Title page

Classification: Biological Sciences (Genetics)

Early farmers from across Europe directly descended from Neolithic Aegeans

Authors: Zuzana Hofmanová[a,1], Susanne Kreutzer[a,1], Garrett Hellenthal[b], Christian Sell[a], Yoan Diekmann[b], David Díez-del-Molino[b], Lucy van Dorp[b], Saioa López[b], Athanasios Kousathanas[c,d], Vivian Link[c,d], Karola Kirsanow[a], Lara M. Cassidy[e], Rui Martiniano[e], Melanie Strobel[a], Amelie Scheu[a,e], Kostas Kotsakis[f], Paul Halstead[g], Sevi Triantaphyllou[f], Nina Kyparissi-Apostolika[h], Dushanka-Christina Urem-Kotsou[i], Christina Ziota[j], Fotini Adaktylou[k], Shyamalika Gopalan[l], Dean M. Bobo[l], Laura Winkelbach[a], Jens Blöcher[a], Martina Unterländer[a], Christoph Leuenberger[m], Çiler Çilingiroğlu[n], Barbara Horejs[o], Fokke Gerritsen[p], Stephen Shennan[q], Daniel G. Bradley[e], Mathias Currat[r], Krishna R. Veeramah[l], Daniel Wegmann[c,d], Mark G. Thomas[b], Christina Papageorgopoulou[s,2] and Joachim Burger[a,2]

[1]These authors contributed equally to this work
[2] Corresponding authors (cpapage@he.duth.gr; jburger@uni-mainz.de)
JB: Institute of Anthropology, Johannes Gutenberg-University Mainz, Anselm-Franz-von-Bentzel-Weg 7, D-55099 Mainz, Germany, +49 6131 392 0981
CP: Laboratory of Anthropology, Department of History and Ethnology, Demokritus University of Thrace, P. Tsaldari 1, 69100 Komotini, Greece, +30 25310 39996

Affiliations
[a] Palaeogenetics Group, Johannes Gutenberg University Mainz, 55099 Mainz, Germany.
[b] Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK.
[c] Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland.
[d] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.
[e] Molecular Population Genetics, Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland.
[f]Aristotle University of Thessaloniki, Faculty of Philosophy, School of History and Archaeology, 54124 Thessaloniki, Greece.
[g] Department of Archaeology, University of Sheffield, Northgate House, West Street, Sheffield S1 4ET, UK.
[h] Honorary Ephor of Antiquities, Hellenic Ministry of Culture & Sports, 106 82 Athens, Greece.
[i] Department of History and Ethnology, Democritus University of Thrace, 69100 Komotini, Greece.
[j]Ephorate of Antiquities of Florina, Hellenic Ministry of Culture & Sports, 106 82 Athens, Greece.
[k] Ephorate of Antiquities of Chalkidiki and Mount Athos, Hellenic Ministry of Culture & Sports, 106 82 Athens, Greece.
[l] Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York, 11794-5245, USA.
[m] Department of Mathematics, University of Fribourg, 1700 Fribourg, Switzerland.
[n] Ege University, Faculty of Letters, Department of Protohistory and Near Eastern Archeology, 35100 Bornova, Izmir, Turkey.
[o] Institute of Oriental and European Archaeology, Austrian Academy of Sciences, 1010 Vienna, Austria.
[p] Netherlands Institute in Turkey, P.K. 132, Beyoğlu 34433, İstanbul, Turkey.
[q] Institute of Archaeology, University College London, London WC1H 0PY, UK.

[r] Department of Genetics & Evolution-Anthropology Unit, University of Geneva, 1211 Geneva, Switzerland.
[s] Laboratory of Anthropology, Department of History and Ethnology, Democritus University of Thrace, 69100 Komotini, Greece.

Keywords: paleogenomics, Neolithic, Mesolithic, Greece, Anatolia

Abstract

Farming and sedentism first appear in southwest Asia during the early Holocene and later spread to neighboring regions, including Europe, along multiple dispersal routes. Conspicuous uncertainties remain about the relative roles of migration, cultural diffusion and admixture with local foragers in the early Neolithisation of Europe. Here we present paleogenomic data for five Neolithic individuals from northern Greece and northwestern Turkey spanning the time and region of the earliest spread of farming into Europe. We employ a novel approach to re-calibrate raw reads and call genotypes from ancient DNA and observe striking genetic similarity both among Aegean early farmers and with those from across Europe. Our study demonstrates a direct genetic link between Mediterranean and Central European early farmers and those of Greece and Anatolia, extending the European Neolithic migratory chain all the way back to southwestern Asia.

