bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Inferring Migration Networks with Time-Lagged F_2 Statistics

Giulio Isacchini[1,4], **Takashi Okada**[1,3,8], **Clemens Schmid**[5], **Divyaratan Popli**[5], **Benjamin M Peter**[5,7,9], **Stephan Schiffels**[5], and **Oskar Hallatschek**[1,2,4,*]

[1]Department of Physics, University of California Berkeley, Berkeley, CA, USA
[2]Department of Integrative Biology, University of California Berkeley, Berkeley, CA, USA
[3]Institute for Life and Medical Sciences, Kyoto University, Japan
[4]Peter Debye Institute for Soft Matter Physics, Leipzig University, 04103 Leipzig, Germany
[5]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[7]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles CA, USA
[8]Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Wako, Japan
[9]Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles, CA, USA
[*]Corresponding Author

Major demographic events, such as population bottlenecks, founder effects, range expansions, and admixture events, have left lasting imprints on human genetic diversity. Ancient DNA (aDNA) sequencing now makes it increasingly possible to observe these signals across time, opening new avenues to address long-standing questions in human demographic history. Yet, deciphering this genetic archive of demographic history is challenging due to high levels of noise, and the complex ways in which demographic processes shape genetic variation. Here, leveraging the linear time evolution of the expectation of neutral allele frequencies, we develop a method to uncover systematic patterns of gene flow from metapopulation time-series data. We show that directional migration rates can be inferred via linear regression on time-dependent genetic dissimilarity between populations, quantified by an extended F_2 statistic evaluated between successive time points. Despite small sample sizes, the method reliably infers migration rates from simulated data by integrating information across multiple time slices. Applied to aDNA sampled from the last 6000 years, we recover signals of well-documented migrations and infer an ancient pan-European migration network. While complementing existing tools that estimate static ancestry proportions, our framework tracks how ancestry is dynamically redistributed through time.

Correspondence: ***ohallats@berkeley.edu***

Introduction

Ancient genomes preserve a noisy record of past migrations, expansions, replacements, and extinctions—events once thought irretrievable but now increasingly accessible through ancient DNA (aDNA) analysis. Over the past decade, extensive work has reconstructed aspects of human demographic history from such data (1–7). Yet most studies have focused on specific spatial or temporal contexts or on time-aggregated signals, such as ancestry proportions inferred from present-day or temporally averaged populations (8–12). How mobility patterns evolve continuously through time remains poorly understood, in part because sparse and uneven genome sampling limits statistical power, and the high levels of missing data typical in aDNA complicate cross-sample comparisons.

A recent study (13) showed that allele frequency changes over the past 5,000 years can be explained by gene flow alone, without invoking detectable natural selection. Motivated by this result, we develop a method to infer temporal patterns of gene flow under a neutral evolutionary model, using allele frequency time series. Our framework operates directly on genomic data and avoids assumptions about the existence or composition of ancestral source populations that underlie admixture proportion estimates (14).

Previous efforts have mapped effective migration surfaces from genetic data (2, 4, 15), including large-scale applications to human diversity (5), but these methods omit temporal information and cannot resolve changes in migration over time. Spatio-temporal interpolation of genetic profiles helps resolve individual-level migration events (16), with limited quantitative interpretation at the population-level. Alternatively, continuous diffusive spread has been modeled via partial differential equations (17), a formulation that struggles to capture discrete, long-distance migration events.

We exploit the fact that gene-flow impacts the covariance of neutral allele-frequency changes between populations over time. Without interaction, allele frequency changes should be uncorrelated, resulting in zero covariance. At the other extreme, if two populations are highly connected - effectively behaving as a single mixed population - their allele frequencies will be strongly correlated (Fig. 1). In between these two extremes, our method seeks to infer a gene flow matrix $\mathbf{A}_{ij}$ by measuring the covariance of allele frequencies across different regions.

This general approach has already proven effective in reconstructing transmission networks during the Covid-19 pandemic (18, 19), where SARS-CoV-2 lineage frequency data was used to estimate inter-regional importation rates using a Kalman Filter that accounted for sampling noise and genetic drift under Gaussian assumptions (20). While such assumptions were appropriate given the dense sequencing data, they are less applicable to ancient DNA (aDNA), where sparse sampling results in many regions (currently) lacking data for extended periods. Nonetheless, the larger human genome and the presence of recombination yield a far greater number of approximately independent lineages compared to the SARS-CoV-2 dataset, potentially enabling inference even with far fewer sampled genomes.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

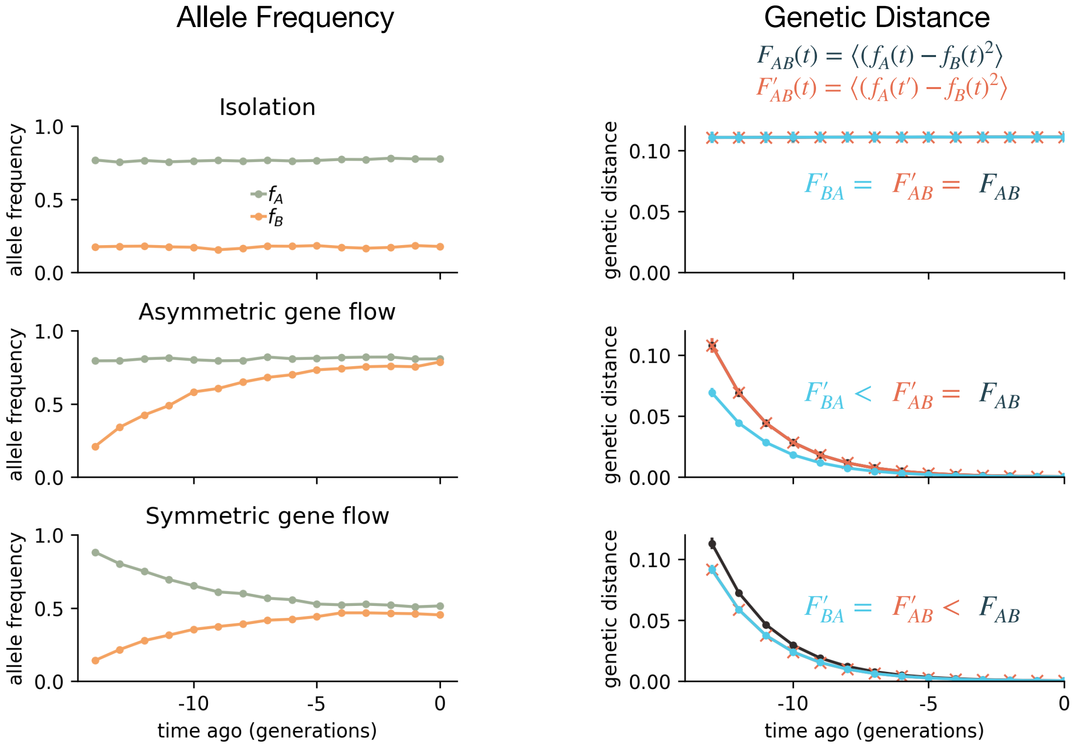


Fig. 1. A decline in genetic distance between populations signals gene flow. In ancient DNA research, genetic distance is often measured as the mean squared allele frequency difference between populations, known as the F-statistic. Here, we introduce a time-lagged extension, F', which quantifies the genetic distance between populations sampled at different time points (see main text for definitions). We illustrate how directional gene flow is reflected in the temporal decay of F and F' by simulating allele frequency trajectories for two populations, A and B, under three scenarios: no migration (top), strong unidirectional migration (middle), and symmetric migration (bottom). All alleles were initialized with $f_A = 0.8$ and $f_B = 0.2$. Migration reduces pre-existing allele frequency differences (left panels), leading to a decay in all genetic distance measures (right panels). Comparing F and F' reveals directionality: with unidirectional flow from A to B (middle), F'_{BA} declines faster than F_{AB} and F'_{AB} (right middle panel), whereas symmetric migration produces no such asymmetry (bottom panels). Simulation details are provided in the SI. Default parameters: $\sigma = 1$, $T = 15$, $l = -1$, $N_D = 2$, $N_{\text{traj}} = 5000$, $N_{\text{pop}} = 1000$, $N_S = 5000$. Migration matrices: $\mathbf{A} = I$ for isolation, $\mathbf{A} = [(1,0),(0.2,0.8)]$ for asymmetric flow, and $\mathbf{A} = [(0.9,0.1),(0.1,0.9)]$ for symmetric flow. Only one representative lineage pair is shown.

