions that had biallelic calls of quality > 30 and were contained in all modern genomes from dataset (*82*) and the Zoroastrian genomes (Table S26). We estimated allele frequencies and employed them in forward drift simulations as described in (*4*), using the mean calibrated age of the genome to estimate the number of generations (Table S26). We explored two parameters, ancient (N_ea) and modern (N_em) effective population sizes, assuming exponential growth between them. For each simulation we sampled a diploid genome from the initial frequency vector and another one from each simulated final frequency vector. To compare the simulated genomes sampled that way with the observed ancient and modern genomes, we extended the six allelic sharing classes described in (*4*) to nine in order accommodate all combinations of biallelic genotypes in each position (Table S27). An overall population continuity rejection p-value for each combination of parameters was then calculated combining the individual p-values for each allelic sharing fraction using Fisher's and Voight's methods as described in (*4*).

<u>Results</u>

We explored the plausible parameter space of N_ea and N_em for each ancient genome and modern population sample by performing the test on a 30x30 grid of assumed effective population sizes, ranging from 10 to 10^7 on a log scale. We performed 1,000 simulations for each of the 900 combinations of parameter values. The ranges of effective population sizes in which continuity could not be rejected were examined by slicing the grid at 1, 5, 10 and 20% of the modern population sizes (Table S28) and the p-values of the test for each ancient effective size were reported.

As suggested by other analyses (See section 9), WC1 may be more closely related to modern Iranian, Pakistani and Afghan populations than to other modern Eurasians. In order to examine the extent to which some of these populations can be considered continuous all the way back to the population from which WC1 was sampled, we applied the test between WC1 and Balochi, Brahui, Makrani and Iranians. Grid results indicate that continuity between the population WC1 was sampled from and these four modern populations could be rejected unless unrealistically small ancient population sizes were assumed (ranges at 10% of modern population size: 108-452, 108-728, 281-4894, 67-452, respectively; Fig. S39 and S40). For modern populations with small population sizes, such as the Makrani, fixing the effective population size to a small percentage of the modern size resulted in failure to reject population continuity for a wide range of ancient sizes. However, these ranges are not consistent with other plausible modern effective sizes and therefore they should not be considered as evidence of population continuity.

The results of the test between WC1 and Iranian Zoroastrians indicate that, for all assumed population size combinations, WC1 was not sampled from an ancient population continuous with them (Fig. S41 and S42). Likewise, by comparing WC1 and modern Caucasian (Megrelians and Armenians) and Anatolians (Turkish Balikesir, from Western Anatolia, and Turkish Adana, for the Southeast), we found that population continuity could be rejected unless the assumed ancient population sizes were very small (ranges at 10% of modern population, 67-1887, 67-728, 42-281, 42-1172, respectively; Fig. S43-46). However, we note that for Turkish Adana, Megrelians and to a lesser extent, Armenians, population continuity could not be rejected under a somewhat wider range of assumed ancient population sizes, possibly indicating a closer relationship to those populations than to Turkish Balikesir in Western Anatolia.

Because Sardinians are considered to be the living population most closely related to the early European farmers (*17, 82, 125*) we used them as surrogates to test the relationship between WC1 and early Neolithic farmers in Europe. Our results show no evidence of continuity between WC1 and modern Sardinians for any combination of assumed ancient and modern population sizes, consistent with WC1 not representing a population directly ancestral to early European farmers (Fig. S43 and S44).

S11. Functional SNPS

Karola Kirsanow

Methods

We assessed the five Iranian individuals for whom nuclear genomic data was available (WC1, AH1, AH2, AH4, and F38) at a panel of single nucleotide polymorphisms (SNPs) having known functional associations in modern populations, focusing particularly on sites for which evidence of recent natural selection has been found in modern or ancient human populations. Diploid genotypes were assessed using the SNP-calling method described S.6, and further validated through direct observation of BAM files using samtools *tview*. Only the WC1 individual had sufficient coverage to produce diploid genotypes at most positions; data from the other four individuals is nevertheless provided to note the presence of derived alleles at certain loci of interest.

Markers associated with pigmentation phenotype

The Iranian individuals carry the ancestral (darker pigmentation allele) at most of the sites in the Hirisplex complex, indicating that they are likely to have had darker eyes and hair (*126*) (Table S29). Reconstruction of the WC1 individual using the Hirisplex tool returns the highest probabilities for brown eyes (95%), dark skin (97%), and black hair (74%). Although the ancient Iranians display derived alleles at rs6119471 in *ASIP,* which are nearly fixed in Europe, East Asia, and South Asia, examination of the eye and skin pigmentation loci in the 8-plex system (*127*) suggests that WC1 is unlikely to have had depigmented skin and irises. Interestingly, derived alleles can be observed in 4 of the 5 Iranian individuals at rs1426654 in *SLC24A5*, a strongly selected site (*128*) associated with skin depigmentation (*129*) (rs1426654 is not included in the Hirisplex assay). WC1 is heterozygous at rs12913832 in *HERC2*, the causal mutation for iris depigmentation, and at rs1129038, the SNP with the highest linkage with rs12913832 in modern populations (*130*). Examination of the *HERC2* (*131*) and *SLC24A5* haplotypes (*132*) in WC1 further support the inference that this individual carried at least one copy of the derived allele at these focal loci (Table S29).

Derived *SLC24A5* rs1426654 alleles can also be observed in the AH1, AH4, and F38 individuals, although coverage at this site is too low to make a conclusive genotype determination. The AH1 and F38 individuals also carried derived alleles at the HERC2 rs1129038 locus in beh2, and at rs4778138 in beh1, suggesting that the HERC2 mutation may have been segregating in these populations. The rs1426654 and rs12913832 observations represent early occurrences of these selected depigmentation mutations outside of central Europe.

The WC1 individual was homozygous for the ancestral allele at rs16891982 in *SLC45A2*, another strongly-selected pigmentation-related locus, producing the second-greatest effect on skin pigmentation differences observed between Europeans and African-admixed populations (after *SLC24A5* rs1426654) (*128, 133, 134*). Allelic state at this locus could not be assessed in the other Iranian individuals.

Markers associated with metabolic phenotype

We genotyped the 5 ancient Iranian individuals at 8 sites in the *NAT2* region associated with acetylation phenotype (slow>intermediate>rapid acetylation of xenobiotics), a metabolic

characteristic associated with the risk of developing certain cancers and the efficacy of pharmaceuticals (*135*) (Table S30). The *NAT2* gene region may have been the target of selective pressure associated with the adoption of novel dietary lifeways in prehistory (*136-139*).

We inferred the most probable acetylation phenotype, where possible, using three models developed for modern humans: a 7-SNP method (*140*), a 2-SNP subset of the 7-SNP method (*135*) and an independent tag-SNP method (*141*). All three methods agree that WC1 is most likely to have been a rapid metabolizer, which is the ancestral phenotype. The 2-SNP and tag-SNP methods indicate that AH1 is most likely to have been a slow metabolizer. The acetylation phenotypes of the AH2, AH4, and F38 individuals could not be determined.

We also genotyped the ancient Iranians at a panel of loci within the *CYP3A4* and *CYP3A5* gene regions (Table S30). These regions are associated with the metabolization of a broad spectrum of hormones and pharmaceutical compounds, have highly variable expression in modern humans, and show evidence for natural selection (*142-145*). The *CYP3A5* region is believed to have undergone positive selection, possibly related to environmental factors such as aridity through its enzyme's role in salt retention and water reabsorption (*142*). The posited *CYP3A5* selective gradient favors the ancestral expresser phenotype in equatorial regions, and low/non-activity alleles at higher latitudes, where expresser phenotypes are adversely affected by salt-sensitivity hypertension (*146, 147*) WC1 is homozygous for the derived allele at rs776746, which defines the *CYP3A5*3* haplotype, as well at the closely-linked marker rs15524, indicating that WC1 is likely to have had reduced *CYP3A5* expression. WC1 is also homozygous for the derived allele at rs2740574 in *CYP3A4*; derived alleles can also be observed in the AH1, AH2, and F38 individuals at this locus. The *CYP3A4*1* ancestral allele is selected against in modern non-Africans, possibly related to its role in vitamin D metabolism (*143*).

We additionally assessed the ancient Iranians at a panel of sites in the *AGT*, *ADRB2*, *ENaCα*, *ENaCγ* and *GNB3* gene regions which also appear to display latitudinal gradients in allele frequency similar to those observed for the *CYP3A5* mutations (Table S30). Similar to *CYP3A5*, it has been proposed that these regions have experienced a positive selection gradient related to heat adaptation, salt retention and susceptibility to hypertension (*148, 149*). The WC1 individual is heterozygous for 7/9 of these positions, except for rs4762 and rs5049 in *AGT*, where WC1 is homozygous for the ancestral hypertension-non risk allele.

The selective events affecting the *CYP3A4*, *CYP3A5*, and *AGT* are believed to predate the onset of strong selective pressure on *NAT2*: selection on *CYP3A4*, *CYP3A5*, and *AGT* may be correlated with the expansion out of Africa and colonization of higher latitudes, while selection on *NAT2* may have occurred later, possibly in response to pressures stemming from the adoption of an agriculturalist diet (*136-139, 142, 143, 148*).

We genotyped the ancient individuals at several loci involved in the metabolism of lactose and alcohol, dietary inputs which are believed to have become more important with the Neolithic dietary transition (Table S30). No derived (alcoholism-preventative) alleles were observed in the Iranian dataset. We did, however, observe derived lactase persistence - associated alleles in 3 of the 5 individuals: WC1 is heterozygous at rs4988235 (C/T(-13910)) and AH1 and AH4 display derived alleles at rs182549. Because this is an unexpected result, we examined these positions in greater detail using samtools tview. Analysis of the specific read groups from which these reads derived along with their base quality and mapping scores did not show signs of contamination. However, the small number of derived reads

present at each *LCT* site, particularly the ratio of derived/ancestral reads (1:10) in WC1, leads us to suspect they may be erroneous.

The Neolithic individual WC1 may have carried a copy of a derived haplotype in the *TCF7L2* gene region (defined by rs7903146 C and rs10885406 A and tagged by rs7924080 T) selected in several different worldwide populations, which is protective against type 2 diabetes (Table S31.1) (*150*). In Europeans, the selective sweep affecting this haplotype is estimated to have occurred c. 11.9 kya, predating this sample. This protective haplotype may have also occurred in at least two of the Neolithic Aegeans described in (*4*).

Finally, we genotyped the ancient Iranians at a panel of 62 SNPs comprising a genotype risk score model used to predict type 2 diabetes (T2D) risk in modern populations (*151*) (Table S31.2). The predictive power of this model is not sufficient to infer diabetes risk in ancient individuals; instead we use it to characterize variation across a number of markers involved in a complex phenotype. In a comparison of the WC1 Neolithic Iranian, the Bar8 Neolithic Aegean from (*4*) and the Loschbour and Stuttgart individuals from Lazaridis *et al.* (*82*), the WC1 individual has the highest number of T2D-susceptibility alleles, and the highest weighted genotype risk score (no. susceptibility alleles × odds ratio for each locus). In comparison, the average weighted genotype risk score for modern non-diabetic individuals of European ancestry was 66.3+/-5.1 in one study cohort, and 66.7+/-5.2 in a second cohort (*151*). Modern diabetics averaged 68.4+/-4.9 in the first cohort and 68.7+/-5.2 in the second, illustrating the marginal predictive power of this assay. Nevertheless, all of the ancient individuals are at the outside of the range of modern non-diabetic individuals.

<u>Markers associated with susceptibility to infectious and non-infectious diseases</u>

We assessed the ancient Iranians at a panel of SNPs related to pathogen resistance and susceptibility to infectious and inflammatory diseases (Table S32). Malaria has been historically endemic in the broader geographic region from which these samples originate, and derived malaria-protective alleles imparting susceptibility to various forms of thalassemia segregate at appreciable frequency in the region (*152, 153*). We genotyped the ancient Iranians at several malaria-protective loci in the *HBB* gene region for which derived alleles are at appreciable frequency in modern Iran (*152*). We did not observe any derived alleles in our ancient sample.

A suite of markers related to infectious and inflammatory disease susceptibility has been identified as the target of relatively recent (post-Neolithic) selection, possibly related to novel dietary inputs and increased residential density (*154*). In this putatively selected suite, the ancient Iranians display derived alleles at a number of positions, including 2/8 sites (in *IL7R* and *STAT3*) identified as belonging to a protein-protein interaction network which underwent coordinated selection in Europeans 2.6-1.2 kya (*154*). This, together with the observation of derived alleles at several additional loci for which selective-sweep estimates postdate the ages of these samples (Table S32).

We additionally assessed the *SLC22A4/SLC22A5* IBD5 haplotype associated with Crohn's disease, which is estimated to have undergone a selective sweep in Europeans c. 12.5 kya (*155, 156*), It has been proposed that the derived haplotype in this region swept to high frequency in Europeans in relation to the novel metabolic requirements of an agriculturalist diet (*155*). The derived deleterious alleles were not observed in our sample.

<u>Additional putatively selected markers</u>

The Neolithic Iranians displayed derived alleles at 7/12 sites identified as selected in ancient Eurasians by Mathieson *et al.* 2015 ((*3*), Table S33). Derived alleles could be observed in Neolithic individuals at two of three sites associated with depigmentation (*HERC2, GRM5* but not S*LC45A2*), in addition to *SLC24A5* (mentioned above), suggesting that selection favoring depigmentation was significant in the early Neolithic populations represented by this sample. Interestingly, derived alleles of S*LC45A2* rs16891982 can be observed in 4/5 of the Neolithic Aegeans described in (*4*), supporting the inference that this site was also under selection in the Neolithic farmers ancestral to European early farmers. Derived alleles could be observed at 7/12 putatively selected sites in both the Neolithic Iranian and Neolithic Aegean samples. No derived alleles were observed for SNPs in the *TLR1/TLR6/TLR10*, *ATXN/SHB3*, *LCT* (but see above), or *ZKSCAN* gene regions in either sample.

S12. G-PhoCS analysis of Neolithic and Hunter-gatherer genomes

Krishna Veeramah

In order to estimate times of divergence between Anatolian/European Neolithic, European Hunter-Gatherer and Iranian Neolithic populations, we used G-PhoCS (*64*), a Markov Chain Monte Carlo Bayesian method that infers population demographic parameters such as divergence times in terms of branch lengths (τ) in units expected number of mutations, and effective population size in terms of the population genetic parameter θ. In addition, migration bands can be included in the inference but we did not utilize this feature for this analysis.

Loci and Samples

Gronau *et al*. (*64*) previously identified a set of 37,574 1-kb 'neutral loci' orientated to human reference genome build 36. To apply these loci to the ancient genomes mapped in this study we performed a LiftOver to build 37. Four Loci were lost during the liftover. We then identified nine paleogenomic genomes with a mean coverage >7x and extracted the neutral loci using the diploid genotypes calls described in section S6. When extracting genotype calls we only considered those with read depth >=7x, otherwise sites were masked. In addition, for any heterozygote genotype call with genotype quality less than 30 we utilized the next most probably homozygote genotype call. We also included a high coverage Yoruba individual, NA19238, sequenced as part of the 1000 Genomes Project (*103*). All sites containing at least one CpG dinucleotide were masked.

Neighbor-Joining Tree Analysis

In order to guide the appropriate topology for G-PhoCS, we used the ~30,000 loci to calculate a matrix of sample pairwise sequence divergence amongst the ten samples using the distance estimator described in Freedman *et al*. (*157*) and constructed a Neighbor-Joining (NJ) tree using MEGA 6.06 (*158*). We then visualized this tree using FigTree v1.40 (Fig. S47).

As expected, the Yoruban sample is the most external branch. Consistent with various *f* statistics analyses in Section S8, Anatolian/European Neolithic farmers (blue branches) and European hunter-gatherers (red branches) form a sister clade and the Neolithic Iranian sample WC1 forms a sister clade with the Caucuses hunter-gatherer sample Kotias (Green branches). Ust'-Ishim, a 45 kya hunter-gatherer from Siberia, is external to both these pairs of sister clades (though this may be a result of admixture in the WC1/Kotias clade, see ADMIXTUREGRAPH results).

G-PhoCS setup

Given the above NJ tree, we chose to consider the population topology shown in Fig. S48 for our G-PhoCS analyses along with the following θ and τ demographic parameters.

We primarily use Bar8 to represent Neolithic Anatolians/Europeans in our analyses as this should be the least admixed and best representative of this population based on (*4*) but as this is the lowest coverage genome we also explore the impact of using the higher coverage NE1. Loschbour was chosen as it was the highest coverage European hunter-gatherer sample, and as Caucasus hunter-gatherers may have experienced admixture with European hunter-gatherers we excluded this sample. Note that our topology is different to Jones *et al*. (*8*), which grouped Neolithic farmers and Caucuses hunter-gatherers with European hunter-gatherers as an external

branch. Though not consistent with our f statistics and NJ analysis, we did this test the alternative topology.

The newest version of G-PhoCS allows the inclusion of ancient samples for which their age from the present can be preset. However, this must be done in units of expected numbers of mutations. Based on new pedigree-based mutation rates from whole genome sequencing of trios and quartets (*12*), we assumed a mutation rate, μ_y, of 0.5 x 10^{-9} mutations per site per year. We multiplied this value by 0.9 to take into account the removal of CpG sites (*159*) and set τ_A, τ_B, and τ_C, based on the mean ^{14}C age of each ancient sample. We ran G-PhoCS either keeping these τ values fixed or allowing them to be estimated by the program.

We utilized a gamma distribution for all parameters. For all θ priors we used $\alpha=1$ and $\beta=5,000$. $\alpha=1$ was also used for all τ parameters. However, we varied β for τ_{AC}, τ_{ACB}, and τ_{ACBD} as 80,000, 40,000, 20,000 respectively, which assuming the same μ_y as above, is equivalent to a mean of ~25 kya, ~50 kya and ~100 kya respectively. To ensure that our results for τ_{AC}, τ_{ACB} were not over-constrained by these priors, we also set $\beta= 10,000$ (~222 kya) for both of them, which is an unfeasibly large number.

Mutation rate was set as fixed across loci in the results reported below, but we note that allowing this to vary had little effect. We also ran G-PhoCS with and without the Yoruban sample. G-PhoCS was always run for 1 million iterations of the Markov chain, which was sufficient for convergence of the τ parameters. Burn-in for each run was determined by inspecting the MCMC trace using the software Tracer v1.5, which was also used to process the results. We used G-PhoCS's auto-tune feature to initially determine the MCMC update steps, and, likely because our model is fairly simple with regard to parameters (3 or 4 modern populations and no migration), we did not find it necessary to adjust these manually for any run (i.e. we observed good acceptance ratios, rapid convergence and good exploration of the parameter space). An example trace showing typical mixing and convergence for the τ parameters is shown in Fig. S49.

Results

As G-PhoCS is computationally intensive, we report here results using only the first 5,000 loci from the total set, as we tested many different combinations of parameters, except for the results in Table S35, where all loci were used. The results were very similar using the subset or full set of loci for this iteration. Tables S34 and S35 show parameter estimates when including and excluding the Yoruban individual and the age of the ancient samples was fixed. The results when ages of ancient samples were inferred were almost identical and thus are not reported here. Assuming a μ_y of 0.5 x 10^{-9}, times of divergence are slightly reduced when excluding the Yoruban individual, and across both analyses the mean split time for the Anatolian/European farmers and hunter-gatherers ranges from 33-39 kya (Combined 95% CI 16-61), and for the Neolithic Iranians 46-77 kya (Combined 95% CI 38-104). The European hunter-gatherers, as represented by Loschbour, were inferred to have a markedly reduced N_e compared to Neolithic farmers (~10%), with WC1 having a more comparable level of diversity to Bar8. When NE1 is used rather than Bar8, divergence times are somewhat reduced (Tables S36) but the divergence time of WC1 is still ~50% greater than that between Loschbour and NE1.

To ensure that our divergence times were not overly constrained by the priors, we also performed a G-PhoCS analysis (excluding the Yoruban individual) where the mean for the τ_{AC} and τ_{AC} priors was equivalent to ~200 kya, a highly unlikely value given that anatomically modern humans are believed to have only emerged at this time. Despite the Markov chain

starting off close to this value for both τ estimates, the chain quickly converged to similar values as described in Table S35 (Fig. S50).

Finally, we also attempted an alternative phylogeny with the Iranian Neolithic and Anatolian/European Neolithic samples forming a sister clade and with Loschbour being an external branch. However, convergence for the two τ parameters was very poor (Fig. S51), with the first proposed divergence event appearing to try and reach values that are equivalent to the older event, suggesting that the topology of the population tree was incorrect.

ADMIXTUREGRAPH analysis of WC1

Fu *et al*. 2016 (*14*) recently reported the analysis of 51 ancient Eurasian genomes dating from ~45,000–7,000 years ago. In this study they produced a base model of Eurasian population demography without admixture from 4 high coverage genomes (Ust-'Ishim, Malta1, Kostenki14, and GoyetQ116-1) and a Mbuti outgroup using ADMIXTUREGRAPH (qpgraph in the Admixtools package)(*6*), which performs an approximate likelihood maximization to best fit *f2, f3, f4* statistics for all combinations of populations.

We merged the original ancient data from Fu *et al*. (2,144,502 SNPs in 51 individuals) with genotype data at the same SNPs from WC1 as well as Mbuti sequenced as part of the Simons Genome Diversity Project (*104*) using the convertf and mergit tools in in the Admixtools package.

We then attempted to examine WC1 within the context of Fu *et al*.'s base model by either iteratively placing it as sister to each node of the tree (including the root) (i.e. unadmixed) or as a combination of any two nodes (i.e. admixed). 288,557 SNPs were usable by ADMIXTUREGRAPH when considering the call rate across all samples in the analysis. *Z*-scores were generated for the fit for each *f* statistics using a weighted block jackknife with a block size of 5cM.

The only graph topology that did not produce *f* statistics outliers (fitted value within three standard errors of estimated value) (Table S37) involved WC1 being derived from an admixture event between a basal Eurasian population (62%) and Ancient North Eurasians (38%) (Fig. S52). The largest |Z| score for *f4* was 2.945. We note that the estimates of drift along the branch lengths for the base model are very similar to those originally estimated by Fu *et al*. (*14*).

SUPPLEMENTARY FIGURES
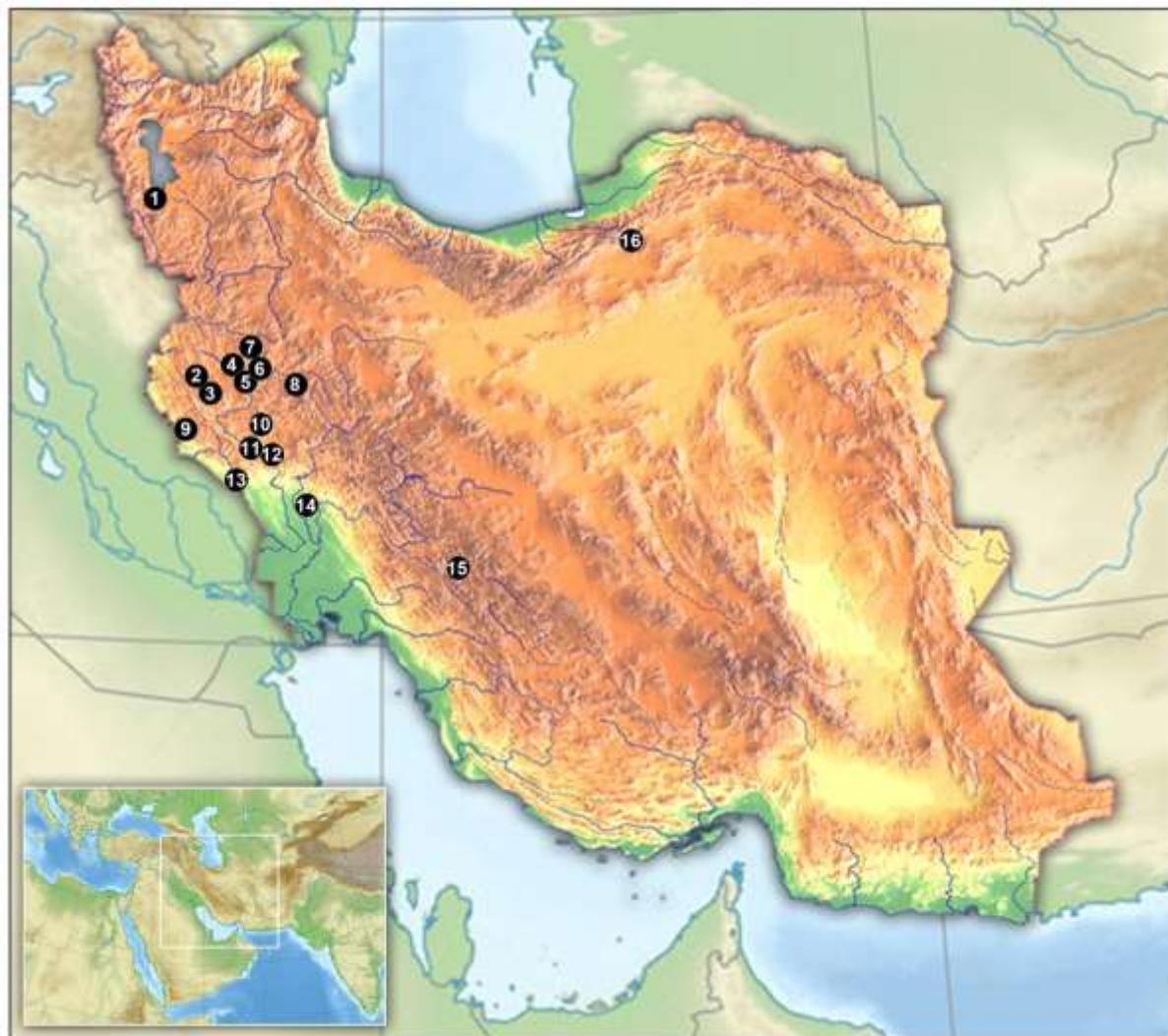
Fig. S1

Map of Iran with the location of the quoted sites. 1. Hasanlu, 2. Wezmeh cave, 3. Jani,
4. Sarab, 5. Asiab, 6. Ganj Dareh, 7. Sheikhi Abad, 8. Tepe Abdul Hosein, 9. Chogha Golan,
10. Guran, 11. East Chia Sabz, 12. Kelek Asad Morad, 13. Chogha Sefid, 14. Chogha Bonut,
15. Qaleh Rostam, 16. Tappeh Sang e Chakhmaq.
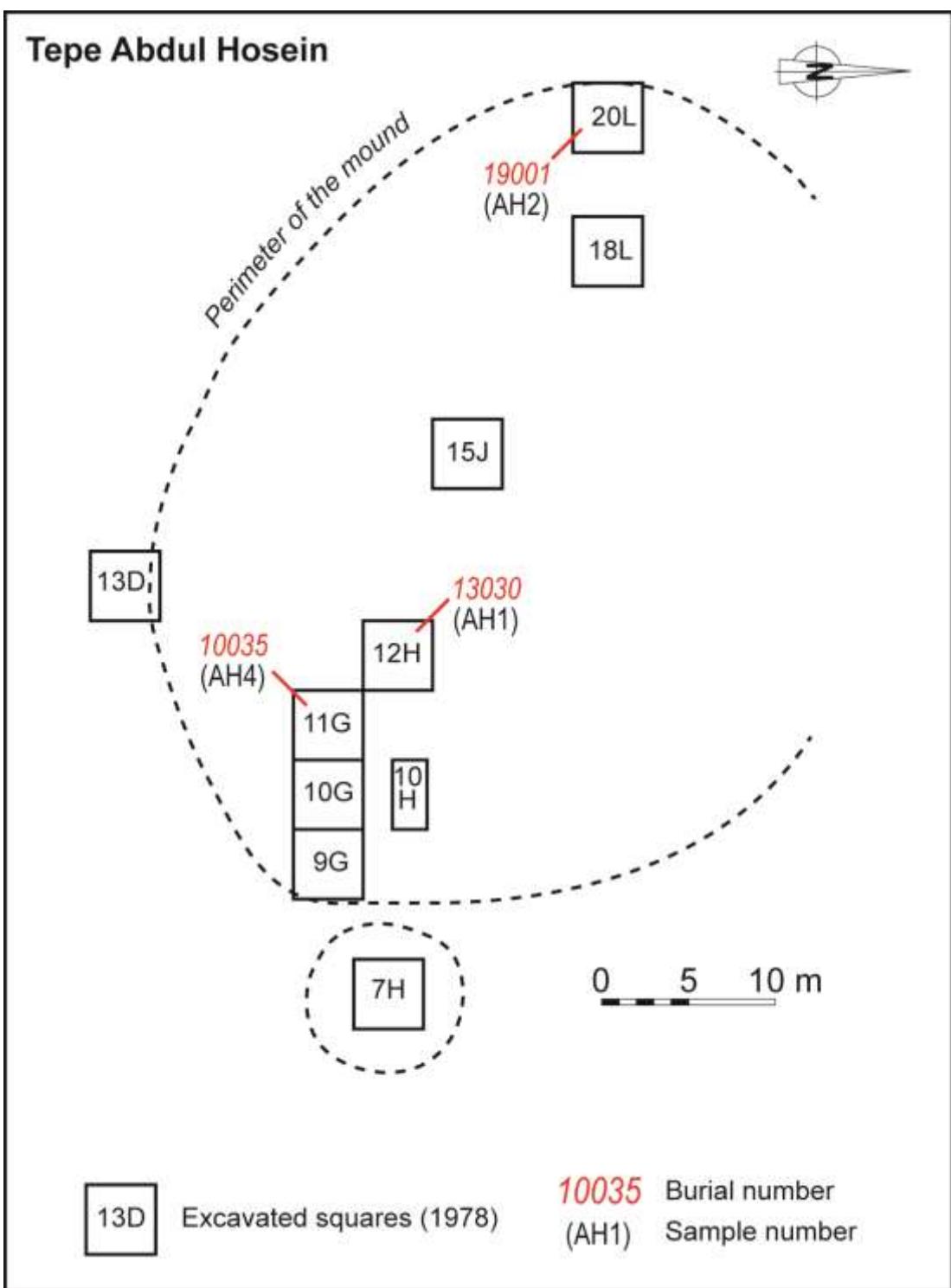
Fig S2
Localization of the sampled skeletons in the excavated squares from Tepe Abdul Hosein (O. Munoz, after (*23*): Fig. 2).
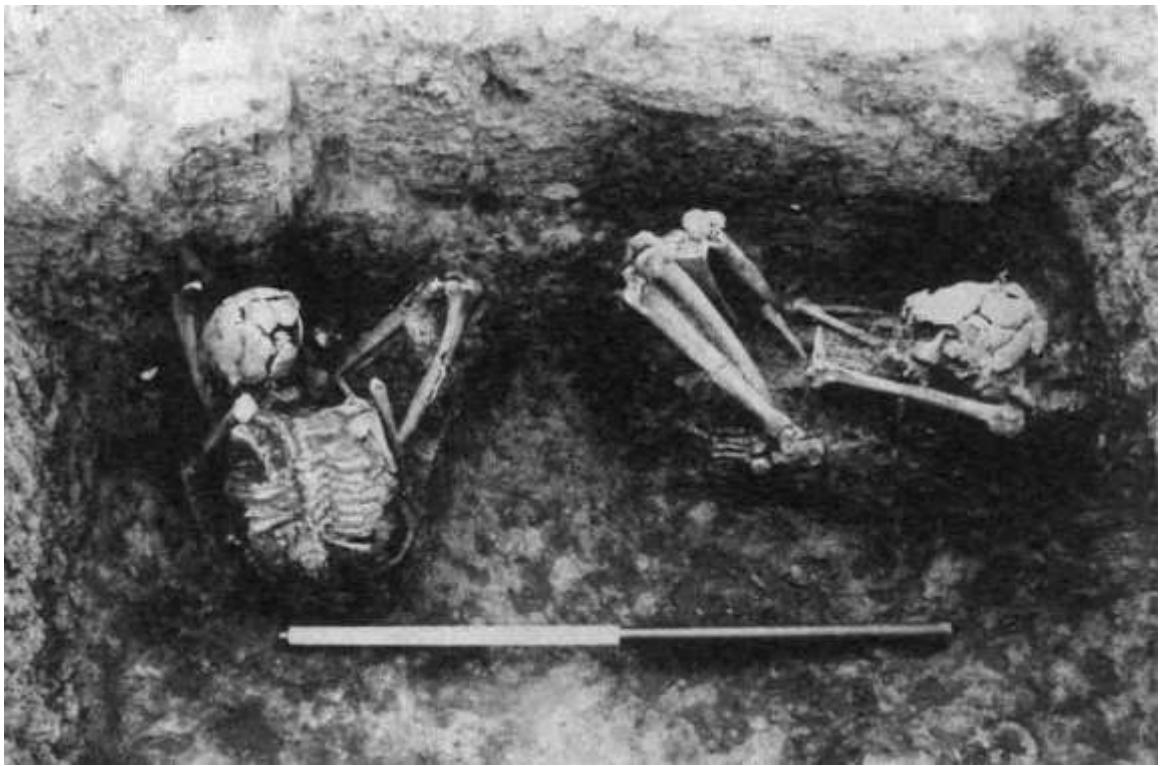
Fig. S3
Burials 13029 (right) and 13030 (sample AH1, left) in square 12H ((*23*): Plate 12).

Fig. S4
Position of the skeleton 19001#2 in square 20L ((*23*): plate 14, p. 247).
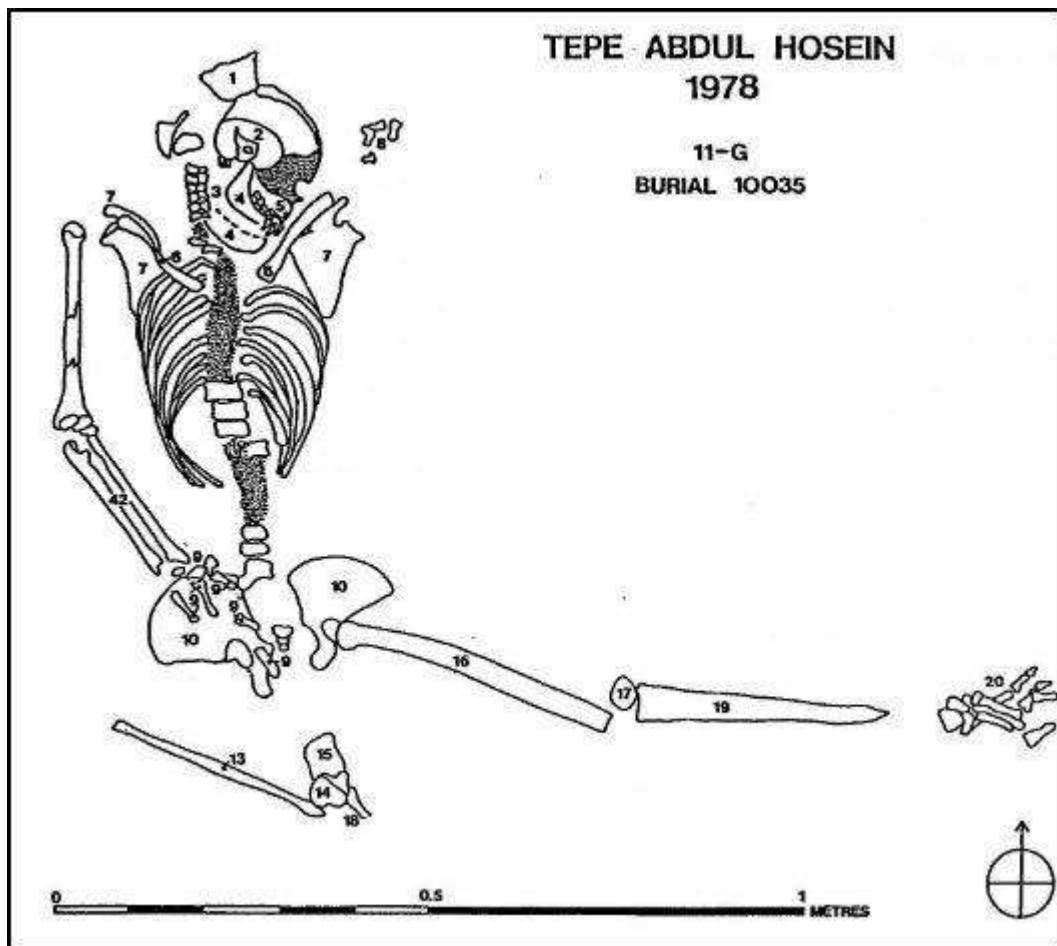
Fig. S5
Position of the skeletal remains from Burial 10035 (sample AH4) in square 11G ((*23*): Fig. 17)

49
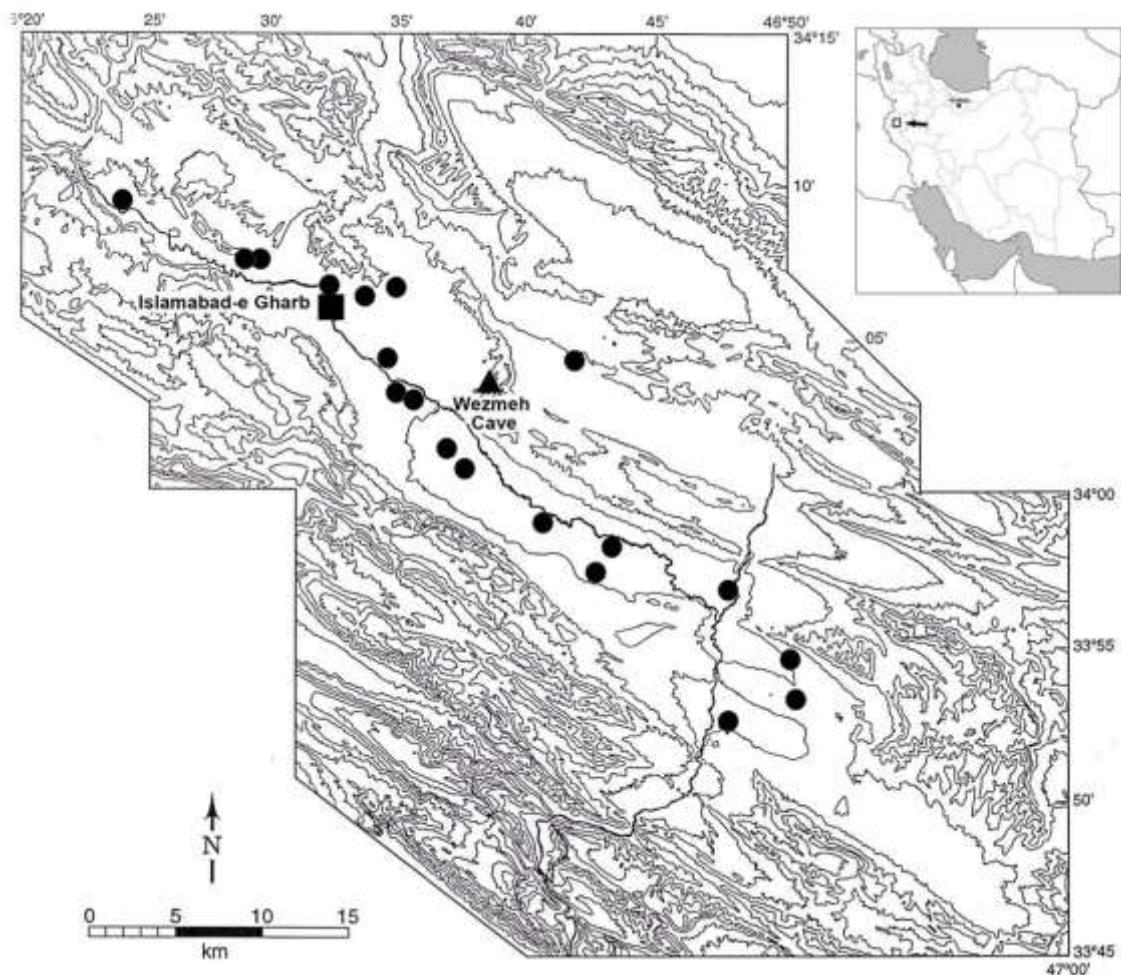
Fig. S6
Map of the Islamabad Plain with the location of Wezmeh Cave. Solid black circles display Neolithic mound sites surrounding Wezmeh Cave (F. Biglari after (*48*): Figs. 10-12).
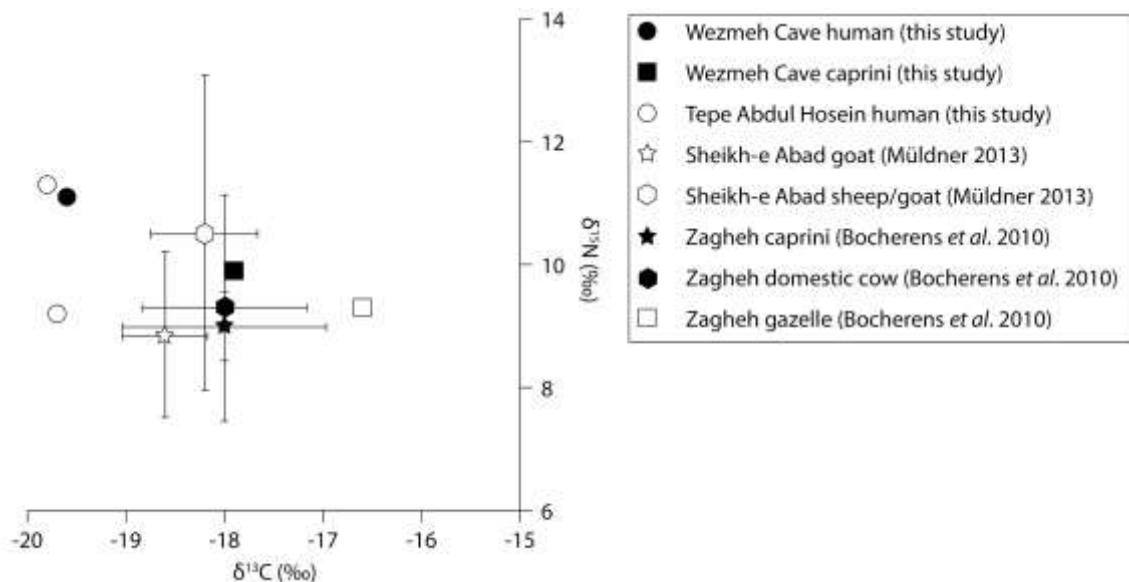
Fig. S7

