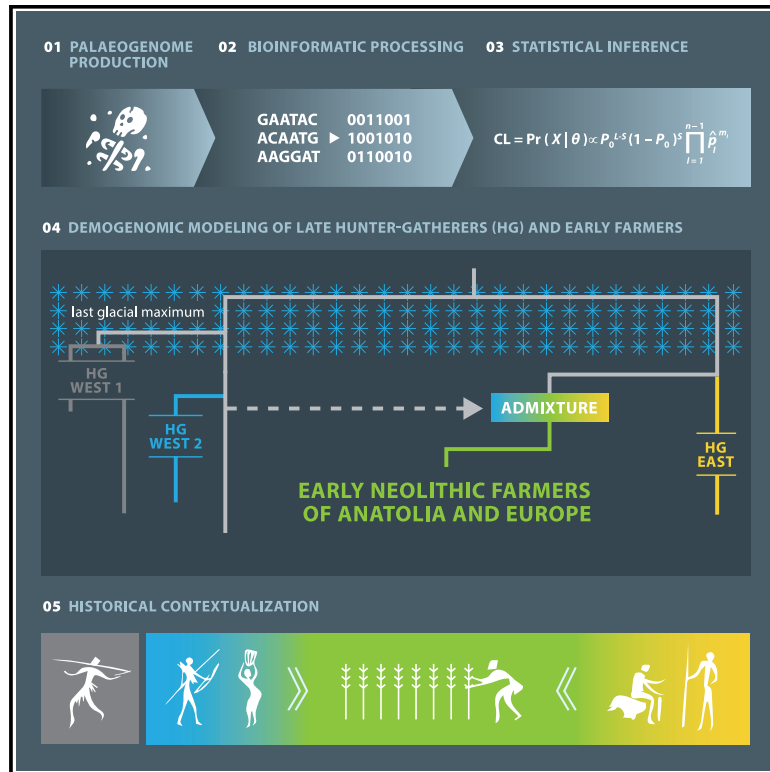


# The genomic origins of the world's first farmers

## Graphical abstract



## Authors

Nina Marchi, Laura Winkelbach, Ilektra Schulz, ..., Daniel Wegmann, Joachim Burger, Laurent Excoffier

## Correspondence

daniel.wegmann@unifr.ch (D.W.),  
jburger@uni-mainz.de (J.B.),  
laurent.excoffier@iee.unibe.ch (L.E.)

## In brief

Ancient DNA analysis and evolutionary modeling have allowed for the ancestral tracing of the Neolithic populations of Southwest Asia and Europe to resolve the genetic origins of the world's first sedentary farmers.

## Highlights

- European HGs diverged from SW Asian HGs during the LGM
- Low genetic diversity of European HGs is due to a strong LGM demographic bottleneck
- Ancestors of western early farmers emerged after repeated post-LGM admixtures
- EFs strongly diverged from SW Asians during their expansion through Anatolia



Article

# The genomic origins of the world's first farmers

Nina Marchi,<sup>1,2,22</sup> Laura Winkelbach,<sup>3,22</sup> Ilektra Schulz,<sup>2,4,22</sup> Maxime Brami,<sup>3,22</sup> Zuzana Hofmanová,<sup>2,4,5,6</sup> Jens Blöcher,<sup>3</sup> Carlos S. Reyna-Blanco,<sup>2,4</sup> Yoan Diekmann,<sup>3,7</sup> Alexandre Thiéry,<sup>1,2,24</sup> Adamandia Kapopoulou,<sup>1,2</sup> Vivian Link,<sup>2,4</sup> Valérie Piuz,<sup>1</sup> Susanne Kreutzer,<sup>3,25</sup> Sylwia M. Figarska,<sup>3</sup> Elissavet Ganiatsou,<sup>8</sup> Albert Pukaj,<sup>3</sup> Travis J. Struck,<sup>9</sup> Ryan N. Gutenkunst,<sup>9</sup> Necmi Karul,<sup>10</sup> Fokke Gerritsen,<sup>11,12</sup> Joachim Pechtl,<sup>13</sup> Joris Peters,<sup>14,15</sup> Andrea Zeeb-Lanz,<sup>16</sup> Eva Lenneis,<sup>17</sup> Maria Teschler-Nicola,<sup>18,19</sup> Sevasti Triantaphyllou,<sup>20</sup> Sofija Stefanović,<sup>21</sup> Christina Papageorgopoulou,<sup>8</sup> Daniel Wegmann,<sup>2,4,23,\*</sup> Joachim Burger,<sup>3,23,\*</sup> and Laurent Excoffier<sup>1,2,23,26,\*</sup>

<sup>1</sup>CMPG, Institute for Ecology and Evolution, University of Berne, 3012 Berne, Switzerland

<sup>2</sup>Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

<sup>3</sup>Palaeogenetics Group, Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg University Mainz, 55099 Mainz, Germany

<sup>4</sup>Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland

<sup>5</sup>Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

<sup>6</sup>Department of Archaeology and Museology, Faculty of Arts, Masaryk University, 60200 Brno, Czech Republic

<sup>7</sup>Research Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK

<sup>8</sup>Laboratory of Physical Anthropology, Department of History and Ethnology, Democritus University of Thrace, 69100 Komotini, Greece

<sup>9</sup>Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

<sup>10</sup>Department of Prehistory, Istanbul University, 34434 Istanbul, Turkey

<sup>11</sup>Netherlands Institute in Turkey, 34433 Istanbul, Turkey

<sup>12</sup>Vrije Universiteit, 1081 HV Amsterdam, the Netherlands

<sup>13</sup>Institute of Archaeology, Innsbruck University, 6020 Innsbruck, Austria

<sup>14</sup>ArchaeoBioCenter and Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, LMU Munich, 80539 Munich, Germany

<sup>15</sup>SNSB, State Collection of Palaeoanatomy Munich, 80333 Munich, Germany

<sup>16</sup>Generaldirektion Kulturelles Erbe Rheinland-Pfalz, Dir. Landesarchäologie, 67346 Speyer, Germany

<sup>17</sup>Department of Prehistoric and Historical Archaeology, University of Vienna, 1190 Vienna, Austria

<sup>18</sup>Department of Anthropology, Natural History Museum of Vienna, 1010 Vienna, Austria

<sup>19</sup>Department of Evolutionary Anthropology, University of Vienna, 1030 Vienna, Austria

<sup>20</sup>Department of History and Archaeology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

<sup>21</sup>Department of Archaeology, Laboratory for Bioarchaeology, University of Belgrade, 11000 Belgrade, Serbia

<sup>22</sup>These authors contributed equally

<sup>23</sup>Senior author

<sup>24</sup>Present address: Cardio-CARE AG, Davos-Wolfgang, Switzerland

<sup>25</sup>Present address: Functional Genomics Center Zurich/GEML, ETH Zurich, Zurich, Switzerland

<sup>26</sup>Lead contact

\*Correspondence: [daniel.wegmann@unifr.ch](mailto:daniel.wegmann@unifr.ch) (D.W.), [jburger@uni-mainz.de](mailto:jburger@uni-mainz.de) (J.B.), [laurent.excoffier@iee.unibe.ch](mailto:laurent.excoffier@iee.unibe.ch) (L.E.)  
<https://doi.org/10.1016/j.cell.2022.04.008>

## SUMMARY

The precise genetic origins of the first Neolithic farming populations in Europe and Southwest Asia, as well as the processes and the timing of their differentiation, remain largely unknown. Demogenomic modeling of high-quality ancient genomes reveals that the early farmers of Anatolia and Europe emerged from a multi-phase mixing of a Southwest Asian population with a strongly bottlenecked western hunter-gatherer population after the last glacial maximum. Moreover, the ancestors of the first farmers of Europe and Anatolia went through a period of extreme genetic drift during their westward range expansion, contributing highly to their genetic distinctiveness. This modeling elucidates the demographic processes at the root of the Neolithic transition and leads to a spatial interpretation of the population history of Southwest Asia and Europe during the late Pleistocene and early Holocene.

## INTRODUCTION

Genetic analyses of skeletal remains from prehistoric sites have greatly enriched our knowledge of the changes that brought sedentism and food-production, along with new people,

material culture and practices, to Europe approximately 8.6 kya (kiloyears ago) through processes often described in archaeology as the “Neolithic revolution” (Childe, 1936). These processes are thought to have reached a tipping point ~11.7 kya in Southwest Asia, where plants and animals were first



domesticated (Fuller et al., 2011; Peters et al., 1999). From this region, it is widely agreed that farming spread into Europe along two main routes, the “Mediterranean” route and the “Danubian” route (Shennan, 2018); it is the latter route that forms the focus of the research described here. Despite this well-developed archaeological narrative, the genetic origins of the world’s first farmers and the spatiotemporal scope of the processes involved remain elusive, in large part due to the lack of high-quality ancient genomes derived from the populations involved in the crucial initial phases of farming expansion.

To date, palaeogenetic studies have established that European early farmers (EFs) were genetically distinct from contemporary Holocene hunter-gatherers (HGs) inhabiting the continent (Bramanti et al., 2009; Skoglund et al., 2012). Apparently, genetic exchange between EFs and HGs appear to have been limited in the early phases of the agricultural expansion, with more intense exchange taking place in the later stages (Lazaridis et al., 2014; Mathieson et al., 2015). Most farmers present in Continental Europe around 9 kya appear to have descended from populations inhabiting the Aegean basin and the Eastern Marmara region (Hofmanová et al., 2016), but their ultimate genetic and geographic origins are still a matter of debate.

Early Neolithic (EN) farmers from the Aegean are clearly related to Central Anatolian farmers (Kılınç et al., 2017), but they also show affinities with Pre-Pottery Neolithic farmers of the Southern Levant (Lazaridis et al., 2016). This suggests a common origin of all these populations prior to the westward spread of agriculture (Kılınç et al., 2016), potentially in the Fertile Crescent area, an archaeologically significant region that contained parts of modern-day Iran, Iraq, Israel, Palestine, Jordan, Lebanon, Syria, and Turkey. However, research has also revealed that Aegean farmers are genetically distinct from early farming populations from the eastern wing of the Fertile Crescent, the Zagros region of Iran and northern Iraq, which may indicate parallel adoption of farming practices by genetically distant groups of HGs across Southwest Asia (Broushaki et al., 2016). Furthermore, there is some evidence of genetic continuity between Epipalaeolithic and Neolithic populations of Central Anatolia (Feldman et al., 2019), suggesting local transitions to agriculture without major gene flow. To make the picture even more complex, some Central Anatolian EFs also show genetic affinities to Caucasus HGs as represented by a 10th millennium before present (BP) genome from Kotias in Western Georgia (Jones et al., 2015; Kılınç et al., 2016; Skourtanoti et al., 2020). Caucasus HGs are themselves thought to be closely related to early Iranian Neolithic farmers (Lazaridis et al., 2016) as well as to later Pontic-Caspian steppe pastoralists (Lazaridis, 2018; Mathieson et al., 2018; Narasimhan et al., 2019).

Despite efforts to understand the genomic history of early farming populations in Europe and Southwest Asia, we still lack a detailed historical scenario of population demography, divergence, and migration embedded in geographic and temporal frameworks. Only a few analyses have previously estimated divergence times between ancestors of Neolithic and HG groups (Broushaki et al., 2016; Jones et al., 2015); the models used were additionally simplistic. The palaeogenetic conclusions outlined above mainly derive from the interpretation of descriptive analyses and summary statistics (e.g., principal

component and admixture analyses,  $f$ -statistics), usually computed on low coverage genomes and/or on ascertained sets of SNPs (Patterson et al., 2012). Such data are difficult to integrate into “demogenomic” (shorthand for demographic modeling using genomic information) analyses. The goal of this paper is to reconstruct the ancestry of Southwest Asian and European EFs and the processes that contributed to their differentiation from HGs. To do so, we produced 15 high-resolution genomes of early Holocene farmers and HGs distributed along a geographical and temporal transect reaching from Southwest Asia to the Rhine valley in West-Central Europe (Figure 1). DNA was extracted from skeletons recovered from some of the most important archaeological sites in early Holocene Europe and Anatolia, including the first farming villages in the Aegean basin (Barcın, Aktopraklık, Nea Nikomedea); Mesolithic and Neolithic sites from the Iron Gates gorges and other areas of the Central Balkans (Lepenski Vir, Vlasac, Grad-Starčevo, Vinča-Belo Brdo); and the oldest cemeteries and mass grave or “massacre” sites of the Central European EN (Kleinhadersdorf, Asparn-Schletz, Essenbach-Ammerbreite, Dillingen-Steinheim, Herxheim) (Table 1; Figure 1).

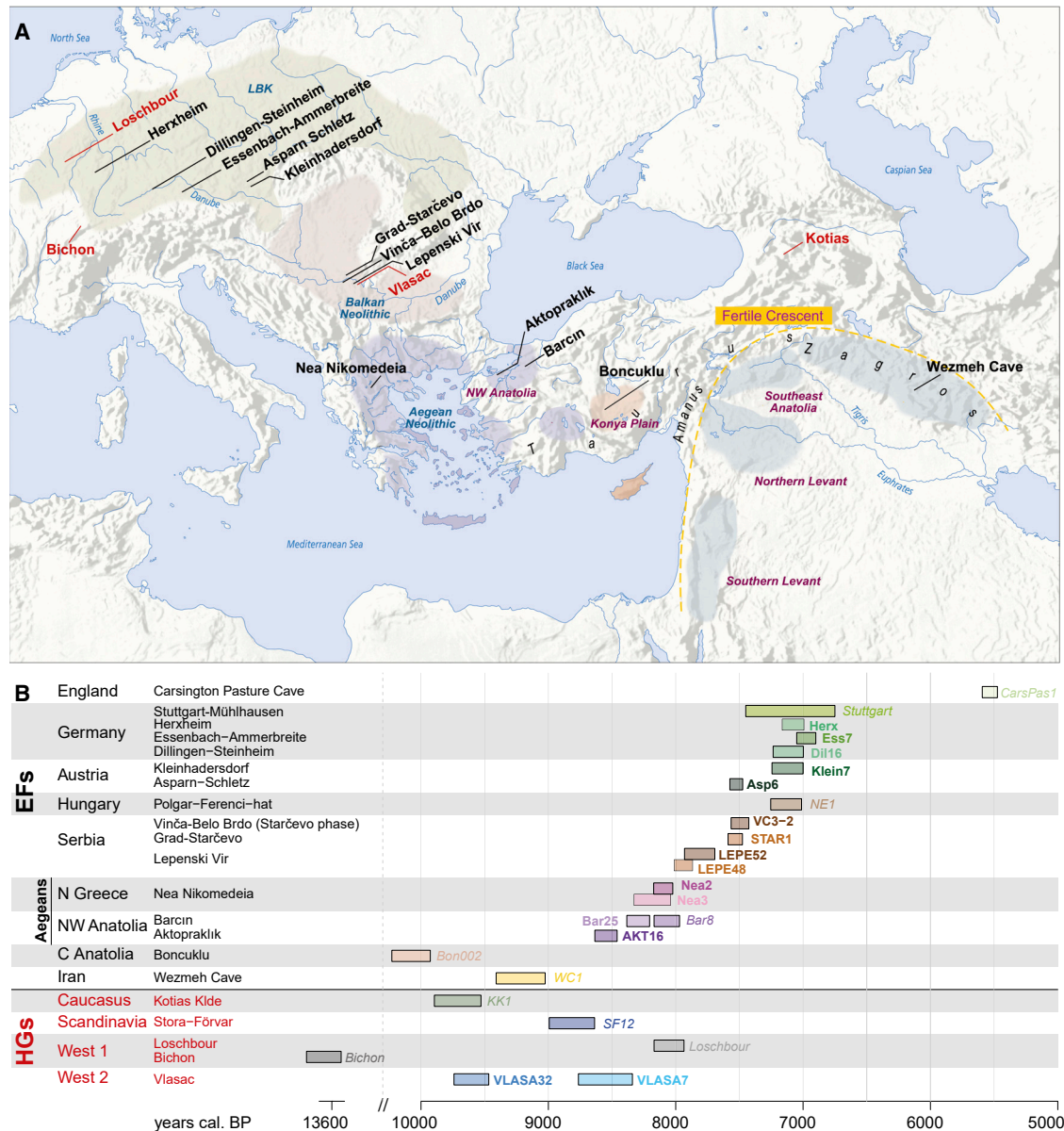
## RESULTS

### Genetic structure and affinities of ancient individuals

Multidimensional scaling (MDS) performed on the neutral portion of the genome (Figure 2A) of ancient individuals and modern reference populations reveals three clusters of ancient individuals: (1) European HGs, (2) western EFs, i.e., EFs from Europe and Anatolia, (3) an EF individual from Iran (WC1) and a Mesolithic HG from Caucasus (KK1). Consistent with previous analyses based on ascertained SNPs (Marcus et al., 2020; Skoglund et al., 2012), the western EFs show strongest affinities with modern Sardinians, with the exception of one English (CarsPas1) and two Northwest Anatolian EFs (Bar8 and AKT16), who are found to cluster with modern individuals from other parts of Southern Europe. In contrast, Palaeolithic and Mesolithic European HGs are genetically well differentiated from all modern Europeans. The Iranian EF and the Caucasus HG appear to be genetically close to modern populations from their sampling area, in keeping with some long-term genetic continuity in those regions. This observation is even more pronounced when performing a MDS analysis on the whole genome, including sites potentially affected by selection (Figure M1\_5 from Methods S1). Generally, ancient individuals appear to be closer to modern ones, once the MDS is computed on the whole genome instead of just neutral sites. This could be explained by a slower evolution of genomic regions influenced by background selection (Pouyet et al., 2018). Another striking difference visible on the whole-genome MDS plot (Figure M1\_5) is that western EFs are closer to some other Southern Europeans than to Sardinians.

### Demogenomic modeling assumptions

To progress our research, we first contrasted alternative models of population differentiation using high-resolution ancient genomes drawn from each of the three MDS clusters described above. Sampled individuals from each cluster were assumed to derive from populations that belong to a large structured



**Figure 1. Spatial and temporal distribution of the ancient genomes analyzed in this study**

(A) Location of archaeological sites with newly sequenced genomes and additional genomes used for modeling: Neolithic (black) and Mesolithic or Palaeolithic (red); different chronological phases of Neolithic expansion (colored areas) and archaeological cultures (blue) along the Danubian route of Neolithization; geographical areas (purple).

(B) Chronological distribution of the 25 genomes analyzed in this study, with the 15 newly sequenced genomes in bold, and the previously published genomes in italics (details in Tables 1 and S3). We also list the cultural groups (EFS, early farmers; HGs, hunter-gatherers), the regions and the archaeological sites where ancient individuals were sampled. The chronological interval at 2 sigma (95.4% probability) is shown for each directly  $^{14}\text{C}$ -dated sample, except for Stuttgart and Ess7, for which approximate dates are given based on the archaeological context.

metapopulation, consisting of interconnected but mostly unsampled populations. Such a model is described in the literature as the “continent-island” model (Excoffier, 2004, 2013) with sampled populations (the islands) receiving a single pulse of gene flow from the metapopulation (the continent) shortly before their sampling time. The three metapopulations were designated *Western*, *Central*, and *Eastern* to reflect the geographic distribu-

tion of the ancient genomes across Europe and SW Asia. These three metapopulations represent the pools of western European HGs, western EFS from Europe and Anatolia, and Iranian EFS, respectively. Additional ancient individuals were then gradually added to the initial model to infer their ancestry and relationships with other populations (see Methods S1—Demogenomic inference with *fastsimcoal2*). Thus, increasingly complex models



**Table 1. Archaeological and genetic information for newly sequenced genomes**

Individual	Period (culture)	Site	Country	Age (cal. BP)	Mean depth (X)	Genetic sex	Haplogroups mtDNA	Haplogroups Y
VLASA7	LM	Vlasac	Serbia	8,764–8,340	15.21	M	U5a2a	I2
VLASA32	LM	Vlasac	Serbia	9,741–9,468	12.65	M	U5a2a	R1b1
AKT16	EN	Aktopraklık	Turkey	8,635–8,460	12.25	F	K1a3	–
Bar25	EN	Barcın	Turkey	8,384–8,205	12.65	M	N1a1a1	G2a2b2a1
Nea3	EN	Nea Nikomedeia	Greece	8,327–8,040	11.57	F	K1a2c	–
Nea2	EN	Nea Nikomedeia	Greece	8,173–8,023	12.51	F	K1a	–
LEPE48	TEN	Lepenski Vir	Serbia	8,012–7,867	10.92	M	K1a1	C1a2b
LEPE52	E-MN	Lepenski Vir	Serbia	7,931–7,693	12.37	M	H3	G2a2b2a1a1c
STAR1	EN (Starčevo)	Grad-Starčevo	Serbia	7,589–7,476	10.55	F	T2e2	–
VC3-2	EN (Starčevo)	Vinča-Belo Brdo	Serbia	7,565–7,426	11.22	M	HV-16311	G2a2a1a3
Asp6	EN (LBK)	Asparn-Schletz	Austria	7,575–7,474	12.11	M	U5a1c1	G2a2b2a3
Klein7	EN (LBK)	Kleinhadersdorf	Austria	7,244–7,000	11.30	F	W1-119	–
Dil16	EN (LBK)	Dillingen-Steinheim	Germany	7,235–6,998	10.60	M	J1c6	C1a2b
Ess7	EN (LBK)	Essenbach-Ammerbreite	Germany	(7,050–6,900 BP)	12.34	M	U5b2c1	G2a2b2a1a1
Herx	EN (LBK)	Herxheim	Germany	7,164–6,993	11.46	F	K1a4a1i	–

LM, late Mesolithic; EN, early Neolithic; TEN, transformational/early Neolithic; E-MN, early-middle Neolithic; LBK, Linearbandkeramik. Samples with genetic sex determined as XX and XY are noted as F and M, respectively.

The samples' ages are based on <sup>14</sup>C dating (95.4% probability), except Ess7, for which an approximate date is given based on the archaeological context.

could be built and tested, resulting in the demographic scenario shown in [Figure 3](#).

### All western EFs share a remote common ancestry with Caucasus HGs

In contrast with previous studies ([Lazaridis et al., 2016](#)), we find that Caucasus HGs (represented by KK1) and western EFs are all descended from a population ancestral to the *Central* metapopulation. This is in line with a recent genetic study showing that KK1 was more closely related to EFs than to western European HGs ([Speidel et al., 2021](#)). This ancestral *Central* metapopulation received about 14% (95% CI 8–26) of its gene pool from the *Western* metapopulation some 14.2 kya (95% CI 13.7–19.0, [Figures 3](#) and M1\_18). Ancestors of the Iranian Neolithic population (represented by WC1) were not affected by this initial admixture: they rather diverged from the *Eastern* metapopulation 13.6 kya (95% CI 11–24.6) after its split from the *Central* metapopulation ~15.8 kya (95% CI 14.3–25.6). Even though Caucasus HGs show closer genetic affinities with EFs from Iran ([Figures 2A](#) and [2B](#)), our analyses suggest that they share a common ancestry with all western EFs.

