A genomic dating tool for ancient genomes resolves the origins of hundreds of Eurasian genomes

Authors: U. Esposito[1*], G. Holland[1,2], G. Alshehab[3,4], A. M. Dobre[1,5], M. Pirooznia[6], C. S. Brimacombe[1,7], E. Elhaik[1,8*]

Affiliations:

[1]University of Sheffield, Department of Animal and Plant Sciences, Sheffield, UK.

[2]Independent investigator, Cambridge, UK.

[3]University of Sheffield, Department of Automatic Control and Systems Engineering, Sheffield, UK.

[4]OSIsoft Technologies Middle East S.P.C., Almoayyed Tower, Al Seef District, Area 428, Bahrain.

[5]Independent investigator, Berkhamsted, UK.

[6]National Heart, Lung, and Blood Institute (NHLBI), US

[7]University of Bristol, School of Arts, Bristol, UK.

[8]Lund University, Department of Biology, Lund, Sweden.

*Correspondence to: eran_elhaik@biol.lu.se

Abstract: Radiocarbon dating is the gold-standard in archaeology to estimate the age of skeletons, a key to studying their origins. Half of all published ancient human genomes lack reliable and direct dates, which results in obscure and contradictory reports. We developed Temporal Population Structure (TPS), the first DNA-based dating method for ancient genomes ranging from the Upper Palaeolithic to modern-day samples and applied it to all 961 ancient Eurasians. We show that TPS predictions for radiocarbon-dated skeletons align with their known dates and correctly account for kin relationships. The TPS-dating of 359 poorly dated Eurasian samples resolves conflicts and sheds new light on disputed findings as illustrated by five test cases. We discuss the phenotypic traits of the Time Informative Markers (TIMs) that underlie TPS.

Summary: TPS is a novel method to date humans from the Upper Palaeolithic to modern time from their DNA sequences.

Main Text: Ancient DNA (aDNA) has transformed the study of human demographic history, allowing us to directly analyze patterns of past genetic variation rather than infer them *post factum (1)*. The last years have witnessed a conspicuous increase in the volumes of ancient skeletal DNA and studies attempting to trace their origins (2). Dating ancient remains is of key importance to produce meaningful and reliable historical reconstructions particularly in light of the growing medicalization of the field.

In the second half of the 20th century, radiocarbon dating dramatically changed the field of archaeology *(3)* and became the gold standard to date ancient organic materials *(4)*. Radiocarbon dating is based on the observation that living beings exchange ^{14}C with their biosphere while alive and cease to do so when dead. At that point, their ^{14}C atoms decay into ^{14}N with a half-life of ~5,700 y, whereas their ^{12}C concentration remains constant *(5, 6)*. Thereby, assuming that the initial ratio of carbon isotopes in the biosphere remained constant over time, measuring the ^{14}C to ^{12}C ratio allows inferring the sample's age. Over the years, improvements to the original method were made *(6)*, including pre-treatment of the samples' bones to eliminate contamination by recent carbon *(7)* and introduction of accelerator mass spectrometry which advanced the measurement of the decaying process *(8)*. Improvements in the knowledge of Earth's past environment and the quantification of reservoir effects have led to more accurate calibration curves of the past biosphere isotopes levels *(6, 9-11)*. Consequently, radiocarbon dating has improved in accuracy and reliability over the years and has been used to resolve conflicting interpretations. For instance, the bones recovered from Repton (England) were associated with the Viking Great Army from 873-874 CE based on the archaeological context, however early radiocarbon results predated some of them to the seventh and eighth centuries CE *(12)*. Only a later radiocarbon analysis that considered the marine reservoir effects found that all dated remains are consistent with a single late ninth-century event, in line with the numismatic evidence *(13)*.

A major limitation of radiocarbon dating is the high amount of collagen extraction (500 mg) involved in the process *(14)*. To date, of the ~2,600 ancient skeletons whose aDNA was successfully sequenced and published, only half were radiocarbon-dated. The remaining skeletons have either been dated according to the archaeological materials found alongside the sample or remain undated. The subjective interpretation of skeletal data already led to major misunderstandings on numerous occasions. For instance, a bone from the Darra-i-Kur cave in Afghanistan, initially assumed to be from the Palaeolithic (30,000 BP) *(15)* and often cited as one of the very few Pleistocene human fossils from Central Asia, was recently radiocarbon-dated to the Neolithic (4,500 BP) *(16)*. Similarly, one of the Brandysek site individuals (RISE569) was originally attributed to the Bell Beaker period (4,800-3,800 BP) *(17)*, but a later radiocarbon dating showed that the skeleton largely post-dated this culture (1,400-1,100 BP) *(18)*. As misattributions can lead to erroneous conclusions, the uncertainty in the age of nearly half of the aDNA samples poses a considerable risk of misinterpretation to the field, which calls into question their cost effectiveness and overall usefulness.

Here, we present the Temporal Population Structure (TPS) tool, the first DNA-based dating method suitable for samples younger than 15,000 BP. Compared to radiocarbon dating, DNA analyses require much less materials (x1/5) *(19)*, which makes DNA-based dating methods more viable in case of limited materials. Thus, TPS can be used to directly date skeletal aDNA for which no radiocarbon date is available or as an independent validation approach to existing results. The only other DNA-based method for dating was developed for genomes older than 30,000-20,000 BP and estimates the age of a sample from their Neanderthal ancestry *(20, 21)*. This technique,

tested and demonstrated on five of the most ancient human genomes (45,000-12,000 BP), becomes unstable for more recent dates *(20)* and is unsuitable for dating the vast majority of the ancient genomes which are younger than 10,000 BP or non-Europeans genomes in general. By contrast, TPS was calibrated using existing radiocarbon-dated individuals, which constrains its accuracy to the range of reliable radiocarbon dating.

Since within a population most allele frequencies vary due to random mutations associated with demographic processes that change over time (22), we hypothesized that there exists a group of markers that exhibit substantially different allele frequencies between different time periods and can be used to estimate temporal trends. We termed these markers *Time Informative Markers* (TIMs). Conceptually, TIMs are reminiscent of ancient Ancestry Informative Markers (aAIMs) *(23),* except that they operate along the time axis. Random time-dependent events create unique allele frequencies in the TIMs that are characteristic of the historical period in which individuals lived, which we termed *temporal components*. Briefly, to date a sample, we first identified the temporal components from a database of ancient Eurasian genomes as the expression of their allele frequency variation over time (not over space). We then calibrated the temporal profiles using radiocarbon-dated ancient genomes and constructed a reference panel of temporal populations. TPS calculates the temporal components of the undated sample and compares them with those of the temporal components of the reference populations. TPS then dates the sample based on the radiocarbon dates of the temporal populations with the most similar dating component to those of the test sample.

The Temporal Population Structure (TPS) model

We curated a database DB_{Anc} of 961 ancient Eurasian genomes (Fig. 1; Table S1) from the Late Upper Palaeolithic to the Anglo-Saxon Age (14,000-1,000 BP). Of these, 602 (63%) were radiocarbon-dated (DB_{RD}) and 321 (33%) were dated indirectly (usually from the archaeological context) (DB_{OD}). The dates of the remaining 38 (4%) individuals were either unknown or unclear (DB_{UD}). The first database was used to build the TPS model and test it via cross-validation. TPS was then applied to date the individuals in the remaining two databases.

To make TPS applicable to more modern genomes, we also curated a database DB_{Mod} of 250 individuals from Europe, Asia and Africa (Supplementary Materials). We merged it with a random subset of 300 DB_{Anc} genomes (Table S2) and applied *unsupervised* ADMIXTURE *(24)* with a various number of K components (Fig. S1; Supplementary Materials). Five ancient and three modern temporal components captured the temporal trends of the combined database. (Fig. S2; Supplementary Materials). The allele frequencies of the temporal components were used to simulate the DNA of putative "populations" that represented typical genomes from various periods (Figs. S3-S4; Table S2).

To construct the predictive model, we carried out a *supervised* ADMIXTURE analysis for all DB_{Anc} samples against the temporal components. Predictably, the results for the DB_{RD} individuals (Fig. 2; Table S2) showed that each temporal component predominates a delimited time interval. Skeletons with comparable dates but different geographical locations exhibited a higher similarity than samples from different time periods, regardless of their burial location.

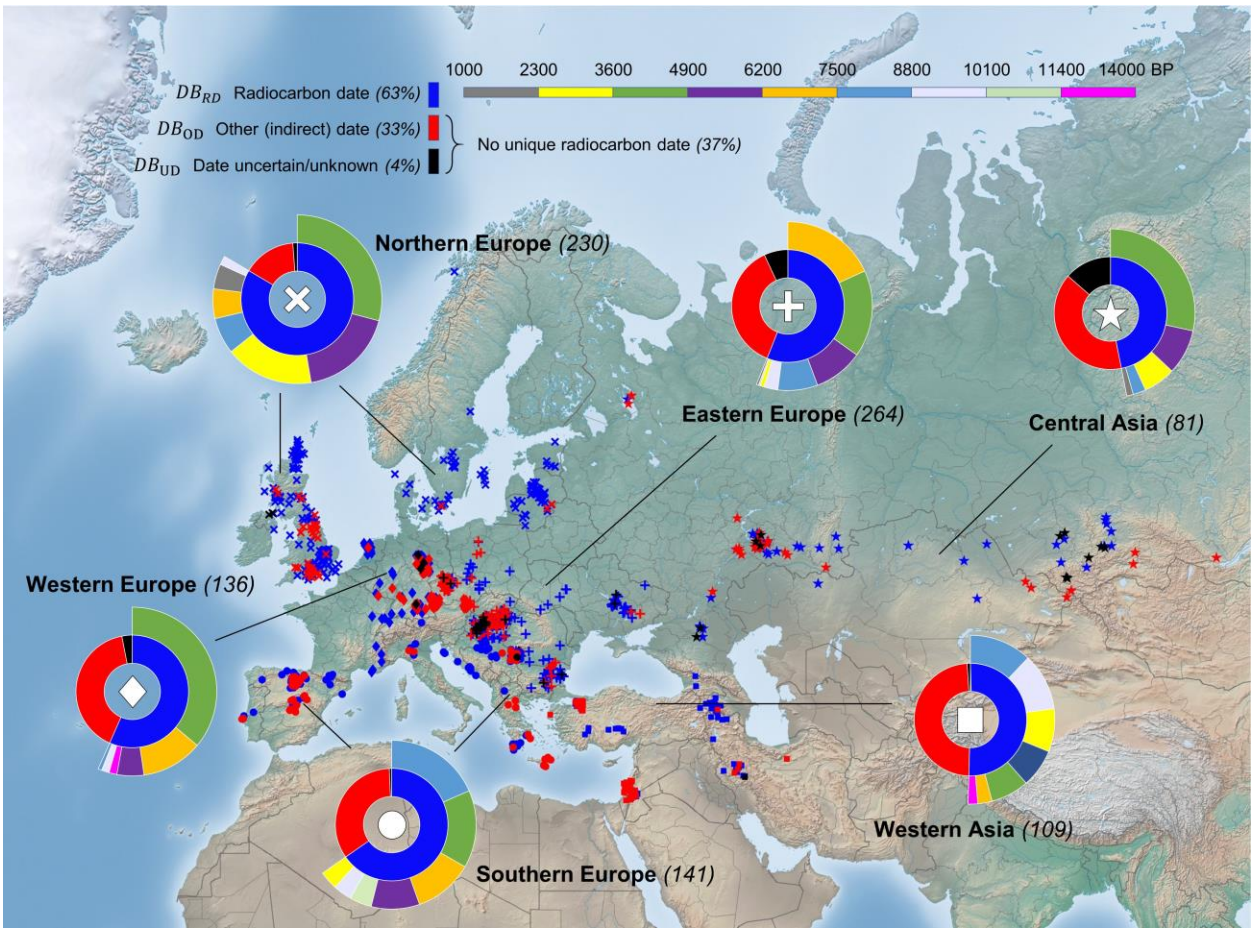


Fig. 1. *Location and dating of the 961 ancient individuals (D_{Anc}) used in this paper. Symbols mark geographical macro-areas where samples were found. Colours represent their dating method (D_{RD} (blue), D_{OD} (red), and D_{UD} (black)). The inner circle of each sunburst shows the proportion of samples dated with different dating methods and the outer circle marks the distribution of radiocarbon dates for the local samples. Radiocarbon dating is divided into nine temporal bins of 1,300 years (top bar) covering the timeline of our database, excepting the oldest bin that is 2,600 years wide.*

The TPS reference panel is derived from radiocarbon-dated individuals whose date was predicted well by TPS in a randomized setting. To identify those individuals, we assigned DB_{RD} individuals to non-overlapping bins of 500 years. TPS was applied 500 times to each individual. In each run, a random set of 400 individuals representing the diversity of all temporal profiles was selected and used to construct a reference panel of temporal populations. TPS then calculated the Euclidean distances between the temporal components of the test individual and those of the reference individuals and identified the three closest matching temporal individuals. The final date is obtained from the linear combination of the dates of these three individuals with weights proportional to the inverse of their Euclidean distances (Supplementary Materials). Based on these data, TPS then reports the average date and confidence intervals. By the end of this procedure, TPS dates per individual were averaged across all runs and individuals whose average TPS prediction deviated from their radiocarbon date by over 1,000 years were discarded from the

reference panel. The final TPS reference panel comprised of 495 ancient individuals and 150 modern individuals.