Significance statement

One of the most enduring and widely debated questions in prehistoric archaeology concerns the origins of Europe's earliest farmers; were they the descendants of local hunter-gatherers or did they migrate in from southwestern Asia, where farming began? We recover genome-wide DNA sequences from early farmers on both the European and Asian sides of the Aegean to reveal an unbroken chain of ancestry leading from central and southwestern Europe back to Greece and northwestern Anatolia. Our study provides the coup de grâce to the notion that farming spread into and across Europe via the dissemination of ideas but without, or with only a limited, migration of people.

\body

Archaeological Background

It is well established that farming was introduced to Europe from Anatolia, but the extent to which its spread was mediated by demic expansion of Anatolian farmers, or by the transmission of farming technologies and lifeways to indigenous hunter-gatherers without a major concomitant migration of people, has been the subject of considerable debate. Paleogenetic studies [1-4] of late hunter-gatherers and early farmers indicate a dominant role for migration in the transition to farming in central and northern Europe, with evidence of only limited hunter-gatherer admixture into early Neolithic populations, but increasing towards the late Neolithic. However, the exact origin of central and western Europe's early farmers, in the Balkans, Greece or Anatolia remains an open question.

Recent radiocarbon dating indicates that by 6,600 to 6,500 cal BCE, sedentary farming communities were established in northwest Anatolia, at sites such as Barcın, Menteşe, and Aktopraklık C, and in coastal west Anatolia, at sites like Çukuriçi and Ulucak (Fig. 1), but did not expand north or west of the Aegean for another several hundred years [5]. All these sites show material culture affinities with the central and southwest Anatolian Neolithic [6].

Early Greek Neolithic sites, such as Franchthi Cave in the Peloponnese, Knossos in Crete and Mauropigi, Paliambela and Revenia in northern Greece date to a similar period [7-9]. The distribution of obsidian from the Cycladic islands, as well as similarities in material culture, suggest extensive interactions since the Mesolithic and a coeval Neolithic on both sides of the Aegean [8]. While it has been argued that *in situ* Aegean Mesolithic hunter-gatherers played a major role in the 'Neolithisation' of Greece [7], the presence of domesticated forms of plants and animals indicates non-local Neolithic dispersals into the area.

We present five ancient genomes from both the European and Asian sides of the northern Aegean (Fig. 1); despite their origin from non-temperate regions, three of them were sequenced to relatively high coverage (~2-7x) enabling diploid calls using a novel SNP calling method that accurately accounts for post-mortem damage (SI Appendix, section 5). Two of the higher coverage genomes are from Barcın, south of the Marmara Sea in Turkey, one of the earliest Neolithic sites in northwestern Anatolia (Bar8 and Bar31). On the European side of the Aegean, one genome is from the early Neolithic site of Revenia (Rev5), and the remaining two are from the late and final Neolithic sites of Paliambela (Pal7) and Kleitos (Klei10), dating to approximately 2,000 years later (Table 1). Estimates of mitochondrial contamination were low (0.006 – 1.772% for shotgun data; Table 1, SI Appendix, section 4). We found unprecedented deamination rates of up to 56% in petrous bone samples, indicating a prehistoric origin for our sequence data from non-temperate environments (SI Appendix, Table S5).

Uniparental genetic systems

The mtDNA haplogroups of all five Neolithic individuals are typical of those found in central European Neolithic farmers and modern Europeans, but not of European Mesolithic hunter-gatherers [1]. Likewise, the Y-chromosomes of the two male individuals belong to haplogroup G2a2, which has been observed in European Neolithic farmers [3, 10], Ötzi, the Tyrolean Iceman [11], and modern western and southwestern Eurasian populations, but not in any pre-Neolithic European hunter-gatherers [12]. The mitochondrial haplogroups of two additional less well-preserved Greek Mesolithic individuals (Theo1, Theo5; Appendix Table S6) belong to lineages observed in Neolithic farmers from across Europe, consistent with Aegean Neolithic populations, unlike central European Neolithic populations, being the direct descendants of the preceding Mesolithic peoples that inhabited broadly the same region. However, we caution against over-interpretation of the Aegean Mesolithic mtDNA data; additional genome-level data will be required to identify the Mesolithic source population(s) of the early Aegean farmers.

Functional variation

Sequences in and around genes underlying phenotypes hypothesized to have undergone positive selection in Europeans indicate that the Neolithic Aegeans were unlikely to have been lactase persistent, and carried the derived *SLC24A5* rs1426654 and *SLC45A2* rs16891982 alleles associated with reduced skin pigmentation. Since our Aegean samples predate the period when the rs4988235 T-allele associated with lactase persistence in Eurasia reached an appreciable frequency in Europe, around 4kya [12- 14], and because this allele remains at relatively low frequencies (<0.15) in modern Greek, Turkish, and Sardinian populations [15], this observation is unsurprising. However, despite their relatively low-latitude, four of the Aegean individuals are homozygous for the derived rs1426654 T-allele in the *SLC24A5* gene, and four carry at least one copy of the derived rs16891982 G-allele in the *SLC45A2* gene. This suggests that these reduced pigmentation-associated alleles were at appreciable frequency in Neolithic Aegeans, and that skin depigmentation was not solely a high-latitude phenomenon (SI Appendix, section 12). The derived rs12913832 G-allele in the *HERC2* domain of the *OCA2* gene was heterozygous in one individual (Klei10), but all other Aegeans for whom the allelic state at this locus could be determined were homozygous for the ancestral allele, indicating a lack of iris depigmentation in these individuals.