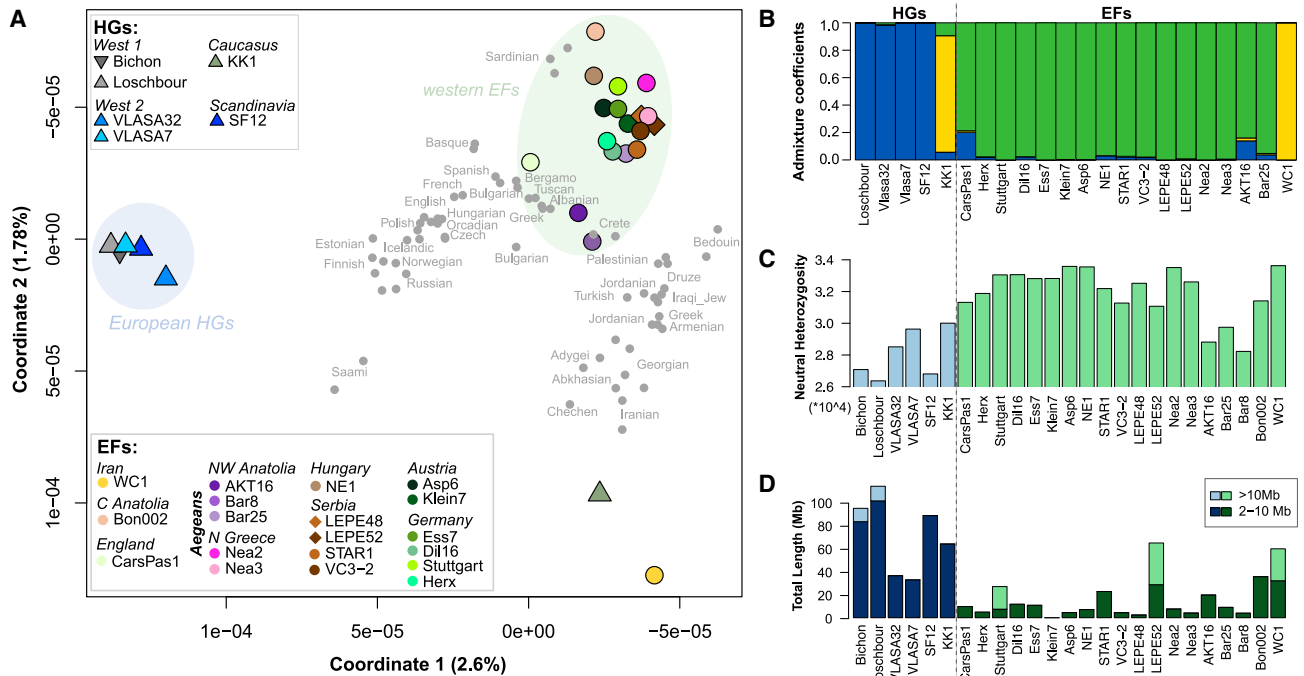
### Ancestors of western EFs admixed twice with western HGs

We find that the ancestors of western EFs received a second pulse of gene flow (15%, 95% CI 6–17) from the *Western* metapopulation ~12.9 kya (95% CI 9.4–13.9), while Caucasus HGs did not ([Figure M1\\_20B](#)). Models that do not include this additional admixture have a lower likelihood and are therefore re-

jected ([Figure M1\\_20A](#)). Thus, the ancestors of western EFs are the product of repeated episodes of gene flow from the *Western* metapopulation. These populations have then diverged from Caucasus HGs due to an intense period of genetic drift between 12.9 and 9.1 kya ([Figures 3](#) and [4](#)). Indeed, we find that their effective population size was reduced to 620 individuals (95% CI 72–2,150) during this relatively long period of drift, which caused them to not only diverge genetically from their ancestral population but also from Caucasus and European HGs, and from Iranian EFs ([Figure 4](#)).

### Anatolian and Aegean farmers differentiation

Populations from Northwest Anatolia (the archaeological sites of Aktopraklık and Barcın) and Northern Greece (Nea Nikomedeia) appear to have diverged from one another at about the same time ~9.1–9.3 kya (95% CI 9.1–12, [Figures M1\\_20](#) and [M1\\_22](#)), potentially during the colonization of the wider Aegean area by EFs. In contrast, EFs from Central Anatolia (represented by a genome from Boncuklu) diverged at least 1,000 years earlier ~10.5 kya (95% CI 10.5–11, [Figure M1\\_24](#)). That Anatolian and Aegean populations show varying amounts of recent gene flow from the *Western* metapopulation suggests different levels of interaction with surrounding HGs. Indeed, genomes from Northern Greece show a lower degree of further HG introgression (3%, 95% CI 1–11) than those of Boncuklu (10%, 95% CI 3–15), Barcın (12%, 95% CI 6–16), and especially Aktopraklık (17%, 95% CI 11–18) ([Table S4](#)). The high level of *Western* metapopulation admixture found in Aktopraklık, a site previously described as influenced by both Epipalaeolithic and Neolithic traditions ([Özdoğan, 2011](#)), is in line with the admixture analysis ([Figure 2B](#))



**Figure 2. Genetic structure and affinities of ancient individuals**

(A) Multidimensional scaling (MDS) analysis performed on the neutrally evolving portion of ancient ( $n = 25$ ; large filled circles and squares: early farmers; triangles: hunter-gatherers) and modern ( $n = 65$ , shown as small circles) genomes from Europe and SW Asia. Ellipses highlight two clusters of ancient individuals: the European hunter-gatherers (HGs) and the European and Anatolians early farmers (i.e., western EFs).

(B) Admixture analyses ( $K = 3$ ) performed on 22 ancient genomes (three genomes with the lowest quality were discarded: Bichon, Bon002, and Bar8). Note that AKT16 in NW Anatolia is more admixed than a neighbor genome from the Barcin site (Bar25), in keeping with  $f$ -statistics analyses (see Figure S1), which has led us to consider them as originating from independent populations in our demographic modeling.

(C) Heterozygosity computed at neutral sites in ancient genomes (HGs in blue, EFs in green).

(D) Runs of homozygosity (ROHs) computed on imputed ancient whole genomes for intermediate ROHs (2–10 Mb, dark color) or long ROHs (>10 Mb, light color), indicative of background relatedness in small populations and close inbreeding, respectively.

See also Figure S1.

and the  $f$ -statistics that show larger genetic affinities of AKT16 with European HGs than other western EFs (Figure S1).

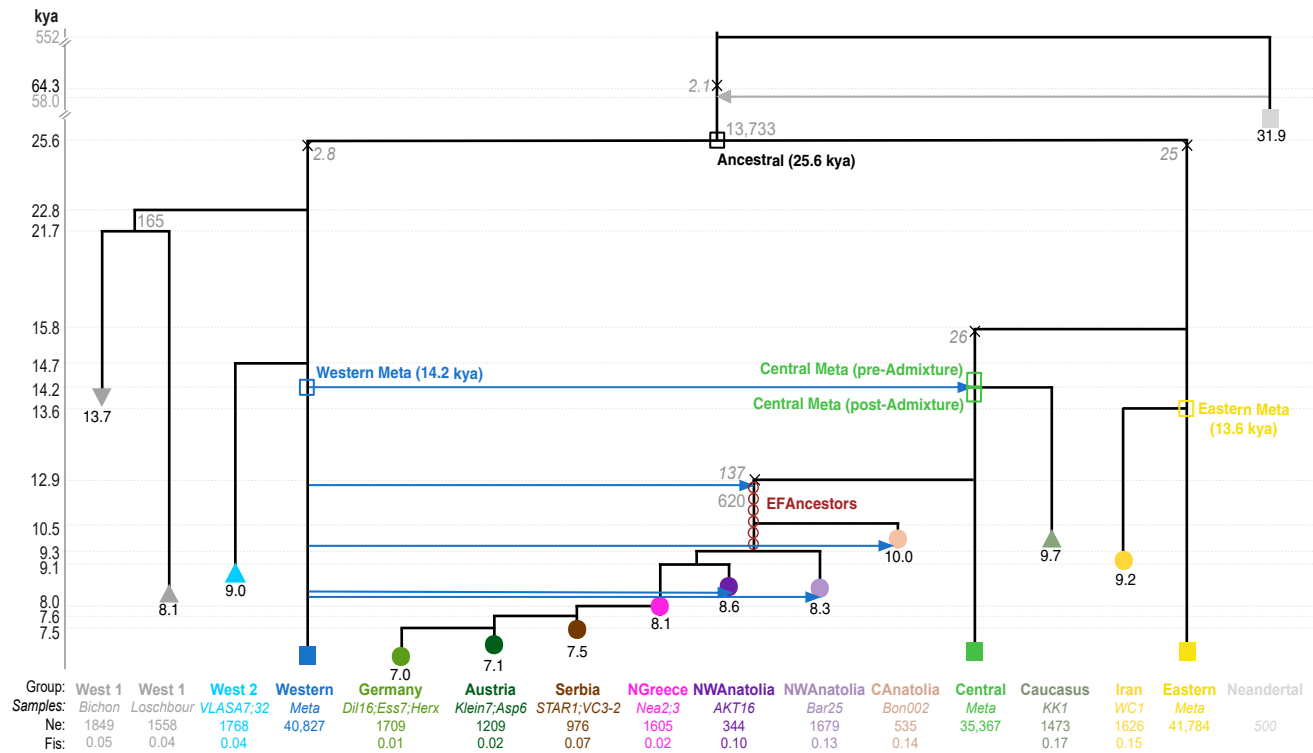
### A stepwise, demic expansion of Neolithic farmers into Central Europe

To understand the progressive spread of EFs into Europe, we expanded the existing demographic model by including three EN populations from Serbia, Austria, and Germany (Figure M1\_25). Even though the model is not spatially explicit, aspects of the EFs spread can be inferred from the spatial and temporal distribution of the archaeological sites. We find that a simple model with a strict stepwise migration of EFs originating in the wider Aegean region (Northwest Anatolia or Northern Greece) and extending to Serbia through the Balkans and along the so-called Danubian corridor, then to Austria, and eventually Germany, is better supported than a scenario allowing for long-distance migration from the Aegean directly to Austria (Figure M1\_26A; Table S4). We also find that early farming communities incorporated a few HG individuals at each modeled stage of their dispersal along the Danubian corridor (2%–7%), compatible with previous estimates of 3%–9% (Hofmanová et al., 2016; Lipson et al., 2017; Nikitin et al., 2019; Figure M1\_26B).

Scenarios without *Western* metapopulation introgression into EF populations have a lower likelihood than scenarios with introgression (Figure M1\_26A). Even though we modeled this introgression from the *Western* metapopulation that is closely related to the newly sequenced genomes from the Mesolithic site of Vlasac, we cannot exclude that it actually occurred from other European HG groups like those related to Loschbour and Bichon. Previous genetic studies have indeed suggested that different Mesolithic backgrounds could have introgressed into the EF gene pool in different regions of Europe (Lipson et al., 2017).

### A last glacial maximum divergence between Eastern and Western metapopulations

Our model also provides important insights regarding the deep branching of pre-Neolithic populations. The divergence between the ancestors of the *Western* and *Eastern* metapopulations is estimated to date back to ~25.6 kya (95% CI 17.3–31.3, Figure M1\_18). This is much younger than the previously inferred divergence time between the ancestors of western European HGs and either Iranian EFs (46–77 kya; Broushaki et al., 2016) or European EFs (46 kya; Jones et al., 2015). However, these



**Figure 3. Demographic scenario inferred from genomic modeling**

This demographic history was obtained by compiling the best models of all tested scenarios (see [Methods S1](#)— Demogenomic inference with *fastsimcoal2*— Final model). Times of the events (y axis) and population ages (shown below their symbols) are indicated in ky BP. Under each population name, we indicate their sampled genomes, their associated inbreeding coefficients (*Fis*), and their diploid effective population sizes (*Ne*). Unfilled symbols indicate ancestral populations that we simulated after or before key events (split times or admixture events). The X symbols indicate bottlenecks that occurred on ancestral branches, modeled as a one-generation bottleneck through a population with its effective size shown in italics. Admixture proportions >10% from the *Western* metapopulation are indicated by blue arrows.

See also [Figure S2](#) for effective populations sizes inferred by *MSMC2*.

previous divergence times were obtained using relatively simple models without bottlenecks and assuming topologies with only recent or even no admixture at all.

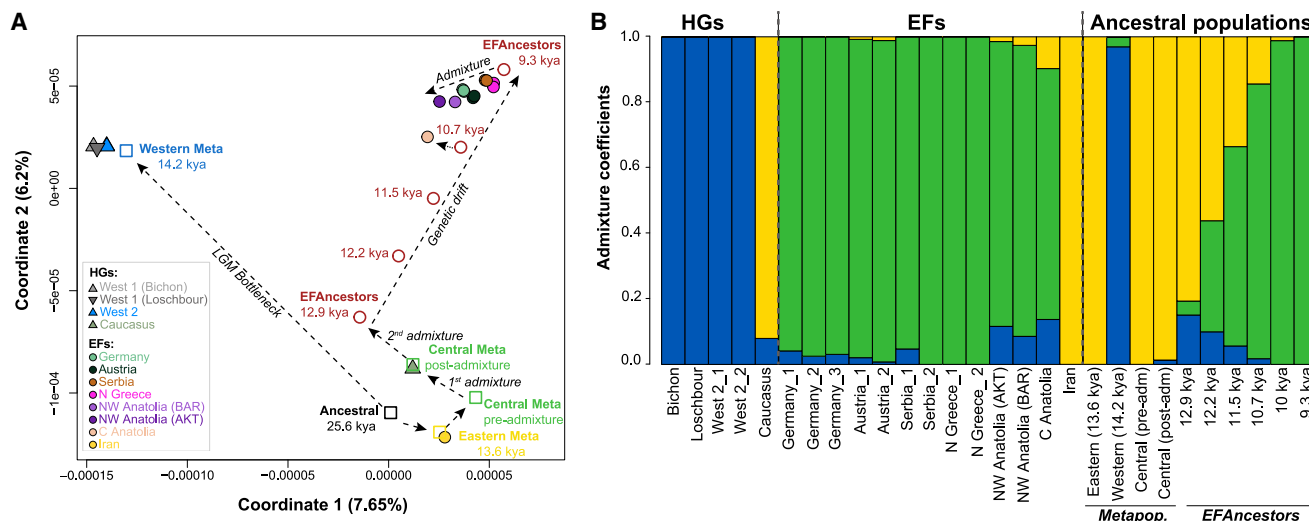
We have explored additional scenarios to evaluate the effects of these simplifications on metapopulation divergence times. As expected, a model without bottlenecks on the metapopulation branches leads to a much older divergence time of 39 kya between *Eastern* and *Western* metapopulations ([Table S4](#)), which is more in line with previous estimates, but this model is inherently less likely than the original one ([Figure M1\\_18A](#)). On the other hand, models without any admixture between the *Western* HG metapopulation and the ancestors of western EFs lead to a much younger divergence time between the *Eastern* and *Western* HG metapopulations (16 kya, [Table S4](#)), but the fit with the data is very poor ([Figure M1\\_18A](#)).

Comparing two western European HGs from the sites of Bichon and Loschbour with our newly sequenced Mesolithic individuals from Vlasac further reveals that European HG populations had already split during the last glacial maximum (LGM) ~22.8 kya (95% CI 16.7–24.7, [Figure M1\\_28](#)), with Bichon and Loschbour populations subsequently diverging from one another, approximately 1,000 years later.

### The reduced diversity in European HGs is due to a massive LGM bottleneck

Genetic diversity as quantified by the heterozygosity at neutral sites is much lower in HGs than in EFs ([Figure 2C](#)), excepting Northwest Anatolian EFs in line with previous studies ([Kilınç et al., 2016](#); [Kousathanas et al., 2017](#)). HG genomes furthermore show a generally larger proportion of intermediate runs of homozygosity (ROHs) (2–10 Mb ROHs, [Figure 2D](#); [Figure M1\\_7](#); [Ringbauer et al., 2021](#)) indicative of background relatedness within European HGs, usually attributed to small population size ([Ceballos et al., 2021](#))—a small population size is also observed in *MSMC2* analyses ([Figure S2](#)).

However, it is interesting to note that we estimate in our model HG effective population sizes to be larger than those of most EFs, particularly those from Anatolia (Bon002, AKT16, and Bar25), which show effective population sizes of a few hundred individuals only ([Figure 3](#)), consistent with their high proportion of intermediate length ROHs ([Figure 2D](#)). The lower diversity observed among European HGs thus appears to be rather due to a very strong LGM bottleneck ([Figures 3](#) and [M1\\_18](#) and see [discussion](#)) than to individuals living in small isolated groups.



**Figure 4. Evolutionary insights gained from the demographic scenario shown in Figure 3**

(A) MDS analysis done on 12 populations used in the demogenomic analyses and on simulated ancestral populations (unfilled symbols) sampled at key moments of their history, as defined in Figure 3: (1) on the ancestral branch before the split between *Western* and *Eastern* metapopulations 25.6 kya; (2) on the *Central* metapopulation branch just before and after the admixture occurring 14.2 kya; (3) on the *Western* metapopulation branch just before this admixture; (4) on the *Eastern* metapopulation branch at the time of split of the Iranian population 13.6 kya; (5) at the top of the western EFs ancestors branch just after its admixture with the *Western* metapopulation (12.9 kya), and then every 25 generations until the split of the Aegean populations 9.3 kya. Arrows indicate the trajectory of the populations caused by important demographic events (i.e., admixture events, bottlenecks, episodes of drift).

(B) Admixture plot for  $K = 3$  performed on sampled and ancestral populations.

See also Figure S3 for corresponding admixture plots done on observed and simulated individuals.

### The inferred model reproduces key features of genomic data

Genomic data simulated under the most complete demographic scenario (Figure 3) lead to population relationships (Figure 4A) that are very similar to those observed on the MDS plot performed on real data (Figure 2A). Three clusters are clearly visible, with European HGs all in proximity to each other, and markedly distinct from the western EFs; Caucasus HGs, and Iran EFs in contrast remain distinct. Furthermore, a simplified admixture graph (Patterson et al., 2012) fitted on data simulated under the model presented in Figure 3 leads to  $f$ -statistics consistent with those calculated on the real data (compare Figures M1\_29 and M1\_30). All population relationships inferred in our model of Figure 3 are thus confirmed by  $f$ -statistics. Finally, an admixture analysis performed on data simulated according to our model leads to admixture proportions that are very similar to those observed, as shown in Figures 4B and S3. In particular, the Caucasus HG shows a large yellow component shared with the Iranian EF, in line with their proximity on the MDS plot. The good match between observed and simulated data thus provides an *a posteriori* validation of our model-based approach.

## DISCUSSION

### Evolutionary insights gained from explicit demographic modeling

Our sequencing of ancient genomes at  $>10\times$  has not only tripled the number of high-quality whole genomes available for the early Holocene in Europe but also allowed us to perform genetic ana-

lyses on an unbiased set of markers minimally impacted by selection. Such neutral markers are ideally suited for reconstructing the population history of Europe and Southwest Asia from the Late Pleistocene to the Early Holocene. In addition to confirming long-held assumptions and interpretations, our modeling provides several key insights about the demographic processes that preceded the Neolithic transition and its expansion to the west.

### A split of European HGs triggered by the LGM

Our model suggests that European HGs had already split into two subgroups (*West 1* and *West 2* in Figure 5C)  $\sim 23$  kya, after experiencing a very severe bottleneck during the LGM, responsible for their low level of genetic diversity (Figure 2C). In contrast to previous studies (Ceballos et al., 2021; Günther et al., 2018), HGs were found to have generally larger effective population sizes than contemporary EFs (Figure 3). Such relatively large effective population sizes can lead to slow population differentiation, which might explain why the different HG groups show close genetic affinities (Figures 2A and 4A) despite long divergence times and a wide geographic distribution. Large HG effective population sizes could be due to long-distance genetic exchanges between groups. Contrastingly, the inferred low effective population size of EFs (despite obvious large census sizes) suggests that the Neolithic transition was linked to a reduction in local EF effective population sizes, potentially due to “sedentarization” or commitment to place (Aimé et al., 2013) and restricted gene flow among small-scale farming communities, as observed at the aceramic Neolithic sites of Boncu-klu and Aşıklı (Yaka et al., 2021).



### Ancestors of western EFs are related to Caucasus HGs

We have shown that HGs from the Caucasus share a common ancestor with western EFs, as both show traces of an ancestral admixture event between the *Western* and *Eastern* metapopulations (Figure 3). Despite this historical relationship, our model still manages to predict the apparent affinity of the Caucasus HG genome with the Iranian EF (e.g. the yellow component shared by KK1 and WC1 found for real and simulated data in Figures 2B, 4B, and S3, as well as their proximities on the MDS plots of Figures 2A and 4A). This shows that observed patterns of genetic similarities are not sufficient to reconstruct actual ancestry relationships.

### Specific demographic processes explain EF and HG differentiation from their ancestral population

Through our demogenomic modeling, we can demonstrate how specific demographic processes gave rise to the genetic divergence of past populations. We can indeed not only simulate the genomic diversity of sampled ancient individuals but also that of individuals drawn from ancestral populations at any time point and thus predict their relationships with sampled individuals (Figure 4A). For instance, we infer that the population ancestral to all our ancient sampled individuals was genetically close to Iranian EFs and Caucasus HGs. As also shown in Figure 4A, the ancestors of European HGs (the *Western* metapopulation) considerably diverged from the ancestral population due to a LGM bottleneck, explaining their outlier position on the top left of the MDS plot.

The ancestors of western EFs (i.e., the *Central* metapopulation) were initially close to the ancestral population and to the Iranian EFs, but the two consecutive admixture events with the *Western* HG metapopulation put them closer to the European HGs on the MDS plot. Nonetheless, it was the >2,500 years of intense genetic drift that made them particularly distinct from all other groups, resulting in them occupying the upper right corner of the MDS plot in Figure 4A. We postulate that this rapid divergence process could be due to recurrent founder effects having occurred during their dispersal through Anatolia. Although western EFs were previously described as genetically intermediate between other SW Asian groups (Feldman et al., 2019), or as a mixture of Iranian and Southern Levant Neolithic populations with western HGs (Lazaridis et al., 2016), this initial admixture signal remains hidden to classical admixture analyses because it is progressively eroded by the genetic drift that occurred during the migration of these populations through Anatolia (Figure 4B). Whereas populations simulated shortly after the two main admixture events with the *Western* metapopulation appear as admixed (i.e., *EFAncestors* [12.9 kya] in Figure 4B), this signal progressively disappears through time with the emergence (and migration) of western EFs who more and more appear as having a completely unrelated gene pool (e.g., green component in the admixture analyses of Figure 4B). Their slightly more central position on the MDS plot (Figure 4A) in later generations is then due to admixture with the *Western* metapopulation and surrounding farmers modeled as the *Central* metapopulation.

### A spatial interpretation of population differentiation

The timing and sequence of demographic events that emerge from our model suggest a scenario of population differentiation

with a clear geographic and chronological resolution from the LGM to the early Holocene (Figure 5).

### HG structuration and divergence induced by the LGM

We find that the divergence between the *Eastern* and *Western* HG metapopulations has been initiated by the LGM some 26 kya (Figures 5A and 5B), probably due to a deterioration of the habitat and a contraction into LGM refugia potentially located in milder regions (Figure 5B). The archaeological, environmental, and climatic records indeed suggest that large tracts of Eurasia were deserted at the height of the LGM, ~26 kya, when ice sheets were at their maximum extent (Clark et al., 2009; French, 2021; Jöris et al., 2009), and both human and animal populations survived in glacial refugia located in more southern regions of Europe (Sommer et al., 2008).

In our initial model, the LGM divergence is immediately followed by a bottleneck of very strong intensity in the population ancestral to European HGs. The modeled intensity  $I$  of this bottleneck depends on the bottleneck duration ( $t$ ) and its size ( $N_{bot}$ ) as  $I = t/(2N_{bot})$ . If the bottleneck had lasted 4,000 years (corresponding to 138 generations of 29 years), our estimated intensity  $I = 0.18$  would correspond to an effective bottleneck population size of 383 individuals. This low number is in line with the archaeological record suggesting a 60% decline in census population size in the latter part of the Gravettian,  $^{14}\text{C}$ -dated to 29,000–25,000 cal. BP, with total population size in Europe as low as 700–1,550 individuals (Maier, 2017). We have explored an alternative demographic scenario where the bottleneck was decoupled from the divergence between the *Eastern* and *Western* HG metapopulations. In that case (Figure M1\_17; Table S4), we find a slightly older metapopulation divergence (27 kya) occurring during the phase of decrease in northern summer solar insolation, i.e., between 32 and 26 kya (Clark et al., 2009), and a more recent bottleneck (23 kya) sitting in the middle of the LGM, confirming that the two events are related to this extreme cold phase.

