

# Reconstructing the spatiotemporal patterns of admixture during the European Holocene using a novel genomic dating method

Manjusha Chintalapati<sup>1#</sup>, Nick Patterson<sup>2#</sup>, Priya Moorjani<sup>1,3#</sup>

<sup>1</sup>Department of Genetics, University of California, Berkeley, CA 94620 USA

<sup>2</sup>Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA

<sup>3</sup>Center for Computational Biology, University of California, Berkeley, CA 94620 USA

<sup>#</sup>To whom correspondence should be addressed: Manjusha Chintalapati <[m\\_chintalapati@berkeley.edu](mailto:m_chintalapati@berkeley.edu)>, Nick Patterson <[nickp@broadinstitute.org](mailto:nickp@broadinstitute.org)> and Priya Moorjani <[moorjani@berkeley.edu](mailto:moorjani@berkeley.edu)>

## Abstract

Recent studies have shown that gene flow or admixture has been pervasive throughout human history. While several methods exist for dating admixture in contemporary populations, they are not suitable for sparse, low coverage data available from ancient specimens. To overcome this limitation, we developed *DATES* that leverages ancestry covariance patterns across the genome of a single individual to infer the timing of admixture. By performing simulations, we show that *DATES* provides reliable results under a range of demographic scenarios and outperforms available methods for ancient DNA applications. We apply *DATES* to ~1,100 ancient genomes to reconstruct gene flow events during the European Holocene. Present-day Europeans derive ancestry from three distinct groups, local Mesolithic hunter-gatherers, Anatolian farmers, and Yamnaya Steppe pastoralists. These ancestral groups were themselves admixed. By studying the formation of Anatolian farmers, we infer that the gene flow related to Iranian Neolithic farmers occurred no later than 9,600 BCE, predating agriculture in Anatolia. We estimate the early Steppe pastoralist groups genetically formed more than a millennium before the start of steppe pastoralism, providing new insights about the history of proto-Yamnaya cultures and the origin of Indo-European languages. Using ancient genomes across sixteen regions in Europe, we provide a detailed chronology of the Neolithization across Europe that occurred from ~6,400–4,300 BCE. This movement was followed by a rapid spread of steppe ancestry from ~3,200–2,500 BCE. Our analyses highlight the power of genomic dating methods to elucidate the legacy of human migrations, providing insights complementary to archaeological and linguistic evidence.

## Keywords

genomic clocks, admixture, ancient DNA, European Holocene, molecular clock, migration, Neolithic, Bronze Age, Yamnaya

## 38 **Significance**

39

40 **The European continent was subject to two major migrations during the Holocene: the**  
41 **movement of Near Eastern farmers during the Neolithic and the migration of Steppe**  
42 **pastoralists during the Bronze Age. To understand the timing and dynamics of these**  
43 **movements, we developed *DATES* that leverages ancestry covariance patterns across the**  
44 **genome of a single individual to infer the timing of admixture. Using ~1,100 ancient genomes**  
45 **spanning ~8,000–350 BCE, we reconstruct the chronology of the formation of the ancestral**  
46 **populations and the fine-scale details of the spread of Neolithic farming and Steppe**  
47 **pastoralist-related ancestry to Europe. Our analysis demonstrates the power of genomic**  
48 **dating methods to provide an independent and complementary timeline of population origins**  
49 **and movements using genetic data.**

## 50 **Introduction**

51

52         Recent studies have shown that population mixture (or “admixture”) is pervasive  
53 throughout human history, including mixture between the ancestors of modern humans and archaic  
54 hominins (i.e., Neanderthals and Denisovans), as well as in the history of many contemporary  
55 human groups such as African Americans, South Asians and Europeans (1, 2). Many admixed  
56 groups are formed due to population movements involving ancient migrations that pre-date  
57 historical records. The recent availability of genomic data for a large number of present-day and  
58 ancient genomes provides an unprecedented opportunity to reconstruct population events using  
59 genetic data, providing evidence complementary to linguistics and archaeology. Understanding  
60 the timing and signatures of admixture offers insights into the historical context in which the  
61 mixture occurred and enables the characterization of the evolutionary and functional impact of the  
62 gene flow.

63         To characterize patterns of admixture, genetic methods use the insight that the genome of  
64 an admixed individual is a mosaic of chromosomal segments inherited from distinct ancestral  
65 populations (3). Due to recombination, these ancestral segments get shuffled in each generation  
66 and become smaller and smaller over time. The length of the segments is inversely proportional to  
67 the time elapsed since the mixture (3, 4). Several genetic approaches—ROLLOFF (4), ALDER  
68 (5), Globetrotter (2), and Tracts (6)—have been developed that use this insight by characterizing  
69 patterns of admixture linkage disequilibrium (LD) or haplotype lengths across the genome to infer  
70 the timing of mixture. Haplotype-based methods perform chromosome painting or local ancestry  
71 inference at each locus in the genome and characterize the distribution of ancestry tract lengths to  
72 estimate the time of mixture (2, 6). This requires accurate phasing and inference of local ancestry,  
73 which is often difficult when the admixture events are old (as ancestry blocks become smaller over  
74 time) or when reference data from ancestral populations is unavailable. Admixture LD-based  
75 methods, on the other hand, measure the extent of the allelic correlation across markers to infer  
76 the time of admixture (4, 5). They do not require phased data from the target or reference  
77 populations and work reliably for dating older admixture events (>100 generations). However,  
78 they tend to be less efficient in characterizing admixture events between closely related ancestral  
79 groups.

80         While highly accurate for dating admixture events using data from present-day samples,  
81 current methods do not work reliably for dating admixture events using ancient genomes. Ancient  
82 DNA samples often have high rates of DNA degradation, contamination (from human and other  
83 sources) and low sequencing depth, leading to a large proportion of missing variants and uneven  
84 coverage across the genome. Additionally, most studies generate pseudo-homozygous genotype  
85 calls—consisting of a single allele call at each diploid site—that can lead to some issues in the  
86 inference. In such sparse datasets, estimating admixture LD can be noisy and biased (see  
87 Simulations below). Moreover, haplotype-based methods require phased data from both admixed  
88 and reference populations which remains challenging for ancient DNA specimens.

89 An extension of admixture LD-based methods, recently introduced by Moorjani et al.  
90 (2016), leverages ancestry covariance patterns that can be measured in a single sample using low  
91 coverage data. This approach measures the allelic correlation across neighboring sites, but instead  
92 of measuring admixture LD across multiple samples, it integrates data across markers within a  
93 single diploid genome. Using a set of ascertained markers that are informative for Neanderthal  
94 ancestry (where sub-Saharan Africans are fixed for the ancestral alleles and Neanderthals have a  
95 derived allele), Moorjani et al. (2016) inferred the timing of Neanderthal gene flow in Upper  
96 Paleolithic Eurasian samples and showed the approach works accurately in ancient DNA samples  
97 (1). However, this approach is inapplicable for dating admixture events within modern human  
98 populations, as there are very few fixed differences across populations (7).

99 Motivated by the single sample statistic in Moorjani et al. (2016), we developed *DATES*  
100 (*Distribution of Ancestry Tracts of Evolutionary Signals*) that measures the ancestry covariance  
101 across the genome in a single admixed individual, weighted by the allele frequency difference  
102 between two ancestral populations. This method was first introduced in Narasimhan et al. (2019),  
103 where it was used to infer the date of gene flow between groups related to Ancient Ancestral South  
104 Indians, Iranian farmers, and Steppe pastoralists in ancient South and Central Asian populations  
105 (8). In this study, we evaluate the performance of *DATES* by performing extensive simulations for  
106 a range of demographic scenarios and compare the approach to other published genomic dating  
107 methods. We then apply *DATES* to infer the chronology of the genetic formation of the ancestral  
108 populations of Europeans and the spatiotemporal patterns of admixture during the European  
109 Holocene using data from ~1,100 ancient DNA specimens spanning ~8,000–350 BCE.

## 110 **Results**

111

### 112 **Overview of *DATES*: Model and simulations**

113

114 *DATES* estimates the time of admixture by measuring the weighted ancestry covariance  
115 across the genome using data from a single diploid genome and two reference populations  
116 (representing the ancestral source populations). *DATES* works like haplotype-based methods as it  
117 is applicable to dating admixture in a single genome and not like admixture LD-based methods,  
118 which by definition require multiple genomes to be co-analyzed; but unlike haplotype-based  
119 methods, it is more flexible as it does not require local ancestry inference. There are three main  
120 steps in *DATES*: we start by first learning the genome-wide ancestry proportions by performing a  
121 simple regression analysis to model the observed genotypes in an admixed individual as a linear  
122 mix of allele frequencies from the two reference populations. For each marker, we then compute  
123 the likelihood of the observed genotype in the admixed individual using the estimated ancestry  
124 proportions and allele frequencies in each reference population (this is similar in spirit to local  
125 ancestry inference). This information is, in turn, used to compute the joint likelihood for two  
126 neighboring markers to test if they derive ancestry from the same ancestral group, accounting for  
127 the probability of recombination between the two markers. Finally, we compute the covariance

128 across pairs of markers located at a particular genetic distance, weighted by the allele frequency  
129 differences in the reference populations (Note S1).