Examination of several SNPs in the *TCF7L2* gene region indicate that the two Neolithic Anatolian individuals, Bar8 and Bar31, are likely to have carried at least one copy of a haplotype conferring reduced susceptibility to type 2 diabetes (T2D); the Klei10 and Rev5 individuals also carry a tag allele associated with this haplotype. Consistent with these observations, it has been previously estimated that this T2D

protective haplotype, which shows evidence for selection in Europeans, East Asians, and West Africans, originated approximately 11,900 years ago in Europe [16].

A number of loci associated with inflammatory disease displayed the derived alleles, including rs2188962 C>T in the *SLC22A5/IRF1* region, associated with Crohn's disease; rs3184504 C>T in the *SH2B3/ATXN2* region, associated with rheumatoid arthritis, celiac disease, and type 1 diabetes; and rs6822844 G>T in the *IL2/IL21* region, associated with rheumatoid arthritis, celiac disease, and ulcerative colitis. Interestingly, we observe derived states for 6 of 8 loci in a protein-protein interaction network inferred to have undergone concerted positive selection 2.6-1.2 kya in Europeans [17], suggesting that any recent selection on these loci acted on standing variation present at already appreciable frequency (see SI Appendix, section 12).

PCA, f-statistics, and mixture modeling

The first two dimensions of variation from principal components analysis (PCA) reveal a tight clustering of all five Aegean Neolithic genomes with Early Neolithic genomes from central and southern Europe [2, 3, 13] (Fig. 2). This cluster remains well-defined when the third dimension of variation is also considered (3D-figure S4, SI Appendix). Two recently published pre-Neolithic genomes from the Caucasus [18] appear to be highly differentiated from the genomes presented here and most likely represent a forager population distinct from the Epipaleolithic/Mesolithic precursors of the early Aegean farmers.

To examine this clustering of Early Neolithic farmers in more detail, we calculated outgroup *f*3 statistics [19] of the form *f*3(‡Khomani; TEST, Greek/Anatolian), where TEST is one of the available ancient European genomes (SI Appendix, section 7 and Figures S8-S10, Dataset S2); ‡Khomani San were selected as an outgroup as they are considered to be the most diverged extant human population. Consistent with their PCA clustering, the northern Aegean genomes share high levels of genetic drift amongst each other, and with all other previously characterized European Neolithic genomes, including early Neolithic from northern Spain, Hungary and central Europe. Given the archaeological context of the different samples, the most parsimonious explanation for this shared drift is migration of early European farmers from the northern Aegean into and across Europe [12].

To better characterize this inferred migration, we modeled ancient and modern genomes as mixtures of DNA from other ancient and/or modern genomes; a flexible approach that characterizes the amount of ancestry sharing among multiple groups simultaneously [20, 21] (Fig. 3, SI Appendix, section 10). Briefly we first represented each ancient or modern 'target' group by the (weighted) number of alleles they share in common with individuals from a fixed set of sampled populations (i.e. the "unlinked" approach described in [20]), which we refer to as the 'allele matching profile' for that target group. To cope with

issues such as unequal sample sizes, we then used a linear model [22] to fit the allele matching profile of the target group as a mixture of that of other sampled groups. Sampled groups that contribute most to this mixture indicate a high degree of shared ancestry with the target group relative to other groups. Under this framework the oldest Anatolian genome (Bar31) was inferred to contribute the highest amount of genetic ancestry (39-53%) to the Early Neolithic genomes from Hungary [13] and Germany [2] compared to any other ancient or modern samples, with the next highest contributors being other ancient Aegean genomes (Klei10, Pal7, Bar8) (Fig. S23, Fig. S24, Fig. S29). This pattern is not symmetric in that we infer smaller contributions from the German (<26%) and Hungarian (<43%) Neolithic genomes to any of the Anatolian or Greek ancient genomes. Furthermore, in this analysis modern samples from Europe and surrounding regions are inferred to be relatively more genetically related to the Aegean Neolithic genomes than to the Neolithic genomes from Germany and Hungary (Fig. 3, SI Appendix, section 10). These patterns are indicative of founder effects [23] in the German and possibly Hungarian Neolithic samples, from a source that appears to be most genetically similar to the Aegean Neolithic samples (specifically Bar31) and which distinguishes them from the ancestors of modern groups. Consistent with this, we found fewer short runs of homozygosity (ROH between 1-2Mb) in our high coverage Anatolian sample (Bar8) than in Early Neolithic genomes from Germany and Hungary (SI Appendix, section 11 and Fig. S31). However, it is not possible to infer a direction for dispersal within the Aegean with statistical confidence since both the Greek and Anatolian genomes copy from each other to a similar extent. We therefore see the origins of European farmers equally well represented by Early Neolithic Greek and northwestern Anatolian genomes.