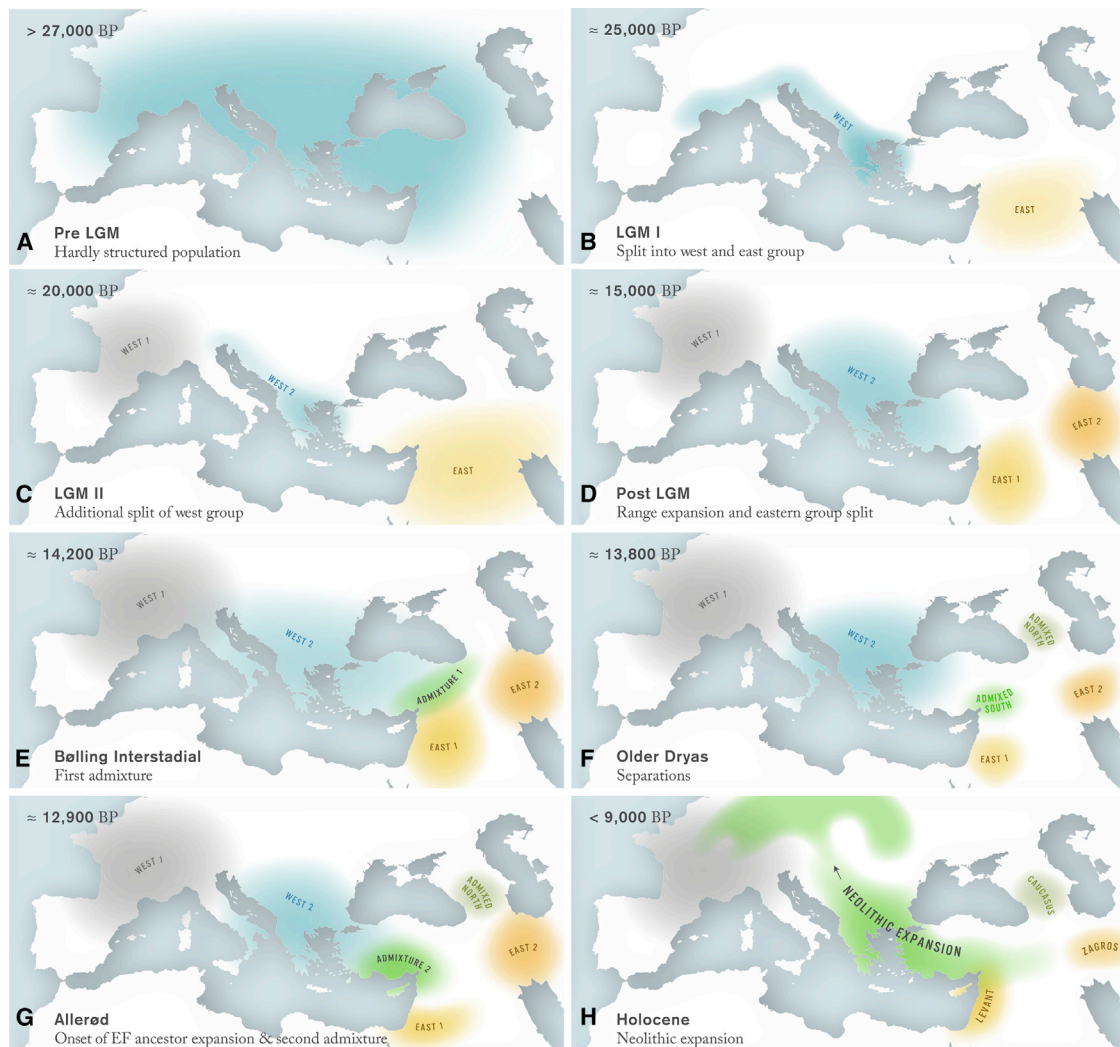
Our analyses further indicate that European HGs differentiated in two separate refugia by the end of the LGM 21.7 kya (Figures 3 and 5C), perhaps corresponding to what archaeologists traditionally identify as the areas of distribution of Solutrean and Epigravettian technocomplexes (Smith, 1966; Koziowski and Kaczanowska, 2004).

### Post-LGM range expansions and admixture

Following a period of range expansion after the LGM (Figure 5D), representatives of the *Central* HG metapopulation, who were likely descendants of the Epigravettian refugial population, mixed ~14.2 kya with the population ancestral to both Caucasus HGs and western EFs (called *East 1* in Figure 5E). Given the assumed geographical distribution of the preceding glacial refugia, this Bölling interstadial period admixture likely happened in a region encompassing Southeastern Anatolia and the Northern Levant or even in neighboring regions such as Central and Eastern Anatolia or the Turkish South coast.

### Differentiation of Anatolian and Aegean EF groups

The demographic processes behind the differentiation between Central Anatolian and Aegean EF populations are more difficult to pinpoint and locate with precision. The inferred low population size of the ancestors of western EFs after their split from the *Central* metapopulation during the Older Dryas (Figure 5F) could



**Figure 5. A spatiotemporal interpretation of population differentiation in SW Asia and Europe based on our model and the geographic distribution of the genomes**

Figure360► For a Figure360 author presentation of this figure, see <https://doi.org/10.1016/j.cell.2022.04.008>.

Colored shaded areas indicate approximate putative distributions of populations at different time points. The letters (A) to (H) indicate the chronological order of events; see main text for a detailed description. Note that warmer periods (Bölling and Allerød interstadials, Holocene) correspond to population range expansions while colder periods (LGM, Older Dryas) are associated with contractions.

See also [Figure S4](#) for additional *f*-statistics analyses supporting alternative connections between the Levant and the Aegean/Greece area.

be due to a westward range expansion and associated recurrent founder effects during the Allerød interstadial. This period was characterized by relatively favorable climatic conditions, which may have allowed the ancestors of western EFs to further mix with Epipalaeolithic HGs in Anatolia (~12.9 kya; [Figure 5G](#)).

The fact that Central Anatolian EFs share the same admixture event and drift with Aegean EFs suggests that they were part of the same expansion wave, and that Central Anatolia was settled by EFs before they reached the Aegean area, potentially by a different route. However, a previous study ([Feldman et al., 2019](#)) has shown that Central Anatolian EFs and Epipalaeolithic HGs were genetically similar, which indicates that admixed groups existed in Central Anatolia prior to the Neolithic transition. Alterna-

tively, admixed HG populations could have moved there from the Fertile Crescent, adopting fully developed farming practices at a later stage, which is in line with the observation that early aceramic sites such as Boncuklu and Aşıklı on the Anatolian Plateau show experiments in crop cultivation and caprine management ~9.7 kya ([Buitenhuis et al., 2018](#); [Ergun et al., 2018](#)).

By contrast, the migration to NW Anatolia ([Figure 5H](#)) likely occurred at the time of the fully developed ceramic Neolithic characterized by the establishment of widespread mixed farming ([Bogaard et al., 2017](#)). Further support for such a demic diffusion scenario to NW Anatolia by a direct (coastal) route and to a lesser extent via the Konya plain region of Central Anatolia comes from *f*-statistics showing Southern Levant populations

sharing more drift with Aegean Neolithic individuals than with Central Anatolian ones (Figure S4). This signal could be due to either (1) long-distance gene flow between Aegean and Levantine communities, as suggested by archaeologists based on similarities in material culture (Perlès, 2001; Horejs et al., 2015) (2) a higher level of Western metapopulation admixture in Central Anatolia as observed in Boncuklu (Figure S1B) and inferred in our model (Figure 3), or (3) an early migration of the Boncuklu ancestors from the Fertile Crescent to Central Anatolia, combined with some later gene flow between populations from the Fertile Crescent and the ancestors of Aegean EFs. However, *f*-statistics analyses reveal that EFs from the Aegean and NW Anatolia are rather heterogeneous in their levels of shared drift with several populations, including HGs from the Levant and EFs from Iran (Figure S4; Table S5), suggesting that the Neolithization of the Aegean was a more complex process.

### Neolithic expansion occurred through a mixture of cultural and demic diffusion

Even though the initial spread of the Neolithic must have been through cultural diffusion in the Fertile Crescent among genetically well differentiated groups, our results indicate that expansion to Northwestern Anatolia, the Aegean Basin and the Danubian corridor occurred primarily through demic diffusion (Figure 5H). The initial spread of populations beyond the Fertile Crescent was certainly far from linear and associated with multiple genetic influences from the Levant, some of which were not linked to the emergence of a fully developed Neolithic economy. As soon as Neolithic lifeways reached Europe, the mode of dispersal of populations to Central Europe became more linear and can essentially be modeled as a stepping-stone migration (Figure M1\_26).

### Limitations of the study

Although our demogenomic model (Figure 3) can clarify observed population affinities, it has exposed temporal and geographical gaps, which can only be filled by producing additional genomic data of similarly high quality. In particular, high-quality genomes from pre-Neolithic and Neolithic populations of the Western Fertile Crescent, as well as reference genomes for Central Anatolia and HGs from Eastern Europe are needed to confirm our conclusions and render them more complete. Whereas the demographic scenarios we have explored are much more complex than those previously investigated (Broushaki et al., 2016; Jones et al., 2015), they are certainly still very schematic and reality must have been more complex. In particular, we have used simplified assumptions such as constant mutation rates and generation times, which may have actually changed over time (Coll Macià et al., 2021; Harris, 2015), or we have modeled genetic interactions between populations as direct and unique pulses of gene flow, whereas gene flow could have occurred over prolonged periods, or could have occurred indirectly through unsampled populations. However, since our model still captures and reproduces most observed genetic patterns (Figure 4), it seems robust enough to provide key insights into past demographic processes. Finally, while we have analyzed genetic evidence relating to European EFs along the Danubian corridor, the spread of EFs along the Mediterranean coast remains to be investigated to obtain a more complete picture of the Neolithic settlement of Europe.

### Conclusions

Population modeling using the framework outlined here allowed us to extract key, unexpected, but complementary and far more detailed information on population affinities than could be concluded from summary statistics and multivariate analysis alone. In addition, this model provides a time frame for the differentiation of the major groups populating Southwest Asia and Europe from the LGM until the introduction of agriculture and highlights the crucial role of climatic changes as driver of population fragmentation and admixture events (Lahr and Foley, 1998). Although the world's first farmers look genetically very different from European HGs, our simulation based on high-quality genomes shows that some European and Southwest Asian populations in fact shared a recent common history marked by repeated interactions since the end of the last ice age. Strong drift during their expansion through Anatolia contributed to making western EFs look more dissimilar than they actually were and somehow concealed their hybrid nature. In summary, the idea of a single cultural and genetic origin of all farmers in the Fertile Crescent, without significant initial contribution of European-like HGs, is no longer tenable in its current form.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Data generation
  - Bioinformatics pipeline
  - Dataset assembling
  - Phenotype predictions
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Population genetics analyses
  - Demographic Analyses

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2022.04.008>.

### ACKNOWLEDGMENTS

We are grateful to Martina Unterländer and Aleksandra Žegarac for help with sample preparation. Lara Cassidy and Kay Prüfer kindly provided access to unpublished raw sequencing data. We thank Ourania Palli and Franz Pieler for useful archaeological information. We thank Katie Meheux for her careful proofreading of the manuscript and Jan Grenner for his voice on the soundtrack of Figure 360. We thank the General Directorate of Cultural Heritage Rhineland-Palatinate, the Anthropological Department of the Natural History Museum Vienna, Michaela Harbeck, and the Bavarian State Collection for Anthropology and Palaeoanatomy, Munich, for providing skeletal samples. We also acknowledge the use of the sequencing platform at the University of Berne for whole-genome sequencing services and support, the IBU cluster



of the University of Berne for NGS data analysis (<https://www.bioinformatics.unibe.ch/>), the UBELIX HPC cluster of the University of Berne (<http://www.id.unibe.ch/hpc>) for the demographic analyses, and the supercomputer Mogen of Johannes Gutenberg University Mainz (<https://hpc.uni-mainz.de>). We finally thank the three anonymous reviewers for their constructive comments, which greatly helped us to improve the manuscript. Fundings: Swiss NSF grant no. 310030\_188883 (L.E. and N.M.); Swiss NSF grant no. 31003A\_173062 (D.W., I.S., V.L., and C.S.R.-B.); Swiss NSF grant no. 310030\_200420 (D.W.); German Science Foundation BU 1403/6-1 (J. Burger, C.P., and S.K.); Humboldt foundation (J. Burger and C.P.); Greek-German bilateral agreement (GSRT and BMBF) project “BIOMUSE” 5030121 (C.P., J. Burger, L.W., E.G., and Y.D.); Serbian Ministry of Science project ID III47001 (S.S.); Czech Grant Agency—GACR 21-17092X (Z.H.); National Institutes of Health grant R01GM127348 (R.N.G., T.J.S.), ERC BIRTH project 640557 (S.S.); European Research Council under the European Union’s Horizon 2020 research and innovation program—ERC-2019-SyG n°856453 (Z.H.); Marie Skłodowska-Curie actions ITN “BEAN” (J. Burger, S.S., S.M.F., C.P., and Z.H.); Marie Skłodowska-Curie individual fellowship (M.B.: 793893 “ODYSSEA”); EMBO Long-Term Fellowship (Z.H.: ALTF 445-2017); Seal of Excellence Fund grant from the University of Berne (N.M.: SELF2018-04); Mainz University (L.W.).

#### AUTHORS CONTRIBUTIONS

Conceptualization, J. Burger, D.W., and L.E.; resources (samples, archaeological, and anthropological context), S.S., M.B., C.P., S.T., N.K., F.G., A.Z.-L., J. Pechtl, J. Peters, E.L., and M.T.-N.; data production, L.W., S.M.F., and S.K.; data curation, I.S., V.L., A.T., and A.K.; formal analyses, investigation, and visualization, N.M., L.E., A.K., E.G., T.J.S., R.N.G., V.P., J. Burger, J. Blöcher, Y.D., Z.H., I.S., A.P., C.S.R.-B., and D.W.; writing – original draft, N.M., M.B., D.W., J. Burger, and L.E.; writing – review & editing: all co-authors.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 13, 2021

Revised: March 4, 2022

Accepted: April 6, 2022

Published: May 12, 2022

#### SUPPORTING CITATIONS

The following references appear in the supplemental information: Adam, (2007); Alpaslan-Roodenberg and Roodenberg, (2020); Alpaslan-Roodenberg, (2011); Ambrose, 1991; Ameur et al., (2012); Angel, (1973); Arbuckle et al., (2014); Atakuman et al., (2020); Baird, (2012); Baird et al., (2013, 2018); Bar-Yosef, (2011); Beerli, (2004); Beichman et al., (2017); Benjamin et al., (2017); Bickle, (2018); Bickle et al., (2015); Bocquet-Appel and Bar-Yosef, (2008); Bonsall et al., (2016); Bonsall et al., (1997, 2008); Borić, (2009); Borić, (2011); Borić, (2016); Borić, (2019); Borić and Dimitrijević, (2007); Borić et al., (2008); Borić and Price, (2013); Boulestin and Coupey, (2015); Boulestin et al., (2009); Brami and Horejs, (2019); Brink-Kloke, (1990); Brotherton et al., (2013); Buckley et al., (2017); Budd et al., (2018); Budd et al., (2013); Burger et al., (2020); Çakırlar, (2013); Carter, (2016); Carter et al., (2019); Cavalli-Sforza et al., (2004); Chapman, (2000); Çilingiroğlu and Çakırlar, (2013); Çilingiroğlu et al., (2020); Clason, (1980); Colledge et al., (2019); Cox et al., (2019); Cristiani et al., (2018); Davison et al., (2006); de Beceidelièvre et al., 2020; de Manuel et al., (2016); Diaz-Papkovich et al., (2019); Diaz-Papkovich et al., (2021); Dietrich and Kocumak, (2001); Douka et al., (2017); Efstratiou et al., (2014); Ehrich, (1977); Enattah et al., (2002); Esin et al., (2016); Fogel et al., (1989); Fujimoto et al., (2008); Fuller et al., (2006); Fumagalli et al., (2015); Gallego-Llorente et al., (2016); Gazal et al., (2014); Gerritsen and Özbal, (2019); Gerritsen et al., 2020; González-Fortes et al., (2017); Haack, (2016); Halstead and Isaakidou, (2020); Harding et al., (2000); Hershkovitz et al., (1997); Hofmanová, (2017); Ivanova et al., (2018); Jovanović et al., (2019); Jovanović, (2017); Kartal, (2003); Karul and Avci, (2011); Kozłowski, (1999); Kozłowski, (2005); Krauß et al., (2018); Kreutzer, (2017); Lagia et al., (2007); Lam-

beck, (1995); Latreille et al., (2009); Lazaridis et al., 2017; Li et al., (2009); Malspinas et al., (2016); Marchi and Excoffier, (2020); Marchi et al., (2021); Marouli et al., 2017; Wild et al., (2004); Mateiciucová, (2015); Mittnik et al., (2019); Nehlich et al., (2010); Neugebauer-Maresch and Lenneis, (2015); Nieszery, (1995); Ohashi et al., (2011); Olalde et al., (2018, 2019); Orschiedt and Haidle, (2012); Ortner, (2003); Özdoğan, (1997); Papathanasiou, (2003); Papathanasiou, (2011, 2015); Papoulia, (2016); Park et al., (2012); Pechtl, (2015); Perlès, (2000); Perlès et al., (2013); Porčić et al., (2020); Pyke and Yiouni, (1996); Quinlan and Hall, (2010); Richards and Hedges, (1999); Riedhammer, (2019); Rodden, (1964, 1965); Rodden et al., (1962); Rodden and Rodden, (1964); Rosenstock et al., (2019); Runnels, (1988); Runnels and Özdoğan, (2001); Schoeninger and DeNiro, (1982); Shang et al., (2013); Slatkin, (2005); Stadler, (2015); Stefanović, (2016); Stefanović and Borić, (2008); Tasić et al., (2015); Tasić et al., (1990, 2016); Teschler-Nicola, (2012); Tiefenböck and Teschler-Nicola, (2015); Tourloukis and Harvati, (2018); Triantaphyllou, (2001); Turck, (2019); Vaiglova et al., (2014); van Andel and Runnels, (1995); Vigne et al., (2012); Wahlund, (1928); Weiberg et al., (2019); Weninger et al., (2014); Whittle et al., (2002); Windl, (1999, 2009); Xue et al., (2009); Yamaguchi et al., (2012); Zeder, (2017); Zeeb-Lanz, (2016); Zeeb-Lanz, (2019a); Zeeb-Lanz, (2019b); Zoledziewska et al., (2015).

#### REFERENCES

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.

Adam, E. (2007). Looking out for the Gravettian in Greece. *Paléo* 19, 145–158.

Aimé, C., Laval, G., Patin, E., Verdu, P., Ségurel, L., Chaix, R., Hegay, T., Quintana-Murci, L., Heyer, E., and Austerlitz, F. (2013). Human genetic data reveal contrasting demographic patterns between sedentary and nomadic populations that predate the emergence of farming. *Mol. Biol. Evol.* 30, 2629–2644.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723.

Alpaslan-Roodenberg, M.S., and Roodenberg, J. (2020). In the light of new data: The population of the first farming communities in the eastern Marmara region. *Prähistorische Z.* 95, 48–77.

Alpaslan-Roodenberg, S.M. (2011). A preliminary study of the burials from late neolithic-early chalcolithic Aktopraklık. *Anatolica* 37, 17–43.

Ambrose, S.H. (1991). Effects of diet, climate and physiology on nitrogen isotope abundances in terrestrial foodwebs. *J. Archaeol. Sci.* 18, 293–317.

Ameur, A., Enroth, S., Johansson, A., Zabolli, G., Igl, W., Johansson, A.C.V., Rivas, M.A., Daly, M.J., Schmitz, G., Hicks, A.A., et al. (2012). Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am. J. Hum. Genet.* 90, 809–820.

Angel, J.L. (1973). Early Neolithic people of Nea Nikomedeia. In *Die Anfänge des Neolithikums vom Orient bis Nordeuropa. Teil VIII Anthropologie*, I. Schwidetzky, ed. (Böhlau Verlag), pp. 103–112.

Arbuckle, B.S., Kansa, S.W., Kansa, E., Orton, D., Çakırlar, C., Gourichon, L., Atici, L., Galik, A., Marciniak, A., Mulville, J., et al. (2014). Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey. *PLoS One* 9, e99845.

Atakuman, Ç., Erdoğan, B., Gemici, H.C., Baykara, İ., Karakoç, M., Biagi, P., Starnini, E., Guilbeau, D., Yücel, N., Turan, D., et al. (2020). Before the Neolithic in the Aegean: the Pleistocene and the Early Holocene record of Bozburun—Southwest Turkey. *J. Isl. Coastal Archaeol.* <https://doi.org/10.1080/15564894.2020.1803458>.

Baird, D. (2012). The late Epipaleolithic, Neolithic, and chalcolithic of the Anatolian plateau. In *A companion to the archaeology of the ancient Near East, 13,000–4000 BC*, D.T. Potts, ed. (Wiley Online Library), pp. 431–465.