Bone collagen carbon and nitrogen stable isotope values measured from humans (WC1, FB2 and FB4) and fauna (WM2) from Wezmeh Cave and Tepe Abdul Hosein compared with goats (n = 9) and sheep/goats (n = 4) from Sheikh-e Abad (*160*) and Bos taurus (n = 2), Caprini (n = 4) and Gazella (n = 1) from Zagheh (*161*). Plot shows mean values for fauna and 1 s.d. from the mean.
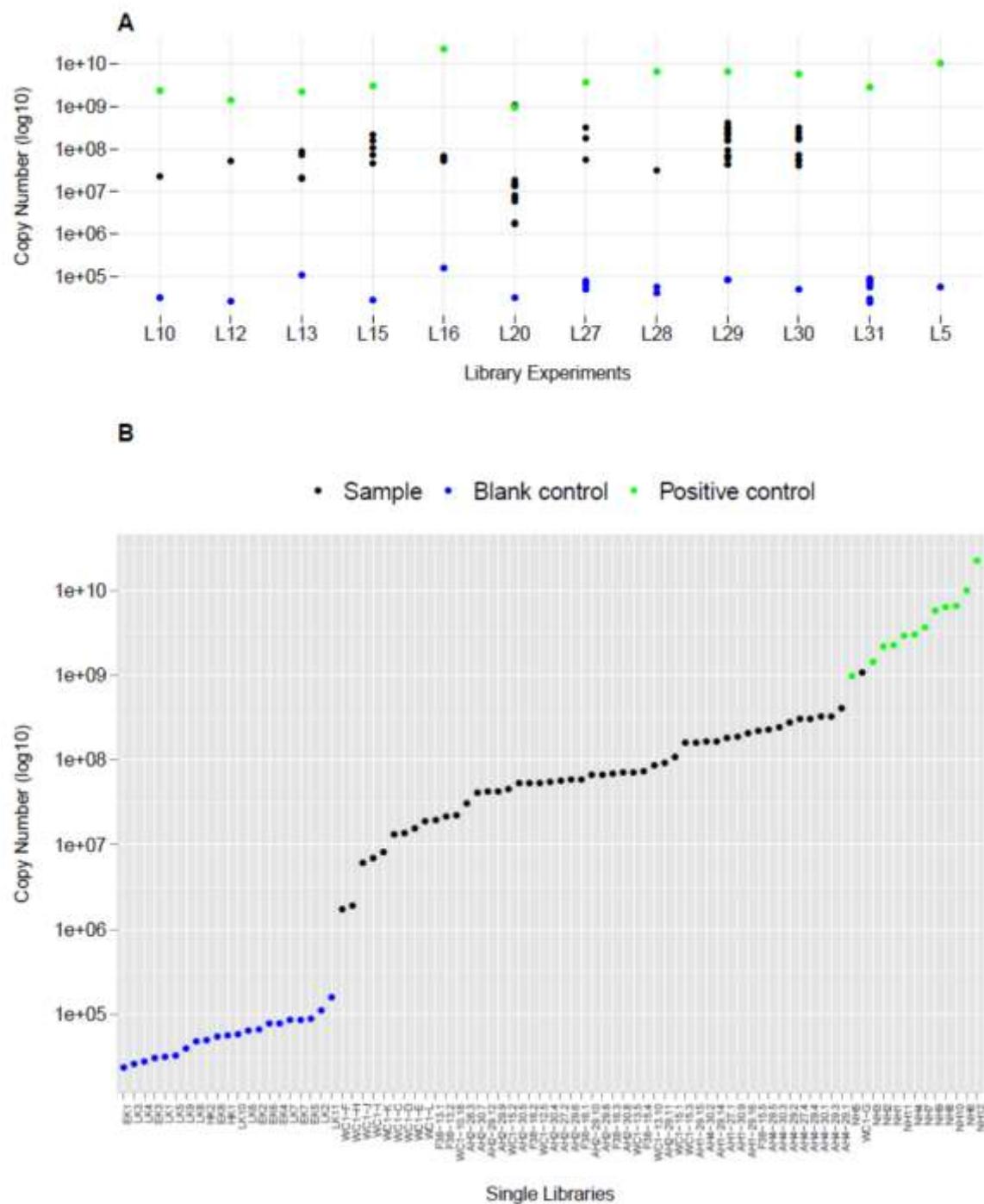
Fig. S8

Copy number comparison of all libraries. Samples, blanks, and positive controls are highlighted in black, blue, and green, respectively. **A**. Comparison of Molecule numbers of each library experiment and corresponding blank and positive control **B**. Molecule number of single libraries, blanks, and positive controls.
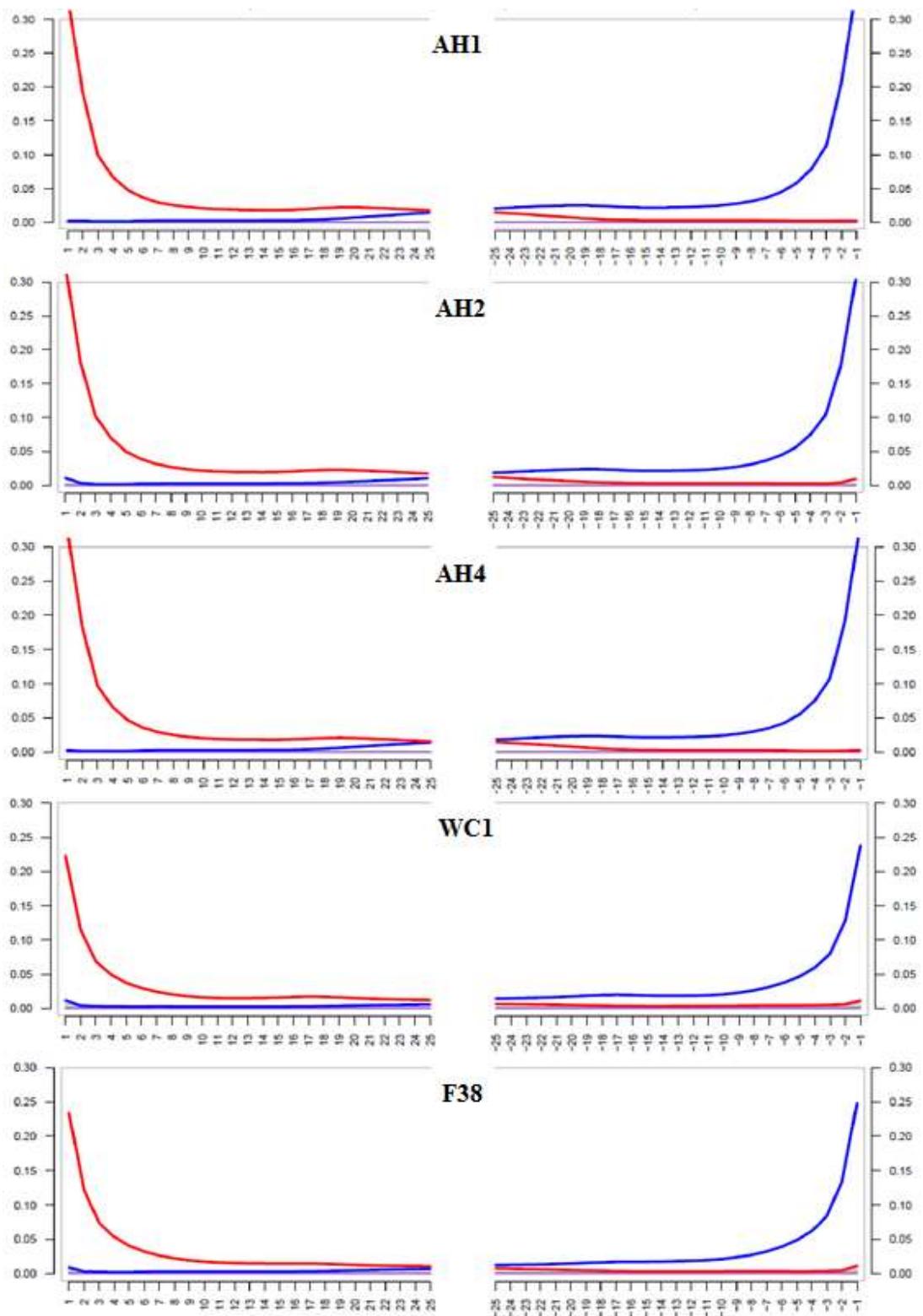
Fig. S9
C/T deamination patterns over all mapped reads per sample
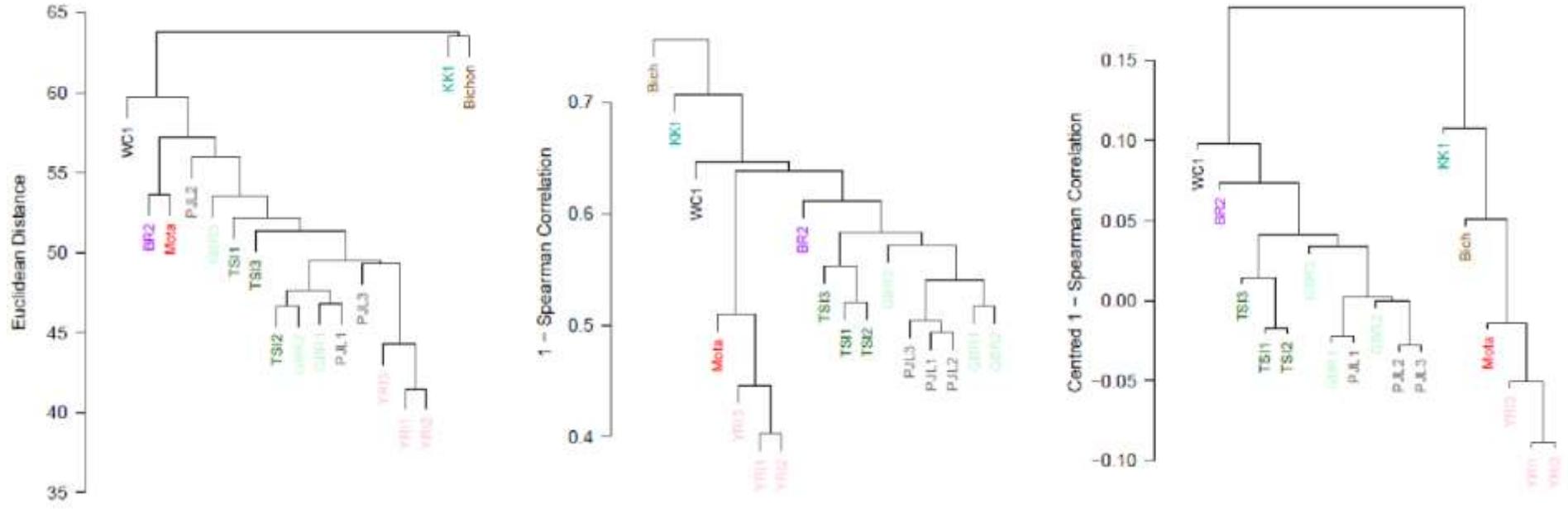
Fig. S10

Results of a cluster analysis showing similarities in heterozygosity ($\hat{\theta}$) estimates across modern and ancient samples. Clustering in the left most panel was based on Euclidean distances and hence reflects primarily similarities in over levels of diversity. In the center panel, individuals were clustered based on 1-Spearman correlations and hence reflect the distribution of diversity in the genome, rather than the overall level. To further reduce noise in the estimates, we also clustered individuals based on the centered 1-Spearman correlations (right most panel)

.
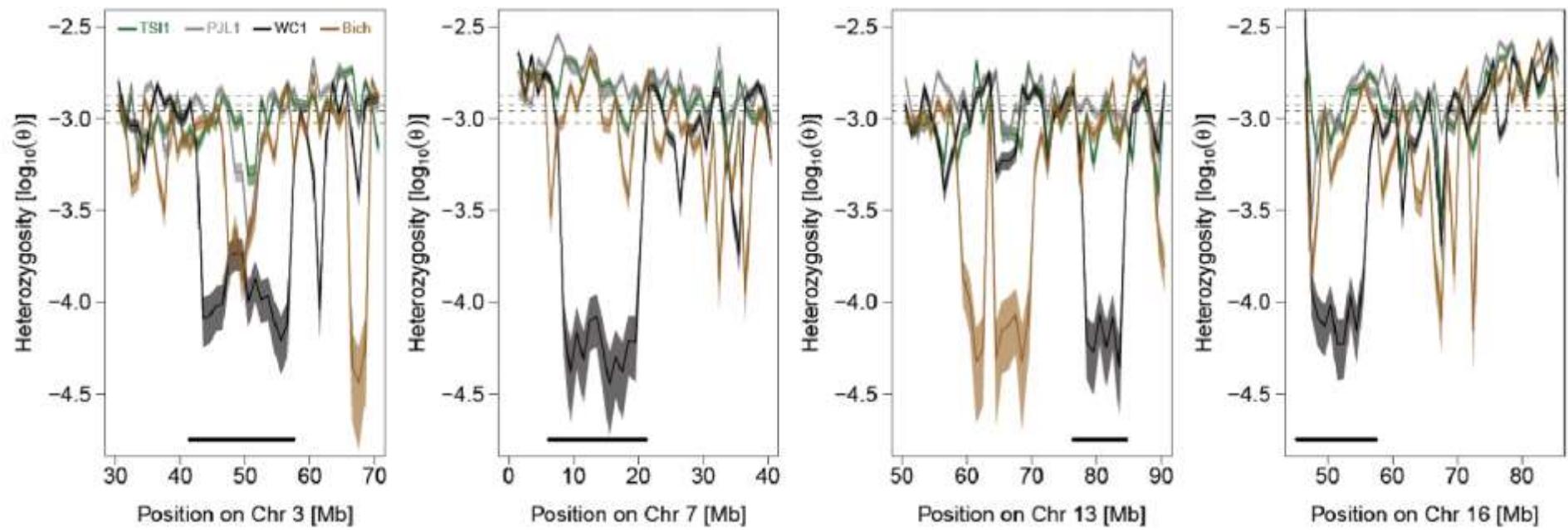
Fig. S11

Maximum Likelihood Estimates of local heterozygosity ($\hat{\theta}$, solid lines) and 95% confidence intervals (shaded areas) around the four segments (black bar at bottom) of low diversity as identified by ROH analysis
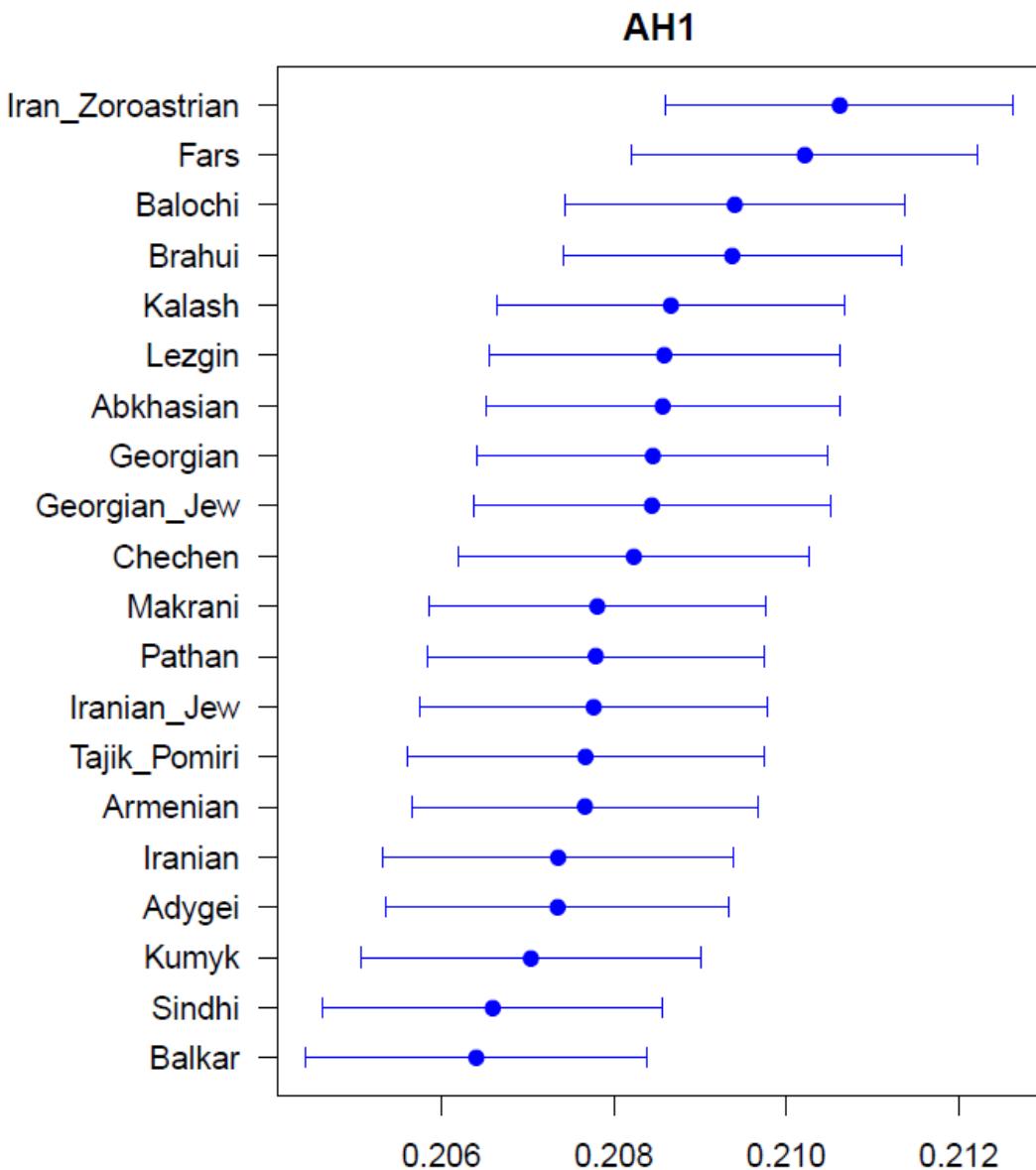
Fig. S12

f3(‡Khomani; Modern_population, AH1). The highest 20 values shown.
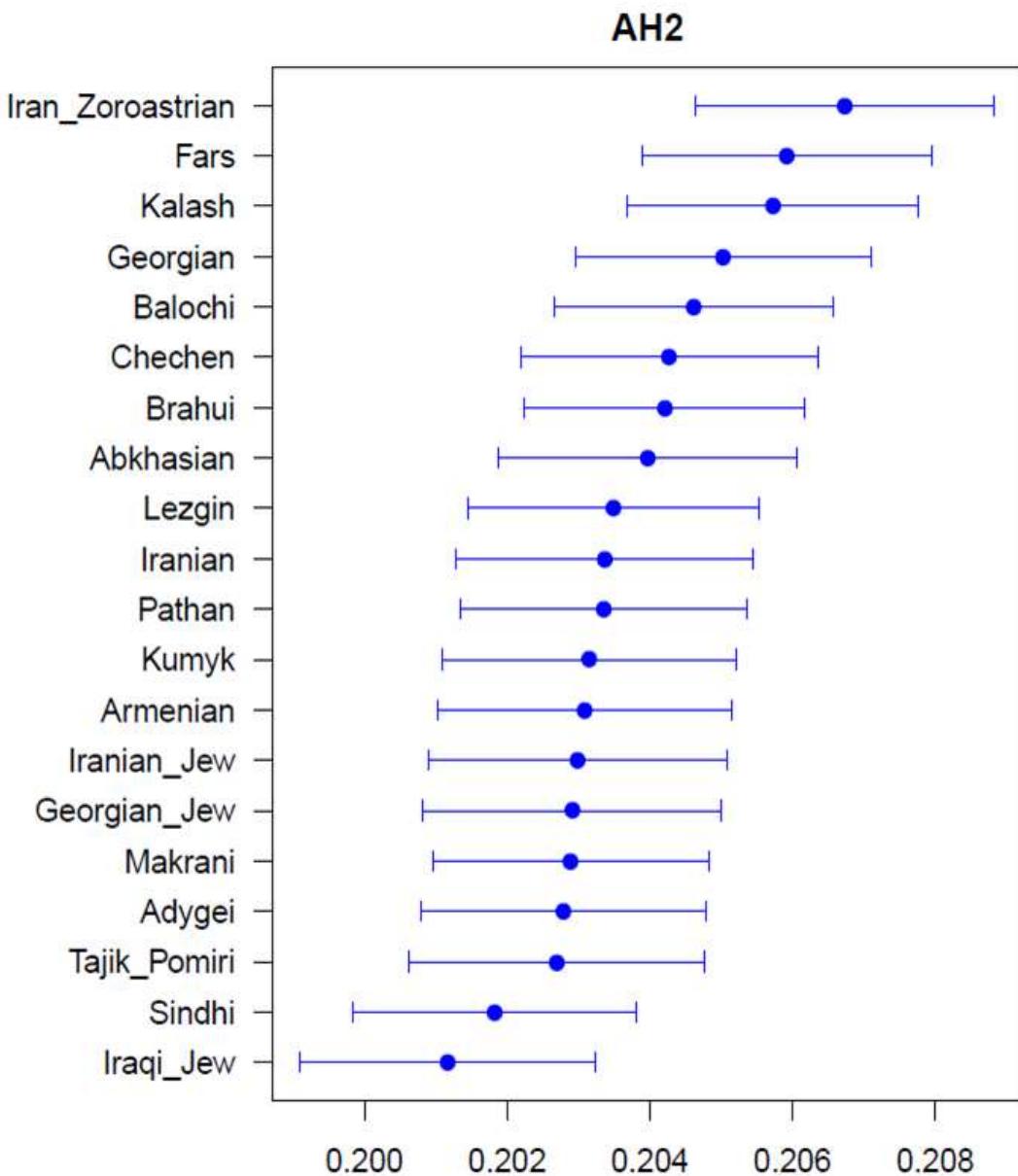
Fig. S13

f3(‡Khomani; Modern_population, AH2). The highest 20 values shown.
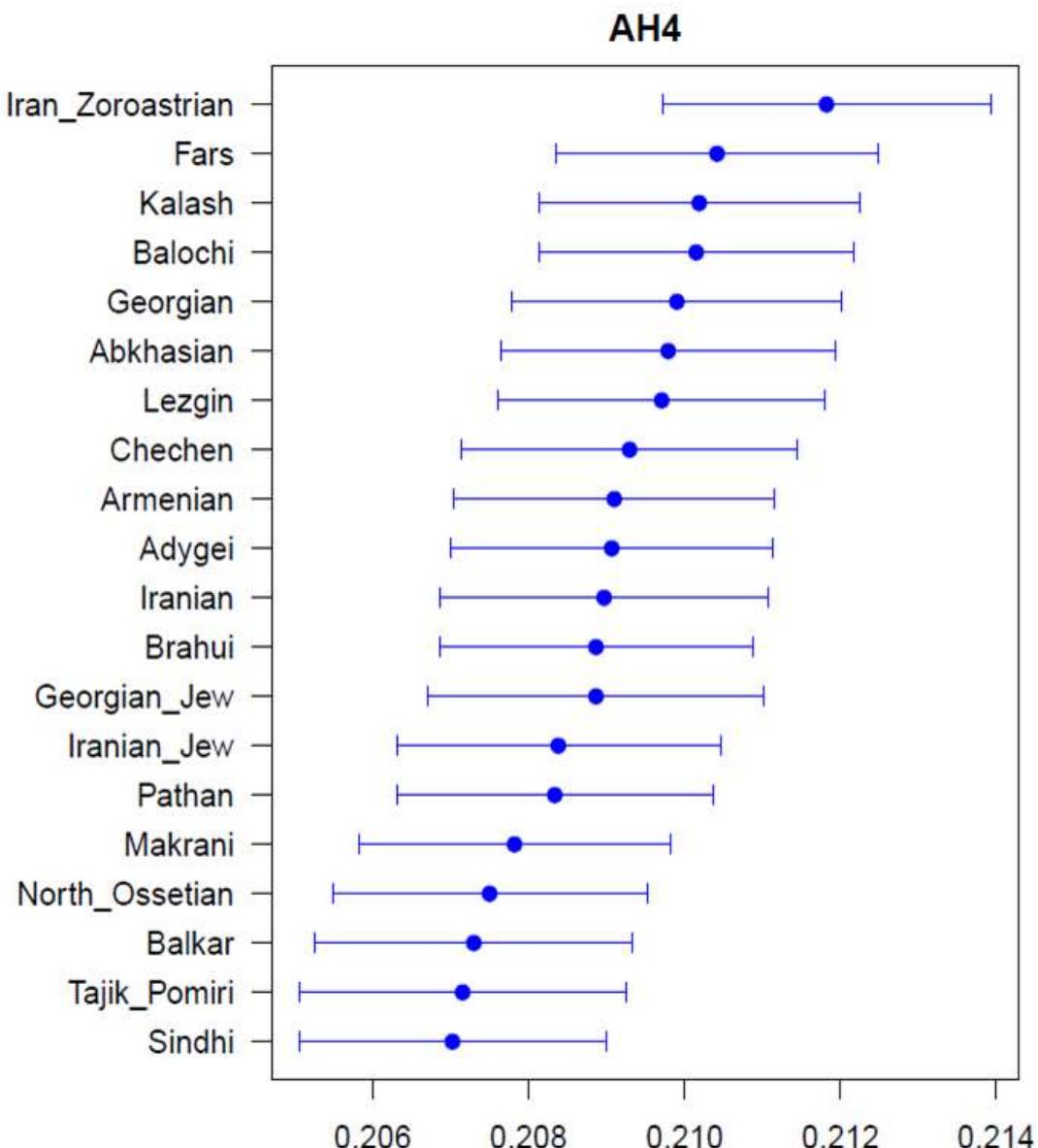
Fig. S14
ƒ3(‡Khomani; Modern_population, AH4). The highest 20 values shown.
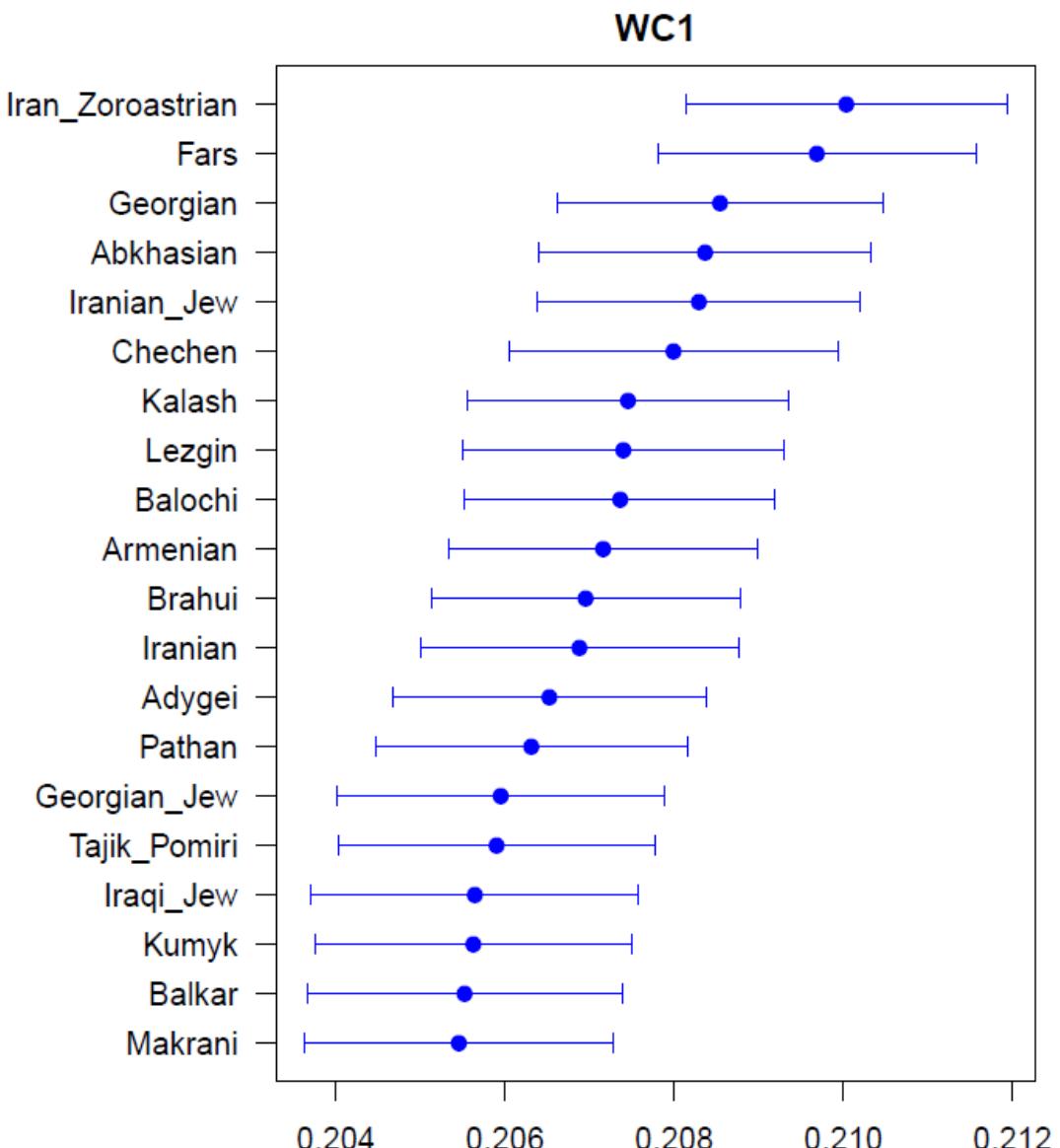
Fig. S15

f3(‡Khomani; Modern_population, WC1). The highest 20 values shown.
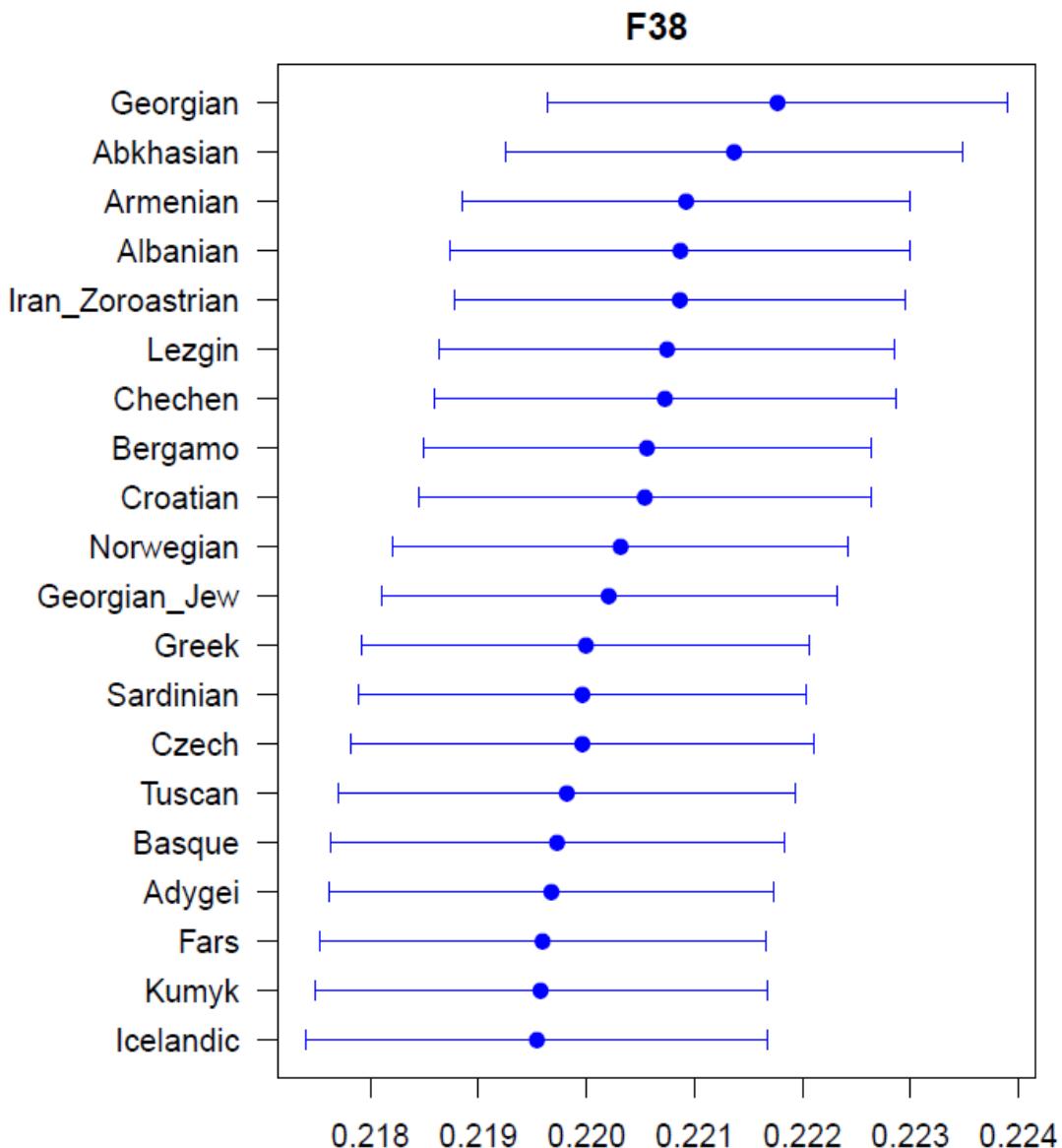
Fig. S16

f3(‡Khomani; Modern_population, F38). The highest 20 values shown.
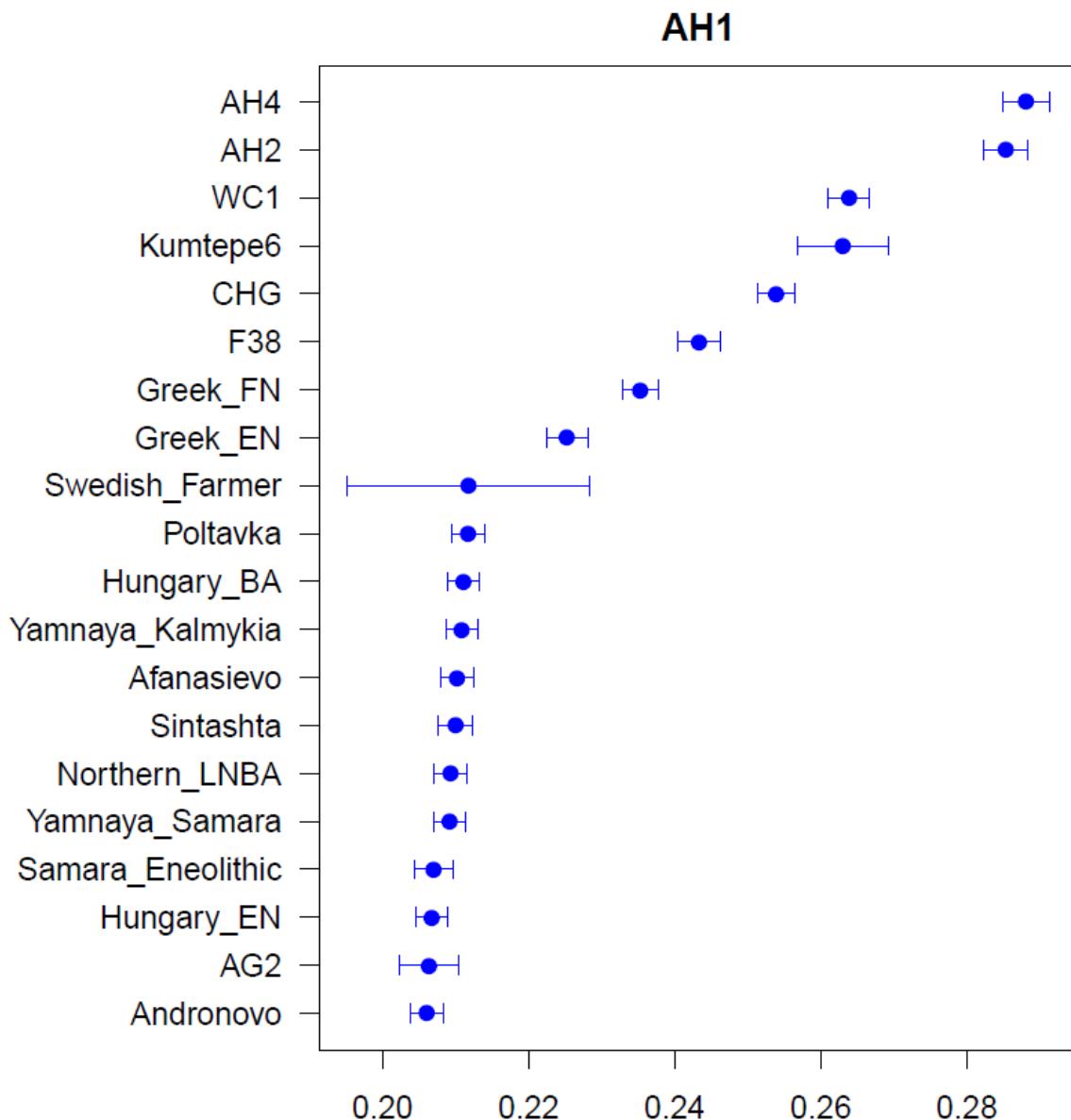
Fig. S17