To harness this abundance of lineages, we aggregate information across genomes to quantify allele frequency covariance between populations. These covariances are expressed through the F_2 statistics, which measure the genetic distance between populations (21, 22). Widely used in aDNA studies (23, 24), F_2 statistics are typically computed pairwise across populations as a preprocessing step for dimensionality reduction via multidimensional scaling (25). Their empirical estimators naturally incorporate binomial sampling noise, and their calculation using block-bootstrap methods is robust to linkage disequilibrium (26).

We show that the F_2-distance between populations sampled at different times, which we call time-lagged F_2' statistics, is linearly related to the inter-population genetic distances between populations sampled at the same time. The linear coefficients characterize gene flow and can be efficiently inferred via a maximum likelihood approach. We estimate confidence intervals by taking into account uncertainty and potential biases due to sparse sampling, temporal dating and genetic linkage. The strength of this method lies in its efficiency and scalability. As the efforts in sequencing will continue to expand the pool of aDNA samples available, our method will be able to extend to longer time scales than those studied in the present work.

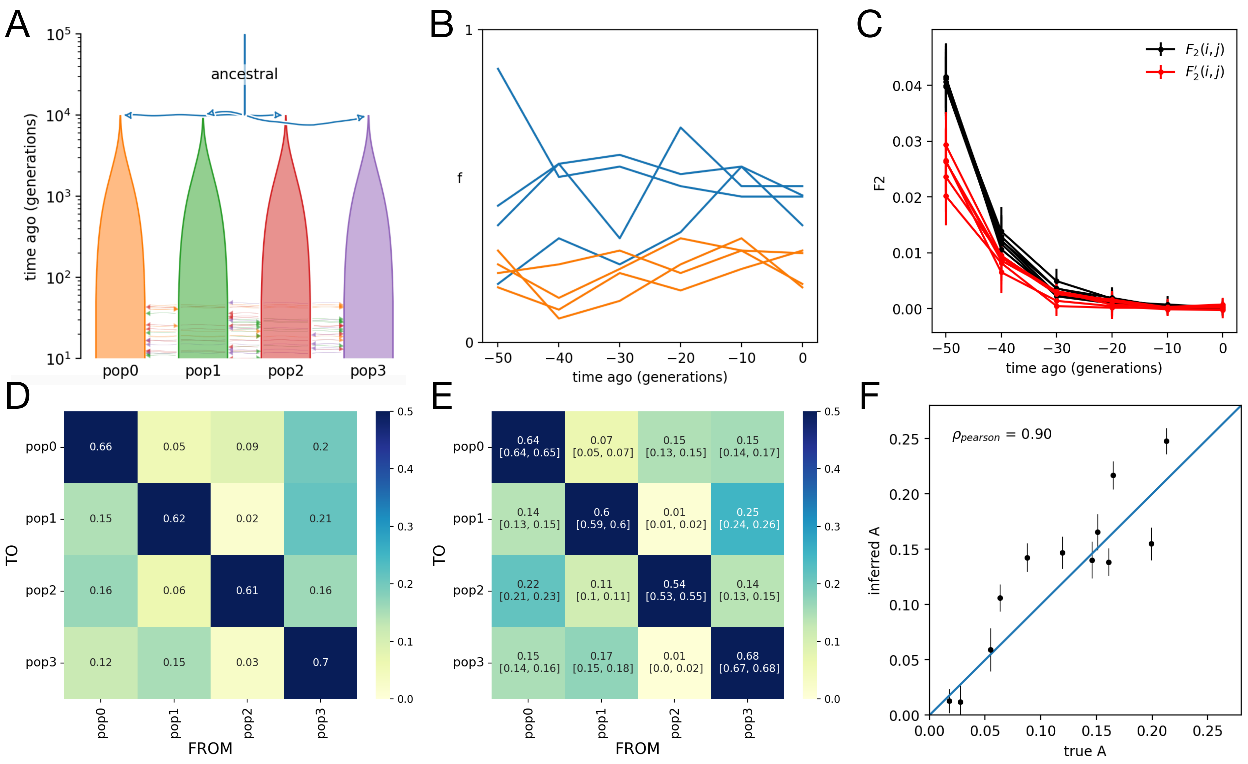
Results

Model and Inference. Genetic dissimilarity between populations can be robustly quantified via F_2-statistics (21, 22). Given two populations A and B sampled at time t, the F_2 statistic is defined as $F_{AB}(t) = \mathbb{E}\left[(f_A(t) - f_B(t))^2\right]$ where the expectation is taken over all sites in the genome, and f_K indicates the frequency of the allele at the site in population K. Typically, analyses involve comparing predetermined populations in an all-to-all fashion, without considering the dating of the samples.

Here, we study how F_2-statistics change over time and demonstrate that it is possible to deduce directional gene flow from these temporal changes. We introduce an additional quantity, that we refer to as time-delayed F-statistics and it is defined as $F'_{AB}(t) = \mathbb{E}\left[(f_A(t') - f_B(t))^2\right]$ where $t' = t + \Delta t$ is delayed by a factor Δt. Fig. 1 illustrates how the strength, direction, and asymmetry of gene flow each imprint characteristic signatures on the statistics F_{AB}, F'_{AB}, and F_{BA} - the key signal leveraged by our inference framework.

We model the dynamics of neutral alleles in a population composed of n sub-populations as a function of genetic drift and migration, similarly to (18). Due to the sample sparsity of aDNA, accurate estimation of single allele frequen-

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



D

	pop0	pop1	pop2	pop3
pop0	0.66	0.05	0.09	0.2
pop1	0.15	0.62	0.02	0.21
pop2	0.16	0.06	0.61	0.16
pop3	0.12	0.15	0.03	0.7

E

	pop0	pop1	pop2	pop3
pop0	0.64 [0.64, 0.65]	0.07 [0.05, 0.07]	0.15 [0.13, 0.15]	0.15 [0.14, 0.17]
pop1	0.14 [0.13, 0.15]	0.6 [0.59, 0.6]	0.01 [0.01, 0.02]	0.25 [0.24, 0.26]
pop2	0.22 [0.21, 0.23]	0.11 [0.1, 0.11]	0.54 [0.53, 0.55]	0.14 [0.13, 0.15]
pop3	0.15 [0.14, 0.16]	0.17 [0.15, 0.18]	0.01 [0.0, 0.02]	0.68 [0.67, 0.68]

Fig. 2. Simulations to assess inference on sparse, incomplete data. (A) Four diploid human populations are simulated as descendants of a single ancestral population of size $N = 10^4$. A single chromosome with a uniform recombination rate is simulated using standard `stdpopsim` parameters, with 20% of data randomly removed. Initial isolation and exponential expansion increase divergence between populations. Migration, defined by a random matrix $\mathbf{A}$, begins 100 generations before the present. Sampling starts at this time and is repeated every 10 generations, matching the time discretization of the empirical analysis. **(B)** Example allele frequency trajectories for two loci (blue and orange); each line of the same colour corresponds to a different population. **(C)** Genetic divergence (F_2 and F_2' statistics) decreases during the migration period. **(D)** Reference migration matrix used in the simulation. **(E)** Inferred migration matrix with confidence intervals given by lower and upper quartiles. **(F)** The inferred matrix closely matches the reference; the Pearson correlation between off-diagonal true and average inferred coefficients is $\rho_{\text{Pearson}} = 0.9$.