Baird, D., Asouti, E., Astruc, L., Baysal, A., Baysal, E., Carruthers, D., Fairbairn, A., Kabukcu, C., Jenkins, E., Lorentz, K., et al. (2013). 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.
- Baird, D., Fairbairn, A., Jenkins, E., Martin, L., Middleton, C., Pearson, J., Asouti, E., Edwards, Y., Kabukcu, C., Mustafaoğlu, G., et al. (2018). Agricultural origins on the Anatolian plateau. *Proc. Natl. Acad. Sci. USA* 115, E3077–E3086.
- Bar-Yosef, O. (2011). Climatic fluctuations and early farming in West and East Asia. *Curr. Anthropol.* 52, S175–S193.
- Beerli, P. (2004). Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Mol. Ecol.* 13, 827–836.
- Beichman, A.C., Phung, T.N., and Lohmueller, K.E. (2017). Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3 (Bethesda)* 7, 3605–3620.
- Benjamin, J., Rovere, A., Fontana, A., Furlani, S., Vacchi, M., Inglis, R.H., Galili, E., Antonioli, F., Sivan, D., Miko, S., et al. (2017). Late Quaternary sea-level changes and early human societies in the central and eastern Mediterranean Basin: an interdisciplinary review. *Quat. Int.* 449, 29–57.
- Bickle, P. (2018). Stable isotopes and dynamic diets: the Mesolithic-Neolithic dietary transition in terrestrial central Europe. *J. Archaeol. Sci.: Rep.* 22, 444–451.
- Bickle, P., Hofmann, D., Bentley, R.A., Hedges, R., Hamilton, J., Lainghas, F., Nowell, G., Pearson, D.G., and Whittle, A. (2015). The isotope results from Kleinhadersdorf. In *Mitteilungen der Prähistorischen Kommission, D.L.G. von Kleinhadersdorf, C. Neugebauer-Maresch, and E. Lenneis, eds. (Verlag der Österreichischen Akademie der Wissenschaften), pp. 173–177.*
- Bocquet-Appel, J.-P., and Bar-Yosef, O. (2008). The Neolithic demographic transition and its consequences (Springer Science and Business Media).
- Bogaard, A., Filipović, D., Fairbairn, A., Green, L., Stroud, E., Fuller, D., and Charles, M. (2017). Agricultural innovation and resilience in a long-lived early farming community: the 1,500-year sequence at Neolithic to early Chalcolithic Çatalhöyük, central Anatolia. *Anatol. Stud.* 67, 1–28.
- Bollongino, R., Nehlich, O., Richards, M.P., Orschiedt, J., Thomas, M.G., Sell, C., Fajkosová, Z., Powell, A., and Burger, J. (2013). 2000 years of parallel societies in Stone Age Central Europe. *Science* 342, 479–481.
- Bonsall, C., Boroneanț, A., Simalcsik, A., and Higham, T. (2016). Radiocarbon dating of Mesolithic burials from Ostrovul Corbului, southwest Romania. In *Southeast Europe and Anatolia in prehistory. Essays in Honor of Vassil Nikolov on His 65th Anniversary, K. Bacvarov and R. Gleser, eds. (Verlag Dr. Rudolf Habelt), pp. 41–50.*
- Bonsall, C., Boroneanț, V., and Radovanović, I. (2008). The Iron Gates in prehistory: new perspectives (Archaeopress).
- Bonsall, C., Lennon, R., McSweeney, K., Stewart, C., Harkness, D., Boronean, V., Bartosiewicz, L., Payton, R., and Chapman, J. (1997). Mesolithic and early Neolithic in the iron gates: a palaeodietary perspective. *J. Eur. Archaeol.* 5, 50–92.
- Bonsall, C., Vasić, R., Boroneanț, A., Roksandic, M., Soficaru, A., McSweeney, K., Evatt, A., Agurajua, Ü., Pickard, C., Dimitrijević, V., et al. (2015). New AMS 14C dates for human remains from Stone Age sites in the Iron Gates reach of the Danube, southeast Europe. *Radiocarbon* 57, 33–46.
- Borić, D. (2009). Absolute dating of metallurgical innovations in the Vinča Culture of the Balkans. In *Metals and societies: studies in honour of Barbara S. Ottaway, T.L. Kienlin and B.W. Roberts, eds. (Verlag Dr. Rudolf Habelt), pp. 191–245.*
- Borić, D. (2011). Adaptations and transformations of the Danube Gorges foragers (c. 13,000–5500 cal. BC): an overview. In *Beginnings – New Research in the Appearance of the Neolithic between Northwest Anatolia and the Carpathian Basin. Papers of the International Workshop 8th - 9th April 2009, D. Ciobotaru, B. Horejs, R. Krauß, and R. Krauß, eds. (Verlag Marie Leidorf GmbH), pp. 157–203.*
- Borić, D. (2016). Deathways at Lepenski Vir: patterns in mortuary practice (Serbian Archaeological Society).
- Borić, D. (2019). Lepenski Vir chronology and stratigraphy revisited. *Starinar*, 9–60.
- Borić, D., and Dimitrijević, V. (2007). When did the “Neolithic package” reach Lepenski Vir? Radiometric and faunal evidence. *Doc. Praehistorica* 34, 53–71.
- Borić, D., French, C., and Dimitrijević, V. (2008). Vlasac revisited: formation processes, stratigraphy and dating. *Doc. Praehistorica* 35, 261–287.
- Borić, D., and Price, T.D. (2013). Strontium isotopes document greater human mobility at the start of the Balkan Neolithic. *Proc. Natl. Acad. Sci. USA* 110, 3298–3303.
- Boulestin, B., and Coupey, A.-S. (2015). Cannibalism in the linear pottery culture: the human remains from Herxheim (Archaeopress).
- Boulestin, B., Zeeb-Lanz, A., Jeunesse, C., Haack, F., Arbogast, R.-M., and Denaire, A. (2009). Mass cannibalism in the Linear Pottery Culture at Herxheim (Palatinate, Germany). *Antiquity* 83, 968–982.
- Brace, S., Diekmann, Y., Booth, T.J., van Dorp, L., Faltyskova, Z., Rohland, N., Mallick, S., Olalde, I., Ferry, M., Michel, M., et al. (2019). Ancient genomes indicate population replacement in Early Neolithic Britain. *Nat. Ecol. Evol.* 3, 765–771.
- Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.-J., et al. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137–140.
- Brami, M., and Horejs, B., eds. (2019). The central/western Anatolian farming frontier. In *Proceedings of the Neolithic Workshop held at 10th ICAANE in Vienna (Austrian Academy of Sciences Press).*
- Brink-Kloke, H. (1990). Das linienbandkeramische Gräberfeld von Essenbach-Ammerbreite. *Ldkr (Germania)* 68, 427–481.
- Bronk Ramsey, C.B. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon* 51, 337–360.
- Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S.M., Der Sarkissian, C., Ganslmeier, R., Friederich, S., et al. (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* 4, 1764.
- Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Diez-del-Molino, D., et al. (2016). Early Neolithic genomes from the eastern Fertile Crescent. *Science* 353, 499–503.
- Browning, B.L., and Browning, S.R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* 93, 840–851.
- Buckley, M.T., Racimo, F., Allentoft, M.E., Jensen, M.K., Jonsson, A., Huang, H., Hormozdiari, F., Sikora, M., Marnetto, D., Eskin, E., et al. (2017). Selection in Europeans on fatty acid desaturases associated with dietary changes. *Mol. Biol. Evol.* 34, 1307–1318.
- Budd, C., Karul, N., Alpaslan-Roodenberg, S., Galik, A., Schulting, R., and Lillie, M. (2018). Diet uniformity at an early farming community in northwest Anatolia (Turkey): carbon and nitrogen isotope studies of bone collagen at Aktopraklık. *Archaeol. Anthropol. Sci.* 10, 2123–2135.
- Budd, C., Lillie, M., Alpaslan-Roodenberg, S., Karul, N., and Pinhasi, R. (2013). Stable isotope analysis of Neolithic and Chalcolithic populations from Aktopraklık, northern Anatolia. *J. Archaeol. Sci.* 40, 860–867.
- Buitenhuis, H., Peters, J., Pöllath, N., Stiner, M.C., Munro, N.D., and Saritaş, Ö. (2018). The faunal remains from levels 3 and 2 of Aşıklı Höyük: evidence for emerging management practices. In *The early settlement at Aşıklı Höyük. Essays in Honor of Ufuk Esin, M. Özbaşaran, G. Duru, and M. Stiner, eds. (Yayınları), pp. 281–324.*
- Burger, J., Link, V., Blöcher, J., Schulz, A., Sell, C., Pochon, Z., Diekmann, Y., Žegarac, A., Hofmanová, Z., Winkelbach, L., et al. (2020). Low prevalence of lactase persistence in Bronze Age Europe indicates ongoing strong selection over the last 3,000 years. *Curr. Biol.* 30, 4307–4315.e13.
- Çakırlar, C. (2013). Rethinking Neolithic subsistence at the gateway to Europe with new archaeozoological evidence from Istanbul. In *Barely surviving or*



- more than enough?, M. Groot, D. Lentjes, and J. Zeiler, eds. (Sidestone Press), pp. 59–79.
- Carter, T. (2016). Obsidian consumption in the Late Pleistocene–Early Holocene Aegean: contextualising new data from Mesolithic Crete. *Annu. Br. Sch. Athens* 111, 13–34.
- Carter, T., Contreras, D.A., Holcomb, J., Mihailović, D.D., Karkanas, P., Guérin, G., Taffin, N., Athanasoulis, D., and Lahaye, C. (2019). Earliest occupation of the Central Aegean (Naxos), Greece: implications for hominin and Homo sapiens' behavior and dispersals. *Sci. Adv.* 5, eaax0997.
- Cavalli-Sforza, L.L., Moroni, A., and Zei, G. (2004). *Consanguinity, inbreeding, and genetic drift in Italy* (Princeton University Press).
- Ceballos, F.C., Gürün, K., Altınışık, N.E., Gemici, H.C., Karamurat, C., Koptekin, D., Vural, K.B., Mapelli, I., Sağlıcan, E., Sürer, E., et al. (2021). Human inbreeding has decreased in time through the Holocene. *Curr. Biol.* 31, 3925–3934.e8.
- Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., Sijen, T., de Knijff, P., Liu, F., Branicki, W., et al. (2018). The HlrisPlex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation. *Forensic Sci. Int. Genet.* 35, 123–135.
- Chan, Y., Salem, R.M., Hsu, Y.-H.H., McMahon, G., Pers, T.H., Vedantam, S., Esko, T., Guo, M.H., Lim, E.T., et al.; GIANT Consortium (2015). Genome-wide analysis of body proportion classifies height-associated variants by mechanism of action and implicates genes important for skeletal development. *Am. J. Hum. Genet.* 96, 695–708.
- Chapman, J. (2000). *Fragmentation in archaeology. People, places and broken objects in the prehistory of south eastern Europe* (Routledge).
- Childe, V.G. (1936). *Man makes himself* (Watts & Company).
- Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69–87.
- Çilingiroğlu, Ç., and Çakırlar, C. (2013). Towards configuring the neolithisation of Aegean Turkey. *Doc. Praehistorica* 40, 21–29.
- Çilingiroğlu, Ç., Kaczanowska, M., Kozłowski, J.K., Dinçer, B., Çakırlar, C., and Turan, D. (2020). Between Anatolia and the Aegean: Epipalaeolithic and Mesolithic foragers of the Karaburun Peninsula. *J. Field Archaeol.* 45, 479–497.
- Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. (2009). The Last Glacial maximum. *Science* 325, 710–714.
- Clason, A.T. (1980). Padina and Starčevo: game, fish and cattle. *Palaeohistoria* 22, 141–173.
- Coll Macià, M., Skov, L., Peter, B.M., and Schierup, M.H. (2021). Different historical generation intervals in human populations inferred from Neanderthal fragment lengths and mutation signatures. *Nat. Commun.* 12, 5317.
- Colledge, S., Conolly, J., Crema, E., and Shennan, S. (2019). Neolithic population crash in northwest Europe associated with agricultural crisis. *Quat. Res.* 92, 686–707.
- Cook, G.T., Bonsall, C., Hedges, R.E.M., McSweeney, K., Boronean, V., and Pettitt, P.B. (2001). A freshwater diet-derived  $^{14}\text{C}$  reservoir effect at the Stone Age sites in the Iron Gates gorge. *Radiocarbon* 43, 453–460.
- Cox, S.L., Ruff, C.B., Maier, R.M., and Mathieson, I. (2019). Genetic contributions to variation in human stature in prehistoric Europe. *Proc. Natl. Acad. Sci. USA* 116, 21484–21492.
- Cristiani, E., Radini, A., Borić, D., Robson, H.K., Caricola, I., Carra, M., Mutri, G., Oxilia, G., Zupancich, A., Šlaus, M., et al. (2018). Dental calculus and isotopes provide direct evidence of fish and plant consumption in Mesolithic Mediterranean. *Sci. Rep.* 8, 8147.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10.
- Davison, K., Dolukhanov, P., Sarson, G.R., and Shukurov, A. (2006). The role of waterways in the spread of the Neolithic. *J. Archaeol. Sci.* 33, 641–652.
- de Bececlière, C., Jovanović, J., Hofmanová, Z., Goude, G., and Stefanović, S. (2020). Direct insight into dietary adaptations and the individual experience of Neolithisation: comparing subsistence, provenance and ancestry of Early Neolithic humans from the Danube Gorges c. 6200–5500 cal BC. In *Farmers at the frontier: a pan European perspective on Neolithisation*, K.J. Gron, L. Sørensen, and P. Rowley-Conwy, eds. (Oxbow Books), pp. 45–76.
- de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V.C., Desai, T., Prado-Martinez, J., Hernandez-Rodriguez, J., Dupanloup, I., Lao, O., Hallast, P., et al. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* 354, 477–481.
- Delaneau, O., Zagury, J.F., Robinson, M.R., Marchini, J.L., and Dermitzakis, E.T. (2019). Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* 10, 5436.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
- Diaz-Papkovich, A., Anderson-Trocme, L., Ben-Eghan, C., and Gravel, S. (2019). UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genet.* 15, e1008432.
- Diaz-Papkovich, A., Anderson-Trocme, L., and Gravel, S. (2021). A review of UMAP in population genetics. *J. Hum. Genet.* 66, 85–91.
- Dietrich, H., and Kociumaka, C. (2001). Jungsteinzeitliche Befunde aus Steinhilf: Stadt und Landkreis Dillingen ad Donau, Schwaben. *Das Archäol. Jahrb. Bayern* 2000, 32–35.
- Dimitrieva, S., and Bucher, P. (2013). UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res.* 41, D101–D109.
- Douka, K., Efstratiou, N., Hald, M.M., Henriksen, P.S., and Karetso, A. (2017). Dating Knossos and the arrival of the earliest Neolithic in the southern Aegean. *Antiquity* 91, 304–321.
- Efstratiou, N., Biagi, P., and Starnini, E. (2014). The Epipalaeolithic site of Ouriakos on the island of Lemnos and its place in the Late Pleistocene peopling of the east Mediterranean region. *Adalya* 17, 1–13.
- Ehrich, R.W. (1977). Starčevo revisited. In *Ancient Europe and the Mediterranean*, V. Markotić, ed. (Aris & Phillips), pp. 59–67.
- Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L., and Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30, 233–237.
- Ergun, M., Tengberg, M., Willcox, G., and Douché, C. (2018). Plants of Aşıklı Höyük and changes through time: first archaeobotanical results from the 2010–14 excavation seasons. In *The early settlement at Aşıklı Höyük. Essays in Honor of Ufuk Esin*, M. Özbaşaran, G. Duru, and M. Stiner, eds. (Yayınları), pp. 191–217.
- Esin, N.V., Esin, N.I., and Yanko-Hombach, V. (2016). The Black Sea basin filling by the Mediterranean salt water during the Holocene. *Quat. Int.* 409, 33–38.
- Excoffier, L. (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol. Ecol.* 13, 853–864.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9, e1003905.
- Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A., and Sousa, V.C. (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* 37, 4882–4885.
- Feldman, M., Fernández-Domínguez, E., Reynolds, L., Baird, D., Pearson, J., Hershkovitz, I., May, H., Goring-Morris, N., Benz, M., Gresky, J., et al. (2019). Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat. Commun.* 10, 1218.

- Fenner, J.N. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423.
- Fogel, M.L., Tuross, N., and Owsley, D.W. (1989). Nitrogen isotope tracers of human lactation in modern and archaeological populations. *Carnegie Inst. Wash. Yearb.* **88**, 111–117.
- French, J.C. (2021). *Palaeolithic Europe: A Demographic and Social Prehistory* (Cambridge University Press).
- Frichot, E., and François, O. (2015). LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929.
- Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lai, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559.
- Fujimoto, A., Kimura, R., Ohashi, J., Omi, K., Yuliwulandari, R., Batubara, L., Mustofa, M.S., Samakkarn, U., Settheetham-Ishida, W., Ishida, T., et al. (2008). A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* **17**, 835–843.
- Fuller, B.T., Fuller, J.L., Harris, D.A., and Hedges, R.E.M. (2006). Detection of breastfeeding and weaning in modern human infants with carbon and nitrogen stable isotope ratios. *Am. J. Phys. Anthropol.* **129**, 279–293.
- Fuller, D.Q., Willcox, G., and Allaby, R.G. (2011). Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeol.* **43**, 628–652.
- Fumagalli, M., Moltke, I., Grarup, N., Racimo, F., Bjerregaard, P., Jørgensen, M.E., Korneliussen, T.S., Gerbault, P., Skotte, L., Linneberg, A., et al. (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343–1347.
- Gallego-Llorente, M., Connell, S., Jones, E.R., Merrett, D.C., Jeon, Y., Eriksson, A., Siska, V., Gamba, C., Meiklejohn, C., Beyer, R., et al. (2016). The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Sci. Rep.* **6**, 31326.
- Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kóvári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257.
- Gazal, S., Sahbatou, M., Perdry, H., Letort, S., Génin, E., and Leutenegger, A.-L. (2014). Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III. *Hum. Hered.* **77**, 49–62.
- Gerritsen, F., and Özbal, R. (2019). Barcın Höyük, a seventh millennium settlement in the Eastern Marmara region of Turkey. *Doc. Praehistorica* **46**, 58–67.
- Gerritsen, F., Özbal, R., and Gerrits, P. (2020). A red floor at Neolithic Barcın Höyük: special or Not? In *Metallurgica Anatolica: Festschrift für Ünsal Yalçın anlässlich seines 65. Geburtstags / Ünsal Yalçın 65. Yaşgünü Armağan Kitabı*, H.G. Yalçın and O. Stegemeier, eds. (Ege Yayınları), pp. 35–43.
- González-Fortes, G., Jones, E.R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Garralda, M.D., Drak, L., Siska, V., Simalcsik, A., et al. (2017). Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. *Curr. Biol.* **27**, 1801–1810.e10.
- Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., and 1000 Genomes Project, and Bustamante, C.D. (2011). Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. Sci. USA* **108**, 11983–11988.
- Günther, T., Malmström, H., Svensson, E.M., Omrak, A., Sánchez-Quinto, F., Kılınc, G.M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., et al. (2018). Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* **16**, e2003703.
- Günther, T., and Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet* **15**, e1008302.
- Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695.
- Haack, F. (2016). The early Neolithic ditched enclosure of Herxheim - architecture, fill formation processes and service life. In *Ritualised destruction in the early Neolithic - the exceptional site of Herxheim (palatinate, Germany)*, A. Zeeb-Lanz, ed. (Generaldirektion Kulturelles Erbe Rheinland-Pfalz), pp. 19–118.
- Halstead, P., and Isaakidou, V. (2020). Pioneer farming in earlier Neolithic Greece. In *Farmers at the frontier: a pan European perspective on Neolithisation*, K.J. Gron, L. Sørensen, and P. Rowley-Conwy, eds. (Oxbow Books), pp. 77–100.
- Harding, R.M., Healy, E., Ray, A.J., Ellis, N.S., Flanagan, N., Todd, C., Dixon, C., Sajantila, A., Jackson, I.J., Birch-Machin, M.A., et al. (2000). Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66**, 1351–1361.
- Harris, K. (2015). Evidence for recent, population-specific evolution of the human mutation rate. *Proc. Natl. Acad. Sci. USA* **112**, 3439–3444.
- Hershkovitz, I., Rothschild, B.M., Latimer, B., Dutour, O., Léonetti, G., Greenwald, C.M., Rothschild, C., and Jellema, L.M. (1997). Recognition of sickle cell anemia in skeletal remains of children. *Am. J. Phys. Anthropol.* **104**, 213–226.
- Hofmanová, Z. (2017). *Palaeogenomic and Biostatistical Analysis of Ancient DNA Data from Mesolithic and Neolithic Skeletal Remains (Johannes Gutenberg-Universität Mainz)*.
- Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. USA* **113**, 6886–6891.
- Horejs, B., Milić, B., Ostmann, F., Thanheiser, U., Weninger, B., and Galik, A. (2015). The Aegean in the Early 7th Millennium BC: maritime Networks and Colonization. *J. World Prehist.* **28**, 289–330.
- Huang, X., Fortier, A.L., Coffman, A.J., Struck, T.J., Irby, M.N., James, J.E., León-Burguete, J.E., Ragsdale, A.P., and Gutenkunst, R.N. (2021). Inferring genome-wide correlations of mutation fitness effects between populations. *Mol. Biol. Evol.* **38**, 4588–4602.
- Ivanova, M., De Cupere, B., Ethier, J., and Marinova, E. (2018). Pioneer farming in southeast Europe during the early sixth millennium BC: climate-related adaptations in the exploitation of plants and animals. *PLoS One* **13**, e0197225.
- Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912.
- Jörös, O., Street, M., and Sirocko, F. (2009). Siedlungsleere – das Kältemaximum der Letzten Kaltzeit (24.000 - 16.000 BP). In *Von der Eiszeit bis ins 21. Wetter, Klima, and Menschheitsentwicklung*, F.S. Jahrhundert, ed. (WBG), pp. 83–87.
- Jovanović, J., de Bececlievre, C., Stefanović, S., Živaljević, I., Dimitrijević, V., and Goude, G. (2019). Last hunters - first farmers: new insight into subsistence strategies in the Central Balkans through multi-isotopic analysis. *Archaeol. Anthropol. Sci.* **11**, 3279–3298.
- Jovanović, J.D. (2017). *The diet and health status of the early Neolithic communities of the central Balkans (6200–5200 BC)* (University of Belgrade). PhD thesis.
- Kartal, M. (2003). Anatolian epi-paleolithic period assemblages: problems, suggestions, evaluations and various approaches. *Anadolu* **24**, 45–62.
- Karul, N., and Avci, M.B. (2011). Neolithic communities in the Eastern Marmara region: Aktopraklik C. *Anatolica* **37**, 1–15.
- Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066.
- Kılınc, G.M., Koptekin, D., Atakuman, Ç., Sümer, A.P., Dönertaş, H.M., Yaka, R., Bilgin, C.C., Büyükkarakaya, A.M., Baird, D., Altınışık, E., et al. (2017). Archaeogenomic analysis of the first steps of Neolithization in Anatolia and the Aegean. *Proc. Biol. Sci.* **284**.
- Kılınc, G.M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A.M., Bıçakçı, E., Baird, D., Dönertaş, H.M., Ghalichi, A., Yaka, R., et al. (2016). The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666.

- Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* *40*, e3.
- Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* *28*, 2520–2522.
- Kousathanas, A., Leuenberger, C., Link, V., Sell, C., Burger, J., and Wegmann, D. (2017). Inferring heterozygosity from ancient and low coverage genomes. *Genetics* *205*, 317–332.
- Kozłowski, J.K. (1999). Gravettian/Epigravettian sequences in the Balkans: environment, technologies, hunting strategies and raw material procurement. *Br. Sch. Athens Stud.* *3*, 319–329.
- Kozłowski, J.K. (2005). The importance of the Aegean basin for the Neolithization of South-Eastern Europe. *J. Isr. Prehist. Soc.* *35*, 409–424.
- Kozłowski, J.K., and Kaczanowska, M. (2004). Gravettian/Epigravettian sequences in the Balkans and Anatolia. *Mediterr. Archaeol. Archaeom.* *4*, 5–18.
- Krauß, R., Marinova, E., De Brue, H., and Weninger, B. (2018). The rapid spread of early farming from the Aegean into the Balkans via the Sub-Mediterranean-Aegean Vegetation Zone. *Quat. Int.* *496*, 24–41.
- Kreutzer, S. (2017). Populationsgenetische Analyse prähistorischer Individuen aus Griechenland.
- Lagia, A., Eliopoulos, C., and Manolis, S. (2007). Thalassemia: macroscopic and radiological study of a case. *Int. J. Osteoarchaeol.* *17*, 269–285.
- Lahr, M.M., and Foley, R.A. (1998). Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am. J. Phys. Anthropol.* *27*, 137–176.
- Lambeck, K. (1995). Late Pleistocene and Holocene sea-level change in Greece and south-western Turkey: a separation of eustatic, isostatic and tectonic contributions. *Geophys. J. Int.* *122*, 1022–1044.
- Latreille, J., Ezzedine, K., Elfakir, A., Ambroisine, L., Gardinier, S., Galan, P., Hercberg, S., Gruber, F., Rees, J., Tschachler, E., et al. (2009). MC1R gene polymorphism affects skin color and phenotypic features related to sun sensitivity in a population of French adult women. *Photochem. Photobiol.* *85*, 1451–1458.
- Lazaridis, I. (2018). The evolutionary history of human populations in Europe. *Curr. Opin. Genet. Dev.* *53*, 21–27.
- Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., et al. (2017). Genetic origins of the Minoans and Mycenaeans. *Nature* *548*, 214–218.
- Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* *536*, 419–424.
- Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* *513*, 409–413.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. 1303.3997v2.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., and Wegmann, D. ATLAS: analysis tools for low-depth and ancient samples. Preprint at bioRxiv, /10.1101/105346.
- Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* *551*, 368–372.
- MacHugh, D.E., Edwards, C.J., Bailey, J.F., Bancroft, D.R., and Bradley, D.G. (2000). The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies. *Anc. Biomol.* *3*, 81–102.
- Maier, A. (2017). Population and settlement dynamics from the Gravettian to the Magdalenian. *Mitt. Ges. Urgeschichte* *26*, 83–101.
- Malaspinas, A.-S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., Crawford, J.E., et al. (2016). A genomic history of Aboriginal Australia. *Nature* *538*, 207–214.
- Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* *538*, 201–206.
- Marchi, N., and Excoffier, L. (2020). Gene flow as a simple cause for an excess of high-frequency-derived alleles. *Evol. Appl.* *13*, 2254–2263.
- Marchi, N., Schlichta, F., and Excoffier, L. (2021). Demographic inference. *Curr. Biol.* *31*, R276–R279.
- Marcus, J.H., Posth, C., Ringbauer, H., Lai, L., Skeates, R., Sidore, C., Beckett, J., Furtwängler, A., Olivieri, A., Chiang, C.W.K., et al. (2020). Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. *Nat. Commun.* *11*, 939.
- Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., et al. (2017). Rare and low-frequency coding variants alter human adult height. *Nature* *542*, 186–190.
- Mateiciucová, I. (2015). Silices. In *Das Linearbandkeramische Gräberfeld von Kleinhadersdorf*, C. Neugebauer-Maresch and E. Lenneis, eds. (Verlag der Österreichischen Akademie der Wissenschaften), pp. 111–122.
- Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. *Nature* *555*, 197–203.
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* *528*, 499–503.
- Matthey-Doret, R., and Whitlock, M.C. (2019). Background selection and FST: consequences for detecting local adaptation. *Mol. Ecol.* *28*, 3902–3914.
- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* *48*, 1279–1283.
- McEvoy, B.P., and Visscher, P.M. (2009). Genetics of human height. *Econ. Hum. Biol.* *7*, 294–306.
- Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* *2010*. pdb.prot5448.
- Mittnik, A., Massy, K., Knipper, C., Wittenborn, F., Friedrich, R., Pfrengle, S., Burri, M., Carlichi-Witjes, N., Deeg, H., Furtwängler, A., et al. (2019). Kinship-based social inequality in Bronze Age Europe. *Science* *366*, 731–734.
- Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. *Science* *365*, eaat7487.
- Navarro-Gomez, D., Leipzig, J., Shen, L., Lott, M., Stassen, A.P.M., Wallace, D.C., Wiggs, J.L., Falk, M.J., van Oven, M., and Gai, X. (2015). Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier. *Bioinformatics* *31*, 1310–1312.
- Nehlich, O., Borić, D., Stefanović, S., and Richards, M.P. (2010). Sulphur isotope evidence for freshwater fish consumption: a case study from the Danube Gorges, SE Europe. *J. Archaeol. Sci.* *37*, 1131–1139.
- Nei, M., and Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* *76*, 5269–5273.
- Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics* *28*, 1919–1920.