f3(ǂKhomani; Ancient_population, AH1). The highest 20 values shown.
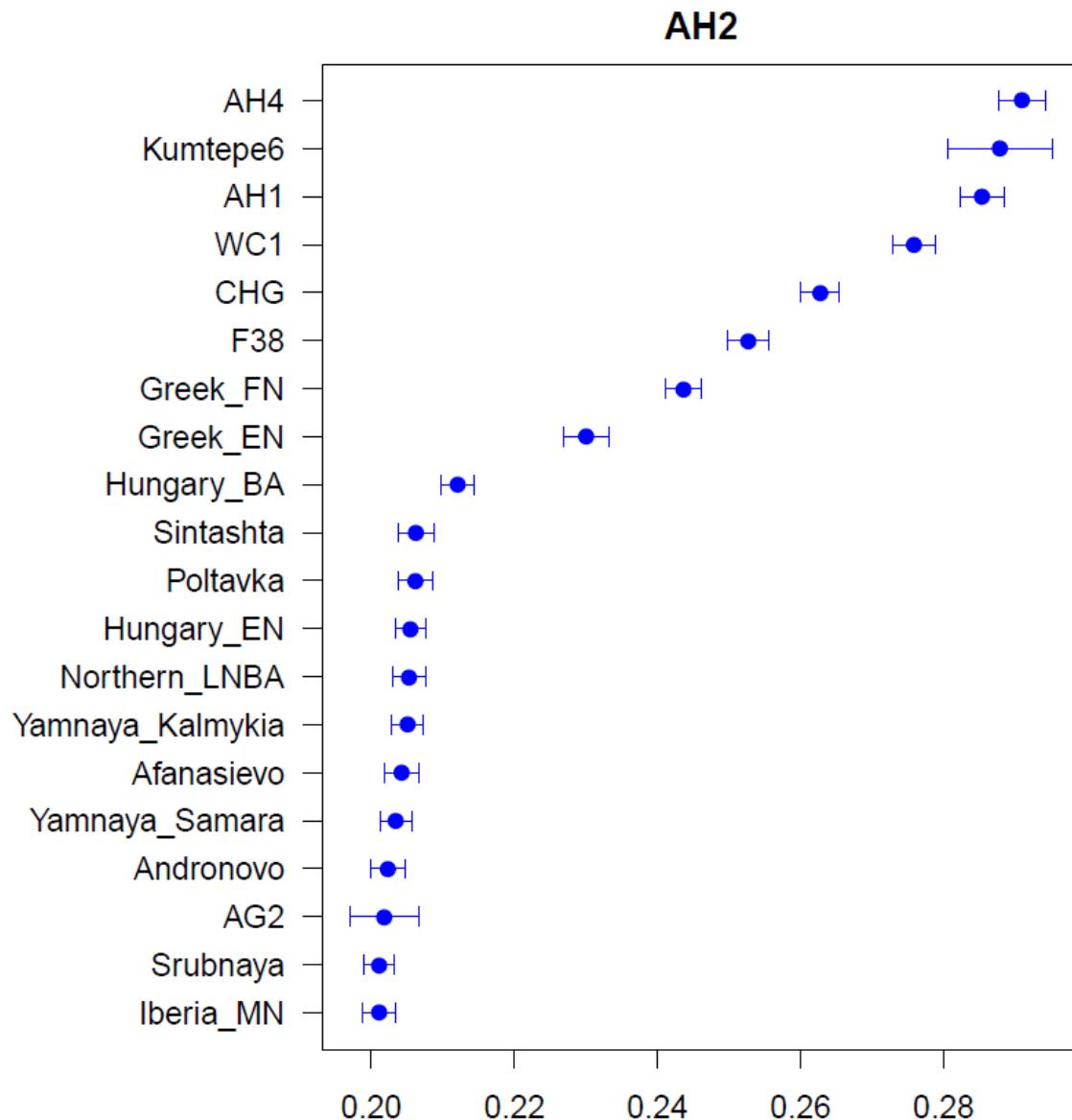
Fig. S18

f3($\natural$Khomani; Ancient_population, AH2). The highest 20 values shown.
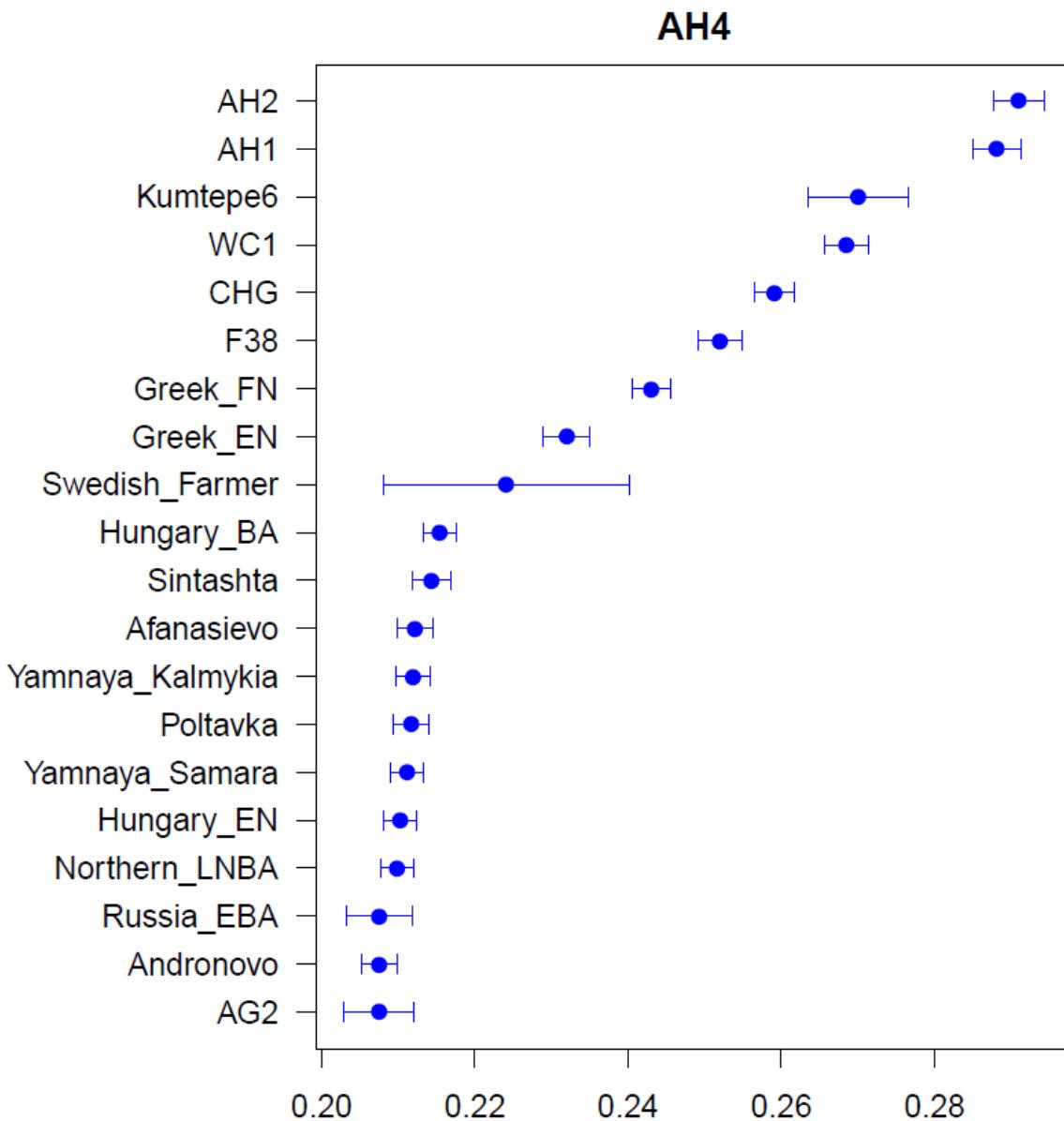
Fig. S19

f3(‡Khomani; Ancient_population, AH4). The highest 20 values shown.
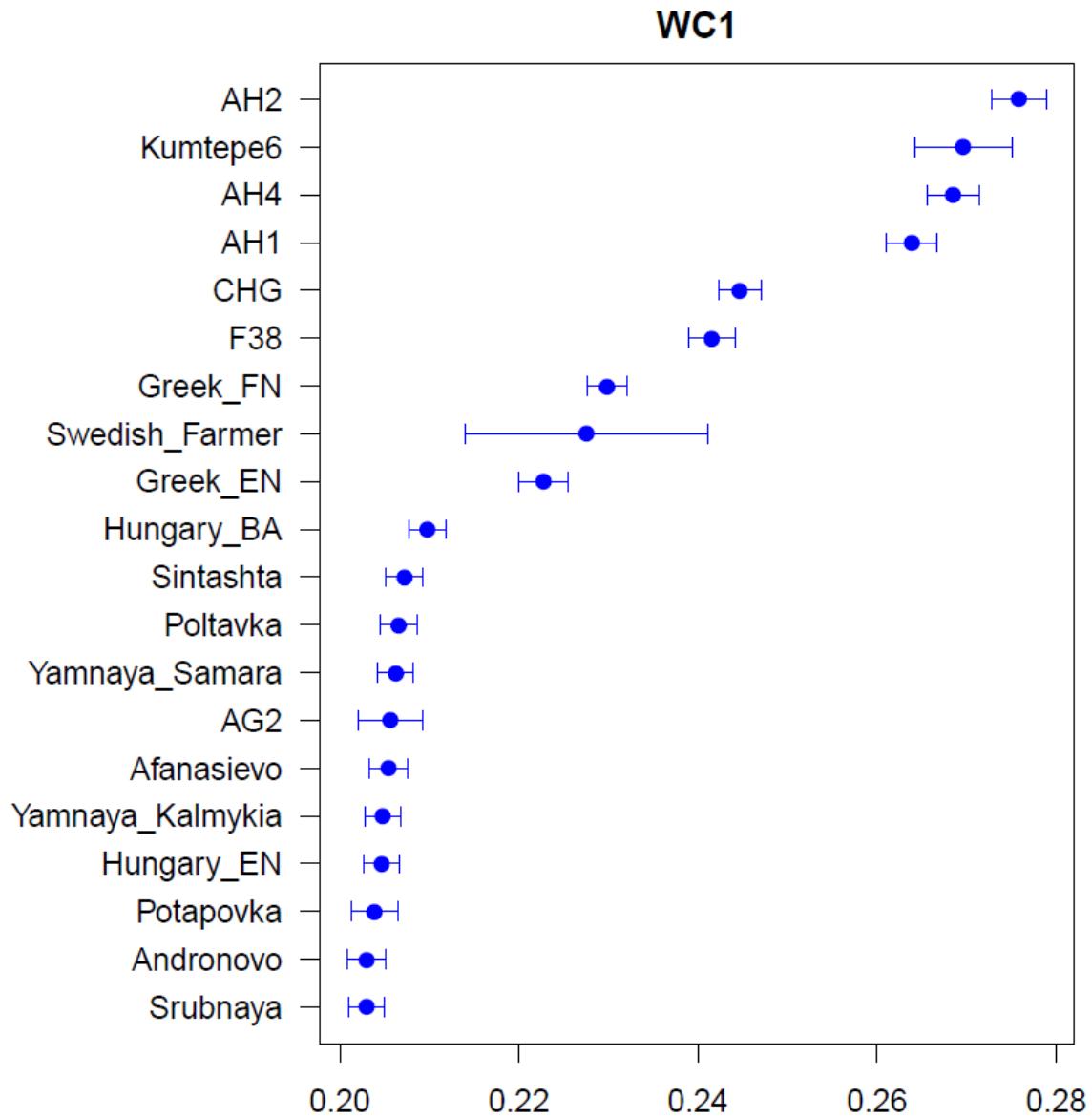
Fig. S20

f3(ǂKhomani; Ancient_population, WC1). The highest 20 values shown.
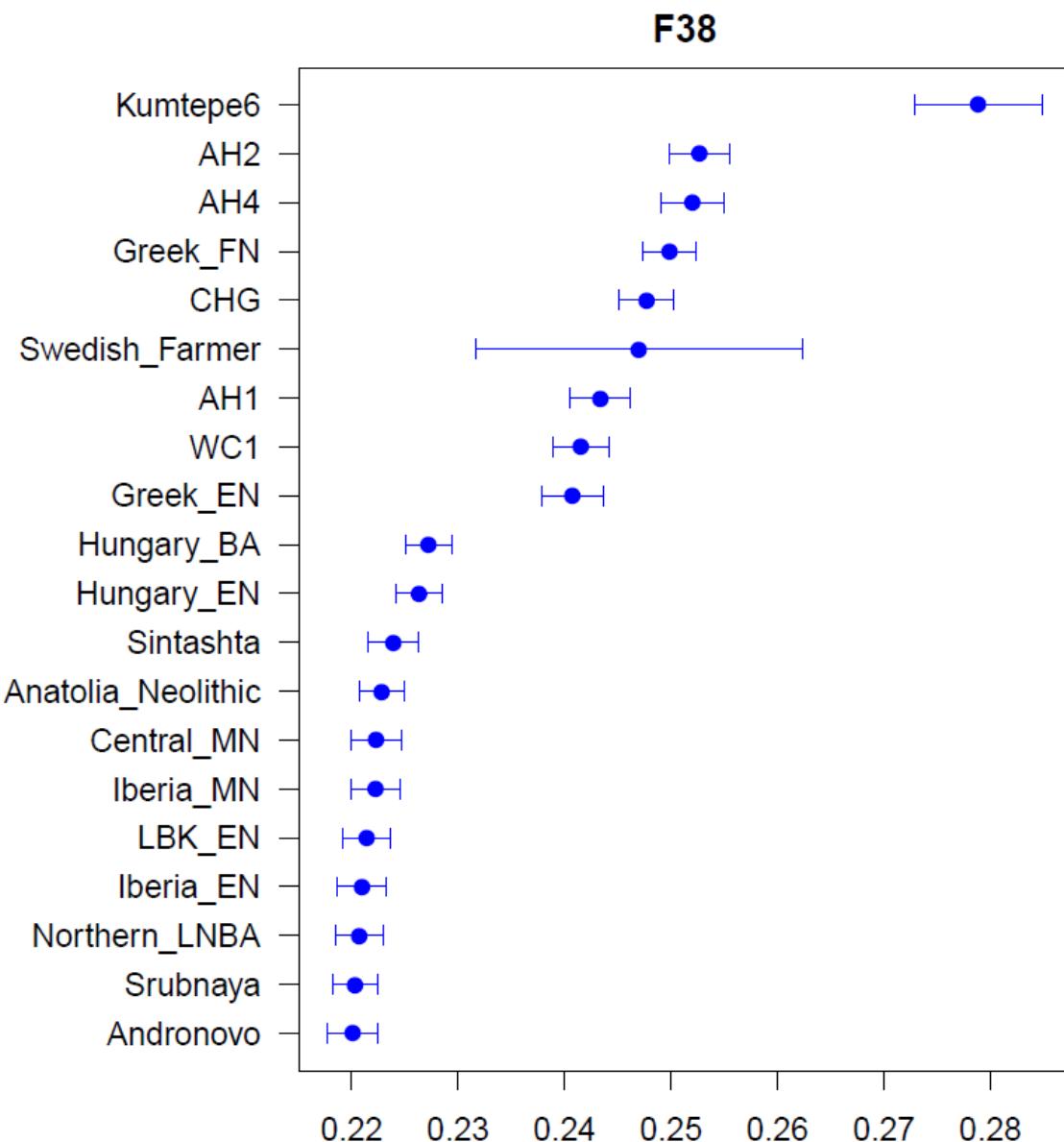
Fig. S21

f3(ǂKhomani; Ancient_population, F38). The highest 20 values shown.
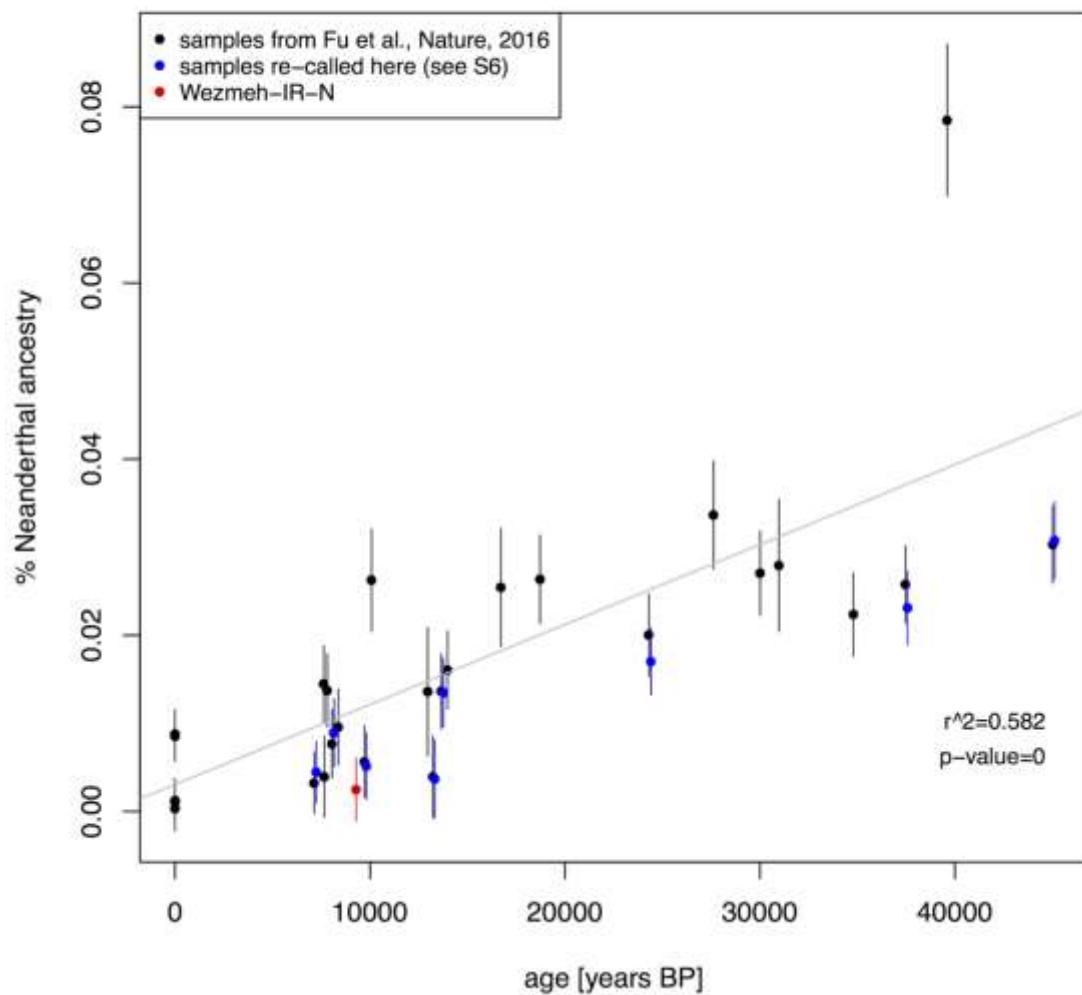
Fig. S22

Neanderthal ancestry proportions estimated by *f4*-ratio statistics. Linear regression includes 'Oase1' (14) corresponding to the outlier with recent Neanderthal ancestry.
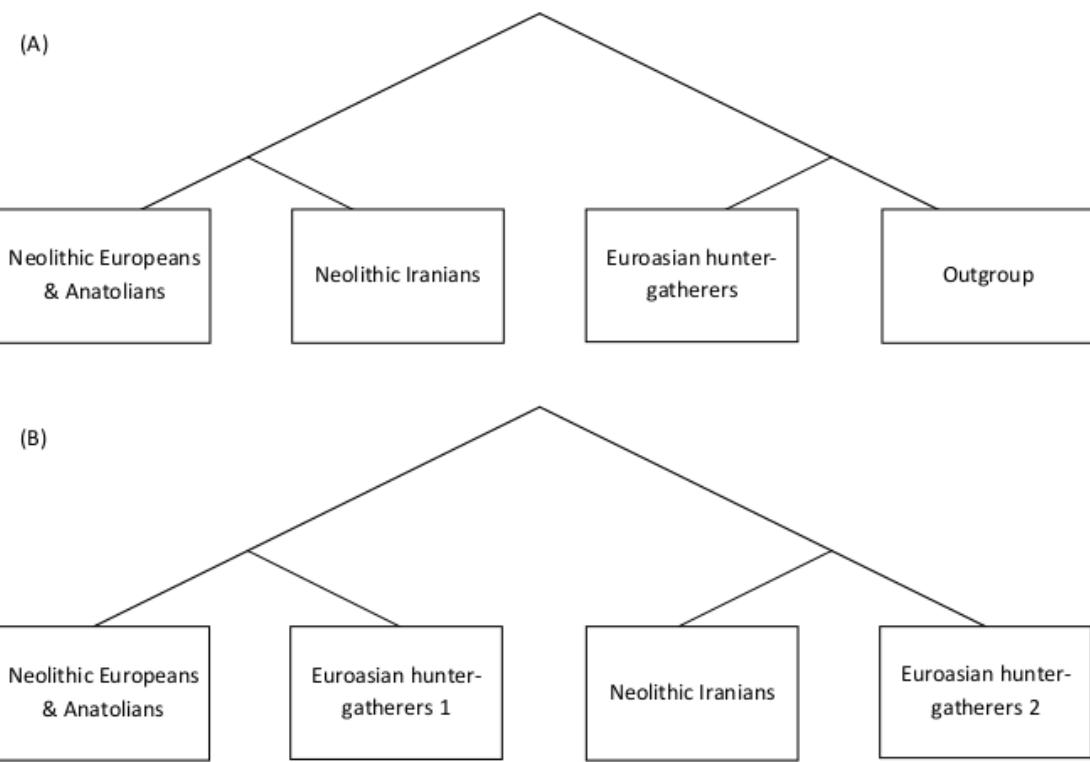
Fig. S23
Different models of relatedness for Near Eastern Neolithic Farmers.
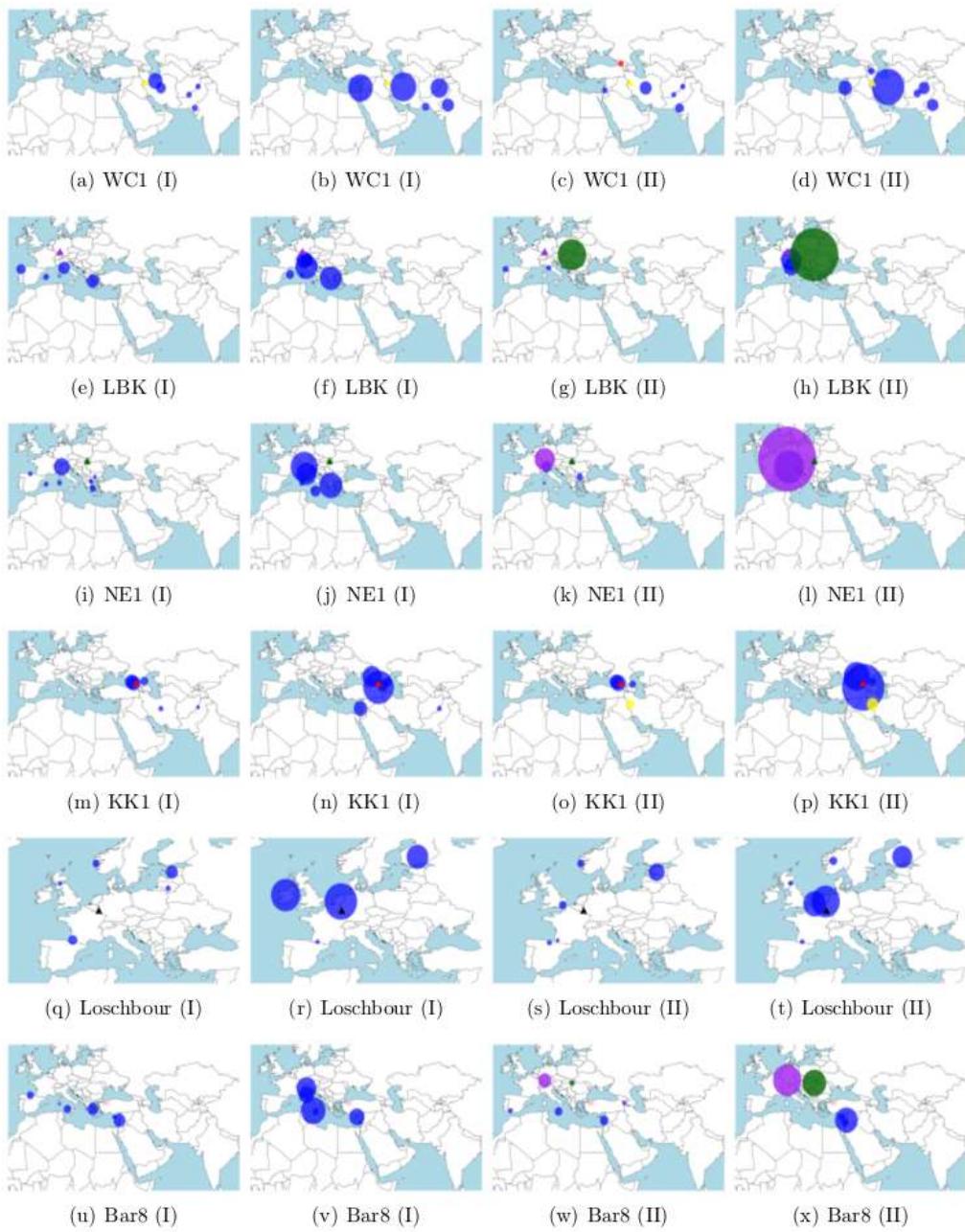
Fig. S24

Inferred proportions of ancestry for each Neolithic sample when using all moderns (analysis (I)) versus all modern+ancient groups (analysis (II)) as surrogates in the Lazaridis (columns 1 and 3) and Busby (columns 2 and 4) merges. Circles are proportional to the inferred proportions from modern samples (blue) and aDNA samples from Iran (*WC1*; yellow), Stuttgart (*LBK*; purple), Hungary (*NE1*; dark green), Georgia (*KK1*; red),Luxembourg (*Loschbour*: black) and Neolithic Anatolian (Bar8: pink). Triangles represent the sampling location of the depicted target sample (and also provide the key for that sample's color).

(a) Mota (I) (b) Mota (I)

(c) Mota (II) (d) Mota (II)

(e) Ust'-Ishim (I) (f) Ust'-Ishim (I)

(g) Ust'-Ishim (II) (h) Ust'-Ishim (II)
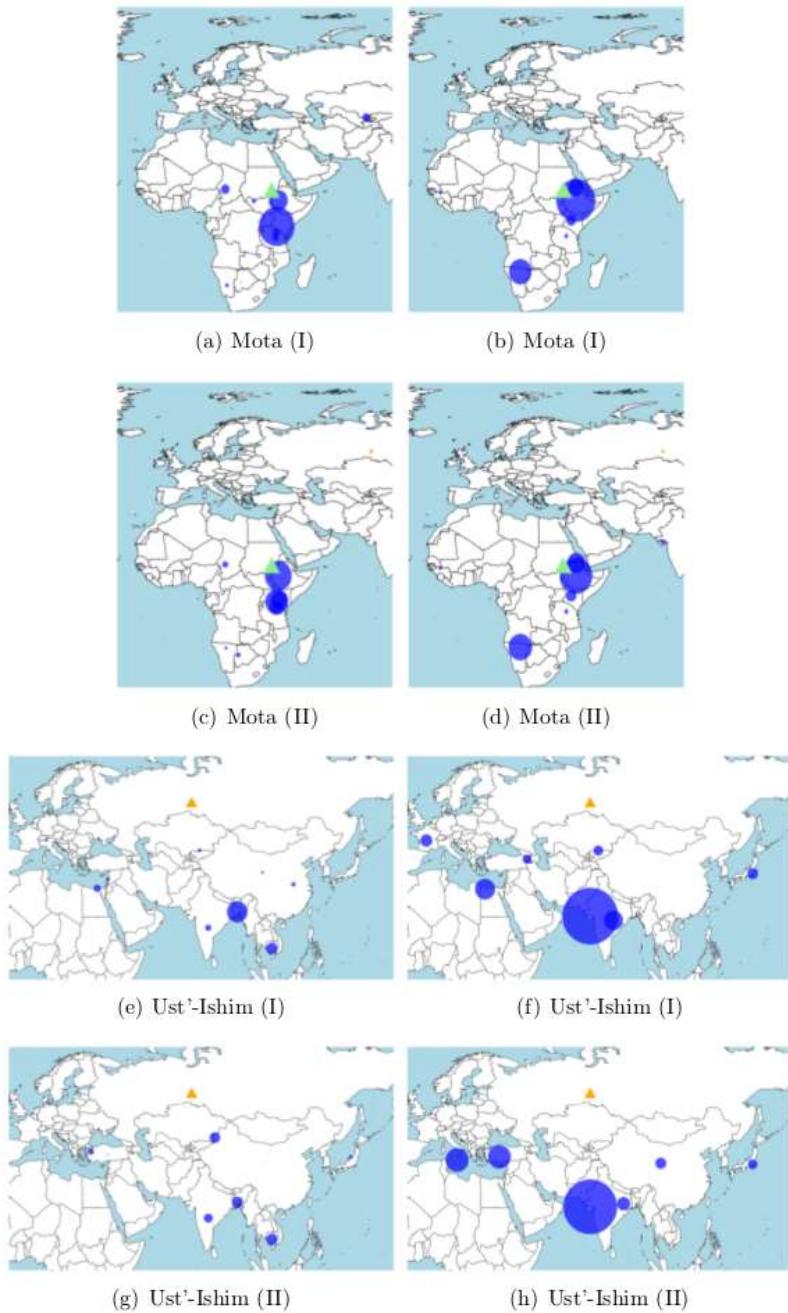
Fig. S25

Inferred proportions of ancestry for each Neolithic sample when using all moderns (analysis (I)) versus all modern+ancient groups (analysis (II)) as surrogates in the Lazaridis (column 1) and Busby (column 2) merges. Circles are proportional to the inferred proportions from modern samples (blue) and aDNA samples from a 4,500 year old genome sampled from Ethiopia (*Mota*; light green) and a 45,000 year old western Siberian (*Ust'-Ishim*: orange). Triangles represent the sampling location of the depicted target sample (and also provide the key for that sample's color).

14
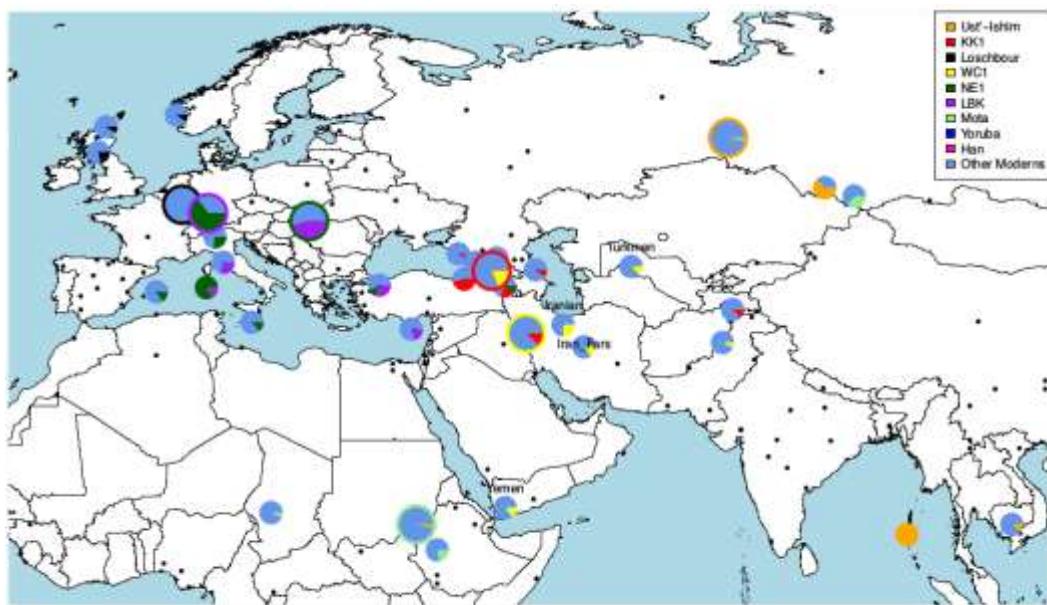
(a) Lazaridis merge - analysis (II)
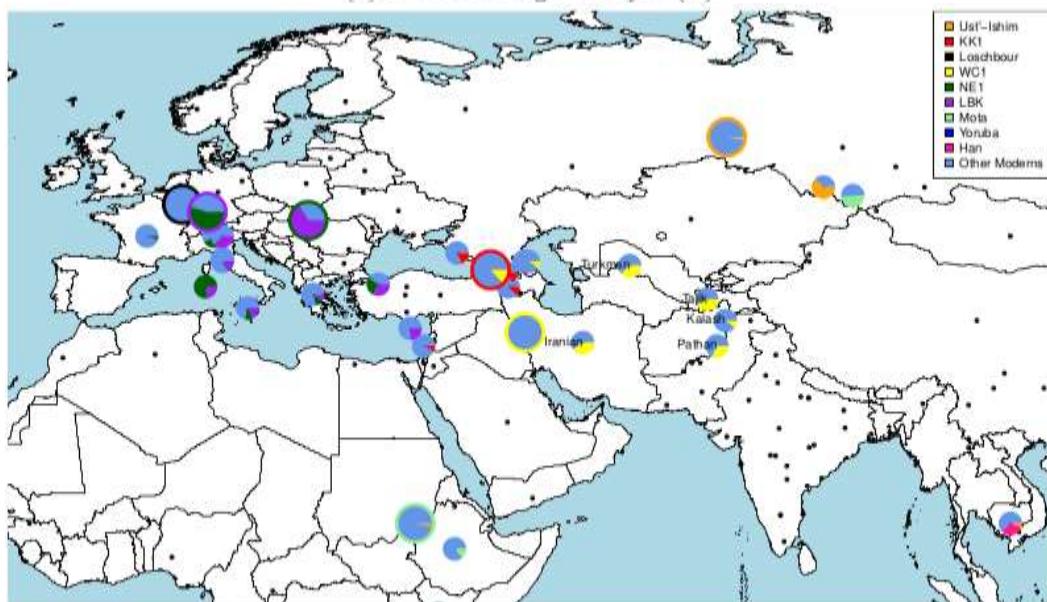
(b) Busby merge - analysis (II)

Fig. S26

Inferred proportions of ancestry for modern and ancient groups when using modern groups and all ancient samples as surrogates under analysis (II) for the Lazaridis (top) and Busby (bottom) merges. Pies are only shown for those modern groups receiving a > 5% contribution from any aDNA sample. All other modern groups are shown as a black dot. Population labels are printed for modern groups where the contribution from WC1 is > 10%. The larger pie charts represent the aDNA samples. The border of these pies corresponds to the legend at the right.

