Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia

Michal Feldman[1], Eva Fernández-Domínguez[2*], Luke Reynolds[3], Douglas Baird[4], Jessica Pearson[4], Israel Hershkovitz[5,6], Hila May[5,6], Nigel Goring-Morris[7], Marion Benz[8], Julia Gresky[9], Raffaela A. Bianco[1], Andrew Fairbairn[10], Gökhan Mustafaoğlu[11], Philipp W. Stockhammer[1,12], Cosimo Posth[1], Wolfgang Haak[1], Choongwon Jeong[1*] and Johannes Krause[1*]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Strasse 10, D-07745, Jena, Germany.

[2]Department of Archaeology, Durham University, Durham, South Road, DH1 3LE, United Kingdom.

[3]School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, Byrom Street, L3 3AF, United Kingdom.

[4]Department of Archaeology, Classics and Egyptology, 8-14 Abercromby Square, University of Liverpool, L69 7WZ, United Kingdom.

[5]Department of Anatomy and Anthropology, The Dan David Center for Human Evolution and Biohistory Research and The Shmunis Family Anthropology Institute, Sackler Faculty of Medicine, Tel Aviv University, Post Office Box 39040, Tel Aviv 6997801, Israel.

[6]The Steinhardt Museum of Natural History, Tel Aviv University, Post Office Box 39040, Tel Aviv 6997801, Israel

[7]Department of Prehistory, Institute of Archaeology, The Hebrew University of Jerusalem, Jerusalem 91955, Israel.

[8]Department of Near Eastern Archaeology, Free University Berlin, Fabeckstrasse 23-25, 14195 Berlin.

[9]Department of Natural Sciences, German Archaeological Institute, 14195 Berlin, Germany.

[10]School of Social Science, The University of Queensland, Michie Building, St Lucia, Brisbane, Queensland, Australia.

[11]Department of Archaeology, Zonguldak Bülent Ecevit University, 67100 Incivez, Zonguldak, Turkey.

[12]Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie Ludwig-Maximilians-Universität München München, Schellingstrasse 12, D-80799, Germany.

*Correspondence to: eva.fernandez@durham.ac.uk (E.F.D); jeong@shh.mpg.de (CJ); krause@shh.mpg.de (JK)

Abstract: Anatolia was home to some of the earliest farming communities. It has been long debated whether a migration of farming groups introduced agriculture to central Anatolia. Here, we report the first genome-wide data from a 15,000-year-old Anatolian hunter-gatherer and from seven Anatolian and Levantine early farmers. We find high genetic continuity (~80-90%) between the hunter-gatherer and early farmers of Anatolia and detect two distinct incoming ancestries: an early Iranian/Caucasus related one and a later one linked to the ancient Levant. Finally, we observe a genetic link between southern Europe and the Near East predating 15,000 years ago that extends to central Europe during the post-last-glacial maximum period. Our results suggest a limited role of human migration in the emergence of agriculture in central Anatolia.

The practice of agriculture began in the Fertile Crescent of Southwest Asia as early as 10,000 to 9,000 BCE. Subsequently, it spread across western Eurasia while increasingly replacing local hunting and gathering subsistence practices, reaching central Anatolia by c. 8300 BCE[1-3].

Recent genetic studies have shown that in mainland Europe, farming was introduced by an expansion of early farmers from Anatolia that replaced much of the local populations[4, 5]. Such mode of spread is often referred to as the demic diffusion model. In contrast, in regions of the Fertile Crescent such as the southern Levant and the Zagros Mountains (located between present-day eastern Iraq and western Iran) the population structure persists throughout the Neolithic

transition[6], indicating that the hunter-gatherers of these regions locally transitioned to a food producing subsistence strategy.

Central Anatolia has some of the earliest evidence of agricultural societies outside the Fertile Crescent[3] and thus is a key region in understanding the early spread of farming. While archaeological evidence points to cultural continuity in central Anatolia[3], due to the lack of genetic data from pre-farming individuals it remains an open question whether and to what scale the development of the Anatolian Neolithic involved immigrants from earlier farming centers admixing with the local hunter-gatherers.

Here, we report new genome wide data from eight prehistoric humans (Fig.1A, Table 1, table S1), including the first Epipalaeolithic Anatolian hunter-gatherer sequenced to date (labeled AHG; directly dated to 13,642-13,073 BCE, excavated from the site of Pınarbaşı, Turkey), 5 early Neolithic Aceramic Anatolian farmers (labeled AAF; c. 8300-7800 cal BCE, one directly dated to 8269-8210 cal BCE[3], from the site of Boncuklu, Turkey), adding to previously published genomes from this site[7], and two Early Neolithic (PPNB) farmers from the southern Levant (One labeled KFH2, directly dated to c. 7,700-7,600 BCE; from the site of Kfar HaHoresh, Israel and the second labeled BAJ001, c. 7027-6685 BCE, from the site of Ba'ja, Jordan). This data comprises a genetic record stretching from the Epiplaeolithic into the Early Holocene, spanning the advent of agriculture in the region.

By analyzing this data, we find that the Anatolian hunter-gatherers are genetically distinct from other reported late Pleistocene populations and thus represent a previously undescribed population. We reveal that Neolithic Anatolian populations derive a large fraction of their ancestry from the Epipaleolithic Anatolian population, suggesting farming was adopted locally by the hunter-gatherers of central Anatolia. We also detect distinct genetic interactions between

the populations of central Anatolia and earlier farming centers to the east, during the late Pleistocene/early Holocene as well as with European hunter-gatherers to the west during the Late Pleistocene.

Results

We extracted DNA from the ancient human remains and prepared it for next-generation sequencing[8, 9] which resulted in human DNA yields lower than 2% (data table S1), comparable with low DNA preservation previously reported in the region[6, 7]. To generate genome wide data, despite the low DNA yields we performed in-solution DNA enrichment targeting 1.24 million genome-wide single nucleotide polymorphisms (SNPs) ('1240k capture')[10], which resulted in 129,406 to 917,473 covered SNPs per individual. We estimated low mitochondrial contamination levels for all eight individuals (1-6%; Materials and Methods and table S2) and could further test the males for nuclear contamination, resulting in low estimates (0.05-2.23%; table S2). For population genetic analyses, we merged genotype data of the new individuals with previously published datasets from 587 ancient individuals and 254 present-day populations (data table S2).