- Neugebauer-Maresch, C., and Lenneis, E. (2015). Das Linearbandkeramische Gräberfeld von Kleinhadersdorf. *Mitteilungen der Prähistorischen Kommission*, 82 (Verlag der Österreichischen Akademie der Wissenschaften).
- Nieszery, N. (1995). *Linearbandkeramische Gräberfelder in Bayern* (verlag Marie Leidorf).
- Nikitin, A.G., Stadler, P., Kotova, N., Teschler-Nicola, M., Price, T.D., Hoover, J., Kennett, D.J., Lazaridis, I., Rohland, N., Lipson, M., et al. (2019). Interactions between earliest Linearbandkeramik farmers and central European hunter gatherers at the dawn of European Neolithization. *Sci. Rep.* 9, 19544.
- Ohashi, J., Naka, I., and Tsuchiya, N. (2011). The impact of natural selection on an ABCC11 SNP determining earwax type. *Mol. Biol. Evol.* 28, 849–857.
- Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555, 190–196.
- Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C.J., Gandini, F., Pala, M., et al. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363, 1230–1234.
- Orschiedt, J., and Haidle, M.N. (2012). Violence against the living, violence against the dead on the human remains from Herxheim, Germany. Evidence of a crisis and mass cannibalism? In *Sticks, stones and broken bones: Neolithic violence in a European perspective*, R.J. Schulting and L. Fibiger, eds. (Oxford University Press), pp. 121–138.
- Ortner, D.J. (2003). *Identification of pathological conditions in human skeletal remains* (Academic Press).
- Özdoğan, M. (1997). Anatolia from the Last Glacial Maximum to the Holocene Climatic Optimum: cultural formations and the impact of the environmental setting. *Paléorient* 23, 25–38.
- Özdoğan, M. (2011). Archaeological evidence on the westward expansion of farming communities from eastern Anatolia to the Aegean and the Balkans. *Curr. Anthropol.* 52, S415–S430.
- Papathanasiou, A. (2003). Stable isotope analysis in Neolithic Greece and possible implications on human health. *Int. J. Osteoarchaeol.* 13, 314–324.
- Papathanasiou, A. (2011). Health, diet and social implications in Neolithic Greece from the study of human osteological material. In *Human Bioarchaeology of the Transition to Agriculture*, R. Pinhasi and J. Stock, eds. (John Wiley & Sons, Ltd), pp. 87–106.
- Papathanasiou, A. (2015). Stable isotope analyses in Neolithic and Bronze Age Greece: an overview. *Hesperia* 49, 25–55.
- Papoulia, C. (2016). Late Pleistocene to Early Holocene Sea-crossings in the Aegean: direct, indirect and controversial evidence. In *Géochronologie Des Îles de méditerranée* (CNRS Editions), pp. 33–46.
- Park, J.-H., Yamaguchi, T., Watanabe, C., Kawaguchi, A., Haneji, K., Takeda, M., Kim, Y.-I., Tomoyasu, Y., Watanabe, M., Oota, H., et al. (2012). Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *J. Hum. Genet.* 57, 508–514.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093.
- Pechtl, J. (2015). Verorten in Raum, Zeit und Umwelt. Ein Forschungsprojekt zur ersten bäuerlichen Kultur in Bayerisch-Schwaben. *Denkmalpflege Informationen* 161, 14–17.
- Perlès, C. (2000). Greece, 30,000–20,000 bp. In *Hunters of the golden Age. The mid-upper Palaeolithic of Eurasia 30,000 - 20,000 BP*, W. Roebroeks, M. Mussi, J. Svoboda, and K. Fennema, eds. (University of Leiden), pp. 375–397.
- Perlès, C. (2001). *The early Neolithic in Greece: the first farming communities in Europe* (Cambridge University Press).
- Perlès, C., Quiles, A., and Valladas, H. (2013). Early seventh-millennium AMS dates from domestic seeds in the Initial Neolithic at Franchthi Cave (Argolid, Greece). *Antiquity* 87, 1001–1015.
- Peters, J., Helmer, D., Von Den Driesch, A., and Saña Seguí, M. (1999). Early animal husbandry in the northern Levant. *Paléorient* 25, 27–48.
- Petr, M., Pääbo, S., Kelso, J., and Vernot, B. (2019). Limits of long-term selection against Neandertal introgression. *Proc. Natl. Acad. Sci. USA* 116, 1639–1644.
- Porčić, M., Blagojević, T., Pendić, J., and Stefanović, S. (2020). The Neolithic demographic transition in the Central Balkans: population dynamics reconstruction based on new radiocarbon evidence. *Phil. Trans. R. Soc. Lond. B.* 376, 1816.
- Pouyet, F., Aeschbacher, S., Thiéry, A., and Excoffier, L. (2018). Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife* 7, e36317.
- Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* 358, 655–658.
- Pyke, G., and Youni, P. (1996). Nea Nikomedeia I: The excavation of an early Neolithic village in Northern Greece 1961–1964. The excavation and the ceramic assemblage (The British School at Athens).
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing). <https://www.R-project.org/>.
- Racimo, F., Sikora, M., Vander Linden, M.V., Schroeder, H., and Lalueza-Fox, C. (2020). Beyond broad strokes: sociocultural insights from the study of ancient genomes. *Nat. Rev. Genet.* 21, 355–366.
- Ragsdale, A.P., Coffman, A.J., Hsieh, P., Struck, T.J., and Gutenkunst, R.N. (2016). Triallelic population genomics for inferring correlated fitness effects of same site nonsynonymous mutations. *Genetics* 203, 513–523.
- Ralf, A., Montiel González, D.M., Zhong, K., and Kayser, M. (2018). Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. *Mol. Biol. Evol.* 35, 1291–1294.
- Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334, 94–98.
- Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* 62, 725–757.
- Richards, M.P., and Hedges, R.E.M. (1999). Stable isotope evidence for similarities in the types of marine foods used by late Mesolithic humans at sites along the Atlantic Coast of Europe. *J. Archaeol. Sci.* 26, 717–722.
- Riedhammer, K. (2019). The radiocarbon dates from Herxheim and their archaeological interpretation. In *Ritualised destruction in the early Neolithic – the exceptional site of Herxheim (palatinate, Germany)*, 2, A. Zeeb-Lanz, ed. (Generaldirektion Kulturelles Erbe, Direktion Landesarchäologie), pp. 285–304.
- Ringbauer, H., Novembre, J., and Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* 12, 5425.
- Rodden, R.J. (1964). Recent discoveries from prehistoric Macedonia: an interim report. *Balk. Stud.* 5, 109–124.
- Rodden, R.J. (1965). An early Neolithic village in Greece. *Sci. Am.* 212, 82–92.
- Rodden, R.J., Dibleby, G.W., Western, A.C., Willis, E.H., Higgs, E.S., and Clench, W.J. (1962). Excavations at the Early Neolithic Site at Nea Nikomedeia, Greek Macedonia (1961 season). *Proc. Prehist. Soc.* 28, 267–288.
- Rodden, R.J., and Rodden, J.M. (1964). A European Link with Chatal Huyuk: uncovering a 7th Millennium Settlement in Macedonia. Part I - Site and pottery (Illustrated London News), pp. 564–567.
- Rosenstock, E., Ebert, J., Martin, R., Hicketier, A., Walter, P., and Groß, M. (2019). Human stature in the Near East and Europe ca. 10,000–1000 BC: its



- spatiotemporal development in a Bayesian errors-in-variables model. *Archaeol. Anthropol. Sci.* **11**, 5657–5690.
- Runnels, C. (1988). A prehistoric survey of Thessaly: new light on the Greek middle paleolithic. *J. Field Archaeol.* **15**, 277–290.
- Runnels, C., and Özdoğan, M. (2001). The Palaeolithic of the Bosphorus region NW Turkey. *J. Field Archaeol.* **28**, 69–92.
- Scally, A., and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753.
- Scheu, A., Powell, A., Bollongino, R., Vigne, J.-D., Tresset, A., Çakırlar, C., Benecke, N., and Burger, J. (2015). The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet* **16**, 54.
- Schiffels, S., and Wang, K. (2020). MSMC and MSMC2: the multiple sequentially markovian coalescent. *Methods Mol. Biol.* **2090**, 147–166.
- Schoeninger, M.J., and DeNiro, M.J. (1982). Carbon isotope ratios of apatite from fossil bone cannot be used to reconstruct diets of animals. *Nature* **297**, 577–578.
- Shang, D., Zhang, X., Sun, M., Wei, Y., and Wen, Y. (2013). Strong association of the SNP rs17822931 with wet earwax and bromhidrosis in a Chinese family. *J. Genet.* **92**, 289–291.
- Shennan, S. (2018). *The first farmers of Europe: an evolutionary perspective* (Cambridge University Press).
- Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliusen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662.
- Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T., Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469.
- Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482.
- Skourtanioti, E., Erdal, Y.S., Frangipane, M., Balossi Restelli, F., Yener, K.A., Pinnock, F., Matthiae, P., Özbal, R., Schoop, U.-D., Guliyev, F., et al. (2020). Genomic history of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* **181**, 1158–1175.e28.
- Slatkin, M. (2005). Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Mol. Ecol.* **14**, 67–73.
- Smith. (1966). *Le Solutréen en France*. Publications de l'Institut de Préhistoire de l'Université de Bordeaux, Mém. n° 5 (Delmas), p. 449.
- Sommer, R.S., Zachos, F.E., Street, M., Jöris, O., Skog, A., and Benecke, N. (2008). Late Quaternary distribution dynamics and phylogeography of the red deer (*Cervus elaphus*) in Europe. *Quat. Sci. Rev.* **27**, 714–733.
- Speidel, L., Cassidy, L., Davies, R.W., Hellenthal, G., Skoglund, P., and Myers, S.R. (2021). Inferring population histories for ancient genomes using genome-wide genealogies. *Mol. Biol. Evol.* **38**, 3497–3511.
- Stadler, P. (2015). Versuch einer Auswertung der <sup>14</sup>C-Proben von Kleinhadersdorf mittels Bayes'scher Statistik. In *Mitteilungen der Prähistorischen Kommission 82*, D.L.G. von Kleinhadersdorf, C. Neugebauer-Maresch, and E. Lenneis, eds. (Verlag der Österreichischen Akademie der Wissenschaften), pp. 149–151.
- Stefanović, S. (2016). *People of the Lepenski Vir: Bioanthropological Analysis of Human Skeletal Remains* (Laboratory for Bioarchaeology, Faculty of Philosophy).
- Stefanović, S., and Borić, D. (2008). The newborn infant burials from Lepenski Vir: in pursuit of contextual meanings. In *The Iron Gates in Prehistory: New Perspectives*, BAR Int. Series 1893, C. Bonsall, V. Boroneanț, and I. Radovanović, eds. (Archaeopress), pp. 131–169.
- Tasić, N., Marić, M., Ramsey, C.B., Kromer, B., Barclay, A., Bayliss, A., Beavan, N., Gaydarska, B., and Whittle, A. (2015). Vinča-Belo Brdo, Serbia: the times of a tell. *Germanica* **93**, 1–75.
- Tasić, N., Marić, M., Ramsey, C.B., Kromer, B., Barclay, A., Bayliss, A., Beavan, N., Gaydarska, B., and Whittle, A. (2016). Vinča-Belo Brdo, Serbia: the times of a tell (Dataset) (Heidelberg Research Data Repository (University of Heidelberg)).
- Tasić, N., Srejić, D., and Stojanović, B. (1990). *Vinča: Centre of the Neolithic Culture of the Danubian Region* (project Rastko).
- Teschler-Nicola, M. (2012). The Early Neolithic site Asparn/Schletz (Lower Austria). In *Sticks, stones and broken bones: Neolithic violence in a European perspective*, R.J. Schulting and L. Fibiger, eds. (Oxford University Press), pp. 101–120.
- Tiefenböck, B., and Teschler-Nicola, M. (2015). In *Das Linearbandkeramische Gräberfeld von Kleinhadersdorf*. *Mitteilungen der Prähistorischen Kommission-Teil II: Anthropologie*, **82**, C. Neugebauer-Maresch and E. Lenneis, eds. (Verlag der Österreichischen Akademie der Wissenschaften), pp. 297–392.
- Tourloukis, V., and Harvati, K. (2018). The Palaeolithic record of Greece: a synthesis of the evidence and a research agenda for the future. *Quat. Int.* **466**, 48–65.
- Triantaphyllou, S. (2001). A bioarchaeological approach to prehistoric cemetery populations from western and central Greek Macedonia (Archaeopress).
- Turck, R. (2019). Where did the dead from Herxheim originate? Isotope analyses of human individuals from the find concentrations in the ditches, Direktion Landesarchäologie. In *Ritualised destruction in the early Neolithic – The Exceptional Site of Herxheim (palatinate, Germany)*, 2, A. Zeeb-Lanz, ed. (Generaldirektion Kulturelles Erbe), pp. 313–421.
- Vaiglova, P., Bogaard, A., Collins, M., Cavanagh, W., Mee, C., Renard, J., Lamb, A., Gardeisen, A., and Fraser, R. (2014). An integrated stable isotope study of plants and animals from Kouphovouno, southern Greece: a new look at Neolithic farming. *J. Archaeol. Sci.* **42**, 201–215.
- van Andel, T.H., and Runnels, C.N. (1995). The earliest farmers in Europe. *Antiquity* **69**, 481–500.
- Veeramah, K.R., Rott, A., Groß, M., van Dorp, L., López, S., Kirsanow, K., Sell, C., Blöcher, J., Wegmann, D., Link, V., et al. (2018). Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proc. Natl. Acad. Sci. USA* **115**, 3494–3499.
- Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delsler, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., et al. (2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* **365**, 173–176.
- Vigne, J.-D., Briois, F., Zazzo, A., Willcox, G., Cucchi, T., Thiébaud, S., Carrère, I., Franel, Y., Touquet, R., Martin, C., et al. (2012). First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proc. Natl. Acad. Sci. USA* **109**, 8445–8449.
- Wahlund, S. (1928). Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* **11**, 65–106.
- Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., and Kayser, M. (2013). The HlrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7**, 98–115.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164.
- Weiberg, E., Bevan, A., Kouli, K., Katsianis, M., Woodbridge, J., Bonnier, A., Engel, M., Finné, M., Fyfe, R., Maniatis, Y., et al. (2019). Long-term trends of land use and demography in Greece: a comparative study. *Holocene* **29**, 742–760.
- Weninger, B., Clare, L., Gerritsen, F., Horejs, B., Krauß, R., Linstädter, J., Özbal, R., and Rohling, E.J. (2014). Neolithisation of the Aegean and southeast Europe during the 6600–6000 calBC period of Rapid Climate Change. *Doc. Praehistorica* **41**, 1–31.
- Whittle, A., Bartosiewicz, L., Borić, D., Pettitt, P., and Richards, M. (2002). In the beginning: new radiocarbon dates for the Early Neolithic in Northern Serbia and South-East Hungary. *Antaeus* **25**, 63–117.



- Wild, E.M., Stadler, P., Häußler, A., Kutschera, W., Steier, P., Teschler-Nicola, M., Wahl, J., and Windl, H.J. (2004). Neolithic massacres: local skirmishes or General Warfare in Europe? *Radiocarbon* 46, 377–385.
- Windl, H.J. (1999). Makabres Ende einer Kultur? *Archäol. Dtschl.* 1, 54–57.
- Windl, H.J. (2009). Zur Stratigraphie der bandkeramischen Grabenwerke von Asparn an der Zaya-Schletz. In *Zum Ende der Bandkeramik in Mitteleuropa. Beiträge der Internationalen Tagung in Herxheim bei Landau (Pfalz) vom 14. Krisen., Kulturwandel., Kontinuitäten., and A. Zeeb-Lanz, eds. (verlag Marie Leidorf)*, pp. 191–196.
- Xue, Y., Zhang, X., Huang, N., Daly, A., Gillson, C.J., Macarthur, D.G., Yngvadottir, B., Nica, A.C., Woodwark, C., Chen, Y., et al. (2009). Population differentiation as an indicator of recent positive selection in humans: an empirical evaluation. *Genetics* 183, 1065–1077.
- Yaka, R., Mapelli, I., Kaptan, D., Doğu, A., Chyleński, M., Erdal, Ö.D., Koptekin, D., Vural, K.B., Bayliss, A., Mazzucato, C., et al. (2021). Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. *Curr. Biol.* 31, 2455–2468.e18.
- Yamaguchi, K., Watanabe, C., Kawaguchi, A., Sato, T., Naka, I., Shindo, M., Moromizato, K., Aoki, K., Ishida, H., and Kimura, R. (2012). Association of melanocortin 1 receptor gene (MC1R) polymorphisms with skin reflectance and freckles in Japanese. *J. Hum. Genet.* 57, 700–708.
- Yang, D.Y., Eng, B., Wayne, J.S., Dudar, J.C., and Saunders, S.R. (1998). Technical note: Improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* 105, 539–543.
- Zeder, M.A. (2017). Out of the Fertile Crescent: the dispersal of domestic livestock through Europe and Africa. In *Human Dispersal and Species Movement: From Prehistory to the Present*, N. Boivin, R. Crassard, and M. Petraglia, eds. (Cambridge University Press), pp. 261–303.
- Zeeb-anz, A., ed. (2016). Ritualised Destruction in the Early Neolithic – the Exceptional Site of Herxheim (Palatinate, Germany) Vol. 1. In *Forschungen zur Pfälzischen Archäologie*, 8.1 (Generaldirektion Kulturelles Erbe, Direktion Landesarchäologie).
- Zeeb-Lanz, A. (2019a). The Herxheim ritual enclosure – a synthesis of results and interpretative approaches. In *Ritualised destruction in the early Neolithic – the exceptional site of Herxheim (palatinate, Germany) Vol. 2*, A. Zeeb-Lanz, ed. (Generaldirektion Kulturelles Erbe, Direktion Landesarchäologie), pp. 423–482.
- Zeeb-anz, A., ed., (2019b). Ritualised Destruction in the Early Neolithic – the Exceptional Site of Herxheim (Palatinate, Germany), Vol. 2. In *Forschungen Zur Pfälzischen Archäologie 8.2* (Generaldirektion Kulturelles Erbe, Direktion Landesarchäologie).
- Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328.
- Zoledziewska, M., UK10K Consortium, Sidore, C., Chiang, C.W.K., Sanna, S., Mulas, A., Steri, M., Busonero, F., Marcus, J.H., Marongiu, M., et al. (2015). Height-reducing variants and selection for short stature in Sardinia. *Nat. Genet.* 47, 1352–1356.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Ancient human bone material	This study	AKT16; ERS10598167
Ancient human bone material	This study	Bar25; ERS10598177
Ancient human bone material	This study	Nea2; ERS10598179
Ancient human bone material	This study	Nea3; ERS10598178
Ancient human bone material	This study	VLASA32; ERS10598181
Ancient human bone material	This study	VLASA7; ERS10598180
Ancient human bone material	This study	LEPE48; ERS10598171
Ancient human bone material	This study	LEPE52; ERS10598168
Ancient human bone material	This study	STAR1; ERS10598170
Ancient human bone material	This study	VC3-2; ERS10598169
Ancient human bone material	This study	Asp6; ERS10598172
Ancient human bone material	This study	Klein7; ERS10598176
Ancient human bone material	This study	Ess7; ERS10598174
Ancient human bone material	This study	Dil16; ERS10598175
Ancient human bone material	This study	Herx; ERS10598173
<b>Chemicals, peptides, and recombinant proteins</b>		
AccuPrime™ Pfx SuperMix	Invitrogen	Cat#12344040
Bst Polymerase, Large Fragment (8 U/μl)	New England Biolabs GmbH	Cat#M0275S
dNTPs (each 10 mM)	QIAGEN, Hilden, Germany	Cat#201901
EDTA (0.5 M, pH 8.0)	Ambion/Applied Biosystems, Life Technologies	Cat#AM9262
Lauroylsarcosine, Sodium Salt	Merck Millipore, Merck KGaA, Darmstadt, Germany	Cat#428010
NEBNext End Repair Enzyme Mix	New England Biolabs GmbH	Cat#E6050L
NEBNext End Repair Reaction Buffer (10X)	New England Biolabs GmbH	Cat#E6050L
Nuclease-free H <sub>2</sub> O	Life Technologies	Cat#AM9932
PEG-4000	Thermo Scientific	Cat#EL0011
Proteinase K	Roche Diagnostics, Mannheim, Germany	Cat#3115828001
T4 DNA Ligase (5 U/μl)	Thermo Scientific	Cat#EL0011
T4 DNA Ligase Buffer (10X)	Thermo Scientific	Cat#EL0011
ThermoPol Buffer (10X)	New England Biolabs GmbH	Cat#M0275S
Tris-EDTA	Sigma-Aldrich	Cat#T9285
Tris-HCl (1M, pH 8.0)	Life Technologies	Cat#15568025
USER™ enzyme	New England Biolabs GmbH	Cat#M5505L
<b>Critical commercial assays</b>		
Agilent 2100 Expert Bioanalyzer System and High Sensitivity DNA Analysis Kit	Agilent Technologies	Cat#5067-4626 (kit)
Qubit Fluorometric quantitation and dsDNA HS Assay Kit	Invitrogen	Cat#Q32854 (kit); Cat#Q32856 (tubes)
<b>Deposited data</b>		
Sequencing data and aligned BAMfiles	This study	ENA: PRJEB50857
Code and input-files connected to this study	This study	<a href="https://doi.org/10.5281/zenodo.6367517">https://doi.org/10.5281/zenodo.6367517</a> for the release of <a href="https://github.com/CMPG/originsEarlyFarmers/">https://github.com/CMPG/originsEarlyFarmers/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ancient genome - Demultiplexed FASTQ files kindly provided by the authors (contact: Zuzana Hofmanová, co-author of this study)	(Hofmanová et al., 2016)	Bar8
Ancient genome - Demultiplexed FASTQ files	(Jones et al., 2015)	Bichon; ERR1078331 - ERR1078351
Ancient genome - FASTQ files (produced from BAMfiles by ENA)	(Kılınç et al., 2016)	Bon002; ERR1514027
Ancient genome - Demultiplexed FASTQ files kindly provided by the authors (contact: Kay Prüfer)	(Lazaridis et al., 2014)	Loschbour
Ancient genome - Raw sequencing FASTQ file kindly provided by the authors (contact: Lara Cassidy)	(Gamba et al., 2014)	NE1
Ancient genome - Aligned BAM file	(Günther et al., 2018)	SF12; ERR2060277
Ancient genome - Demultiplexed FASTQ files kindly provided by the authors (contact: Kay Prüfer)	(Lazaridis et al., 2014)	Stuttgart
Ancient genome - Demultiplexed FASTQ file kindly provided by the authors (contact: Yoan Diekmann, co-author of this study)	(Brace et al., 2019)	CarsPas1
Ancient genome - Demultiplexed FASTQ files kindly provided by the authors (contact: Jens Blöcher, co-author of this study)	(Broushaki et al., 2016)	WC1
Ancient genome - Demultiplexed FASTQ files	(Jones et al., 2015)	KK1; ERR1078321-ERR1078325
Modern genomes - Aligned BAMfiles from 77 modern individuals	The Simons Genome Diversity Project (SGDP) (Mallick et al., 2016)	<a href="https://www.simonsfoundation.org/simons-genome-diversity-project/Individual%20identifiers">https://www.simonsfoundation.org/simons-genome-diversity-project/Individual identifiers</a> , see Table S2.
1000 Genomes phase3 per chromosome VCFs	(1000 Genomes Project Consortium et al., 2015)	<a href="http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/">http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/</a>
Allen Ancient DNA Resource v42.4 and v37.2	Reich lab public data release	<a href="https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/index_v42.4.html">https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/index_v42.4.html</a> ; <a href="https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data">https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data</a>
Assembly gaps and centromeres	UCSC genome browser	<a href="http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/gap.txt.gz">http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/gap.txt.gz</a>
Chimpanzee hg19 nucleotide states	UCSC genome browser	<a href="http://hgdownload.cse.ucsc.edu/goldenpath/hg19/vsPanTro4/axtNet/">http://hgdownload.cse.ucsc.edu/goldenpath/hg19/vsPanTro4/axtNet/</a>
Chimpanzee Reference Genome (panTro4)	The Chimpanzee Sequencing and Analysis Consortium (2005)	GenBank assembly accession: GCA_000001515.4
CpG islands list	UCSC genome browser	(CpG Islands (cpgIslandExt) Track)
Ensembl Compara 71 genome FASTA files	Ensembl Compara	<a href="http://ftp.ensembl.org/pub/release-71/fasta/ancestral_alleles/homo_sapiens_ancestor_GRCh37_e71.tar.bz2">http://ftp.ensembl.org/pub/release-71/fasta/ancestral_alleles/homo_sapiens_ancestor_GRCh37_e71.tar.bz2</a>
Gorilla hg19 nucleotide states	UCSC genome browser	<a href="http://hgdownload.cse.ucsc.edu/goldenpath/hg19/vsGorGor5/axtNet/">http://hgdownload.cse.ucsc.edu/goldenpath/hg19/vsGorGor5/axtNet/</a>
Haplotype Reference Consortium dataset	(McCarthy et al., 2016)	accession number EGAD00001002729 on the European Genome-phenome Archive
HapMap file for chrX (ANGSD) HapMapChrX.gz	(Rasmussen et al., 2011)	<a href="http://www.popgen.dk/angsd/index.php/ANGSD">http://www.popgen.dk/angsd/index.php/ANGSD</a>
HapMap phase II b37 genetic map	N/A	<a href="https://github.com/odelaneau/shapeit4/tree/master/maps">https://github.com/odelaneau/shapeit4/tree/master/maps</a>
Human reference sequence hs37d5	(1000 Genomes Project Consortium et al., 2015)	<a href="ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz">ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Known InDel positions	(1000 Genomes Project Consortium et al., 2015)	<a href="ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/b37/1000G_phase1.indels.b37.vcf.gz">ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/b37/1000G_phase1.indels.b37.vcf.gz</a>
Known InDel positions	(1000 Genomes Project Consortium et al., 2015)	<a href="ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/b37/Mills_and_1000G_gold_standard.indels.b37.vcf.gz">ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/b37/Mills_and_1000G_gold_standard.indels.b37.vcf.gz</a>
Per chromosome mappability mask for human reference genome hs37d5	The Simons Genome Diversity Project (SGDP) (Mallick et al., 2016)	<a href="https://github.com/wangke16/MSMC-IM/blob/master/masks/um75-hs37d5.bed.gz">https://github.com/wangke16/MSMC-IM/blob/master/masks/um75-hs37d5.bed.gz</a>
Recombination Map for YRI population	(1000 Genomes Project Consortium et al., 2015)	<a href="ftp://1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_rates/">ftp://1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_rates/</a>
Ultraconserved sites for recalibration step	(Dimitrieva and Bucher, 2013)	<a href="https://ccg.epfl.ch/UCNEbase/data/download/ucnes/hg19_UCNE_coord.bed">https://ccg.epfl.ch/UCNEbase/data/download/ucnes/hg19_UCNE_coord.bed</a>