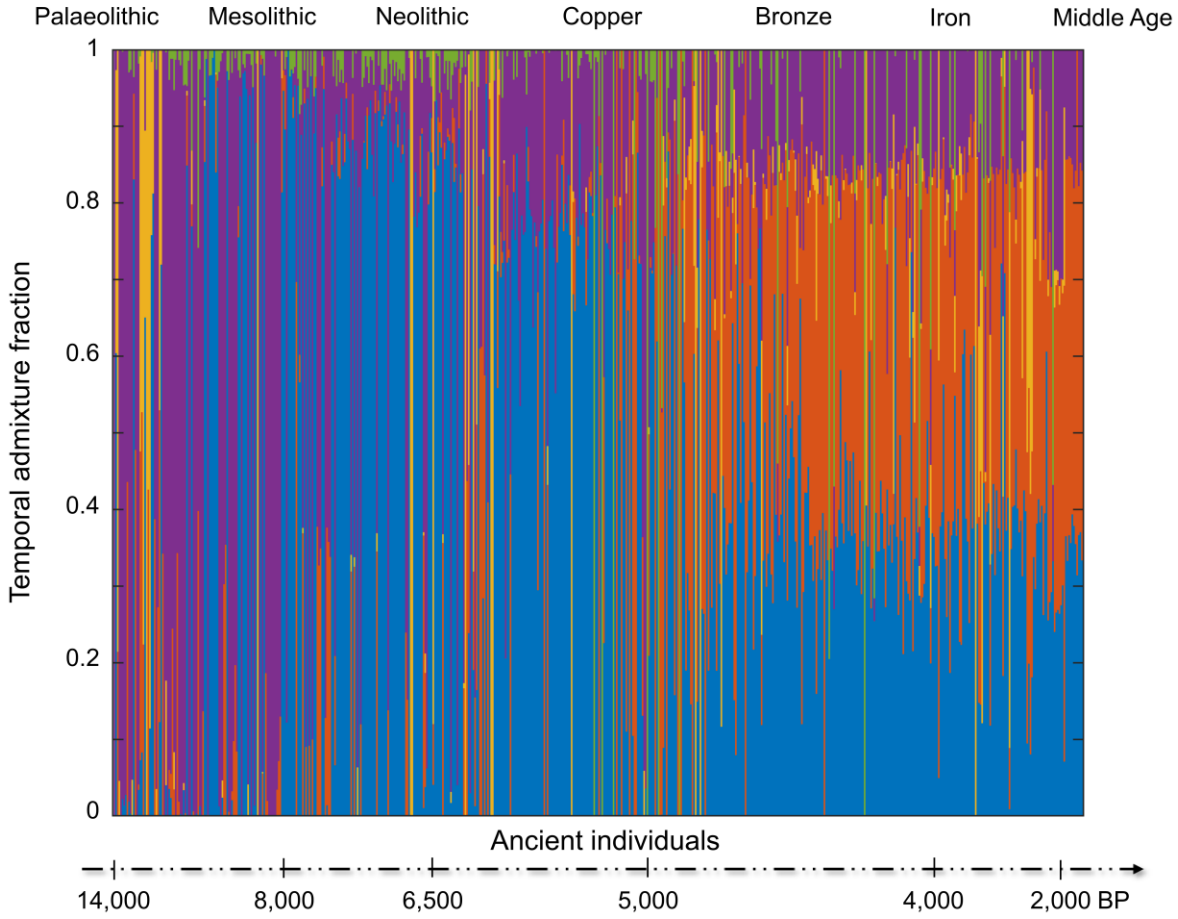


Fig. 2. *Ancient temporal components for the 602 radiocarbon-dated individuals (DB_{RD}). Each vertical stacked bar represents an individual. Colours correspond to the five ancient temporal components. The three modern temporal components are meagre for these samples and were omitted for coherency. Individuals are sorted by their age (note that the time x-axis is not linear). All the components are reported in Table S2.*

Identifying Time Informative Markers (TIMs)

To identify the genomic markers that underlie the ancient temporal components employed by TPS, the components were sorted from oldest to youngest (Fig. S5) creating a temporal variation profile of allele frequencies for every SNP. A time series analysis (Supplementary Materials) identified 62,371 SNPs whose frequency either decreased or increased over at least 3,000 years (Fig. S6), which we termed Time Informative Markers (TIMs) (Table S3) and 24,311 non-TIMs whose allele frequency exhibited little or no variation over time. TPS results for the non-TIMs were comparable to the results of two null models (Figs. S7-S8). By contrast, TPS results for the TIMs (Fig. S9) yielded similar results to those of the full set of SNPs (Fig. 3 below). Unsurprisingly, most of the

TIMs (53%) are intronic and 13% are intergenic *(25)*; 52% of the coding variants are missense (Fig. S10). Due to the high missingness of the dataset, to avoid omitting samples (Supplementary Materials), we used the entire SNP set for the remaining analyses.

Evaluating the accuracy of TPS

We next compared the radiocarbon dates for the 495 individuals in the reference panel with their TPS predictions obtained through a testing procedure where we excluded one individual at a time from the reference panel and used the remaining 494 to calculate its date. (Fig. 3). The two dating measures were highly correlated (*t-test*, N = 495: r = 0.96, *95% CI*: [0.947, 0.963], *p-value* = 1.4e-263) and in a close alignment to the ideal bisecting line $y = x$ (Fig. 3A). To gauge the reliability of TPS predictions, we defined the accuracy per individual as the absolute difference between TPS result and radiocarbon date. The average accuracy for this testing procedure was 417 years and 68% of the individuals were assigned a TPS date within 500 years from their radiocarbon date. Similarly, the 95% confidence intervals of TPS (Supplementary Materials) and the radiocarbon dates (Table S4) overlapped for 66% of the individuals. Only 6% of the samples were TPS-dated over 1,000 years from their radiocarbon date (Fig. 3B; Table S4). The general uniformity in the distribution of the accuracy across the different time periods suggests the absence of temporal biases towards any particular period, with the exception of the oldest samples for which performance is below average most likely due to the small number of samples. Overall, TPS results matched the radiocarbon dates (Fig. 3C), with the exception of Western Asia, likely due to the over representation of very old individuals in this region. TPS predictions of same-country individuals did not cluster around a single time period and were spread over the timeline following their radiocarbon dates, confirming that the temporal components represent temporal rather than geographical variation (Fig. S11). TPS accuracy was not correlated with the genomic coverage (*t-test*, N = 433: r = -0.0085, *95% CI*: [-0.1026, 0.0858], *p-value* = 0.8607) (Fig. S12), suggesting that TPS is robust to missing SNPs (within the set limit of including samples with over 15k SNPs). TPS predictions for the individuals discarded from the reference panel (Table S4) correlated reasonably well with their radiocarbon dates (*t-test*, N = 107: r = 0.4458, *95% CI*: [0.2796, 0.5860], *p-value* = 1.5e-06), albeit with higher deviation (average accuracy of 2,306 years) due to insufficient data for these time periods and graphical regions.

To develop a confidence measure in TPS predictions, we discretized DB_{RD} individuals according to the accuracy of their TPS predictions into five classes representing five levels of confidence, labelled 1-5 (5 is the highest confidence corresponding to deviations of 0 to 1,500 years from the mean radiocarbon date). We then used these labels to train a random forest classifier with confidence intervals and genetic distances between the predicted and radiocarbon date as input features (Supplementary Materials). The model correctly classified 82% of the samples, i.e., their confidence scores correctly matched their accuracy. Specifically, 75% of TPS dates were classified as having the highest confidence score (5), 19% in the range [4,5), 5% in [3,4), 1% in [2,3), and only one sample with confidence score in [1,2) (Table S4).