To estimate how the ancient individuals relate to the known west Eurasian genetic variation, we projected them onto the top two dimensions of present-day principal component analysis (PCA)[6] (Fig. 1B). Strikingly, the AHG individual is positioned near both AAF and later Anatolian Ceramic farmers[10] (labeled ACF; 7,000 - 6,000 cal BCE) with a slight leftward shift. These three prehistoric Anatolian populations (AHG, AAF and ACF), that represent a temporal transect spanning the transition into farming, are positioned between Mesolithic western European hunter-gatherers (WHG)[4, 10, 11] at the far left and Levantine Epipalaeolithic Natufians[6]

at the far right. The newly reported Neolithic farmers[6] (BAJ001 and KFH2) are positioned near the published ones[6] (Supplementary Text S2). In ADMIXTURE analysis, AHG, AAF and ACF are all modeled as a mixture of two components that are each maximized in Natufians and WHG, consistent with their positions in PCA (fig. S1).

5   Based on our observations in PCA and ADMIXTURE analysis we formally tested the ancestral compositions of the three Anatolian populations. We first characterized the ancestry of AHG. As expected from AHG's intermediate position on PCA between Epipaleolithic/Neolithic Levantines and WHG, Patterson's D- statistics[12] of the form *D (AHG, WHG; Natufian/Levant_N, Mbuti)* $\geq$ 4.8 SE (standard error) and *D (AHG, Natufian/Levant_N; WHG, Mbuti)* $\geq$ 9.0 SE (table

10   S3) indicates that AHG is distinct from both the WHG and Epipaleolithic/Neolithic Levantine populations and yet shares extra affinity with each when compared to the other. Accordingly, we find an adequate two-way admixture model using *qpAdm*[12] ($\chi^2 p = 0.158$), in which AHG derives around half of his ancestry from a Neolithic Levantine-related gene pool (48.0 $\pm$ 4.5 %; estimate $\pm$ 1 SE) and the rest from the WHG-related one (tables S4 and S5). These results support a late

15   Pleistocene presence of both ancestries in a mixed form in central Anatolia. Notably, the genetic connection with the Levant predates the advent of farming in this region by at least five millennia and potentially correlates with evidence of human interactions between central Anatolia and the Levant during the Epipalaeolithic[13].

In turn, AAF is slightly shifted upwards compared to AHG in the PCA, to the direction

20   where ancient and modern Caucasus and Iranian groups are located. Likewise, when compared to AHG by *D(AAF, AHG; test, Mbuti)*, the AAF early farmers show extra affinity with early Holocene populations from Iran or Caucasus and with present-day South Asians, who have also been genetically linked with Iranian/Caucasus ancestry[14, 15] (Fig. 2A, fig. S2 and data table S3).

A mixture of AHG and Neolithic Iranians provides a good fit to AAF in our *qpAdm* modeling ($\chi^2 p = 0.296$), in which they derive most of their ancestry (89.7 ± 3.9 %) from a population related to AHG (tables S4 and S6). This suggests a long-term genetic stability in central Anatolia over five millennia despite changes in climate and subsistence strategy. Given that our admixture model for AHG does not require the Neolithic Iranian ancestry, it presumably diffused into central Anatolia during the final stages of the Pleistocene or early Holocene, most likely via contact through eastern Anatolia. This provides evidence of interactions between eastern and central Anatolia in the Younger Dryas or the first millennium of the Holocene, currently poorly documented archaeologically.

In contrast, we find that the later ACF individuals share more alleles with the early Holocene Levantines than AAF do, as shown by positive *D(ACF, AAF; Natufian/Levant_N, Mbuti)* ≥ 3.84 SE (Fig. 2B, fig. S3 and data table S3). Ancient Iran/Caucasus populations and contemporary South Asians do not share more alleles with ACF ($|D| < 3.3$ SE). Likewise, qpAdm modeling suggests that the AAF gene pool still constitutes more than 3/4 of the ancestry of ACF 2,000 years later (78.7 ± 3.5 %; tables S4 and S7) with additional ancestry well modeled by the Neolithic Levantines ($\chi^2 p = 0.115$) but not by the Neolithic Iranians ($\chi^2 p = 0.076$; the model estimated infeasible negative mixture proportions) (tables S4 and S7). These results suggest gene flow from the Levant to Anatolia during the early Neolithic. In turn, Levantine early farmers (Levant_Neol) that are temporally intermediate between AAF and ACF could be modeled as a two-way mixture of Natufians and AHG or AAF (18.2 ± 6.4 % AHG or 21.3 ± 6.3 % AAF ancestry; tables S4 and S8 and data table S4), confirming previous reports of an Anatolian-like ancestry contributing to the Levantine Neolithic gene pool[6]. These two distinct

detected gene flows support a reciprocal genetic exchange between the Levant and Anatolia during the early stages of the transition to farming.

Anatolian hunter-gatherers experienced climatic changes during the last glaciation[16] and inhabited a region that connects Europe to the Near East. However, interactions between Anatolia and Southeastern Europe in the later Upper Palaeolithic/Epipalaeolithic are so far not well documented archaeologically. Interestingly, a previous genomic study showed that present-day Near-Easterners share more alleles with European hunter-gatherers younger than 14,000 BP ('Later European HG') than with earlier ones ('Earlier European HG')[17]. With ancient genomic data available, we could directly compare the Near-Eastern hunter-gatherers (AHG and Natufian) with the European ones. As is the case for present-day Near-Easterners, the Near-Eastern hunter-gatherers share more alleles with the Later European HG than with the Earlier European HG, shown by the significantly positive statistic *D(Later European HG, Earlier European HG; AHG/Natufian, Mbuti)* (Fig. 3A and data table S5). Among the Later European HG, recently reported Mesolithic hunter-gatherers from the Balkan peninsula, which geographically connects Anatolia and central Europe ('Iron Gates HG')[18], are genetically closer to AHG when compared to all the other European hunter-gatherers, as shown in the significantly positive statistic *D(Iron_Gates_HG, European hunter-gatherers; AHG, Mbuti/Altai)*. Iron Gates HG are followed by Epigravettian and Mesolithic individuals from Italy and France (Villabruna14 and Ranchot88 respectively[17]) as the next two European hunter-gatherers genetically closest to AHG (Fig. 3A and data table S5). Iron Gates HG have been suggested to be genetically intermediate between WHG and eastern European hunter-gatherers (EHG) with an additional unknown ancestral component[18]. We find that Iron Gates HG can be modeled as a three-way mixture of Near-Eastern hunter-gatherers (25.8 ± 5.0 % AHG or 11.1 ± 2.2 % Natufian), WHG (62.9 ± 7.4