**Oligonucleotides**

P5 and P7 adapters	(Meyer and Kircher, 2010) IDT, Leuven, Belgium	N/A
P5 and P7 indexing primers with index sequences (8 bp) from the NexteraXT index Kit v2	(Kircher et al., 2012) and Illumina, San Diego, California, United States IDT, Leuven, Belgium	N/A

**Software and algorithms**

ADMIXTOOLS package v7300	(Patterson et al., 2012)	<a href="https://github.com/DReichLab/AdmixTools">https://github.com/DReichLab/AdmixTools</a>
ANGSD - version 0.917	(Rasmussen et al., 2011)	<a href="http://www.popgen.dk/angsd/index.php/ANGSD">http://www.popgen.dk/angsd/index.php/ANGSD</a>
ANNOVAR	(Wang et al., 2010)	<a href="https://annovar.openbioinformatics.org">https://annovar.openbioinformatics.org</a>
ATLAS - commits 6bd2482 & 7cfc900	(Link et al., 2017)	<a href="https://bitbucket.org/wegmannlab/atlas/">https://bitbucket.org/wegmannlab/atlas/</a>
ATLAS-Pipeline, commit 6df90e7	Wegmann lab, Ilektra Schulz	<a href="https://bitbucket.org/wegmannlab/atlas-pipeline">bitbucket.org/wegmannlab/atlas-pipeline</a>
bcftools versions: 1.9 and 0.1.15	(Danecek et al., 2021)	<a href="https://samtools.github.io/bcftools/howtos/index.html">https://samtools.github.io/bcftools/howtos/index.html</a>
bwa - Burrows-Wheeler Alignment Tool - versions 0.7.15 and 0.7.17	(Li, 2013)	<a href="http://bio-bwa.sourceforge.net">bio-bwa.sourceforge.net</a>
BEDOPS v2.4.40	(Neph et al., 2012)	<a href="https://bedops.readthedocs.io/en/latest/">https://bedops.readthedocs.io/en/latest/</a>
Bedtools 2.25.0	(Quinlan and Hall, 2010)	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
ContamMix - version 1.0	(Fu et al., 2013)	<a href="https://science.umd.edu/biology/plfj/">https://science.umd.edu/biology/plfj/</a>
dadi	(Gutenkunst et al., 2009)	<a href="https://bitbucket.org/gutenkunstlab/dadi">https://bitbucket.org/gutenkunstlab/dadi</a>
fastsimcoal2.7	(Excoffier et al., 2013, 2021)	<a href="http://cmpg.unibe.ch/software/fastsimcoal2/">http://cmpg.unibe.ch/software/fastsimcoal2/</a>
fastqc - version 0.11.5	Babraham Bioinformatics	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
GATK - version 3.7	(DePristo et al., 2011)	<a href="https://gatk.broadinstitute.org">https://gatk.broadinstitute.org</a>
HirisPlex-S webtool	(Chaitanya et al., 2018; Walsh et al., 2013)	<a href="https://hirisplex.erasmusmc.nl/">https://hirisplex.erasmusmc.nl/</a>
IBDSeq v. r1206	(Browning and Browning, 2013)	<a href="http://faculty.washington.edu/browning/ibdseq.html">http://faculty.washington.edu/browning/ibdseq.html</a>
LEA R package v.2.6.0	(Frichot and François, 2015)	<a href="https://bioconductor.org/packages/release/bioc/html/LEA.html">https://bioconductor.org/packages/release/bioc/html/LEA.html</a>
mafft - version 7.31	(Katoh et al., 2002)	<a href="https://mafft.cbrc.jp/alignment/software/linux.html">https://mafft.cbrc.jp/alignment/software/linux.html</a>
MIA (Mapping Iterative Assembler) - version 1.0	MPI EVA Bioinformatics	<a href="https://github.com/mpieva/mapping-iterative-assembler">https://github.com/mpieva/mapping-iterative-assembler</a>
MSMC2	(Schiffels and Wang, 2020)	<a href="https://github.com/stschiff/msmc2">https://github.com/stschiff/msmc2</a>
MSMC-tools, commit 07bc8a9	(Schiffels and Wang, 2020)	<a href="https://github.com/stschiff/msmc-tools">https://github.com/stschiff/msmc-tools</a>
phy-mer	(Navarro-Gomez et al., 2015)	<a href="https://github.com/MEEIBioinformaticsCenter/phy-mer/">https://github.com/MEEIBioinformaticsCenter/phy-mer/</a>
Picard-tools - version 2.9	Broad Institute	<a href="http://broadinstitute.github.io/picard/">http://broadinstitute.github.io/picard/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R - version 4.0, 3.7, and 3.6.1	R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria	<a href="https://www.R-project.org/">https://www.R-project.org/</a>
SAMtools - version 1.3	(Li et al., 2009)	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
Samtools 1.9	(Danecek et al., 2021)	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
seqtk - version 1.2	N/A	<a href="https://github.com/lh3/seqtk">https://github.com/lh3/seqtk</a>
SHAPEIT4 v1.2	(Delaneau et al., 2019)	<a href="https://odelaneau.github.io/shapeit4/">https://odelaneau.github.io/shapeit4/</a>
Snakemake - version 4.0	(Köster and Rahmann, 2012)	<a href="https://snakemake.readthedocs.io/en/stable/">https://snakemake.readthedocs.io/en/stable/</a>
Trim Galore! - version 0.4.3	Babraham Bioinformatics	<a href="http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a>
Yjasc_3752_ry_compute.py, version 0.4	(Skoglund et al., 2013)	<a href="https://ars.els-cdn.com/content/image/1-s2.0-S0305440313002495-mmc1.zip">https://ars.els-cdn.com/content/image/1-s2.0-S0305440313002495-mmc1.zip</a>
Yleaf	(Ralf et al., 2018)	<a href="https://github.com/genid/Yleaf">https://github.com/genid/Yleaf</a>
<b>Other</b>		
Agencourt® AMPure® XP beads	Beckmann Coulter	Cat#A63880
Amicon Ultra-4 Centrifugal Filter Units, 30kDa	Merck Millipore, Darmstadt, Germany	Cat#UFC803096
Hydroxyapatite	Sigma-Aldrich	Cat#21223
MinElute PCR Purification Kit	QIAGEN, Hilden, Germany	Cat#28006
MSB® Spin PCRapace	Invitex Molecular GmbH, Berlin, Germany	Cat#1020220400
Spezial-Edelkorund (EW60/250µ)	Harnisch+Rieth, Winterbach, Germany	Cat#75250
Spezial-Edelkorund (30B/50µ)	Harnisch+Rieth, Winterbach, Germany	Cat#75308

**RESOURCE AVAILABILITY**

**Lead contact**

- Further information and requests should be directed to and will be fulfilled by the lead contact Laurent Excoffier ([laurent.excoffier@iee.unibe.ch](mailto:laurent.excoffier@iee.unibe.ch)).

**Materials availability**

- This study did not generate new unique reagents.

**Data and code availability**

- Raw sequencing data (FASTQ-files) and aligned BAM-files generated in this study have been deposited to European Nucleotide Archive (ENA: PRJEB50857) and are publicly available as of the date of publication. Individual Accession numbers are listed in the key resources table. Filtered VCF-files have been deposited to European Variant Archive (EVA: PRJEB51919) and are publicly available as of the date of publication. Further archaeological information and analyses additional to the present article are available ([Data S1](#); [Methods S1](#)), as well as Supplemental Tables and Figures.
- All original code has been deposited at <https://github.com/CMPG/originsEarlyFarmers> and is publicly available as of the date of publication. under DOI <https://doi.org/10.5281/zenodo.6367517>.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

In this study, we present new whole-genome sequences for 15 ancient human individuals (see [Table 1](#), Figure D1\_1 from [Data S1](#), and [key resources table](#) for an overview). Our sample set consists of two Mesolithic individuals from Vlasac (Serbia), two individuals from Transitional and Neolithic layers at Lepenski Vir (Serbia) and 11 early Neolithic individuals originating from Aktopraklık and



Barçın in Turkey (one individual each), from Nea Nikomedeia in Greece (two individuals), from Vinča-Belo Brdo and Grad-Starčevo in Serbia (one individual each), from Kleinhadersdorf and Asparn-Schletz in Austria (one individual each) and Essenbach-Ammerbreite, Dillingen-Steinheim and Herxheim in Germany (one individual each). For detailed information on the archaeological background, please refer to the [Data S1](#). Age estimated from anthropological analyses and genetic sex of the individuals are reported in [Table 1](#) and [Data S1](#).

The ancient skeletal material used for this study was collected and analysed in collaboration with the anthropologists/archaeologists in charge of the archaeological sites, in accordance with the laws of the respective countries. Specifically, for the Greek samples, we had permission by the Greek Ministry of Culture and Sports for the sampling, DNA extraction, and radiocarbon dating according to Greek law for destructive sampling of archaeological material (N.3028/02).

## METHOD DETAILS

### Data generation

#### Sample preparation

All laboratory analyses were conducted in the dedicated ancient DNA facilities of the Palaeogenetics Group Mainz (Institute of Organismic and Molecular Evolution, Johannes Gutenberg University, Mainz), according to strict ancient DNA protocols to prevent contamination with modern DNA as well as cross contamination between samples as previously described ([Bollongino et al., 2013](#); [Bramanti et al., 2009](#); [Scheu et al., 2015](#)). These ancient DNA standards include decontamination of workspace, labware and samples, independent DNA extractions and the processing of blank controls during sample pulverisation, DNA extraction, library preparation and PCR reactions to monitor contaminations.

All steps prior to PCR amplification (sample preparation, DNA extraction and library preparation) were carried out in the dedicated ancient DNA facilities of the Palaeogenetics Group, which are separated from post-PCR areas.

The petrous bone samples were prepared as described in [Hofmanová et al. \(2016\)](#). In detail, after documentation, samples were sterilised under ultraviolet light (254 nm) from 2 sides for 45 minutes per side. In order to remove superficial contaminants, soil remains and the outer bone surface were removed using a sandblasting machine (P-G 400, Harnisch+Rieth) with Spezial-Edelkorund (EW60/250 $\mu$  and 30B/50 $\mu$ ; Harnisch+Rieth). The densest, inner part of the petrous bone was then isolated and cut in cubes using a disk saw (Electer E<sub>max</sub> IH-300, MAFRA). After irradiation with ultraviolet light (254 nm) from two sides, for 45 minutes per side, the densest bone cubes were pulverised using a mixer mill (MM200, Retsch). To control contamination, blank milling controls containing hydroxyapatite were processed in parallel.

#### Radiocarbon dating and stable isotope analysis

New radiocarbon dates and stable isotope values (carbon  $\delta^{13}\text{C}$  and nitrogen  $\delta^{15}\text{N}$ ) were obtained for seven individuals (Table D1\_2). The analyses were performed at the Curt-Engelhorn-Zentrum Archäometrie gGmbH (Mannheim, Germany). The collagen used for the analyses was extracted from fragments of the petrous bones used for palaeogenomic analysis.

All dates were uniformly re-calibrated in OxCal 4.4.2 ([Bronk Ramsey, 2009](#)) using the IntCal20 calibration curve ([Reimer et al., 2020](#)). The isotope results were used as a basis for palaeodietary inference (Figure D1\_11) and for freshwater reservoir effect correction in the case of the Vlasac sample (Table D1\_2). Radiocarbon dates on human bones from archaeological sites in the Danube Gorges are known to be influenced by a large freshwater  $^{14}\text{C}$  reservoir effect, due to high intake of freshwater food, and need to be corrected before calibration ([Cook et al., 2001](#)). We did it following the method described by [Cook et al. \(2001\)](#), later amended by [Bonsall et al. \(2015\)](#): for human bones with  $\delta^{15}\text{N} \geq 8.3\text{‰}$ , uncalibrated dates of the form  $\mu \pm \sigma$  were adjusted to  $(\mu - 545c) \pm$

$$\sqrt{\sigma^2 + (70c)^2} \text{ where } c = \frac{\delta^{15}\text{N} - 8.3}{17 - 8.3}.$$

#### DNA extraction

DNA extraction followed the protocol by [Yang et al. \(1998\)](#) with the modifications described in [MacHugh et al. \(2000\)](#) and [Gamba et al. \(2014\)](#) as well as additional modifications described below. For each sample, 0.15–0.31 g of bone powder was used for extraction.

Prior to extraction, a pre-lysis was performed by adding 1 mL of EDTA (0.5 M, pH8) to the bone powder and incubating at room temperature for 10 minutes. The solution was centrifuged at maximum speed to pellet the powder and the supernatant was removed.

Lysis was performed on rocking shakers at 37°C for 24 hours (900 rpm) using 1 ml of extraction buffer containing EDTA (950  $\mu\text{l}$ , 0.5 M, pH8), Tris-HCl (20  $\mu\text{l}$ , 1 M, pH8), N-Lauroylsarcosine (17  $\mu\text{l}$ , 5%) and Proteinase K (13  $\mu\text{l}$ ; 20 mg/ml). After 24 hours of incubation, the samples were centrifuged for 10 minutes at 10,000 rpm, the supernatant was removed, transferred into new tubes and stored in a fridge until further processing. A second lysis step was performed following the same procedure.

Following lysis, the supernatants from the two lysis steps were merged on an Amicon Filter (Amicon Ultra-4 30 kDa, 15 ml) and centrifuged for 10 min at 2500 rpm. The DNA was then washed twice with 3 ml 1X Tris-EDTA, followed by centrifugation at 2500 rpm for 20 minutes and discarding of the flow-through in between. After washing, the extract was concentrated to 100  $\mu\text{l}$  and subsequently purified with the MinElute PCR Purification Kit following the manufacturer's instructions but incubating for 5 minutes during elution with 44  $\mu\text{l}$  elution buffer (preheated to 65°C).

Blank controls were processed during DNA extraction and incorporated into all further steps of the analyses.

#### Library preparation

Double-indexed libraries were prepared according to the protocol by [Kircher et al. \(2012\)](#) with slight modifications.

Prior to library preparation the DNA extracts were treated with USER<sup>TM</sup> enzyme: 5 µl of USER<sup>TM</sup> enzyme was added to 16.25 µl of DNA extract (exception: 17 µl of two merged extracts were used for library SL3\_5 of sample AKT16, see [Table S1](#)) and the mixture was incubated for three hours at 37°C ([Verdugo et al., 2019](#)).

The blunt-end repair step followed immediately and was performed using the NEBNext End Repair Module: the DNA extract was mixed with NEBNext End Repair Reaction Buffer (10X, 7 µl), NEBNext End Repair Enzyme Mix (3.5 µl) and nuclease-free water (38.25 µl; for a final reaction volume of 70 µl) and incubated for 15 minutes at 25°C followed by 5 minutes at 12°C. Deviating from this procedure, two libraries (SL2.12\_MU and SL3.12\_MU of sample VC3-2) were produced using 20 µl DNA extract and without USER<sup>TM</sup> enzyme treatment of the DNA extract (see [Table S1](#)). In the adapter ligation step hybridised adapters P5 and P7 ([Meyer and Kircher, 2010](#)) were used at a concentration of 0.75 µM. 3 µl of Fill-In product (total volume: 40 µl) were amplified using the AccuPrime<sup>TM</sup> Pfx SuperMix (20 µl) in one PCR parallel adding unique and sample-specific index combinations to the library molecules (final reaction volume: 25 µl; final primer concentration: 200 nM or 160 nM each). Double indexing followed [Kircher et al. \(2012\)](#), but using index sequences from the NexteraXT index Kit v2 (Illumina; barcode length 8 bp; ordered at IDT, Leuven, Belgium). The PCR was performed in 9-13 cycles; the PCR temperature profile followed the manufacturer's recommendations but using an annealing temperature of 60°C, extending for 30 seconds during each cycle and performing a final elongation step for 5 minutes.

For purification during library preparation the MinElute PCR Purification Kit was used, while amplified libraries were purified with the MSB® Spin PCRapace. Libraries were quantified by Qubit Fluorometric quantitation (dsDNA HS assay) and measurement on the Agilent 2100 Bioanalyser System (High Sensitivity DNA Analysis).

Blank controls as well as positive controls (nonsense hybrids) of known concentration, were processed in every library step including PCR amplification to verify the success of the library preparation and to monitor contamination. Quantification of blank controls did not indicate significant contamination during any laboratory step (pulverisation, DNA extraction, library preparation and amplification).

### Sample Screening

Using shallow shotgun sequencing on an Illumina MiSeq<sup>TM</sup> platform at StarSEQ GmbH (Mainz, Germany), all samples were screened for their endogenous DNA preservation. Libraries were equimolarly pooled, subsequently purified with Agencourt® AMPure® XP beads and sequenced in single-end runs with 50 bp read length. Demultiplexing was performed by the sequencing facility using the MiSeq Reporter and allowing one mismatch in the barcode. Blank controls from DNA extraction, library preparation and PCR-steps were sequenced alongside to estimate the fraction of potentially contaminating reads introduced in the lab (for details see [Table S1](#)).

### Whole-genome sequencing

For deeper shotgun sequencing, 2-5 DNA extracts and 3-9 libraries of each sample were prepared (as detailed above). The libraries were amplified in up to 13 PCR parallels, each with an individual index combination, to increase the complexity of the libraries. Subsequently, PCR parallels were purified and quantified individually as described above. For sequencing, PCR parallels were pooled equimolarly according to their concentrations measured on Qubit® and purified with Agencourt® AMPure® XP beads. The samples were sequenced on an Illumina HiSeq3000 (SE, 100 cycles or PE, 150 cycles) at the Next Generation Sequencing Platform at the University of Berne, Switzerland. A summary and details of the laboratory work and the sequencing strategy for each sample can be found in [Table S1](#).

### Bioinformatics pipeline

The bioinformatic steps within this section were conducted with the following programs and versions unless specified otherwise:

- ANGSD - version 0.917 ([Rasmussen et al., 2011](#))
- ATLAS - version 1.0, commit 6bd2482 ([Link et al., 2017](#))
- ATLAS-Pipeline, commit 6df90e7 ([bitbucket.org/wegmannlab/atlas-pipeline](http://bitbucket.org/wegmannlab/atlas-pipeline))
- BWA - Burrows-Wheeler Alignment Tool - version 0.7.15 ([Li, 2013](#))
- ContamMix - version 1.0 ([Fu et al., 2013](#))
- fastqc - version 0.11.5 ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/))
- GATK - version 3.7 ([DePristo et al., 2011](#))
- mafft - version 7.31 ([Katoh et al., 2002](#))
- MIA (Mapping Iterative Assembler) - version 1.0 (<https://github.com/mpieva/mapping-iterative-assembler>)
- Picard-tools - version 2.9 (<http://broadinstitute.github.io/picard/>)
- SAMtools - version 1.3 ([Li et al., 2009](#))
- seqtk - version 1.2 (<https://github.com/lh3/seqtk>)
- Snakemake - version 4.0 ([Köster and Rahmann, 2012](#))
- Trim Galore! - version 0.4.3 (<https://github.com/FelixKrueger/TrimGalore>)

### Raw data handling

We committed to process all 102 samples (15 ancient genomes from this study, 10 previously published ancient genomes from [Brace et al., 2019](#); [Broushaki et al., 2016](#); [Gamba et al., 2014](#); [Günther et al., 2018](#); [Hofmanová et al., 2016](#); [Jones et al., 2015](#); [Kılınc et al., 2016](#); [Lazaridis et al., 2016](#), and 77 modern SGDP genomes from [Mallick et al., 2016](#); see [Table D1\\_1](#) and [Table S2](#) for the complete

list) with the same bioinformatic pipeline, albeit minor changes due to variation in how the raw data were obtained. We therefore first describe the pipeline used for the 15 samples sequenced in this study, followed by a description of the differences in how previously published data were analysed:

The sequencing quality was checked with *fastqc*. Raw reads were trimmed using *Trim Galore* with no quality filter and a length filter of 30 bp (`-q0, -length 30, -a 'AGATCGGAAGAGCACACGTCTGAACTCC'`). For paired-end libraries, the additional option `-retain unpaired` was used and the second adapter sequence (`-a2 'AGATCGGAAGAGCGTCGTGTAGGGAAAG'`) was provided. A second *fastqc* analysis was performed to verify trimming and the quality of the remaining sequences. Reads were then aligned to the 1000 Genomes project version of the human reference genome *hs37d5* (1000 Genomes Project Consortium et al., 2015) using *bwa -mem* with options `-t 8 -M`. Reads with a mapping quality below 30 were filtered out with *SAMtools -q30*. Duplicate reads were marked but not removed using *picard-tools MarkDuplicates* with `VALIDATION_STRINGENCY=SILENT` and `AS=TRUE`. Unmapped reads were removed with *SAMtools* with option `-F4`. Informative read groups were added to keep track of individual PCR-parallels with *picard-tools AddOrReplaceReadGroups*. All necessary sorting and indexing in-between the steps was performed with *SAMtools index* and *sort*, library-files from the same samples were merged using *SAMtools merge*.