15
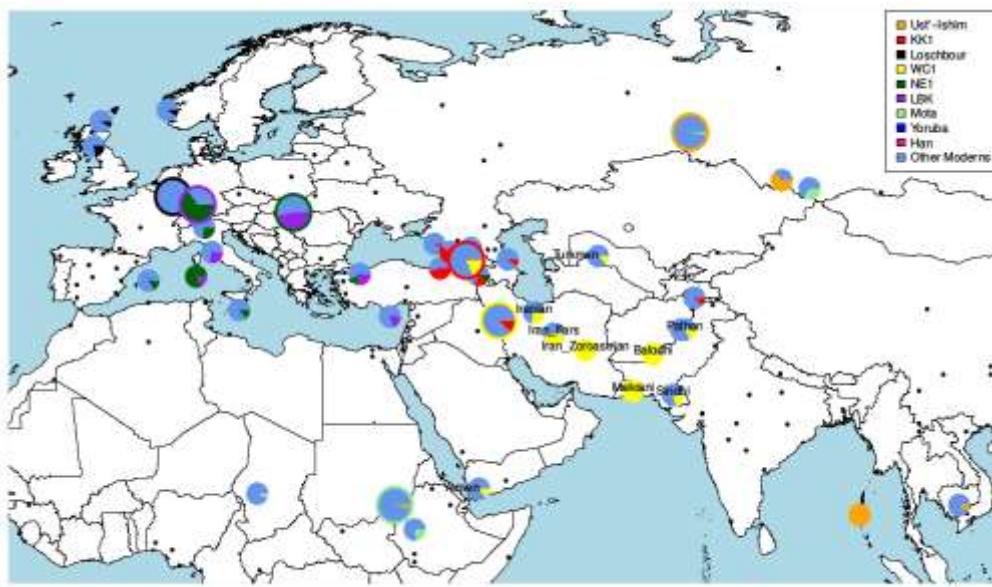
(a) Lazaridis merge - analysis (II) and (II-)
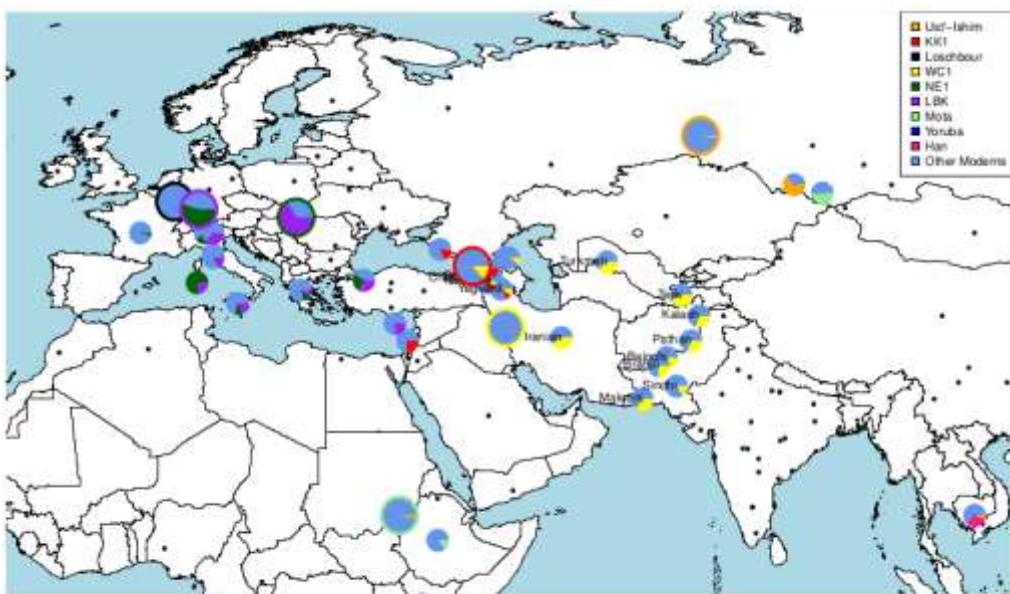
(b) Busby merge - analysis (II) and (II-)

Fig. S27

Inferred proportions of ancestry for modern and ancient groups when using modern groups and all ancient samples as surrogates under analysis (II) for the Lazaridis (top) and Busby (bottom) merges. In contrast to Fig. S26, results for analysis (II-) are shown where there is a high density of neighbouring groups (see methods). For the Lazaridis merge, analysis (II-) results are shown for separate analyses of Iranian, Pakistani, Indian and Georgian populations. For the Busby merge, analysis (II-) results are shown for separate analyses of Armenian, Pakistani and Indian populations. See caption of Fig. S26 for further details.
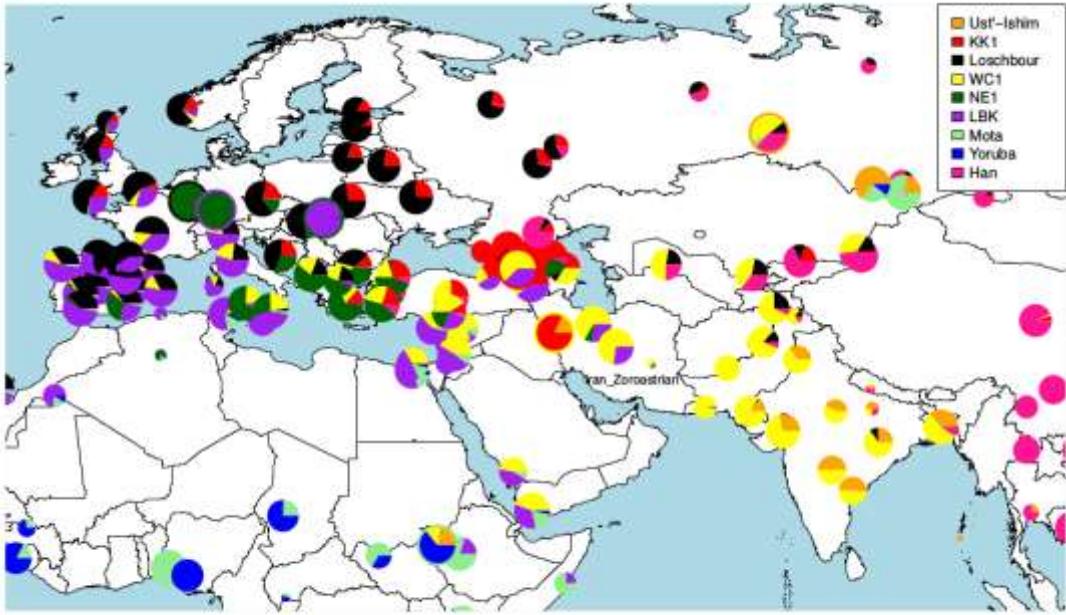
16

(a) Lazaridis merge - Ancients + Yoruba + Han as surrogates (III)
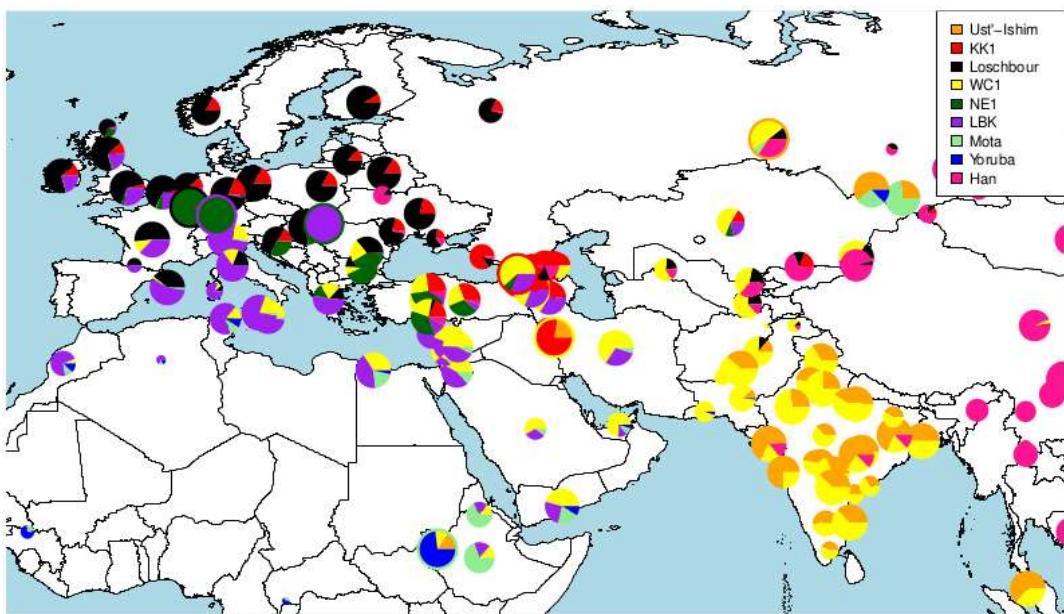
(b) Busby merge - Ancients + Yoruba + Han as surrogates (III)

Fig. S28

Inferred proportions of ancestry for modern and ancient groups when using all ancients and the modern Yoruba and Han samples (analysis (III)) as surrogates for the Lazaridis (top) and Busby (bottom) merges. For each modern population, the size of the pie charts is inversely proportional to the degree of 'self-copying' experienced by this population. This can be thought of as the smaller the circle, the more that modern group is drifted (or different) from the mixture represented by their pie chart. The bordered pie charts represent the aDNA samples, with borders corresponding to the legend at the right.
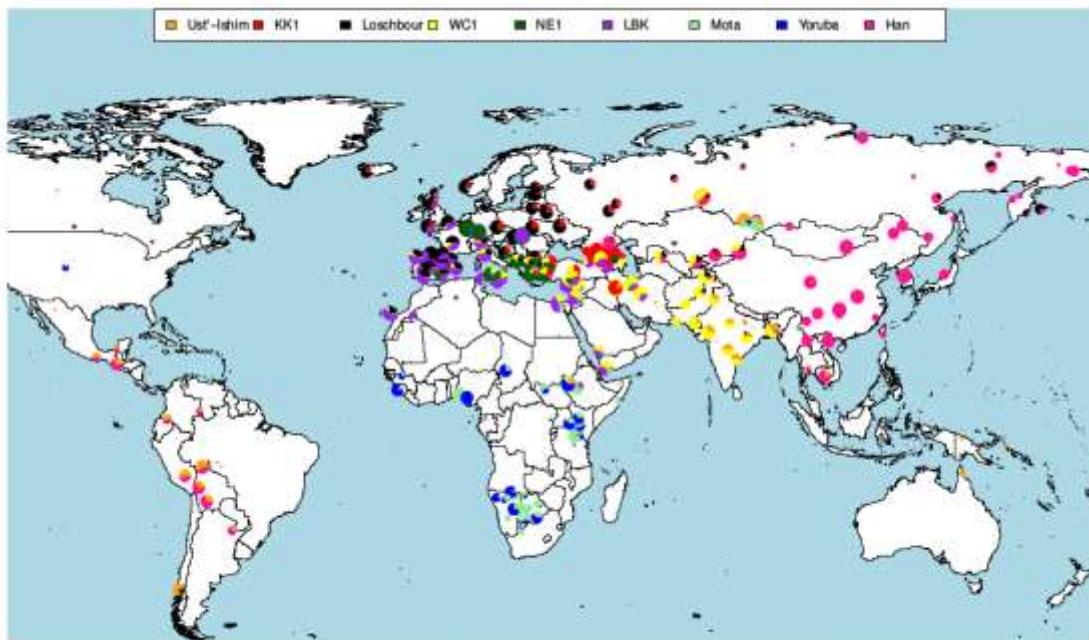
17

(a) Lazaridis merge - Ancients + Yoruba + Han as surrogates (III)
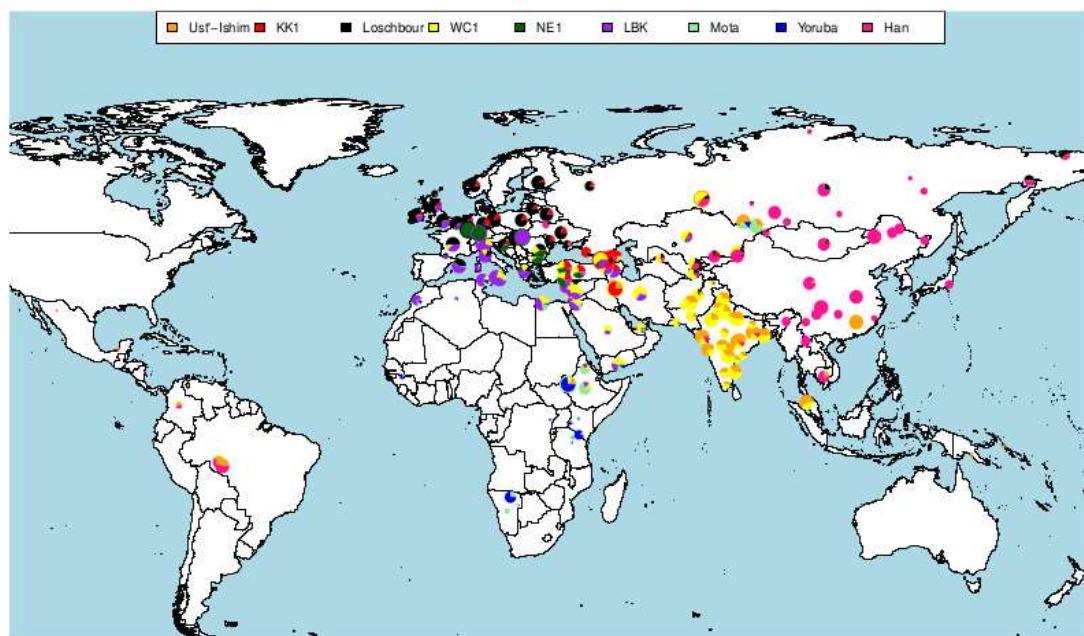
(b) Busby merge - Ancients + Yoruba + Han as surrogates (III)

Fig. S29

Inferred proportions of ancestry for modern and ancient groups when using all ancients and the modern Yoruba and Han samples (analysis (III)) as surrogates for the Lazaridis (top) and Busby (bottom) merges - world view. See caption of Fig. S28 for more detail.

18
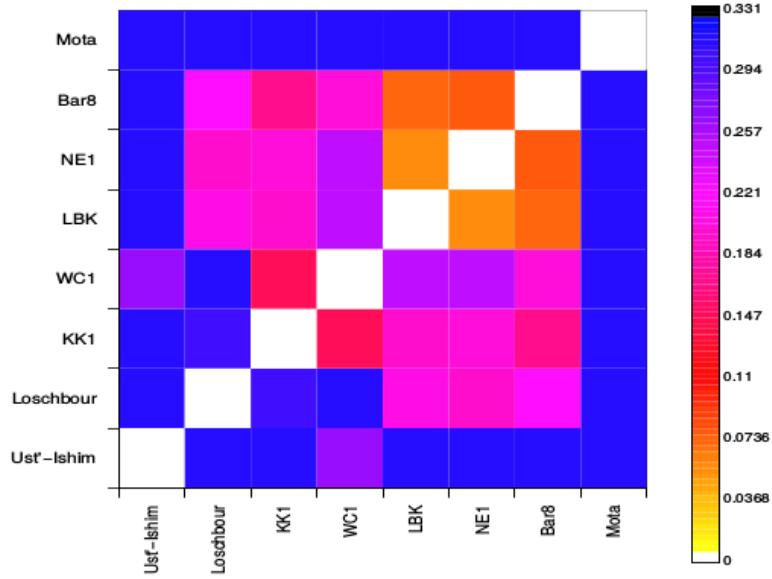
(a) TVD Lazaridis merge
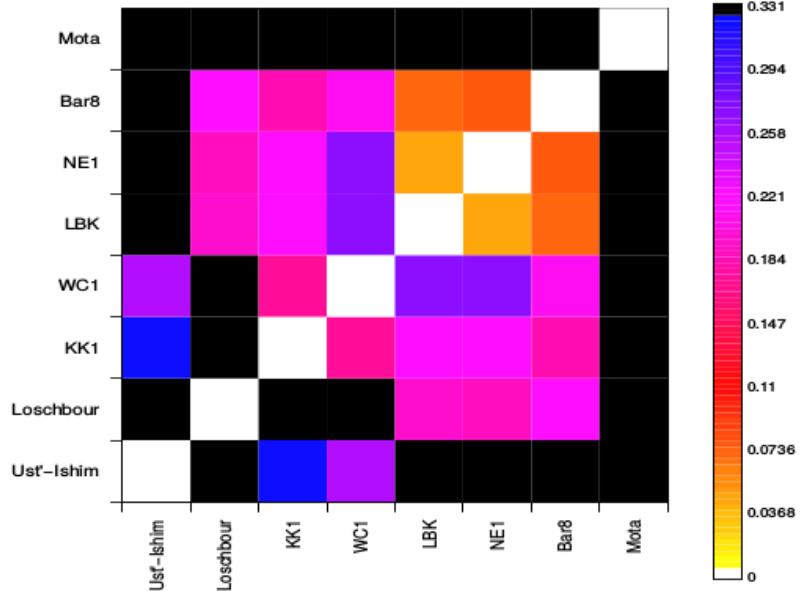
(b) TVD Busby merge

Fig. S30

Differences in inferred painting profiles (TV D) amongst ancient samples for the Lazaridis (top) and Busby (bottom) merges under analysis (I).

19
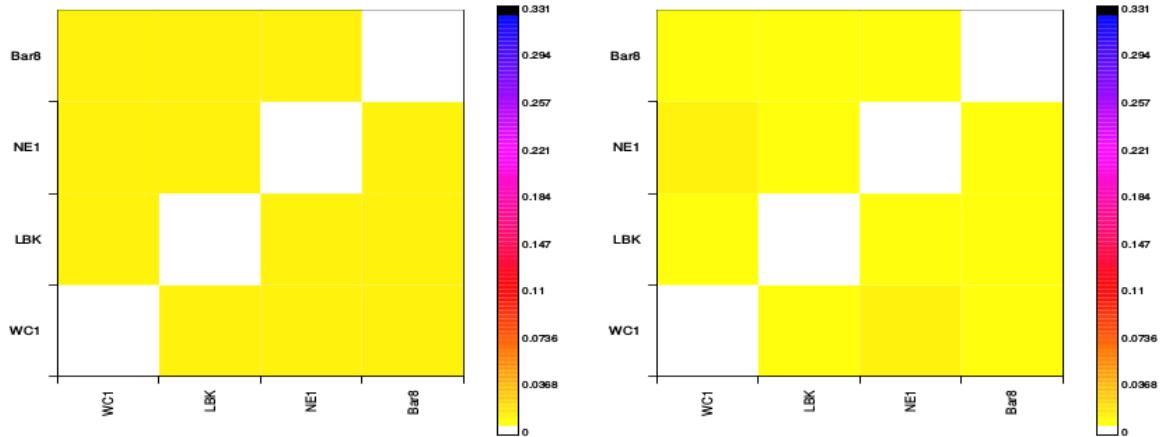
(a) TVD Lazaridis merge
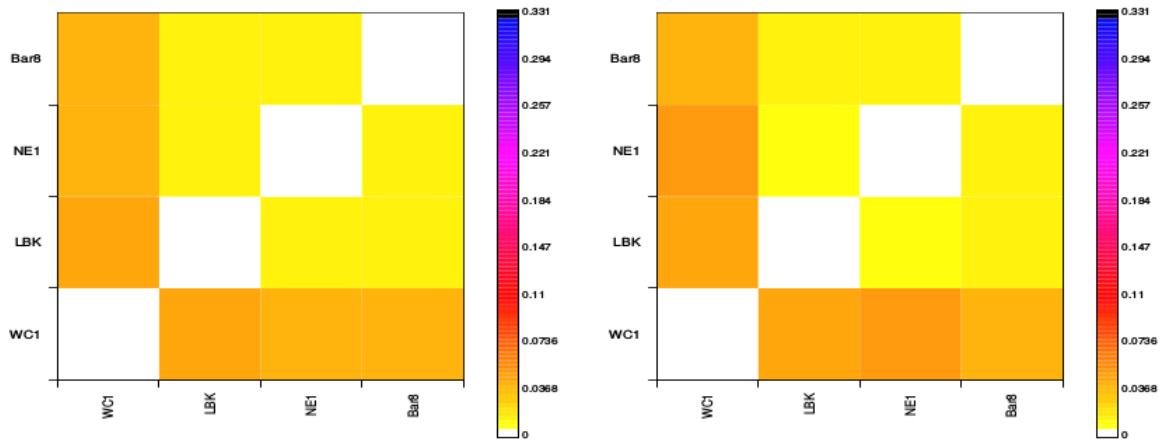
(b) TVD Busby merge

(c) TVD Lazaridis merge - plus 1 Sardinian

(d) TVD Busby merge - plus 1 Sardinian

Fig. S31

a-b) Differences in inferred painting profiles (TVD) amongst ancient samples for the Lazaridis (a) and Busby (b) merges when painted in relation to African donor individuals with no reported evidence of recent West Eurasian admixture. For the Lazaridis merge this included African individuals from Mende_Sierra_Leone, Esan_Nigeria, Bantu_SA, MbutiPygmy, BiakaPygmy, Yoruba and *Mota*. For the Busby merge this included African individuals from BantuSA, Biakapygmy, MbutiPygmy, Yoruba and *Mota*. c-d) TVD when one Sardinian individual is also included to donors used in (a) and (b), respectively.
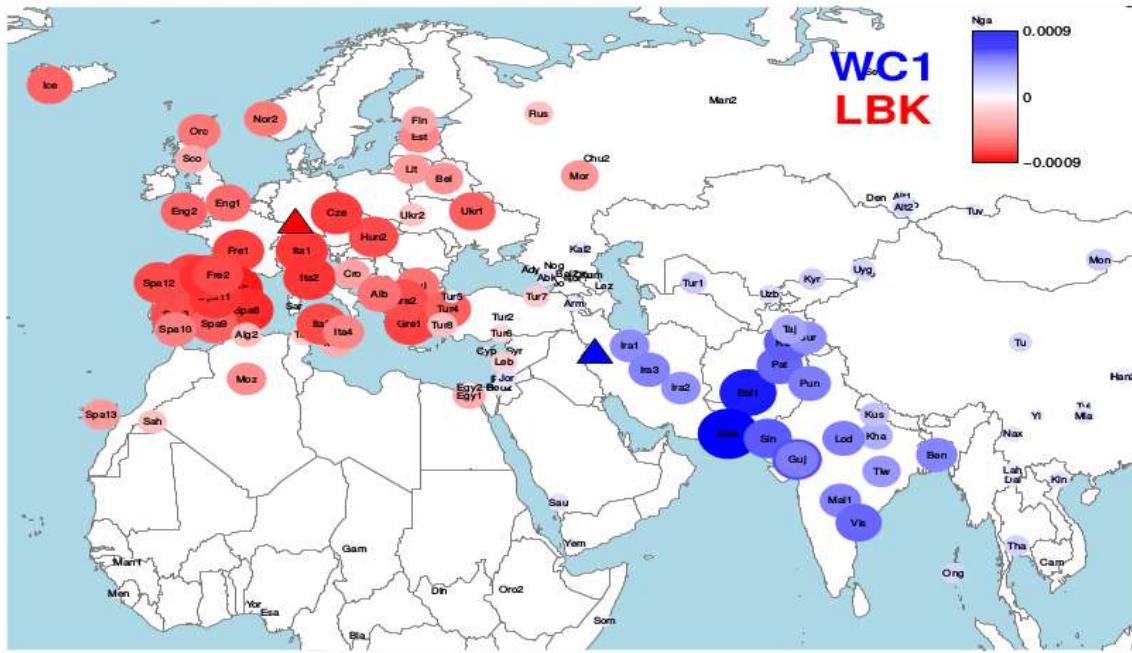
(a) Lazaridis merge - WC1 vs LBK
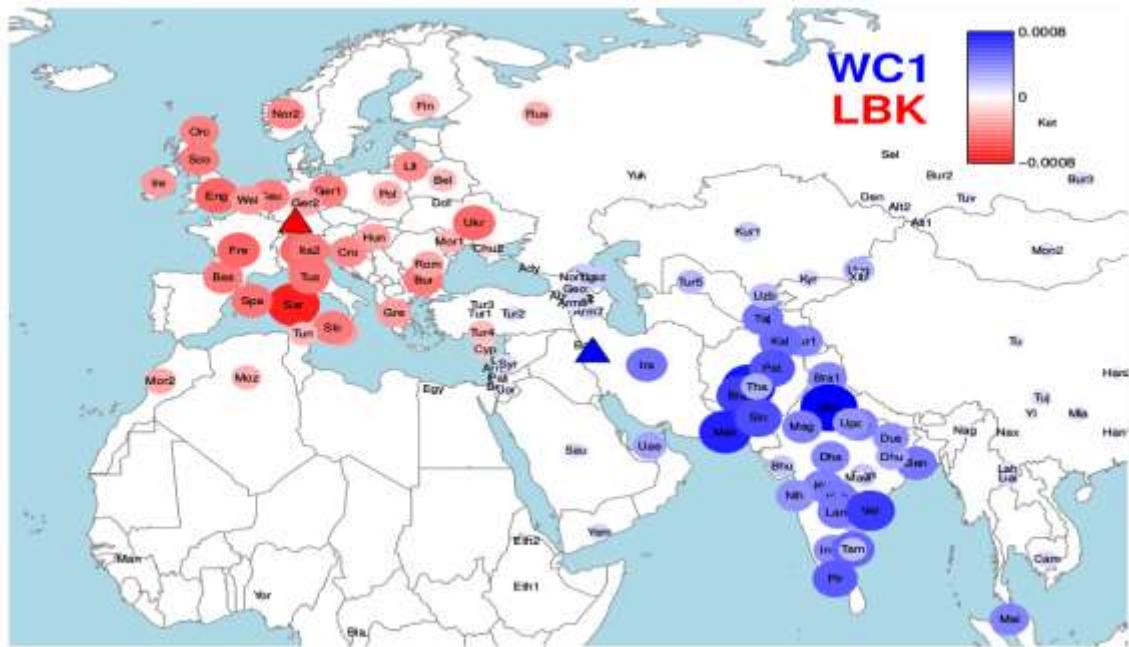
(b) Busby merge - WC1 vs LBK

Fig. S32

Difference in CHROMOPAINTER inferred painting profiles for pairwise comparison of WC1 (blue) and LBK (red) for the Lazaridis merge (top) and the Busby merge (bottom). The size and intensity of colour of the depicted circles corresponds to the degree of difference between the two samples being compared as indicated by the colour scale in the top right (e.g. a higher proportion of matching in WC1 is indicated with darker blue).
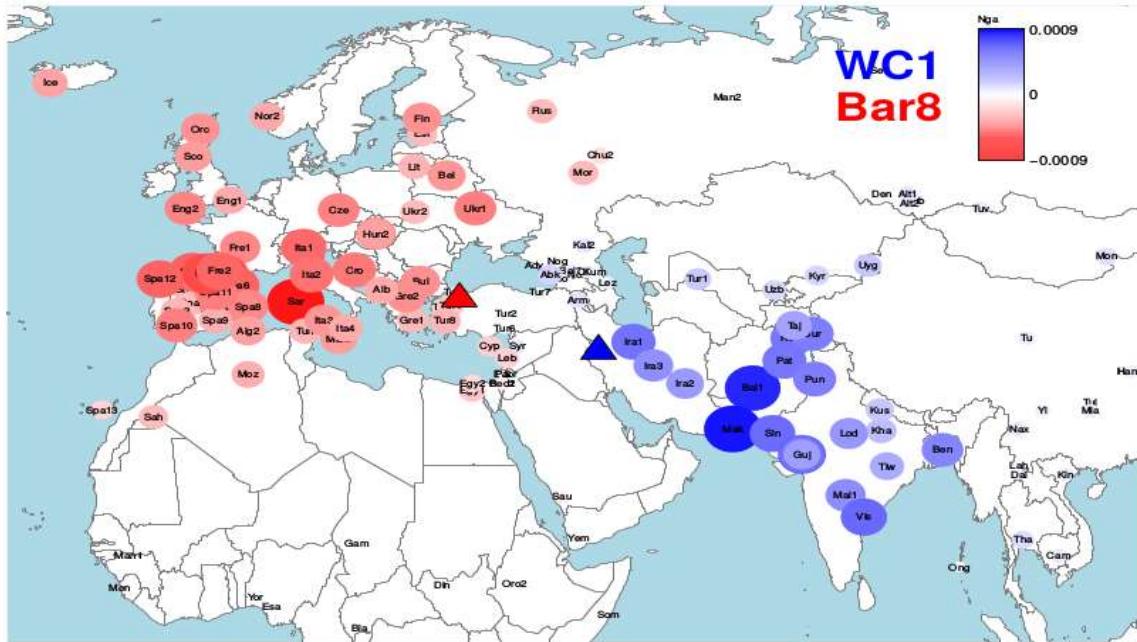
(a) Lazaridis merge - WC1 vs Bar8
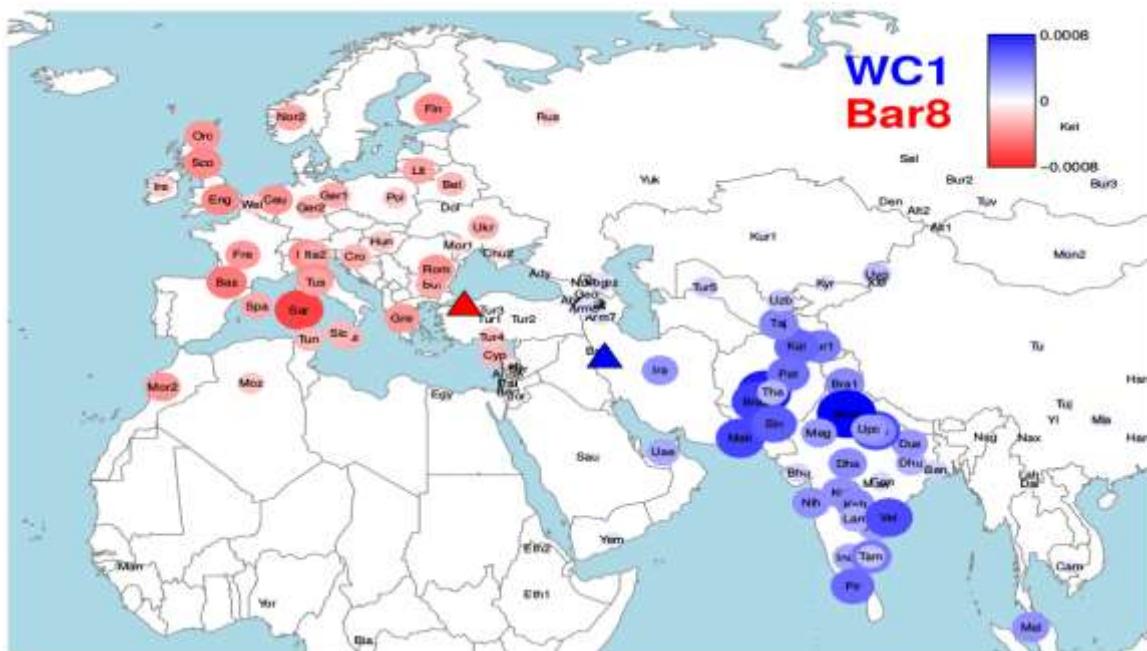
Fig. S33

Difference in CHROMOPAINTER inferred painting profiles for pairwise comparison of WC1 (blue) and Bar8 (red) for the Lazaridis merge (top) and the Busby merge (bottom).
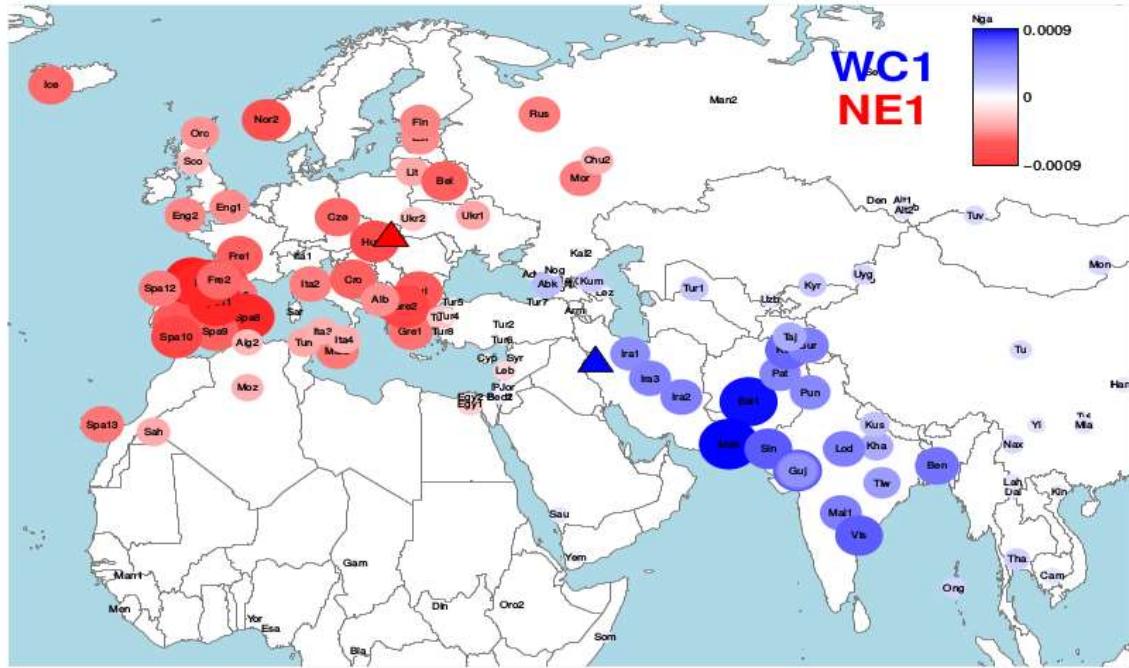
(a) Lazaridis merge - WC1 vs NE1
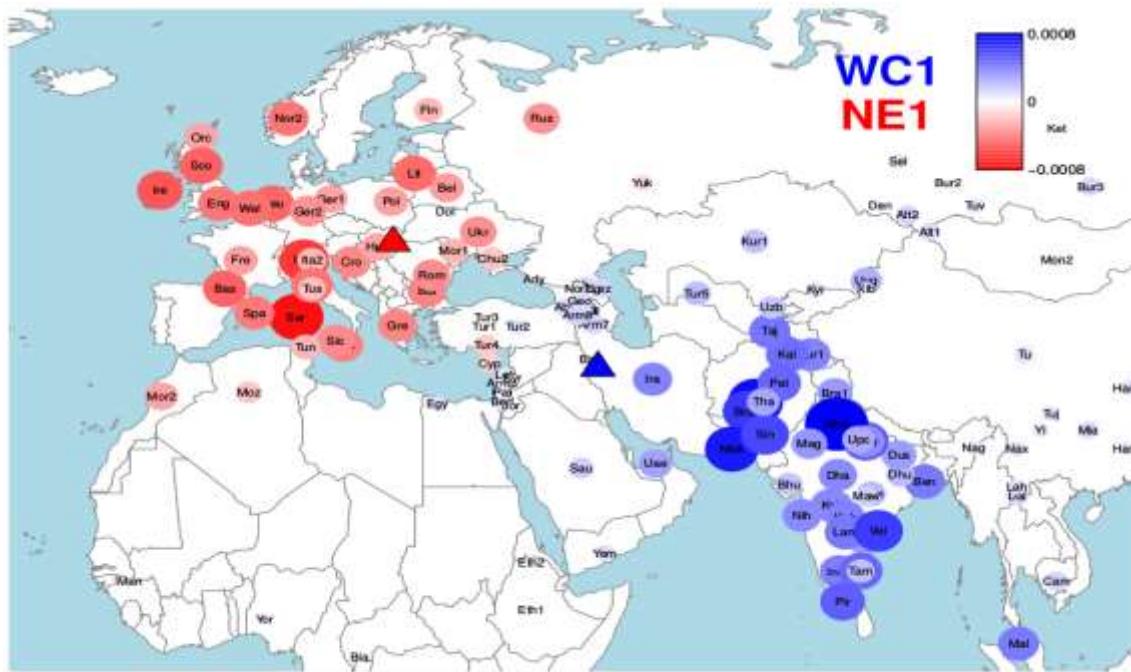
(b) Busby merge - WC1 vs NE1

Fig. S34
Difference in CHROMOPAINTER inferred painting profiles for pairwise comparison of
WC1 (blue) and NE1 (red) for the Lazaridis merge (top) and the Busby merge (bottom).

23
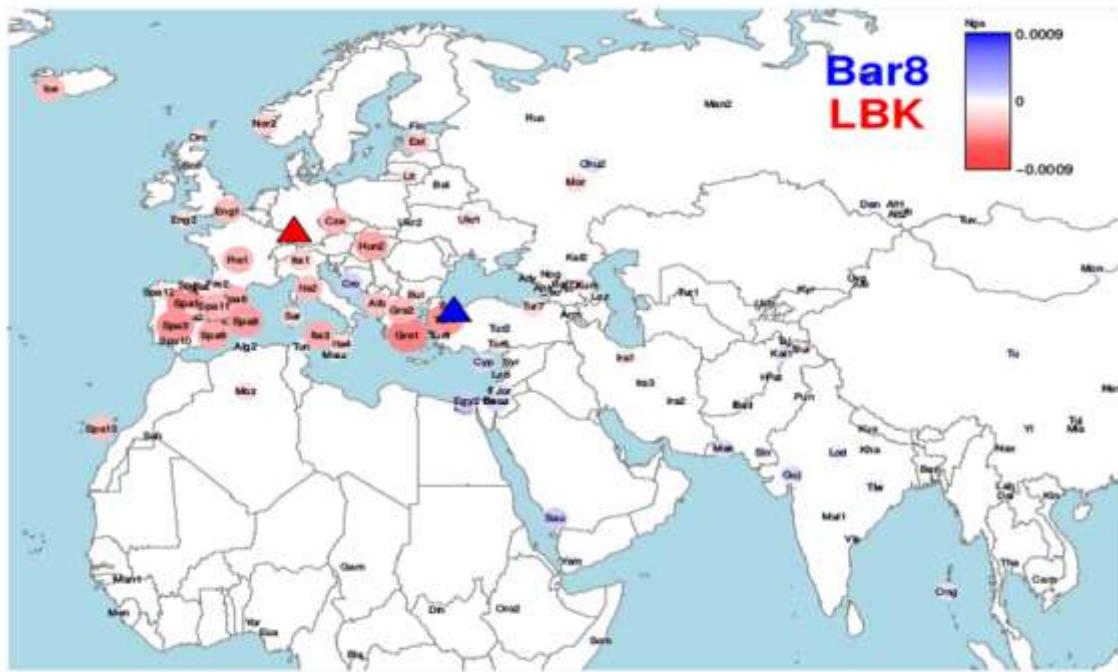
(a) Lazaridis merge - Bar8 vs LBK
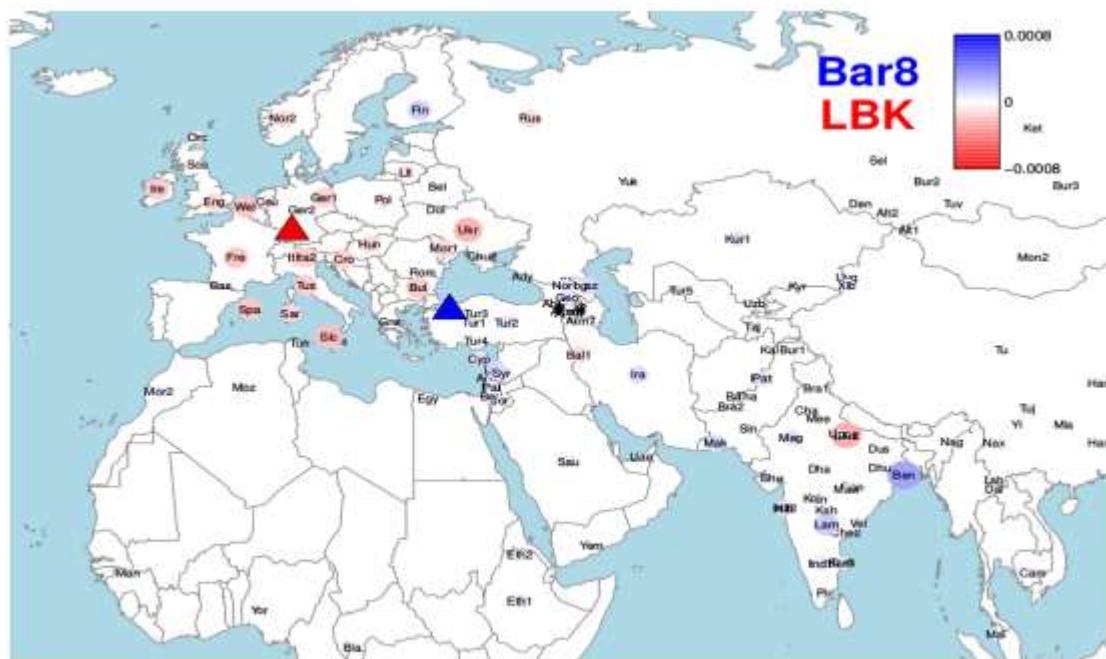
(b) Busby merge - Bar8 vs LBK

Fig. S35

Difference in CHROMOPAINTER inferred painting profiles for pairwise comparison of Bar8 (blue) and LBK (red) for the Lazaridis merge (top) and the Busby merge (bottom).
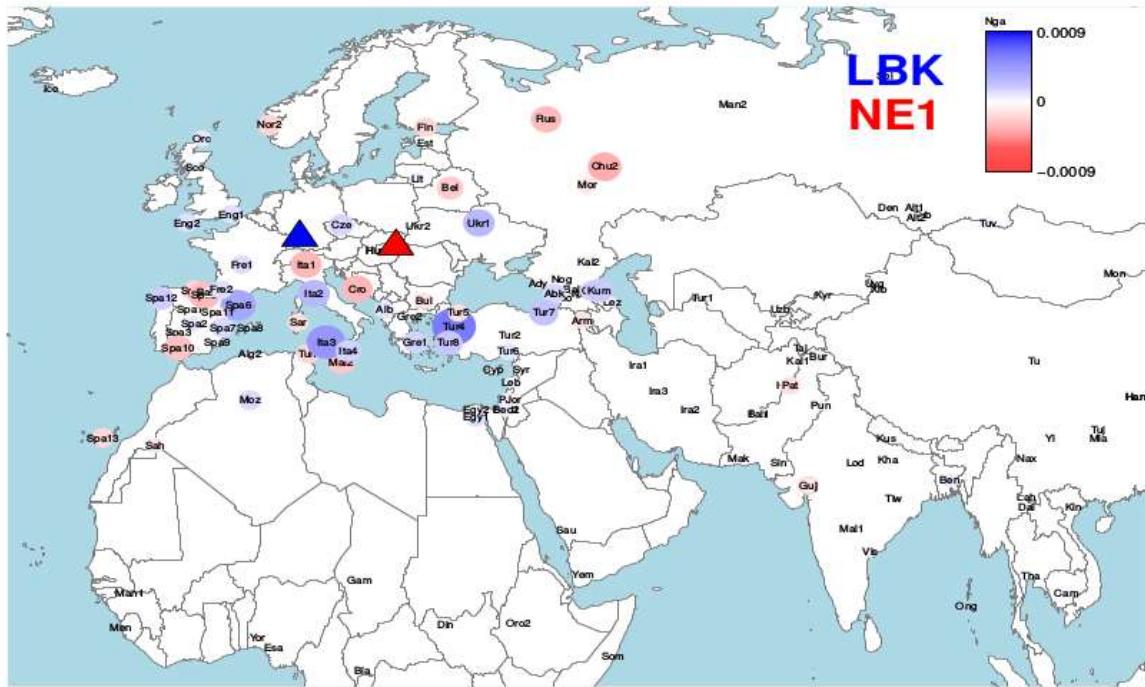
(a) Lazaridis merge - LBK vs NE1

(b) Busby merge - LBK vs NE1

Fig. S36
Difference in CHROMOPAINTER inferred painting profiles for pairwise comparison of
LBK (blue) and NE1 (red) for the Lazaridis merge (top) and the Busby merge (bottom).

(a) Lazaridis merge - Bar8 vs NE1

(b) Busby merge - Bar8 vs NE1

Fig. S37
Difference in CHROMOPAINTER inferred painting profiles for pairwise comparison of
Bar8 (blue) and NE1 (red) for the Lazaridis merge (top) and the Busby merge (bottom).