7

% or 78.0 ± 4.6 % respectively) and EHG (11.3 ± 3.3 % or 10.9 ± 3 % respectively); (tables S4 and S9). The affinity detected by the above D-statistic can be explained by gene flow from Near-Eastern hunter-gatherers into the ancestors of Iron Gates or by a gene flow from a population ancestral to Iron Gates into the Near-Eastern hunter-gatherers as well as by a combination of both. To distinguish the direction of the gene flow, we examined the Basal Eurasian ancestry component (α), which is prevalent in the Near East [6] but undetectable in European hunter-gatherers[17]. Following a published approach[6], we estimated α to be 24.8 ± 5.5 % in AHG and 38.5 ± 5.0 % in Natufians (Fig. 3B, table S10), consistent with previous estimates for the latter[6]. Under the model of unidirectional gene flow from Anatolia to Europe, 6.4 % is expected for α of Iron Gates by calculating (% AHG in Iron Gates HG) × (α in AHG). However, Iron Gates can be modeled without any Basal Eurasian ancestry or with a non-significant proportion of 1.6 ± 2.8 % (Fig. 3B, table S10), suggesting that unidirectional gene flow from the Near East to Europe alone is insufficient to explain the extra affinity between the Iron Gates HG and the Near-Eastern hunter-gatherers. Thus, it is plausible to assume that prior to 15,000 years ago there was either a bidirectional gene flow between populations ancestral to Southeastern Europeans of the early Holocene and Anatolians of the late glacial or a dispersal of Southeastern Europeans into the Near East. Presumably, this Southeastern European ancestral population later spread into central Europe during the post-last-glacial maximum (LGM) period, resulting in the observed late Pleistocene genetic affinity between the Near East and Europe.

The uniparental marker analysis placed AHG within the mitochondrial sub-haplogroup K2b and within the Y-chromosome haplogroup C1a2, both rare in present-day Eurasians (Table 1 and data table S6). Mitochondrial Haplogroup K has so far not been found in Paleolithic hunter-gatherers[19]. However, Y-haplogroup C1a2 has been reported in some of the earliest

European hunter-gatherers[17, 20, 21]. The early farmers belong to common early Neolithic mitochondrial (N1a, U3 and K1a) and Y chromosome types (C and G2a), with the exception of the Levantine BAJ001 which represents the earliest reported individual carrying the mitochondrial N1b group (Table 1 and data table S6).

We examined alleles related to phenotypic traits in the ancient genomes (data table S7). Notably, three of the AAF carry the derived allele for rs12193832 in the HERC2 (hect domain and RLD2) gene that is primarily responsible for lighter eye color in Europeans[22]. The derived allele is observed as early as 14,000 -13,000 years ago in individuals from Italy and the Caucasus[17, 23] but had not yet been reported in early farmers or hunter-gatherers from the Near East.

Discussion

By analyzing genome-wide-data from pre- and early-Neolithic Anatolians and Levantines, we describe the demographic developments leading to the formation of the Anatolian early farmer population that later replaced most of the European hunter-gatherers and represents the largest ancestral component in present-day Europeans[4, 5].

We report a long-term persistence of the local Anatolian hunter-gatherer gene pool over seven millennia and throughout the transition from foraging to farming. This demographic pattern is similar to those previously observed in earlier farming centers of the Fertile Crescent[6] and differs from the pattern of the demic diffusion-based spread of farming into Europe[4, 5]. Our results provide a genetic support for archaeological evidence[3] suggesting that Anatolia was not merely a stepping stone in a movement of early farmers from the Fertile Crescent into Europe

but rather a place where local hunter-gatherers adopted ideas, plants and technology that led to agricultural subsistence.

Interestingly, while the local population structure remains highly stable, a pattern of genetic interactions with neighboring regions is observed from as early as the Late Pleistocene and into the early Holocene. External genetic contributions, associated with two distinct early farming populations of the Fertile Crescent, substituted about 10% and 20% of indigenous ancestry each. The earlier one is associated with Neolithic Iran/Caucasus and the later one with Neolithic Levant. Wide temporal gaps between available genomes currently limit our ability to distinguish the mode of transfer. Obtaining additional genomic data from these regions as well as from geographically intermediate populations of eastern Anatolia and the greater Mesopotamia region could help determine how these gene flows were introduced into central Anatolia: e.g., whether by a short-term massive migration or a low-level background gene flow in an "isolation by distance" manner.

To the west, a genetic link is observed between the Anatolian and Southeastern European Pleistocene hunter gatherers. Inspection of the shared genetic components between these two populations provides us with a model explaining the genetic affinity between late Pleistocene Europeans and present-day Near Eastern populations[17]. Further sampling in Anatolia and Southeastern Europe is needed to specify the spatiotemporal extent of the genetic interactions that we observe.

Materials and Methods

aDNA analysis

10

We extracted and prepared DNA for next generation sequencing in two different dedicated ancient DNA facilities (Liverpool and Jena).

Liverpool, UK. Sampling and extraction steps for the individuals from Pınarbaşı and Boncuklu were carried out in the aDNA labs at Liverpool John Moores University. The outer layer of the bone was removed using powdered aluminium oxide in a sandblasting instrument. Then, the bone was UV irradiated for 10 minutes on each side and ground into fine powder using a cryogenic grinder Freezer/Mill. DNA was extracted from 100 mg of bone powder following an established protocol[8]. The extracts were then shipped to Jena, Germany where downstream processing was performed.