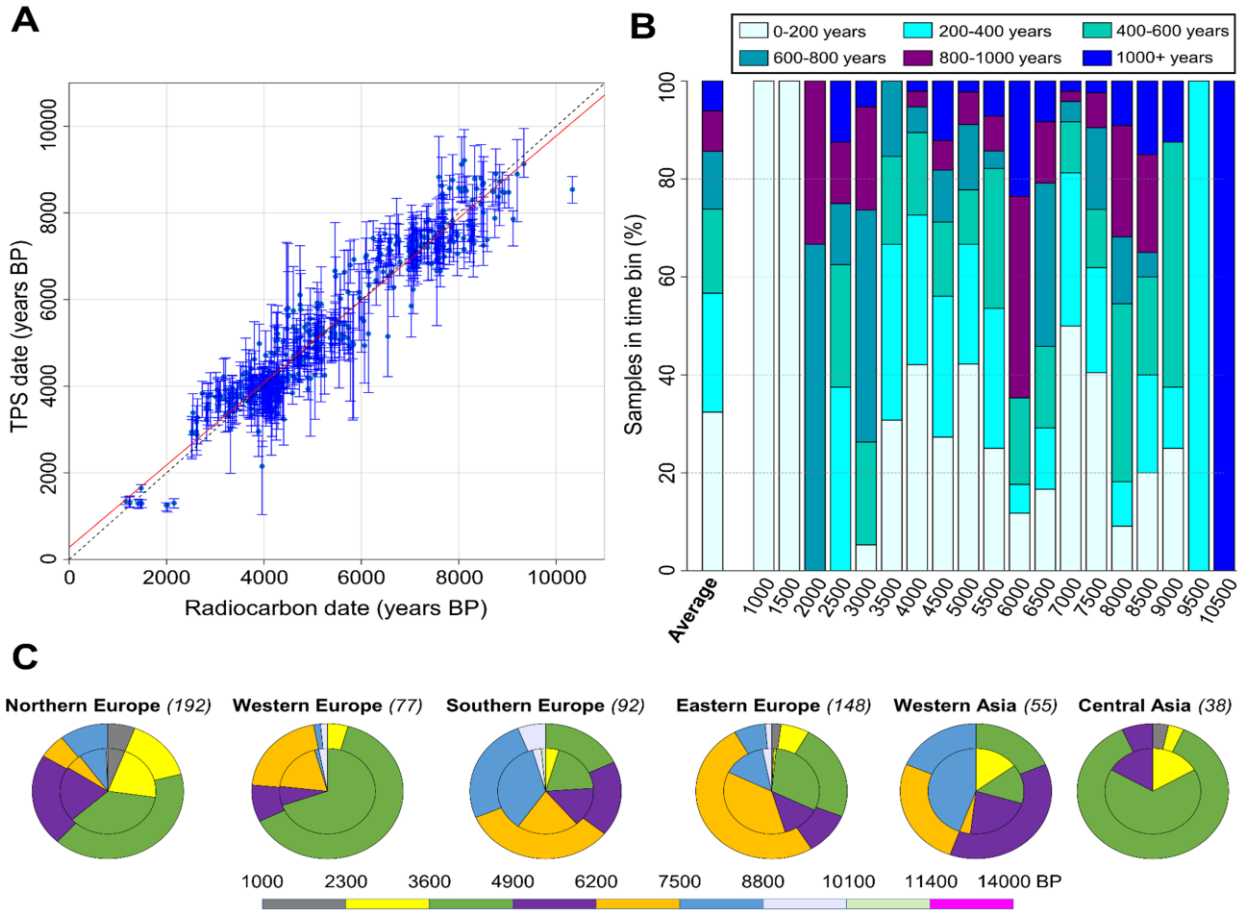


Fig. 3. *TPS results for 495 radiocarbon-dated Eurasian samples averaged across the 500 bootstrapping runs. A) The correlation between the radiocarbon dates and TPS results for the testing dataset (t-test, N = 495: r = 0.96, 95% CI: [0.947, 0.963], p-value = 1.4e-263). Error bars correspond to 95% confidence interval. Red line represents the linear fit against the $y = x$ line (dashed black). B) TPS aggregated accuracy. Individuals are sorted into 500-year time period bins according to their radiocarbon dates (e.g., the 4,000 BP bin represents individuals dated from 4,250 to 3,750 BP). The accuracy is calculated as the difference in years between the mean TPS prediction and mean radiocarbon date per individual. 32% and 56% of the samples were predicted within 200 and 400 years from their radiocarbon dates, respectively. C) TPS predicted dates split into 1,300 years (outer pie charts) and the corresponding radiocarbon dates (inner pie charts).*

We next applied TPS to the 250 modern test samples. TPS dated 249 individuals to the present (1,950 CE) and one UK individual (HG00099) whose temporal profile was very similar to the Anglo-Saxons samples (1,500-1,000 BP) to 1,243-1,688 CE.

Finally, we TPS-dated the 359 D_{OD} and D_{UD} individuals. The majority of TPS predicted dates (91%) were of high confidence (score in the range [4.5, 5]), with the remaining 3% in [4, 4.5), 5% in [3, 4), and 1% in [2, 3) (Table S4). The difference between the TPS and assigned dates for these samples was 940 years with a median of 484.91 years. The confidence intervals of the different dating methods did not overlap for 151 samples.

TPS application to five case studies

Settling inaccurate dates for parent-offspring pairs. Dates for the 26 first degree relatives in our database are not always consistent with the observed kinship. In 25 cases, TPS dates for the pairs overlapped (95% confidence intervals) with the 26th pair dated ~150 years apart compared to alternative dating methods where overlap was achieved only for 23 pairs. Radiocarbon dates overlapped in 8 out of 10 cases and the mean age difference of the pairs (275 years) was over twice that of TPS (125 years) (Table S5). For example, the two parent-child pairs I5236-I5241 and I1378-I1732 whose radiocarbon dates were 858 and 1188 years apart were TPS-dated to overlapping time periods, with the pairs being only 150 and 75 years apart, respectively (Table S5).

Dating the 13 Brandysek individuals, Czechia. Two individuals from (RISE568 and RISE569) the Brandysek site, were originally attributed by archaeological context to the Bell Beaker period (4,800-3,800 BP) *(17)*. Olalde et al. *(18)* later dated RISE569 (1,290-1,180 BP) and RISE568 (1,350-1,150 BP) based on the radiocarbon and archaeologically associated materials, respectively. As these post-dated the Bell Beaker Culture, the authors have omitted the two individuals from their analysis of the Brandysek individuals (4,850-4,150 BP). TPS dated RISE569 to 2,741-1,566 BP, RISE568 to 4,040-3,628 BP, and the remaining Brandysek individuals (excepting I7272) to 4,489-3,344 BP.

Three observations are noteworthy for these individuals. First, TPS's date of 4,040-3,628 BP for RISE568 contradicts the archaeologically-derived and much younger date obtained by Olalde et al. *(18)* and confirms the Brandysek Bell Beaker identity of RISE568. Secondly, TPS's date for RISE569 confirms Olalde et al.'s *(18)* assessment that this individual predates the Bell Beaker period by at least ~2,000 years. Third, of the remaining eleven individuals, ten were TPS-dated to fit within the Bell Beaker and Corded Ware period (4,850-4,150 BP). Individual I7272 was much older (6,800-4,964) and predated the Corded Ware culture as is evidenced by two additional features: I7272 lacked the ancient temporal component present in all other individuals at this site, which is ubiquitous among Bell Beaker samples and associated with the time period following the Yamnaya invasion (ancient temporal component 2) (Table S4). I7272 Y haplogroup is I2, whereas all the other males at that site, including the other two attributed to Corded Ware, are R1. Haplogroup R1 dominates post-Yamnaya migration populations *(26-28)* while I2 is primarily associated with Palaeolithic and Neolithic Europe *(29, 30)*. TPS dating, temporal components, and Y-haplogroup – all suggest that I7272 is related to an earlier Neolithic occupation at this site. We also note that the site consists of architectural features that are not usually associated with Bell Beaker burials, like the use of stone in graves *(18)*. Such indefinite archaeological materials should not be used for dating.

Establishing the Neolithic identity of Oxie 7 sample. Oxie 7 is an established Swedish Neolithic site dated to 4,900 BP *(31)*, which is often used as a type-site for the Neolithic *(32)*. In their analysis of Neolithic individuals, Allentoft et al. *(17)* employed archaeological context to date RISE174 to the Iron Age (1,523-1,340 BP) and excluded this individual from their analysis. TPS date for RISE174 is 4,347-3,533 BP, consistent with the Neolithic attribution of the site.

Correcting the radiocarbon date at Kyndeløse, Denmark. RISE61 was radiocarbon-dated to 4,801-4,442 BP *(17)*, which is suggestive of a Middle Neolithic origin ascribed to 5,400-4,700 BP in Scandinavia. However, Allentoft et al. *(17)* also cautioned that there may be a marine reservoir

effect, since the $^{13}C/^{14}N$ stable isotope data were consistent with a heavily marine diet, which would potentially shift the carbon age older than a terrestrial diet by several hundred years. TPS dated RISE61 to the Late Neolithic (4,352-3,830 BP), consistent with the difference predicted by a marine reservoir effect on this individual. This finding suggests that the Passage Grave Culture at Kyndeløse *(33)*, which began to appear around 5,300 BP in Denmark *(34)* may have existed as late as 4,000 BP.

Tracing the origins of the Longobards. The Longobards appeared during Roman times as a barbarian tribe who lived north of the Danube in present-day Hungary. From there, they established themselves in the Roman province of Pannonia at the beginning of the sixth century CE. They subsequently invaded Italy in 568 CE and were eventually conquered by Charlemagne in 774 CE *(35)*. A total of 10 samples were excavated from Pannonia at Szólád, a site considered to be of a Longobard type in terms of grave goods, location, and burial practices *(36)*. Amorim et al. *(36)* suggested that the graves of nine of the samples (SZ2-SZ5, SZ11, SZ15, SZ36, SZ43, SZ45) are from the mid-sixth century after radiocarbon-dated one individual (SZ43: 1,475-1,355 BP). We TPS-dated all the nine individuals. TPS confidence intervals overlapped with the sixth century Longobardian association (1,415-1,199 BP), with an intermediate confidence score in the range of 3-3.91 (Table S4). Amorim et al. assigned the tenth individual SZ1 to the older Bronze Age (3,800-2,900 BP) *(36)*, likely based on the archaeological context. TPS, instead, dated SZ1 along with the first nine individuals between the sixth and eighth centuries.

Phenotypic traits connected to the TIMs

Some TIMs can be associated with phenotypes, such as those harbored in the HERC2, OCA2, and TYR genes involved in skin, eye, and hair pigmentation (Fig. 4A, Fig. S13A-B). Since the Mesolithic, these traits were under selective pressure in favor of variants associated with lighter pigmentation *(37)*. rs2269424 (G/A) is another TIM adjacent to PPT2 and EGFL8, genes associated with immunity. A strong evidence of selection has been found at this marker *(30, 38)*.

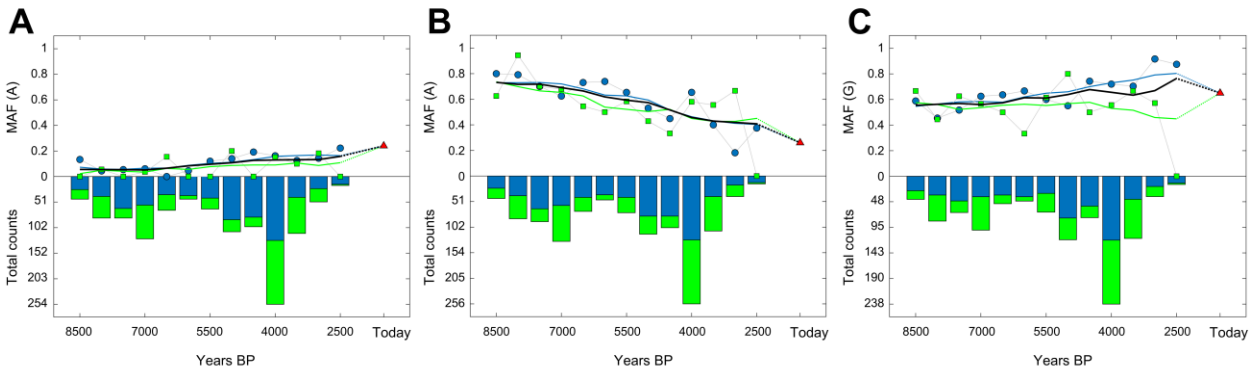


Fig. 4. Temporal variation in the allele frequencies of three TIMs. Bars show the number of individuals genotyped for that TIM and lines show the changes in the minor allele frequency. Blue referring to radiocarbon-dated samples and green colour referring to TPS dated samples with the highest confidence score ([4,5]). The black line shows the weighted average of the two MAF measures. A) rs1393350 (G/A) in the TYR gene involved in pigmentation. B) rs2269424 (G/A) is adjacent to PPT2 and EGFL8 genes and associated with immunity. C) rs2073711 (A/G) in the CILP gene.

Our temporal trend (Fig. 4B) supports these findings and suggests the presence of negative selection. TIMs rs2073711 (A/G) (Fig. 4C) and rs1800562 (G/A) (Fig. S13C) located in the CILP and HFE genes are associated with cartilage scaffolding *(39)* and hemochromatosis *(40)*, respectively. In both cases, the temporal variation trends indicate an increase in the frequency of these risk alleles possibly due to the transition to a sedentary lifestyle and diet changes.