Ongoing gene flow into and across the Aegean is also indicated in the genome of a Chalcolithic individual from Kumtepe (Kum6; [24]); a site geographically close to Barcın but dating to approx. 1,600 years later. While archaeological evidence indicates a cultural break in many Aegean and West Anatolian settlements around 5,700/5,600 cal BCE (i.e. spanning this 1,600 year period [25]), Kum6 shows affinities to the Barcın genomes in "outgroup" *f3*-statistics in the form *f3*(‡Khomani; TEST, Greek/Anatolian). The shared drift between Kum6 and both the early and late Neolithic Aegeans is similar in extent to the drift Aegeans share amongst one another. However, *f4* statistics of the form *f4*(Aegean, Kum6, Early farmer, ‡Khomani) were often significantly positive (see SI Appendix, Table S22 and Dataset S2), suggesting that European Neolithic farmers (namely Linearbandkeramik (LBK), Starcevo and Early Hungarian Neolithic farmers) share some ancestry with early Neolithic Aegeans that is absent in Kum6. This is consistent with population structure in the Early Neolithic Aegean, or Kum6 being sampled from a population that differentiated from early Neolithic Aegeans after they expanded into the rest of Europe. Accordingly, when compared with Barcın, Kum6 shares unique drift with the Late Neolithic genomes from Greece (Klei10 and Pal7), consistent with ongoing gene flow across the Aegean during the 5th millennium, and with archaeological evidence demonstrating similarities in Kumtepe ceramic types with the Greek Late

Neolithic [26]. Finally, the Kum6, Klei10 and Pal7 genomes show signals of Caucasus hunter-gatherer [18] admixture that is absent in the Barcın genomes, suggesting post early Neolithic gene-flow into the Aegean from the east.

It is widely believed that farming spread into Europe along both Mediterranean and central European routes, but the extent to which this process involved multiple dispersals from the Aegean has long been a matter of debate [27]. We calculated *f4* statistics to examine whether the Aegean Neolithic farmers shared drift with genomes from the Spanish Epicardial site Els Trocs in the Pyrenees [3, 12] that is distinct from that shared with Early Neolithic genomes from Germany and Hungary. In a test of the form *f4*(Germany/Hungary EN, Spain EN, Aegean, ‡Khomani), we observe significant unique drift among Neolithic Aegeans (not significantly in Bar8) and Early Neolithic Spain to the exclusion of Hungarian and German Neolithic genomes (see SI Appendix, Table S21). The best explanation for this observation is that migration to southwestern Europe started in the Aegean but was independent from the movement to Germany via Hungary. This is also supported by other genetic inferences [28] and archaeological evidence [29]. An alternative scenario is a very rapid colonization along a single route with subsequent gene flow back to Greece from Spain. Potentially, pre-existing hunter-gatherer networks along the western Mediterranean could have produced a similar pattern, but this is not supported by archaeological data. Interestingly, Ötzi the Tyrolean Iceman [11] shows unique shared drift with Aegeans to the exclusion of Hungarian Early Neolithic farmers and Late and Post Neolithic European genomes, and feasibly represents a relict of Early Neolithic Aegeans (SI Appendix, section 7 and Table S18).

Hunter-gatherer admixture

Given the Aegean is the likely origin of European Neolithic farmers, we utilized Bar8 and Bar31 as putative sources to assess the extent of hunter-gatherer admixture in European farmers through the Neolithic. *f4* statistics of the form *f4*(Neolithic farmer, Anatolian, HG, ‡Khomani) indicated small but significant amounts of hunter-gatherer admixture into both Spanish and Hungarian early farmer genomes, and interestingly, the Early Neolithic Greek genome. Our mixture modeling analysis also inferred a small genetic contribution from the Loschbour hunter-gatherer genome (3-9%) to each of the Early Neolithic Hungarian and German genomes, but evidence of a smaller contribution to any Aegean genomes (0-6%). These results suggest that mixing between migrating farmers and local hunter-gatherers occurred sporadically at low levels throughout the continent even in the earliest stages of the Neolithic. However, consistent with previous findings [3], both *f4* statistics and ADMIXTURE analysis indicate a substantial increase in hunter-gatherer ancestry transitioning into the Middle Neolithic across Europe, while Late

Neolithic farmers also demonstrate a considerable input of ancestry from steppe populations (SI Appendix, section 8 and Figure S32).