Jena, Germany. All pre-amplification steps were performed in dedicated aDNA facilities of the Max Planck Institute for the Science of Human History (MPI-SHH). The inner ear part of the petrous bones of the individuals from Kfar HaHoresh and Ba'ja was sampled by drilling[24] and DNA was extracted from 76 - 109 mg of the bone powder. An extraction of ~100 mg pulverized bone from the Pınarbaşı individual ZBC was done in the Jena facility in addition to the Liverpool extraction (the sequenced data from the two extracts of individual ZBC were merged in downstream analysis after passing the quality control step). All extractions followed the same protocol as cited for Liverpool. A 20 µl aliquot from each extract was used to prepare an Illumina double stranded, double indexed DNA library following established protocols[9, 25]. Deaminated cytosines that result from DNA damage were partially removed using uracil-DNA glycosylase and endonuclease VIII but still retained in terminal read positions as a measure of aDNA authentication[26]. Indexed libraries were amplified using Herculase II Fusion DNA polymerase following the manufacturer's protocol and used for two previously published downstream in-solution enrichments: a protocol targeting 1,237,207 genome-wide SNPs ('1240k

capture'[10]) and one targeting the entire human mitochondrial genome[27]. Both the initial shotgun and target-enriched libraries were single-end sequenced on an Illumina Hiseq 4000 platform (1 x 75 + 8 + 8 cycles). Sequenced reads were demultiplexed allowing one mismatch in each index and further processed using EAGER (v 1.92.54)[28]. First, adapter sequences were clipped and reads shorter than 30 bp were discarded using AdapterRemoval (v 2.2.0)[29]. Adapter-clipped reads were subsequently mapped with the BWA aln/samse programs (v 0.7.12)[30] against the UCSC genome browser's human genome reference hg19 with a lenient stringency parameter ("-n 0.01"). We retained reads with Phred-scaled mapping quality scores ≥ 20 and ≥ 30 for the whole genome and the mitochondrial genome, respectively. Duplicate reads were subsequently removed using DeDup v0.12.2[28]. Pseudo-diploid genotypes were generated for each individual using pileupCaller which randomly draws a high quality base (Phred-scaled base quality score ≥ 30) mapping to each targeted SNP position (https://github.com/stschiff/sequenceTools). To prevent false SNP calls due to retained DNA damage, two terminal positions in each read were clipped prior to genotyping. The genotyping produced between 129,406 to 917,473 covered targeted SNPs and a mean coverage ranging between 0.16 and 2.9 fold per individual (Table 1).

Dataset

We merged the newly reported ancient data and data reported by *Mathieson et al. 2017* [18] with a dataset that has been described elsewhere[6]. This dataset includes 587 published ancient genomes[6, 7, 10, 11, 14, 17, 20, 23, 31-34] and genomes from 2,706 individuals, representing world-wide present-day populations[6, 35] that were genotyped on the Affymetrix Axiom™ Genome-Wide Human Origins 1 array[4] ('HO dataset') with a total of 597,573 SNP sites in the merged dataset. To minimize bias from differences in analysis pipelines, we re-processed the raw read data

deposited for previously published Neolithic Anatolian genomes[7] (labeled Tepecik_pub and Boncuklu_pub) in the same way as described for the newly reported individuals.

aDNA authentication and quality control

We estimated authenticity of the ancient data using multiple measures. First, blank controls were included and analyzed for extractions as well as library preparations (Data table S8). Second, we assessed levels of DNA damage in the mapped reads using mapDamage (v 2.0)[36]. Third, we estimated human DNA contamination on the mitochondrial DNA using schmutzi[37]. Last, we estimated nuclear contamination in males with ANGSD (v 0.910)[38], which utilizes haploid X chromosome markers in males by comparing mismatch rates of polymorphic sites and adjacent ones (that are likely to be monomorphic). The genetic sex of the reported individuals was determined by comparing the genomic coverage of X and Y chromosomes normalized by the autosomal average coverage. To avoid bias caused by grouping closely related individuals into a population, we calculated the pairwise mismatch rates of the Boncuklu individuals following a previously reported method[39] (Data table S9).

Five of the twelve individuals reported here were excluded from the population genetic analysis: two due to a high genomic contamination level (> 5 %) and three due to low amount of analyzable data (< 10,000 SNPs covered); (Data table S1).

Principal component analysis (PCA)

We used the smartpca software from the EIGENSOFT package (v 6.0.1)[40] with the lsqproject option to construct the principal components of 67 present-day west Eurasian groups and project the ancient individuals on the first two components (fig. S4).

ADMIXTURE analysis

We used ADMIXTURE (v 1.3.0)[41] to perform an unsupervised clustering of 3293 ancient and present-day individuals in the HO merged dataset, allowing the number of clusters (k) to range between 2 and 20. Pruning for linkage disequilibrium (LD) was done by randomly removing one SNP from each pair with genotype $r^2 \geq 0.2$, using PLINK (v 1.90)[42, 43]; (--indep-pairwise 200 25 0.2). The analysis was replicated five times for each k value with random seeds and the highest likelihood replicate is reported (fig. S1 and S5) Five-fold cross-validation errors were calculated for each run.

D-statistics

To estimate allele frequency correlations between populations, *D*- statistics were computed using the *qpDstat* program (v 701) of the ADMIXTOOLS package[44] (v 4.1) with default parameters. In order to determine whether a test population is symmetrically related to populations X and Y, the *D*- statistic *D* (*X, Y; Test, Outgroup*) was used. In particular, when comparing the affinity of different European hunter-gatherers to Near-Eastern ones in the *D*-statistic of the form *D* (*European HG1, European HG2; Near Eastern HG, Outgroup*), both the central African *Mbuti* and the Altai Neanderthal (*Altai_published.DG*) were used to check if the differing level of Neanderthal ancestry in these hunter-gatherers affects the results. Otherwise, Mbuti was used as the single outgroup. The above statistics are reported when more than 30,000 SNP positions were overlapping between the four analyzed populations.