cies is challenging. For this reason, we rewrite the dynamics in terms of F_2 summary statistics, which can be robustly estimated (22). In Methods, we show that the temporal change in F_2 statistics is given by

$$F'_{ik} - F'_{ii} = \sum_j \mathbf{A}_{ij} \left(F_{jk} - F_{ji} \right) \,, \tag{1}$$

where $\mathbf{A}_{ij}$ is an $n \times n$ migration matrix: its elements are non-negative and, within each row, sum up to one, $\sum_j \mathbf{A}_{ij} = 1$. The coefficient $\mathbf{A}_{ij}$ represents the proportion of individuals in population i that are replaced by migrants from population j within the considered time period, thus capturing migration rates between regions. $\mathbf{A}_{ij}$ can also be interpreted as a matrix of backward-in-time probabilities that a lineage moves from population i to population j in the time period Δt. The linearity of Eq. 1 implies that the migration matrix $\mathbf{A}_{ij}$ can be inferred via linear regression. In constructing confidence intervals, we take into account uncertainty and potential biases due to random sampling of individuals, uncertainty in the temporal assignment, and genetic linkage across loci, see Fig. S1 and Methods for additional details.

Simulation-based validation of inference method. We study the applicability of our method to sparse, incomplete aDNA data in a simulated scenario. We perform simulations using msprime (27, 28) and population parameter specifications from the stdpopsim project (29, 30) for a diploid human population with a single chromosome with uniform recombination rate. As shown in Fig. 2A, we simulate a scenario where an ancestral human population of size $N_e = 10^4$ splits into 4 deeply diverged populations, which begin to expand exponentially for 5000 generations. Migration is defined via a random migration matrix $\mathbf{A}_{\text{true}}$ whose rows sum up to 1 and begin 50 generations before the end of the simulations. At the same time, we begin sampling 20 individuals from each population at intervals of 10 generations apart. Given the approximate generation time of humans is $\sim 30y$, this reproduces sampling times of $\Delta t = 300y$, similar to those in the aDNA data sets analyzed below. We sample one chromosome per individual and introduce missing data at 20% probability, to reproduce the sampling noise of typical pseudohaploid aDNA data. Fig. 2B presents an example of the neutral allele frequency trajectories, which serve as the input for our approach. Fig. 2C illustrates the temporal variation of the F statistics during the sampling period. We refer to Fig. S2 , for an overview of how genetic divergence gets established in the time period when the populations are isolated.

In our simulation scenario, migration significantly reduces the genetic divergence between populations — divergence that initially emerged through genetic drift and population expansion. The correlation plot in Fig. 2F demonstrates strong agreement between inferred and true migration coefficients. An extended benchmark across a broad range of parameters using Wright–Fisher simulations is presented in Fig. S3.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

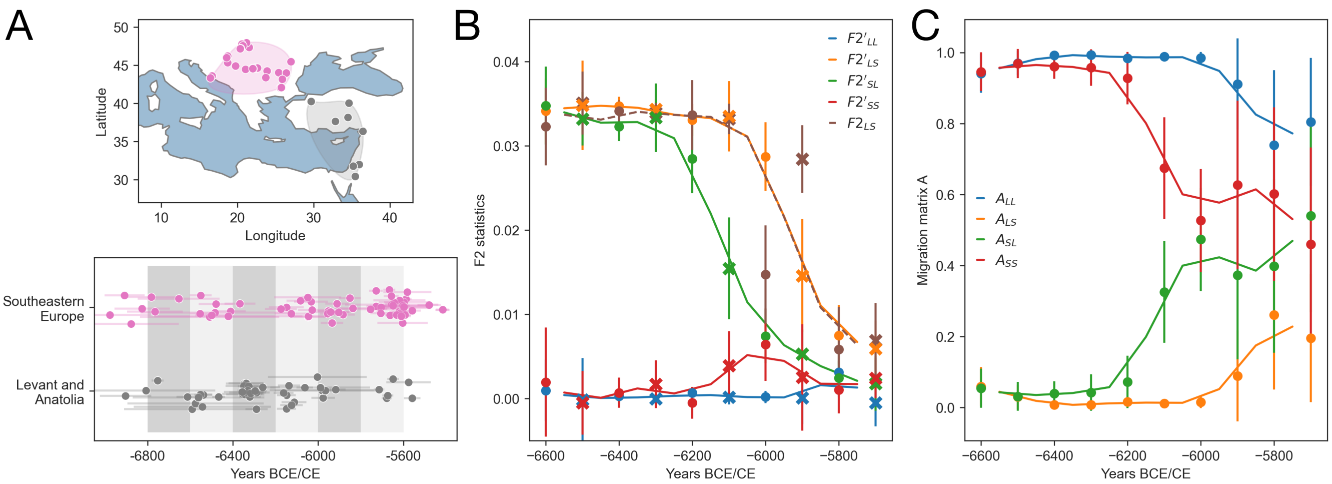


Fig. 3. Early Neolithic expansion through time. (A) Spatio-temporal distribution of the samples analyzed, focusing on the Early Neolithic expansion from the Levant and Anatolia into Southeastern Europe. Time windows of $\Delta t = 200$ y from 6800 to 5600 BCE yield six time steps, shown in shades of gray. **(B)** F_2 distances across time windows. The solid line identifies the average between the values computed between the six time steps and the corresponding values with respect to the overlapping delayed time windows computed at $dt = 100y$. **(C)** Inferred migration across time windows. As in the previous figure, the solid line represents the average across overlapping delayed time windows.

Early Neolithic expansion. As an initial application to real data, we examine the early Neolithic expansion—a well-documented migration from the Levant into Europe supported by both archaeological (31–33) and archaeogenetic evidence (34–37). For this analysis, we selected ancient genomic samples dated between 6800 and 5600 BCE from Southeastern Europe (SE Europe) and the Levant/Anatolia region (LevA) (see Methods for details on the data preparation (16, 38, 39)). We organized these samples into two distinct populations and established time intervals of 200 years each, as shown in Fig. 3A. We calculated F_2-statistics and time-delayed F_2'-statistics for all possible combinations of these populations.

In this simple scenario with just two populations, the patterns present in the F-statistics are directly represented in the inferred migration rates (Fig. 3C): The inferred migration rate from LevA to SE Europe (indicated by the green line) rises significantly above zero, while the coefficient from SE Europe to LevA (represented by the orange line) remains close to zero throughout the entire period. This observation aligns with the presence of substantial gene flow from the Levant to Southeastern Europe, consistent with the archaeological record as discussed in (40–42). For an estimation of the relaxation times, see Fig. S5.

The analysis indicates that observing the temporal dynamics of F_2-statistics can be helpful in characterizing the source of change in the genetic composition of a population and that our method can effectively identify known migration patterns.

Reconstructing Gene Flow Dynamics Across Nine Regions. In the previous section, we tested the validity of our method in a limited and well-understood spatio-temporal context. We now seek to characterize the change in the genetic composition of nine populations spanning a duration of 3000 years, ranging from 4000 to 1000 BCE, see Fig. 4A-B for the spatiotemporal distribution of the samples. In this time period, linguistic, archaeological and – most recently – genetic evidence hints towards a major migration from the Pontic steppes into Central Europe, potentially corresponding to the spread of the Corded Ware culture and Indo-European languages into Central Europe (8, 9, 12). Incidentally, the significance and magnitude of this event had not been appreciated by archaeologists prior to the insights gained through ancient DNA (43).