Relation to modern populations

Most of the modern Anatolian and Aegean populations do not appear to be the direct descendents of Neolithic peoples from the same region. Indeed, our mixture model comparison of the Aegean genomes to >200 modern groups [2] indicates low affinity between the two Anatolian Neolithic genomes and six of eight modern Turkish samples; the other two were sampled near the Aegean Sea at a location close to the site of the Neolithic genomes. Furthermore, when we form each Anatolian Neolithic genome as a mixture of all modern groups, we infer no contributions from groups in southeastern Anatolia and the Levant, where the earliest Neolithic sites are found (SI Appendix, Fig. S22, Fig. S30 and Table S30). Similarly, comparison of allele sharing between ancient and modern genomes to those expected under population continuity indicates Neolithic to modern discontinuity in Greece and western Anatolia, unless ancestral populations were unrealistically small (SI Appendix, section 9). Instead, our mixing analysis shows that each Aegean Neolithic genome closely corresponds to modern Mediterraneans (>68% contributions from southern Europe), and in particular Sardinians (>25%), as also seen in the PCA and outgroup *f3* statistics, with few substantial contributions from elsewhere. Modern groups matching to Neolithics – mostly from the Mediterranean and North Africa – strikingly match more to Bar8 from Northwestern Anatolia than to the LBK genome from Stuttgart in Germany, indicating that the LBK genome experienced processes such as drift and admixture that were independent from the Mediterranean expansion route, consistent with the dual expansion model.

Concluding remarks

Over the last 7 years, ancient DNA studies have transformed our understanding of the European Neolithic transition [1-4, 12, 13], demonstrating a crucial role for migration in central and southwestern Europe. Our results further advance this transformative understanding by extending the unbroken trail of ancestry and migration all the way back to southwestern Asia.

The high levels of shared drift between Aegean and all available Early Neolithic genomes in Europe, together with the inferred unique drift between Neolithic Aegeans and Early Neolithic genomes from Northern Spain to the exclusion of Early Neolithic genomes from central Europe indicates that Aegean Neolithic populations can be considered the root for all early European farmers, and that at least two independent colonization routes were followed.

A key remaining question is whether this unbroken trail of ancestry and migration extends all the way back to southeastern Anatolia and the Fertile Crescent, where the earliest Neolithic sites in the world are found. Regardless of whether the Aegean early farmers ultimately descended from western or central Anatolian, or even Levantine hunter-gatherers, the differences between the ancient genomes presented here and those from the Caucasus [18] indicates that there was considerable structuring of forager populations in southwest Asia prior to the transition to farming. The dissimilarity and lack of continuity of the Early Neolithic Aegean genomes to most modern Turkish and Levantine populations, in contrast to those of early central and southwestern European farmer and modern Mediterraneans, is best explained by subsequent gene-flow into Anatolia from yet unknown sources.

Methods

Ancient DNA extraction and sequencing. Five Neolithic and two Mesolithic samples from both sides of the Aegean were selected for ancient DNA extraction and sequencing (Table 1). DNA was extracted and Next Generation Sequencing libraries were constructed in dedicated ancient DNA facilities as previously described [30] with slight modifications. DNA quality and quantity of all samples were derived from the combination of estimates of endogenous DNA content based on the percentage of reads mapping to the reference genome (GRCh37/hg19) after shallow Illumina Miseq sequencing and estimating the DNA copy number of extracts by quantitative PCR. The five Neolithic samples (Bar8, Bar31, Rev5, Klei10 and Pal7) showed endogenous DNA contents between 8.80 and 60.83 % and underwent deep Illumina whole genome resequencing. The two Mesolithic samples (Theo1 and Theo5) showed endogenous DNA content of only 0.05 and 0.62 %, respectively, and were used to capture the full mitochondrial genome. SI Appendix Fig. S1 displays the relationship between endogenous DNA content and copy number for each sample and DNA extraction. The enrichment of the mitochondrial genome was carried out with Agilent's SureSelect^{XT} in-solution target enrichment kit. The protocol for the preparation of further libraries for shotgun sequencing and capture was modified according to previously estimated sample quality, whereby some libraries from samples Bar8, Bar31, and Rev5 were prepared with USER™ treated DNA extract. Detailed information about the experimental setup is described in SI Appendix, section 2.

Bioinformatics. All sequence reads underwent 3' adaptor trimming and were filtered for low quality bases. For paired-end sequences only pairs with overlapping sequence were retained and merged into a single sequence. All sequences were aligned against the human reference build GRCh37/hg19 using BWA [31] and realigned using the Genome Analysis Toolkit [32] (SI Appendix, section 3). For genotyping, we developed a novel method to recalibrate quality scores and call genotypes that probabilistically accounts for post-mortem damage (PMD) patterns as estimated in mapDamage2.0 [33]. For low-coverage genomes, we further developed a Bayesian haploid caller to reliably identify the most likely allele call for each site (code available on request from DW). For further details see SI Appendix, section 5.