Modeling ancestry proportions

We used the qpWave (v400) and qpAdm (v 632) programs of ADMIXTOOLS[6, 12] to test

and model admixture proportions in a studied population from potential source populations

(reference populations). As the explicit phylogeny is unknown, a diverse set of outgroup

populations (Supplementary Information sections 1.2-1.4) was used to distinguish the ancestry of

the reference populations.

For estimating admixture proportions in the tested populations, we used a basic set of seven

outgroups including present-day populations (Han, Onge, Mbuti, Mala, Mixe) that represent a

global genetic variation and published ancient populations such as Natufian[6], that represents a

Levantine gene pool outside of modern genetic variation and the European Upper Palaeolithic

individual Kostenki14[20]. As a prerequisite for the admixture modeling of the target population

we tested whether the corresponding set of reference populations can be distinguished by the

chosen outgroups using qpWave[6] (Supplementary text S3). In some cases, when a reference

population did not significantly contribute to the target in the attempted admixture models, it was

removed from the reference set and added to the basic outgroup set in order to increase the

power to distinguish the references. In cases where "Natufian" was used as a reference

population, we instead used the present-day Near-Eastern population "BedouinB" as an

outgroup.

For estimations of Basal Eurasian ancestry, we followed a previously described qpAdm

approach[6] that does not require a proper proxy for the Basal Eurasian ancestry, which is currently

not available in unadmixed form. This framework relies on the basal phylogenetic position of

both Basal Eurasian and an African reference (the ancient Ethiopian *Mota* genome[45]) relative to

other non-Africans. Thus, by using a set of outgroups that includes eastern non-African

populations (Han; Onge; Papuan) and Upper Palaeolithic Eurasian genomes (*Ust'-Ishim*[46],

Kostenki14, *MA-1*[47]) but neither west Eurasians with detectable basal Eurasian ancestry nor

Africans, the mixture proportion computed for *Mota* (α) can be used indirectly to estimate the

Basal Eurasian mixture proportion of west Eurasian populations.

5   ## Mitochondrial DNA analysis

The endogenous mitochondrial consensus sequences were inferred from the output of

schmutzi[37], using its log2fasta program and a quality cutoff of 10. Mitochondrial haplotypes

were established by aligning these consensuses to rCRS[48] using the online tool haplosearch[49].

The coverage of each of the reported SNPs was confirmed by visually inspecting the bam pileup

10   in Geneious (v11.0.4)[50]. The resulting consensus sequences were then analyzed with HaploFind[51]

and Haplogrep[52] to assign mitochondrial haplogroups and double-checked with the rCRS

oriented version of Phylotree[53].

Y-chromosome analysis

15   To assign Y-chromosome haplogroups we used yHaplo[54]. Each male individual was

genotyped at 13,508 ISOGG consortium SNP positions (strand-ambiguous SNPs were excluded)

by randomly drawing a single base mapping to the SNP position, using the same quality filters as

for the HO dataset. In addition to the yHaplo automated haplogroup designations, we manually

verified the presence of derived alleles supporting the haplogroup assignment.

20

Phenotypic traits analyses

We tested for the presence of alleles related to biological traits that could be of interest in

the geographical and temporal context of the reported ancient populations, including lactase

persistence[55, 56], Malaria resistance[57, 58], Glucose-6-phosphate dehydrogenase (G6PD)

deficiency[59, 60] and skin pigmentation[23, 61, 62]. The allele distribution for the SNP positions listed

in Data table S7 was tabulated for each individual using Samtools mpileup (v 1.3).

Carbon dating

The phalanx bone from individual ZBC (Pinarbaşı) and the petrous bone from individual

KFH2 (Kfar HaHoresh) were each sampled and directly radiocarbon dated at the CEZ

Archaeometry gGmbH, Mannheim, Germany (table S1). Collagen was extracted from the bone

samples, purified by ultrafiltration (fraction >30kD), freeze-dried and combusted to CO_2 in an

Elemental Analyzer (EA). CO_2 was converted catalytically to graphite. The dating was

performed using the MICADAS-AMS of the Klaus-Tschira-Archäometrie-Zentrum. The

resulting ^{14}C ages were normalized to d13C=-25% [63] and calibrated using the dataset

INTCAL13[64] and the software SwissCal 1.0[65].

Data availability

Alignment data (BAM) is deposited in the European Nucleotide Archive (ENA) under

the accession numbers (Study PRJEB24794).

References

1. Bellwood, P. First farmers. *The Origins of Agricultural Societies. Wiley-Blackwell* (2005).

2. Baird, D. The Late Epipaleolithic, Neolithic, and Chalcolithic of the Anatolian Plateau, 13,000–4000 BC. *A companion to the archaeology of the Ancient Near East*, 431-465 (2012).

3. Baird, D. *et al*. Agricultural origins on the Anatolian plateau. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3077-E3086 (2018).

4. Lazaridis, I. *et al*. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413 (2014).

5. Bramanti, B. *et al*. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137-140 (2009).

6. Lazaridis, I. *et al*. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424 (2016).

7. Kılınç, G. M. *et al*. The demographic development of the first farmers in Anatolia. *Current Biology* **26**, 2659-2666 (2016).

8. Dabney, J. *et al*. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758-15763 (2013).

9. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**, pdb. prot5448 (2010).

10. Mathieson, I. *et al*. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).

11. Jones, E. R. *et al*. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).

12. Haak, W. *et al*. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).

13. Baird, D. *et al*. Juniper smoke, skulls and wolves' tails. The Epipalaeolithic of the Anatolian plateau in its South-west Asian context; insights from Pınarbaşı. *Levant* **45**, 175-209 (2013).

14. Broushaki, F. *et al*. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499-503 (2016).

15. Narasimhan, V. M. *et al*. The Genomic Formation of South and Central Asia. *bioRxiv*, 292581 (2018).

16. Gamble, C., Davies, W., Pettitt, P. & Richards, M. Climate change and evolving human diversity in Europe during the last glacial. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**, 243-53; discussion 253-4 (2004).