130 Following (1), we bin the markers that occur at a similar genetic distance across the  
131 genome, rather than estimating admixture LD for each pair of markers, and compute the covariance  
132 across increasing genetic distance between markers. The estimated covariance is expected to decay  
133 exponentially with genetic distance, and the rate of decay is informative of the time of the mixture  
134 (4). Assuming the gene flow occurred instantaneously, we infer the average date of gene flow by  
135 fitting an exponential distribution to the decay pattern (Methods). In cases where data for multiple  
136 individuals is available, we compute the likelihood by summing over all individuals. To make  
137 *DATES* computationally tractable, we implemented the fast Fourier transform (FFT) for  
138 calculating ancestry covariance as described in ALDER (5). This provides a speedup from  $O(n^2)$   
139 to  $O(n \log n)$ , which reduces the typical runtimes from hours to seconds (Note S1).

140 To assess the reliability of *DATES*, we performed simulations where we constructed ten  
141 admixed diploid genomes by randomly sampling haplotypes from two source populations (Note  
142 S2). Briefly, we simulated individual genomes with 20% European and 80% African ancestry by  
143 using phased haplotypes of Northern Europeans (Utah European Americans, CEU) and west  
144 Africans (Yoruba from Nigeria, YRI) from the 1000 Genomes Project respectively (7). As  
145 reference populations in *DATES*, we used closely related surrogate populations of French and  
146 Yoruba respectively, from the Human Genome Diversity Panel (9). We first investigated the  
147 accuracy of *DATES* by varying the time of admixture between 10–300 generations. For  
148 comparison, we also applied ALDER (5) to these simulations. Both methods reliably recovered  
149 the time of admixture up to 200 generations or ~5,600 years ago, assuming a generation time of  
150 28 years (1), though *DATES* was more precise than ALDER for older admixture events (>100  
151 generations) (Table S2.4). Further, *DATES* shows accurate results even for single samples (Figure  
152 1).

153 Next, we tested *DATES* for features such as varying admixture proportions and use of surrogate  
154 populations as reference groups. By varying of European ancestry proportion between ~1–50%  
155 (the rest derived from west Africans), we observed *DATES* accurately estimated the timing in all  
156 cases (Figure S2.2A). However, the inferred admixture proportion was overestimated was lower  
157 admixture proportions (<10%) (Figure S2.2B). Thus, we caution against using *DATES* for  
158 estimating ancestry proportions and recommend other methods based on  $f$ -statistics (10). Using  
159 reference populations which are divergent from true admixing source, we found that the inferred  
160 dates were accurate even when we used Khomani San instead of Yoruba as the reference  
161 population ( $F_{ST} \sim 0.1$ ) (Figure S2.5). We also found that using the admixed samples themselves as  
162 one of the reference populations also works reliably as ALDER (i.e., single reference setup) (5).

163 An important feature of *DATES* is that it does not require phased data and is applicable to  
164 datasets with small sample sizes, making it in principle useful for ancient DNA applications. To  
165 test the reliability of *DATES* for ancient genomes, we simulated data mimicking the relevant  
166 features of ancient genomes, namely with large proportions of missing genotypes (between 10–  
167 60%), and pseudo-homozygous genotype calls (instead of diploid genotype calls). *DATES* showed

168 reliable results in both cases, even only a single admixed individual was available (Figure S2.7).  
169 In contrast, admixture LD-based methods require more than one sample and do not work reliably  
170 with missing data. For example, ALDER estimates were very unstable for simulations with >40%  
171 missing data. For older dates (>100 generations), we observed slight bias even with >10% missing  
172 genotypes (Figure S2.17). As LD calculations leverage shared patterns across samples, variable  
173 missingness of genotypes across individuals leads to substantial loss of data leading to unstable  
174 and noisy inference. This highlights a major advantage of *DATES* for ancient DNA studies as it  
175 provides reliable results even in sparse datasets (Note S2.5).

176 *DATES* assumes a model of instantaneous gene flow with a single pulse of mixture between  
177 two source populations. However, many human populations have a history of multiple pulses of  
178 gene flow. To test the performance of *DATES* for multi-way admixture events, we generated  
179 admixed individuals with ancestry from three sources (East Asians, Africans, and Europeans)  
180 where the gene flow occurred at two distinct time points (Note S2, Figure S2.10). By applying  
181 *DATES* with pairs of reference populations at a time and fitting a single exponential to the ancestry  
182 covariance patterns, we observed that *DATES* recovered both admixture times in case of equal  
183 ancestry proportion from the three ancestral groups when the associated reference groups were  
184 used for dating (Figure S2.11). In the case of unequal admixture proportions from three ancestral  
185 groups, *DATES* inferred the timing of the recent admixture event in most cases, though some  
186 confounding was observed, especially when the ancestry proportion of the recent event was low  
187 (Figure S2.12). However, if the reference populations were set up to match the model of gene flow,  
188 we observed that we could reliably recover the time of the recent gene flow event. For example,  
189 there is limited confounding if the two references used in *DATES* include (i) the source population  
190 for the recent event and (ii) either the pooled ancestral populations contributing to the first (or  
191 earlier) event or the intermediate admixed group formed after the first event (Table S2.1). This  
192 highlights how the choice of reference populations can help to tune the method to infer the timing  
193 of specific admixture events reliably.

194 Finally, we explored the impact of more complex demographic events, including  
195 continuous admixture and founder events using coalescent simulations (Note S2). In the case of  
196 continuous admixture, *DATES* inferred an intermediate timing between the start and the end of the  
197 gene flow period, similar to other methods like ALDER and Globetrotter (2, 5) (Table S2.2). In  
198 the case of populations with founder events, we inferred unbiased dates of admixture in most cases  
199 except when the founder event was extreme ( $N_e \sim 10$ ) or the population had maintained a low  
200 population size ( $N_e < 100$ ) until present (i.e., no recovery bottleneck) (Figure S2.13, Table S2.3).  
201 In humans, few populations have such extreme founder events, and thus, in most other cases, our  
202 inferred admixture dates should be robust to founder events (11). We note that while *DATES* is not  
203 a formal test of admixture, in simulations, we find that in the absence of gene flow, the method  
204 does not infer significant dates of admixture even when the target has a complex demographic  
205 history (Figure S2.15, S2.16).

206

## 207 **Comparison to other methods**

208  
209 We assessed the reliability of *DATES* in real data by comparing our results with published  
210 methods: Globetrotter, ALDER, and ROLLOFF. These methods are designed for the analysis of  
211 present-day samples that typically have high-quality data with limited missing variants. In  
212 addition, Globetrotter uses phased data which is challenging for ancient DNA samples. Thus,  
213 instead of rerunning other methods, we took advantage of the published results for contemporary  
214 samples presented in Hellenthal et al., (2014) (2). Following (2), we created a merged dataset  
215 including individuals from Human Genome Diversity Panel (9), Behar et al. (2010) (12), and Henn  
216 et al. (2012) (13) (Methods). We applied *DATES* and ALDER to 29 target groups using the  
217 reference populations reported in Hellenthal et al. 2014 (Table S12), excluding one group where  
218 the population label was unclear. Interestingly, the majority of these groups (25/29) failed  
219 ALDER's formal test of admixture; either because the results of the single reference and two  
220 reference analyses yielded inconsistent estimates or because the target had long-range shared LD  
221 with one of the reference populations (Table S4.1). Using *DATES*, we inferred significant dates of  
222 admixture in 20 groups, and 14 of those were consistent with estimates based on Globetrotter. In  
223 most remaining cases, recent studies suggest the target populations may have ancestry from  
224 multiple gene flow events, either involving the same source populations or additional ancestral  
225 groups. The estimated admixture timing based on *DATES*, ROLLOFF, and ALDER (assuming  
226 two-way admixture regardless of the formal test results) were found to be highly concordant (Table  
227 S4.1).

228

## 229 **Fine-scale patterns of population mixtures in ancient Europe**

230

231 Recent ancient DNA studies have shown that present-day Europeans derive ancestry from three  
232 distinct sources: (a) hunter-gatherer-related ancestry that is closely related to Mesolithic hunter-  
233 gatherers (HG) from Europe; (b) Anatolian farmer-related ancestry related to Neolithic farmers  
234 from the Near East and associated to the spread of farming to Europe; and (c) Steppe pastoralist-  
235 related ancestry that is related to the Yamnaya pastoralists from Russia and Ukraine (16–19). Many  
236 open questions remain about the timing and dynamics of these population interactions, in particular  
237 related to the formation of the ancestral groups (which were themselves admixed) and their  
238 expansion across Europe. To characterize the spatial and temporal patterns of mixtures in Europe  
239 in the past 10,000 years, we used 1,096 ancient European samples from 152 groups from the  
240 publicly available Allen Ancient DNA Resource (AADR) spanning a time range of ~8,000–350  
241 BCE (Methods, Table SA). Using *DATES*, we characterized the timing of the various gene flow  
242 events, and below, we describe the key events in chronological order focusing on three main  
243 periods.

244

245 ***Holocene to Mesolithic:*** Pre-Neolithic Europe was inhabited by hunter-gatherers until the arrival  
246 of the first farmers from the Near East (20, 21). There was large diversity among hunter-gatherers  
247 with four main groups— western hunter-gatherers (WHG) that were related to the Villabruna

248 cluster in central Europe, eastern hunter-gatherers (EHG) from Russia and Ukraine related to the  
249 Upper Paleolithic group of Ancestral North Eurasians (ANE) ancestry, Caucasus hunter-gatherers  
250 (CHG) from Georgia associated to the first farmers from Iran, and the GoyetQ2-cluster associated  
251 to the Magdalenian culture in Spain and Portugal (18, 22–25). Most Mesolithic HGs fall on two  
252 main clines of relatedness: one cline that extends from Scandinavia to central Europe showing  
253 variable WHG–EHG ancestry, and the other in southern Europe with WHG–GoyetQ2 ancestry  
254 (23). This ancestry is already present in the 17,000 BCE *El Mirón* individual from Spain,  
255 suggesting that the GoyetQ2-related gene flow occurred well before the Holocene. However, the  
256 WHG–EHG cline was formed more recently during the Mesolithic period, though the precise  
257 timing of the spread of EHG ancestry remains less well understood.