Ancient DNA authenticity. The assessment of ancient DNA authenticity was performed using the sequence reads mapping to the mitochondrial genome following the likelihood approach described in Fu 2013 [34] (SI Appendix, section 4). Post-mortem damage deamination rates were estimated using mapDamage 2.0 [33] and are displayed together with distribution of DNA fragment lengths of each sample (SI Appendix, Fig. S3). We used ANGSD [35] to determine X-chromosome contamination in male samples (SI Appendix, section 4).

Analysis of uniparental markers. Mitochondrial haplogroups were determined using HaploFind [36]. Consensus sequences in FASTA format were created from alignments with SAMtools [37] (SI Appendix, section 4).

In order to determine patrilineal lineages in ancient samples, we used clean_tree [38]. This software requires BAM format files as input, and alleles are called with SAMtools mpileup at given SNP positions. These SNP positions were provided with the clean_tree software and contain 539 SNPs used for haplogroup determination. (SI Appendix, section 4).

PCA. Principal component analysis was performed with LASER v2.02 [39]. First, a reference space was generated on genotype data of modern individuals. For Figure 2, we used European and Middle Eastern populations from a merged dataset published as part of Hellenthal *et al.* 2014 and Busby *et al.* 2015 [21,40]. In a second step, ancient samples provided as BAM files were projected into the reference space via a Procrustes analysis. See SI Appendix, section 6 for details.

D-statistics, Admixture. *f3* and *f4* statistics and the associated Z-scores (via block jackknife with default options) were calculated using the ADMIXTOOLS package [41] on haploid calls (SI Appendix, section 7). Samples from this study were compared to the Haak *et al.* 2015 dataset [3] containing 2,076 contemporary and ancient individuals. Additionally, ADMIXTURE analysis [42] was performed on a subset of this data containing all Eurasian ancient samples that predate the Bronze Age (n=77) and additionally with Caucasus hunter-gatherers (n=79) and Yamnaya (n=89) (SI Appendix, section 8).

Mixture modeling. To compose a target group as mixtures of other sampled groups, we used the following two-step procedure. First we used a previously described technique [20] to infer an 'allele matching profile' for each target group by comparing its allele frequencies independently at each SNP to that of a set of 'donor' groups. In particular, at a given SNP for each chromosome in our target group, we identified all X non-missing donor chromosomes that shared the same allele type as the target and assigned each of these donors a score of 1/X and all other donors a score of 0. We did this for each SNP and each target individual, and then summed up these scores across SNPs and target individuals to give an allele matching profile for the target group conditional on that set of donors. For each target group, the contributions from each donor group were rescaled to sum to 1. For analyses presented here, our donor groups consisted of modern individuals [2] (plus Neanderthal and Denisova). Our target groups included all modern and ancient groups. Next, as in [21], we performed a multiple linear regression using the target group's allele matching profile as a response and a set of allele matching profiles of different 'surrogate' groups as predictors. In all analyses, we used three different sets of surrogate groups: (i) all (or a subset of) modern groups, (ii) all ancient groups and all (or a subset of) modern groups, (iii) the modern Yoruba plus all (or a subset of) ancient groups. Mixture coefficients were inferred using non-negative least squares (nnls) in R, with a slight modification to ensure the coefficients sum to 1 (SI Appendix, section 10).

Population continuity test. We used a forward simulation approach to test for population continuity. For

our purposes a continuous population is defined as a single panmictic population without admixture from other populations. Our approach is designed to test continuity using a single ancient genome and a set of modern genomes. We designate alleles as ancestral or derived by comparing them to the chimpanzee genome (panTro2), and consider only haploid calls for the ancient genome to avoid genotype calling biases. We examine the proportion of allele sharing between the ancient haploid and modern diploid genome calls that fall into each of the following six classes: A/AA; D/DD; A/DD; D/AA; A/AD; D/AD (where A=ancestral and D=derived alleles in the ancient haploid / modern diploid genomes, respectively). To generate expected proportions of these allele sharing classes we forward simulate genetic drift by binomial sampling from a set of allele frequency vectors based on the modern site frequency spectrum. Finally, we use Fisher's method to combine two-tailed p-values for the observed sharing class fractions falling into the simulated ranges, and compare the resultant χ^2 values to those obtained by comparing each simulation against the set of all other simulations [43] to obtain a p-value. We performed this test for a range of assumed ancient and modern effective population sizes (SI Appendix, section 9).

Runs of homozygosity. The distribution of runs of homozygosity (ROH) for 5 ancient [2,13] and 2,527 modern individuals [44] was determined with PLINK v1.90 [45] following the specifications used in [13], with a set of 1,447,024 transversion SNPs called securely across all ancient samples (SI Appendix, section 11).