17. Fu, Q. *et al*. The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).

18. Mathieson, I. *et al*. The Genomic History of Southeastern Europe. *Nature* (2017).

19. Posth, C. *et al*. Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. *Current Biology* **26**, 827-833 (2016).

20. Seguin-Orlando, A. *et al*. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118 (2014).

21. Sikora, M. *et al*. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* (2017).

22. Sturm, R. A. *et al*. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *The American Journal of Human Genetics* **82**, 424-431 (2008).

23. Olalde, I. *et al*. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228 (2014).

24. Pinhasi, R. *et al*. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PloS one* **10**, e0129102 (2015).

25. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3-e3 (2011).

26. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**, 20130624 (2015).

27. Fu, Q. *et al*. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219 (2015).

28. Peltzer, A. *et al*. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).

29. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes* **9**, 88 (2016).

30. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

31. Prüfer, K. *et al*. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49 (2014).

32. Hofmanova, Z. *et al*. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886-6891 (2016).

33. Gallego-Llorente, M. *et al*. The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Scientific reports* **6**, 31326 (2016).

34. Omrak, A. *et al*. Genomic evidence establishes Anatolia as the source of the European Neolithic gene pool. *Current Biology* **26**, 270-275 (2016).

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

36. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-1684 (2013).

37. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).

38. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

39. Kennett, D. J. *et al*. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nature Communications* **8**, 14115 (2017).

40. Patterson, N., Price, A. L. & Reich, D. Population structure and eigen analysis. *PLoS genetics* **2**, e190 (2006).

41. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655-1664 (2009).

42. Purcell, S. *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).

43. Chang, C. C. *et al*. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

44. Patterson, N. *et al*. Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).

45. Gallego Llorente, M. *et al*. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820-822 (2015).

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

47. Raghavan, M. *et al*. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91 (2014).

48. Andrews, R. M. *et al*. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
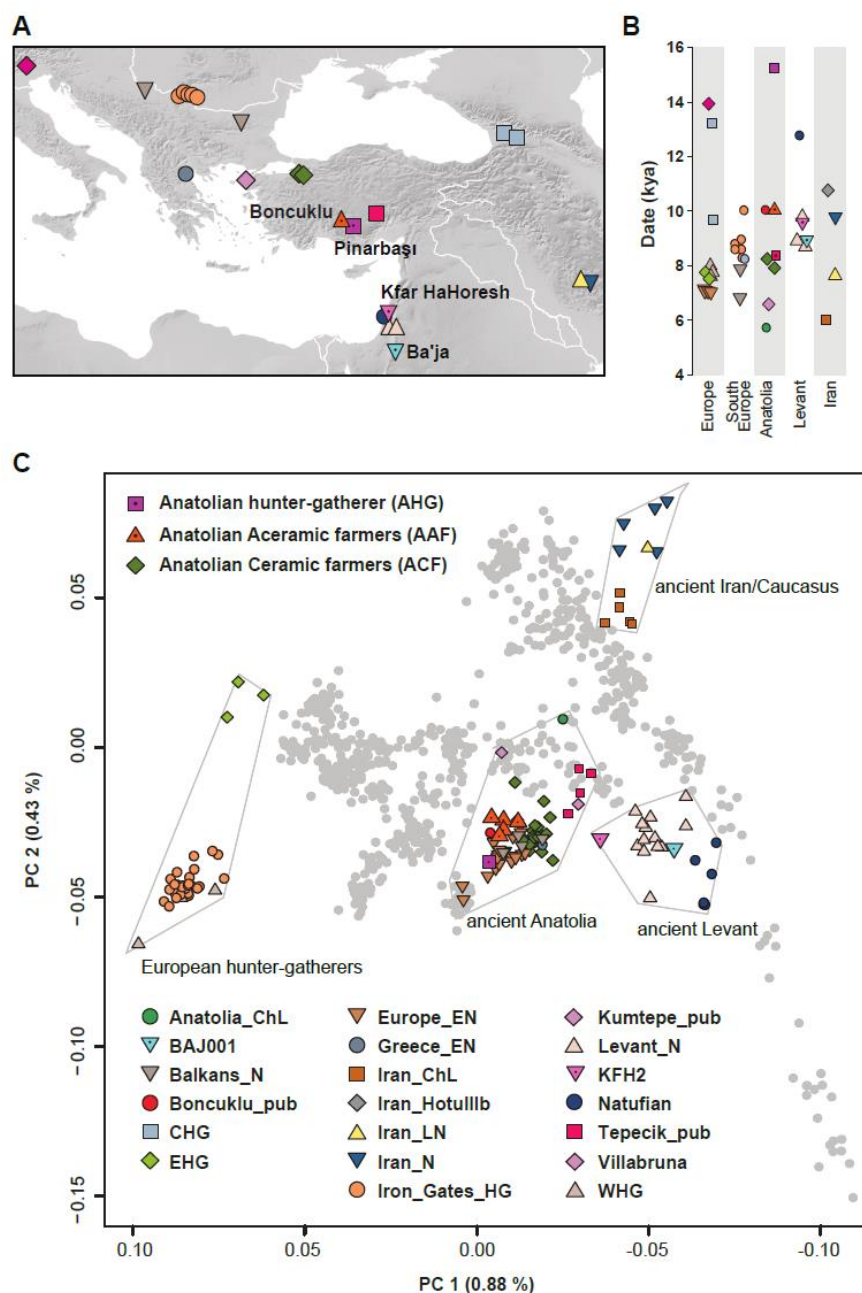
49. Fregel, R. & Delgado, S. HaploSearch: a tool for haplotype-sequence two-way transformation. *Mitochondrion* **11**, 366-367 (2011).

50. Kearse, M. *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).

51. Vianello, D. *et al*. HAPLOFIND: a New method for high-throughput mtDNA Haplogroup assignment. *Hum. Mutat.* **34**, 1189-1194 (2013).

52. Kloss-Brandstätter, A. *et al*. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25-32 (2011).

53. Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30** (2009).