258 To characterize the formation of the WHG–EHG cline, we used genomic data from 16  
259 ancient HG groups ( $n=101$ ) with estimated ages of  $\sim 7,500$ – $3,600$  BCE. We first verified the  
260 ancestry of each HG group using *qpAdm* that compares the allele frequency correlations between  
261 the target and a set of source populations to formally test the model of admixture and then infer  
262 the ancestry proportions for the best fitted model (16). For each target population, we chose the  
263 most parsimonious model, i.e., fitting the data with the minimum number of source populations.  
264 Consistent with previous studies, our *qpAdm* analysis showed that most HGs from Scandinavia,  
265 the Baltic Sea region, and central Europe could be modeled as a two-way mixture of WHG and  
266 EHG-related ancestry (Table S5.1, Note S5). To confirm that the target populations do not harbor  
267 Anatolian farmer-related ancestry (that could lead to some confounding in estimated admixture  
268 dates), we applied *D*-statistics of the form  $D(\text{Mbuti}, \text{target}, \text{WHG}, \text{Anatolian farmers})$  where *target*  
269 = Mesolithic HGs. We observed that none of the target groups had a stronger affinity to Anatolian  
270 farmers than WHG, suggesting that the mixtures we date below reflect pre-Neolithic contacts  
271 between the HGs (Table S5.2).

272 To infer the timing of the mixtures in the history of Mesolithic European HGs, we applied  
273 *DATES* to hunter-gatherers from Scandinavia, the Baltic regions, and central Europe. We inferred  
274 that the earliest admixture occurred in Scandinavian HGs from Norway and Sweden around  $\sim 80$ –  
275 113 generations before the samples lived (Figure SB). Accounting for the average sampling age  
276 of the specimens and the mean human generation time of 28 years (1), this translates to a timing  
277 of admixture of  $\sim 10,200$  BCE for Norway and Sweden Mesolithic individuals, though dates are  
278 more recent ( $\sim 8,000$  BCE) in the Motala HG's. In the Baltic region, we inferred admixture dates  
279 of  $\sim 8,700$ – $6,000$  BCE in Latvia and Lithuania HGs, postdating the mixture in Scandinavia (Figure  
280 3). In southeast Europe, the Iron Gates region of the Danube Basin shows widespread evidence of  
281 mixtures between hunter-gatherer groups and, in the case of some outliers, mixture of hunter-  
282 gatherers and Anatolian farmer-related ancestry as early as the Mesolithic period (26). Further,  
283 these groups showed strong affinity to the WHG-related ancestry in Anatolian populations,  
284 suggesting ancient interactions with Near Eastern populations (26). We applied *qpAdm* to test the  
285 model of admixture in Iron Gates HG and found that the parsimonious model with WHG and EHG  
286 provides a good fit to the data. Further, when we tested the model with Anatolian-related ancestry  
287 using Anatolian HG (AHG) as an additional source population, AHG was not required as the AHG



288 ancestry proportion was not significant (Table S5.1.1 and S5.1.2). Applying *DATES* to Iron Gates  
289 HG with WHG and EHG as reference populations, we inferred this group was genetically formed  
290 in ~10,000–8400 BCE. Our samples of the Iron Gates HGs include a wide range of C14 dates  
291 between 8,800–5,700 BCE. We confirmed our dates were robust to the sampling age of the  
292 individuals as we obtained statistically consistent dates when all samples were combined as one  
293 group or when subsets of samples were grouped in bins of 500 years (Figure SA). The most recent  
294 dates of ~7,500 BCE were inferred in eastern Europe in Ukraine HGs, highlighting how the WHG-  
295 EHG cline was formed over a period ~2000–3000 years (Figure 3, Table SC).

296  
297 **Early to middle Neolithic:** Neolithic farming began in the Near East—the Levant, Anatolia, and  
298 Iran—and spread to Europe and other parts of the world (18, 20, 27). The first farmers of Europe  
299 were related to Anatolian farmers, whose origin remains unclear. The early Neolithic Anatolian  
300 farmers (Aceramic Anatolian farmers) had majority ancestry from AHG with some gene flow from  
301 the first farmers from Iran (26). AHG, in turn, had ancestry from Levant HG (Natufians) and some  
302 mysterious hunter-gatherer group related to the ancestors of WHG individuals from central  
303 Europe—a gene flow event that likely occurred in the late Pleistocene (26). Using *qpAdm*, we  
304 confirmed that early Anatolian farmers could be modeled as a mixture of AHG and Iran Neolithic  
305 farmer-related groups (Note S5). To learn about the timing of the genetic formation of early  
306 Anatolian farmers, we applied *DATES* using one reference group as a set of pooled individuals of  
307 WHG-related and Levant Neolithic farmers-related individuals as a proxy of AHG ancestry and  
308 the second reference group containing pooled Iran Neolithic farmer-related individuals. We note  
309 that the application of *DATES* to three-way admixed groups can lead to intermediate dates between  
310 the first and second pulse of gene flow unless the reference populations are chosen carefully (Table  
311 S2.1). Our setup for early Anatolian farmers should have limited confounding and should recover  
312 the timing of the most recent event (in this case, the gene flow from CHG or Iran Neolithic-related  
313 groups) reliably (Table S2.1). We infer the Iran Neolithic farmer-related gene flow occurred  
314 ~10,900 BCE (12,200–9,600 BCE), predating the origin of farming in Anatolia (28). During the  
315 subsequent millennia, these early farmers further admixed with Levant Neolithic groups to form  
316 Anatolian Neolithic farmers who spread towards the west to Europe and in the east to mix with  
317 Iran Neolithic farmers, forming the Chalcolithic groups of Seh Gabi and Hajji Firuz. Using  
318 *DATES*, we inferred the Chalcolithic groups were genetically formed in ~7,600–5,700 BCE (Table  
319 SC).

320 In Europe, the Anatolian Neolithic farmers mixed with the local indigenous hunter-  
321 gatherers replacing between ~3-50% ancestry of Neolithic Europeans. To elucidate the fine-scale  
322 patterns and regional dynamics of these mixtures, we applied *DATES* to time transect samples  
323 from 94 groups ( $n=657$ ) sampled from sixteen regions in Europe, ranging from ~6,000-1,900 BCE  
324 and encompassing individuals from the early Neolithic to Chalcolithic periods (Table SB). Using  
325 *qpAdm*, we first confirmed that the Neolithic Europeans could be modeled as a mixture of  
326 European hunter-gatherer-related ancestry and Anatolian farmer-related ancestry and inferred their  
327 ancestry proportions (Table SD). For most target populations (~80%), we found the model of gene

328 flow between Anatolian farmer-related and WHG-related ancestry provided a good fit to the data  
329 ( $p$ -value > 0.05). In some populations, we found variation in the source of the HG-related ancestry  
330 and including either EHG or GoyetQ2 improved the fit of the model. In five groups, none of the  
331 models fit, despite excluding outlier individuals whose ancestry profile differed from the majority  
332 of the individuals in the group (Table SD, Table SE). To confirm that the target populations do not  
333 harbor Steppe pastoralist-related ancestry, we applied  $D$ -statistics of the form  $D(\text{Mbuti}, \text{target},$   
334 Anatolian farmers, Steppe pastoralists) where  $\text{target} =$  Neolithic European groups. We observed  
335 that four groups had a stronger affinity to Steppe pastoralists compared to Anatolian farmers, and  
336 hence we excluded these from further analysis (Table SF). After filtering, we applied *DATES* to  
337 86 European Neolithic groups using WHG-related individuals and Anatolian farmers as reference  
338 populations.

339 Earlier analysis has suggested that farming spread along two main routes in Europe, from  
340 southeast to central Europe ('continental route') and along the Mediterranean coastline to Iberia  
341 ('coastal route') (23, 29, 30). Consistent with this, we inferred one of the earliest timings of gene  
342 flow in the Balkans around 6,400 BCE. Using the most comprehensive time-transect in Hungary  
343 with 19 groups ( $n=63$ ) spanning from middle Neolithic to late Chalcolithic, we inferred that the  
344 admixture occurred between ~6,100–4,500 BCE. Under a model of a single shared gene flow event  
345 in the common ancestors of all individuals, we would expect to obtain similar dates of admixture  
346 (before present) after accounting for the age of the ancient specimens. Similar to Lipson et al.  
347 (2017), we observed that the estimated dates in middle Neolithic individuals were substantially  
348 older than those inferred in late Neolithic or Chalcolithic individuals (Figure 3). This would be  
349 expected if the underlying model of gene flow involved multiple pulses of gene flow, such that the  
350 timing in the middle Neolithic samples reflects the initial two-way mixture and the timing in the  
351 Chalcolithic samples captures both recent and older events. Interestingly, Lipson et al. (2017) and  
352 other recent studies have documented increasing HG ancestry from ~3-15% from the Neolithic to  
353 Chalcolithic period (16, 23, 31), suggesting that there was additional HG gene flow after the initial  
354 mixture. This highlights that the interactions between local hunter-gatherers and incoming  
355 Anatolian farmers were complex with multiple gene flow events between these two groups, which  
356 explains the increasing HG ancestry and more recent dates in Chalcolithic individuals (Table SD).