Discussion

That half of ancient skeletal samples are imprecisely dated with no alternative to radiocarbon dating is the most daunting problem for paleogenomics. Radiocarbon may be widely accepted as the benchmark standard for dating ancient remains *(6, 41)*, but its reliance on large amounts of organic material renders many samples undatable

Genomic dating methods rely on aDNA sequences and are free from the environmental biases to which radiocarbon dating is exposed *(19)*, making it possible to perform direct dating on the skeletons whose radiocarbon dating cannot be established. Motivated by our observations that allele frequencies show temporal variability over time we introduced the concept of Time Informative Markers (TIMs). We showed that TIMs, like rs1393350 associated with pigmentation, have increased or decreased their frequency over time as reported elsewhere *(30, 37, 38, 40)* and can be used as biomarkers for specific time periods. Based on this concept, we developed the Temporal Population Structure (TPS) tool, which utilises TIMs to date ancient skeletons as far back as the Late Upper Palaeolithic using their genomic data without prior assumptions. TPS models genomes as consisting of eight temporal components and calculates the dates by comparing the proportion of these components to those of radiocarbon-dated reference samples that fit the model. We demonstrated the accuracy of TPS by showing its ability to correctly predict radiocarbon-dated individuals ($r = 0.96$), modern individuals, and ancient family relatives. Finally, TPS dated 359 skeletons with unreliable or missing dates.

TPS is a novel instrument to the growing toolkit of paleogeneticists that can be used to address contradictory findings in the Paleogenomic literature. We envision that, with time, direct dating methods will become more accurate for broader worldwide communities with the increase in the number of sequenced populations. Therefore, our results should be considered a lower bound to the full potential of TPS for biodating. Since TPS was calibrated on Eurasians radiocarbon-dated to the period 14,000-1,000 BP and modern-day individuals, it cannot be used for older non-Eurasian samples due to the sparsity of such radiocarbon-dated samples. This limitation can be addressed when more data will become available.

Material and Methods

Curating ancient genomic databases. We obtained a database DB_{Anc} of 961 ancient individuals and 147,229 SNPs after filtering a published ancient DNA database with 150,278 SNPs *(23)* and enforcing the following criteria: i) missingness lower than 0.9 (i.e., individuals with less than 15k SNPs were removed); ii) dates in the range 14,000-1,000 BP; and iii) burial coordinates within Eurasia (Table S1). We also obtained full annotation for these individuals, such as burial locations and dates. Throughout the paper, unless explicitly specified, we used mean radiocarbon dates, defined as the middle point of the dating interval obtained with the samples' annotation.

Constructing the temporal components. To identify the temporal components, we randomly selected 300 samples from DB_{Anc} that were merged with 250 individuals (DB_{Mod}) selected evenly from five present-day populations (Chinese CHB, Yoruba YRI, Finnish FIN, British GBR, and Tuscan TSI) from the 1000 Genomes database *(42)*, and filtered them for the same 147,229 SNPs. We applied *unsupervised* ADMIXTURE *(24)* for *K's* ranging from 4 to 12. Individuals were sorted by time and temporal plots were produced for each run (Fig. S1). For each plot, we selected putative temporal components that were characteristic of time (not geography), i.e., components that were evenly distributed in all samples within a certain time period. Ten and three putative modern and ancient components were identified, respectively with a clear split between the ancient and modern components. We continued refining them separately.

For the ancient components, using the allele frequencies output of ADMIXTURE (*p*-file), 15 synthetic samples associated with each temporal component candidate were generated *(23, 43)*. A Principal Component Analysis (PCA) plot of these 150 samples showed the components to be separated, except for two that could not be resolved. We discarded one of the two and were left with 135 synthetic samples from nine components (Fig. S2). We merged these 135 synthetic samples with the 300 ancient individuals and ran ADMIXTURE in a *supervised* mode against the synthetic samples. In following *supervised* ADMIXTURE analyses, we dropped one component at the time and compared the resulting plots, reducing the number of overlapping components. Eventually, five ancient temporal components were retained (Fig. S3; Table S2). For the modern components, we found that three components best described our samples, and that when merging them with the ancient individuals, *K*=8 had minimum noise, smoother profiles and high ancient-modern sample separation (Fig. S4).

Calculating the TPS reference panel. Based on their date, radiocarbon-dated individuals DB_{RD} were grouped into non-overlapping bins of 500 years ([14,250-13,750] … [1,250-750] BP). For each of 500 runs, 400 samples were chosen at random with stratified sampling to represent all the temporal bins. We then applied, for these samples, an iterative combination of *k*-means (starting from $k = 2$) and pairwise *F*-tests within each group with at least 3 reference individuals, to identify possible substructures based on their temporal profiles. For each group and *k*, we applied *k*-means to the temporal coefficients of the reference individuals to identify the clusters and their centroids. A pairwise *F*-test between centroids was used to verify the hypothesis that each cluster was indeed a distinct entity from all the others. If this hypothesis was verified for all the clusters within the group, the next value of *k* was tested. Alternatively, the previous *k* represented the optimal number of subgroups that describe the genetic variability within the temporal bin. Each subgroup was called a temporal population, characterized by an average temporal signature and an average date T with the corresponding error calculated as a standard deviation of the components. Temporal populations of only one individual were discarded. The remaining temporal populations were added to the reference panel (Table S4).

Predictive algorithm. To date a test sample, a *supervised* ADMIXTURE infers its temporal components against the temporal populations of the reference panel. TPS then calculates the genetic distances $\{g_i\}_{i=1\ldots N}$ between the individual and the temporal populations as the Euclidean distance between two temporal vectors. Finally, TPS selects the three temporal populations P_1, P_2, P_3 with the closest genetic distance to the test sample and applies a weighted average to their average dates $\tau = w_1T_1 + w_2T_2 + w_3T_3$, where the weights ($w_i$) are the inverse of the genetic distances normalized to their sum

$$w_i = \frac{g_i^{-1}}{g_1^{-1}+g_2^{-1}+g_3^{-1}}.$$

and τ is the TPS date for the test sample. This procedure is repeated 500 times and TPS returns the average date and the confidence intervals.

Identifying Time Informative Markers (TIMs). Provided the per-SNP allele frequencies (ADMIXTURE's *p*-file) which comprise the five ancient temporal components and represent different periods, it was possible to associate the temporal components with time periods (Figs. S3, S5). Sorting the five ancient temporal components from old to new, we used the allele frequencies of each component (Table S3) to detect SNPs whose allele frequencies show directed behavior over time. For that, we constructed time series with the date ranges assigned to the temporal components in 500-year bins (from 14,000 to 3,500 BP), resulting in 22 data points (Fig. S5). Overlaps in the assigned dates ranges of the temporal components were averaged to construct the time series, and the resulting temporal trends were smoothed using a moving average filter to reduce noise. A total of 62,371 SNPs showing global increasing or decreasing trends or displaying local behavior over sub-intervals of at least 3,000 years were considered TIMs (Fig. S6; Table S3). 24,311 SNPs whose allele frequencies exhibit little or no variation over time (absolute change <=0.1) were considered non-TIMs. 715 individuals with at least 15k TIMs were analyzed.

Null models. We gauged the extent to which the choice of markers affected TPS's accuracy by constructing two null models. The first one comprised of 24,311 non-TIMs and 385 individuals with at least 15k non-TIMs. The second one, comprised of the complete SNP-set. Here, we used random temporal components drawn from a uniform distribution.

Confidence estimate in TPS results. In addition to reporting TPS's average date and confidence interval developed a measure of confidence in the overall estimate. For that, we discretized TPS's accuracy (absolute difference between mean radiocarbon date and average TPS date) computed for the 602 radiocarbon-dated individuals into 5 bins of 1,500 years. Each bin was assigned a confidence score, from the highest 5 (corresponding to a distance of less than 1,500 years) to the lowest 1 (corresponding to a distance of 6,000-7,500 years). The confidence score was used to train a random forest classification model (with the Python library Scikit-learn) to assess the reliability of TPS predictions based on the confidence intervals and the genetic distances generated by TPS itself as Euclidean distance between temporal coefficients. Cross-validation and randomization were employed on the 602 individuals in order to obtain multiple classification trees, and the final results were averaged. This procedure was then applied to samples without radiocarbon dating to obtain their confidence score (Table S4).

Supplementary materials

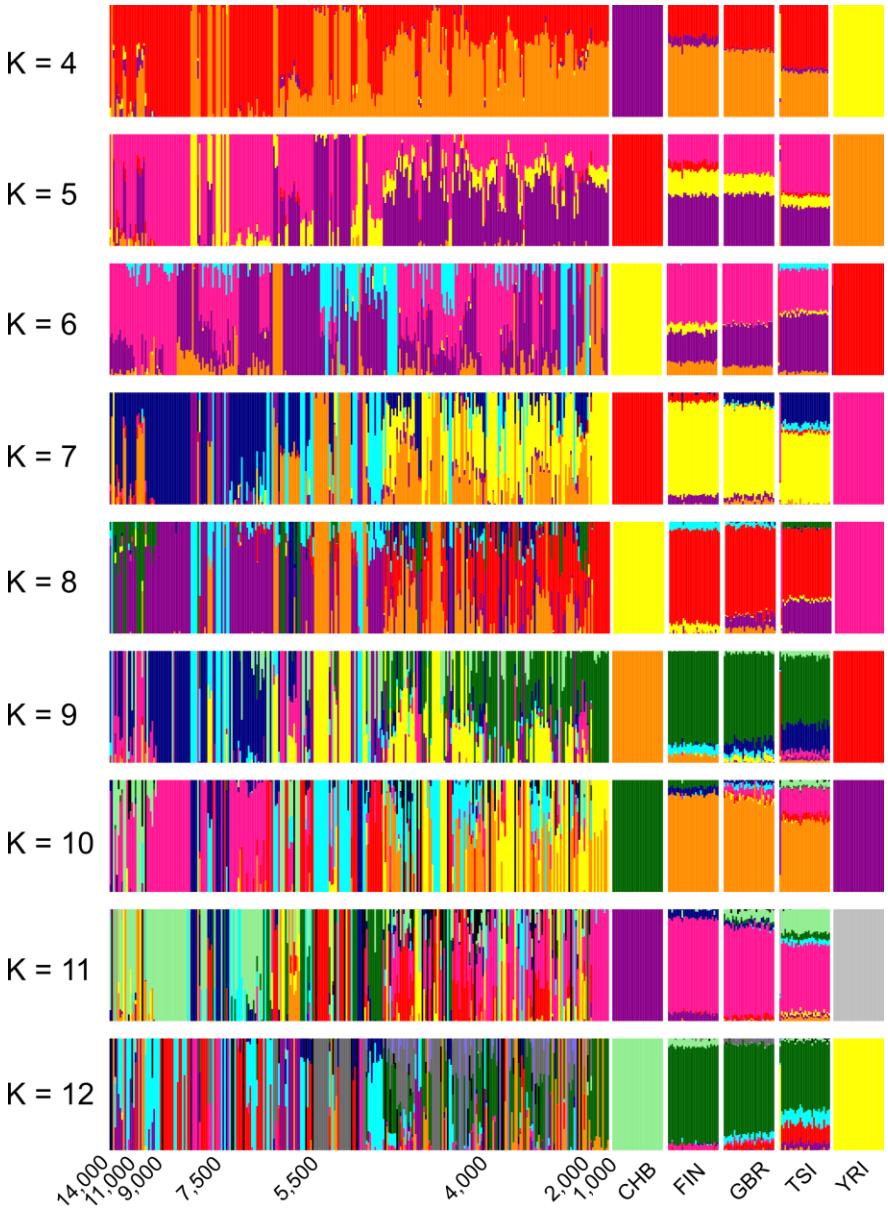


Fig. S1. Temporal plots of 300 ancient and 250 modern individuals sorted by age. The plots were obtained using *unsupervised* ADMIXTURE with *K's* ranging from 4 to 12. Each vertical stacked bar represents an individual. Colours correspond to the temporal components.