To overcome the scarcity of samples in the time frame under examination, we implement an interpolation technique for both the F_2 and F_2' statistics, see the methods section for additional details and supplementary Fig. S6. We infer the migration matrix $\mathbf{A}(t)$ on the interpolated values throughout the entire time span. In Fig. 4C we summarise the results of our inference. We use a resampling approach to display the variability of the inferred matrix elements and use a goodness-of-fit test to characterize time periods where our method can predict the temporal dynamics of F-statistics at sufficient accuracy, see Methods and Supplementary Fig. S7. We observe that our method works well within the period of 3500 to 2000 BCE. Outside this period, the scarcity of samples, in particular for populations of the Western Pontic Steppe and Scandinavia, impedes an accurate interpolation. In Fig. 4D we visualize the average inferred matrix elements more intuitively. In each subplot, we visualize a single row i of $\mathbf{A}$, which represents the proportion of migrations from all other populations j into population i through time.

In the time period of sufficient sample certainty, between 3500 and 2000 BCE, we find that Britain (blue) receives most of its genetic input from Central Europe (both East and West) and, by proxy, the Western Pontic Steppe. That is consistent with the well attested arrival of Steppe ancestry in Britain and Ireland in the third millennium BCE, together with the cultural markers of the Bell Beaker complex (10, 44, 45). Notably, the peak of Steppe-related inflow into Central Europe East occurs several hundred years earlier than the corresponding peak in Britain, consistent with a scenario in which

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

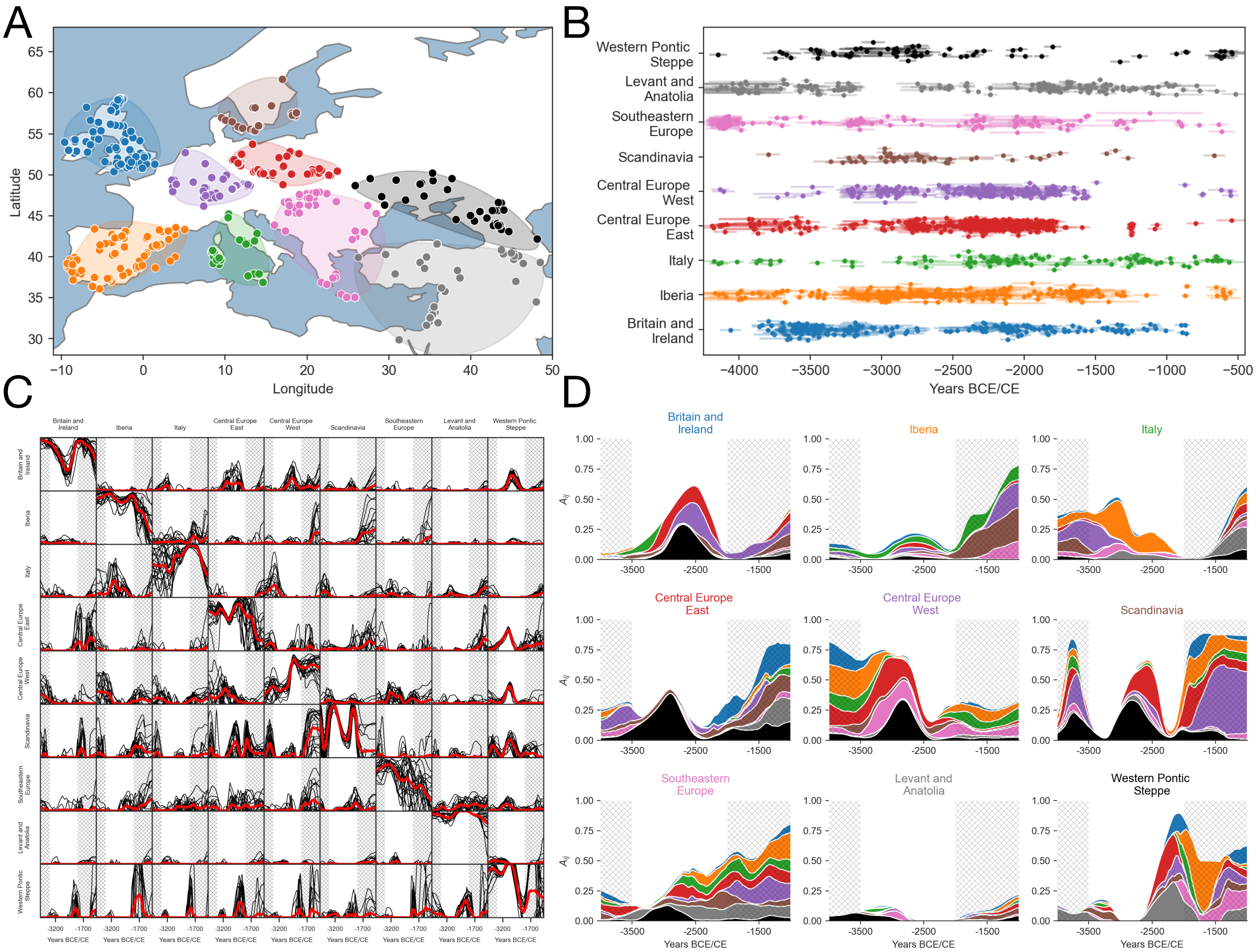


Fig. 4. Time-resolved reconstruction of migration networks across nine regions (A–B) Spatial and temporal distribution of the samples analyzed. **(C)** Time series for individual elements of the inferred migration matrix $\mathbf{A}$, where $\mathbf{A}_{ij}$ denotes the fraction of genetic material that population j receives from population i within a 300-year interval. Black curves show interpolations from a single resampling iteration; the red curve represents the mean across iterations. Shaded areas indicate periods of lower confidence, based on a goodness-of-fit test (see main text and Supplementary Information, Fig. S7). **(D)** Each subplot shows the fraction of genetic material a focal region imports from all other regions over 300-year intervals. Titles indicate the focal region; colors indicate source regions. Values are means over resampling iterations (red line in subfigure C). The x-axis shows time.

migrating Steppe groups moved through Central Europe before reaching the North Atlantic coastline. By the time gene flow into Britain peaks, the inferred influx comes predominantly via Central Europe East and West, with only a modest direct contribution from the Western Pontic Steppe. This pattern suggests that the groups reaching Britain were often descendants of Steppe herders who had already acquired substantial ancestry from resident European farmer populations during their westward expansion.

Within our approach we can also integrate information across multiple time points to infer migratory events over extended periods. In Fig. S8-S9, we apply this coarse-graining method to the Steppe-related migration event and, once again, observe a clear east-to-west genetic flow; see also the SI text for additional information.

For Italy (green) we surprisingly observe relatively strong gene flow from Iberia around 3000 BCE, with smaller contributions from Southeastern Europe and Central Europe West. However, given few data points and our aggregation of regions with a relatively distinct genomic history (Sardinia, Sicily, Northern Italy), uncertainty is high in this case (see Fig. 4C). From published research we know that an influx of Steppe-related ancestry is only expected later in the second half of the third millennium (46), in Sicily then indeed possibly from Iberia (47). Steppe ancestry does not arrive in sizable proportions in Sardinia (48). Iberia (orange) itself receives little genetic input during this period, which is in line with previous research (49). Steppe-related ancestry appears there occasionally from 2500 BCE onward (45, 47) and becomes a stable component only by the end of the third millennium (50).

Central Europe East (red) shows strong genetic influx from the Western Pontic Steppe, consistent with the steppe migration event and the emergence of the Corded Ware cultural complex from 3000 BCE (51–53). Similarly and to no surprise, Central Europe West (violet) receives genetic input primarily from the Western Pontic Steppe, Central Europe East, and Southeastern Europe in the first half of the third millen-

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

nium (8, 9, 54–56). The strong signal of ancestry flux from the West – from Britain, Italy and primarily Iberia – at around 3500 BCE is curious and potentially a methodological artifact. A similar signal has been observed already in (16), possibly explained by low levels of genomic differentiation among different Neolithic populations, creating a large geographic area of very similar genetic ancestry that is hard to resolve. Scandinavia (brown) shows contributions from both the Western Pontic Steppe and Central Europe East, most likely induced by the Steppe migration phenomenon before and after 2500 BCE (57, 58).