357 Mirroring the pattern in Hungary, we documented the resurgence of HG ancestry in the  
358 Czech Republic, France, Germany, and southern Europe. In central Europe, we inferred that the  
359 Anatolian farmer-related gene flow occurred ~5,600-5,000 BCE, with some exceptions. In the  
360 Blätterhöhle site from Germany, we inferred the gene flow occurred more recently (~4,000 BCE),  
361 consistent with the occupation of both hunter-gatherers and farmers in this region until the late  
362 Neolithic (31). In eastern Europe, using samples related to the Funnel Beaker culture (TRB; from  
363 German *Trichterbecher*) from Poland, we dated the Anatolian farmer-related gene flow occurred  
364 ~5,300–4,200 BCE. Following the TRB decline, the Baden culture and the Globular Amphora  
365 culture appeared in many areas of Poland and Ukraine (25). These cultures had close contacts with  
366 Corded Ware complex and steppe societies, though we did not find any evidence of Steppe  
367 pastoralist-related ancestry in the GAC individuals (Table SD). Applying *DATES*, we inferred the

368 Anatolian farmer-related and HG mixture occurred ~5,200-3,100 BCE, predating the spread of  
369 Steppe pastoralists to eastern Europe (16, 19).

370 Along the Mediterranean route, we characterized Anatolian farmer-related gene flow in  
371 Italy, Iberia, France, and the British Isles. Using samples from five groups in Italy, we inferred the  
372 earliest dates of Anatolian farmer-related gene flow of ~6,100 BCE, and within the millennium,  
373 the ancestry spread from Sardinia to Sicily (Figure 3). In Iberia, the Anatolian farmer-related  
374 mixture occurred ~6,000–3,400 BCE and showed evidence for an increase in HG ancestry from  
375 ~9–20% after the initial gene flow. In France, previous studies have shown that Anatolian farmer-  
376 related ancestry came from both routes, along the Danubian in the north and along the  
377 Mediterranean in the south (23). This is reflected in the source of the HG ancestry, which is  
378 predominantly EHG and WHG-related in the north and includes WHG and Goyet-Q2 ancestry in  
379 the south (23). Consistently, we also observed that the admixture dates in France were structured  
380 along these routes, with the median estimate of ~5,100 BCE in the east and much older ~5,500  
381 BCE in the south (Table SC). In Scandinavia, we inferred markedly more recent dates of admixture  
382 of ~4,300 BCE using samples from Sweden associated with the TRB culture and Ansarve  
383 Megalithic tombs, consistent with a late introduction of farming to Scandinavia (33).

384 Finally, we inferred recent dates of admixture in Neolithic samples from the British Isles  
385 (England, Scotland, and Ireland) with the median timing of ~5,000 BCE across the three regions.  
386 Interestingly, unlike in western and southern Europe, there was no resurgence in HG ancestry  
387 during the Neolithic in Britain (34). This suggests our dates can be interpreted as the time of the  
388 main mixture of HGs and Anatolian farmers in this region, implying that the farmer-related  
389 ancestry reached Britain a millennium after its arrival in continental Europe. By 4,300 BCE, we  
390 find that Anatolian farmer-related ancestry is present in nearly all regions in Europe.

391  
392 **Late Neolithic to Bronze Age:** The beginning of the Bronze Age was a period of major cultural  
393 and demographic change in Eurasia, accompanied by the spread of Yamnaya Steppe Pastoralist-  
394 related ancestry from Pontic-Caspian steppes into Europe and South Asia (16). The archaeological  
395 record documents that the early Steppe pastoralists cultures of Yamnaya and Afanasievo, with  
396 characteristic burial styles and pottery, appeared around ~3,300 to 2,600 BCE (35). These groups  
397 were likely the result of a genetic admixture between the descendants of EHG-related groups and  
398 CHG-related groups associated with the first farmers from Iran (8, 22, 36). Using *qpAdm*, we first  
399 tested how well this model fits the data from 8 early Steppe pastoralist groups, including seven  
400 groups associated with Yamnaya culture and one group related to the Afanasievo culture  
401 (Methods). For all but two Yamnaya groups (from Hungary Baden and Russia Kalmykia), we  
402 found this model provides a good fit to the data (Table S5.4). We note that the samples from  
403 Kalmykia in our dataset were shotgun sequenced, and in the *qpAdm* analysis, we are mixing  
404 shotgun and capture data that could potentially lead to technical issues. To understand the timing  
405 of the formation of the early Steppe pastoralist-related groups, we applied *DATES* using pooled  
406 EHG and pooled Iranian Neolithic farmers. Focusing on the groups with the largest sample sizes,  
407 Yamnaya Samara ( $n=10$ ) and Afanasievo ( $n=19$ ), we inferred the admixture occurred between 40–

408 45 generations before the individuals lived, translating to an admixture timing of ~4,100 BCE  
409 (Table S6.1). We obtained qualitatively similar dates across four Yamnaya and one Afanasievo  
410 groups, consistent with the findings that these groups descend from a recent common ancestor (for  
411 Oзера samples from Ukraine, the dates were not significant). This is also further supported by the  
412 insight that the genetic differentiation across early Steppe pastoralist groups is very low ( $F_{ST} \sim$   
413 0.000-0.006) (Table S6.2). Thus, we combined all early Steppe pastoralist individuals in one group  
414 to obtain a more precise estimate for the genetic formation of proto-Yamnaya of ~4,400 to 4,000  
415 BCE (Figure 2). These dates are noteworthy as they pre-date the archaeological evidence by more  
416 than a millennium (37) and have important implications for understanding the origin of proto-  
417 Pontic Caspian cultures and their spread to Europe and South Asia.

418 Over the following millennium, the Yamnaya-derived groups of the Corded Ware Complex  
419 (CWC) and Bell Beaker complex (BBC) cultures brought Steppe pastoralist-related ancestry to  
420 Europe. Present-day Europeans derive between ~10-60% Steppe pastoralist-related ancestry,  
421 which was not seen in Neolithic samples. To obtain a precise chronology of the spread of Steppe  
422 pastoralist-related ancestry across Europe, we analyzed 109 late Neolithic, Chalcolithic, and BA  
423 samples dated between 3,000-750 CE from 18 regions, including samples associated with the  
424 CWC and BBC cultures. We first confirmed that most target samples had Steppe pastoralist-related  
425 ancestry, in addition to European HG-related and Anatolian farmer-related ancestry using *qpAdm*.  
426 We excluded 20 groups that could not be parsimoniously modeled as a three-way mixture even  
427 after removing individual outliers. After filtering, we retained 79 groups for dating Steppe  
428 pastoralist-related gene flow across Europe (Note S5, Table SH). As Bronze Age Europeans have  
429 ancestry from three distinct groups, we applied *DATES* using the following two reference  
430 populations, one group including early Steppe pastoralists (Yamnaya and Afanasievo) and the  
431 other group with pooled samples of WHG-related and Anatolian farmer-related individuals, which  
432 is the proxy for the ancestral Neolithic Europe population.

433 To learn about the spread of CWC culture across Europe, we used seven late Neolithic and  
434 Bronze age groups, including five associated with CWC artifacts. Using *DATES*, we inferred that  
435 the oldest date of Steppe pastoralists gene flow in Europe was ~3,200 BCE in Scandinavia in  
436 samples associated with Battle Axe Culture in Sweden and Single Grave Culture in Denmark that  
437 were both contemporary to CWC. The samples from Scandinavia showed large heterogeneity in  
438 ancestry, including some individuals with majority Steppe pastoralist-related ancestry (and  
439 negligible amounts of Anatolian farmer-related ancestry), consistent with patterns expected from  
440 recent gene flow (38). Strikingly, we inferred the timing of admixture in central Europe (Germany  
441 and the Czech Republic) and eastern Europe (Estonia and Poland) to be remarkably similar. These  
442 dates fall within a narrow range of ~3,000–2,900 BCE across diverse regions, suggesting that the  
443 mixed population associated with the Corded Ware culture formed over a short time and spread  
444 across Europe rapidly with very little further mixture (Table SC).

445 Following the Corded Ware culture, from around 2,800 to 2,300 BCE, Bell Beaker pottery  
446 became widespread across Europe (39). Using 19 Chalcolithic and Bronze Age samples, including  
447 ten associated with Beaker-complex artifacts, we inferred the dynamics of the spread of the Beaker

448 complex across Europe. We inferred the oldest date of Steppe pastoralist-related admixture was  
449 ~3,200 BCE (3600–2800 BCE) in EBA Mallorca samples from Iberia. We note the EBA Mallorca  
450 sample is not directly associated with Beaker culture, but *qpAdm* modeling suggests that this  
451 individual is clade with the small subset of Iberian Beaker-complex-associated individuals who  
452 carried Steppe pastoralist-related (40). Most individuals from Iberia, however, had negligible  
453 Steppe pastoralist-related ancestry suggesting the Beaker culture was not accompanied by major  
454 gene flow in Iberia despite the earliest dates (Table SH). In central and western Europe, where  
455 steppe gene flow was more pervasive, we inferred the median date of the mixture was ~2,700 BCE  
456 with the oldest dates in the Netherlands, followed by Germany and France (Figure 3). There was,  
457 however, large heterogeneity in the dates across Europe and even within the same region. For  
458 example, comparing two BA groups from the Netherlands suggests a wide range of dates ~3,000  
459 BCE and 2,500 BCE, and four groups from Germany indicate a range of ~2,900–2,700 BCE. From  
460 central Europe, the Steppe pastoralist-related ancestry spread quickly to the British Isles, where  
461 people with steppe ancestry replaced 90% of the genetic ancestry of individuals from Britain. Our  
462 estimates for the time of gene flow in Bell Beakers samples from England suggest that the gene  
463 flow occurred ~2,700 BCE (2770-2550 BCE). Our estimated dates of admixture are older than the  
464 dates of arrival of this ancestry in Britain (41) and, interestingly, overlap the dates in central  
465 Europe. Given that a significant fraction of the Beaker individuals were recent migrants from  
466 central Europe, we interpret our dates reflect the admixture into ancestors of the British Beaker  
467 people, occurring in mainland Europe (41).