26

(a) Busby merge - WC1 vs LBK

(b) Busby merge - WC1 vs NE1

(c) Busby merge - LBK vs NE1

(d) Busby merge - WC1 vs NE1 - match coverage

(e) *TVD* Busby merge - modern donors (f) *TVD* Busby merge - African donors
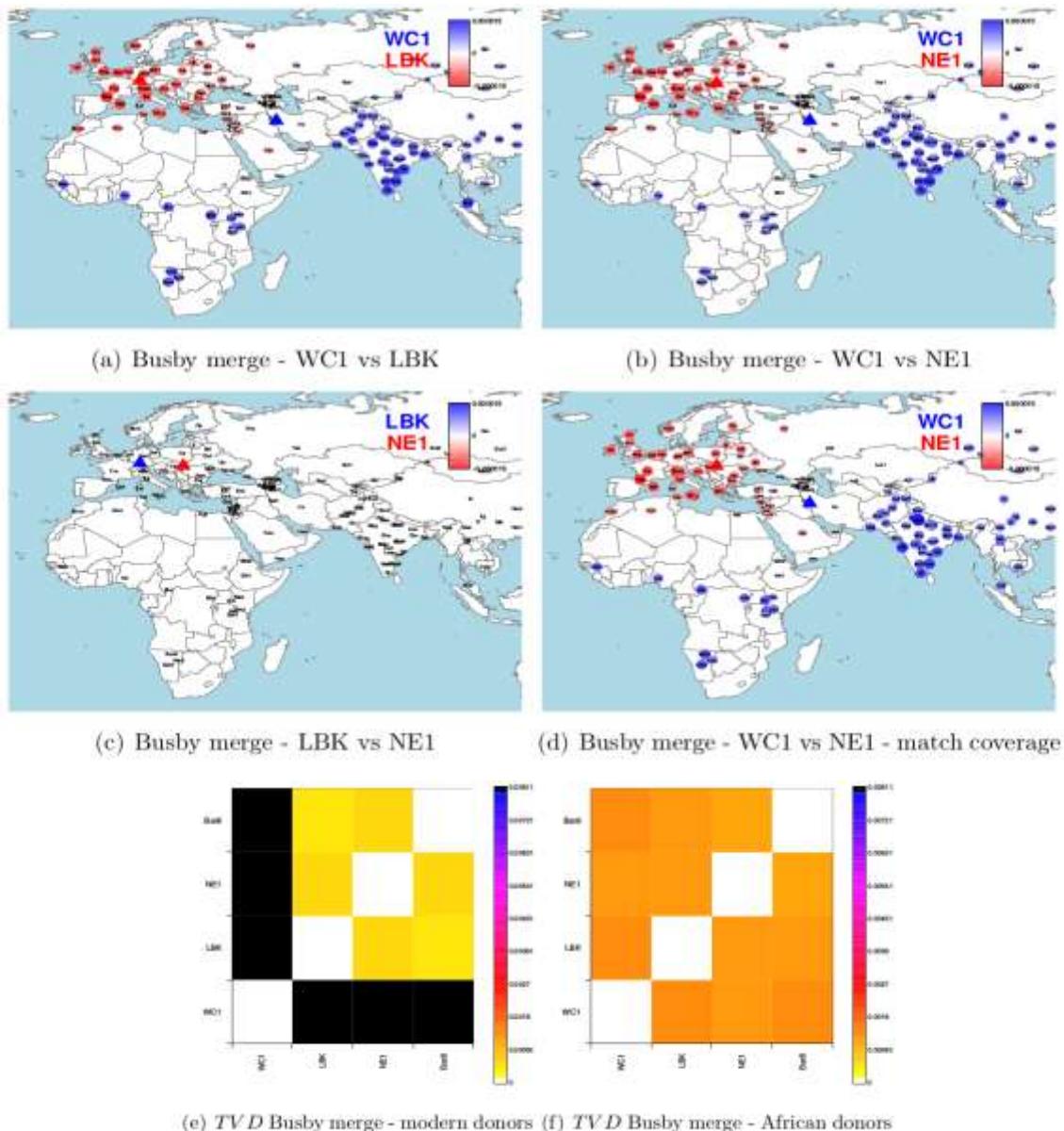
Fig. S38

(a)-(c) Differences in CHROMOPAINTER inferred allele frequency profiles for pairwise comparisons of *WC1*, *NE1* and *LBK* based on the merge described in the Section "Allele presence merge and inferring allele frequency profiles". The size and colour intensity of the depicted circles corresponds to the degree of difference between the two samples being compared as indicated by the colour scale in the top right. (d) as in (a)-(c) but with the coverage of NE1 matched to that of WC1. (e) Heatmap of the differences in inferred allele frequency profiles (TVD) amongst ancient samples *WC1*, *LBK*, *NE1*, *Bar8*. (f) Heatmap of the differences in inferred allele frequency profiles (TVD) amongst ancient samples *WC1, LBK, NE1, Bar8* when painted in relation to African donor individuals with no reported evidence of admixture (as listed in Fig. S31). Results are presented for the Busby merge but were consistent for the Lazaridis merge.

27
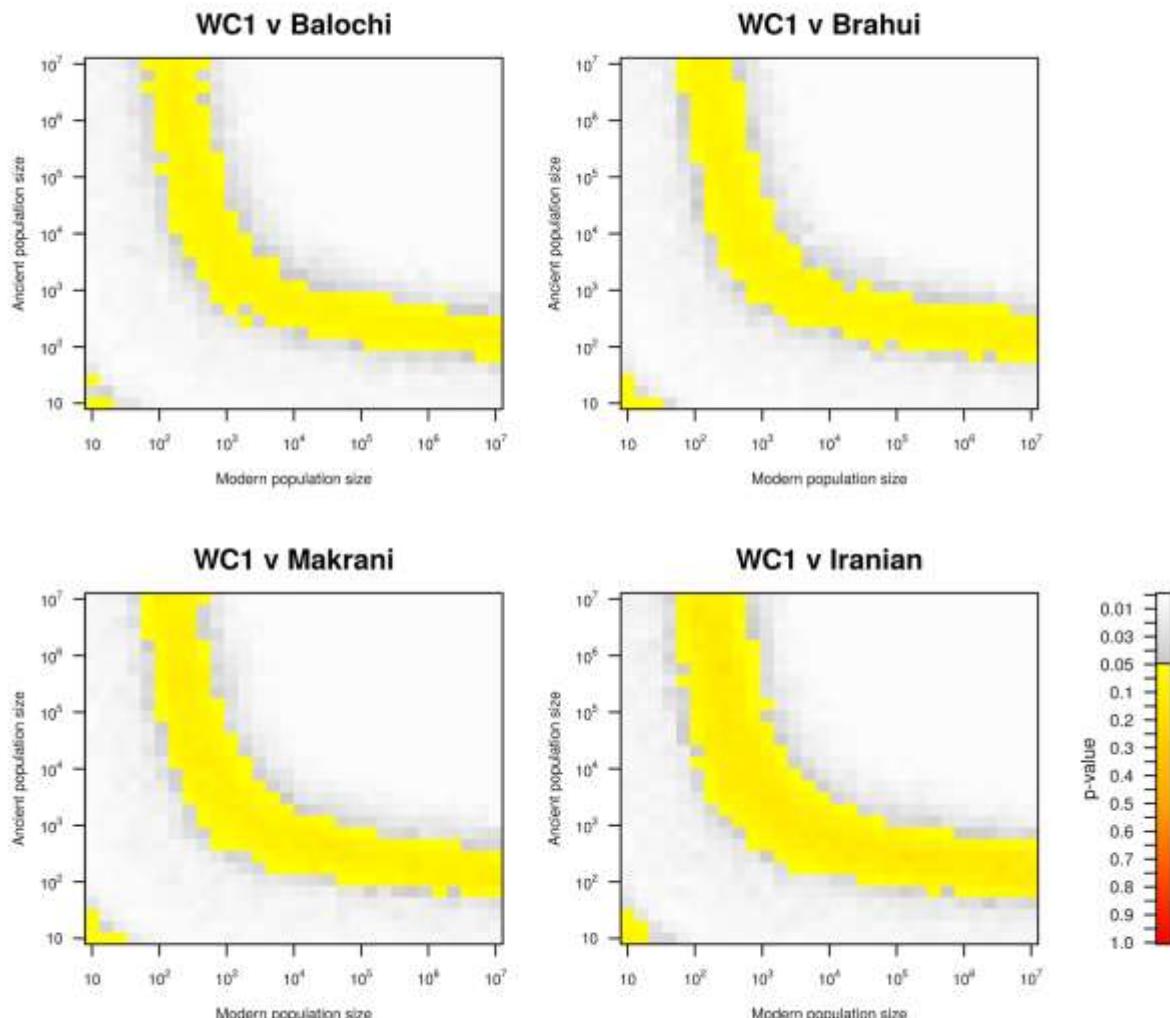
Fig. S39
Continuity grids between WC1 and modern Balochi, Brahui, Makrani and Iranian.
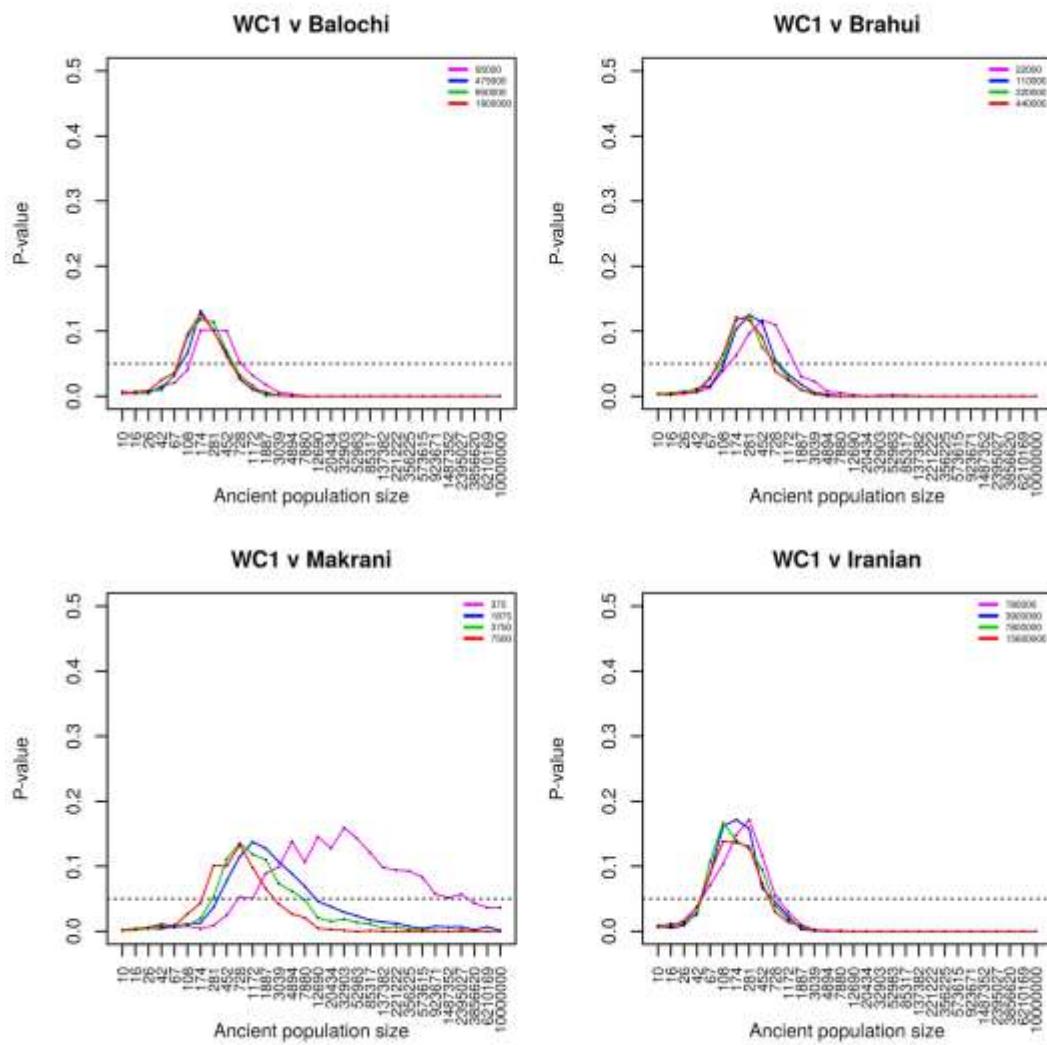
Fig. S40

Continuity grid slices between WC1 and modern Balochi, Brahui, Makrani and Iranian for 1, 5, 10 and 20% of their population size. Dotted lines represent a p-value threshold of 0.05 above which continuity cannot be rejected.
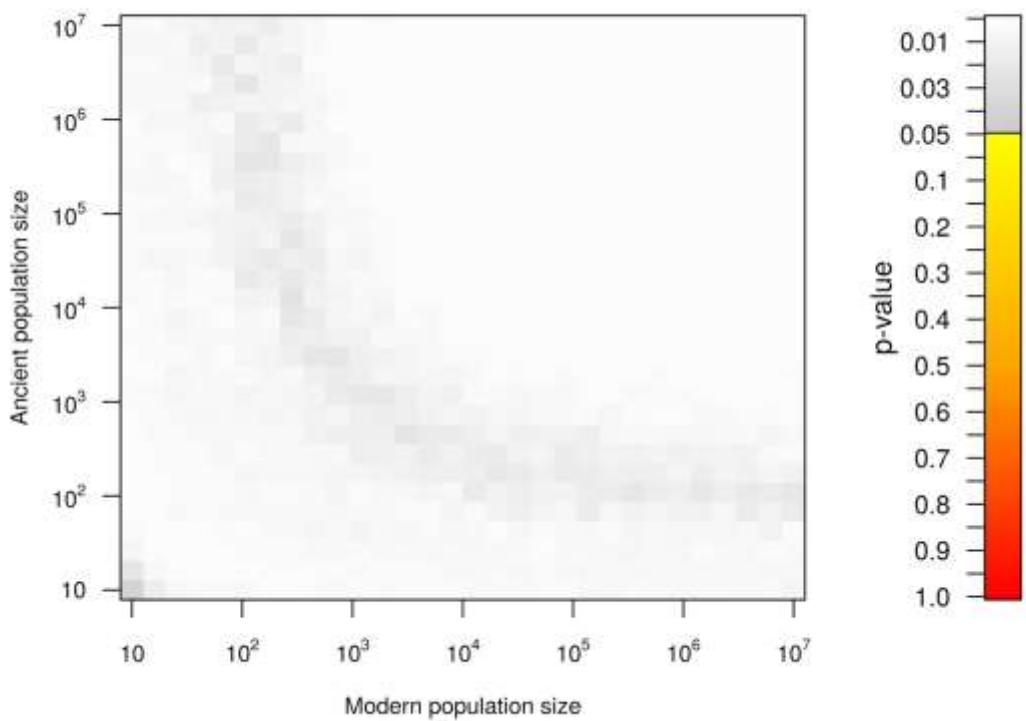
Fig. S41
Continuity grids between WC1 and Iranian Zoroastrians.

WC1 v Iran Zoroastrian
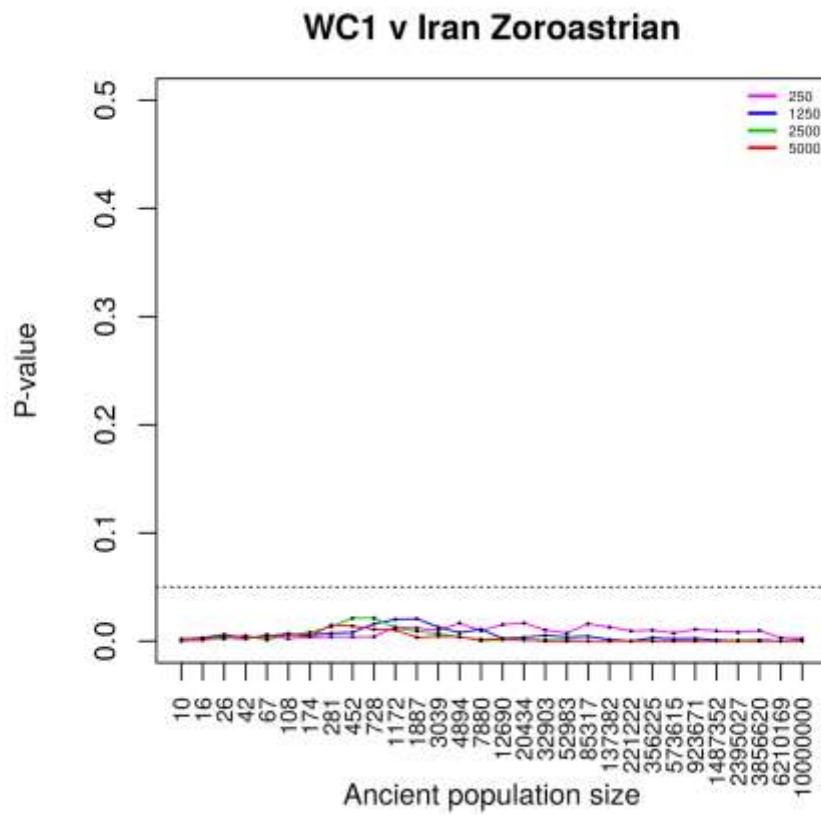
Fig. S42

Continuity grid slices between WC1 and Iranian Zoroastrians for 1, 5, 10 and 20% of their population size. Dotted lines represent a p-value threshold of 0.05 above which continuity cannot be rejected.
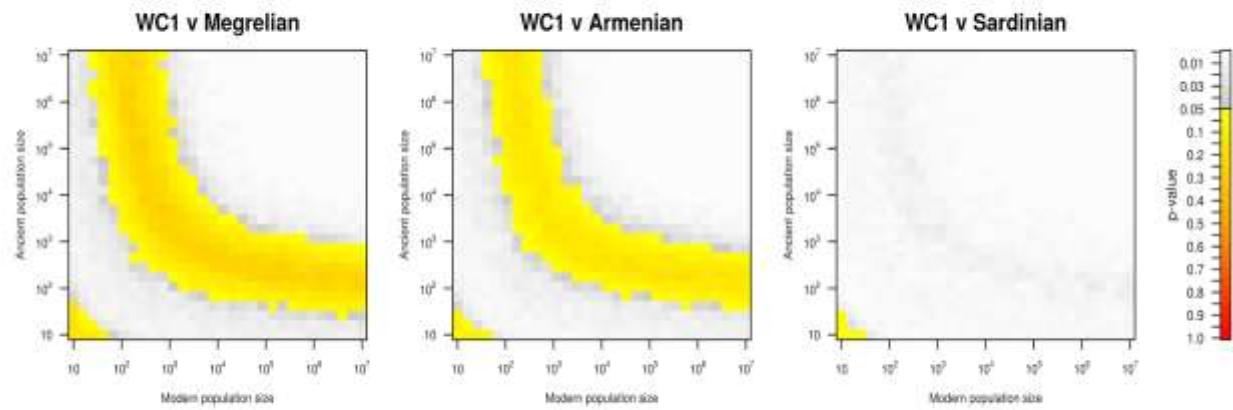
Fig. S43
Continuity grids between WC1 and modern Megrelians, Armenians and Sardinians.
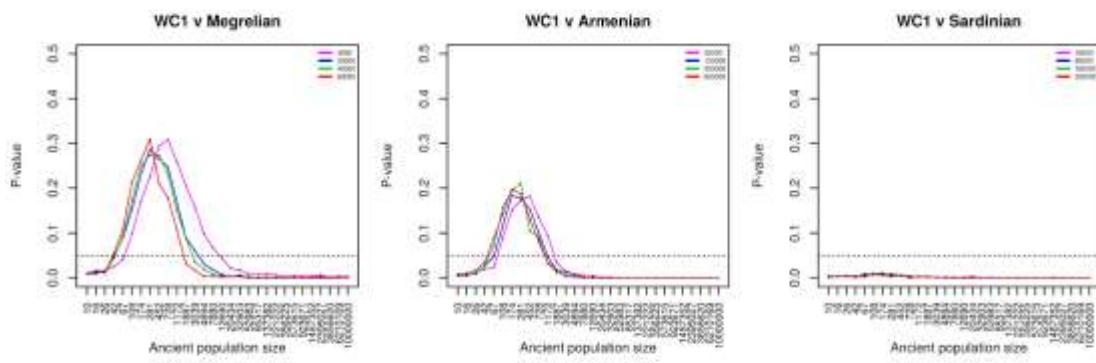
Fig. S44

Continuity grid slices between WC1 and modern Megrelians, Armenians and Sardinians for 1, 5, 10 and 20% of their population size. Dotted lines represent a p-value threshold of 0.05 above which continuity cannot be rejected.
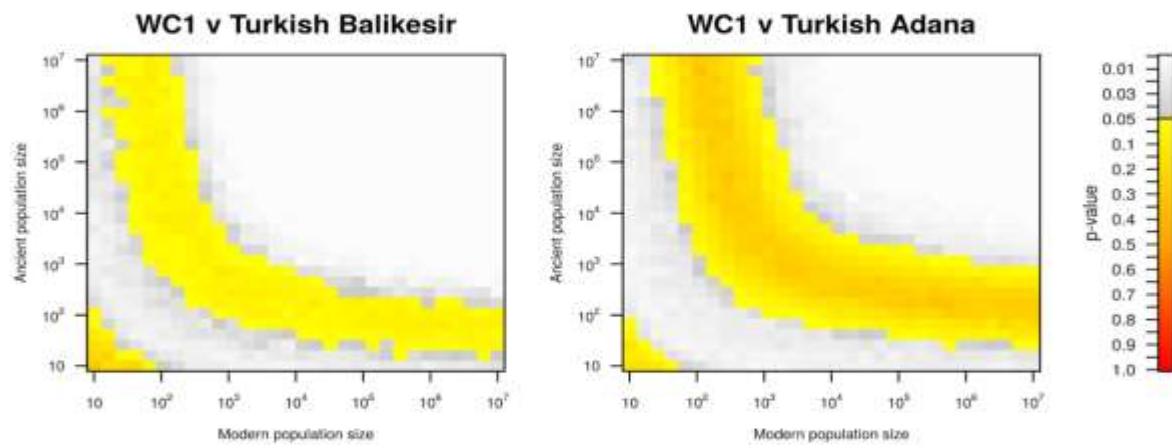
Fig. S45

Continuity grids between WC1 and modern Turkish.
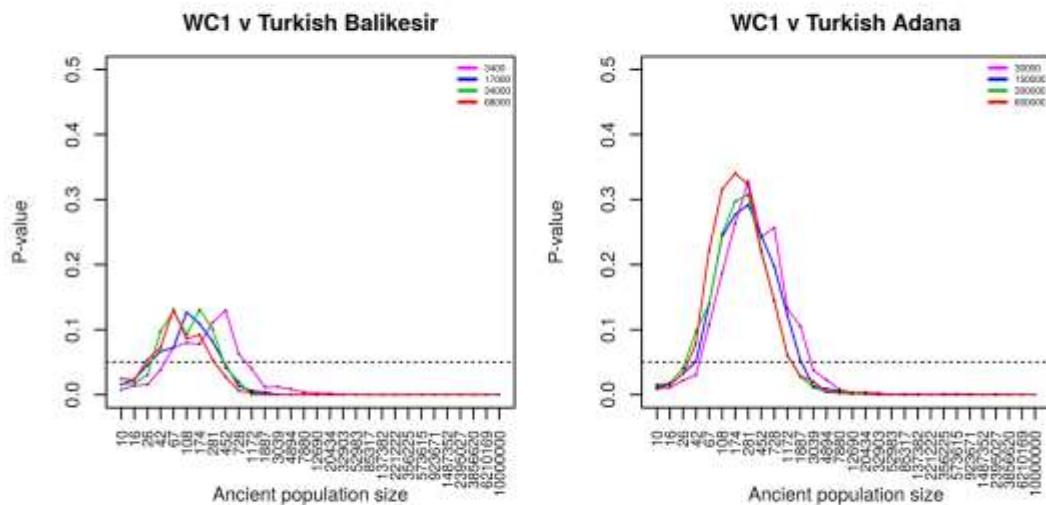
Fig. S46

Continuity grid slices between WC1 and modern Turkish for 1, 5, 10 and 20% of their population size. Dotted lines represent a p-value threshold of 0.05 above which continuity cannot be rejected.
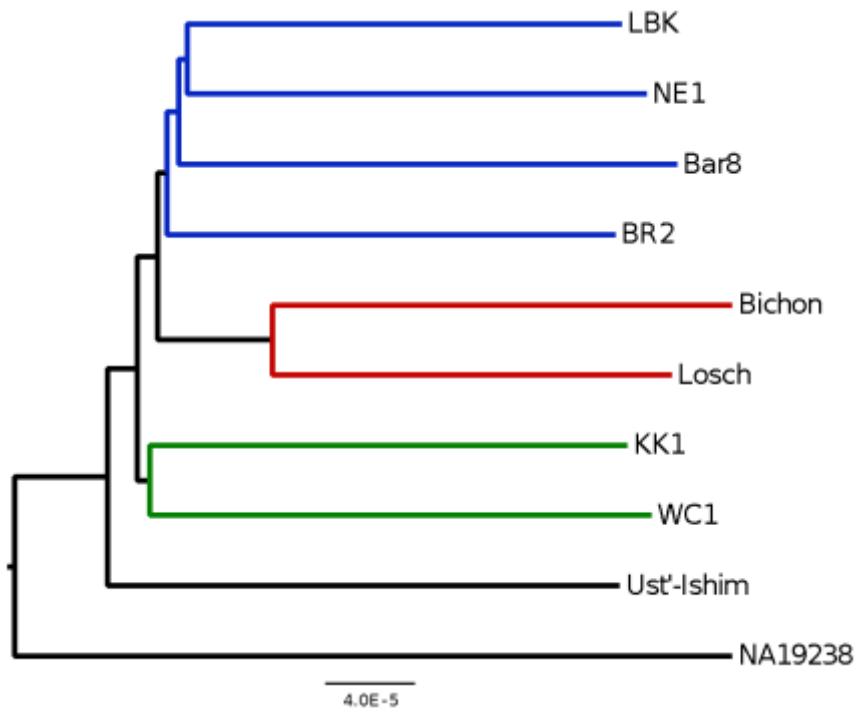
Fig. S47
Neighbor-Joining Tree for nine ancient and one modern genome.
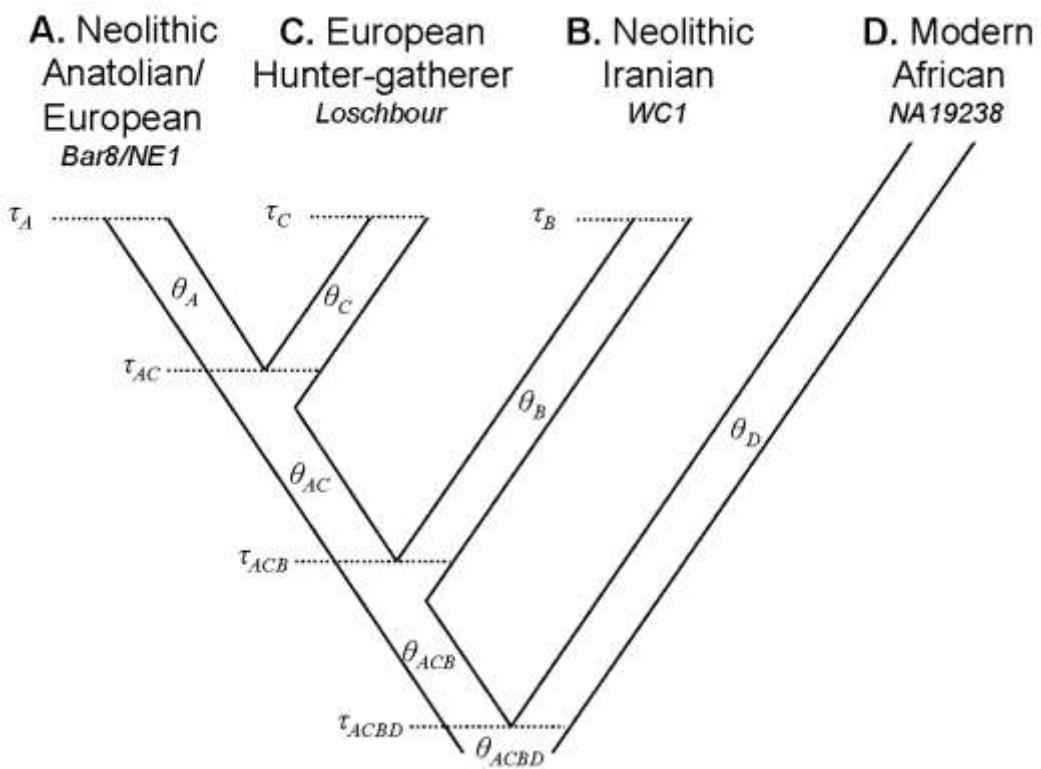
Fig. S48
G-PhoCS population model examined here.
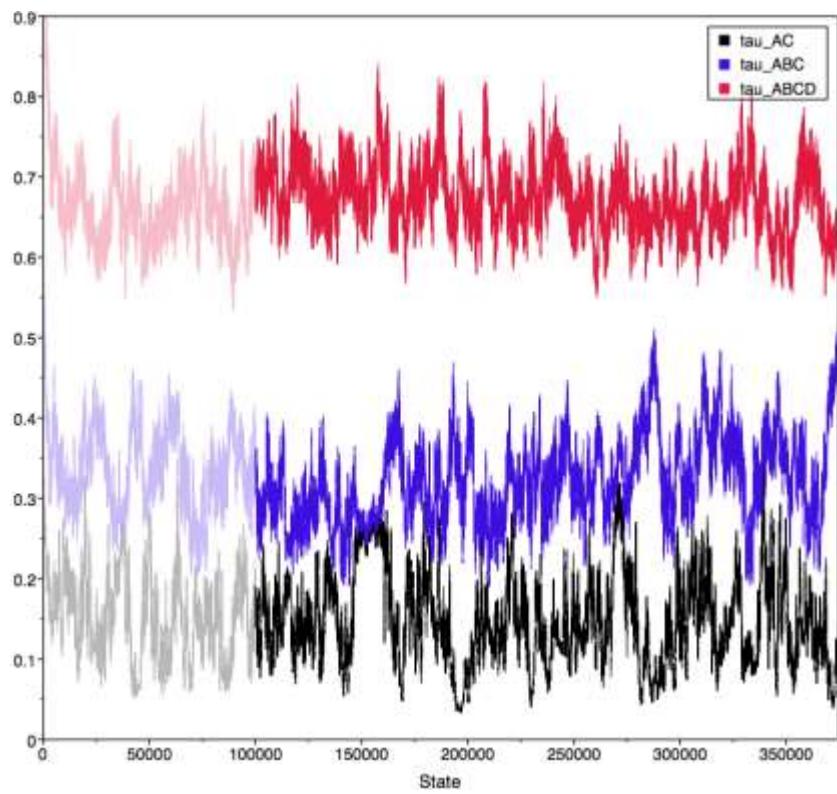
Fig. S49
MCMC trace for the τ parameters when including the Yoruba in the analyses.
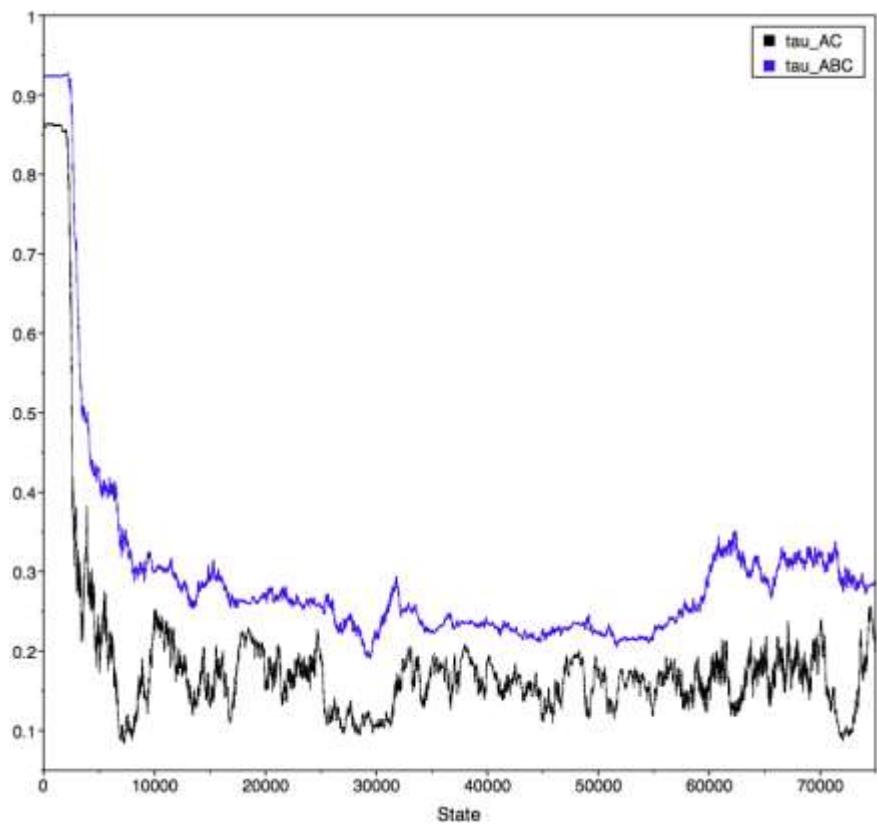
Fig. S50

MCMC trace for the τ parameters when excluding the Yoruba in the analyses. Note how the Markov chain starts at very high values close to 1.0 because of the prior, but quickly converges to values observed in Table S29.
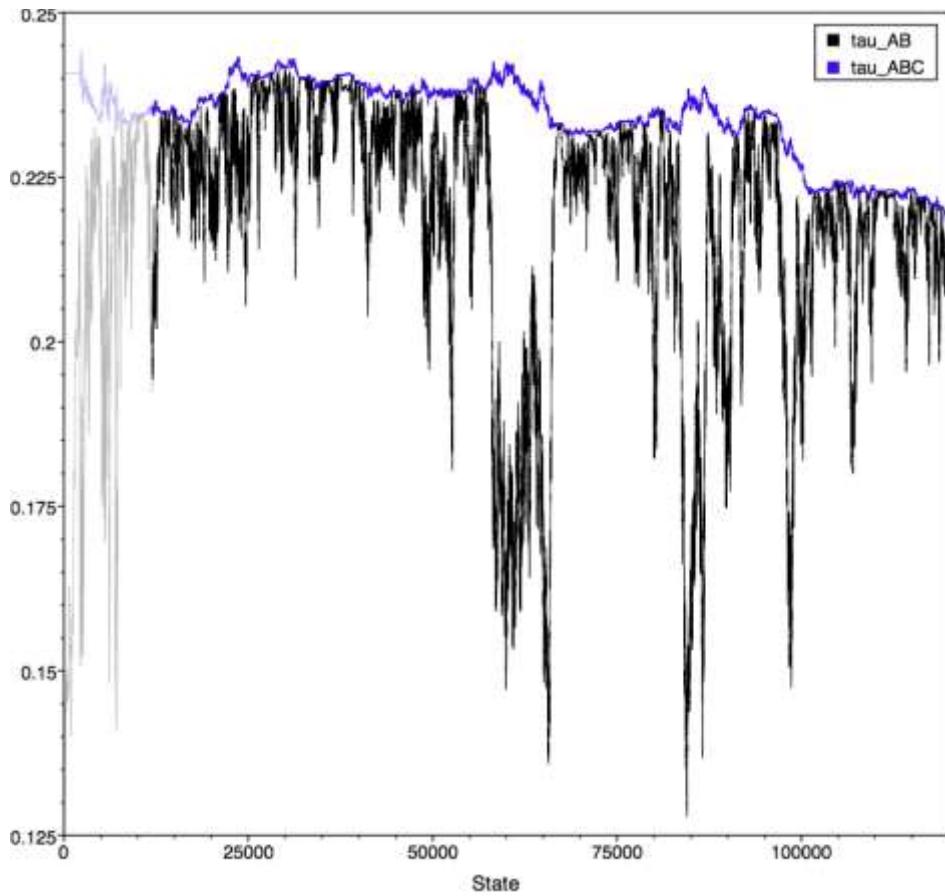
Fig. S51

MCMC trace for the τ parameters when excluding Yoruba in the analyses and allowing Bar8 and WC1 to be the most recent divergence event. Note how τ_{AB} (i.e. the proposed younger event) continues attempting the climb to values already observed for τ_{ABC} (i.e. the proposed older event)
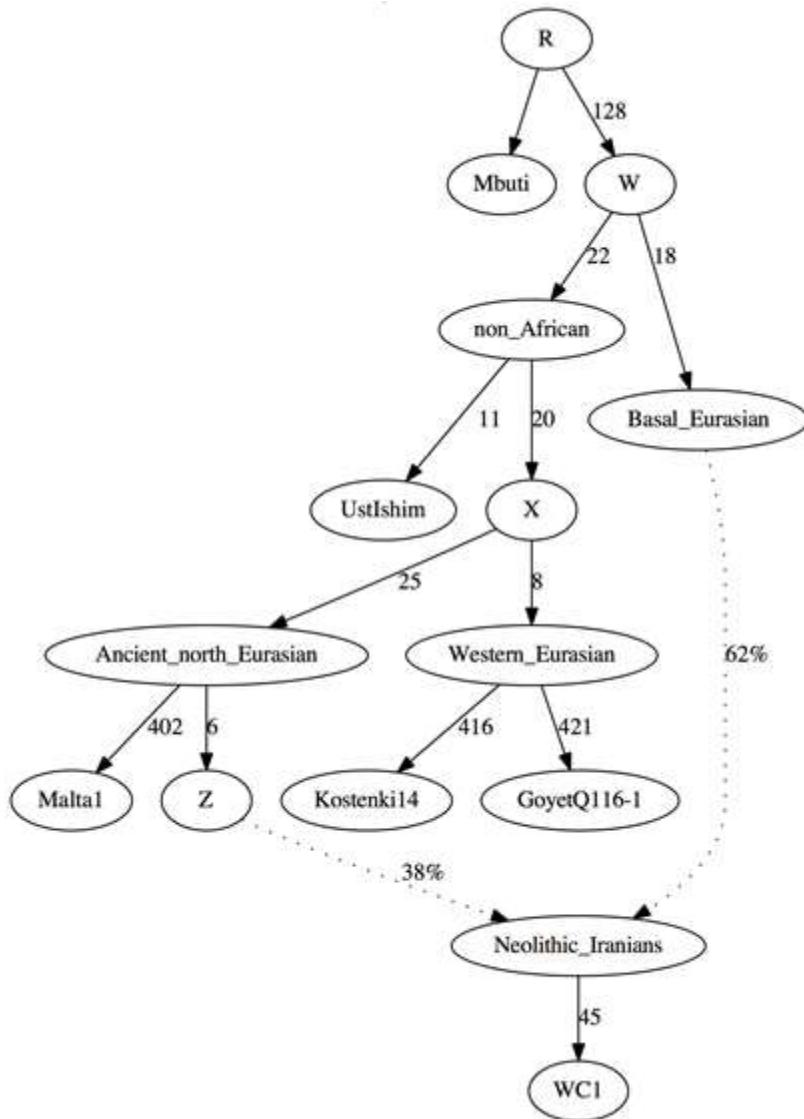
Fig. S52

Best fit Admixture graph when placing WC1 within Fu *et al.*'s base model (*14*). This model had no *f* statistics outliers (determined by fitted value being within 3 standard errors of estimated value)

SUPPLEMENTARY TABLES

Table S1
Samples analyzed. Dates calibrated using Oxcal v4.2.2 and the Intcal13 calibration curve. For details on ^{14}C dating see Table S2.