The result for Southeastern Europe (pink) is surprising in light of published research. The arrival of Steppe ancestry starts earlier here, and is a more gradual process compared to other parts of Europe (59–61). This may explain the weaker signal we observe from the East. The seemingly diverse influx from Northern, Western and Southern Europe throughout the third millennium, though, is hard to understand. Maybe this signal is driven by an elevated hunter-gatherer ancestry component with a potentially Central European source (62, 63). The Levant (grey) remains largely isolated in relation to the other regions considered here, with only a minor early contribution from the Western Pontic Steppe (61, 64–66). Finally, the Western Pontic Steppe (black) itself appears to receive genetic input from all other populations toward the end of the study period, specifically from the South, so Anatolia, the Southern Caucasus or even the Levant. Due to sparse, spatially uneven sampling the error bars are large, indicating lower confidence in this new unexpected result (12, 61).

Discussion

We presented a scalable and computationally efficient algorithm for inferring directional gene flow between populations. Its primary strength lies in its simplicity and minimal reliance on strong assumptions. Unlike methods that require the identification of ancestral populations or estimation of admixture proportions, our approach bypasses such prerequisites, making it broadly applicable and easy to implement. The algorithm's scalability stems from its linear-time interpolation scheme and pairwise-statistic framework, which allow it to handle increasingly large ancient DNA datasets without prohibitive computational costs.

Simulation results demonstrated that the method reliably detects migration signals under realistic sampling constraints and data sparsity. We validated its effectiveness using two well-studied episodes of human history: the Early Neolithic expansion and the Western Steppe Migration. In the first case, we recovered asymmetrical gene flow from the Levant into Southeastern Europe after 7000 BCE, consistent with archaeological expectations. For the Western Steppe Migration, we demonstrated how combining information across multiple time points allows reconstruction of time-varying gene flow from the Pontic Steppe into Central and Western Europe during the third millennium BCE.

The second case also illustrates how secondary waves of migration can dominate the ancestry flux into a region: Britain appears to receive substantial Steppe-derived ancestry via Central Europe rather than directly from the Western Pontic Steppe. This is qualitatively in line with models of range expansions in which invading groups repeatedly pick up ancestry from resident populations along their route, so that later migrants are already strongly admixed by the time they reach the expansion front (67). Our framework makes such multistep, temporally structured ancestry flows directly visible in the time series of migration matrices.

Our current inference scheme relies on pairwise time bin comparisons, which may miss finer-grained temporal structure. We envision that extending the framework to a probabilistic time-series model — such as a Hidden Markov Model — could improve sensitivity and robustness by explicitly incorporating correlations across adjacent time points. This represents a promising avenue for future development.

A limitation of the current model is that it can only detect migration through its effect on reducing preexisting genetic divergence. As such, it cannot capture migration events that do not alter allele frequencies, such as exchanges between already well-mixed populations. Over longer time scales, and in contexts with high mutation rates, incorporating *de novo* mutations into the framework could expand its resolution and applicability. A more explicit incorporation of rare genetic variation may also provide an avenue for future improvement (68).

As genome sequencing efforts continue to improve spatial and temporal sampling density, we expect our method to become a valuable tool for reconstructing the dynamics of gene flow across broad regions and over extended historical periods.

Beyond estimating migration rates, our framework also provides a natural connection to classical models of admixture proportions. In the case where migration matrices $\mathbf{A}(t)$ vary over time, the cumulative effect of gene flow across multiple time steps is captured by the product $\mathbf{A}(t)\mathbf{A}(t-1)\cdots\mathbf{A}(1)$. Each row of this product matrix describes the expected proportion of ancestry in a given population at time t that originates from other populations at the initial time point. This product thus generalizes the notion of admixture proportions (8, 22, 69), allowing for temporally and spatially heterogeneous migration dynamics. In contrast to static models of ancestry composition, our rate-based formulation enables tracking the evolving structure and directionality of gene flow across both space and time.

Methods

Data preparation. For the empirical analysis we reused a dataset compiled for and described in (16), based on the Allen Ancient DNA Resource (AADR, V50.0, Dataverse V4.0, Oct 10 2021) (38, 39). This dataset includes samples from ancient human individuals genotyped at up to 1.23 million positions. Starting from the AADR we extracted samples with >25,000 recovered autosomal single-nucleotide polymorphism (SNPs), low contamination estimates, and favourable AADR quality assessment. Among closely related individuals or samples from the same individ-

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ual we only kept the best sample. We removed SNPs in sections of the genome prone to linkage disequilibrium (70, 71) and below a minor allele frequency threshold of 5%.

For the spatial binning of samples into subpopulations, we defined spatial polygons based on Western-Eurasian geographical macro-regions. Their temporal binning was based on their archaeological dating information. Dating uncertainty was addressed with age resampling as in (16): If an ancient genomic sample came with one or multiple radiocarbon ages, then we drew age samples from the post-calibration distribution. If the dating information was limited to a range determined from contextual, archaeological information, then we drew samples from a uniform distribution within this range.

The dynamics of neutral alleles in a metapopulation. Consider a population decomposed into n sub-populations, distinguished by location (different cities or districts) or any other feature, such as age or ethnicity. We assume to have available frequency time series of many unlinked alleles.

Under neutrality, we can assume that the frequency $X_i(t+\Delta t)$ of a particular neutral allele in population i at time $t+1$ depends *linearly* on the allele frequencies $\{X_j(t)\}_{j=1\dots n}$ at some earlier time t,

$$X_i(t+\Delta t) = \sum_{j=1}^{n} \mathbf{A}_{ij} X_j(t) + \eta_i \,, \tag{2}$$

where $\mathbf{A}$ is the backward-migration matrix, i.e. $\mathbf{A}_{ij}$ is the proportion of individuals in population i at time $t+1$ that originated from population j at time t. Thus, the elements are non-negative and, within each row, sum up to one, $\sum_j \mathbf{A}_{ij} = 1$. η_i represents random genetic drift. For now, we do not need to know anything about the noise term, except that its expectation vanishes.

The linearity of Eq. 2 follows from the extensivity of the conditional expectation of the frequency of neutral alleles: Starting with half as many mutants at earlier time is expected to lead to half as many mutants at later time, i.e., $\mathbb{E}[X_i(t+\Delta t)|X_j(t) = \alpha x_j] = \alpha\, \mathbb{E}[X_i(t+\Delta t)|X_j(t) = x_j]$ for any $0 < \alpha < 1$.

The rows of $\mathbf{A}$ have to sum up to 1 because neutral alleles are not expected to change if their frequency is the same in all populations, $\mathbb{E}[X_i(t+\Delta t)|X_j(t) = x] = x$ for any $0 < x < 1$. Finally, negative matrix elements are excluded because they can generate negative expectations.

An alternative way of writing Eq. 2 is

$$X_i(t+\Delta t) - X_i(t) = \sum_{j=1}^{n} \mathbf{A}_{ij}\,[X_j(t) - X_i(t)] + \eta_i \,, \tag{3}$$

which explicitly shows that (i) the frequency in j only influences the frequency in i if $X_i \neq X_j$, and (ii) that a larger value of $\mathbf{A}_{ij} > 0$ leads to a faster convergence of X_i to X_j. The coefficient $\mathbf{A}_{ij}$ thus measures how population j influences population i.

Expressing the dynamics in terms of F_2 statistics. Our goal to infer $\mathbf{A}_{ij}$ can in principle be done directly using the dynamical equation for individual allele frequencies, Eq. 2, see also (18), via a state-space model approach that directly takes into account sampling noise, or via (total) least squares estimation by comparing predicted and observed lineage frequencies through time,

$$\mathbf{A}^{(\mathrm{LS})} = \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{\mu,|t'-t|\le T,i} \left[X_i^{\mu}\left(t'+\Delta t\right) - \sum_j \mathbf{A}_{ij} X_j^{\mu}\left(t'\right) \right]^2 , \tag{4}$$

where the index μ labels the different genomic sites.