468 The middle to late Bronze age led to the final integration of Steppe pastoralist-related  
469 ancestry in Europe. In southern Europe, early BA samples had limited Steppe pastoralist-related  
470 ancestry, though present-day individuals have between ~5–30% steppe ancestry (16). Using  
471 pooled samples of middle to late BA from Spain, we inferred major mixture occurred ~2,500 BCE  
472 in Iberia. We inferred a similar timing in Italy using individuals associated with the Bell Beaker  
473 culture and early BA samples from Sicily (Table SC). In Sardinia, a majority of the BA samples  
474 do not have Steppe pastoralist-related ancestry. In a few individuals, we found evidence for steppe  
475 ancestry, though in most cases, the Steppe pastoralist-related ancestry proportion overlapped 0,  
476 and the dates were very noisy (Table SH). Using Iron Age samples from Sardinia, we inferred the  
477 gene flow occurred ~2,600 BCE, though there is large uncertainty associated with this estimate  
478 (2,614 +/- 560 BCE). In other parts of continental Europe and the British Isles, the Steppe  
479 pastoralist-related gene flow got diluted over time, as evidenced by more recent dates in LBA than  
480 EBA or MBA samples in Germany, England, and Scotland, and increase in Neolithic farmer  
481 ancestry during this period (42) (Table SC).

482 Finally, the Corded Ware Complex expanded to the east to form the archaeological  
483 complexes of Sintashta, Srubnaya, Andronovo, and the Bronze Age cultures of Kazakhstan.  
484 Samples associated with these cultures harbor mixed ancestry from the Yamnaya Steppe  
485 pastoralist-related groups (CWC, in some cases) and Neolithic individuals from central Europe  
486 (Table S5.5) (8). Applying *DATES* to 8 Middle to late Bronze Age (MLBA) Steppe pastoralist  
487 groups, we inferred the precise timing for the formation of these groups beginning in the third

488 millennium BCE. These groups were formed chronologically, with the date of genetic formation  
489 of ~3,200 BCE for Sintashta culture, followed by ~2,900 BCE for Srubnaya and Andronovo  
490 cultures. In the central Steppe region (present-day Kazakhstan), we obtained median dates of  
491 ~2,800 BCE for the expansion of Steppe pastoralist-related ancestry in four Kazakh cultures of  
492 Maitan Alakul, Aktogai, and Kairan. By ~2,700 BCE, most of these cultures had almost 60-70%  
493 Yamnaya Steppe pastoralist-related ancestry (Table SC). These groups, in turn, expanded  
494 eastwards, transforming the genetic composition of populations in South Asia.

## 495 **Discussion**

496  
497 We developed *DATES*, a novel method to measure ancestry covariance in a single diploid  
498 individual genome to estimate the time of admixture. Using extensive simulations, we show that  
499 *DATES* provides accurate estimates of the timing of admixture for a range of demographic  
500 scenarios. Application of *DATES* to present-day samples shows that the results are concordant with  
501 published methods—Rolloff, ALDER, and Globetrotter. For sparse datasets, *DATES* outperforms  
502 published methods as it does not require phased data and works reliably with limited samples,  
503 large proportions of missing variants as well as pseudodiploid genotypes. This makes *DATES*  
504 ideally suited for the analysis of ancient DNA samples.

505 We illustrate the application of *DATES* by reconstructing population movements and  
506 admixtures during the European Holocene. The European continent was subject to two major  
507 migrations during the Holocene: the movement of Anatolian farmers during the Neolithic and the  
508 migration of Yamnaya Steppe pastoralists during the Bronze Age. First, we document that the  
509 Mesolithic hunter-gatherers formed as a mixture of WHG and EHG ancestry ~10,200 to 7400  
510 BCE. These dates are consistent with the archeological evidence for the appearance of lithic  
511 technology associated with eastern HGs in Scandinavia and the Baltic regions and the spread of  
512 WHG ancestry to east (17, 43, 44). Next, we studied the timing of the genetic formation of  
513 Anatolian farmers. The earliest evidence of agriculture comes from the Fertile Crescent, the  
514 southern Levant, and the Zagros Mountains of Iran and dates to around 10,000 BCE. In central  
515 Anatolia, farming has been documented c. 8,300 BCE (45, 46). It has been long debated if  
516 Neolithic farming groups from Iran and the Levant introduced agriculture to Anatolia or hunter-  
517 gatherers in the region locally adopted agricultural practices. The early Anatolian farmers can be  
518 modeled as a mixture of local hunter-gatherers people related to Caucasus hunter-gatherers or first  
519 farmers from Iran (26). By applying *DATES* (assuming single instantaneous admixture), we  
520 inferred that the Iran Neolithic gene flow occurred around 10,900 BCE (~12,200–9,600 BCE). An  
521 alternate possibility is that there was a long period of gradual gene flow between the two groups  
522 and our dates reflect intermediate dates between the start and end of the gene flow. An upper bound  
523 for such mixture comes from the lack of Iran Neolithic ancestry in Anatolian HGs at 13,000 BCE,  
524 and a lower bound comes from the C14 dates of early Anatolian farmers, one of which is directly  
525 dated at 8269–8210 BCE (26). In either case (instantaneous admixture or gradual gene flow), the  
526 genetic mixture that formed Anatolian farmers predates the advent of agriculture in this region.  
527 This supports the model that Anatolian hunter-gatherers locally transitioned to agricultural

528 subsistence, and most probably, there was cultural diffusion from other regions in Near East (Iran  
529 and Levant) (26). Future studies with more dense temporal sampling will shed light on the  
530 demographic processes that led to the transition from foraging to farming in the Near East, and in  
531 turn, elucidate the relative roles of demic and cultural diffusion in the dispersal of technologies  
532 like agriculture across populations.

533 Using data from sixteen regions in Europe, we reconstruct a detailed chronology and  
534 dynamics of the expansion and admixture of Anatolian farmers during the Neolithic period. We  
535 infer that starting in ~6,400 BCE, gene flow from Anatolian farmers became widespread across  
536 Europe, and by ~4,300 BCE, it was present in almost all parts of continental Europe and the British  
537 Isles. These dates are significantly more recent than the estimates of farming based on  
538 archaeological evidence in some parts of Europe, suggesting that the local hunter-gatherers and  
539 farmers co-existed for more than a millennium before the mixture occurred (16, 31). In many  
540 regions, after the initial mixture, there was a resurgence of HG ancestry, highlighting the  
541 complexities of these ancient interactions. We note that our results are consistent with two previous  
542 genetic studies, Lipson et al. (2017) and Rivollat et al. (2020), that applied genetic dating methods  
543 to a subset of samples we used in our analysis. Lipson et al. (2017) used a modified version of  
544 ALDER to infer the timing of admixture in three regions ( $n=151$ ), and we obtained statistically  
545 consistent results for all overlapping samples (within two standard errors). An advantage of our  
546 approach over the modified ALDER approach is that we do not rely on helper samples (higher  
547 coverage individuals combined with the target group) for dating; unless these have a similar  
548 ancestry profile, they could bias the inferred dates. Our results are concordant with Rivollat et al.  
549 (2020) that used a previous version of *DATES* to infer the timing of Neolithic gene flow in 32  
550 groups (vs. 86 groups in our study). We find the performance of both versions of *DATES* is similar,  
551 though some implementation details have improved (see Note S3, Table S3.3).

552 The second major migration occurred when populations associated with the Yamnaya  
553 culture in the Pontic-Caspian steppe expanded to central and western Europe from far eastern  
554 Europe. Our analysis reveals the precise timing of the genetic formation of these early Steppe  
555 pastoralists groups—Yamnaya and Afanasievo—occurred ~4,400-4,000 BCE. This estimate  
556 predates the archaeological evidence by more than a millennium (37) and suggests the presence of  
557 an ancient “ghost” population of proto-Yamnaya around this time. Understanding the source and  
558 location of this ghost population will provide deep insights into the history of Pontic-Caspian  
559 cultures and the origin of Indo-European languages that have been associated to have spread with  
560 Steppe pastoralists ancestry to Europe and South Asia (16, 47). Starting in ~3,200 BCE, the  
561 Yamnaya-derived cultures of Corded Ware Complex and Bell Beaker complex spread westwards,  
562 bringing steppe ancestry to Europe. Our analysis reveals striking differences in the spread of these  
563 three cultures: the Yamnaya were genetically formed a millennium before the evidence for  
564 pastoralism, while CWC formation is coincident with the archaeological dates and similar across  
565 diverse regions separated by thousands of kilometers, suggesting a rapid spread after the initial  
566 formation of this group. In contrast, the formation and expansion of people with Steppe pastoralist-  
567 related ancestry associated with Bell Beakers cultural artifacts are much more complex and

568 heterogeneous across regions. We find the earliest evidence of Steppe pastoralist-related ancestry  
569 in Iberia around 3200 BCE, though this ancestry only becomes widespread after 2,500 BCE. In  
570 central Europe, the gene flow occurred simultaneously with archaeological evidence and was  
571 coexisting with the Corded Ware complex in some parts (41, 48). Finally, in the British Isles, the  
572 Bell Beaker culture spreads rapidly from central Europe and replaces almost 90% of the ancestry  
573 of individuals in this region (41).