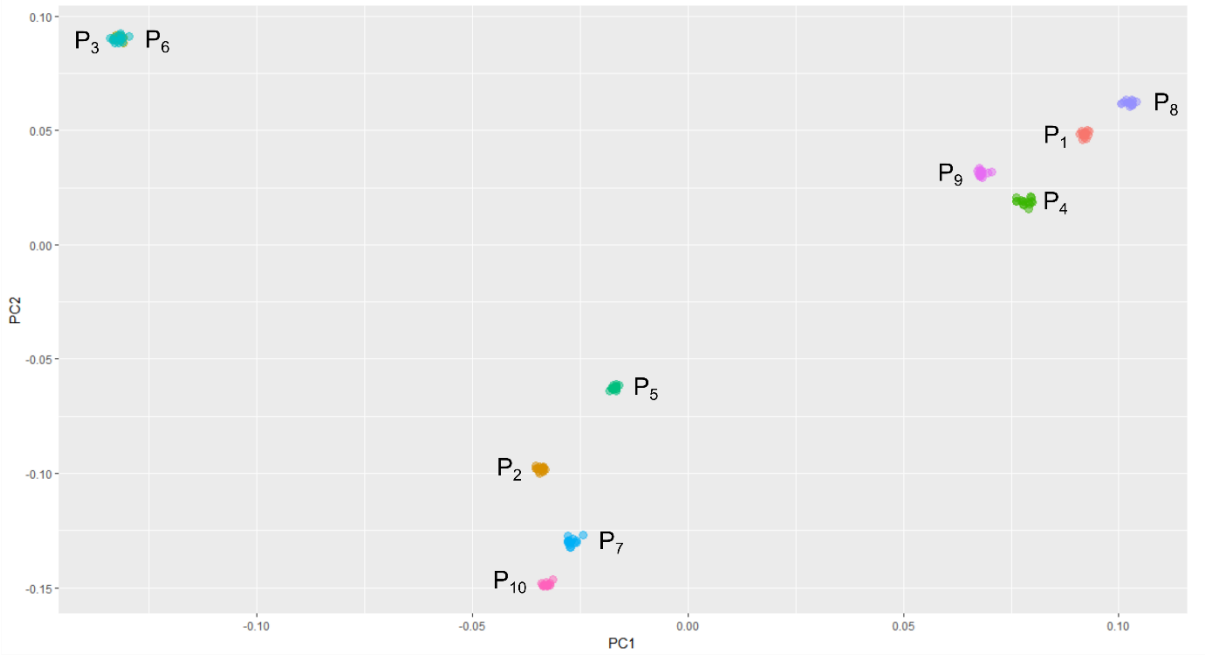


Fig. S2. Principal Component Analysis (PCA) plot for 150 simulated samples derived from ten ancient temporal component candidates (15 samples per component). Two clusters (P_3 and P_6) showed a complete overlap (top left corner) and one of them was removed.

5

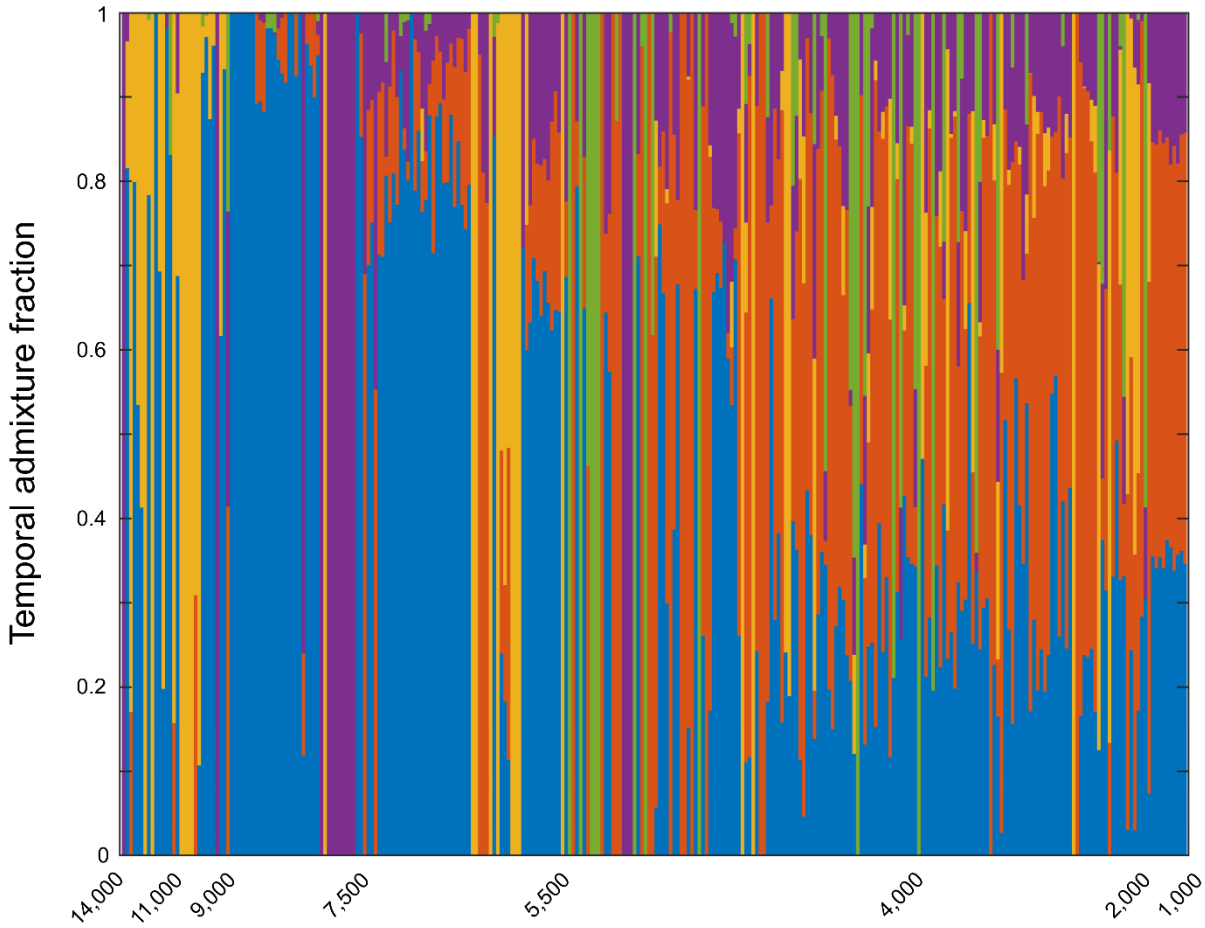


Fig. S3. Ancient temporal components of 300 ancient individuals sorted by age. Results were obtained using *supervised* ADMIXTURE with five ancient temporal components. Each vertical stacked bar represents an individual. Colours correspond to the temporal components.

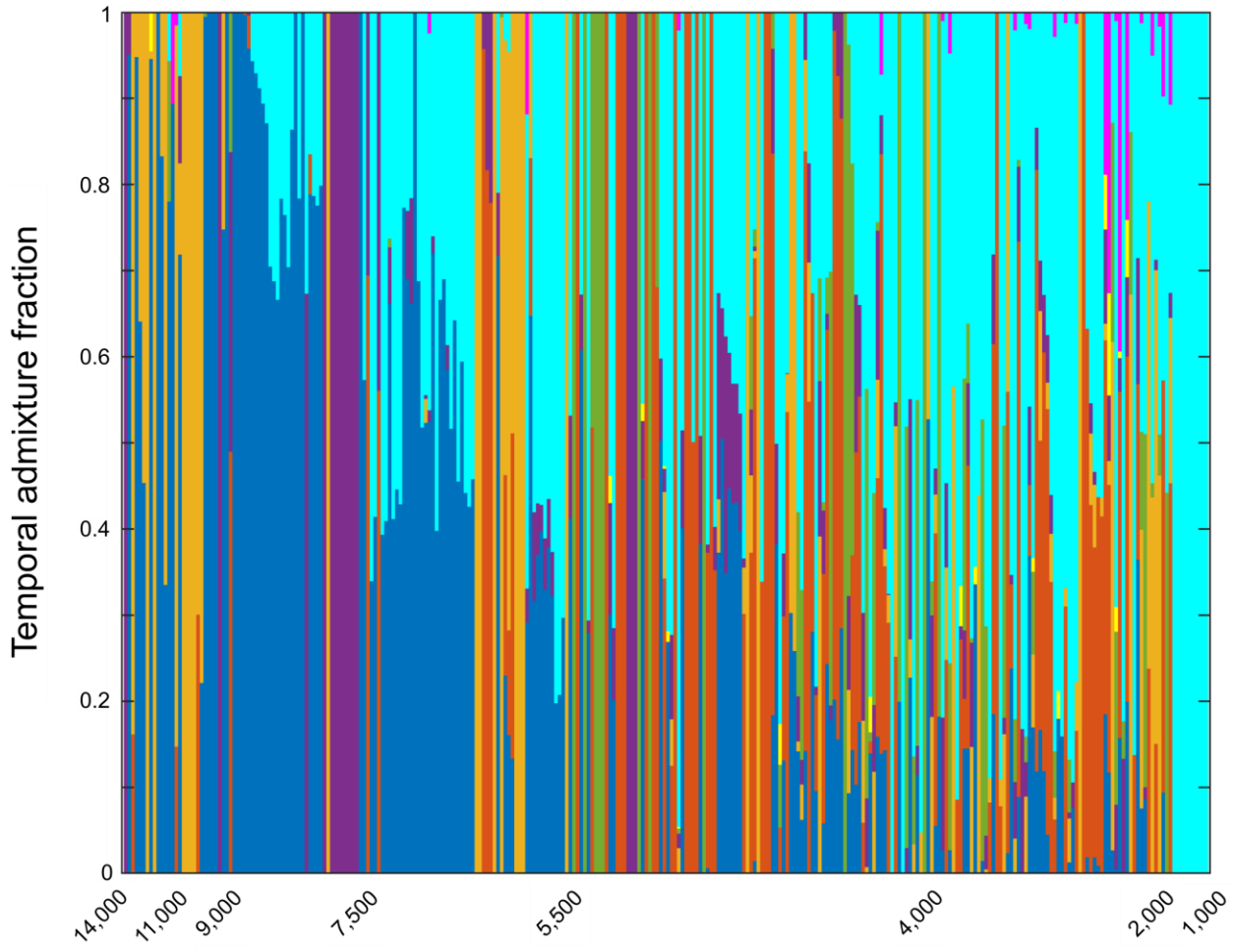


Fig. S4. All temporal components for 300 ancient individuals sorted by age. The results were obtained using *supervised* ADMIXTURE with the five ancient and three modern temporal components. Each vertical stacked bar represents an individual. Colours correspond to the temporal components.