54. Poznik, G. D. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv*, 088716 (2016).

55. Ingram, C. J. *et al*. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum. Genet.* **120**, 779-788 (2007).

56. Tishkoff, S. A. *et al*. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31-40 (2007).

57. Network, M. G. E. *et al*. Reappraisal of known malaria resistance loci in a large multicenter study. *Nat. Genet.* **46**, 1197 (2014).

58. Timmann, C. *et al*. Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**, 443-446 (2012).

59. Sepúlveda, N. *et al*. Malaria Host Candidate Genes Validated by Association With Current, Recent, and Historical Measures of Transmission Intensity. *J. Infect. Dis.* **216**, 45-54 (2017).

60. Maiga, B. *et al*. Glucose-6-phosphate dehydrogenase polymorphisms and susceptibility to mild malaria in Dogon and Fulani, Mali. *Malaria journal* **13**, 270 (2014).

61. Norton, H. L. *et al*. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710-722 (2006).

62. Wilde, S. *et al*. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4832-4837 (2014).

63. Stuiver, M. & Polach, H. A. Discussion reporting of 14 C data. *Radiocarbon* **19**, 355-363 (1977).

64. Reimer, P. J. *et al*. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869-1887 (2013).

65. Christl, M. *et al*. The ETH Zurich AMS facilities: Performance parameters and reference materials. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **294**, 29-38 (2013).

Fig. 1. Location, age and PCA of analyzed individuals. (A) Locations of newly reported and selected published genomes. Archaeological sites from which new data is reported are annotated. Symbols for the analyzed groups are annotated in (C). **(B)** Average ages of ancient groups. **(C)** Ancient genomes (marked with color-filled symbols) projected onto the principal components computed from present-day west Eurasians (grey circles) (fig. S4). The geographic location of each ancient group is marked in (A). Ancient individuals newly reported in this study are additionally marked with a black dot inside the symbol.
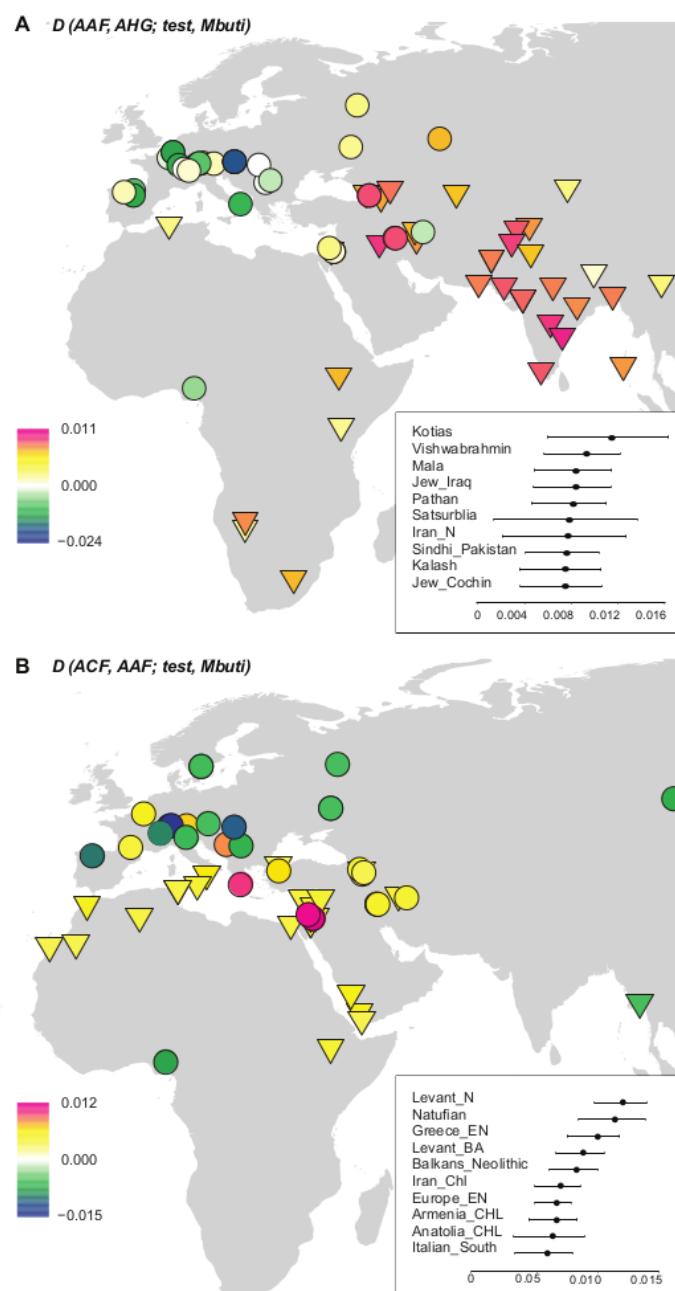
Fig. 2. Differences between the ancient Anatolian populations in their genetic affinity to worldwide present-day and ancient populations. We plot the highest and lowest 40 values of *D(population 1, population 2; test, Mbuti)* on the map. Circles mark ancient populations and triangles present-day ones. "test" share more alleles with *population 1* when values are positive and with *population 2* when negative. The statistics and SEs are found in figs S2-S3 and data table S3. (**A**) Early Holocene Iranian and Caucasus populations, as well as present-day South Asians, share more alleles with AAF than AHG, measured by positive *D(AAF, AHG; test, mbuti)*. The top 10 values with ±1 SE are shown in the upper box. (**B**) Ancient Levantine populations share more alleles with ACF than AAF, measured by positive *D(ACF, AAG; test, Mbuti)*. The top 10 values with ±1 SE are shown in the lower box.
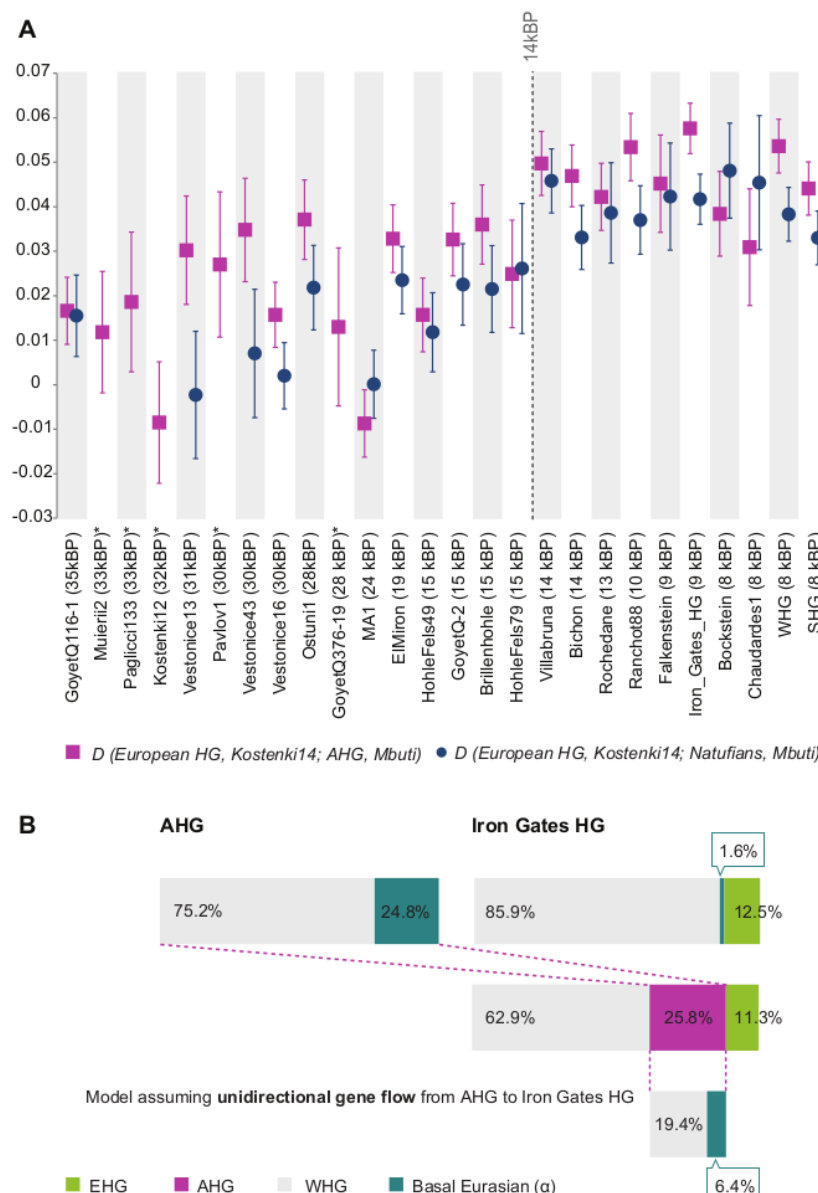
Fig. 3. Genetic links between Near-Eastern and European hunter-gatherers. (A) Genetic affinity between Near-Eastern and European hunter-gatherers increases after 14,000 years ago as measured by the statistic *D(European HG, Kostenki14; Natufian/AHG, Mbuti)* Vertical lines mark ±1 SE. *Kostenki14* serves here as a baseline for the earlier European hunter-gatherers. Statistics including all analyzed European hunter-gatherers are listed in data table S5. Individuals marked with an asterisk did not reach the analysis threshold of over 30,000 SNPs overlapping with *Natufian/AHG*. **(B)** Basal Eurasian ancestry proportions (α) as a marker for Near-Eastern gene flow. Mixture proportions inferred by qpAdm for AHG and the Iron Gates HG are schematically represented [6]. The lower schematic shows the expected α in Iron Gates HG under assumption of unidirectional gene flow, inferred from α in the AHG source population. The observed α for Iron Gates HG is considerably smaller than expected thus, the unidirectional gene flow from the Near East to Europe is not sufficient to explain the above affinity.