574 Recent analysis has shown remarkable parallels in the history of Europe and South Asia;  
575 with both groups deriving ancestry from local indigenous HGs, Near Eastern farmers, and Steppe  
576 pastoralist-related groups (8). Interestingly, however, the timing of the two major migrations  
577 events differs across the two subcontinents. Both mixtures occurred in Europe almost a millennium  
578 before they occurred in South Asia. In Europe, the Neolithic migrations primarily involved  
579 Anatolian farmers, while the source of Neolithic ancestry is closer to Iran Neolithic farmers in  
580 South Asia. The Steppe pastoralist-related gene flow occurred in the context of the spread of CWC  
581 and BBC cultures in Europe around 3,200-2,500 BCE; in South Asia, this ancestry arrived with  
582 Steppe MLBA cultures in 1,800-1,500 BCE (8). The Steppe MLBA groups were genetically  
583 formed as an admixture of Steppe pastoralist-derived groups and European Neolithic farmers  
584 following the eastward expansion of CWC groups between ~3,200–2,700 BCE. Understanding the  
585 origin and migration paths of the ancestral groups thus helps to illuminate the differences in the  
586 timeline of the spread of steppe genetics across the two subcontinents of Eurasia.

587 Genomic dating methods like *DATES* provide an independent and complementary  
588 approach for reconstructing population history. By focusing on genetic clocks like recombination  
589 rate, we provide an independent estimate of the timing of evolutionary events up to several  
590 thousands of years. Our analysis also has advantages over temporal sampling of ancient DNA, in  
591 that we can obtain direct estimates of when a population was formed, rather than inferring putative  
592 bounds for the timing based on the absence/presence of a particular ancestry signature (which may  
593 be sensitive to sampling choice and density). Genetic approaches provide complementary evidence  
594 to archaeology and linguistics as they date the time of gene flow and not migration. Both dates are  
595 similar in many contemporary populations like African Americans and Latinos, though this may  
596 not be generally true (2). This is underscored by our dates for the Neolithic farmer mixture, which  
597 post-dates evidence of material culture related to agriculture by almost two millennia in some  
598 regions. This suggests that European HGs and farmers resided side by side for several thousand  
599 years before gene exchange (49, 50). This highlights how genetic dates can provide  
600 complementary evidence to archaeology and help to build a comprehensive picture of population  
601 origins and movements.

## 602 **Methods and Materials**

603

### 604 **Dataset**

605

606 We analyzed 1,096 ancient European samples from 152 groups restricting to data from 1,233,013  
607 autosomal SNP positions that were genotyped using the Affymetrix Human Origins array (the



608 V44.3 release of the Allen Ancient DNA Resource (AADR); <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>). We  
609 filtered this dataset to remove samples that were marked as contaminated, low coverage, outliers,  
610 duplicates or first- or second-degree relatives (Table SB). We grouped individuals together from  
611 a particular culture or region. Details of sample affiliation and grouping used is described in Table  
612 SA.  
613

614

### 615 **Modeling admixture history**

616

617 We applied *qpAdm* from ADMIXTOOLS to identify the best fitting model and estimate the  
618 ancestry proportions in a target population modeled as a mixture of  $n$  “reference” populations using  
619 a set of “Outgroup” populations (16). We set the details: YES parameter, which reports a normally  
620 distributed Z-score to evaluate the goodness of fit of the model (estimated with a Block Jackknife).  
621 For each target population, we chose the most parsimonious model, i.e., fitting the data with the  
622 minimum number of source populations. We excluded models where the  $p$ -value  $< 0.05$  indicating  
623 a poor fit to the data. Details of the *qpAdm* analysis for each group are reported in Note S5. We  
624 also applied  $D$ -statistics in some cases using *qpDstat* in ADMIXTOOLS with default parameters.  
625

626

### 626 **Dating admixture events**

627

628 We applied *DATES* to infer the time of admixture for a given target population. We present the  
629 details of the model and implementation in Note S1. We applied *DATES* using genome-wide SNP  
630 data from the target population and two reference populations. To infer the allele frequency in the  
631 ancestral populations more reliably, where specified, we pooled individuals deriving the majority  
632 of their ancestry from the population of interest (Table SA). We computed the weighted ancestry  
633 covariance between 0.45cM (to minimize the impact of background LD) to 100 cM, with a bin  
634 size of 0.1 cM. We plotted the weighted covariance with genetic distance and obtained a date by  
635 fitting an exponential function with an affine term  $y = Ae^{-\lambda d} + c$ , where  $d$  is the genetic distance  
636 in Morgans and  $\lambda = (t+1)$  is the number of generations since admixture ( $t$ ). The factor of  $(t+1)$  is  
637 because in the first-generation following admixture, the admixed population derives one  
638 chromosome from each ancestral group. The mixing of chromosomes only begins in the following  
639 generations as the chromosomes recombine. We computed standard errors using weighted block  
640 jackknife, where one chromosome was removed in each run (51). We examined the quality of the  
641 exponential fit by computing the normalized root-mean-square deviation (NRMSD) between the  
642 empirical ancestry covariance values  $z$  and the fitted ones  $\hat{z}$ , across all the genetic distance bins  
643 (11).

644 The estimated dates of admixture were considered significant if the Z-score  $> 2$ ,  $\lambda < 200$   
645 generations and NRMSD  $< 0.7$ . We converted the inferred dates from generations to years by  
646 assuming a mean generation time of 28 years (1). For ancient samples, we added the sampling age  
647 of the ancient specimen (Table SA). When multiple individuals were available, we used the

648 average sampling ages to offset the admixture dates. We report dates in BCE by assuming the 1950  
649 convention.

650

### 651 **Software availability**

652

653 The executable and source code for *DATES* will be available on GitHub:

654 [https://github.com/MoorjaniLab/DATES\\_v3600](https://github.com/MoorjaniLab/DATES_v3600)