Site, region in Iran	Culture	Skeletal code	Sample ID	Age (cal. BCE)	Genomic coverage (+/-)	Stable isotope ratio			mtDNA hg	Sex	Yhg
						$\delta^{13}C(‰)$	$\delta^{15}N$ (‰)	C:N (atomic)			
Wezmeh Cave, Central Zagros	Early Neolithic	n-10	WC1	7,455-7,082	10.42 +/- 10.46	-19.6	11.1	3.2	J1d6	m	G2b
Tepe Abdul Hosein, Central Zagros	Early Neolithic	13030	AH1	-	1.226 +/- 1.736	-	-	-	R2	f	-
Tepe Abdul Hosein, Central Zagros	Early Neolithic	19001-SK#1	AH2	8,205-7,756	0.673 +/- 1.174	-19.7	9.2	3.2	R2	m	-
Tepe Abdul Hosein, Central Zagros	Early Neolithic	10035	AH4	8,204-7,755	0.907 +/- 1.017	-19.9	11.3	3.2	T2c	f	-
Tepe Hasanlu, Northern Zagros	Iron Age	HAS70	F38	971-832	1.941 +/- 2.295	-	-	-	N1a3a	m	R1b1a2a2
Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	caprine	WM2	-	-	-17.9	9.9	3.4	-	-	-

Table S2

^{14}C radiocarbon dates.

Sample	^{14}C laboratory ID	^{14}C age +/-	C:N %	collagen %	Calibration curve	Area enclosed %	^{14}C date cal. BCE
WC1	UBA-25840	8,296-8,184	3.2	2.1	OxCal 4.2	95.4 sigma 2	7,455-7,082
AH1	MAMS-25471	Not enough collagen					
AH2	MAMS-25472	8,874-8,792	3.2	5.6	OxCal 4.2	95.5 sigma 2	8,205-7,756
AH4	MAMS-25473	8,873-8,791	3.2	2.0	OxCal 4.2	95.4 sigma 2	8,204-7,755
F38	MAMS-22352	2,774-2,732	-	5.1	OxCal 4.2	95.4 sigma 2	971-832

Table S3

Human and faunal bone samples and isotope data.

Sample ID	Site, region in Iran	Culture	Taxa	δ¹³C	% C	δ¹⁵N	% N	C:N atomic ratio	Collagen yield (wt%)
Successful samples									
WC1	Wezmeh Cave, Central Zagros	Early Neolithic	*Homo sapiens*	-19.6 +/- 0.1	47.0 +/- 3.3	11.1 +/- 0.2	17.0 +/- 1.2	3.2	6.7
AH2	Tepe Abdul Hosein, Central Zagros	Early Neolithic	*Homo sapiens*	-19.7 +/- 0.1	46.6 +/- 1.8	9.2 +/- 0.3	17.0 +/- 0.7	3.2	-
AH4	Tepe Abdul Hosein, Central Zagros	Early Neolithic	*Homo sapiens*	-19.8 +/- 0.1	40.2 +/- 0.9	11.3 +/- 0.2	14.7 +/- 0.3	3.2	-
WM2	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Caprini*	-17.9 +/- 0.1	45.5 +/- 7.5	9.9 +/- 0.1	15.7 +/- 0.1	3.4	0.6
Unsuccessful samples									
WM6	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Caprini*	No collagen preserved					
WM7	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Cervus elaphus*	No collagen preserved					
WM8	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Cervus elaphus*	No collagen preserved					
WM12	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Vulpes vulpes*	No collagen preserved					
WM13	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Ursus actos*	No collagen preserved					
WM14	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Crocuta spelea*	No collagen preserved					
WM15	Wezmeh Cave, Central Zagros	Palaeolithic/ Neolithic	*Crocuta*	No collagen preserved					

Table S4

MiSeq screening results. LK = library control; EK = extraction control; HK = milling control. Library controls highlighted in blue correspond to libraries which have not been screened and were directly sequenced deeper. PCR cycles = number of cycles used to amplify the library. Raw reads = number of raw reads in a screened pool corresponding to each library; Mapped-hg19 = reads aligned to the human genome (hg19); Endog % = endogenous DNA content; Copy number = molecule number of each library.

Sample	Element	Extraction code	Library name	Library code	PCR cycles	Raw reads	Mapped hg19	Endog (%)	Copy number
WC1	Metatarsal	A	WC1-10.18	10	18	681174	354290	52.3	2.25E+07
WC1	Metatarsal	B	WC1-13.5	13	17	572498	266705	47.3	7.09E+07
WC1	Metatarsal	B	WC1-13.10	13	17	496596	231079	47.2	8.73E+07
WC1	Metatarsal	A	WC1-12.5	12	16	506716	219549	43.7	5.33E+07
AH1	petrous bone	D	AH1-27.1	27	15	712963	130293	18.5	1.79E+08
AH1	petrous bone	D	AH1-29.14	29	14	933722	199789	21.95	1.67E+08
AH1	petrous bone	D	AH1-29.16	29	14	1124836	251724	22.93	2.09E+08
AH1	petrous bone	D	AH1-30.9	30	15	704211	155713	22.63	1.87E+08
AH2	petrous bone	D	AH2-27.2	27	15	471874	103928	21.27	5.73E+07
AH2	petrous bone	D	AH2-29.11	29	14	766129	183063	24.45	9.23E+07
AH2	petrous bone	E	AH2-29.8	29	14	951425	258743	27.77	6.61E+07
AH2	petrous bone	F	AH2-30.5	30	15	798747	214635	27.4	5.25E+07
AH4	petrous bone	D	AH4-27.4	27	15	539277	78846	14.8	3.06E+08
AH4	petrous bone	D	AH4-29.1	29	14	1024398	153902	15.4	4.04E+08
AH4	petrous bone	F	AH4-29.3	29	14	875563	147719	17.29	3.23E+08
AH4	petrous bone	F	AH4-29.4	29	14	927694	178769	19.72	3.06E+08
AH4	petrous bone	F	AH4-30.3	30	15	968053	156420	16.59	2.42E+08
F38	Femur	B	F38-13.1	13	17	470817	189355	40.27	1.97E+07
F38	Femur	B	F38-13.2	13	17	605547	218901	36.42	2.17E+07
HK1	-	C	FL31.4	31	16	32708	147	-	5.62E+04
HK2	-	D	FL27.9	27	15	9506	161	-	4.87E+04
EK1	-	C	FL31.3	31	16	22358	67	-	2.38E+04
EK2	-	G	FL31.2	31	16	28015	65	-	6.69E+04
EK3	-	H	FL31.1	31	16	20839	40	-	3.00E+04
EK4	-	A	FL31.5	31	16	5112	8	-	7.81E+04
EK5	-	B	FL31.6	31	16	6545	8	-	8.91E+04
EK6	-	D	FL27.8	27	15	16941	22	-	7.80E+04
EK7	-	F	FL29.17	29	14	637	0	-	8.53E+04
EK8	-	E	FL28.10	28	15	21	0	-	5.49E+04
LK1	-	-	FL10.11	10	18	1	0	-	3.16E+04
LK2	-	-	FL13.11	13	17	5573	7	-	1.10E+05
LK3	-	-	FL12.8	12	16	22150	58	-	2.62E+04
LK4	-	-	FL15.7	15	15	7438	15	-	2.79E+04
LK5	-	-	FL20.11	20	13	13620	12	-	3.22E+04
LK6	-	-	FL27.10	27	15	12990	9	-	6.37E+04
LK7	-	-	FL29.18	29	14	426	2	-	8.43E+04
LK8	-	-	FL30.10	30	15	4038	5	-	4.82E+04
LK9	-	-	FL28.11	28	15	17519	12	-	3.99E+04
LK10	-	-	FL5.10	5	18	0	0	-	5.78E+04
LK11	-	-	FL16.10	16	15	0	0	-	1.60E+05

Table S5

Molecule number of all controls and libraries. Samples that were screened are highlighted in green and samples that were only deeply sequenced are not highlighted. LK = library control; EK= extraction control; HK= milling control; NH= positive control; PCR cycles = number of cycles used to amplify the library; Copy number = molecule number of each library. Samples showing the same extraction code were extracted in parallel.

Sample	Extraction code	Library name	Library code	PCR cycles	Copy number	Sample	Extraction code	Library name	Library code	PCR cycles	Copy number
WC1	A	WC1-10.18	10	18	2,25E+07	F38	B	F38-13.1	13	17	1,97E+07
WC1	A	WC1-12.5	12	16	5,33E+07	F38	B	F38-15.4	15	15	7,36E+07
WC1	A	WC1-C	20	13	1,34E+07	F38	B	F38-13.2	13	17	2,17E+07
WC1	A	WC1-D	20	13	1,37E+07	F38	B	F38-15.5	15	15	2,19E+08
WC1	A	WC1-E	20	13	1,56E+07	F38	H	F38-16.1	16	15	5,90E+07
WC1	B	WC1-13.5	13	17	7,09E+07	F38	H	F38-16.2	16	15	5,33E+07
WC1	B	WC1-13.10	13	17	8,73E+07	F38	H	F38-16.3	16	15	6,76E+07
WC1	B	WC1-15.1	15	15	1,09E+08	HK1	C	FL31.4	31	16	5,62E+04
WC1	B	WC1-15.2	15	15	4,46E+07	HK2	D	FL27.9	27	15	4,87E+04
WC1	B	WC1-15.3	15	15	1,59E+08	EK1	C	FL31.3	31	16	2,38E+04
WC1	C	WC1-L	20	13	1,86E+07	EK2	G	FL31.2	31	16	6,69E+04
WC1	G	WC1-F	20	13	1,72E+06	EK3	H	FL31.1	31	16	3,00E+04
WC1	G	WC1-G	20	13	1,08E+09	EK4	A	FL31.5	31	16	7,81E+04
WC1	G	WC1-H	20	13	1,88E+06	EK5	B	FL31.6	31	16	8,91E+04
WC1	H	WC1-I	20	13	6,98E+06	EK6	D	FL27.8	27	15	7,80E+04
WC1	H	WC1-J	20	13	6,00E+06	EK7	F	FL29.17	29	14	8,53E+04
WC1	H	WC1-K	20	13	8,05E+06	EK8	E	FL28.10	28	15	5,49E+04
AH1	D	AH1-27.1	27	15	1,79E+08	LK1		FL10.11	10	18	3,16E+04
AH1	D	AH1-29.14	29	14	1,67E+08	LK2		FL13.11	13	17	1,10E+05
AH1	D	AH1-29.15	29	14	1,60E+08	LK3		FL12.8	12	16	2,62E+04
AH1	D	AH1-29.16	29	14	2,09E+08	LK4		FL15.7	15	15	2,79E+04
AH1	D	AH1-30.9	30	15	1,87E+08	LK5		FL20.11	20	13	3,22E+04
AH2	D	AH2-27.2	27	15	5,73E+07	LK6		FL27.10	27	15	6,37E+04
AH2	D	AH2-29.10	29	14	6,58E+07	LK7		FL29.18	29	14	8,43E+04
AH2	D	AH2-29.11	29	14	9,23E+07	LK8		FL30.10	30	15	4,82E+04
AH2	D	AH2-29.9	29	14	4,22E+07	LK9		FL28.11	28	15	3,99E+04
AH2	D	AH2-30.7	30	15	4,04E+07	LK10		FL5.10	5	18	5,78E+04
AH2	D	AH2-30.8	30	15	7,01E+07	LK11		FL16.10	16	15	1,60E+05
AH2	E	AH2-28.3	28	15	3,06E+07	NH1		FL10.12	10	18	2,27E+09
AH2	E	AH2-29.6	29	14	5,81E+07	NH2		FL13.12	13	17	2,15E+09
AH2	E	AH2-29.8	29	14	6,61E+07	NH3		FL12.9	12	16	1,42E+09
AH2	F	AH2-29.12	29	14	4,22E+07	NH4		FL15.8	15	15	2,97E+09
AH2	F	AH2-30.4	30	15	5,49E+07	NH5		FL20.12	20	13	9,60E+08
AH2	F	AH2-30.5	30	15	5,25E+07	NH6		FL5.11	5	18	1,01E+10
AH4	D	AH4-27.4	27	15	3,06E+08	NH7		FL27.11	27	15	3,71E+09
AH4	D	AH4-29.1	29	14	4,04E+08	NH8		FL29.19	29	14	6,32E+09
AH4	D	AH4-29.2	29	14	2,75E+08	NH9		FL30.11	30	15	5,80E+09
AH4	D	AH4-30.1	30	15	3,20E+08	NH10		FL28.12	28	15	6,51E+09
AH4	F	AH4-29.3	29	14	3,23E+08	NH11		FL31.7	31	16	2,88E+09
AH4	F	AH4-29.4	29	14	3,06E+08	NH12		FL16.11	16	15	2,26E+10
AH4	F	AH4-29.5	29	14	2,24E+08						
AH4	F	AH4-30.2	30	15	1,65E+08						
AH4	F	AH4-30.3	30	15	2,42E+08						

Table S6

Contamination fraction C for screened libraries and corresponding blank controls.

Screened library	MAP.S / (TR x Ng.S)	Corresponding blank control	MAP.B / (TR x Ng.B)	Contamination fraction
WC1-10.18	0.00741383	LK1	0	0
WC1-12.5	0.003608187	LK3	3.57E-01	0.0989
WC1-13.5	0.005581037	LK2	4.00E-03	0.0007
WC1-13.10	0.004835532	LK2	4.00E-03	0.0008
F38-13.1	0.00396242	LK2	4.00E-03	0.0010
F38-13.2	0.004580696	LK2	4.00E-03	0.0009
AH1-27.1	0.001834169	LK6	1.44E-01	0.0785
		HK2	8.59E-01	0.4681
		EK6	3.52E-01	0.1919
AH2-27.2	0.001463022	LK6	1.44E-01	0.0984
		HK2	8.59E-01	0.5868
		EK6	3.52E-01	0.2406
AH4-27.4	0.001109936	LK6	1.44E-01	0.1297
		HK2	8.59E-01	0.7735
		EK6	3.52E-01	0.3171
AH1-29.14	0.005094796	LK7	3.20E-02	0.0063
		EK7	0	0
AH1-29.16	0.006419185	LK7	3.20E-02	0.0050
		EK7	0	0
AH2-29.11	0.004668269	LK7	3.20E-02	0.0069
		EK7	0	0
AH2-29.8	0.006598176	LK7	3.20E-02	0.0048
		EK7	0	0
AH4-29.1	0.003924637	LK7	3.20E-02	0.0082
		EK7	0	0
AH4-29.3	0.003766965	LK7	3.20E-02	0.0085
		EK7	0	0
AH4-29.4	0.004558768	LK7	3.20E-02	0.0070
		EK7	0	0
AH1-30.9	0.003970819	LK8	6.67E-02	0.0168
AH2-30.5	0.005473383	LK8	6.67E-02	0.0122
AH4-30.3	0.003988849	LK8	6.67E-02	0.0167

Table S7

Sequencing and pooling strategy for deep shotgun sequencing.

Sample	Library	Pool / Single	Sequencing machine	Number of lanes	Single end / Paired end
WC1	wc1-10.18				
WC1	wc1-13.5				
WC1	wc1-13.10				
WC1	wc1-12.5	pooled	Illumina HiSeq 2500	6	paired end
WC1	wc1-15.1				
WC1	wc1-15.2				
WC1	wc1-15.3				
WC1	WC1-C				
WC1	WC1-D				
WC1	WC1-E				
WC1	WC1-F				
WC1	WC1-G	pooled	Illumina HiSeq 2500	2	paired end
WC1	WC1-H				
WC1	WC1-I				
WC1	WC1-J				
WC1	WC1-K				
WC1	WC1-L				
AH1	AH1-27.1				
AH1	AH1-29.14				
AH1	AH1-29.15				
AH1	AH1-29.16				
AH1	AH1-30.9				
AH2	AH2-27.2				
AH2	AH2-28.3				
AH2	AH2-29.10				
AH2	AH2-29.11				
AH2	AH2-29.12				
AH2	AH2-29.6				
AH2	AH2-29.8				
AH2	AH2-29.9	pooled	Illumina HiSeq 2500	8 (2 x 4)	single end
AH2	AH2-30.4				
AH2	AH2-30.5				
AH2	AH2-30.7				
AH2	AH2-30.8				
AH4	AH4-27.4				
AH4	AH4-29.1				
AH4	AH4-29.2				
AH4	AH4-29.3				
AH4	AH4-29.4				
AH4	AH4-29.5				
AH4	AH4-30.1				
AH4	AH4-30.2				
AH4	AH4-30.3				
F38	F38-13.1				
F38	F38-15.4				
F38	F38-13.2				
F38	F38-15.5	pooled	Illumina HiSeq 2500	2	paired end
F38	F38-16.1				
F38	F38-16.2				
F38	F38-16.3				

Table S8

Summary of results for deep shotgun sequencing per sample. %1x cov.hg19 = percentage of the genome which was covered at least once. %2x cov.hg19 = percentage of the genome which was covered at least twice. mD5 = C/T deamination rate at the first position at 5`-end. mD3 = G/A deamination rate at the first position at 3`-end.

Sample	Raw reads	Joined paired end reads	Mapped to hg19	Endogenous DNA %	Mean fragment length	Coverage (+/- SD)	% 1xcov hg19	% 2xcov hg19	mD5	mD3
WC1	3,348,604,194	1,197,310,513	622,993,765	45.68	97.46	10.42 +/- 10.46	89.23	87.94	0.23	0.24
AH1	393,621,637	-	77,311,197	22.95	54.87	1.226 +/- 1-736	59.64	32.9	0.33	0.35
AH2	158,202,579	-	37,029,436	27.72	58.56	0.673 +/- 1.174	42.97	15.3	0.32	0.3
AH4	364,765,318	-	56,043,756	18.13	54.73	0.907 +/- 1.017	51.9	23.47	0.33	0.33
F38	482,336,164	201,434,665	108,474,460	55.90	85.49	1.941 +/- 2.295	74.51	52	0.23	0.25

Table S9

Uniparental marker analysis and contamination estimates. MT = Mitochondria. X chr contamination = X chromosome contamination rate for male samples. M1 & 2= Method 1 & 2. Cont (%) = Contamination percentage. SE = Standard Error. MT contamination estimates were subtracted from 1.

Sample	MT haplogroup	Coverage (+/-)	MT fragment length bp	MT authentic data (%)	MT contamination estimate (%)	Y haplogroup	X chr contamination		
							Cont (%)	SE	p-value
WC1	J1d6	547.56 +/- 253.6	96.5	99.94	0.00 - 0.21	G2b	M1= 0.79	0.0005	6,80E-03
							M2= 0.86	0.0011	7,80E-03
AH1	R2	74.38 +/- 53.69	57.54	99.98	0.00 - 0.34	-	-	-	-
AH2	R2	49.93 +/- 35.1	59.6	99.98	0.00 - 0.45	Insufficient reads	M1= 0.94	0.0034	3,10E-02
							M2= 0.09	0.0029	6.00E-01
AH4	T2c	40.79 +/- 32.5	55.01	99.95	0.01 - 0.84	-	-	-	-
F38	N1a3a	86.1 +/- 45.55	78.97	99.43	0.09 -1.45	R1b1a2a2	M1= 0.98	0.0013	9.01E-13
							M2= 1.14	0.0022	1.65E-06

Table S10

Number of SNPs overlapping phylogenetically informative Y-chromosome SNPs.

Sample	Number of reads	Number of SNPs
WC1	920,541	2,141
F38	183,574	769

Table S11

Derived and ancestral alleles for haplogroup R1b and sub-lineages identified in sample F38 (A = ancestral; D = derived).

Position	Marker	Haplogroup	Mutation	Ancestral	Derived	Coverage	Bases	State
2887824	M343	R1b	C ->A	C	A	2	AA	D
9170545	M415	R1b1	C->A	C	A	3	AAA	D
18656508	P297	R1b1a	G->C	G	C	3	CCC	D
22739367	M269	R1b1a2	T->C	T	C	3	CCC	D
2842212	L49.1	R1b1a2a	T->A	T	A	2	AA	D
6753511	L23	R1b1a2a	G->A	G	A	4	AAAA	D
7186135	CTS1078	R1b1a2a2	G->C	G	C	3	CCC	D
14641193	L52	R1b1a2a1a	C->T	C	T	2	CC	A
8796078	S21	R1b1a2a1a1	C->T	C	T	2	CC	A
6753316	L217.1	R1b1a2a1a1a	C->G	C	G	2	CC	A
4862861	V88	R1b1c	C->T	C	T	2	CC	A

Table S12

Derived alleles for haplogroup G2 and sub-lineages identified in WC1 (A = ancestral; D = derived).

Position	Marker	Haplogroup	Mutation	Ancestral	Derived	Coverage	Bases	State
17,174,741	L156	G2	A->T	A	T	7	TTTTTTT	D
22,072,097	P287	G2	G->T	G	T	7	TTTTTTT	D
6,753,306	L142.2	G2	G->A	G	A	7	AAAAAAA	D
13,811,971	Z8018	G2b	C->T	C	T	21	TTTTTTTTTT (...)	D
14,894,401	M3145	G2b	A->G	A	G	3	GGG	D
18,750,052	Z8019	G2b	G->T	G	T	5	TTTTT	D
22,064,715	M3200	G2b	C->T	C	T	6	TTTTTT	D
23,364,100	Z8021	G2b	A->G	A	G	7	GGGGGGG	D
2,976,942	Z8015	G2b	T->A	T	A	2	AA	D
4,943,744	Z8017	G2b	A->G	A	G	10	GGGGGGGGGG	D
7,322,981	M3115	G2b	G->T	G	T	4	TTTT	D
8,455,657	PF5721	G2b	G->A	G	A	9	AAAAAAAAA	D
14,028,148	L31	G2a	C->A	C	A	4	CCCC	A
23,973,594	U5	G2a	T->G	T	G	3	TTT	A
9,985,022	L293	G2a1	G->C	G	C	35	GGGGGGGGGG (...)	A
15,615,340	M3308	G2a2	C->G	C	G	8	CCCCCCCC	A
14,692,227	L32	G2a2b	T->C	T	C	7	TTTTTTT	A
15,027,433	M377	G2b1	A->G	A	G	5	AAAAA	A
23,243,887	L72	G2b1	G->C	G	C	4	GGGG	A
8,467,136	L183	G2b1	G->C	G	C	7	GGGGGGG	A

Table S13

PCA. For each ancient sample projected into the reference PCA space, the number of sites overlapping with the modern reference panel of SNPs and covered by at least one read is given. In addition, we provide a sample-specific Procrustes similarity score t ranging from 0 and 1 computed by LASER, which quantifies the confidence in the projection, where values closer to 0 indicating higher uncertainty. See (*107*) for details on the statistic.

Sample name	Original sample name	Nb. of sites covered	t	Reference
Ust-Ishim-SIB-UP	Ust'Ishim	511.850	0.999991	Fu *et al.*, Nature, 2014
Wezmeh-IR-N	WC1	511.555	0.999959	
Abd.-Hosein-IR-N	AH1	358.696	0.999322	**this study**
	AH2	259.785	0.99886	
	AH4	311.212	0.999102	
Hasanlu-IR-IA	F38	439.000	0.999599	
European Foragers	SfF11	44.238	0.992717	Skoglund *et al.*, Science, 2014
	LaBrana	465.374	0.99976	Olalde *et al.*, Nature, 2014
	Loschbour	511.845	0.999978	Lazaridis *et al.*, Nature, 2014
	KO1	336.396	0.999172	Gamba *et al.*, Nat Commun, 2014
	Motala1	84.694	0.996211	Lazaridis *et al.*, Nature, 2014
	Motala2	73.325	0.995461	
	Motala3	225.930	0.998572	
	Motala4	35.605	0.991133	
	Motala6	12.808	0.975093	
	Motala9	5.038	0.941636	
	Motala12	461.599	0.999659	
	Ajv52	44.731	0.992796	Skoglund *et al.*, Science, 2014
	Ajv53	13.018	0.976265	
	Ajv58	451.690	0.999627	
	Ajv70	76.608	0.995866	
	Ire8	20.034	0.984106	
	KK1	511.277	0.999962	Jones *et al.*, Nat Commun, 2015
Aegean and European farmers	KO2	44.433	0.992803	Gamba *et al.*, Nat Commun, 2014
	LBK	511.772	0.999976	Lazaridis *et al.*, Nature, 2014
	I0707	421.306	0.999776	Mathieson *et al.*, Nature, 2015
	I0708	407.405	0.999731	
	I0709	408.837	0.999753	
	I0736	330.077	0.999462	
	I0744	367.622	0.99955	
	I0745	413.599	0.999754	
	I0746	417.330	0.999775	
	I1096	320.421	0.999475	
	I1097	319.999	0.999446	
	I1098	336.416	0.99954	
	I1099	232.726	0.998873	
	I1100	129.115	0.997645	
	I1101	282.225	0.999265	
	I1102	178.560	0.998402	
	I1103	259.412	0.99907	
	I1579	353.419	0.999545	
	I1580	388.210	0.999629	
	I1581	355.657	0.999561	
	I1583	411.945	0.999777	
	I1585	354.928	0.999573	
	Bar31	477.922	0.999773	Hofmanová *et al.*, PNAS, 2016
	Bar8	508.614	0.999924	
	I0723	181.657	0.998338	Mathieson *et al.*, Nature, 2015
	I0724	22.428	0.98614	
	I0726	98.379	0.996835	
	I0727	19.649	0.984555	

54

Sample name	Original sample name	Nb. of sites covered	t	Reference
Aegean and European farmers	Rev5	335.781	0.999215	Hofmanová et al., PNAS, 2016
	Gok2	308.123	0.99929	Skoglund *et al.* , Science, 2014
	Gok4	17.022	0.981767	
	Gok5	9.288	0.968401	
	Gok7	5.559	0.951528	
	NE1	511.805	0.999981	Gamba *et al.* , Nat Commun, 2014
	NE2	76.669	0.99573	
	NE3	46.283	0.993036	
	NE4	38.679	0.991627	
	NE5	299.470	0.998985	
	NE6	336.448	0.999164	
	I0409	212.162	0.998643	Mathieson *et al.* , Nature, 2015
	I0410	215.305	0.998931	
	I0412	412.702	0.999782	
	I0413	279.563	0.999269	
	I0405	128.635	0.997673	
	I0406	299.725	0.999463	
	I0407	218.463	0.998911	
	I0408	387.370	0.999737	
	CB13	324.009	0.999159	Olalde *et al.* , MBE, 2015
Kleitos-GR-LN	Klei10	456.169	0.999643	Hofmanová et al., PNAS, 2016
Paliambela-GR-LN	Pal7	378.351	0.99935	
Satsurblia-GE-U	SATP	360.226	0.999277	Jones *et al.* , Nat Commun, 2015
Bichon-CH-U	Bichon	509.871	0.999938	
Kumtepe-TR-CHALC	Kumtepe6	56.682	0.994329	Omrak *et al.* , Curr Biol, 2016
Kostenki-RU-UP	Kostenki	441.747	0.999633	Seguin-Orlando *et al.* , Science, 2014
Mal'ta-SIB-UP	Mal'ta	360.422	0.999298	Raghavan *et al.* , Nature, 2014
Apc-Berek.-HUN-LN	NE7	329.729	0.99911	Gamba *et al.* , Nat Commun, 2014
Apc-Berek.-HUN-CA	CO1	191.977	0.996997	
Lud.-Var.-HUN-BA	BR2	511.935	0.999982	
Komp.-Kig.-HUN-BA	BR1	259.415	0.998809	
Suk.-Term.-RUS-YAM	RISE240	86.229	0.9962	Allentoft *et al.* , Nature, 2015
Temrta-RUS-YAM	RISE546	69.360	0.995219	
	RISE547	271.581	0.998883	
	RISE548	304.190	0.999138	
Peshany-RUS-YAM	RISE550	198.750	0.998404	
Ulan-RUS-YAM	RISE552	401.968	0.999594	
Lopatino-RUS-YAM	I0357	229.426	0.998661	Mathieson *et al.* , Nature, 2015
	I0429	250.954	0.998935	
	I0439	114.131	0.997174	
	I0443	419.386	0.999766	
Ekaterin.-RUS-YAM	I0231	448.821	0.999809	
Ishkinovka-RUS-YAM	I0370	299.709	0.999142	
Luzhki-RUS-YAM	I0438	236.941	0.998807	
Kurman.-RUS-YAM	I0441	40.530	0.992281	
Kutuluk-RUS-YAM	I0444	228.718	0.998718	
Kytmanovo-RUS-AND	RISE500	381.662	0.99949	Allentoft *et al.* , Nature, 2015
	RISE503	290.042	0.998995	
	RISE505	498.839	0.999877	
	RISE512	131.785	0.997545	
Lud.-Var.-HUN-IA	IR2	292.197	0.998724	Gamba *et al.* , Nat Commun, 2014

Table S14

Reference dataset. Additional samples added to (and replaced in) the reference datasets.

Sample ID	Population ID (this study)	Coverage	Country	Reference
Ust'-Ishim	UstIshim	42	Siberia	Fu et al. 2014
K14	Kostenki	2.84	Russia	Seguin-Orlando et al. 2014
Bichon	WHG (= Western Hunter Gatherers)	9.5	Switzerland	Jones et al. 2015
SATP	CHG (= Caucasus Hunter Gatherer)	1.44	Georgia	Jones et al. 2015
KK1	CHG (= Caucasus Hunter Gatherer)	15.38	Georgia	Jones et al. 2015
Bar8, Bar31	Anatolia_Neolithic	7.21, 3.71	Turkey	Hofmanová et al. 2016
Rev5	Greek_EN (= Greek Early Neolithic)	1.17	Greece	Hofmanová et al. 2016
Loschbour	WHG (= Western Hunter Gatherers)	22	Luxembourg	Lazaridis et al. 2015
NE1	Hungary_EN (= Early Neolithic)	22.12	Hungary	Gamba et al. 2014
Stuttgart	LBK_EN (Linear Pottery culture, Early Neolithic)	19	Germany	Lazaridis et al. 2015
Kumtepe6, Kumtepe4	Kumtepe6, Kumtepe4	0.13, 0.01	Turkey	Omrak et al. 2016
Pal7	Greek_FN (= Greek Final Neolithic)	1.29	Greece	Hofmanová et al. 2016
Klei10	Greek_FN (= Greek Final Neolithic)	2.01	Greece	Hofmanová et al. 2016
BR2	Hungary_BA (= Bronze Age)	21.25	Hungary	Gamba et al. 2014
Mota	Mota	12.54	Ethiopia	Llorente et al. 2015

Table S15.1

D statistic results. *D(Neo_Iranian, Neo_Iranian, Test, ‡Khomani)* where *Test* is ancient population (but not ancient Iranian) and values are |Z|>3. For all obtained values, see Table S20.

Neolithic Iranian	Neolithic Iranian	Test	‡Khomani	D-value	Z-score
WC1	AH2	CHG	‡Khomani	-0.0549	-7.592
WC1	AH2	Mota	‡Khomani	-0.0539	-7.21
WC1	AH4	Greek_FN	‡Khomani	-0.0422	-6.156
WC1	AH2	Greek_FN	‡Khomani	-0.0469	-6.102
WC1	AH4	CHG	‡Khomani	-0.0416	-6.009
AH1	AH2	CHG	‡Khomani	-0.0422	-4.648
WC1	AH4	Mota	‡Khomani	-0.0332	-4.588
WC1	AH4	Northern_LNBA	‡Khomani	-0.0253	-4.467
WC1	AH4	Yamnaya_Kalmykia	‡Khomani	-0.0245	-4.357
WC1	AH4	Sintashta	‡Khomani	-0.0266	-4.092
WC1	AH4	Afanasievo	‡Khomani	-0.0213	-3.711
AH1	AH2	Mota	‡Khomani	-0.0336	-3.604
AH1	AH4	Greek_FN	‡Khomani	-0.0305	-3.59
WC1	AH1	CHG	‡Khomani	-0.0238	-3.561
WC1	AH1	Northern_LNBA	‡Khomani	-0.0192	-3.504
AH1	AH2	Greek_FN	‡Khomani	-0.0306	-3.375
WC1	AH2	Kumtepe6	‡Khomani	-0.0886	-3.368
WC1	AH4	Hungary_BA	‡Khomani	-0.0175	-3.246
AH2	AH4	Bell_Beaker_LN	‡Khomani	-0.0222	-3.236
WC1	AH4	Yamnaya_Samara	‡Khomani	-0.0163	-3.21
WC1	AH4	Poltavka	‡Khomani	-0.0173	-3.067
WC1	AH2	Greek_EN	‡Khomani	-0.0323	-3.043
WC1	AH4	Hungary_EN	‡Khomani	-0.0158	-3.034

Table S15.2

D statistic results. Results of *D(Neo_Iranian, Neo_Iranian, Neo_Iranian, ǂKhomani)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

Neolithic Iranian	Neolithic Iranian	Neolithic Iranian	ǂKhomani	D-value	Z-score
AH2	WC1	AH4	ǂKhomani	0.0941	8.707
AH4	WC1	AH1	ǂKhomani	0.0842	8.331
AH2	WC1	AH1	ǂKhomani	0.0823	7.983
AH1	WC1	AH4	ǂKhomani	0.0758	7.702
AH4	WC1	AH2	ǂKhomani	0.0706	6.764
AH1	WC1	AH2	ǂKhomani	0.0458	4.448
AH2	AH1	WC1	ǂKhomani	0.0366	3.427

Table S15.3

D statistic results. Results of *D(Neo_Iranian, Neo_Iranian, F38, ǂKhomani)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

Neolithic Iranian	Neolithic Iranian	Iron Age Iranian	ǂKhomani	D-value	Z-score
AH2	WC1	F38	ǂKhomani	0.0466	5.157
AH4	WC1	F38	ǂKhomani	0.04	4.792
AH2	AH1	F38	ǂKhomani	0.0373	3.294

Table S15.4

D statistic results. Results of *D(Neo_Iranian, Test, F38, ‡Khomani)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

Neolithic Iranian	Test	Iron Age Iranian	‡Khomani	D-value	Z-score
WC1	Kumtepe6	F38	‡Khomani	-0.1147	-5.775
AH1	Kumtepe6	F38	‡Khomani	-0.1118	-4.874
AH4	Kumtepe6	F38	‡Khomani	-0.1115	-4.406
AH2	Kumtepe6	F38	‡Khomani	-0.1187	-3.962
WC1	Greek_FN	F38	‡Khomani	-0.0244	-3.6

Table S15.5

D statistic results. Results of *D(F38, Test, Neo_Iranian, ≠Khoman*i). Only values for |Z|>3 shown. For all obtained values, see Table S20.

Iron Age Iranian	Test	Neolithic Iranian	ǂKhomani	D-value	Z-score
F38	Kumtepe6	AH2	ǂKhomani	-0.1162	-3.961
F38	CHG	AH1	ǂKhomani	-0.0303	-3.711
F38	Kumtepe6	WC1	ǂKhomani	-0.0635	-3.09
F38	Kumtepe6	AH1	ǂKhomani	-0.0746	-3.047

Table S15.6

D statistic results. Results of *D(Neo_Iranian, IA_Iranian,Test, ǂKhomani)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

Neolithic Iranian	Iron Age Iranian	Test	ǂKhomani	D-value	Z-score
AH2	F38	Mota	ǂKhomani	0.0512	6.166
AH2	F38	CHG	ǂKhomani	0.0491	5.799
AH4	F38	Mota	ǂKhomani	0.0357	4.621
AH4	F38	CHG	ǂKhomani	0.0301	3.997
AH1	F38	Mota	ǂKhomani	0.0263	3.664
AH2	F38	Mezmaiskaya	ǂKhomani	0.0673	3.132

Table S16.1

D statistic results. Results of D(Neo_Iranian, CHG, pre-Neolithic, ǂKhomani).