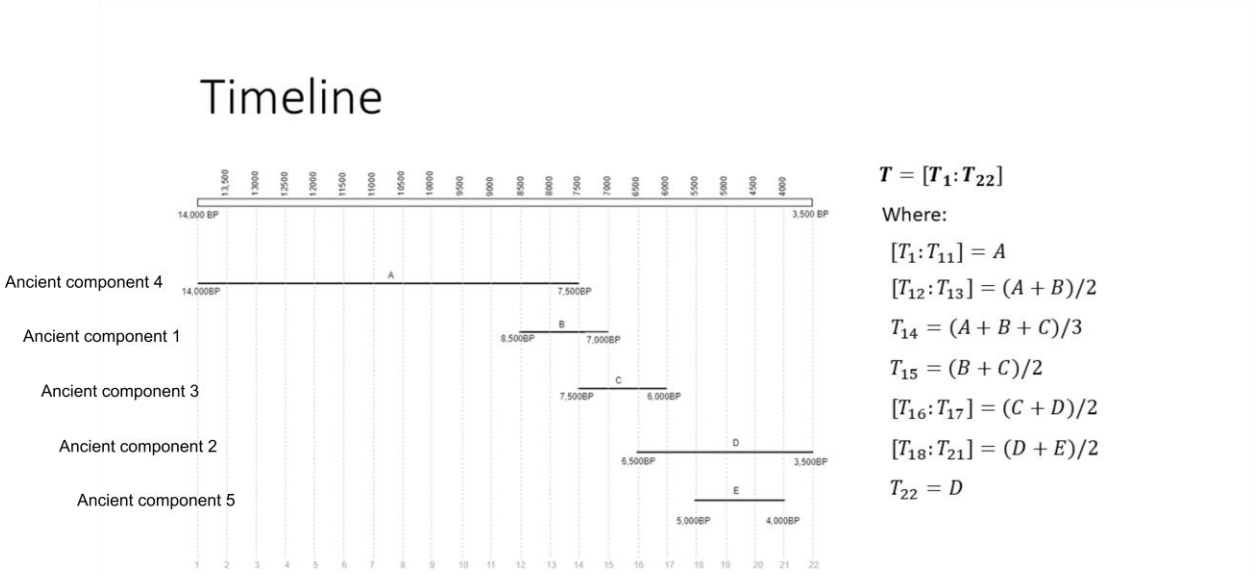


Fig. S5. Construction of time series for the identification of Time Informative Markers (TIMs). The five ancient temporal components (Fig. S3) were assigned a time period in our timeline, as shown by the five horizontal bars. The timeline was divided into 22 bins of 500 years. For each SNP we constructed a time series $T = [T_1:T_{22}]$ that consists of 22 allele frequencies (legend). Values were obtained from the allele frequencies of the five temporal components for that SNP, combined as shown in the legend. For instance, the values of T_1 to T_{11} were equal to the allele frequency of the ancient temporal component 4 (labelled as A), as this component was dominant during this time interval; the values of T_{12} and T_{13} were the average between the allele frequency of ancient temporal components 4 and 1 (labelled as B), as both components were dominant in this time interval. The final time series were smoothed using a moving average filter to reduce noise. For example, following this procedure TIM rs6603791 showed a frequency of 0.71 for the 10,000 BP bin and a frequency of 0.47 for the 3,000 BP bin, with decreasing intermediate values. The time series of the TIMs is shown in Table S3.

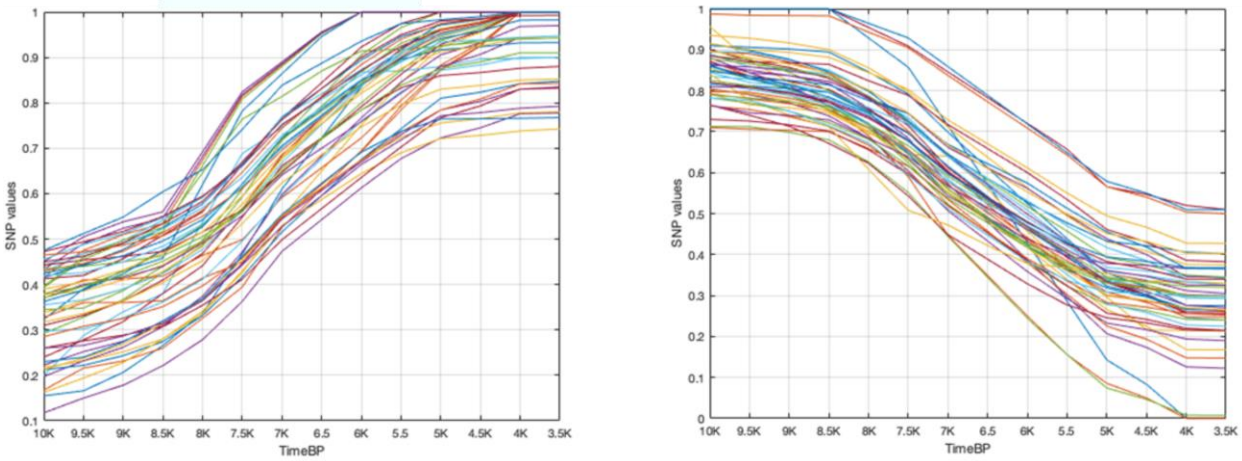


Fig. S6. Time series of alleles frequencies for 100 Time Informative Markers (TIMs) that showed the most pronounced 50 increasing (left) and 50 decreasing (right) trends.

5

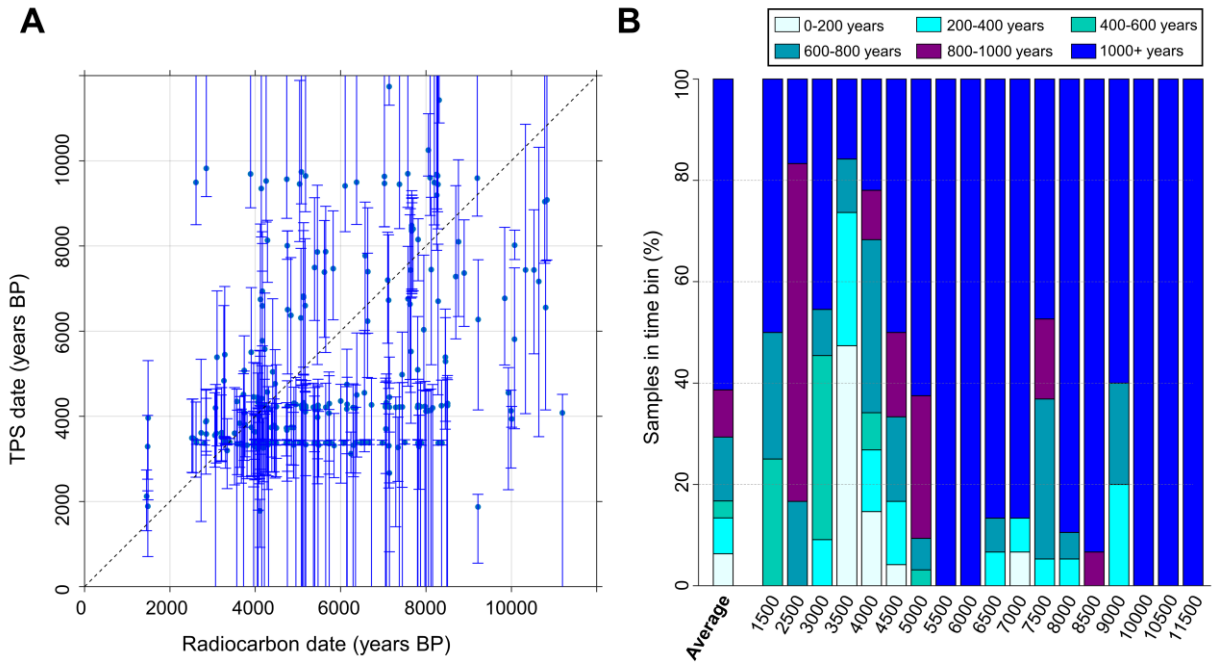


Fig. S7. TPS results for radiocarbon-dated Eurasian samples with the 24,311 non-TIMs, averaged across 500 bootstrapping runs. A) The correlation between radiocarbon dates and TPS results for the testing dataset (*t-test*, $N = 276$: $r = 0.39$, *95% CI*: [0.282, 0.486], *p-value* = 4.1e-11). Error bars correspond to 95% confidence interval. Dashed black line represents the $y = x$ line. B) TPS aggregated accuracy. Individuals are sorted into 500-year bin time periods according to their radiocarbon dates. Note, that due to the higher missingness in the data that resulted from the smaller SNP set, the non-TIMs dataset required dropping additional 219 individuals compared to the all-SNP dataset.

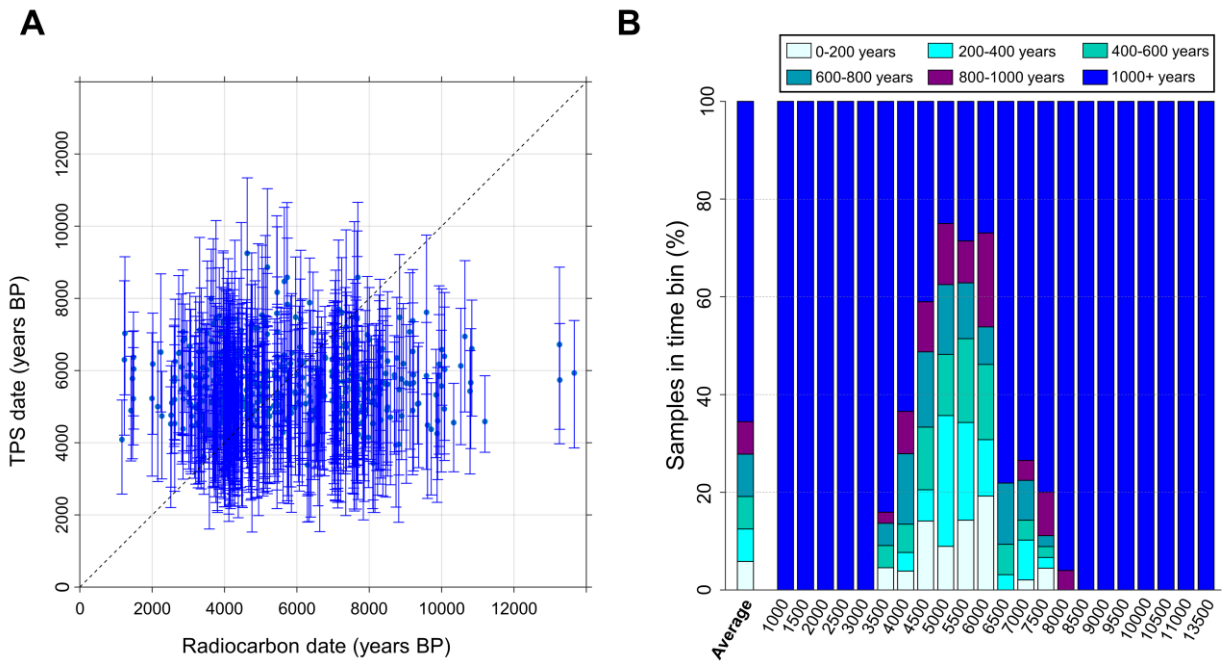


Fig. S8. TPS results for radiocarbon-dated Eurasian samples with a random model, averaged across 500 bootstrapping runs. A) The correlation between radiocarbon dates and TPS results for the testing dataset (*t-test*, N = 602: r = 0.04, *95% CI*: [-0.038, 0.122], *p-value* = 0.295). Error bars correspond to 95% confidence interval. Dashed black line represents the ideal $y = x$. B) TPS aggregated accuracy. Individuals are sorted into 500-year bin time periods according to their radiocarbon dates.