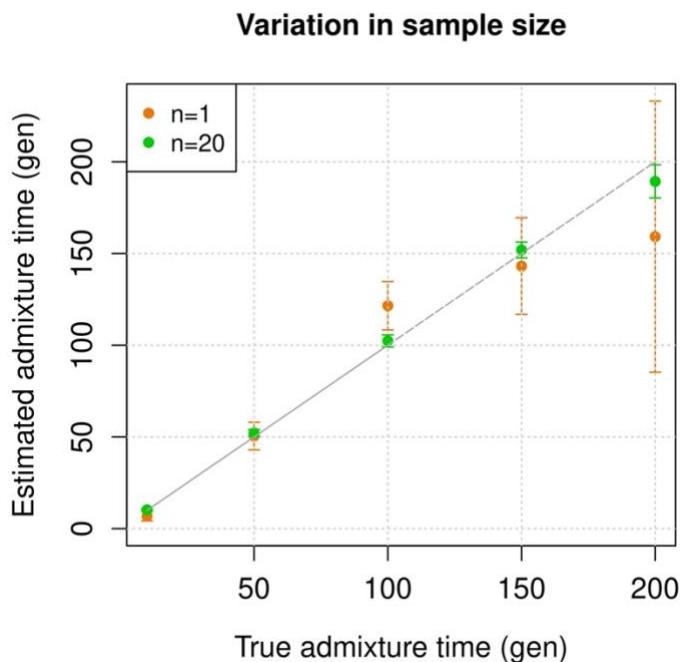
655

### 656 **Acknowledgments**

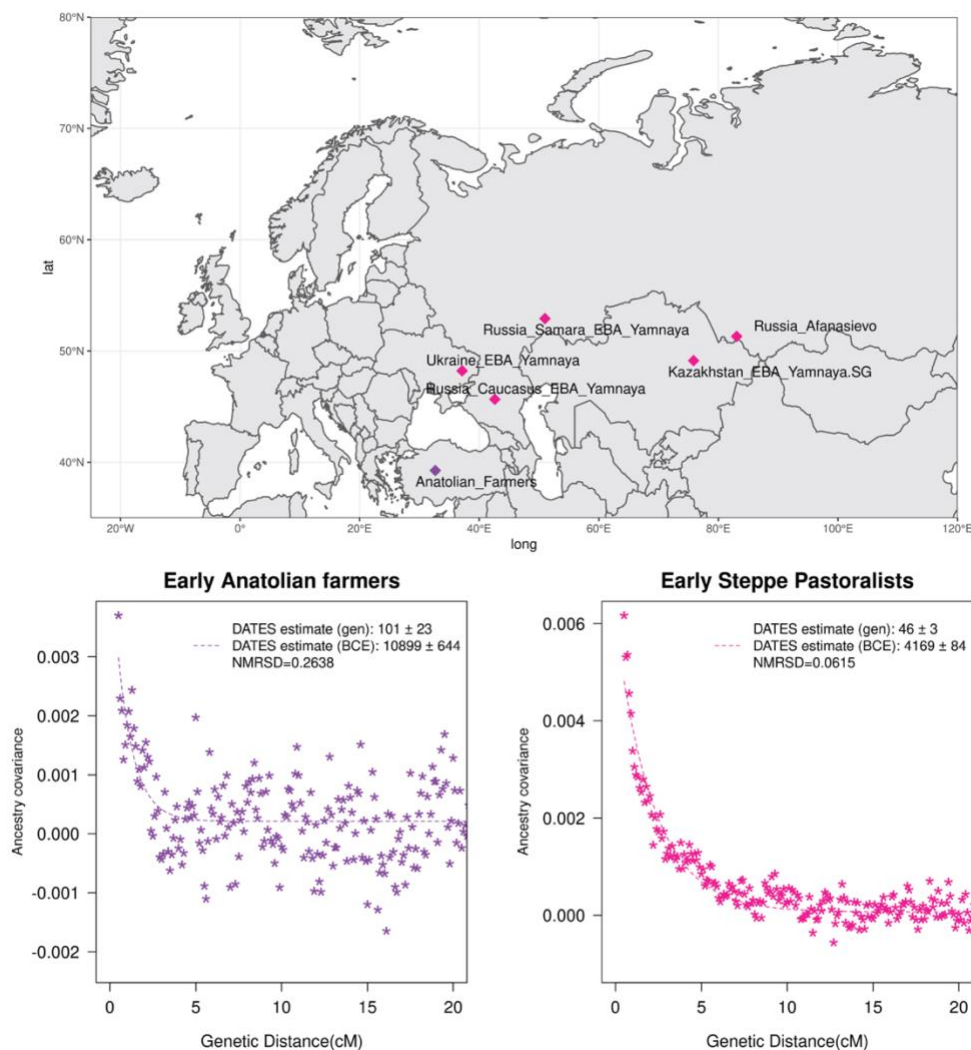
657

658 We thank Monty Slatkin, Ziyue Gao, David Reich, Iosif Lazaridis, and Vagheesh Narasimhan for  
659 their comments on the manuscript. We thank Iosif Lazaridis for helpful discussions about  
660 population models in the Near East, Remi Tournabize for suggestions for evaluating the fit of  
661 exponential decay curves, and Neel Alex for suggestions for implementation of FFT for an earlier  
662 version of *DATES*. We are grateful for support from the Burroughs Wellcome Fund Careers at the  
663 Scientific Interface, Sloan Research Fellowship, and NIH R35GM142978 awarded to PM. NP was  
664 a fellow at the Radcliffe Institute for Advanced Study at Harvard University.

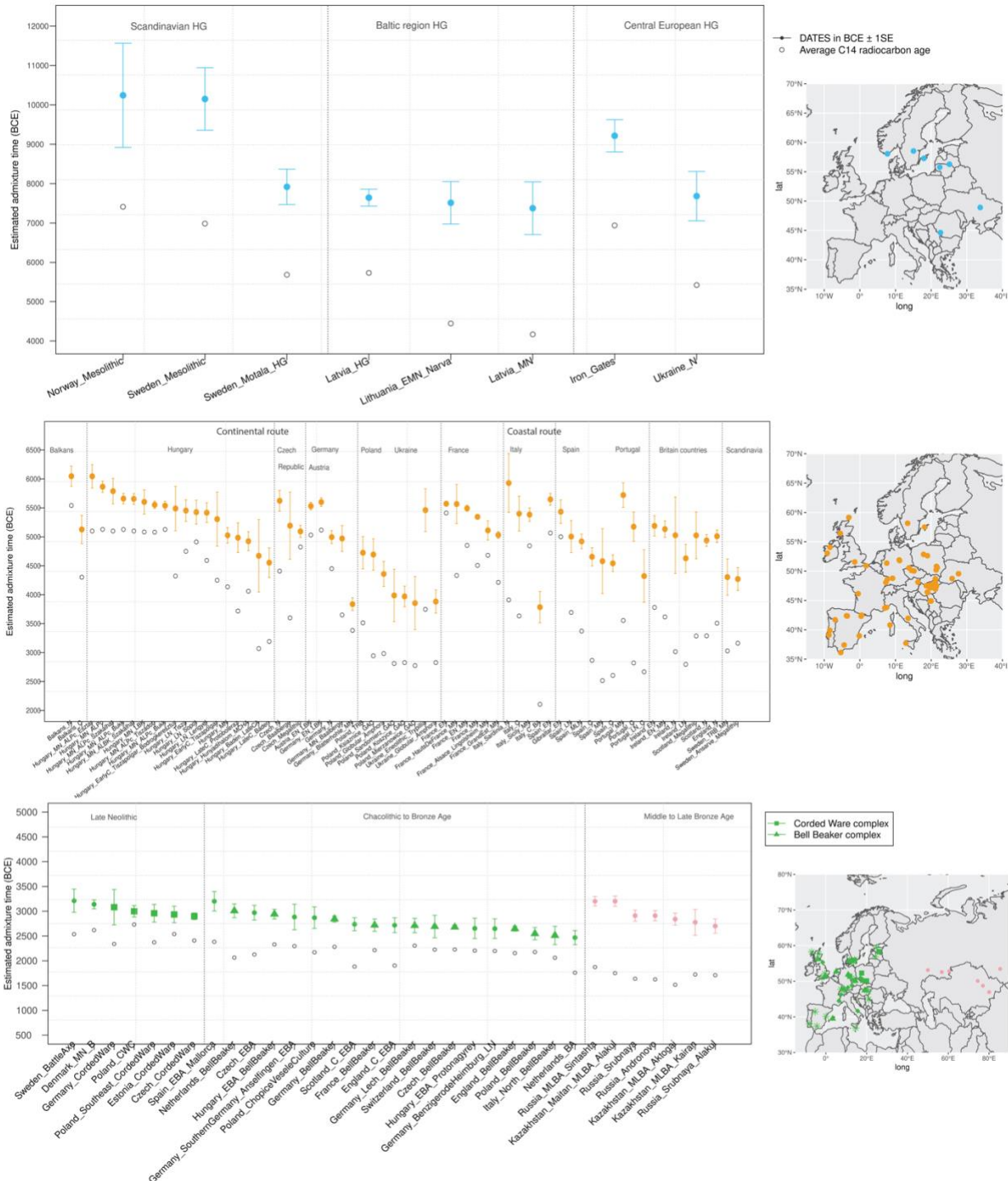
665 **Figures**  
666



667  
668 **Figure 1: Simulation results.** We constructed admixed individuals with 20% European (CEU)  
669 and 80% Africa (YRI) ancestry for admixture dates ranging between 10–200 generations where  
670 the sample size of the target group is as shown in the legend. We applied *DATES* using French and  
671 Yoruba as reference populations. We show the true time of admixture (X-axis, in generations) and  
672 the estimated time of admixture ( $\pm 1$  SE) (Y-axis, in generations). Standard errors were calculated  
673 using a weighted block jackknife approach by removing one chromosome in each run (Methods).



674  
 675 **Figure 2: Genetic formation of early Anatolian farmers and early Bronze Age Steppe**  
 676 **pastoralists.** The top panel shows a map with sampling locations of the target groups analyzed for  
 677 admixture dating. The bottom panels show the inferred times of admixture for each target using  
 678 *DATES* by fitting an exponential function with an affine term  $y = Ae^{-\lambda d} + c$ , where  $d$  is the  
 679 genetic distance in Morgans and  $\lambda = (t+1)$  is the number of generations since admixture ( $t$ )  
 680 (Methods). We start the fit at a genetic distance ( $d$ ) > 0.5cM to minimize confounding with  
 681 background LD and estimate a standard error by performing a weighted block jackknife removing  
 682 one chromosome in each run. For each target, in the legend, we show the inferred average dates of  
 683 admixture ( $\pm 1$  SE) in generations before the individual lived, in BCE accounting for the average  
 684 age of all the individuals and the mean human generation time, and the NRMSD values to assess  
 685 the fit of the exponential curve (Methods). The bottom left shows the ancestry covariance decay  
 686 curve for early Anatolian farmers inferred using one reference group as a set of pooled individuals  
 687 of WHG-related and Levant Neolithic farmers-related individuals as a proxy of AHG ancestry and  
 688 the second reference group containing Iran Neolithic farmer-related individuals. The bottom right  
 689 shows the ancestry covariance decay curve for early Steppe pastoralists groups, including all  
 690 Yamnaya and Afanasievo individuals as the target group and EHG-related and Iran Neolithic  
 691 farmer-related groups as reference populations.  
 692



693  
694  
695  
696  
697  
698  
699  
700

**Figure 3: Timeline of admixture events in ancient Europe.** We applied *DATES* to ancient samples from Europe. In the right panel, we show the sampling locations of the ancient specimens, and in the left panel, we show the admixture dates for each target group listed on the X-axis. The inferred dates in generations were converted to dates in BCE by assuming a mean generation time of 28 years (4) and accounting for the average sampling age (shown as grey dots) of all ancient individuals in the target group (Methods). The top panel shows the formation of WHG–EHG cline (in blue) using Mesolithic hunter-gatherers as the target and EHG and WHG as reference

701 populations. The middle panel shows admixture dates of local HGs and Anatolian farmers (in  
702 orange) using Neolithic European groups as targets and Anatolian farmers-related groups and  
703 WHG-related groups as reference populations. The bottom panel shows the spread of Steppe  
704 pastoralist-related ancestry (in green) estimated using middle and late Neolithic, Chalcolithic, and  
705 Bronze Age samples from Europe as target populations and early Steppe pastoralist-related groups  
706 (Afanasievo and Yamnaya Samara) and a set of Anatolian farmers and WHG-related groups as  
707 reference populations. For the middle to late Bronze Age samples from Eurasia, we used the early  
708 Steppe pastoralist-related groups and the Neolithic European groups as reference populations. The  
709 cultural affiliation (CWC, BBC, or Steppe MLBA cultures) of the individuals is shown in the  
710 legend.