Functional markers. Genotypes were determined using the diploid genotyping method described in SI Appendix, section 5, and further verified through direct observation of BAM files using samtools tview [htslib.org]. We included sites having $\geq$ 2X coverage in the analysis (SI Appendix, section 12).

Acknowledgments

ZH and RM are supported by a Marie Curie Initial Training Network (BEAN / Bridging the European and Anatolian Neolithic, GA No: 289966) awarded to MC, SS, DGB, MGT, and JB. CP, JB, and SK received funding from DFG (BU 1403/6-1). CP and JB received funding from the Alexander von Humboldt Foundation. CS and MS were supported by the EU: SYNTHESYS / Synthesis of Systematic Resources, GA No: 226506-CP-CSA-INFRA, DFG: (BO 4119/1) and Volkswagenstiftung (FKZ: 87161). AS was supported by the EU: CodeX Project No: 295729. KoKo, ST, DCUK, PH, CP were co-financed by the EU Social Fund and Greek national funds, research funding program:THALES. CP, MU, KoKo, ST, DCUK were co-financed by the EU Social Fund (ESF) and the Greek national funds, research funding program: ARISTEIA II. MC was supported by Swiss NSF grant 31003A_156853. AK, DW were supported by Swiss NSF grant 31003A_149920. SL is supported by BBSRC (Grant Number BB/L009382/1). LvD is supported by CoMPLEX via EPSRC (Grant Number EP/F500351/1). GH is supported by a Sir Henry

Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 098386/Z/12/Z) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. MGT and YD are supported by a Wellcome Trust Senior Research Fellowship awarded to MGT. JB is grateful for support by the University of Mainz and the HPC cluster MOGON (funded by DFG; INST 247/602-1 FUGG). FG was supported by grant no. 380-62-005 of the Netherlands Organization for Scientific Research. We thank Songül Alpaslan for help with sampling in Barcın and Eleni Stravopodi for help with sampling in Theopetra.

Accession numbers: Mitochondrial genome sequences are deposited in GenBank (KU171094 - KU171100). Genomic data are available at ENA with the accession number PRJEB11848 in BAM format.

References

1. Bramanti B, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science 326(5949):137-140.
2. Lazaridis I, et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513(7518):409-413.
3. Haak W, et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522(7555):207-211.
4. Skoglund P, et al. (2014) Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science 344(6185):747-750.
5. Weninger B, et al. (2014) Neolithisation of the Aegean and Southeast Europe during the 6600–6000 calBC period of Rapid Climate Change. Documenta Praehistorica 41:1-31.
6. Özdoğan M (2014) A new look at the introduction of the Neolithic way of life in Southeastern Europe. Changing paradigms of the expansion of the Neolithic way of life. Documenta Praehistorica 41:33-49.
7. Perlès C, Quiles A, & Valladas H (2013) Early seventh-millennium AMS dates from domestic seeds in the Initial Neolithic at Franchthi Cave (Argolid, Greece). Antiquity 87(338):1001-1015.
8. Kotsakis K (2014) Domesticating the Periphery - New Research into the Neolithic of Greece. Pharos 20(1):41-73.
9. Maniatis Y (2014) Radiocarbon dating of the major cultural changes in Prehistoric Macedonia: recent developments. 1912-2012. A Century of Research in Prehistoric Macedonia. Proceedings of the International Conference, Archaeological Museum of Thessaloniki, 22-24 November 2012, eds Stefani E, Merousis N, & Dimoula A (Archaeological Museum of Thessaloniki Thessaloniki), pp 205-222.
10. Lacan M, et al. (2011) Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. Proc Natl Acad Sci U S A 108(24):9788-9791.
11. Keller A, et al. (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun 3:698.
12. Mathieson I, et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature doi: 10.1038/nature16152.
13. Gamba C, et al. (2014) Genome flux and stasis in a five millennium transect of European prehistory. Nat Commun 5:5257.
14. Allentoft ME, et al. (2015) Population genomics of Bronze Age Eurasia. Nature 522:167-172, doi:10.1038/nature14507.
15. Itan Y, et al. (2010) A worldwide correlation of lactase persistence phenotype and genotypes. BMC evolutionary biology 10(1): 36.
16. Helgason A, et al. (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nature Genet. 39(2):218-225.
17. Raj T, et al. (2013) Common risk alleles for inflammatory diseases are targets of recent positive selection. Am J Hum Genet 92(4):517-529, doi:10.1016/j.ajhg.2013.03.001.
18. Jones ER, et al. (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun 6:8912.