The downside of these direct approaches is that (I) it is very hard to reliably infer the frequency of individual alleles, (II) in the low sampling regime the noise is non-Gaussian and needs to be treated accordingly, and (III) one has to deal with complicated biases introduced by which set of genomes actually have a given allele. While the last point can perhaps be avoided by focusing on the set of overlapping alleles present in all genomes, this would reduce the amount of data that can be used, especially in ancient DNA.

Here, we take a simpler approach and derive an equation that constrains the mean allele frequency differences between different populations at different times. This leads to F_2 statistics and a dynamical extension thereof, which can be robustly inferred from the data.

For notational simplicity, we introduce the following abbreviations: $X_i \equiv X_i(t)$, $X'_i \equiv X_i(t+\Delta t)$, $F_{ij} \equiv \overline{(X_i - X_j)^2}$, $F'_{ij} \equiv \overline{(X'_i - X_j)^2}$, where $\overline{\xi}$ refers to the combination of taking the expectation value of ξ and averaging over all biallelic sites in a genome. In practice, if we have many sites, averaging over all alleles should generate a narrow Gaussian around the expectation value. Note that $F_{ij} = F_2(X_i, X_j)$ is the standard F_2 statistics (22), which corresponds to the Euclidean mean squared length of the allele frequency vectors of populations i and j normalized by the number of alleles. The primed quantity F' corresponds to the time-dependent version of this statistics, i.e., it is the normalized Euclidean mean squared length of the allele frequency vectors of population i at time τ and population j at time t.

Derivation of the equations of motion in terms of F_2 statistics. We derive F'_{ik} for two different populations, which will give us a simple linear equation that is well suited for the inference of $\mathbf{A}$. We begin with the simple identity

$$X'_i - X_k = (X'_i - X_i) + (X_i - X_k)\,. \tag{5}$$

Taking the square of both sides and averaging over all alleles, we obtain

$$F'_{ik} = F'_{ii} + F_{ik} + 2\overline{(X'_i - X_i)\,(X_i - X_k)}\,. \tag{6}$$

The average on the right hand side can be rewritten using our

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

basic equation of motion Eq. 3 as

$$\begin{aligned} 2\overline{(X_i'-X_i)(X_i-X_k)} &= 2\sum_j \mathbf{A}_{ij}\overline{(X_j-X_i)(X_i-X_k)} \\ &= \sum_j \mathbf{A}_{ij}\left(F_{jk}-F_{ji}-F_{ik}\right) \\ &= -F_{ik}+\sum_j \mathbf{A}_{ij}\left(F_{jk}-F_{ji}\right)\ , \end{aligned} \tag{7}$$

where in the final step we used the row normalization $\sum_j \mathbf{A}_{ij}=1$ and in the second-to-last step we used the identity:

$$\begin{aligned} 2\overline{(X_j-X_i)(X_k-X_i)} &= -\overline{[(X_j-X_i)-(X_k-X_i)]^2} \\ &\quad +F_{ji}+F_{ki} = -F_{jk}+F_{ji}+F_{ki}\ , \end{aligned}$$

which corresponds to the statistics commonly known as $F_3(X_i;X_j,X_k)$. Inserting Eq. 7 into Eq. 6, we obtain the useful linear equation

$$F'_{ik}-F'_{ii} = \sum_j \mathbf{A}_{ij}\left(F_{jk}-F_{ji}\right)\ . \tag{8}$$

This is a set of $n\times(n-1)$ equations to fix the $n\times(n-1)$ components of the $\mathbf{A}$ matrix, given we can observe all F' and all F. Remarkably, these equations do not contain a term from genetic drift, which cancels on the left hand side in the difference of two F' statistics. Note that, while F_{ij} is symmetric in its indices, F'_{ij} is not because of the time delay.

Estimation of F2 statistics from data. Following (22) we compute an unbiased estimator for the F_2 statistics under the assumption of binomial noise at a single site between populations A and B as:

$$\hat{F}_2(A,B) = (a-b)^2 - \frac{n_A(N_A-n_A)}{N_A-1} - \frac{n_B(N_B-n_B)}{N_B-1} \tag{9}$$

where n_K are the allele counts in the population K, N_K is the total counts of both alleles, $k=n_K/N_K$ is their frequency. In order to produce confidence intervals in the inference of $\mathbf{A}_{ij}$ we compute the average of $\hat{F}_2(A,B)$ in blocks of nearby SNPs (typically 5000) and we bootstrap blocks.

Inference of the migration matrix A via convex optimization. The linear equation Eq. 8 suggests a straight forward approach to infer $\mathbf{A}$, namely via linear regression under the constraints $\mathbf{A}_{ij}>0$, $\sum_j \mathbf{A}_{ij}=1$. This will give us an MLE estimate of $\mathbf{A}$ since all F's should be approximately Gaussian.

Minimizing the square difference of left and right hand sides of Eq. 8 for all i and k reduces to the following optimization problem

$$\begin{aligned} \mathbf{A}^{(\mathrm{F2})} = \underset{\mathbf{A}}{\mathrm{argmin}} \quad & \sum\nolimits_{ik}\left[(F'_{ik}-F'_{ii})-\sum\nolimits_j A_{ij}(F_{jk}-F_{ji})\right]^2 \\ \text{s.t.} \quad & \sum\nolimits_j A_{ij}=1 \\ & A_{ij}>0 \end{aligned} \tag{10}$$

We rewrite this constrained least-squares optimization problem to emphasize that each row of the matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$ can be optimized independently. Define the row vector $\mathbf{y}_i \equiv (F'_{i1}-F'_{ii},\ldots,F'_{in}-F'_{ii})$ and the matrix $\mathbf{X}^{(i)}$ with elements $X^{(i)}_{jk}=F_{jk}-F_{ji}$. Then, each row vector $\mathbf{a}_i\equiv(A_{i1},\ldots,A_{in})$ of the matrix $\mathbf{A}$ can be estimated independently via constrained linear regression,

$$\mathbf{a}_i^{(\mathrm{F2})} = \underset{\mathbf{a}_i}{\mathrm{argmin}} \left\|\mathbf{y}_i-\mathbf{a}_i\mathbf{X}^{(i)}\right\|^2 \quad \text{s.t. } \mathbf{a}_i\mathbf{I}=1, \quad \mathbf{a}_i>0$$

This is a standard non-negative least-squares problem with a simplex constraint, and can be solved efficiently and independently for each row $(\mathbf{a}_i)_{i=1,\ldots,n}$.

Estimation of Confidence Intervals. We estimate confidence intervals via multiple iterations of resampling of our data with respect to different sources of potential uncertainty: (I) individual variability, (II) dating uncertainty, (III) genetic uncertainty. For each sample, we have a metapopulation assignment and its date. We first define a binning Δt and a period of interest $\mathbf{p}=(T_0,T_0+\Delta t,T_0+2\Delta t,\ldots)$ for the analysis. As there is an underlying uncertainty in dating, we seek to take it into account via temporal resampling following the same approach of (16). The main inference algorithm is subdivided into the following steps:

- Time resampling of dates.
- Subsampling of $N-1$ individuals from the pool of N available samples in each time bin and metapopulation. If only 2 individuals are available, keep both.
- Assignment of samples to bins in $\mathbf{p}$ and compute count vector $\mathbf{n}_t$ and total count vector $\mathbf{N}_t$ for each time bin.
- Computation of mean $F2$ and $F2'$ statistics via blockwise resampling.
- Infer migration matrix $\mathbf{A}$ via the constrained convex optimization problem as described in eq. 9.

By iteratively following the described process several times, we can compute confidence intervals for the migration matrix elements $\mathbf{A}_{ij}$ that take into account the sampling of individuals, time uncertainty in the samples dating, and uncertainty in the computation of F2 summary statistics.