711 **References**

- 712
- 713 1. P. Moorjani, *et al.*, A genetic method for dating ancient genomes provides a direct estimate  
714 of human generation interval in the last 45,000 years. *Proc. Natl. Acad. Sci.* **113**, 5652  
715 (2016).
- 716 2. G. Hellenthal, *et al.*, A Genetic Atlas of Human Admixture History. *Science* **343**, 747–751  
717 (2014).
- 718 3. R. Chakraborty, K. M. Weiss, Admixture as a tool for finding linked genes and detecting  
719 that difference from allelic association between loci. *Proc. Natl. Acad. Sci.* **85**, 9119–9123  
720 (1988).
- 721 4. P. Moorjani, *et al.*, The History of African Gene Flow into Southern Europeans, Levantines,  
722 and Jews. *PLoS Genet.* **7**, e1001373 (2011).
- 723 5. P.-R. Loh, *et al.*, Inferring Admixture Histories of Human Populations Using Linkage  
724 Disequilibrium. *Genetics* **193**, 1233–1254 (2013).
- 725 6. S. Gravel, Population Genetics Models of Local Ancestry. *Genetics* **191**, 607–619 (2012).
- 726 7. A. Auton, *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- 727 8. V. M. Narasimhan, *et al.*, The formation of human populations in South and Central Asia.  
728 *Science* **365**, eaat7487 (2019).
- 729 9. N. A. Rosenberg, *et al.*, Genetic Structure of Human Populations. *Science* **298**, 2381  
730 (2002).
- 731 10. N. Patterson, *et al.*, Ancient Admixture in Human History. *Genetics* **192**, 1065–1093  
732 (2012).
- 733 11. R. Tournebize, G. Chu, P. Moorjani, Reconstructing the history of founder events using  
734 genome-wide patterns of allele sharing across individuals. *bioRxiv*, 2020.09.07.286450  
735 (2020).
- 736 12. D. M. Behar, *et al.*, The genome-wide structure of the Jewish people. *Nature* **466**, 238–242  
737 (2010).
- 738 13. B. M. Henn, *et al.*, Genomic Ancestry of North Africans Supports Back-to-Africa  
739 Migrations. *PLoS Genet.* **8**, e1002397 (2012).
- 740 14. L. Pagani, V. Colonna, C. Tyler-Smith, Q. Ayub, An Ethnolinguistic and Genetic  
741 Perspective on the Origins of the Dravidian-Speaking Brahui in Pakistan. **13** (2017).
- 742 15. A. L. Price, *et al.*, Sensitive Detection of Chromosomal Segments of Distinct Ancestry in  
743 Admixed Populations. *PLoS Genet.* **5**, e1000519 (2009).

- 744 16. W. Haak, *et al.*, Massive migration from the steppe was a source for Indo-European  
745 languages in Europe. *Nature* **522**, 207–211 (2015).
- 746 17. I. Mathieson, *et al.*, The genomic history of southeastern Europe. *Nature* **555**, 197–203  
747 (2018).
- 748 18. P. Skoglund, *et al.*, Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers  
749 in Europe. *Science* **336**, 466–469 (2012).
- 750 19. M. E. Allentoft, *et al.*, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172  
751 (2015).
- 752 20. W. Haak, *et al.*, Ancient DNA from European Early Neolithic Farmers Reveals Their Near  
753 Eastern Affinities. *PLoS Biol.* **8**, e1000536 (2010).
- 754 21. Z. Hofmanová, *et al.*, Early farmers from across Europe directly descended from Neolithic  
755 Aegeans. *Proc. Natl. Acad. Sci.* **113**, 6886–6891 (2016).
- 756 22. E. R. Jones, *et al.*, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat.*  
757 *Commun.* **6**, 8912 (2015).
- 758 23. M. Rivollat, *et al.*, Ancient genome-wide DNA from France highlights the complexity of  
759 interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci. Adv.* **6**,  
760 eaaz5344 (2020).
- 761 24. Q. Fu, *et al.*, The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
- 762 25. D. M. Fernandes, *et al.*, A genomic Neolithic time transect of hunter-farmer admixture in  
763 central Poland. *Sci. Rep.* **8**, 14879 (2018).
- 764 26. M. Feldman, *et al.*, Late Pleistocene human genome suggests a local origin for the first  
765 farmers of central Anatolia. *Nat. Commun.* **10**, 1218 (2019).
- 766 27. G. M. Kılınc, *et al.*, The Demographic Development of the First Farmers in Anatolia. *Curr.*  
767 *Biol.* **26**, 2659–2666 (2016).
- 768 28. B. Bramanti, *et al.*, Genetic discontinuity between local hunter-gatherers and central  
769 Europe's first farmers. *Science* **326**, 137–140 (2009).
- 770 29. D. Gronenborn, *The persistence of hunting and gathering: Neolithic western temperate and*  
771 *Central Europe*, in *The Oxford Handbook of the Archaeology and Anthropology of Hunter-*  
772 *Gatherers*, V. Cummings, P. Jordan, M. Zvelebil, Eds. (Oxford University Press, 2014).
- 773 30. J. Guilaine, *Aspects de la néolithisation en Méditerranée et en France*, in *The Widening*  
774 *Harvest. The Neolithic Transition in Europe: Looking Back, Looking Forward*, A.  
775 *Ammerman, P. Biagi, Eds* (Archaeological Institute of America, 2003).



- 776 31. M. Lipson, *et al.*, Parallel palaeogenomic transects reveal complex genetic history of early  
777 European farmers. *Nature* **551**, 368–372 (2017).
- 778 32. Marija Gimbutas, *The Living Goddesses*. (University of California Press., 2001).
- 779 33. A. Mittnik, *et al.*, The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442  
780 (2018).
- 781 34. S. Brace, *et al.*, Ancient genomes indicate population replacement in Early Neolithic  
782 Britain. *Nat. Ecol. Evol.* **3**, 765–771 (2019).
- 783 35. N. L. Morgunova, O. S. Khokhlova, Chronology and Periodization of the Pit-Grave Culture  
784 in the Region Between the Volga and Ural Rivers Based on Radiocarbon Dating and  
785 Paleopedological Research. *Radiocarbon* **55**, 1286–1296 (2013).
- 786 36. C.-C. Wang, *et al.*, Ancient human genome-wide data from a 3000-year interval in the  
787 Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
- 788 37. D. W. ANTHONY, *The Horse, the Wheel, and Language* (Princeton University Press,  
789 2007) <https://doi.org/10.2307/j.ctt7sjpn> (December 13, 2020).
- 790 38. H. Malmström, *et al.*, The genomic ancestry of the Scandinavian Battle Axe Culture people  
791 and their relation to the broader Corded Ware horizon. *Proc. R. Soc. B Biol. Sci.* **286**,  
792 20191528 (2019).
- 793 39. Fokkens, H. and Nicolis, F, *Background to Beakers. Inquiries into regional cultural*  
794 *backgrounds of the Bell Beaker complex*. (Leiden: Sidestone Press., 2012).
- 795 40. D. M. Fernandes, *et al.*, The spread of steppe and Iranian-related ancestry in the islands of  
796 the western Mediterranean. *Nat. Ecol. Evol.* **4**, 334–345 (2020).
- 797 41. I. Olalde, *et al.*, The Beaker phenomenon and the genomic transformation of northwest  
798 Europe. *Nature* **555**, 190–196 (2018).
- 799 42. I. Olalde, *et al.*, The genomic history of the Iberian Peninsula over the past 8000 years.  
800 *Science* **363**, 1230–1234 (2019).
- 801 43. N. Kashuba, *et al.*, Ancient DNA from mastics solidifies connection between material  
802 culture and genetics of mesolithic hunter–gatherers in Scandinavia. *Commun. Biol.* **2**, 185  
803 (2019).
- 804 44. T. Günther, *et al.*, Population genomics of Mesolithic Scandinavia: Investigating early  
805 postglacial migration routes and high-latitude adaptation. *PLOS Biol.* **16**, e2003703 (2018).
- 806 45. P. Bellwood, *First Farmers. The Origins of Agricultural Societies*, (Wiley-Blackwell,  
807 Chichester (2005).

- 808 46. D. Baird, *et al.*, Agricultural origins on the Anatolian plateau. *Proc. Natl. Acad. Sci.* **115**,  
809 E3077–E3086 (2018).
- 810 47. A. S. Kassian, *et al.*, Rapid radiation of the inner Indo-European languages: an advanced  
811 approach to Indo-European lexicostatistics. *Linguistics* **59**, 949–979 (2021).
- 812 48. Johannes. Müller Willigen, Samuel van. .., New radiocarbon evidence for european Bell  
813 Beakers and the consequences for the diffusion of the Bell Beaker Phenomenon. *Bell*  
814 *Beakers Today* **1** (2001).
- 815 49. R. Bollongino, *et al.*, 2000 years of parallel societies in Stone Age Central Europe. *Science*  
816 **342**, 479–481 (2013).
- 817 50. P. Skoglund, *et al.*, Genomic Diversity and Admixture Differs for Stone-Age Scandinavian  
818 Foragers and Farmers. *Science* **344**, 747–750 (2014).
- 819 51. F. M. T. A. Busing, E. Meijer, R. V. D. Leeden, Delete-m Jackknife for Unequal m. *Stat.*  
820 *Comput.* **9**, 3–8 (1999).