19. Raghavan M, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505(7481):87-91.
20. Lawson DJ, Hellenthal G, Myers S, & Falush D (2012) Inference of population structure using dense haplotype data. PLoS Genet 8(1):e1002453.
21. Hellenthal G, et al. (2014) A genetic atlas of human admixture history. Science 343(6172):747-751.
22. Leslie S, et al. (2015) The fine scale genetic structure of the British population. Nature 519:309–314.
23. van Dorp L, et al. (2015) Evidence for a Common Origin of Blacksmiths and Cultivators in the Ethiopian Ari within the Last 4500 Years: Lessons for Clustering-Based Inference. PLoS Genet 11(8):e1005397.
24. Omrak A, et al. (2015) Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. Current Biology 26:270–275.
25. Schoop UD (2005) Das anatolische Chalkolithikum. Eine chronologische Untersuchung zur vorbronzezeitlichen Kultursequenz im nördlichen Zentralanatolien und den angrenzenden Gebieten. (Urgeschichtliche Studien I) (Verlag Bernhard Albert Greiner, Remshalden).
26. Korfmann M, et al. (1995) Kumtepe 1993. Bericht über die Rettungsgrabung. Stud Troica 5:237-289.
27. Whittle AWR (1996) Europe in the Neolithic - The Creation of New Worlds (Cambridge Univ. Press, Cambridge) pp 1-460.
28. Olalde I, et al. (2015) A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. Mol Biol Evol., doi:10.1093/molbev/msv181.
29. Özdoğan M (2010) Westward expansion of the Neolithic way of life - sorting the Neolithic package into distinct packages. Proceedings of the 6th International Congress of the Archaeology of the Ancient Near East, eds Matthiae P, Pinnock F, Nigro L, & Marchetti N (Harrassowitz Verlag, Wiesbaden), Vol 1, pp 883-897.
30. Kircher M, Sawyer S, & Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research 40(1):e3.
31. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754-1760.
32. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5):491-498.
33. Jonsson H, Ginolhac A, Schubert M, Johnson PL, & Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29(13):1682-1684.
34. Fu Q, et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol 23(7):553-559.
35. Korneliussen TS, Albrechtsen A, & Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15.
36. Vianello D, et al. (2013) HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. Hum Mutat 34(9):1189-1194.
37. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-2079.
38. Ralf A, van Oven M, Zhong KY, & Kayser M (2015) Simultaneous Analysis of Hundreds of Y-Chromosomal SNPs for High-Resolution Paternal Lineage Classification using Targeted Semiconductor Sequencing. Human Mutation 36(1):151-159.
39. Wang C, Zhan X, Liang L, Abecasis GR, & Lin X (2015) Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. Am J Hum Genet 96(6):926-937.
40. Busby GB, et al. (2015) The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. Curr Biol 25(19):2518-2526.
41. Patterson N, et al. (2012) Ancient admixture in human history. Genetics 192(3):1065-1093.
42. Alexander DH, Novembre J, & Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19(9):1655-1664.
43. Voight BF, et al. (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc Natl Acad Sci U S A 102(51):18508-18513.
44. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526(7571):68-74.
45. Chang C, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4(1):7.
46. Fu Q, et al. (2014) Genome Sequence of a 45,000-year-old modern human from western Sibiria. Nature 514(7523):445-449.

47. Seguin-Orlando A, et al. (2014) Genomic structure in Europeans dating back at least 36,200 years. Science 346(6213):1113-1118.
48. Olalde I, et al. (2014) Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European." Nature 507(7491):225-228.

Figure Legends

Figure 1: North Aegean archaeological sites investigated in Turkey and Greece.

Figure 2: PCA of modern reference populations [21, 40] and projected ancient individuals. The Greek and Anatolian samples reported here cluster tightly with other European farmers close to modern-day Sardinians, however, are clearly distinct from previously published Caucasian hunter-gatherers [18]. This excludes the latter as potential ancestral source population for early European farmers and suggests strong genetic structure in hunter-gatherers of southwest Asia.

Abbreviations: Fennoscandian (Fennosc.), British Isles (Brit. Isles), Central and East European (C./E. Eur.), South European (South Eur.), Ancient DNA data: Pleistocene hunter-gatherer (Plei. HG, [18, 46 47]), Holocene hunter-gatherer (Holocene HG, [2, 4, 13, 18, 48]), Neolithic [2, 4, 12, 13, 28], Late Neolithic/Chalcolithic/Copper Age (LN/Chalc./CA [13, 24]), Bronze Age [13]. Ancient samples are abbreviated consistently using the nomenclature "site-country code-culture", see SI Appendix, Table S14 and Dataset S1 for more information. A three-dimensional PCA-plot can be viewed as 3D Figure (see 3D-figure S4, SI Appendix).

Figure 3: Inferred mixture coefficients when forming each modern (small pies) and ancient (large pies, enclosed by borders matching key at left) group as a mixture of the modern-day Yoruba from Africa and the ancient samples shown in the key at left.