Table 1. An overview of ancient genomes reported in this study. For each individual the analysis group is given (AHG = Anatolian hunter-gatherer; AAF = Anatolian Aceramic early farmer; Levant_Neol = Levantine early farmer). When ^{14}C dating results are available the date is given in cal BCE in 2-sigma range, otherwise a date based on the archaeological context is provided (detailed dating information is provided in Supplementary text S1 and table S1). The proportion of human DNA and the mean coverage on 1240K target sites in the '1240K' enriched libraries are given. Uniparental haplogroups (mt = mitochondrial; Ychr = Y chromosome) are listed. Detailed information on the uniparental analysis can be found in Supplementary text S1 and data table S6.

ID	Library name	Analysis group	Estimated date	Site	Sampled tissue	Total sequenced reads (x10⁶)	Human DNA (%)	Mean Coverage (fold)	Genetic sex	mt	Ychr
ZBC	IPB001.B/C0101	AHG	13,642-13,073 cal BCE	Pınarbaşı	Intermediate phalanx	126.7	33	2.9	Male	K2b	C1a2
ZHAG	BON004.A0101	AAF	8300-7800 BCE	Boncuklu	Petrous	92.0	38	1.48	Female	N1a1a1	
ZMOJ	BON014.A0101	AAF	8300-7800 BCE	Boncuklu	Third molar	77.9	27	0.8	Male	K1a	C
ZKO	BON001.A0101	AAF	8300-7800 BCE	Boncuklu	Petrous	84.8	31	0.9	Male	U3	G2a2b2b
ZHJ	BON024.A0101	AAF	8300-7800 BCE	Boncuklu	Third molar	87.7	38	0.76	Female	U3	
ZHAJ	BON034.A0101	AAF	8269-8210 cal BCE	Boncuklu	Petrous	75.4	30	0.69	Female	U3	
KFH2	KFH002.A0101	Levant_Neol	7712-7589 cal BCE	Kfar HaHoresh	Petrous	342.0	8	0.16	Female	N1a1b	
BAJ001	BAJ001.A0101	Levant_Neol	7027-6685 cal BCE	Ba'ja	Petrous	17.3	45	0.75	Female	N1b1a	