Neolithic Iranian	CHG	pre-Neolithic	ǂKhomani	D-value	Z-score
WC1	CHG	WHG	ǂKhomani	-0.0376	-7.606
WC1	CHG	EHG	ǂKhomani	-0.0347	-6.37
AH2	CHG	SHG	ǂKhomani	-0.0458	-6.246
AH2	CHG	EHG	ǂKhomani	-0.0416	-6.179
AH1	CHG	WHG	ǂKhomani	-0.0339	-6.072
AH1	CHG	SHG	ǂKhomani	-0.0389	-5.906
AH2	CHG	WHG	ǂKhomani	-0.0344	-5.624
WC1	CHG	SHG	ǂKhomani	-0.0297	-5.056
AH1	CHG	EHG	ǂKhomani	-0.0295	-4.851
WC1	CHG	Kostenki	ǂKhomani	-0.0278	-4.293
AH4	CHG	EHG	ǂKhomani	-0.0275	-4.255
AH4	CHG	WHG	ǂKhomani	-0.0222	-3.919
AH4	CHG	SHG	ǂKhomani	-0.0248	-3.569
AH2	CHG	Kostenki	ǂKhomani	-0.0276	-3.422
WC1	CHG	MA1	ǂKhomani	-0.0194	-2.981
AH2	CHG	AG2	ǂKhomani	-0.0422	-2.943
AH2	CHG	AG2	ǂKhomani	-0.0422	-2.943
AH1	CHG	Kostenki	ǂKhomani	-0.0215	-2.877
AH4	CHG	Kostenki	ǂKhomani	-0.0189	-2.512
WC1	CHG	AG2	ǂKhomani	-0.02	-2.043
WC1	CHG	AG2	ǂKhomani	-0.02	-2.043
AH2	CHG	Ust'-Ishim	ǂKhomani	-0.0147	-1.991
AH2	CHG	MA1	ǂKhomani	-0.0168	-1.94
AH1	CHG	AG2	ǂKhomani	-0.0178	-1.47
AH1	CHG	AG2	ǂKhomani	-0.0178	-1.47
AH1	CHG	MA1	ǂKhomani	-0.0106	-1.425
AH4	CHG	Ust'-Ishim	ǂKhomani	-0.0093	-1.31
WC1	CHG	Ust'-Ishim	ǂKhomani	-0.007	-1.228
AH4	CHG	AG2	ǂKhomani	-0.0156	-1.183
AH4	CHG	AG2	ǂKhomani	-0.0156	-1.183
AH4	CHG	MA1	ǂKhomani	-0.0058	-0.723
AH1	CHG	Ust'-Ishim	ǂKhomani	-0.0012	-0.166

Table S16.2

D statistic results. Results of *D(Neo_Iranian, pre-Neolithic, Ust'-Ishim, ǂKhomani)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

Neolithic Iranian	pre-Neolithic	Ust'-Ishim	ǂKhomani	D-value	Z-score
WC1	AG2	Ust'-Ishim	ǂKhomani	-0.0412	-4.552
WC1	AG2	Ust'-Ishim	ǂKhomani	-0.0412	-4.552
AH4	AG2	Ust'-Ishim	ǂKhomani	-0.0466	-3.55
AH4	AG2	Ust'-Ishim	ǂKhomani	-0.0466	-3.55
WC1	EHG	Ust'-Ishim	ǂKhomani	-0.0311	-5.501
AH2	EHG	Ust'-Ishim	ǂKhomani	-0.0371	-5.236
AH4	EHG	Ust'-Ishim	ǂKhomani	-0.0314	-4.468
AH1	EHG	Ust'-Ishim	ǂKhomani	-0.0232	-3.519
WC1	Kostenki	Ust'-Ishim	ǂKhomani	-0.0477	-6.943
AH4	Kostenki	Ust'-Ishim	ǂKhomani	-0.0573	-6.843
AH2	Kostenki	Ust'-Ishim	ǂKhomani	-0.0473	-5.701
AH1	Kostenki	Ust'-Ishim	ǂKhomani	-0.0385	-4.895
WC1	MA1	Ust'-Ishim	ǂKhomani	-0.0428	-6.585
AH4	MA1	Ust'-Ishim	ǂKhomani	-0.0389	-4.908
AH2	MA1	Ust'-Ishim	ǂKhomani	-0.0392	-4.486
AH1	MA1	Ust'-Ishim	ǂKhomani	-0.0316	-4.189
WC1	SHG	Ust'-Ishim	ǂKhomani	-0.0324	-5.51
AH4	SHG	Ust'-Ishim	ǂKhomani	-0.0353	-4.977
AH2	SHG	Ust'-Ishim	ǂKhomani	-0.0366	-4.8
WC1	WHG	Ust'-Ishim	ǂKhomani	-0.0334	-6.36
AH4	WHG	Ust'-Ishim	ǂKhomani	-0.0367	-5.727
AH2	WHG	Ust'-Ishim	ǂKhomani	-0.037	-5.635
AH1	WHG	Ust'-Ishim	ǂKhomani	-0.0235	-3.748

Table S16.3

D statistic results. Results of *D(Neo_Iranian, pre-Neolithic HG, Ust'-Ishim, Chimp)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

Neolithic Iranian	pre-Neolithic	Ust'-Ishim	Chimp	D-value	Z-score
WC1	AG2	Ust'-Ishim	Chimp	-0.0643	-5.839
AH4	AG2	Ust'-Ishim	Chimp	-0.0726	-4.611
AH2	AG2	Ust'-Ishim	Chimp	-0.0707	-4.194
AH2	CHG	Ust'-Ishim	Chimp	-0.0349	-4.042
WC1	EHG	Ust'-Ishim	Chimp	-0.0428	-6.821
AH2	EHG	Ust'-Ishim	Chimp	-0.0551	-6.811
AH4	EHG	Ust'-Ishim	Chimp	-0.0429	-5.293
AH1	EHG	Ust'-Ishim	Chimp	-0.0278	-3.666
WC1	Kostenki	Ust'-Ishim	Chimp	-0.0514	-6.913
AH4	Kostenki	Ust'-Ishim	Chimp	-0.0642	-6.755
AH2	Kostenki	Ust'-Ishim	Chimp	-0.0564	-5.826
AH1	Kostenki	Ust'-Ishim	Chimp	-0.0363	-3.926
WC1	MA1	Ust'-Ishim	Chimp	-0.0459	-6.53
AH2	MA1	Ust'-Ishim	Chimp	-0.0507	-4.947
AH4	MA1	Ust'-Ishim	Chimp	-0.0413	-4.365
AH1	MA1	Ust'-Ishim	Chimp	-0.0276	-3.12
WC1	SHG	Ust'-Ishim	Chimp	-0.047	-7.207
AH2	SHG	Ust'-Ishim	Chimp	-0.0551	-6.382
AH4	SHG	Ust'-Ishim	Chimp	-0.0512	-6.191
AH1	SHG	Ust'-Ishim	Chimp	-0.0289	-3.668
AH2	WHG	Ust'-Ishim	Chimp	-0.0581	-7.788
WC1	WHG	Ust'-Ishim	Chimp	-0.045	-7.655
AH4	WHG	Ust'-Ishim	Chimp	-0.049	-6.599
AH1	WHG	Ust'-Ishim	Chimp	-0.0292	-4.143

Table S16.4

D statistic results. Results of *D(Neo_Iranian, Pre-neolithic, Neanderthal, Chimp)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

Neolithic Iranian	pre-Neolithic	Neanderthal	Chimp	D-value	Z-score
AH4	AG2	Vindija_light	Chimp	-0.0747	-3.237
AH4	EHG	Vindija_light	Chimp	-0.0444	-4.738
WC1	EHG	Vindija_light	Chimp	-0.0265	-3.6
AH2	EHG	Vindija_light	Chimp	-0.0364	-3.566
AH1	EHG	Vindija_light	Chimp	-0.0293	-3.126
AH4	Kostenki	Vindija_light	Chimp	-0.0511	-4.592
AH1	Kostenki	Vindija_light	Chimp	-0.0388	-3.589
AH4	MA1	Vindija_light	Chimp	-0.0514	-4.18
AH1	MA1	Vindija_light	Chimp	-0.04	-3.513
WC1	MA1	Vindija_light	Chimp	-0.0298	-3.477
AH2	MA1	Vindija_light	Chimp	-0.0401	-3.098
AH4	SHG	Vindija_light	Chimp	-0.0435	-4.118
AH1	SHG	Vindija_light	Chimp	-0.0311	-3.2
WC1	SHG	Vindija_light	Chimp	-0.0255	-3.093
AH1	Ust'-Ishim	Vindija_light	Chimp	-0.042	-4.033
AH4	Ust'-Ishim	Vindija_light	Chimp	-0.0399	-3.855
WC1	Ust'-Ishim	Vindija_light	Chimp	-0.0271	-3.267
AH4	WHG	Vindija_light	Chimp	-0.0423	-5.134
AH1	WHG	Vindija_light	Chimp	-0.0303	-3.601
WC1	WHG	Vindija_light	Chimp	-0.0238	-3.496
AH2	WHG	Vindija_light	Chimp	-0.0291	-3.179

Table S16.5

D statistic results. Results of *D(Neo_Iranian, Mota, Neanderthal, Chimp)*

Neolithic Iranian	Mota	Neanderthal	Chimp	D-value	Z-score
WC1	Mota	Vindija_light	Chimp	0.0284	3.738
WC1	Mota	Altai	Chimp	0.0167	2.186
AH1	Mota	Vindija_light	Chimp	0.0193	2.051
AH1	Mota	Altai	Chimp	0.0166	1.753
WC1	Mota	Mezmaiskaya	Chimp	0.0269	1.494
AH1	Mota	Mezmaiskaya	Chimp	0.0313	1.405
AH2	Mota	Altai	Chimp	0.0133	1.211
AH4	Mota	Vindija_light	Chimp	0.0111	1.071
AH2	Mota	Mezmaiskaya	Chimp	0.0297	1.048
AH4	Mota	Altai	Chimp	0.01	0.947
AH4	Mota	Mezmaiskaya	Chimp	0.0172	0.643
AH2	Mota	Vindija_light	Chimp	0.0063	0.61

Table S16.6

D statistic results. Results of *D(Neo_Iranian, Pre-neolithic, Mota, Chimp).* All values shown.

Neolithic Iranian	pre-Neolithic	Mota	Chimp	D-value	Z-score
AH4	EHG	Mota	Chimp	0.1148	15.425
AH1	EHG	Mota	Chimp	0.1085	14.983
AH2	Ust'-Ishim	Mota	Chimp	0.1181	14.089
AH2	EHG	Mota	Chimp	0.112	13.915
WC1	EHG	Mota	Chimp	0.0789	13.644
AH4	Ust'-Ishim	Mota	Chimp	0.1141	13.572
AH2	SHG	Mota	Chimp	0.1143	13.194
AH1	Ust'-Ishim	Mota	Chimp	0.1021	13.031
AH1	SHG	Mota	Chimp	0.0929	12.618
WC1	Ust'-Ishim	Mota	Chimp	0.0845	12.528
AH4	SHG	Mota	Chimp	0.1015	12.175
WC1	SHG	Mota	Chimp	0.0746	12.144
AH1	WHG	Mota	Chimp	0.0768	11.902
AH4	WHG	Mota	Chimp	0.0838	11.886
AH2	WHG	Mota	Chimp	0.0866	11.752
AH4	MA1	Mota	Chimp	0.1101	11.52
WC1	MA1	Mota	Chimp	0.0806	11.408
AH1	MA1	Mota	Chimp	0.1014	11.328
AH2	MA1	Mota	Chimp	0.1106	11.232
AH2	Kostenki	Mota	Chimp	0.1022	11.005
AH4	Kostenki	Mota	Chimp	0.086	10.002
AH1	Kostenki	Mota	Chimp	0.0819	9.757
WC1	WHG	Mota	Chimp	0.052	9.339
WC1	Kostenki	Mota	Chimp	0.0609	8.949
WC1	AG2	Mota	Chimp	0.0573	4.736
AH1	AG2	Mota	Chimp	0.077	4.701
AH2	AG2	Mota	Chimp	0.0821	4.31
AH4	AG2	Mota	Chimp	0.0702	4.093
AH2	CHG	Mota	Chimp	0.0279	3.117
AH4	CHG	Mota	Chimp	0.0246	2.986
AH1	CHG	Mota	Chimp	0.0149	1.917
WC1	CHG	Mota	Chimp	-0.0064	-0.988

Table S17.1

D statistic results. Results of *D(EN_farmer, Neo_Iranian, CHG, ‡Khomani)* . All data shown. For additional populations, see Table S20.

Early Neolithic Farmer	Neolithic Iranian	CHG	‡Khomani	D-value	Z-score
Anatolia_Neolithic	AH1	CHG	‡Khomani	-0.1094	-21.516
Anatolia_Neolithic	AH2	CHG	‡Khomani	-0.1391	-26.545
Anatolia_Neolithic	AH4	CHG	‡Khomani	-0.128	-25.843
Anatolia_Neolithic	WC1	CHG	‡Khomani	-0.0795	-18.689
LBK_EN	AH1	CHG	‡Khomani	-0.1142	-22.022
LBK_EN	AH2	CHG	‡Khomani	-0.1445	-27.341
LBK_EN	AH4	CHG	‡Khomani	-0.1301	-25.498
LBK_EN	WC1	CHG	‡Khomani	-0.0834	-19.555
Hungary_EN	AH1	CHG	‡Khomani	-0.1036	-19.475
Hungary_EN	AH2	CHG	‡Khomani	-0.1331	-24.092
Hungary_EN	AH4	CHG	‡Khomani	-0.1188	-22.616
Hungary_EN	WC1	CHG	‡Khomani	-0.0729	-16.327
Greek_EN	AH1	CHG	‡Khomani	-0.0632	-7.999
Greek_EN	AH2	CHG	‡Khomani	-0.1021	-11.581
Greek_EN	AH4	CHG	‡Khomani	-0.0803	-10.457
Greek_EN	WC1	CHG	‡Khomani	-0.0376	-5.652
Iberia_EN	AH1	CHG	‡Khomani	-0.1095	-19.752
Iberia_EN	AH2	CHG	‡Khomani	-0.1369	-23.579
Iberia_EN	AH4	CHG	‡Khomani	-0.126	-23.367
Iberia_EN	WC1	CHG	‡Khomani	-0.0796	-17.551

Table S17.2

D statistic results. Results of *D(EN_farmer, Neo_Iranian, WHG, ‡Khomani)*. All values shown. For additional populations, see Table S20.

Early Neolithic Farmer	Neolithic Iranian	WHG	‡Khomani	D-value	Z-score
Anatolia_Neolithic	AH1	WHG	‡Khomani	0.0401	9.534
Anatolia_Neolithic	AH2	WHG	‡Khomani	0.0421	9.289
Anatolia_Neolithic	AH4	WHG	‡Khomani	0.031	7.21
Anatolia_Neolithic	WC1	WHG	‡Khomani	0.0504	13.881
LBK_EN	AH1	WHG	‡Khomani	0.0467	10.817
LBK_EN	AH2	WHG	‡Khomani	0.0494	10.454
LBK_EN	AH4	WHG	‡Khomani	0.0386	8.906
LBK_EN	WC1	WHG	‡Khomani	0.0575	15.263
Hungary_EN	AH1	WHG	‡Khomani	0.0581	13.248
Hungary_EN	AH2	WHG	‡Khomani	0.0601	12.632
Hungary_EN	AH4	WHG	‡Khomani	0.0509	11.271
Hungary_EN	WC1	WHG	‡Khomani	0.0671	17.397
Greek_EN	AH1	WHG	‡Khomani	0.0466	7.085
Greek_EN	AH2	WHG	‡Khomani	0.0483	6.49
Greek_EN	AH4	WHG	‡Khomani	0.0368	5.44
Greek_EN	WC1	WHG	‡Khomani	0.0557	10.086
Iberia_EN	AH1	WHG	‡Khomani	0.0596	12.682
Iberia_EN	AH2	WHG	‡Khomani	0.0616	12.378
Iberia_EN	AH4	WHG	‡Khomani	0.0518	10.868
Iberia_EN	WC1	WHG	‡Khomani	0.0675	16.576

Table S17.3

D statistic results. Results of *D(EN_farmer, Neo_Iranian, EHG/SHG, ‡Khomani)*. All values shown. For additional populations, see Table S20.

Early Neolithic Farmer	Neolithic Iranian	EHG/SHG	‡Khomani	D-value	Z-score
Anatolia_Neolithic	AH1	EHG	‡Khomani	0.0283	6.02
Anatolia_Neolithic	AH2	EHG	‡Khomani	0.0361	7.152
Anatolia_Neolithic	AH4	EHG	‡Khomani	0.0298	5.96
Anatolia_Neolithic	WC1	EHG	‡Khomani	0.0394	9.86
LBK_EN	AH1	EHG	‡Khomani	0.0319	6.667
LBK_EN	AH2	EHG	‡Khomani	0.0378	7.17
LBK_EN	AH4	EHG	‡Khomani	0.0352	6.852
LBK_EN	WC1	EHG	‡Khomani	0.0435	10.83
Hungary_EN	AH1	EHG	‡Khomani	0.0367	7.503
Hungary_EN	AH2	EHG	‡Khomani	0.0462	8.717
Hungary_EN	AH4	EHG	‡Khomani	0.0405	7.854
Hungary_EN	WC1	EHG	‡Khomani	0.0472	11.149
Greek_EN	AH1	EHG	‡Khomani	0.0225	3.025
Greek_EN	AH2	EHG	‡Khomani	0.0251	2.917
Greek_EN	AH4	EHG	‡Khomani	0.0184	2.24
Greek_EN	WC1	EHG	‡Khomani	0.0286	4.493
Iberia_EN	AH1	EHG	‡Khomani	0.0366	7.128
Iberia_EN	AH2	EHG	‡Khomani	0.0438	7.74
Iberia_EN	AH4	EHG	‡Khomani	0.0397	7.397
Iberia_EN	WC1	EHG	‡Khomani	0.0455	10.452
Anatolia_Neolithic	AH1	SHG	‡Khomani	0.0554	10.936
Anatolia_Neolithic	AH2	SHG	‡Khomani	0.0569	10.085
Anatolia_Neolithic	AH4	SHG	‡Khomani	0.0444	8.315
Anatolia_Neolithic	WC1	SHG	‡Khomani	0.0547	12.448
LBK_EN	AH1	SHG	‡Khomani	0.0618	11.885
LBK_EN	AH2	SHG	‡Khomani	0.065	11.179
LBK_EN	AH4	SHG	‡Khomani	0.0511	9.347
LBK_EN	WC1	SHG	‡Khomani	0.0613	13.904
Hungary_EN	AH1	SHG	‡Khomani	0.0653	12.159
Hungary_EN	AH2	SHG	‡Khomani	0.0682	11.768
Hungary_EN	AH4	SHG	‡Khomani	0.0564	10.145
Hungary_EN	WC1	SHG	‡Khomani	0.0627	13.55
Greek_EN	AH1	SHG	‡Khomani	0.0513	6.253
Greek_EN	AH2	SHG	‡Khomani	0.064	7.036
Greek_EN	AH4	SHG	‡Khomani	0.0453	5.307
Greek_EN	WC1	SHG	‡Khomani	0.0476	7.086
Iberia_EN	AH1	SHG	‡Khomani	0.0701	12.593
Iberia_EN	AH2	SHG	‡Khomani	0.0708	11.667
Iberia_EN	AH4	SHG	‡Khomani	0.0619	10.556
Iberia_EN	WC1	SHG	‡Khomani	0.0675	14.514

Table S17.4

D statistic results. Results of *D(Neo_Iranian, HG, Anatolian_Neolithic, ǂKhomani)*. All values shown. For all obtained values (also for other early farmers than Anatolian) see Table S20.

Neolithic Iranian	HG	Neolithic Anatolian	ǂKhomani	D-value	Z-score
WC1	AG2	Anatolia_Neolithic	ǂKhomani	-0.0277	-4.885
AH2	AG2	Anatolia_Neolithic	ǂKhomani	-0.0334	-3.878
AH1	AG2	Anatolia_Neolithic	ǂKhomani	-0.0239	-3.219
AH4	AG2	Anatolia_Neolithic	ǂKhomani	-0.0159	-2.064
AH2	CHG	Anatolia_Neolithic	ǂKhomani	-0.0357	-7.898
AH1	CHG	Anatolia_Neolithic	ǂKhomani	-0.0329	-7.66
WC1	CHG	Anatolia_Neolithic	ǂKhomani	-0.0268	-7.424
AH4	CHG	Anatolia_Neolithic	ǂKhomani	-0.0229	-5.33
WC1	EHG	Anatolia_Neolithic	ǂKhomani	-0.027	-8.016
AH2	EHG	Anatolia_Neolithic	ǂKhomani	-0.0308	-7.457
AH1	EHG	Anatolia_Neolithic	ǂKhomani	-0.0254	-6.181
AH4	EHG	Anatolia_Neolithic	ǂKhomani	-0.0188	-4.624
WC1	Kostenki	Anatolia_Neolithic	ǂKhomani	-0.0008	-0.182
AH2	Kostenki	Anatolia_Neolithic	ǂKhomani	-0.0005	-0.091
AH1	Kostenki	Anatolia_Neolithic	ǂKhomani	0.0019	0.389
AH4	Kostenki	Anatolia_Neolithic	ǂKhomani	0.0039	0.774
AH2	MA1	Anatolia_Neolithic	ǂKhomani	-0.0015	-0.275
WC1	MA1	Anatolia_Neolithic	ǂKhomani	0	-0.004
AH1	MA1	Anatolia_Neolithic	ǂKhomani	0.0008	0.158
AH4	MA1	Anatolia_Neolithic	ǂKhomani	0.0084	1.735
WC1	SHG	Anatolia_Neolithic	ǂKhomani	-0.0384	-10.132
AH2	SHG	Anatolia_Neolithic	ǂKhomani	-0.0419	-9.552
AH1	SHG	Anatolia_Neolithic	ǂKhomani	-0.0379	-8.669
AH4	SHG	Anatolia_Neolithic	ǂKhomani	-0.0334	-7.473
AH2	Ust'-Ishim	Anatolia_Neolithic	ǂKhomani	0.0443	8.663
AH1	Ust'-Ishim	Anatolia_Neolithic	ǂKhomani	0.0482	9.959
WC1	Ust'-Ishim	Anatolia_Neolithic	ǂKhomani	0.0444	10.634
AH4	Ust'-Ishim	Anatolia_Neolithic	ǂKhomani	0.0544	10.952
WC1	WHG	Anatolia_Neolithic	ǂKhomani	-0.0401	-12.736
AH2	WHG	Anatolia_Neolithic	ǂKhomani	-0.0449	-11.428
AH1	WHG	Anatolia_Neolithic	ǂKhomani	-0.0401	-10.512
AH4	WHG	Anatolia_Neolithic	ǂKhomani	-0.0356	-9.16

Table S17.5

D statistic results. Results of *D(Anatolian_Neolithic, HG, Neo_Iranian, ‡Khomani)*. All values shown. For additional populations, see Table S20.

Neolithic Anatolian	HG	Neolithic Iranian	‡Khomani	D-value	Z-score
Anatolia_Neolithic	AG2	AH1	‡Khomani	-0.0062	-0.733
Anatolia_Neolithic	AG2	AH2	‡Khomani	-0.0072	-0.697
Anatolia_Neolithic	AG2	AH4	‡Khomani	-0.0039	-0.408
Anatolia_Neolithic	AG2	WC1	‡Khomani	-0.0025	-0.354
Anatolia_Neolithic	CHG	AH1	‡Khomani	-0.1417	-30.594
Anatolia_Neolithic	CHG	AH2	‡Khomani	-0.174	-35.69
Anatolia_Neolithic	CHG	AH4	‡Khomani	-0.1505	-32.272
Anatolia_Neolithic	CHG	WC1	‡Khomani	-0.1061	-26.448
Anatolia_Neolithic	EHG	AH1	‡Khomani	0.0029	0.666
Anatolia_Neolithic	EHG	AH2	‡Khomani	0.0053	1.067
Anatolia_Neolithic	EHG	AH4	‡Khomani	0.011	2.193
Anatolia_Neolithic	EHG	WC1	‡Khomani	0.0125	3.233
Anatolia_Neolithic	Kostenki	AH1	‡Khomani	0.0165	2.784
Anatolia_Neolithic	Kostenki	AH2	‡Khomani	0.0185	2.845
Anatolia_Neolithic	Kostenki	AH4	‡Khomani	0.02	3.447
Anatolia_Neolithic	Kostenki	WC1	‡Khomani	0.028	5.527
Anatolia_Neolithic	MA1	AH1	‡Khomani	0.0101	1.733
Anatolia_Neolithic	MA1	AH2	‡Khomani	0.0125	1.873
Anatolia_Neolithic	MA1	AH4	‡Khomani	0.0098	1.668
Anatolia_Neolithic	MA1	WC1	‡Khomani	0.0183	3.713
Anatolia_Neolithic	SHG	AH1	‡Khomani	0.0176	3.672
Anatolia_Neolithic	SHG	AH2	‡Khomani	0.015	2.885
Anatolia_Neolithic	SHG	AH4	‡Khomani	0.011	2.114
Anatolia_Neolithic	SHG	WC1	‡Khomani	0.0163	3.898
Anatolia_Neolithic	Ust'-Ishim	AH1	‡Khomani	0.052	9.147
Anatolia_Neolithic	Ust'-Ishim	AH2	‡Khomani	0.0625	10.326
Anatolia_Neolithic	Ust'-Ishim	AH4	‡Khomani	0.0693	11.689
Anatolia_Neolithic	Ust'-Ishim	WC1	‡Khomani	0.0603	12.103
Anatolia_Neolithic	WHG	AH1	‡Khomani	0	0.012
Anatolia_Neolithic	WHG	AH2	‡Khomani	-0.0028	-0.684
Anatolia_Neolithic	WHG	AH4	‡Khomani	-0.0046	-1.186
Anatolia_Neolithic	WHG	WC1	‡Khomani	0.0103	3.157

Table S17.6

f3 statistic. Results of *f3*(*Neo_Iranian, HG; Anatolia_Neolithic*).

Neolithic Iranian	Hunter-gatherer	Neolithic Anatolian	f3-value	std.err.	Z-score
Neo_Iranian	CHG	Anatolia_Neolithic	0.092674	0.001477	62.756
Neo_Iranian	EHG	Anatolia_Neolithic	0.030124	0.001262	23.877
Neo_Iranian	SHG	Anatolia_Neolithic	0.02172	0.001365	15.907
Neo_Iranian	WHG	Anatolia_Neolithic	0.026526	0.000999	26.562

Table S18.1

D statistic results. Results of *D*(*European_Early_farmer1, Neo_Iranians, European_Early_farmer2, ǂKhomani*). Only values for |Z|>3 shown. For all obtained values, see Table S20.

European Early farmer	Neolithic Iranian	European Early farmer	ǂKhomani	D-value	Z-score
Anatolia_Neolithic	WC1	Kumtepe6	ǂKhomani	-0.1154	-10.353
Anatolia_Neolithic	AH2	Kumtepe6	ǂKhomani	-0.1781	-10.145
Anatolia_Neolithic	AH1	Kumtepe6	ǂKhomani	-0.1166	-8.166
Anatolia_Neolithic	AH4	Kumtepe6	ǂKhomani	-0.1237	-7.657
Bell_Beaker_LN	WC1	Kumtepe6	ǂKhomani	-0.1657	-14.438
Bell_Beaker_LN	AH2	Kumtepe6	ǂKhomani	-0.2247	-12.482
Bell_Beaker_LN	AH1	Kumtepe6	ǂKhomani	-0.1646	-11.142
Bell_Beaker_LN	AH2	Greek_FN	ǂKhomani	-0.0613	-11.009
Bell_Beaker_LN	AH4	Greek_FN	ǂKhomani	-0.0563	-10.43
Bell_Beaker_LN	AH4	Kumtepe6	ǂKhomani	-0.1664	-10.283
Bell_Beaker_LN	AH1	Greek_FN	ǂKhomani	-0.0356	-6.609
Bell_Beaker_LN	WC1	Greek_FN	ǂKhomani	-0.0155	-3.775
Bell_Beaker_LN	AH4	Greek_EN	ǂKhomani	-0.0203	-3.031
Central_MN	WC1	Kumtepe6	ǂKhomani	-0.1402	-10.553
Central_MN	AH2	Kumtepe6	ǂKhomani	-0.2024	-10.115
Central_MN	AH1	Kumtepe6	ǂKhomani	-0.1361	-8.497
Central_MN	AH4	Kumtepe6	ǂKhomani	-0.1483	-8.005
Central_MN	AH2	Greek_FN	ǂKhomani	-0.0231	-3.658
Central_MN	AH4	Greek_FN	ǂKhomani	-0.0214	-3.583
Hungary_EN	AH2	Kumtepe6	ǂKhomani	-0.1524	-8.222
Hungary_EN	WC1	Kumtepe6	ǂKhomani	-0.0941	-7.971
Hungary_EN	AH1	Kumtepe6	ǂKhomani	-0.1005	-6.718
Hungary_EN	AH4	Kumtepe6	ǂKhomani	-0.1125	-6.598
Iberia_Chalcolithic	AH2	Kumtepe6	ǂKhomani	-0.1937	-9.664
Iberia_Chalcolithic	WC1	Kumtepe6	ǂKhomani	-0.1157	-8.604
Iberia_Chalcolithic	AH1	Kumtepe6	ǂKhomani	-0.111	-6.834
Iberia_Chalcolithic	AH4	Kumtepe6	ǂKhomani	-0.1213	-6.566
Iberia_Chalcolithic	AH2	Greek_FN	ǂKhomani	-0.0271	-4.493
Iberia_Chalcolithic	AH4	Greek_FN	ǂKhomani	-0.0194	-3.312
Iberia_EN	WC1	Kumtepe6	ǂKhomani	-0.1183	-9.295
Iberia_EN	AH2	Kumtepe6	ǂKhomani	-0.1851	-9.216
Iberia_EN	AH1	Kumtepe6	ǂKhomani	-0.1225	-7.578
Iberia_EN	AH4	Kumtepe6	ǂKhomani	-0.1213	-6.653
Iberia_MN	AH2	Kumtepe6	ǂKhomani	-0.1937	-9.519
Iberia_MN	WC1	Kumtepe6	ǂKhomani	-0.1164	-8.576
Iberia_MN	AH1	Kumtepe6	ǂKhomani	-0.1231	-7.245
Iberia_MN	AH4	Kumtepe6	ǂKhomani	-0.126	-6.716
Iberia_MN	AH2	Greek_FN	ǂKhomani	-0.0216	-3.451
Iceman	WC1	Kumtepe6	ǂKhomani	-0.146	-10.403
Iceman	AH2	Kumtepe6	ǂKhomani	-0.2065	-9.618
Iceman	AH1	Kumtepe6	ǂKhomani	-0.1365	-7.572
Iceman	AH4	Kumtepe6	ǂKhomani	-0.1429	-7.166
Iceman	AH2	Greek_FN	ǂKhomani	-0.0299	-4.272
Iceman	AH4	Greek_FN	ǂKhomani	-0.0262	-3.961
Kumtepe4	AH2	Kumtepe6	ǂKhomani	-0.7431	-3.092
LBK_EN	AH2	Kumtepe6	ǂKhomani	-0.1902	-10.68
LBK_EN	WC1	Kumtepe6	ǂKhomani	-0.1215	-10.527
LBK_EN	AH1	Kumtepe6	ǂKhomani	-0.1232	-8.278
LBK_EN	AH4	Kumtepe6	ǂKhomani	-0.1264	-7.698
LBK_EN	AH2	Greek_FN	ǂKhomani	-0.0192	-3.446

Table S18.2

D statistic results. Results of *D*(*Steppe, EHG, Neo_Iranian, ǂKhomani*) . Only values for |Z|>3 shown. For all obtained values, see Table S20.

Steppe	EHG	Neolithic Iranian	ǂKhomani	D-value	Z-score
Sintashta	EHG	AH4	ǂKhomani	0.043	7.395
Yamnaya_Kalmykia	EHG	AH4	ǂKhomani	0.0353	6.603
Yamnaya_Samara	EHG	AH4	ǂKhomani	0.0319	6.302
Sintashta	EHG	WC1	ǂKhomani	0.0273	6.234
Afanasievo	EHG	AH4	ǂKhomani	0.0343	6.143
Yamnaya_Kalmykia	EHG	AH1	ǂKhomani	0.0291	5.984
Poltavka	EHG	AH4	ǂKhomani	0.0328	5.962
Poltavka	EHG	AH1	ǂKhomani	0.0294	5.884
Poltavka	EHG	WC1	ǂKhomani	0.0232	5.572
Yamnaya_Samara	EHG	WC1	ǂKhomani	0.0217	5.44
Sintashta	EHG	AH1	ǂKhomani	0.0272	5.273
Northern_LNBA	EHG	AH4	ǂKhomani	0.028	5.155
Afanasievo	EHG	AH1	ǂKhomani	0.0264	5.129
Yamnaya_Samara	EHG	AH1	ǂKhomani	0.0223	5.084
Northern_LNBA	EHG	AH1	ǂKhomani	0.0232	4.82
Afanasievo	EHG	WC1	ǂKhomani	0.0206	4.786
Poltavka	EHG	AH2	ǂKhomani	0.0254	4.689
Yamnaya_Kalmykia	EHG	WC1	ǂKhomani	0.0191	4.678
Sintashta	EHG	AH2	ǂKhomani	0.0271	4.478
Srubnaya	EHG	AH4	ǂKhomani	0.0206	4.128
Yamnaya_Kalmykia	EHG	AH2	ǂKhomani	0.0226	4.108
Northern_LNBA	EHG	AH2	ǂKhomani	0.0233	4.054
Andronovo	EHG	AH4	ǂKhomani	0.0211	3.815
Afanasievo	EHG	AH2	ǂKhomani	0.0196	3.617
Yamnaya_Samara	EHG	AH2	ǂKhomani	0.0181	3.608
Srubnaya	EHG	WC1	ǂKhomani	0.0129	3.355
Andronovo	EHG	WC1	ǂKhomani	0.0129	3.137
Potapovka	EHG	WC1	ǂKhomani	0.0176	3.115
Samara_Eneolithic	EHG	AH1	ǂKhomani	0.0184	3.075
Srubnaya	EHG	AH1	ǂKhomani	0.0136	3.068

Table S18.3

D statistic results. Results of *D(Steppe, F38, Anatolia_Neolithic, ǂKhomani)*. Only values for |Z|>3 shown. For all obtained values and additional populations, see Table S20.

Steppe	Iron Age Iranian	Neolithic Anatolian	ǂKhomani	D-value	Z-score
Yamnaya_Kalmykia	F38	Anatolia_Neolithic	ǂKhomani	-0.0199	-5.763
Afanasievo	F38	Anatolia_Neolithic	ǂKhomani	-0.0198	-5.567
Samara_Eneolithic	F38	Anatolia_Neolithic	ǂKhomani	-0.0213	-5.124
Russia_EBA	F38	Anatolia_Neolithic	ǂKhomani	-0.028	-4.52
Yamnaya_Samara	F38	Anatolia_Neolithic	ǂKhomani	-0.0145	-4.27
Scythian_IA	F38	Anatolia_Neolithic	ǂKhomani	-0.0153	-3.579

Table S18.4

f4 statistic results. Results of *f4*(*European_Early_farmer, F38, Steppe, ‡Khomani*). Only values for |Z|>3 shown. For all obtained values, see Table S20.

European Early farmer	Iron Age Iranian	Steppe	‡Khomani	D-value	Z-score
Kumtepe6	F38	Samara_Eneolithic	‡Khomani	-0.0752	-3.556
Greek_EN	F38	Samara_Eneolithic	‡Khomani	-0.0276	-3.395
Kumtepe6	F38	Srubnaya	‡Khomani	-0.0471	-3.674
Kumtepe6	F38	Yamnaya_Kalmykia	‡Khomani	-0.0491	-3.294
Greek_EN	F38	Yamnaya_Samara	‡Khomani	-0.0203	-3.621

Table S18.5

D statistic results. Results of *D(European_Early_farmer, Steppe, F38, ǂKhomani)*. Only values for |Z|>3 shown. For all obtained values, see Table S20.

European Early farmer	Steppe	Iron Age Iranian	ǂKhomani	D-value	Z-score
Remedello	Sintashta	F38	ǂKhomani	-0.0292	-5
Bell_Beaker_LN	Sintashta	F38	ǂKhomani	-0.0176	-4.949
Iceman	Sintashta	F38	ǂKhomani	-0.0214	-3.773
Central_LNBA	Sintashta	F38	ǂKhomani	-0.013	-3.772
Remedello	Srubnaya	F38	ǂKhomani	-0.0183	-3.686
Remedello	Andronovo	F38	ǂKhomani	-0.0195	-3.579
Remedello	Yamnaya_Samara	F38	ǂKhomani	-0.0168	-3.272
Remedello	Poltavka	F38	ǂKhomani	-0.017	-3.115

Table S18.6

D statistic results. Results of *D(Sintashta, European_Early_farmer, F38, ǂKhomani)*.
Only values for |Z|>3 shown. For all obtained values, see Table S20.

Sintashta	European Early farmer	Iron Age Iranian	ǂKhomani	D-value	Z-score
Sintashta	Greek_FN	F38	ǂKhomani	-0.0699	-12.682
Sintashta	Kumtepe6	F38	ǂKhomani	-0.1479	-9.053
Sintashta	Greek_EN	F38	ǂKhomani	-0.0426	-6.059

Test S18.7

f3 statistic results. Results of *f3*(*Neo_Iranian, Test; Steppe*). Only values for |Z|>3 shown. For all obtained values and additional populations, see Table S20.

Neolithic Iranian	Test	Steppe	f3-value	std.err.	Z-score
Neo_Iranian	EHG	Yamnaya_Samara	-0.008409	0.001284	-6.551
Neo_Iranian	EHG	Yamnaya_Kalmykia	-0.009865	0.001617	-6.101
Neo_Iranian	EHG	Srubnaya	-0.007386	0.001217	-6.068
Neo_Iranian	EHG	Poltavka	-0.010458	0.001738	-6.016
Neo_Iranian	SHG	Srubnaya	-0.007037	0.001424	-4.941
Neo_Iranian	SHG	Poltavka	-0.007474	0.002058	-3.632
Neo_Iranian	SHG	Andronovo	-0.006651	0.002033	-3.271
Neo_Iranian	Samara_Eneolithic	Poltavka	-0.006906	0.002165	-3.189

Table S19

f3 statisti*c* results. Results of *f3*(*Neo_Iranian,*
Anatolia_Neolithic/Steppe/EHG/WHG/CHG; Test). Only values for |Z|>3 shown. For all
obtained values and additional populations, see Table S20.

Neolithic Iranian	Ancient population	Test	f3-value	std.err.	Z-score
Neo_Iranian	Anatolia_Neolithic	Fars	-0.006625	0.000466	-14.23
Neo_Iranian	Anatolia_Neolithic	Turkish	-0.003942	0.000397	-9.938
Neo_Iranian	Anatolia_Neolithic	Armenian	-0.005134	0.000577	-8.897
Neo_Iranian	Anatolia_Neolithic	Lebanese	-0.004732	0.000566	-8.363
Neo_Iranian	Anatolia_Neolithic	Jordanian	-0.003759	0.000553	-6.799
Neo_Iranian	Anatolia_Neolithic	Syrian	-0.003784	0.000588	-6.43
Neo_Iranian	Anatolia_Neolithic	Cypriot	-0.003504	0.000583	-6.005
Neo_Iranian	Anatolia_Neolithic	Turkish_Jew	-0.002687	0.0006	-4.477
Neo_Iranian	Anatolia_Neolithic	Iraqi_Jew	-0.002282	0.000696	-3.279
Neo_Iranian	Anatolia_Neolithic	Kumyk	-0.001929	0.000616	-3.13
Neo_Iranian	EHG	Pathan	-0.009051	0.000974	-9.294
Neo_Iranian	EHG	Tajik_Pomiri	-0.009284	0.001083	-8.573
Neo_Iranian	EHG	Balochi	-0.006701	0.000947	-7.078
Neo_Iranian	EHG	Fars	-0.006261	0.000979	-6.396
Neo_Iranian	EHG	Sindhi	-0.005516	0.000957	-5.765
Neo_Iranian	EHG	Makrani	-0.005386	0.000967	-5.568
Neo_Iranian	EHG	GujaratiA	-0.005533	0.001276	-4.337
Neo_Iranian	EHG	Lezgin	-0.004568	0.001143	-3.998
Neo_Iranian	EHG	Kumyk	-0.00427	0.001147	-3.722
Neo_Iranian	Potapovka	Fars	-0.005181	0.001691	-3.063
Neo_Iranian	Samara_Eneolithic	Tajik_Pomiri	-0.008022	0.001378	-5.822
Neo_Iranian	Samara_Eneolithic	Pathan	-0.006552	0.001213	-5.402
Neo_Iranian	Samara_Eneolithic	Fars	-0.005079	0.00126	-4.029
Neo_Iranian	Samara_Eneolithic	Makrani	-0.004898	0.001259	-3.891
Neo_Iranian	Samara_Eneolithic	Balochi	-0.00472	0.001231	-3.834
Neo_Iranian	Srubnaya	Fars	-0.005548	0.000532	-10.432
Neo_Iranian	Srubnaya	Balochi	-0.003122	0.00053	-5.893
Neo_Iranian	Srubnaya	Pathan	-0.002835	0.000533	-5.315
Neo_Iranian	Srubnaya	Makrani	-0.002589	0.000535	-4.84
Neo_Iranian	Srubnaya	Tajik_Pomiri	-0.002903	0.000671	-4.325
Neo_Iranian	Srubnaya	Kumyk	-0.002591	0.000664	-3.902
Neo_Iranian	Yamnaya_Samara	Fars	-0.001999	0.000605	-3.306

Table S20

Detailed results of the *f3* and *D* statistics.(Available online as a separate excel file).

Table S21

Neanderthal ancestry proportions. Neanderthal ancestry proportions estimated by *f4*-ratio statistics. 'PAT_gt' means that the corresponding sample has been recalled with our genotype caller described in Supplementary section 6.