Interpolation Analysis. The interpolation analysis algorithm comprises several steps, which are delineated as follows: (I) Sampling multiple F and F' values across the entire period via temporal resampling. (II) Interpolating using

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a convolution filter. For the initial sampling step (I), we do the following:

- Randomly sample the start of the first time bin from the uniform distribution $\mathcal{U}(-4150, -3850)$.
- Sample dates for all individuals and allocate them into corresponding bins.
- Calculate F and F' for each pair of populations across all time bins.

By iteratively conducting the aforementioned process, two sets of tuples, denoted as $D_F = (t, F)$ and $D_{F'} = (t, F')$, are constructed, where t represents the start time of the time bin. In the subsequent step, resampling and interpolation of the points from D_F and $D_{F'}$ are performed to assess the robustness of the interpolation line. The interpolation algorithm (II) entails the following steps:

- Define dense binning of size $30y$ across the period.
- Sample $N = 50$ tuples from D_F and $D_{F'}$ and allocate the values to the bins.
- Compute the average within each bin and infer the value from neighboring bins if no value is present within the bin.
- Perform multiple (N=5) convolutions with a uniform filter of size of 10 bins to perform smoothing.

Although using convolutions for smoothing can be susceptible to edge issues, we can mitigate this problem by extending the time period under consideration. While more advanced interpolation techniques could be applied, such as Gaussian processes or Kalman smoothing, we leave this for future research and development.

Goodness-of-fit test. The inference of migration matrix elements is carried out one row at a time. For each row, the model's performance can be assessed by calculating the Pearson correlation between the left and right-hand sides of Eq. 2 using the inferred matrix. In Fig. S7A, we conduct this goodness-of-fit test across the entire time span for each row of $\mathbf{A}$. We observe that, especially for the Western Pontic Steppe and Scandinavia metapopulations, there is a period where the method shows reduced accuracy. In Figure S7B, we plot the left and right-hand sides of Eq. 2 during the period of high accuracy (3500 - 2000 BCE) to provide a clearer visualization of the model's predictions.

Lower Dimensional Representation. The representation of genetic samples into lower dimensions to better visualize genetic relationships is typically performed via PCA projection. This procedure is equivalent to Multidimensional Scaling on the matrix of all pairwise F-statistics (25). While this procedure is typically performed independently of temporal sampling information, it is possible to extend this idea to include time. At each time point t, we perform centering of the matrix $\mathbf{M}(t) = -0.5\mathbf{C}\mathbf{F}(t)\mathbf{C}$ with $C = \mathbf{I} - (1/n)\mathbf{1}$ is a centering matrix and $\mathbf{I}$, $\mathbf{1}$ the identity matrix and a matrix of ones. We calculate the average of the first two eigenvectors of $\mathbf{M}(t)$ over time and then project all $\mathbf{M}(t)$ into a two-dimensional space. In Fig. S7C-D, we present the resulting lower-dimensional representation of our populations. As the eigenvectors and eigenvalues of $\mathbf{M}(t)$ evolve over time, as shown in Supplementary Fig. S7E-F-G, we opted to compute the average of the first two eigenvectors across time to determine a consistent projection for all $\mathbf{M}(t)$. This method yields a lower-dimensional representation of the populations' relative genetic trajectories. Populations with a higher genetic exchange between them are brought closer together, while those that remain genetically isolated move further apart.

Code and Data Availability

All code and data to reproduce the results of this paper can be found at 10.5281/zenodo.18740249 and https://github.com/Hallatscheklab/interactions-from-fluctuations/

Acknowledgements

We thank all members of the Hallatschek lab (past and present) for helpful discussions and advice on the project. This work was supported by the National Institute of General Medical Sciences of the NIH under award R01GM149827 and by a Humboldt Professorship of the Alexander von Humboldt Foundation. TO acknowledges support from JSPS KAKENHI (Grant Numbers JP22K03453, JP22K06347, 25K09723, and 25K07168) and the RIKEN iTHEMS Program. GI acknowledges support from Humboldt Research Fellowship.

References

1. Gideon S. Bradburd and Peter L. Ralph. Spatial population genetics: It's about time. *Annual Review of Ecology, Evolution, and Systematics*, 50(1):427–449, 2019. doi: 10.1146/annurev-ecolsys-110316-022659.
2. Desislava Petkova, John Novembre, and Matthew Stephens. Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48(1):94–100, Jan 2016. ISSN 1546-1718. doi: 10.1038/ng.3464.
3. Gideon S. Bradburd, Peter L. Ralph, and Graham M. Coop. A spatial framework for understanding population structure and admixture. *PLOS Genetics*, 12(1):1–38, 01 2016. doi: 10.1371/journal.pgen.1005703.
4. Hussein Al-Asadi, Desislava Petkova, Matthew Stephens, and John Novembre. Estimating recent migration and population-size surfaces. *PLOS Genetics*, 15(1):1–21, 01 2019. doi: 10.1371/journal.pgen.1007908.
5. Benjamin M Peter, Desislava Petkova, and John Novembre. Genetic Landscapes Reveal How Human Genetic Diversity Aligns with Geography. *Molecular Biology and Evolution*, 37(4):943–951, 11 2019. ISSN 0737-4038. doi: 10.1093/molbev/msz280.
6. Ke Wang, Iain Mathieson, Jared O'Connell, and Stephan Schiffels. Tracking human population structure through time from whole genome sequences. *PLOS Genetics*, 16(3):1–24, 03 2020. doi: 10.1371/journal.pgen.1008552.
7. Harald Ringbauer, Alexander Kolesnikov, David L Field, and Nicholas H Barton. Estimating Barriers to Gene Flow from Distorted Isolation-by-Distance Patterns. *Genetics*, 208(3):1231–1245, 03 2018. ISSN 1943-2631. doi: 10.1534/genetics.117.300638.
8. Wolfgang Haak, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, Susanne Nordenfelt, Eadaoin Harney, Kristin Stewardson, Alissa Mittnik, Eszter Bánffy, Christos Economou, Michael Francken, Susanne Friederich, Rafael Garrido Pena, Fredrik Hallgren, Valery Khartanovich, Aleksandr Khokhlov, Michael Kunst, Pavel Kuznetsov, Harald Meller, Oleg Mochalov, Vayacheslav Moiseyev, Nicole Nicklisch, Sandra L. Pichler, Roberto Risch, Manuel A. Rojo Guerra, Christina Roth, Anna Szécsényi-Nagy, Joachim Wahl, Matthias Meyer, Johannes Krause, Dorcas Brown, David Anthony, Alan Cooper, Kurt Werner Alt, and David Reich. Massive migration from the steppe was a source for indo-european languages in europe. *Nature*, 522(7555):207–211, Jun 2015. ISSN 1476-4687. doi: 10.1038/nature14317.
9. Morten E. Allentoft, Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, Peter B. Damgaard, Hannes Schroeder, Torbjörn Ahlström, Lasse Vinner, Anna-Sapfo Malaspinas, Ashot Margaryan, Tom Higham, David Chivall, Niels

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Lynnerup, Lise Harvig, Justyna Baron, Philippe Della Casa, Paweł Dąbrowski, Paul R. Duffy, Alexander V. Ebel, Andrey Epimakhov, Karin Frei, Mirosław Furmanek, Tomasz Gralak, Andrey Gromov, Stanisław Gronkiewicz, Gisela Grupe, Tamás Hajdu, Radosław Jarysz, Valeri Khartanovich, Alexandr Khokhlov, Viktória Kiss, Jan Kolář, Aivar Kriiska, Irena Lasak, Cristina Longhi, George McGlynn, Algimantas Merkevicius, Inga Merkyte, Mait Metspalu, Ruzan Mkrtchyan, Vyacheslav Moiseyev, László Paja, György Pálfi, Dalia Pokutta, Łukasz Pospieszny, T. Douglas Price, Lehti Saag, Mikhail Sablin, Natalia Shishlina, Václav Smrčka, Vasilii I. Soenov, Vajk Szeverényi, Gusztáv Tóth, Synaru V. Trifanova, Liivi Varul, Magdolna Vicze, Levon Yepiskoposyan, Vladislav Zhitenev, Ludovic Orlando, Thomas Sicheritz-Pontén, Søren Brunak, Rasmus Nielsen, Kristian Kristiansen, and Eske Willerslev. Population genomics of bronze age eurasia. *Nature*, 522(7555):167–172, June 2015. ISSN 1476-4687. doi: 10.1038/nature14507.