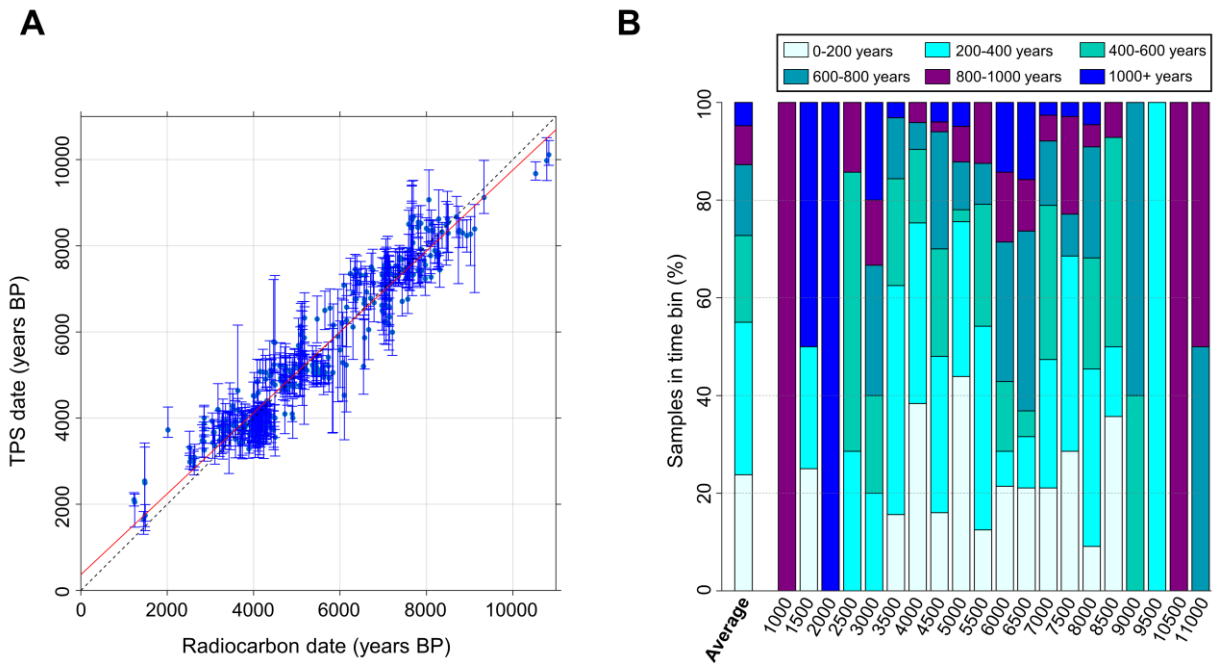
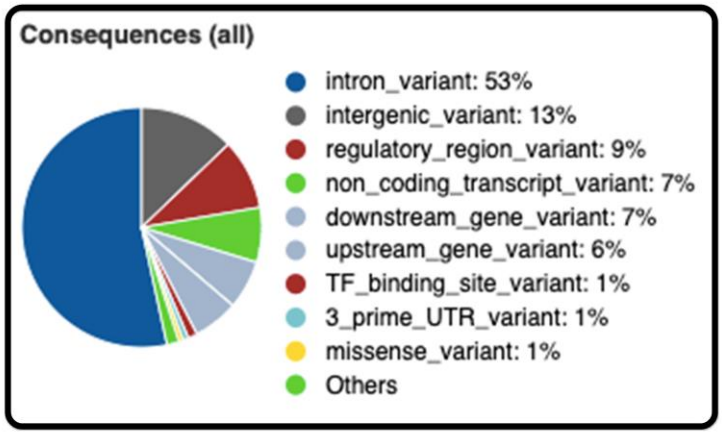


Fig. S9. TPS results for radiocarbon-dated Eurasian samples with TIMs, averaged across the 500 bootstrapping runs. A) The correlation between radiocarbon dates and TPS results for the testing dataset. Error bars correspond to 95% confidence interval. Red line represents the linear fit against the y = x line (dashed black). B) TPS aggregated accuracy. Individuals are sorted into 500-year bin time periods according to their radiocarbon dates. The accuracy is calculated as the difference in years between average TPS prediction and mean radiocarbon date per individual. These results (*t-test*, N = 406: r = 0.96, *95% CI*: [0.948, 0.964], *p-value* = 1.2e-215) replicated those obtained with the full set of markers shown in Fig. 3 (*t-test*, N = 495: r = 0.96, *95% CI*: [0.947, 0.963], *p-value* = 1.4e-263). Note that, due to the higher the missingness, which was a result of the smaller SNP set, the TIMs dataset required dropping additional 89 individuals compared to the All-SNP dataset, so that the final number of individuals predicted in this case is 406.

Category	Count
Variants processed	64640
Variants filtered out	0
Novel / existing variants	162 (0.3) / 64478 (99.7)
Overlapped genes	20597
Overlapped transcripts	42324
Overlapped regulatory features	16094



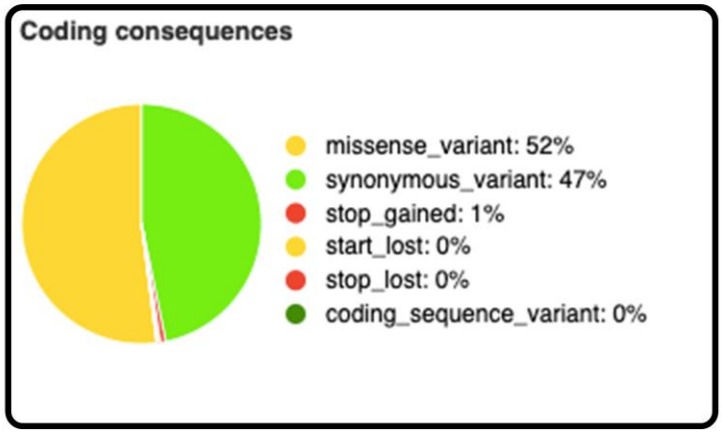


Fig. S10. VEP summary results for the Time Informative Markers (TIMs).

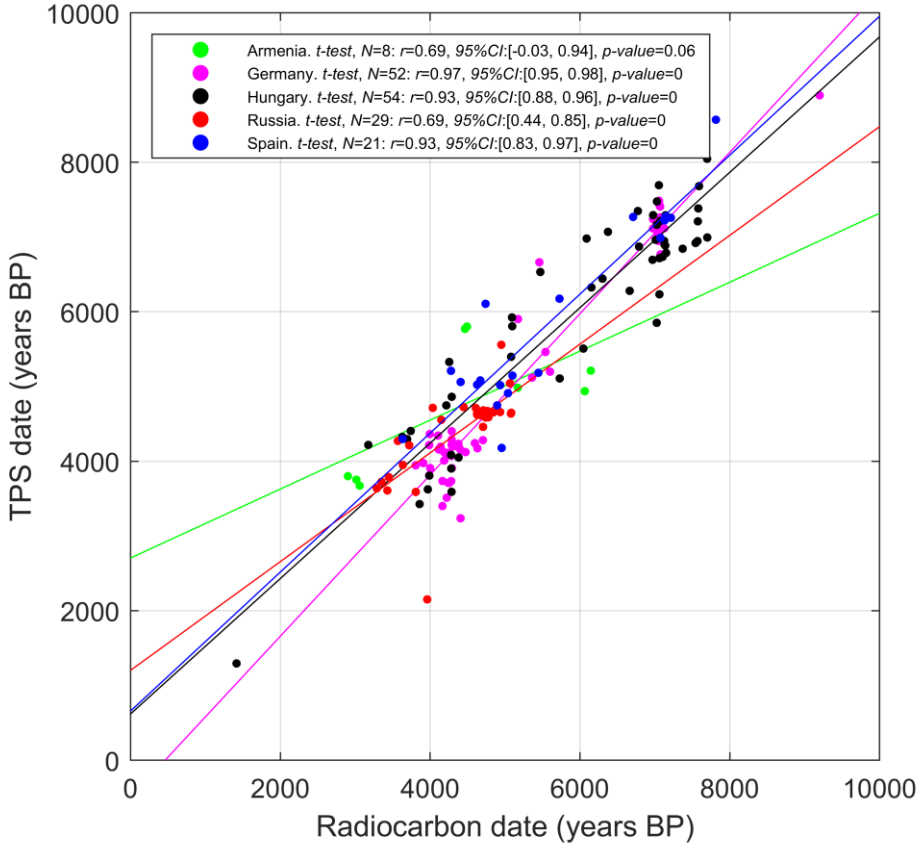


Fig. S11. Correlation between radiocarbon dates and TPS results for same-country radiocarbon-dated individuals (from Fig. 3A) averaged across the bootstrapping runs. TPS results were distributed along the timeline that follows their radiocarbon dates, confirming that the temporal components represent temporal rather than geographical variation. The only countries diverging from this general behaviour are Armenia and Iran, suggesting the possibility that a minor geographical effect might be present in their components (in particular in the 3rd ancient component, see Table S4).

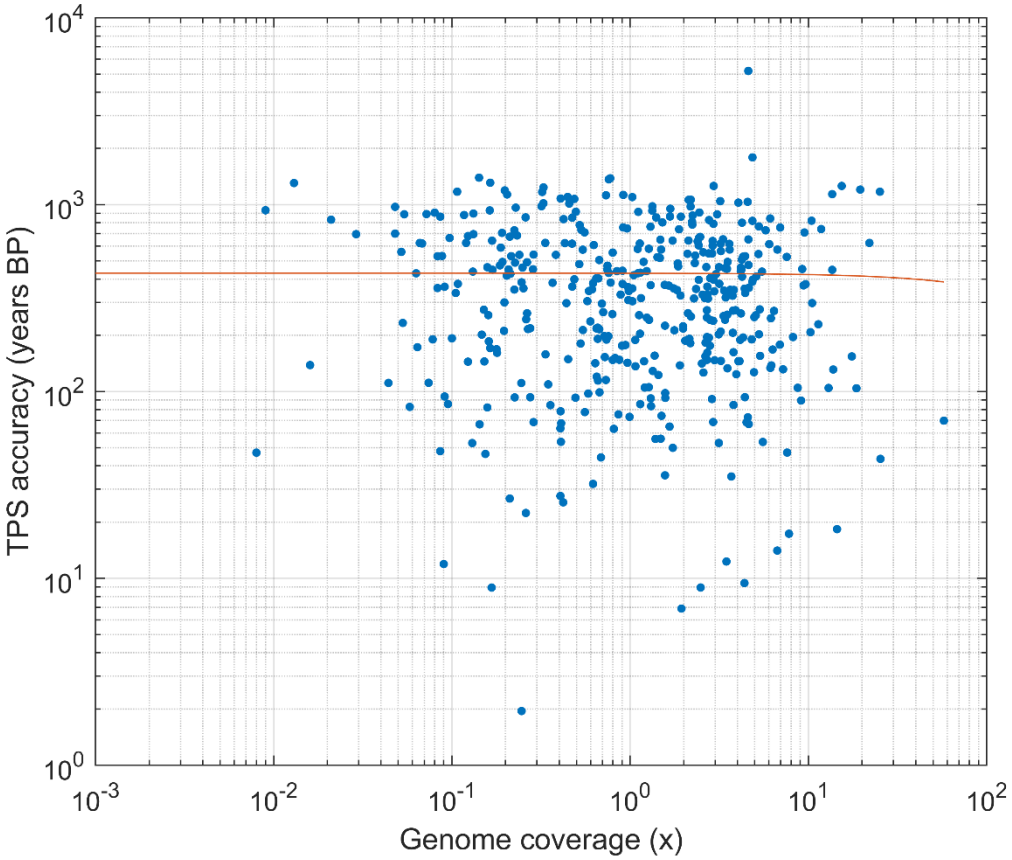


Fig. S12. Correlation between TPS accuracy and genome coverage for the radiocarbon-dated samples for which coverage was reported in the literature. Both axes are logarithmic scaled, red line represents the linear fit. There is no correlation between the two quantities (*t-test*, $N = 433$: $r = -0.0085$, *95% CI*: [-0.1026, 0.0858], *p-value* = 0.8607), suggesting that TPS results are not sensitive to missingness (provided the use of at least 15k SNPs).