Sample ID	Sample name	Dataset	Age	Estimate	Std.-err.	Reference
UstIshim	Ust-Ishim-SIB-UP	Fu (but diploid!)	45.020	0.030297	0.004398	Fu *et al.*, Nature, 2014
Stuttgart	Stutgart-GER-N	Fu (but diploid!)	7.140	0.003252	0.003501	Lazaridis *et al.*, Nature, 2014
Loschbour_Fu	Loschbour-LU-M	Fu (but diploid!)	8.050	0.007679	0.003888	Lazaridis *et al.*, Nature, 2014
Motala12	Motala-SW-M	Fu	7.625	0.014487	0.004326	Lazaridis *et al.*, Nature, 2014
Hungarian.KO1	Tisz.-Doma.-HUN-KOR-HG	Fu	7.660	0.003948	0.004614	Gamba *et al.*, Nat Commun, 2014
LaBrana1	La Brana-ES-M	Fu	7.815	0.01373	0.00417	Olalde *et al.*, Nature, 2014
Karelia		Fu	8.375	0.009587	0.004307	Haak *et al.*, Nature, 2015
Kotias	Kotias-GE-M	Fu	9.720	0.005646	0.004092	Jones *et al.*, Nat Commun, 2015
Ranchot88		Fu	10.085	0.026253	0.005793	Fu *et al.*, Nature, 2016
Rochedane		Fu	12.960	0.013599	0.007299	Fu *et al.*, Nature, 2016
Satsurblia	Satsurblia-GE-U	Fu	13.225	0.003929	0.004654	Jones *et al.*, Nat Commun, 2015
Bichon	Bichon-CH-U	Fu	13.665	0.01364	0.00427	Jones *et al.*, Nat Commun, 2015
Villabruna		Fu	13.980	0.016053	0.004426	Fu *et al.*, Nature, 2016
AfontovaGora3		Fu	16.710	0.025429	0.006708	Fu *et al.*, Nature, 2016
Oase1		Fu	39.610	0.078461	0.008615	Fu *et al.*, Nature, 2015
Kostenki14	Kostenki-RU-UP	Fu	37.470	0.025777	0.004408	Seguin-Orlando *et al.*, Science, 2014
GoyetQ116-1		Fu	34.795	0.022355	0.004731	Fu *et al.*, Nature, 2016
KremsWA3		Fu	30.970	0.027952	0.007507	Fu *et al.*, Nature, 2016
Vestonice16		Fu	30.010	0.027067	0.004791	Fu *et al.*, Nature, 2016
Ostuni1		Fu	27.620	0.033653	0.006116	Fu *et al.*, Nature, 2016
Malta1_Fu	Mal'ta-SIB-UP	Fu	24.305	0.020049	0.00465	Raghavan *et al.*, Nature, 2014
ElMiron		Fu	18.720	0.026349	0.005028	Fu *et al.*, Nature, 2016
Dai		Lazaridis	0	0.008542	0.002828	Lazaridis *et al.*, Nature, 2014
Han		Lazaridis	0	0.008776	0.002763	Lazaridis *et al.*, Nature, 2014
English		Lazaridis	0	0.00123	0.002531	Lazaridis *et al.*, Nature, 2014
French		Lazaridis	0	0.001084	0.002461	Lazaridis *et al.*, Nature, 2014
Sardinian		Lazaridis	0	0.000306	0.00249	Lazaridis *et al.*, Nature, 2014
UstIshim	Ust-Ishim-SIB-UP	PAT_gt	45.020	0.030752	0.004396	Fu *et al.*, Nature, 2014
Kostenki	Kostenki-RU-UP	PAT_gt	37.470	0.023105	0.004184	Seguin-Orlando *et al.*, Science, 2014
SATP	Satsurblia-GE-U	PAT_gt	13.225	0.003676	0.00443	Jones *et al.*, Nat Commun, 2015
Bichon	Bichon-CH-U	PAT_gt	13.665	0.013491	0.003888	Jones *et al.*, Nat Commun, 2015
Mal_ta	Mal'ta-SIB-UP	PAT_gt	24.305	0.017006	0.003778	Raghavan *et al.*, Nature, 2014
KK1	Kotias-GE-M	PAT_gt	9.720	0.005146	0.003761	Jones *et al.*, Nat Commun, 2015
Loschbour	Loschbour-LU-M	PAT_gt	8.050	0.008923	0.003871	Lazaridis *et al.*, Nature, 2014
Stuttgart	Stutgart-GER-N	PAT_gt	7.140	0.004469	0.003482	Lazaridis *et al.*, Nature, 2014
WC1	Wezmeh-IR-N	PAT_gt	9.285	0.002479	0.003589	this study

Table S22

Ancestry proportions. Inferred proportions of ancestry for aDNA samples when using all modern populations as surrogates (analysis (I)) for the Lazaridis (left of \) and Busby (right of \) merges. Modern populations showing contributions to any of the aDNA samples are classified by geographical region as defined below.

	Ust'-Ishim	KK1	Loschbour	WC1	Bar8	NE1	LBK	Mota
S.Europe (1)	8.3\0	0\0	20.8\5.2	0\0	63.9\79	91\100	94\100	0\0
N/C.Europe (2)	0\8.2	0\0	40.3\70.1	0\0	0\0	0\0	6\0	0\0
Cyprus (3)	0\ 0	0\ 0	0\ 0	0\ 0	9.2\ 17.3	0\ 0	0\ 0	0\ 0
E.Europe (4)	0\0	0\0	38.9\24.7	0\0	0\0	9\0	0\0	0\0
NearEast (5)	6.4\13.1	0\0	0\0	3.8\0	26.9\0	0\0	0\0	0\1.8
MiddleEast (6)	0\0	19.1\5.3	0\0	93.2\58.1	0\0	0\0	0\0	0\0
Caucasus (7)	0\6.1	80.9\94.2	0\0	3\27.9	0\3.7	0\0	0\0	0\0
India (8)	37.9\38.2	0\0.4	0\0	0\14	0\0	0\0	0\0	2.8\1.2
C.AsiaSiberia (9)	5.1\6.7	0\0	0\0	0\0	0\0	0\0	0\0	7.1\2.2
E.Asia (10)	24.9\7.5	0\0	0\0	0\0	0\0	0\0	0\0	4.2\0.8
Oceania (11)	6.7\6.5	0\0	0\0	0\0	0\0	0\0	0\0	0\0
America (12)	1.0\0	0\0	0\0	0\0	0\0	0\0	0\0	0\0
N.Africa (13)	9.7\13.8	0\0	0\0	0\0	0\0	0\0	0\0	0\0
C/S.Africa (14)	0\0	0\0	0\0	0\0	0\0	0\0	0\0	85.9\94

Lazaridis (1): Spanish_Canarias_IBS, Spanish_Cantabria_IBS, Sardinian, French_South, Greek_Comas, Greek_Coriell, Spanish_Galicia_IBS, Spanish_Baleares_IBS, Italian_Tuscan, Italian_Bergamo, Italian_WestSilician; Busby (1): Basque, EastSicilian, Greek, ItalianN, ItalianS, Sardinian, SicilianW, Spanish, Tsi, Tuscan;
Lazaridis (2): Albanian, Bulgarian, Scottish_Argyll_Bute_GBR, Icelandic, Norwegian; Busby (2): Ceu, French, Ireland;
Lazaridis (3): Cypriot; Busby (3): Cypriot;
Lazaridis (4): Lithuanian, Kumyk, Estonian; Busby (4): Finnish;
Lazaridis (5): Jordanian, Lebanese; Busby (5): Mawasi;
Lazaridis (6): Balochi, Sindhi, Pathan, Iranian, Iran_Fars; Busby (6): Iranian, Makrani, Pathan;
Lazaridis (7): Armenian, Abkhasian, Georgian_Megrels, Kumyk; Busby (7): Abhkasian, Armenian, Armenian_LebArmenian, Armenian_Yegvard, Armenian_Yerevan, Balkar, Georgian, Kumyk;
Lazaridis (8): Onge, GujaratiB_GIH, GujaratiC_GIH, Mala, Bengali_Bangladesh_BEB; Busby (8): Bhunjia, Dhurwa, Meghawal;
Lazaridis (9): Kyrgyz, Uzbek, Mongola; Busby (9): Kyrgyz, Mongola;
Lazaridis (10): Korean, Tu, Han, Cambodian; Busby (10): Han, Japanese;
Lazaridis (11): Australian_ECCAC, Bougainville, Papuan; Busby (11): Melanesian, Papuan;
Lazaridis (12): Mayan;
Lazaridis (13): Egyptian_Comas; Busby (13): Egyptian
Lazaridis (14): Gana, Hadza_Henn, Haiom, MbutiPygmy, Taa_North, Ju_hoan_North, BiakaPygmy, Nama, Dinka, Sandawe, Datog, Gambian_GWD, Oromo, Masai_Ayodo; Busby (14): BantuSA, BiakaPygmy, Ethiopian, EthiopianJew, Hadza, Maasai, Mandenka, MbutiPygmy, Sandawe, SanNamibia.

Table S23

Inferred proportions of WC1 ancestry under analysis (II), analysis (II-) and analysis (III) for modern populations in the Lazaridis (upper) and Busby (lower) merges that receive a >10% contribution from WC1 under analysis (II-) (as shown in Fig. S27). While contributions tend to increase (as expected) from analyses (II) to (II-) to (III), they do not always. This is perhaps most striking in Iran_Zoroastrian, and appears to be attributable to allowing self-copying that accounts for high levels of group-specific drift (see methods) under analysis (III) only. E.g. allowing such self-copying under analysis (II-) changes the contribution from WC1 in Iran_Zoroastrian to 28.8% and disallowing self-copying under analysis (III) changes the contribution to 100%.

	Analysis(II)	Analysis(II-)	Analysis(III)
Lazaridis merge			
Iranian	24.07	28.31	66.45
Iran_Fars	14.99	44.28	72.54
Iran_Zoroastrian	0	100	72.04
Makrani	0.02	94.84	97.81
Balochi	0	86.7	100
Sindhi	0	19.95	84.37
Pathan	7.06	17.54	80.4
Turkmen	12.68	-	52.14
Yemen	14.44	-	44.48
Busby merge			
Iranian	40.16	-	65.53
Balochi	0	54.54	98.57
Brahui	0	50.35	98.1
Kalash	10.01	55.75	83.96
Makrani	0	68.52	94.48
Pathan	31.56	30.39	76
Sindhi	0	18.2	86.28
Turkmen	38.41	-	61.3
Tajik	44.53	-	63.17
Armenian_Chambarak	0	15.08	39.25
Armenian_Dprabak	0	16.64	39.36
Armenian_Gavar	0	11.93	41.19
Armenian_Martuni	0	18.44	38.92
Armenian_Yegvard	0	14.81	40.48
Armenian_Yerevan	0	11.68	39.28

Table S24

Ancestry proportions. Inferred proportions of ancestry for all ancient individuals and modern populations (Lazaridis merge) under Analysis I, II, II-, and III as described in supplementary section 9 (Available online as a separate excel file)

Table S25

Ancestry proportions. Inferred proportions of ancestry for all ancient individuals and modern populations (Busby merge) under Analysis I, II, II-, and III as described in supplementary section 9 (Available online as a separate excel file).

Table S26

Overlapping positions between the WC1 genome and the (19) dataset. Generation time was calculated by adding 2015 years to the mean cal. age (BCE) and assuming 25 years per generation.

Ancient genome	Mean Cal. Age (BCE)	Epoch	Generations	Overlapping SNPs
WC1	7,424	Early Neolithic	378	454,970

Table S27

Allele sharing classes.The nine possible allele sharing classes for the comparison of the biallelic calls of the ancient genome (t_0) and the biallelic calls of each modern genome (t_n) for the same position.

	Match		Mismatch		Half match				Match
	a	b	a	b	a	b	c	d	het
t_0	AA	DD	AA	DD	AA	DD	AD	AD	AD
t_n	AA	DD	DD	AA	AD	AD	AA	DD	AD

Table S28

Modern populations tested for continuity. Number of samples and modern population size of each modern population tested for population continuity.

Modern population	Number of samples	Population size (x10^4)
Balochi	20	950
Brahui	20	220
Makrani	20	3.75
Iranian	17	7,800
Iran Zoroastrian	30	2.5
Megrelians	10	40
Armenian	13	300
Turkish Balikesir	8	34
Turkish Adana	10	300
Sardinian	27	160

Table S29

Markers associated with skin, hair, and eye pigmentation.

SNP	Gene	Ancestral/ Derived Alleles	WC1	AH1	AH2	AH4	F38
SLC24A5 16-marker haplotype							
rs1834640	*SLC24A5*	G/A	A/G(7X)	-	-	-	-
rs2675345	*SLC24A5*	G/A	A/G (19X)	-	-	-	A/A (3X)
rs2469592	*SLC24A5*	G/A	A/A(9X)	A/A (6X)	A/A (2X)	-	A/A (3X)
rs2470101	*SLC24A5*	C/T	T/T (16X)	-	-	-	-
rs938505	*SLC24A5*	C*/T	C/C(10X)	-	-	-	C/C (2X)
rs2433354	*SLC24A5*	T/C	C/C(11X)	-	C/T(2X)	C/C (3X)	-
rs2459391	*SLC24A5*	G/A	A/A (9X)	A/A (2X)	-	A/A (2X)	A/A (7X)
rs2433356	*SLC24A5*	A/G	G/G (6X)	-	-	-	-
rs2675347	*SLC24A5*	G/A	A/A (16X)	A/A (2X)	-	-	-
rs2675348	*SLC24A5*	G/A	A/A(4X)	-	-	-	A/A (2X)
rs1426654	*SLC24A5*	G/A	A/G (13X)	A/G (4X)	G/G (4X)	A/A (2X)	A/A (4X)
rs2470102	*SLC24A5*	G/A	A/G (11X)	-	G/G (2X)	-	A/A (2X)
rs16960631	*SLC24A5*	A*/G	A/A (17X)	A/A(2X)	A/A (2X)	A/A (4X)	A/A (2X)
rs2675349	*SLC24A5*	G/A	A/A (4X)	-	-	-	A/A(2X)
rs3817315	*SLC24A5*	T/C	C/C (9X)	C/C (2X)	-	C/C (2x)	C/C (2X)
rs7163587	*SLC24A5*	T/C	C/C (7X)	C/C (2X)	-	C/C (2X)	C/C (2X)
*ancestral allele part of the C11 haplotype							
HERC2 haplotypes							
beh1			WC1	AH1	AH2	AH4	F38
rs4778138	*HERC2*	G/A	A/G(8X)	A/A (3X)	-	G/G (2X)	A/A(5X)
rs4778241	*HERC2*	A/C	A/C (13X)	-	-	-	-
rs7495174	*HERC2*	G/A	A/G (13X)	A/A (3X)	-	-	-
beh2			WC1	AH1	AH2	AH4	F38
rs1129038	*HERC2*	C/T	C/T (15X)	C/T (2X)	C/C (2X)	-	T/T (4X)
rs12913832	*HERC2*	A/G	A/G (20X)	-	A/A (3X)	-	-
beh3			WC1	AH1	AH2	AH4	F38
rs916977	*HERC2*	T/C	T/C (20X)	-	T/C (3X)	T/T(2X)	C/C (4X)
rs1667394	*HERC2*	C/T	C/T (11X)	-	-	C/C (2X)	-
HERC2 13-marker haplotype							
rs4778241	*HERC2*	A/C	A/C (13X)	-	-	-	-
rs1129038	*HERC2*	C/T	C/T (15X)	C/T (2X)	C/C (2X)	-	T/T (4X)
rs12593929	*HERC2*	G/A	G/A (9X)	-	G/A (3X)	-	A/A (3X)
rs12913832	*HERC2*	A/G	A/G (20X)	-	A/A (3X)	-	-
rs7183877	*HERC2*	C*/A	C/C(8X)	-	-	-	-
rs3935591	*HERC2*	T/C	T/C (13X)	-	-	-	C/C(4X)
rs7170852	*HERC2*	T/A	T/A (9X)	A/A(6X)	-	-	A/A(4X)
rs2238289	*HERC2*	G/A	A/G (15X)	-	-	-	A/A(4X)
rs3940272	*HERC2*	T/G	T/G (9X)	-	-	-	-
rs8028689	*HERC2*	C/T	T/C(12X)	-	T/T (3X)	T/C (3X)	-
rs2240203	*HERC2*	C/T	T/C (12X)	-	-	C/C (2X)	T/T (4X)
rs11631797	*HERC2*	A/G	A/G (14X)	-	A/A (3X)	-	G/G (2X)
rs916977	*HERC2*	T/C	T/C (20X)	-	T/C (3X)	T/T(2X)	C/C (4X)
*ancestral allele part of the blue-eye-associated haplotype							

Hirisplex SNPs							
SNP	Gene	Ancestral/ Derived Alleles	WC1	AH1	AH2	AH4	F38
n29insa	*MC1R*	C/insA	C/C (16X)	-	C/C (2X)	C/C (3X)	G/G (4X)
rs11547464	*MC1R*	G/A	G/G (15X)	G/G(3X)	G/G (2X)	G/A (2X)	-
rs885479	*MC1R*	G/A	G/G (9X)	G/G (2X)	G/G (2X)	G/G (2X)	G/G (2X)
rs1805008	*MC1R*	C/T	C/C (8X)	C/C (2X)	C/C (2X)	C/C (2X)	-
rs1805005	*MC1R*	G/T	G/G (12X)	G/G (2X)	-	-	G/G (5X)
rs1805006	*MC1R*	C/A	C/C(14X)	-	-	C/C (3X)	C/C (2X)
rs1805007	*MC1R*	C/T	C/C (14X)	C/C (3X)	-	C/C (3X)	-
rs1805009	*MC1R*	G/C	G/G (15X)	-	-	G/G (2X)	-
y152och	*MC1R*	C/A	C/C(14X)		-	C/C(3X)	-
rs2228479	*MC1R*	G/A	G/A (13X)	-	G/G (2X)	G/G (2X)	G/G (4X)
rs1110400	*MC1R*	T/C	T/T (10X)	T/T (2X)	-	T/T(3X)	-
rs28777	*SLC45A2*	C/A	C/C (10X)	-	C/C (3X)	-	A/A(2X)
rs16891982	*SLC45A2*	C/G	C/C(13X)	-	-	-	-
rs12821256	*KITLG*	T/C	T/T(12X)	-	T/T (2X)	-	T/T (2X)
rs4959270	*EXOC2*	C/A	C/A(7X)	-	-	-	A/A(5X)
rs12203592	*IRF4*	C/T	C/C (22X)	-	-	C/C (3X)	C/A(4X)
rs1042602	*TYR*	C/A	C/C (19X)	C/C (2X)	-	-	C/C (3X)
rs1800407	*OCA2*	C/T	C/C (19X)	-	-	-	C/C (2X)
rs2402130	*SLC24A4*	G/A	A/A (15X)	A/A (4X)	A/A (4X)	A/A (4X)	G/A (4X)
rs12913832	*HERC2*	A/G	A/G (20X)	-	A/A (3X)	-	-
rs2378249	*PIGU/ASIP*	A/G	G/A (12X)	-	-	-	G/A (3X)
rs12896399	*SLC24A4*	G/T	T/T (13X)	-	-	-	G/G (2X)
rs1393350	*TYR*	G/A	G/G (16X)	-	-	-	G/G (5X)
rs683	*TYRP1*	A/C	-	-	-	-	-

Table S30

Markers associated with xenobiotic metabolism, alcohol metabolism, and lactase persistence.

SNP	Gene	Ancestral/ Derived alleles	WC1	AH1	AH2	AH4	F38
rs4988235	*LCTa*	G/A	G/G** (11X)	-	-	G/G (3X)	G/G(2X)
rs182549	*LCTb*	C/T	C/C (19X)	C/C** (3X)	-	C/C**(5X)	C/C (2X)
rs3811801	*ADH1Ba*	G/T	G/G (9X)	G/G (8X)	G/G (2X)	-	G/G (3X)
rs1229984	*ADH1Bb*	C/T	C/C(18X)	-	-	-	-
rs671	*ALDH2*	G/A	G/G (17X)	G/G (2X)	G/G (2X)	-	G/G (5X)
rs1801279	*NAT2*	G/A	G/G(13X)	-	-	-	G/G(2X)
rs1041983	*NAT2*	C/T	C/C(9X)	T/T (2X)	-	-	-
rs1801280	*NAT2*	T/C	T/T(12X)	T/T (3X)	-	T/T(2X)	T/T (3X)
rs1799929	*NAT2*	C/T	C/C (5X)	C/C (3X)	-	-	C/C (4X)
rs1799930	*NAT2*	G/A	G/G (5X)	-	A/A (2X)	-	-
rs1208	*NAT2*	A/G	A/A (9X)	-	-	-	-
rs1799931	*NAT2*	G/A	G/G(10X)	G/G (2X)	-	-	-
rs1495741	*NAT2*	G/A	G/G (13X)	A/A (3X)	-	-	-
rs699 (M235T)	*AGT*	G°/A	A/G (16X)	A/G(2X)	-	-	-
rs5051 (A-6G)	*AGT*	T*/C	C/T (7X)	-	-	-	-
rs4762	*AGT*	G/A°	G/G (19X)	-	-	-	G/A (2X)
rs5049 (G217A)	*AGT*	C/T*	C/C (17X)	-	-	-	C/C (3X)
Rs5443 (C825T)	*GNB3*	C/T*	C/T (17X)	T/T (2X)	-	-	C/C (5X)
rs1042713 (G47A)	*ADRB2*	G/A*	G/A (8X)	A/A (4X)	-	A/A (3X)	G/G (2X)
rs1024714 (G79C)	*ADRB2*	G/C*	G/C (8X)	-	-	G/C (2X)	G/C (2X)
rs3759324 (A-946G)	*ENaCα/ SCNN1α*	T/C*	T/C (12X)	-	T/T (2X)	T/T (2X)	T/T (2X)
rs5718 (A-173G)	*ENaCγ / SCNN1γ*	G*/A	A/G (11X)	G/G (2X)	-	-	G/G (2X)

SNP	Gene	Ancestral/ Derived alleles	WC1	AH1	AH2	AH4	F38
rs2740574 (*CYP3A4*1/ CYP3A4*1B*)	*CYP3A4*	C/T	T/T (21X)	T/T (4X)	C/C (2X)	-	T/T (4X)
rs776746 (CYP3A5*1/ CYP3A5*3)	*CYP3A5*	T/C	C/C(17X)	-	C/C(2X)	-	-
rs283656083 (CYP3A5*2)	*CYP3A5*	G/T	G/G(18X)	-	G/G(2X)	G/G (2X)	G/G (2X)
rs54611401 (CYP3A5*4)	*CYP3A5*	T/C	T/T (13X)	T/T (2X)	-	-	T/T (3X)
rs10264272 (CYP3A5*6)	*CYP3A5*	C/T	C/C (14X)	-	-	-	-
rs41303343 (CYP3A5*7)	*CYP3A5*	--/A	T/T(8X)	-	-	T/T (2X)	T/T (3X)
rs55817950 (CYP3A5*8)	*CYP3A5*	G/A	G/G (6X)	G/G(3X)	-	-	G/G (2X)
rs28383479 (CYP3A5*9)	*CYP3A5*	C/T	C/C (14X)	-	-	C/C (2X)	-
rs41279854 (CYP3A5*10)	*CYP3A5*	A/G	A/A (10X)	-	-	-	A/A (2X)
rs72552791 (CYP3A5*11)	*CYP3A5*	T/C	T/T (7X)	T/T (2X)	T/T (3X)	-	T/T (5X)
rs56244447 (CYP3A5*3D)	*CYP3A5*	A/C	A/A (17X)	A/A (2X)	A/A (2X)	A/A (2X)	A/A (3X)
rs28365085 (CYP3A5*3F)	*CYP3A5*	A/G	A/A (12X)	A/A (3X)	A/A (2X)	A/A (2X)	-
rs28383468 (H30Y)	*CYP3A5*	G/A	G/G (5X)	G/G (3X)	-	-	G/G (2X)
rs41279857 (S100Y)	*CYP3A5*	G/A,C,T	G/G (11X)	G/G (2X)	-	-	G/G (2X)
rs15524	*CYP3A5*	G/A	A/A (14X)	A/A(3X)	-	-	-

*allele associated with heat adaptation (149)
°risk allele
** as described in the text, due to the small number of derived reads present at these sites, the ancestral allele is reported here

Table S31.1

Markers related to metabolism and susceptibility to metabolic disease (only Iranian samples).

SNP	Gene	Ancestral/ Derived alleles	WC1	AH1	AH2	AH4	F38	Estimated sweep age (ya) (150)
rs7903146	*TCF7L2*	T/C	C/C (5X)	-	-	-	C/C (2X)	~11,900
rs10885406	*TCF7L2*	G/A	A/G (13X)	-	-	-	G/G (2X)	~11,900
rs12255372	*TCF7L2*	G/T	G/G (22X)	G/T (2X)	-	-	G/T (2X)	~11,900
rs7924080	*TCF7L2*	C/T	T/C (12X)	C/C (2X)	-	-	C/C (6X)	~11,900

Table S31.2

Markers included in a 62-SNP type 2 diabetes (T2D) genetic risk score (*151, 162*).

SNP	Gene	Nonrisk/risk alleles	WC1	Bar8	Loschbour	Stuttgart
rs10923931	NOTCH2	G/T	G/G(13X)	G/T(7X)	G/T (16X)	G/T(20X)
rs2075423	PROX1	T/G	G/G(7X)	G/T (10X)	G/T (26X)	T/T(16X)
rs243088	BCL11A	A/T	T/T(6X)	A/A (8X)	A/A (16X)	A/T(7X)
rs780094	GCKR	T/C	T/C(11X)	T/T (6X)	C/C (22X)	T/T (11X)
rs13389219	GRB14	T/C	C/C(11X)	C/C (8X)	C/C (26X)	T/T (30X)
rs2943640	IRS1	A/C	C/C(7X)	A/A (8X)	A/C(20X)	A/C (17X)
rs7569522	RBMS1	G/A	A/A(12X)	-	A/A(27X)	A/G(12X)
rs10203174	THADA	T/C	C/C(9X)	C/C (5X)	C/C(27X)	C/C (25X)
rs6795735	ADAMTS9	T/C	T/T(10X)	T/T (5X)	C/C(28X)	C/C(24X)
rs11717195	ADCY5	C/T	T/T(16X)	T/T (13X)	T/C(23X)	T/T (28X)
rs4402960	IGF2BP2	G/T	G/T(10X)	T/T (6X)	G/T (15X)	G/G (18X)
rs1801282	PPARG	G/C	C/C(9X)	C/C (5X)	C/C (26X)	C/C (14X)
rs12497268	PSMD6	C/G	G/G(14X)	G/G(6X)	G/C (19X)	G/G (10X)
rs17301514	ST64GAL1	G/A	G/G(7X)	G/G(12X)	G/G(14X)	G/G(11X)
rs1496653	UBE2E2	G/A	A/G(6x)	A/A (8X)	A/A(30X)	A/A (24X)
rs6819243	MAEA	C/T	T/T(19X)	T/T (14X)	T/T (12x)	T/T (21X)
rs4458523	WFS1	T/G	G/G(3X)	G/G(4X)	T/T (8X)	T/G (23X)
rs459193	ANKRD55	A/G	A/G(18X)	G/G(10X)	A/A (23X)	A/G(22X)
rs6878122	ZBED3	A/G	A/A(12X)	A/A (9X)	A/A (30X)	G/A (32X)
rs7756992	CDKAL1	A/G	A/G(13X)	A/G(6X)	A/G (35X)	A/G (27X)
rs3734621	KCNK16	A/C	A/A(14X)	A/A (11X)	A/A (16X)	A/C (20X)
rs4299828	ZFAND3	G/A	A/A(8x)	A/A (12X)	A/A (20X)	G/A (17X)
rs17168486	DGKB	C/T	T/T(16X)	C/T (3X)	C/C (15X)	C/C (12X)
rs17867832	GCC1	G/T	T/T(5X)	T/T (8X)	T/T (28X)	T/T (18X)
rs10278336	GCK	G/A	G/G(9X)	A/G (10X)	A/G (25X)	A/A (18X)
rs849135	JAZF1	A/G	G/A(9X)	G/G (2X)	G/A (14X)	A/A (19X)
rs13233731	KLF14	A/G	G/A(17X)	G/G(5X)	G/G(16X)	G/A (14X)
rs516946	ANK1	T/C	T/C(10X)	C/C (2X)	T/C (11X)	T/C (10X)
rs3802177	SLC30A8	A/G	G/G(6X)	G/G(13X)	G/G (32X)	G/G (16X)
rs7845219	TP53INP1	C/T	T/T(13X)	T/T (11X)	T/T (20X)	T/T (26X)
rs10811661	CDKN2A/B	C/T	T/T(17X)	T/T (6X)	T/C (19X)	T/T (24X)
rs10758593	GLIS3	G/A	G/A(9X)	G/G (20X)	G/A (28X)	G/A (34X)
rs16927668	PTPRD	C/T	C/C(12X)	C/C (8X)	C/C (21X)	C/T (26X)
rs2796441	TLE1	A/G	G/A(12X)	G/G (7X)	G/A (30X)	A/A (38X)
rs17791513	TLE4	G/A	A/A(8X)	A/A (4X)	A/A (22X)	A/A (19X)

SNP	Gene	Nonrisk/risk alleles	WC1	Bar8	Loschbour	Stuttgart
rs11257655	CDC123/CAMK1D	C/T	C/C(11X)	C/T (4X)	C/T (14X)	C/C (9X)
rs1111875	HHEX/IDE	T/C	C/C(4X)	C/C (3X)	C/C (17X)	T/T (18X)
rs7903146	TCF7L2	T/C	C/C(5X)	C/C (3X)	C/C (22X)	C/C (14X)
rs12242953	VPS26A	A/G	G/G(12X)	G/A (3X)	G/A (10X)	G/G (24X)
rs12571751	ZMIZ1	G/A	A/A(16X)	G/G(3X)	A/G (11X)	A/G (18X)
rs1552224	ARAP1/ (CENTD2)	C/A	A/A(18X)	A/A (8X)	A/C (21X)	A/C (25X)
rs2334499	DUSP8	C/T	T/T(7X)	C/C (9X)	C/T (19X)	C/T (14X)
rs5215	KCNJ11	T/C	C/C(19X)	T/T (5X)	C/T (13X)	C/T (17X)
rs163184	KCNQ1	T/G	G/G(13X)	T/G (4X)	G/G (9X)	T/G (11X)
rs10830963	MTNR1B	C/G	C/G(12X)	C/G (4X)	G/G (17X)	C/G (19X)
rs2261181	HMGA2	C/T	C/C(12X)	C/C (8X)	C/C (33X)	C/C (24X)
rs12427353	HNF1A(TCF1)	C/G	G/G(10X)	G/G (4X)	G/G (9X)	C/C (18X)
rs10842994	KLHDC5	T/C	C/C(14X)	T/T (10X)	C/C (20X)	C/T (17X)
rs7955901	TSPAN8/LGR5	T/C	C/C(6X)	C/C (4X)	T/T (12X)	C/T (17X)
rs1359790	SPRY2	A/G	G/G(8X)	G/A (10X)	G/G (31X)	G/A (17X)
rs2007084	AP3S2	A/G	G/G(15X)	G/G (10X)	G/G (21X)	G/G (8X)
rs4502156	C2CD4A	C/T	T/C(8X)	T/C (7X)	T/C (21X)	T/C (24X)
rs7177055	HMG20A	G/A	G/G(13X)	G/A (2X)	A/A(8X)	G/A (16X)
rs12899811	PRC1	A/G	A/A(13X)	A/G (4X)	A/G (17X)	G/G (20X)
rs11634397	ZFAND6	A/G	A/A(9X)	A/A (8X)	A/G (27X)	G/G(32X)
rs7202877	BCAR1	G/T	T/T(13X)	T/T (4X)	T/T (12X)	T/T (19X)
rs9936385	FTO	T/C	T/T(13X)	T/C (9X)	C/C (17X)	T/C (25X)
rs2447090	SRR	G/A	A/G(11X)	G/G (2X)	A/G (10X)	A/A (15X)
rs12970134	MC4R	G/A	G/A(15X)	G/G (8X)	G/A (14X)	G/G (12X)
rs10401969	CILP2	T/C	T/T(4X)	T/T (8X)	T/T (16X)	T/T (26X)
rs8182584	PEPD	G/T	T/G(24X)	G/G (3X)	G/G (19X)	T/G (20X)
rs4812829	HNF4A	G/A	G/A(19X)	G/A(7X)	G/A(11X)	G/G (13X)
	Count GRS		80	69-71*	75	67
	Weighted GRS		87.51	75.92-78.02*	82.06	73.31
*range given to account for missing position						

Table S32

Markers related to pathogen resistance, susceptibility to infectious and non-infectious disease.

SNP	Gene	Ancestral/ Derived alleles	WC1	AH1	AH2	AH4	F38	Estimated sweep age (ya) (154, 155)
rs35004220 (IVS-I-110)	*HBB*	C/T	C/C (18X)	C/C (5X)	-	C/C(2X)	C/C (3X)	_
rs35724775 (IVS-I-6)	*HBB*	A/G	A/A (11X)	A/A (3X)	-	-	A/A (5X)	_
rs35497102 (FSC-8)	*HBB*	TT/-	TT(7X)	TT(2X)	-	TT(4X)	TT(4X)	-
rs33971440 (IVS-I-1)	*HBB*	C/T	C/C (10X)	C/C (3X)	-	-	C/C (5X)	-
rs34690599 (IVS-II-745)	*HBB*	G/C	G/G (12X)	G/G (2X)	-	-	G/G (2X)	-
rs10786436	*HPSE2*	C/T	C/C(16X)	-	-	-	C/C (2X)	~6,560
rs1132200 *	*ARHGAP31/ STAT1*	C/T	C/C (15X)	C/C (4X)	-	C/C (3X)	C/C (3X)	~2,200
rs12638253	*LEKR1*	C/T	T/T (4X)	-	-	-	C/T (3X)	~5,100
rs12722489*	*IL2RA*	C/T	C/C (14X)	-	-	-	C/C (4X)	~1,200
rs17696736*	*TMEM116*	A/G	A/A (9X)	-	-	-	A/A (3X)	~1,400
rs17810546*	*SCHIP1/ IL12A*	A/G	A/A (9X)	A/A (2X)	-	-	A/A (2X)	~2,310
rs2058660	*IL18RAP*	A/G	G/A (10X)	-	-	-	G/A (2X)	~7,500
rs2188962*	*SLC22A5/ IRF1*	C/T	C/C (21X)	C/C (7X)	C/C (2X)	C/C (3X)	C/C (7X)	~1,380
rs2248359	*CYP24A1*	T/C	C/T (13X)	T/T (2X)	-	T/T (2X)	T/T (5X)	~8,500
rs2285795	*TRIM10*	T/C	C/T (11X)	-	C/C (2X)	-	C/C (7X)	~2,280
rs307896	*SAE1*	G/A	G/G (12x)	-	-	-	G/A (4X)	~2,700
rs3129934	*BTNL2*	C/T	C/C (5X)	C/C (3X)	-	-	C/C (2X)	~3,300
rs3131379	*VARS/ LSM2*	G/A	G/G (10X)	-	-	-	G/G (2X)	~1,980
rs3184504	*SH2B3/ ATXN2*	C/T	C/C (15X)	C/C (3X)	-	-	C/C (3X)	~1,500
rs6822844*	*IL2/IL21*	G/T	G/G (14X)	-	-	-	G/T (2X)	~2,150
rs6897932 *	*IL7R*	T/C	C/C (7X)	T/T (2X)	-	-	C/C (2X)	~1,800
rs744166 *	*STAT3*	G/A	A/A (9X)	-	-	-	G/G (3X)	~2,600
rs1050152	*SLC22A4*	C/T	C/C (12X)	C/C (3X)	-	C/C (2X)	C/C (3X)	~12,500
rs2631367	*SLC22A5*	G/C	G/G (6X)	-	G/G (2X)	-	-	~12,500
rs11739623	*IRF1/IL5*	C/T	C/C (13X)	C/C (2X)	C/C (3X)	-	C/C (3X)	~12,500
SNPs in core selected network identified by Raj et al. (154)								

Table S33

Additional markers identified as selected in modern and ancient populations.

SNP	Gene	Ancestral/ Derived alleles	WC1	AH1	AH2	AH4	F38
rs 2269424*	*MHC region*	G/A	G/A (18X)	G/A (2X)	-	-	G/G (2X)
rs 174546*	*FAD1/FADS2*	T/C	C/T (14X)	-	T/T (2X)	C/C (2X)	C/C (2X)
rs 4833103*	*TLR1/TLR6/TLR10*	C/A	C/C (7X)	-	-	-	-
rs 653178*	*ATXN2/SH2B3*	T/C	T/T(12X)	-	-	T/T (2X)	T/T(3X)
rs 7944926*	*DHCR7/NADSYN1*	G/A	A/A (20X)	-	-	-	A/A (2X)
rs 7119749*	*GRM5*	A/G	A/G (15X)	G/G (4x)	-	G/G (3X)	A/G (3X)
rs 272872*	*SLC22A4*	G/A	A/A (16X)	-	A/A (2X)	-	A/A (2X)
rs 6903823*	*ZKSCAN3/ ZSCAN31*	A/G	A/A (12X)	-	-	-	A/A (5X)
rs 1979866*	-	C/A	A/A (9X)	A/A (3X)	-	-	A/A (5X)
* identified as selected in ancient Eurasians by Mathieson *et al.(3)*							

Table S34

G-PhoCS parameter estimates when including Yoruba and fixed age for ancient samples. θ and τ and scaled by 10,000x. N_e and year estimates assume a μy, of 0.5 x 10^9 and human generation time of 25 years.

Parameter	mean	95% lower	95% upper	N_e
θ_A	5.15	0.93	9.55	11,447 (2,064-21,212)
θ_B	4.63	2.65	6.82	10,297 (5,892-15,165)
θ_C	0.66	0.1	1.14	1,464 (227-2,541)
θ_D	46.83	24.71	73.62	104,070 (54,918-163,606)
θ_{AC}	6.17	0.8	11.62	13,715 (1,771-25,831)
θ_{ACB}	2.86	1.51	4.13	6,359 (3,356-9,172)
θ_{ACBD}	6.26	6	6.51	13,916 (13,340-14,470)
				Years
τ_{AC}	0.18	0.07	0.27	39 (15-61)
τ_{ACB}	0.35	0.24	0.47	77 (53-104)
τ_{ACBCD}	0.7	0.62	0.78	156 (138-173)

Table S35

G-PhoCS parameter estimates when excluding Yoruba and fixed age for ancient samples. θ and τ and scaled by 10,000x. N_e and year estimates assume a μ_y, of 0.5 x 10^{-9} and human generation time of 25 years.

Parameter	mean	95% lower	95% upper	N_e
θ_A	3.32	1.63	4.88	7,369 (3,630-10,841)
θ_B	2.39	1.88	2.82	5,306 (4,176-6,275)
θ_C	0.66	0.37	0.92	1,468 (824-2,052)
θ_{AC}	2.77	0.79	5.11	6,162 (1,757-1,1357)
θ_{ACB}	5.57	5.48	5.65	12,374 (12,180-12,566)
				Years
τ_{AC}	0.15	0.11	0.19	33 (23-42)
τ_{ACB}	0.21	0.17	0.23	46 (38-50)

Table S36

G-PhoCS parameter estimates when excluding Yoruba, fixed age for ancient samples and Bar8 is replaced with NE1. θ and τ and scaled by 10,000x. N_e and year estimates assume a μ_y, of 0.5 x 10^9 and human generation time of 25 years.

Parameter	mean	95% lower	95% upper	N_e
θ_A	5.09	1.03	10.22	11,300 (2,298-22,712)
θ_B	2.17	0.68	3.49	4,827 (1,508-7,754)
θ_C	0.34	0.04	0.65	766 (82-1,440)
θ_{AC}	2.35	0.1	4.94	5,233 (225-10,983)
θ_{ACB}	5.52	5.29	5.76	12,258 (11,764-12,804)
				Years
τ_{AC}	0.11	0.06	0.17	25 (12-37)
τ_{ACB}	0.17	0.09	0.22	37 (21-49)

Table S37

Fitted versus estimate *f2* statistics.

Pop1	Pop2	Fitted	Estimated	Difference	Std.err	Z score
Mbuti	UstIshim	0.162	0.161	-2.70E-04	5.40E-03	-0.049
Mbuti	Malta1	0.597	0.596	-1.10E-03	2.60E-03	-0.419
Mbuti	Kostenki14	0.594	0.593	-3.30E-04	2.70E-03	-0.12
Mbuti	GoyetQ116-1	0.599	0.6	5.40E-04	2.60E-03	0.204
Mbuti	WC1	0.19	0.19	-3.90E-05	4.00E-03	-0.01
UstIshim	Malta1	0.458	0.461	3.40E-03	5.90E-03	0.575
UstIshim	Kostenki14	0.455	0.458	2.90E-03	5.90E-03	0.486
UstIshim	GoyetQ116-1	0.46	0.452	-8.30E-03	5.90E-03	-1.412
UstIshim	WC1	0.079	0.08	8.60E-04	6.00E-03	0.144
Malta1	Kostenki14	0.851	0.856	5.50E-03	4.50E-03	1.214
Malta1	GoyetQ116-1	0.856	0.851	-4.90E-03	4.70E-03	-1.048
Malta1	WC1	0.481	0.482	5.40E-04	4.70E-03	0.115
Kostenki14	GoyetQ116-1	0.837	0.837	-2.00E-04	5.30E-03	-0.038
Kostenki14	WC1	0.497	0.494	-2.80E-03	4.40E-03	-0.625
GoyetQ116-1	WC1	0.502	0.506	4.40E-03	5.10E-03	0.865

Animation S1

Interactive 3D version of the PCA of modern and projected ancient samples shown in Fig. 2