### Local Realignment

Local indel-realignment is an important but computationally expensive step in our pipeline. In order to allow for parallelization, we proceeded with the following three steps:

- (i) We identified potential target intervals using *GATK RealignerTargetCreator*, on a set of ten ancient and ten modern samples (Asp6, Bar25, Dil16, Ess7, Klein7, LEPE48, Nea2, Nea3, STAR1, VC3-2, VLASA32, VLASA7, French-1, Sardinian-1, Russian-1, Spanish-1, Georgian-1, Hungarian-1, Icelandic-1, Finnish-1, English-1, Greek-1, Mende-1, Estonian-1, Polish-1), while providing known indel sites identified by the 1000 Genomes project (see [Key Resources table](#)). We refer to the obtained target set as *TargetsBase*.
- (ii) Using *ATLAS task=downsample*, we created a set of reads by downsampling the following 5 ancient and 5 modern samples to a depth of 4X: Bar25, Klein7, STAR1, VLASA32, VLASA7, French-1, Georgian-1, Finnish-1, Greek-1, Polish-1. We refer to this set as the *GuidanceSet*.
- (iii) We run local indel-realignment for each sample individually as follows: we identified private target positions for the sample using *GATK RealignerTargetCreator* while providing the known indel sites from the 1000 Genomes project (as above). We then unified the identified targets with *TargetsBase* to obtain a sample-specific target set. Finally, we run *GATK IndelRealigner* on this sample together with the *GuidanceSet* on the sample-specific targets. We parallelized this step additionally by contig, merging the output with *picard-tools*.

### Filter and Sample Statistics

To reduce spurious alignments, we used *picard-tools FilterSamReads* to filter out reads that contained soft-clipped bases previously identified with *ATLAS task=assessSoftClipping* or did not pass *picard-tools ValidateSamFile*. We further used *SAMtools* to keep only primary alignments and, for paired-end libraries, proper pairs.

We determined read counts, sequencing depth, endogenous DNA-content and other statistics using *ATLAS task=BAMDiagnos-tics* and *SAMtools flagstat*. These statistics are given per library-parallel and for the merged data for all 15 ancient samples presented here in [Table S1](#).

### ATLAS Pipeline

We used commit 6df90e7 of the *snakemake* pipeline *ATLAS-Pipeline*, which is available at [bitbucket.org/wegmannlab/atlas-pipeline](https://bitbucket.org/wegmannlab/atlas-pipeline) and summarised below. Configfiles used are made available at <https://github.com/CMPG/originsEarlyFarmers>.

Since Post-mortem-Damage (PMD) is most prevalent at read ends, the PMD pattern is different if a read spans the entire fragment or not. Following [Kousathanas et al. \(2017\)](#), we split single-end read groups by length using *ATLAS task=splitRGbyLength* with the option `allowForLarger`. We identified reads that span the entire fragment as those shorter than the maximum read length minus 5 bases to account for variation introduced by adapter-trimming. We then identified PMD patterns for each read group using *ATLAS task=estimatePMD*, providing the reference genome with `fasta=hs37d5.fasta` and limiting the analysis to the chromosomes 1-22, X and Y with option `chr`. PMD estimates are shown in Figure M1\_1A. We performed base-quality recalibration as described in [Kousathanas et al. \(2017\)](#), using *ATLAS task=recal* on known ultraconserved sites obtained from [https://ccg.epfl.ch/UCNEbase/data/download/ucnes/hg19\\_UCNE\\_coord.bed](https://ccg.epfl.ch/UCNEbase/data/download/ucnes/hg19_UCNE_coord.bed) ([Dimitrieva and Bucher, 2013](#)). We assumed equal base frequencies (`equalBaseFreq`) and inferred recalibration parameters for each PCR-parallel (`pmdFile`) from all sites with a minimum depth of 2 (`minDepth=2`). Sequencing- and PCR-duplicates were not excluded from the estimation of recalibration parameters (`keepDuplicates`) to add additional power. A quality transformation by base quality recalibration was created with *ATLAS task=qualityTransformation* (Figure M1\_2).

For paired-end sequenced read-groups, the estimated recalibration parameters were directly corrected in the BAM file using *ATLAS task=recalBAM*, providing the recal-parameters (`recal`) and the reference (`fasta=hs37d5.fasta`). Paired-end reads with corrected base qualities were then physically merged with *ATLAS task=mergeReads* to avoid overlapping bases to be counted twice in downstream analysis. Reads that did not pass *picard-tools ValidateSamFile* were omitted. As a result of the physical merging, relative positions of bases within a read have changed, and we thus re-estimated PMD patterns as outlined above.

We called Bayesian genotypes in the reference genome with *ATLAS task=callNew* and parameters *method=Bayesian prior=theta fixedTheta=0.001 infoFields=DP equalBaseFreq*. The reference genome, PMD patterns and recal-patterns (for single-end read groups) were provided with *fasta*, *pmdFile* and *recal*, respectively. The first and last 2 bp of each read were ignored (*trim5=2*, *trim3=2*). We also created recalibrated BAM files from single-end libraries, which we used in any subsequent analysis in which *ATLAS* was not involved (e.g. contamination estimation).

The molecular sex was inferred using a script from (Skoglund et al., 2013) version 0.4, based on the ratio of reads aligning to the X and Y chromosomes (Table 1; Figure M1\_1B).

### Contamination estimation

Blank controls of extraction, library-preparation and index-PCR were analysed alongside the screening process. The concentration of potential contaminants was never higher than 0.81 ng/ $\mu$ l. A *Bioanalyser* analysis (HS DNA, Agilent Technologies, Waldbronn, Germany) and the screening results for extraction- and library-controls confirm the detected DNA to consist of primer- and adapter-dimers with a maximum of 55 aligning reads per blank control (out of a potential share of 200,000 reads; see Table S1).

We used the *ContamMix* R-script *estimate.R* to estimate the amount of authentically mapping mitochondrial reads at 311 modern diagnostic marker positions. We run *ContamMix* with the option *-baseq20* and the following two input-files that we generated from all reads in the recalibrated BAM file that mapped against the MT-Genome: i) An alignment of mitochondrial reads against their own consensus (*-samFn*). This consensus was constructed by iteratively mapping the reads with *MIA (mia)* and mapping the last iteration with *MIA (mia)* to obtain one FASTA-sequence. The MT-reads were mapped against their consensus using *bwa aln* and *samse (v.0.7.17)* and filtered for a minimum mapping quality of 30 with *SAMtools -q30*. ii) A multiple alignment of the consensus genome and a fasta-file containing the diagnostic marker positions against each other (*-malnFn*) obtained with *mafft*.

For male individuals, contamination was additionally estimated using the *ANGSD contamination* script based on haploid X-chromosomal regions (X:5000000-154900000). We run *ANGSD* with the options *-doCounts 1* and *-iCounts 1*, followed by the executable *misc/contamination*, providing the publicly available *HapMap* file *HapMapChrx.gz*.

The contamination estimates obtained with both methods (*ANGSD* and *ContamMix*) are shown in Figure M1\_1C and Table S1.

### Reference samples

To minimise reference biases (Günther and Nettelblad, 2019), we analysed all previously published samples with the same pipeline as outlined above. However, some adaptations were necessary for some samples:

- Blank characters in read names (e.g. from ENA accession numbers) were removed to ensure the proper detection of optical duplicates.
- Quality scores of Bichon FASTQ files were converted from illumina 1.5 to illumina 1.9 with *setqk seq -V 64*.
- Stuttgart and Loschbour were received as unaligned and untrimmed demultiplexed raw data in BAM file format and transformed to unaligned FASTQ format using *picard-tools SamToFastq*.
- For Loschbour files, individual library information was not recoverable. Therefore, duplicates across all sequencing files were marked.
- Raw sequencing data was not obtainable for Bon002 and SF12. For Bon002, we downloaded the FASTQ files that were produced from BAM files by ENA. For SF12, we used published BAM files with already physically merged reads and converted them back to FASTQ files using *picard-tools SamToFastq*. For both, we split the FASTQ files by run identifiers to treat each run separately throughout the pipeline. Since those reads were already pre-processed, we omitted adapter trimming, removed orphaned reads and otherwise followed the pipeline for single-end samples without splitting read groups by length.
- For NE1, one of the FASTQ files was corrupted and only 25% of the contained reads could be recovered. As a result, we could only use 97% of the total number of reads generated for that sample (Gamba et al., 2014). We then demultiplexed the fixed FASTQ-files with a custom bash script, allowing one mismatch at the first position and one mismatch at any other position of the index.
- Since the modern samples from the SGDP dataset were already aligned to the desired reference genome, we abstained from remapping them and analysed them along the ancient samples as described above, starting with filtering with for  $MQ < 30$ , marking optical and PCR duplicates and performing local realignment.

For each ancient sample, all pipeline results and differences from the standard pipeline are detailed in Table S1.

### Data filtering

Individual VCFs obtained after Bayesian genotype calling were first filtered according to their read depth (DP): for each individual we excluded sites that had  $DP < 8$  and those that had a DP bigger or smaller than 2.5 s.d. away from the mean DP for each individual. To minimise effects of outliers, the mean DP was calculated using the *R* function *optimize* for the sites comprised +/-10 of the mode DP and a tolerance of  $10^{-6}$  (see Table S2). We also excluded sites that had poor genotype quality ( $GQ < 30$ ), and we only kept autosomal sites. Furthermore, heterozygous sites were considered as homozygous if they had a significant allelic imbalance ( $p\text{-value} \leq 0.1$ ) tested by Fisher's exact test. After filtering, the modern individuals had on average 2,549,812,798 sites passing the filters against 1,980,736,974 for the ancients (ranging from 333,061,590 to 2,544,170,685). All VCF file processing was performed using *bcftools* v0.1.15 (Danecek et al., 2021).



Applying these filters led to a substantial increase in the quality of our heterozygous sites, reducing the amount of sites with high allelic imbalances and decreasing the asymmetry between singleton and non-singleton sites to a minimum (Figures M1\_3 and M1\_4). To quantify this improvement explicitly, let us denote by  $H_{is}$  the set of loci at which individual  $i$  was called heterozygous either for a singleton (only non-reference allele across all samples,  $s = 1$ ) or not ( $s = 0$ ). We then define the imbalance statistics

$$r_{is} = \frac{\sum_{l \in H_{is}} \mathbb{I}\left(\frac{r_{il}}{d_{il}} > 0.5\right)}{\sum_{l \in H_{is}} \mathbb{I}\left(\frac{r_{il}}{d_{il}} < 0.5\right)},$$

where  $r_{il}$  denotes the number of reference alleles observed in individual  $i$  at site  $l$  out of the total number of observed alleles  $d_{il}$  at this site. To compare singleton ( $s = 1$ ) against non-singleton ( $s = 0$ ) positions, we further quantify  $\rho = r_{i1}/r_{i0}$ . As shown in Figure M1\_4, the chosen filters guarantee comparable imbalances at singleton and non-singleton sites, indicating similar quality of both categories. A striking outlier was sample CarsPas1, which was not included in any demographic analysis.

### Dataset assembling

For further downstream analysis we assembled the following datasets:

- “Ancient”: The filtered individual VCFs from the 25 ancient individuals were merged into a single VCF polarised with the Chimpanzee reference genome (*panTro4* from [The Chimpanzee Sequencing Analysis Consortium, 2005](#)). We only kept the biallelic polymorphic sites (6,986,216) for which individuals can have missing genotypes (pipeline made available in our GitHub repository: <https://github.com/CMPG/originsEarlyFarmers>).
- “102samples”: Same as *Ancient* but with all 102 ancient and modern samples, keeping 14,896,103 biallelic polymorphic sites. On average, the modern genomes show 4% of missing sites with a maximum of 6%, the ancients 29% ranging from 3 to 87%.
- “Neutral”: In order to avoid biases due to background selection (BGS) and biased gene conversion (BGC) when estimating population diversity and relationships ([Matthey-Doret and Whitlock, 2019](#)), we performed most of our genetic and demographic analyses on a “neutral” portion of the genome ([Pouyet et al., 2018](#)). We defined this portion as the restricted set of sites from the *102samples* dataset that had the same reference allele in the chimpanzee and gorilla reference genomes, that were in regions with recombination rate  $>1$  cM/Mb where BGS has little effect, outside of CpG islands, non CpG sites (i.e. when a C is followed by a G in the chimpanzee and gorilla genomes and has a T as alternative allele in our data; or a G preceded by a C in the chimpanzee and gorilla genomes with an A as alternative allele) and with  $A \leftrightarrow T$  and  $G \leftrightarrow C$  mutations which are not affected by BGC or *post-mortem* damages. We kept 473,834 biallelic sites. We obtained the Chimpanzee and the Gorilla *hg19* nucleotide states from the chained and netted alignments <http://hgdownload.cse.ucsc.edu/goldenpath/hg19/vsPanTro4/axtNet/> and <http://hgdownload.cse.ucsc.edu/goldenpath/hg19/vsGorGor5/axtNet/>, respectively; the recombination map for the YRI population from the 1000 Genomes Project ([ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507\\_omni\\_recombination\\_rates/](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_rates/)); the CpG islands listing from the UCSC genome browser (CpG Islands cpGIslandExt Track). Filtering was performed in R, with VCF files read using SNPRelate package ([Zheng et al., 2012](#)).
- “Phased”: We did genotype imputation and phasing for each chromosome on the *102samples* dataset that was polarised back with *hs37d5* human reference genome. We used *SHAPEIT4* v1.2 ([Delaneau et al., 2019](#)) with by-default parameters for sequencing data, the HapMap phase II b37 genetic map provided in *SHAPEIT4* folders, and the Haplotype Reference Consortium ([McCarthy et al., 2016](#)) dataset (accession number EGAD00001002729 on the European Genome-phenome Archive) as reference panel. All the chromosomal phased-imputed VCFs were then concatenated into a single phased VCF with *bcftools* v1.9, which includes 2,795,127,079 sites in total.
- “1240k”: We compared the samples obtained in this study with previously published, target-enriched datasets available through Allen Ancient DNA Resource v42.4 (AADR) at [https://reichdata.hms.harvard.edu/pub/datasets/amh\\_repo/curated\\_releases/index\\_v42.4.html](https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/index_v42.4.html). We use these population labels, but split *Peleponnese\_N* into *Diros\_EN* and *Peleponnese\_LN*. To ensure our data is comparable to the pseudo-haploid calls, we generated majority calls for all our ancient individuals with *ATLAS* (*task=majorityBase*; commit 7cfc900) at the 1,233,013 SNP positions present in this reference dataset. We then combined the calls of the 25 ancient genomes with the AADR and refer to this dataset as *1240k* in the following.
- “1240k\_HOIII”: We extracted the non-functional *HOIII* subset of sites ([Lazaridis et al., 2016](#)) for the *Ancient* dataset. *HOIII* sites were also extracted for additional samples (Altai, Chimp, Vindija and Dinka) retrieved from the target-enriched dataset available through Allen Ancient DNA Resource v37.2 (AADR) at <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>.

### Phenotype predictions

#### Pigmentation

Pigmentation phenotypes of hair, skin and eyes were predicted with the *HirisPlex-S* webtool ([Chaitanya et al., 2018](#); [Walsh et al., 2013](#)) for each of the newly sequenced samples and the previously published samples. In case of missing genotypes for any of

the 41 SNPs used by *HlrisPlex-S*, the sites with low or no depth in the *Ancient* dataset were examined directly in the BAMs and the most abundant allele was chosen (Table S3). In BAMs, we only considered bases at least 3 bases away from either end of the read, with a base quality  $\geq 25$ , and no C $\leftrightarrow$ T and G $\leftrightarrow$ A SNPs to avoid any effect of PMD on the prediction. To account for the uncertainty associated with low-coverage sites, two *HlrisPlex-S* input files were created for each individual: one in which the missing genotype consisted of the allele taken from the BAM and the non-effect allele and one where the genotype was comprised of the allele taken from the BAM and the effect allele. Running *HlrisPlex-S* twice for each sample resulted in ranges of probabilities for each phenotype (Table S3). A prediction was accepted without further explanation if in both runs the same phenotype showed a probability  $\geq 0.7$  (Chaitanya et al., 2018). If predictions differed between runs, the most parsimonious phenotype was chosen, following the approach of Walsh in Brace et al. (2019) (Figure M1\_9).

### Standing height

To assess genetic variation related to standing height, a classical highly heritable polygenic trait (McEvoy and Visscher, 2009), polygenic scores (PS) were computed based on a set of 670 SNPs (Chan et al., 2015) on the *Ancient* dataset (Figure M1\_10). To account for missing data common even in high depth ancient genomes, we applied the generalised risk score approach described by Veeramah et al. (2018) on samples where enough SNPs were present to account for at least 75% of the scores effect size. This led to the exclusion of five ancient individuals (Bar8, Bon002, Bichon, CarsPas1, AKT16) out of the 25 included in this study.

### Other phenotypes

Genotypes for SNPs associated with additional phenotypes of interest were inspected manually for each sample in the *Ancient* dataset or BAM files if necessary: rs4988235 a variant in the *MCM6* gene associated with lactase-persistence in Eurasia; rs3827760 in the *EDAR* gene; rs17822931 in *ABCC11*; seven SNPs located in the *FADS1/2* gene complex (listed in Table M1\_2). A two-sided binomial test was used to compare counts of derived alleles in the ancient individuals with frequencies estimated in the CEU population from phase3 of the 1000 genomes project (1000 Genomes Project Consortium et al., 2015), chosen as a proxy for Central European modern populations (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Population genetics analyses

#### Uniparental haplogroup determination

Mitochondrial haplogroups were determined from the BAM files for the 15 newly-sequenced genomes using *phy-mer* (Navarro-Gomez et al., 2015) with *K-mer minimal number of occurrences* = 10 (Table S3).

For the samples genetically identified as men, Y-chromosomal haplogroups were determined using *Yleaf* (Ralf et al., 2018) with minimal base-quality of 20 (*-q* 20) and base-majority to determine an allele set to 90% (*-b* 90).

#### Neanderthal introgression

We quantified proportions of Vindija Neanderthal (Prüfer et al., 2017) introgression by computing  $f_4$  ratio statistics of the form  $\frac{f_4(\text{Altai\_Chimp}; X, \text{Dinka})}{f_4(\text{Altai\_Chimp}; \text{Vindija}, \text{Dinka})}$  as previously suggested (Petr et al., 2019) with *qpF4ratio* from the *ADMIXTOOLS* package (Patterson et al., 2012) on the 1240k\_HOIII dataset.

#### Genomic heterozygosity

We estimated the level of genetic diversity of each individual as the proportion of heterozygous sites found in the neutrally evolving portion of the genome (Figure 2C; Table S2). Thus, for every genome, we divided the amount of heterozygous sites observed in the *Neutral* dataset by the expected number of neutral sites genotyped for this individual. This number is obtained by considering all the genotyped sites for the considered individual that i) have the same reference allele for both the chimpanzee and gorilla reference genomes, ii) are not CpG sites and out of CpG islands, iii) are in regions with recombination rate  $> 1$  cM/Mb based on Pouyet et al. (2018). Finally, in order to consider only sites unaffected by BGC, we divided the number of these previously defined sites by three, as only one third of all mutations should be BGC free (i.e. A $\leftrightarrow$ T and G $\leftrightarrow$ C polymorphisms).

#### Runs of homozygosity (ROHs)

ROHs were identified on the *Phased* dataset in genomes of 90 European and SW Asian modern and ancient individuals imputed using *IBDSeq v. r1206* (Browning and Browning, 2013) with default parameters but *errormax* = 0.005 and *ibdloc* = 2. We further processed the artificially long tracts spanning assembly gaps or centromeres (genomic locations were obtained from UCSC genome browser: <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/gap.txt.gz>) and split them into shorter tracks excluding the gap stretch, inspired by Sikora et al. (2017). As advised when genomes do not come from a homogeneous population (Browning and Browning, 2013), we focused on intermediate (2-10 Mb) and long ( $>10$  Mb) ROHs similarly to Racimo et al. (2020).

#### Multi-Dimensional scaling (MDS)

Genetic relationships among individuals were estimated from pairwise average nucleotide divergence  $\pi_{XY}$  (Nei and Li., 1979). For each pair of individuals X and Y, we identified the sites showing no missing data for the two individuals, and we computed their average nucleotide divergence  $\pi_{XY}$  over these sites (Table S2). We considered whole-genome sites or neutral sites only (in this case, the number of differences was divided by the expected number of neutral sites as defined above in the heterozygosity paragraph). We then represented the relationships between modern and ancient genomes from Europe and SW Asia (for sites from the

102samples or Neutral data sets; Figure 2A; Figure M1\_5A) or only the ancient genomes (for sites from the Ancient dataset; Figure M1\_5B), using a classical multidimensional scaling (MDS) approach implemented in R (*cmdscale* function from *stats* package).

### Uniform manifold approximation and projection

In order to represent potentially complex relationships between 90 individuals (85 modern and 25 ancient samples) from Western Eurasia, we performed a Uniform Manifold Approximation and Projection (UMAP) dimension reduction on their genotypes obtained at 2035 neutral polymorphic sites (from the Neutral dataset) that did not present missing data for any individual. The analysis was performed using the *umap* function from the *umap* R package (using the options *n\_neighbors*=3, *n\_epochs*=1000, *min\_dist*=0.1, *negative\_sample\_rate*=5, *n\_components*=2, and *metric*="euclidean") (Figure M1\_6).

### Admixture clustering analyses

Admixture coefficients of each ancient genome were estimated using the R package LEA (Frichot and François, 2015) with parameters *K* = 2 or 3, *alpha* = 10, and *number of repetitions* = 5. We used the function *snmf* to calculate the fit (entropy) of each run and we plotted the admixture coefficients from the run with the smallest entropy value. The groups were defined in an unsupervised manner. In order to maximise the number of genomic sites to be used, we excluded for these analyses the three lowest quality Neolithic genomes (Bichon, Bon002, Bar8) from the Ancient dataset (Figure 2B; Figure S3A).

### f-statistics

We first used *f*-statistics to confirm that the assignment of samples to specific populations is not violated by the presence of shared drift with external samples. For this, we calculated all possible *f*-statistics on the 1240k dataset of the form of D(Individual 1 from the population tested, Individual 2 from the same population; Other samples, Outgroup) using ADMIXTOOLS (Patterson et al., 2012) and *Mbuti* as the outgroup. We used *qpGraph* from ADMIXTOOLS v7300 (Patterson et al., 2012) using 1,000 initial conditions to match specificities of the datasets.

To validate the model inferred by *fastsimcoal2*, we aimed at comparing *f*-statistics predicted under the *fastsimcoal2* model against those calculated from the data. For this, we first created a full admixture graph carefully matching the model in Figure 3. To validate this graph, we used data simulated with *fastsimcoal2* under the model shown in Figure 3 (see Methods S1 - Demogenomic inference with *fastsimcoal2* - Final Model" section), which we fitted with *qpGraph* 100 times with 1,000 initial conditions each. However, *qpGraph* failed to identify a suitable set of branch lengths and admixture proportion to explain the simulated data with all runs resulting in predicted *f*-statistics very different from those calculated on the simulated data, suggesting that the SFS captures more information of the full data than *f*-statistics do.

Since the full graph could not be fitted, we next aimed at manually simplifying the graph. Specifically, we removed admixture edges, until we identified the simplest graph (in terms of admixture edges) for which no major differences between the *f*-statistics predicted under the model and those calculated from the simulated data were detected. This graph was then re-fitted with *qpGraph* on both a subset of the Ancient dataset (the sites from the *EurEF* panel shown in Table M1\_4) as well as on the 1240k dataset described above. In both cases, *qpGraph* was run 100 times with 1,000 initial conditions each. To present the fitted graphs and quantify the reliability of the fitting, we calculated the median and 90% quantiles for all branch lengths and admixture odds ratios across the 20 best graphs as judged by the final score. A more detailed description can be found in Methods S1.

### Joint Distribution of Fitness Effects

To create a representative Neolithic model population, we aggregated all Neolithic samples that were newly sequenced, plus the WC1 genome (*N* = 14) from the 102samples dataset. As well, to create a representative modern population, we used the SGDP populations in the 102samples dataset that are in rough proximity to the ancient populations (Polish, Bergamo, Czech, French, Hungarian, Greek, Albanian, Bulgarian, and Turkish). We annotated the synonymous and nonsynonymous SNPs using ANNOVAR (Wang et al., 2010) with hg19/GRCh37 (included with ANNOVAR) as the reference genome. Ancestral alleles were determined based on the Ensembl Compara 71 genome FASTA files. To account for missing data, we projected the joint allele frequency spectra downward to 16 Neolithic chromosomes and 28 modern chromosomes, to roughly maximise the number of segregating synonymous SNPs.