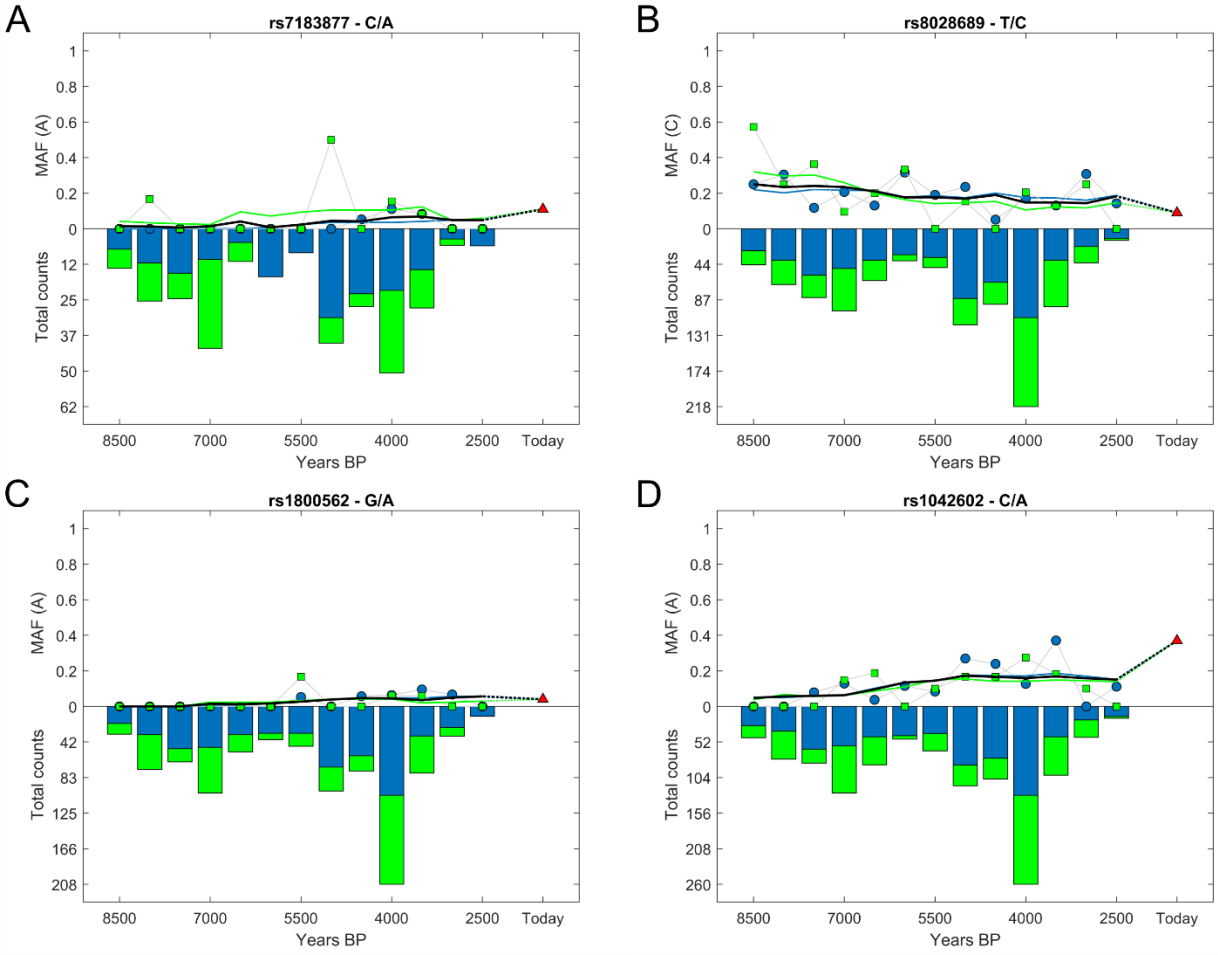


Fig. S13. Temporal variation of the allele frequencies for four TIMs. Radiocarbon-dated samples (blue) and TPS dated samples with the highest confidence score ([4,5]) (green) are shown. Bars show the number of individuals genotyped for each TIM. Lines show the changes in the minor allele frequency over time for each dataset. The black line shows the weighted average of the two MAF. A-B) TIMs rs7183877 (C/A) and rs8028689 (T/C) within the HERC gene are part of a haplotype spanning 166kB on chromosome 15, defined by 13 SNPs, found in 97% of all Caucasians with blue eyes, with values: rs7183877 (C) and rs8028689 (T). The other alleles (A for rs7183877 and C for rs8028689), seem to suggest brown eye colour (44). C) A TIM within the HFE gene associated with hemochromatosis. D) A TIM in the TYR gene. There is a strong evidence that the rs1042602 A allele, associated with the absence of freckles, has been subject to positive selection in European populations (45).

Tables legend

Table S1. (separate xlsx file) Ancient DNA database used.

Table S2. (separate xlsx file) Temporal proportions obtained for: 1) 300 ancient individuals for the determination of the 5 ancient temporal components; 2) 602 radiocarbon-dated individuals with 5 temporal components (plotted in Fig. 2); 3) 300 ancient individuals for the determination of the 5 ancient and 3 modern temporal components.

Table S3. (separate xlsx file) 1) *p*-file output by ADMIXTURE with the allele frequencies associated to the 5 ancient components; 2) Corresponding Plink fam file; 3) Time series obtained for the TIMs.

Table S4. (separate xlsx file) TPS results for the ancient database with eight temporal components. Filtering column P for "No" shows the 495 individuals in the reference panel, from which 400 random ones are selected in each run.

Table S5. (separate xlsx file) TPS results for the 1st degree kinship pairs.

Data Legend

Data S1. (separate zip file)

Plink files with genetic data for the 961 ancient individuals and 147,229 SNPs used in this study.

Data S2. (separate zip file)

Plink files with the 8 temporal components generated in this study.

References and Notes:

1. E. MatisooSmith, K. A. Horsburgh, DNA for Archaeologists. *DNA for Archaeologists*, 1-233 (2012).
2. N. K. Schaefer, B. Shapiro, New middle chapter in the story of human evolution. *Science* **365**, 981-982 (2019).
3. W. F. Libby, E. C. Anderson, J. R. Arnold, Age Determination by Radiocarbon Content: World-Wide Assay of Natural Radiocarbon. *Science* **109**, 227-228 (1949).
4. R. E. Taylor, O. BarYosef, Radiocarbon Dating: An Archaeological Perspective, 2nd Edition. *Radiocarbon Dating: An Archaeological Perspective, 2nd Edition*, 1-404 (2014).
5. M. M. Be, Duchemin, B., Browne, E., Wu, S.C., Chechev, V., Helmer, R., & Schonfeld, E., *Table of radionuclides comments on evaluations* (CEA Saclay, France, 1999).
6. C. B. Ramsey, Radiocarbon dating: Revolutions in understanding. *Archaeometry* **50**, 249-275 (2008).
7. R. M. Jacobi, T. F. G. Higham, C. B. Ramsey, AMS radiocarbon dating of Middle and Upper Palaeolithic bone in the British Isles: improved reliability using ultrafiltration. *J Quaternary Sci* **21**, 557-573 (2006).
8. F. Brock, T. Higham, P. Ditchfield, C. B. Ramsey, Current Pretreatment Methods for Ams Radiocarbon Dating at the Oxford Radiocarbon Accelerator Unit (Orau). *Radiocarbon* **52**, 103-112 (2010).
9. B. Kromer *et al.*, Regional (CO2)-C-14 offsets in the troposphere: Magnitude, mechanisms, and consequences. *Science* **294**, 2529-2532 (2001).
10. E. Q. Alves, K. Macario, P. Ascough, C. BronkRamsey, The Worldwide Marine Radiocarbon Reservoir Effect: Definitions, Mechanisms, and Prospects. *Rev Geophys* **56**, 278-305 (2018).
11. P. Ascough, G. Cook, A. Dugmore, Methodological approaches to determining the marine radiocarbon reservoir effect. *Prog Phys Geog* **29**, 532-547 (2005).
12. M. Biddle, Kjolbye-Biddle, B. (2001) Repton and the `great heathen army'. in *13th Viking congress* (Oxbow, 1997 Nottingham), pp 45-96.
13. C. L. Jarman, M. Biddle, T. Higham, C. B. Ramsey, The Viking Great Army in England: new dates from the Repton charnel. *Antiquity* **92**, 183-199 (2018).
14. T. F. G. Higham, R. M. Jacobi, C. B. Ramsey, AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* **48**, 179-195 (2006).
15. L. Dupree, *Prehistoric research in Afghanistan (1959-1966)*, Transactions of the American Philosophical Society (American Philosophical Society, Philadelphia,, 1972), pp. 84 p.
16. K. Douka *et al.*, Direct radiocarbon dating and DNA analysis of the Darra-i-Kur (Afghanistan) human temporal bone. *J Hum Evol* **107**, 86-93 (2017).
17. M. E. Allentoft *et al.*, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172 (2015).
18. I. Olalde *et al.*, The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190-196 (2018).
19. P. Korlevic, S. Talamo, M. Meyer, A combined method for DNA analysis and radiocarbon dating from a single sample. *Sci Rep-Uk* **8** (2018).
20. P. Moorjani *et al.*, A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *P Natl Acad Sci USA* **113**, 5652-5657 (2016).

21. M. Meyer *et al.*, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226 (2012).
22. D. Graur *et al.*, On the immortality of television sets: "function" in the Human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* **5**, 578-590 (2013).
23. U. Esposito, R. Das, S. Syed, M. Pirooznia, E. Elhaik, Ancient Ancestry Informative Markers for Identifying Fine-Scale Ancient Population Structure in Eurasians. *Genes (Basel)* **9** (2018).
24. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664 (2009).
25. W. McLaren *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
26. N. M. Myres *et al.*, A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet* **19**, 95-101 (2011).
27. P. A. Underhill *et al.*, Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet* **18**, 479-484 (2010).
28. P. A. Underhill *et al.*, The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet* **23**, 124-131 (2015).
29. Q. Fu *et al.*, The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).
30. I. Mathieson *et al.*, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).
31. B. Salomonsen, Die Varby-funde. Ein Beitrag zur Kenntnis der altesten Trichterbecherkultur in Schonen. *Acta Archaeologica*, 55-95 (1970).
32. N. V. Skak-Nielsen, The neolithisation of Scandinavia: How did it happen? *Adoranten* (2004).
33. B. S. Paulsson, Scandinavian Models: Radiocarbon Dates and the Origin and Spreading of Passage Graves in Sweden and Denmark. *Radiocarbon* **52**, 1002-1017 (2010).
34. T. Sørensen, "Passage and Passing: Movement, Boundary, and Presence in Neolithic Mortuary Architecture" in Materialities of Passing: Explorations in Transformation, Transition and Transience, R. A. Bjerrgaard P, Sørensen TF, Ed. (Routledge, New York, 2016), pp. 65-77.
35. S. Gasparri, *Italia longobarda. Il regno, i Franchi, il papato* (Laterza, 2012).
36. C. E. G. Amorim *et al.*, Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat Commun* **9**, 3547 (2018).
37. Z. Hofmanova *et al.*, Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A* **113**, 6886-6891 (2016).
38. F. Broushaki *et al.*, Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499-503 (2016).
39. W. Y. Wang *et al.*, Association Between Cartilage Intermediate Layer Protein and Degeneration of Intervertebral Disc A Meta-analysis. *Spine* **41**, E1244-E1248 (2016).
40. T. Gunther *et al.*, Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* **16**, e2003703 (2018).
41. P. Mellars, A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* **439**, 931-935 (2006).
42. C. Genomes Project *et al.*, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

43. E. Elhaik *et al.*, Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat. Commun.* **5**, 1-12 (2014).
44. P. Sulem *et al.*, Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* **39**, 1443-1452 (2007).
45. H. Eiberg *et al.*, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet* **123**, 177-187 (2008).

Acknowledgments:

Funding: This work was partially supported by the EPSRC Doctoral Training Partnership Grant EP/N509735/1 to G.H. and the MRC (MR/R025126/1) to E.E..

Author contributions: E.E. conceptualized the study. U.E., E.E. led the study. U.E., M.P. curated the data. U.E., G.H., G.A., A.M.D. analyzed the data. U.E., C.B., E.E. interpreted the results. U.E. wrote TPS code. U.E., C.B., E.E. wrote the manuscript.

Competing interests: E.E. Consults the DNA Diagnostics Center.

Data and materials availability: All ancient data are available in the supplementary materials. TPS is available at https://github.com/UEsposito/TPSpaper.

Supplementary Materials:

Materials and Methods

Figures S1-S13

Tables S1-S5 (Supplementary Tables.zip)

Data S1 (Data S1.zip)

Data S2 (Data S2.zip)

References (*43-46*)