The joint Distribution of Fitness Effects (DFE) analysis requires a simpler demographic model than our main inference (Figure 3), and simulations suggest that joint DFE analysis is robust to demographic model details (Huang et al., 2021). Based on prior results (Gravel et al., 2011), we used *dadi* (Gutenkunst et al., 2009) to fit a demographic model to the synonymous data in which the ancestors of the Neolithic population underwent a bottleneck to relative size  $n_B$  followed by exponential growth (Figure M1\_11A). The ancestors of the modern samples diverged  $T_B$  time units after the bottleneck, following the same growth rate to reach final size  $n_F$ .  $T_S$  time units after this divergence, Neolithic chromosomes were sampled,  $T_F$  time units before present. We also included a parameter  $p_{mis}$  to account for potential ancestral state misidentification (Ragsdale et al., 2016). To fit this model, we used grid points of [128, 138, 148], and the best-fit parameters were  $n_B = 0.113$ ,  $n_F = 1.42$ ,  $T_B = 0.106$ ,  $T_S = 0.0054$ ,  $T_F = 0.0232$ ,  $p_{mis} = 0.0634$ .

We then fit a model including this demographic history plus a joint DFE to the nonsynonymous data (Huang et al., 2021). We modelled the DFE as a bivariate lognormal distribution. We assumed that the population-scaled mutation rate  $q$  for nonsynonymous mutations was 2.31 times that for synonymous mutations, and we again included a parameter for ancestral state misidentification. We assumed that selection coefficients were potentially different in the ancestors of the modern individuals (Figure M1\_11E). The best fit parameters for the joint DFE were  $m = 3.29$ ,  $s = 2.61$ , and  $r = 0.9968$ , with 0 misidentification inferred.

## Demographic Analyses

### MSMC2

We used *MSMC2* (Schiffels and Wang, 2020) to infer past effective sizes of ancestral populations and their split times for all high-quality ancient and some representative modern individuals (Table M1\_3) on the *Phased* dataset. To ensure high data quality especially high mappability and genotype quality, we followed (Schiffels and Wang, 2020) and used two masks: 1) per chromosome mappability masks ([um75-hs37d5.bed.gz](https://github.com/wangke16/MSMC-IM/tree/master/masks)) for the human reference genome *hs37d5* downloaded from <https://github.com/wangke16/MSMC-IM/tree/master/masks>; 2) sample-specific masks that we generated as suggested by Schiffels and Wang (2020). We then ran the *generate\_multihetsep.py* script from *MSMC-tools* (<https://github.com/stschiff/msmc-tools>, commit 07bc8a9) to get single, multi-sample and paired population input files for *MSMC2*.

Example of a command line for two, four or eight haplotypes:

```
generate_multihetsep.py -chr 1 -mask ind1.chr1.mask.bed -mask
mappabilityMaskperChr/chr1_m75-hs37d5.bed ind1.chr1.phased.vcf > ind1.chr1.multihetsep.txt
generate_multihetsep.py -chr 1 -mask ind1.chr1.mask.bed -mask ind2.chr1.mask.bed -mask
mappabilityMaskperChr/chr1_m75-hs37d5.bed ind1.chr1.phased.vcf ind2.chr1.phased.vcf > pop1.chr1.multihetsep.txt
generate_multihetsep.py -chr 1 -mask ind1.chr1.mask.bed -mask ind2.chr1.mask.bed -mask
ind3.chr1.mask.bed -mask ind4.chr1.mask.bed -mask mappabilityMaskperChr/chr1_m75-hs37d5.bed ind1.chr1.phased.vcf in-
d2.1.phased.vcf ind3.chr1.phased.vcf ind4.chr1.phased.vcf >
pop1_pop2.chr1.multihetsep.txt
```

We inferred past effective population size for each diploid individual separately (Figure M1\_12) as well as using two samples per population if available (Figure S2), using command lines such as:

```
msmc2 -t11 -s -o ind1.2haps.msmc2 ind1.chr*.multihetsep.txt
msmc2 -t11 -l 0,1,2,3 -s -o pop1.4haps.msmc2 pop1.chr*.multihetsep.txt
```

All *MSMC2* results were scaled using a mutation rate of  $1.25 \times 10^{-8}$  per base pair per generation (Scally and Durbin, 2012; Schiffels and Wang, 2020), and a generation time of 29 years (Fenner, 2005).

To estimate split times between population pairs, we used *MSMC2* 1) to estimate coalescent rates among the samples of the first population, 2) to estimate coalescent rates among the samples of the second population, and 3) to estimate coalescent rates across the two populations. Example command lines used for two populations pop1 and pop2 were:

```
msmc2 -t11 -l 0,1,2,3 -s -o pop1.4haps.msmc2 pops.chr*.multihetsep.txt,
msmc2 -t11 -l 4,5,6,7 -s -o pop2.4haps.msmc2 pops.chr*.multihetsep.txt,
msmc2 -t11 -l 0-4,0-5,0-6,0-7,1-4,1-5,1-6,1-7,2-4,2-5,2-6,2-7,3-4,3-5,3-6,3-7 -s -o
pop1-pop2.8haps.cross.msmc2 pops.chr*.multihetsep.txt.
```

We then used the *MSMC2* script *combineCrossCoal.py* to create a single output file with all three rates:

```
combineCrossCoal.py pop1-pop2.*.cross.msmc2.final.txt pop1.*.msmc2.final.txt pop2.*.msmc2.final.txt
> pop1-pop2.combined.msmc2.final.txt
```

For each population-pair, we then plotted the relative cross-coalescence rate (CCR), which is estimated by taking the ratio of the across-rate and the mean within-rate (Figure M1\_13; Schiffels and Wang, 2020). The relative CCR indicates when two populations were a single population (values around 1) and when they were well separated into two isolated populations (values close to zero) (Schiffels and Wang, 2020). However, translating the relative CCR into estimates of split times is difficult for two reasons: First, if a population split was followed by migration, the relative CCR will remain high even after the split. Second, the uncertainty associated with the different coalescent rates translates into a gradual change in the relative CCR, even under a hard split. As a rough estimate, (Schiffels and Wang, 2020) recommends estimating split times as the time when the relative CCR hits 0.5, which we show in Figure M1\_14 for all pairwise comparisons.

To obtain confidence intervals around coalescence rate estimates, we generated 20 artificial genomes by block-bootstrapping *MSMC2* input files in 5Mb blocks (Figures M1\_15 and M1\_16) as suggested in Schiffels and Wang (2020).

### fastsimcoal2

Demographic inferences were carried out with *fastsimcoal2.7* (Excoffier et al., 2013, 2021) on six different panels of newly sequenced ancient individuals, on the neutral SFS and with neutral mutation rate adjusted for each panel from the basal mutation (Table M1\_4).

**Panels composition and properties:** For each panel, we excluded from the *Ancient* dataset i) sites with any missing data in any individual of the panel using *BEDOPS* (Neph et al., 2012), ii) sites with different reference allele for the chimp and gorilla reference genomes, iii) CpG sites, iv) sites found in CpG islands or v) in genomic regions with recombination rate less than 1 cM/Mb. We thus kept a total of  $T_x$  sites for the panel  $x$ , among which  $S_x$  were polymorphic and  $M_x$  monomorphic (the properties of each panel are found in Table M1\_4). Second, in order to have neutral data best suited for demographic inference (Pouyet et al., 2018), we only kept  $S_{\text{neutral}_x}$  BGC-free A ↔ T and G ↔ C polymorphic sites, which represented  $\alpha_x \sim 0.2$  of the filtered polymorphic sites for any panel. Since we would have expected them to represent one third of all filtered sites if mutation rates had been equal for all mutation types, we computed the ratio  $r_x = \alpha_x / (1/3) = 3 \alpha_x$ , which represents the reduction in mutation rates that has occurred at those sites in panel $_x$ . We then estimated the number of neutrally evolving monomorphic sites in each panel as  $M_{\text{neutral}_x} = M_x/3$ , since we expect that one third of all mutations should be BGC-free.



**Parameter inference via maximum likelihood:** Parameter estimates were obtained by maximising the model likelihood over 50 independent runs of *fastsimcoal2*, 50 expectation conditional maximisation (ECM) cycles per run and 500,000 coalescent simulations per estimation of the expected SFS (except for the *Core* models for which 200,000 simulations were run). The command line used for the estimation was of the type:

```
fsc -t xxx.tpl -n500000 -d -e xxx.est -M -L50 -q -C5 -multiSFS -logprecision 18 -c1 -B1
```

where *fsc* is the *fastsimcoal2.7* program (available on <http://cmpg.unibe.ch/software/fastsimcoal2/>) and *xxx* the generic name of the input files. The *.est* and *.tpl* input files are available in our GitHub repository: <https://github.com/CMPG/originsEarlyFarmers>.

**Likelihood comparison and model choice:** When several models were tested for the same panel, we retained the model with the highest estimated likelihood over 50 runs, and recorded the maximum likelihood (ML) parameters of this model. In order to take into account the variance in the estimation of the likelihoods due the limited number of performed coalescent simulations (here 200,000 or 500,000) to estimate the expected SFS, we also compared the likelihoods of the models estimated on the basis of 10 million coalescent simulations done under the ML parameters. We repeated this procedure 100 times per model to check if the distributions of these likelihoods were overlapping and thus not distinguishable. We used the following command line to get these 100 likelihoods:

```
fsc -i xxx_maxL.par -R100 -n10000000 -d -u -C5 -logprecision 18 -q
```

Finally, we also computed the Akaike criterion (AIC) (Akaike, 1974) and the model relative likelihoods assuming site independence, even though our “neutral” sites may be linked on some chromosomes.

**Confidence intervals:** Confidence intervals around ML parameter point estimates were obtained via a parametric bootstrap approach, for which we first generated 100 SFS covering  $T_{neutral,X}$  nucleotides using estimated ML parameters, with the command line:

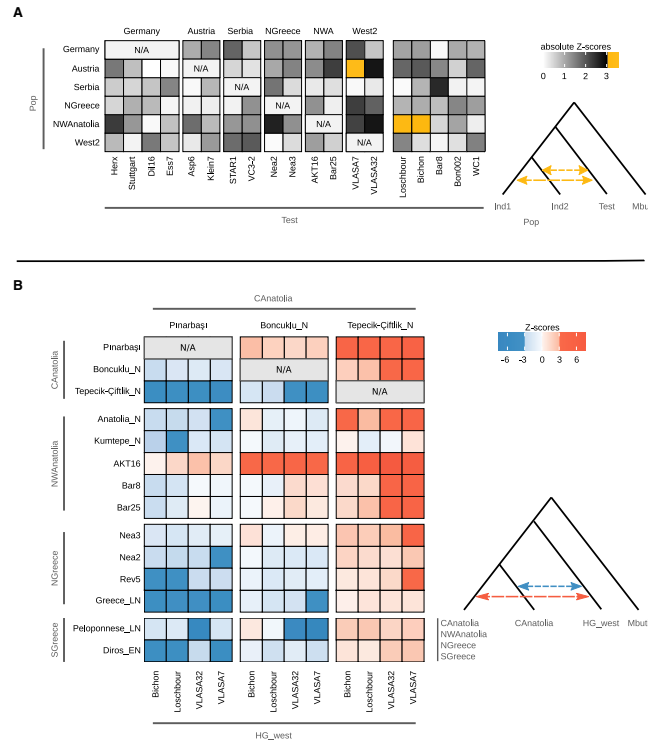
```
fsc -i xxx.par -n100 -j -d -s0 -x -l -q -u
```

Then, for each of these bootstrapped SFS, we re-estimated the parameters of the model using 20 independent runs starting at the ML parameters value (option *-initvalues* in *fastsimcoal2*). We used 60 ECM cycles for each run and performed 500,000 simulations for estimating the expected SFS under a given set of parameter values and to estimate the model likelihood. The *fastsimcoal2* command line used for the bootstrap was of the type:

```
fsc -t xxx.tpl -n500000 -d -e xxx.est -initvalues xxx.pv -M -L60 -q -C5 -multiSFS -logprecision 18
```

The limits of 95% confidence intervals were finally estimated by computing the 2.5% and 97.5% quantiles of the distribution of the 100 newly estimated ML parameter values.

# Supplemental figures

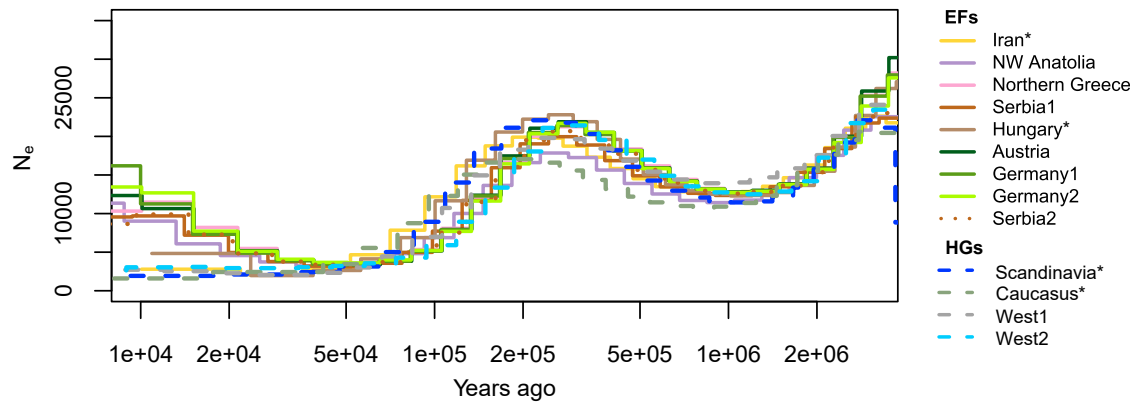


**Figure S1. Population grouping verified with  $f$ -statistics, related to Figure 2 and Methods S1**

These analyses were performed on the 1240k dataset.

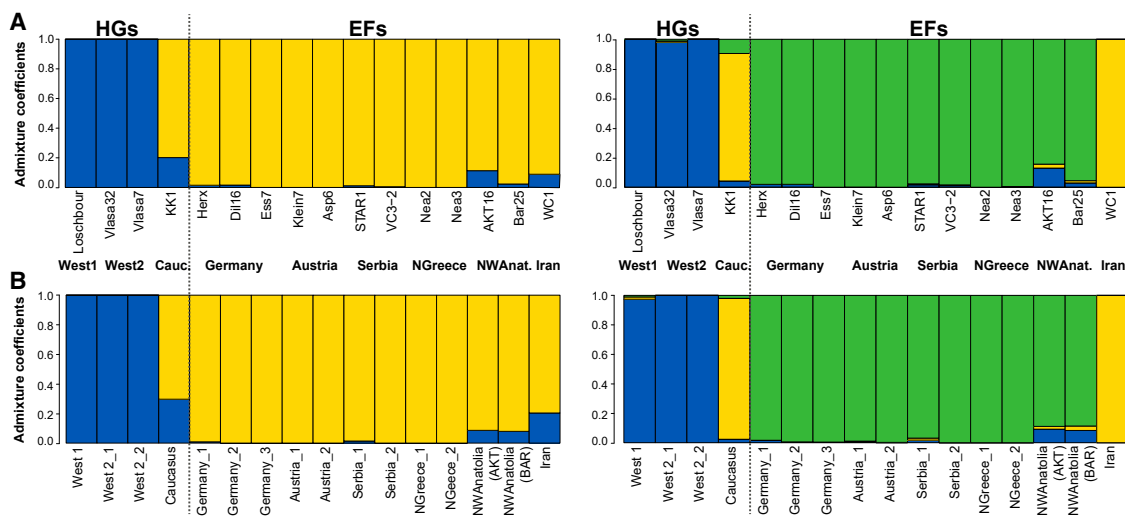
(A) We first tested if some individuals of a specific group had significantly more shared ancestry with individuals of a different group using  $f$ -statistics of the form  $D(Ind1 \text{ from a population, } Ind2 \text{ from the same population; Test, Mbuti [outgroup]})$ . We only found three significant absolute  $Z$  scores ( $>3.0$ , yellow). For Austria, Asp6 appears to share more ancestry with VLASA7 than Klein7. Since variation in European HG ancestry is expected in EF populations due to the ongoing process of admixture and since these samples did not show variation in their affinities to other EF samples, modeling them as a single population seems justified. For NWAnatolia, AKT16 was found to share significantly more ancestry with both Loschbour and Bichon, which we further investigated in (B).

(B) To shed more light on the variation in European HG ancestry among Anatolian and Greek samples, we calculated  $f$ -statistics of the form  $D(NGreece/SGreece/NWAnatolia/CAnatolia, CAnatolia; HG_west, Mbuti [outgroup])$ , where we use  $HG_west$  to denote both West 1 and West 2 European HGs. This test indicates whether the tested individual/population from NGreece, SGreece, NWAnatolia, or CAnatolia (left) shares more (orange,  $Z$  score  $> 0$ ) or less (blue,  $Z$  score  $< 0$ ) ancestry with the tested  $HG_west$  individual (bottom) than the tested individual from CAnatolia (top). Significant  $Z$  scores above 3.0 or below  $-3.0$  are shown with more intense colors. Among the CAnatolian individuals, Pinarbaşı and Boncuklu\_N appear to share excess drift with  $HG_west$  when compared with individuals from Greece and NWAnatolia, in contrast to Tepecik-Çiftlik\_N. Among the individuals from NGreece, SGreece, and NWAnatolia, AKT16 appears closest to  $HG_west$ , and much closer than Bar25. In light of these results, we modeled AKT16 and Bar25 independently in the demographic inferences.



**Figure S2. MSMC2 effective population size estimates, related to Figure 3 and Methods S1**

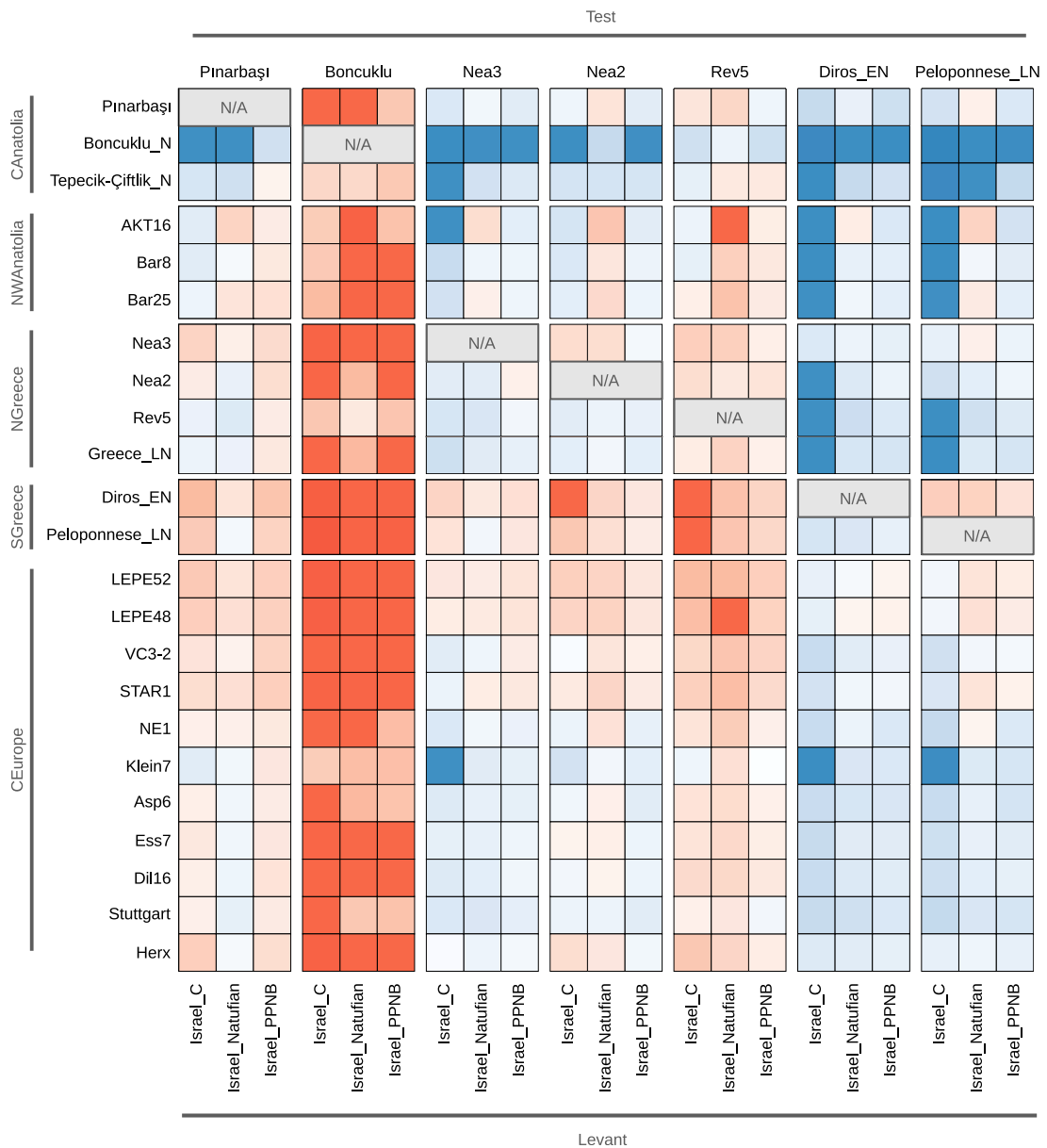
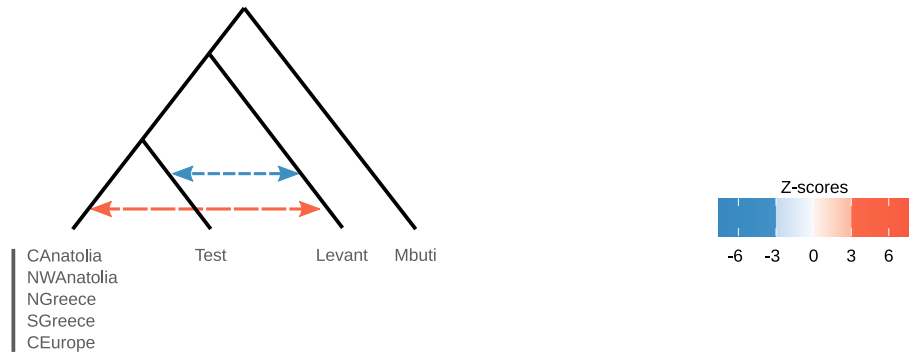
These were obtained for populations either with four haplotypes or two haplotypes when only one ancient individual was available for the population (shown in the legend with \*, details in Table M1\_3, population sizes estimated from single ancient and modern genomes are shown in Figure M1\_12). We used a mutation rate of  $1.25 \times 10^{-8}$  per generation per site and a generation time of 29 years. The analysis suggests smaller effective population sizes in the most recent times for HGs compared with EFs.



**Figure S3. Comparison of observed and simulated Admixture plots, related to Figure 4 and Methods S1.**

Admixture plot for  $K = 2$  (left panel) and  $K = 3$  (right panel) carried out on (A) observed data for 16 ancient genomes included into *fastsimcoal2* demographic inferences (909,688 sites without missing data; Bichon, Bar8, and Bon002 were not included because of lower quality) and on (B) data simulated accordingly to our final model (shown in Figure 3), for the same subset of individuals as in (A).





(legend on next page)

---

**Figure S4. Patterns of population admixture revealed by *f*-statistics, related to Figure 5 and Methods S1.**

Relationship of Anatolian and Greek Neolithics with the Levant using *f*-statistics on the 1240k dataset of the form  $D(\text{CAnatolia/NWAnatolia/NGreece/SGreece/CEurope,Test}; \text{Levant,Mbuti [outgroup]})$ . This test indicates whether Neolithic individuals from *CAnatolia*, *NWAnatolia*, *Greece*, or *CEurope* (left) share less (blue, *Z* score < 0) or more (orange, *Z* score > 0) ancestry with samples from the Levant, namely individuals from contemporary Israel associated with Natufian culture (*Israel\_Natufian*), Pre-Pottery Neolithic B (*Israel\_PPNB*), and Chalcolithic (*Israel\_C*; all bottom) than the *Test* individuals/populations from *Greece* and *CAnatolia*. Significant *Z* scores above 3.0 or below -3.0 are shown with more intense colors. We find a strongly significant excess of shared drift between populations from the *Levant* and *NGreece*, *SGreece*, *NWAnatolia*, and *CEurope* when contrasted to *Boncuklu*. This signal was, however, not replicated when representing *CAnatolia* by *Pınarbaşı*. In contrast, the *SGreece* populations *Diros\_EN* and *Peloponnese\_N* appear to share excess drift with populations from the Levant, and in particular the Chalcolithic *Israel\_C*, when contrasted to samples from *NGreece*, *NWAnatolia*, or *CEurope*.