10. Nick Patterson, Michael Isakov, Thomas Booth, Lindsey Büster, Claire-Elise Fischer, Iñigo Olalde, Harald Ringbauer, Ali Akbari, Olivia Cheronet, Madeleine Bleasdale, Nicole Adamski, Eveline Altena, Rebecca Bernardos, Selina Brace, Nasreen Broomandkhoshbacht, Kimberly Callan, Francesca Candilio, Brendan Culleton, Elizabeth Curtis, Lea Demetz, Kellie Sara Duffett Carlson, Ceiridwen J. Edwards, Daniel M. Fernandes, M. George B. Foody, Suzanne Freilich, Helen Goodchild, Aisling Kearns, Ann Marie Lawson, Iosif Lazaridis, Matthew Mah, Swapan Mallick, Kirsten Mandl, Adam Micco, Megan Michel, Guillermo Bravo Morante, Jonas Oppenheimer, Kadir Toykan Özdoğan, Lijun Qiu, Constanze Schattke, Kristin Stewardson, J. Noah Workman, Fatma Zalzala, Zhao Zhang, Bibiana Agustí, Tim Allen, Katalin Almássy, Luc Amkreutz, Abigail Ash, Christèle Baillif-Ducros, Alistair Barclay, László Bartosiewicz, Katherine Baxter, Zsolt Bernert, Jan Blažek, Mario Bodružić, Philippe Boissinot, Clive Bonsall, Pippa Bradley, Marcus Brittain, Alison Brookes, Fraser Brown, Lisa Brown, Richard Brunning, Chelsea Budd, Josip Burmaz, Sylvain Canet, Silvia Carnicero-Cáceres, Morana Čaušević Bully, Andrew Chamberlain, Sébastien Chauvin, Sharon Clough, Natalija Čondić, Alfredo Coppa, Oliver Craig, Matija Črešnar, Vicki Cummings, Szabolcs Czifra, Alžběta Danielisová, Robin Daniels, Alex Davies, Philip de Jersey, Jody Deacon, Csilla Deminger, Peter W. Ditchfield, Marko Dizdar, Miroslav Dobeš, Miluše Dobisíková, László Domboróczki, Gail Drinkall, Ana Đukić, Michal Ernée, Christopher Evans, Jane Evans, Manuel Fernández-Götz, Slavica Filipović, Andrew Fitzpatrick, Harry Fokkens, Chris Fowler, Allison Fox, Zsolt Gallina, Michelle Gamble, Manuel R. González Morales, Borja González-Rabanal, Adrian Green, Katalin Gyenesei, Diederick Habermehl, Tamás Hajdu, Derek Hamilton, James Harris, Chris Hayden, Joep Hendriks, Bénédicte Hernu, Gill Hey, Milan Horňák, Gábor Ilon, Eszter Istvánovits, Andy M. Jones, Martina Blečić Kavur, Kevin Kazek, Robert A. Kenyon, Amal Khreisheh, Viktória Kiss, Jos Kleijne, Mark Knight, Lisette M. Kootker, Péter F. Kovács, Anita Kozubová, Gabriella Kulcsár, Valéria Kulcsár, Christophe Le Pennec, Michael Legge, Matt Leivers, Louise Loe, Olalla López-Costas, Tom Lord, Dženi Los, James Lyall, Ana B. Marín-Arroyo, Philip Mason, Damir Matošević, Andy Maxted, Lauren McIntyre, Jacqueline McKinley, Kathleen McSweeney, Bernard Meijlink, Balázs G. Mende, Marko Menđušić, Milan Metlička, Sophie Meyer, Kristina Mihovilić, Lidija Milasinovic, Steve Minnitt, Joanna Moore, Geoff Morley, Graham Mullan, Margaréta Musilová, Benjamin Neil, Rebecca Nicholls, Mario Novak, Maria Pala, Martin Papworth, Cécile Paresys, Ricky Patten, Domagoj Perkić, Krisztina Pesti, Alba Petit, Katarína Petriščáková, Coline Pichon, Catriona Pickard, Zoltán Pilling, T. Douglas Price, Siniša Radović, Rebecca Redfern, Branislav Resutík, Daniel T. Rhodes, Martin B. Richards, Amy Roberts, Jean Roefstra, Pavel Sankot, Alena Šefčáková, Alison Sheridan, Sabine Skae, Miroslava Šmolíková, Krisztina Somogyi, Ágnes Somogyvári, Mark Stephens, Géza Szabó, Anna Szécsényi-Nagy, Tamás Szeniczey, Jonathan Tabor, Károly Tankó, Clenis Tavarez Maria, Rachel Terry, Biba Teržan, Maria Teschler-Nicola, Jesús F. Torres-Martínez, Julien Trapp, Ross Turle, Ferenc Ujvári, Menno van der Heiden, Petr Veleminsky, Barbara Veselka, Zdeněk Vytlačil, Clive Waddington, Paula Ware, Paul Wilkinson, Linda Wilson, Rob Wiseman, Eilidh Young, Joško Zaninović, Andrej Žitňan, Carles Lalueza-Fox, Peter de Knijff, Ian Barnes, Peter Halkon, Mark G. Thomas, Douglas J. Kennett, Barry Cunliffe, Malcolm Lillie, Nadin Rohland, Ron Pinhasi, Ian Armit, and David Reich. Large-scale migration into britain during the middle to late bronze age. *Nature*, 601(7894):588–594, December 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-04287-4.

11. Joscha Gretzinger, Duncan Sayer, Pierre Justeau, Eveline Altena, Maria Pala, Katharina Dulias, Ceiridwen J Edwards, Susanne Jodoin, Laura Lacher, Susanna Sabin, Åshild J Vågene, Wolfgang Haak, S Sunna Ebenesersdóttir, Kristjan H S Moore, Rita Radzeviciute, Kara Schmidt, Selina Brace, Martina Abenhus Bager, Nick Patterson, Luka Papac, Nasreen Broomandkhoshbacht, Kimberly Callan, Éadaoin Harney, Lora Iliev, Ann Marie Lawson, Megan Michel, Kristin Stewardson, Fatma Zalzala, Nadin Rohland, Stefanie Kappelhoff-Beckmann, Frank Both, Daniel Winger, Daniel Neumann, Lars Saalow, Stefan Krabath, Sophie Beckett, Melanie Van Twest, Neil Faulkner, Chris Read, Tabatha Barton, Joanna Caruth, John Hines, Ben Krause-Kyora, Ursula Warnke, Verena J Schuenemann, Ian Barnes, Hanna Dahlström, Jane Jark Clausen, Andrew Richardson, Elizabeth Popescu, Natasha Dodwell, Stuart Ladd, Tom Phillips, Richard Mortimer, Faye Sayer, Diana Swales, Allison Stewart, Dominic Powlesland, Robert Kenyon, Lilian Ladle, Christina Peek, Silke Grefen-Peters, Paola Ponce, Robin Daniels, Cecily Spall, Jennifer Woolcock, Andy M Jones, Amy V Roberts, Robert Symmons, Anooshka C Rawden, Alan Cooper, Kirsten I Bos, Tom Booth, Hannes Schroeder, Mark G Thomas, Agnar Helgason, Martin B Richards, David Reich, Johannes Krause, and Stephan Schiffels. The anglo-saxon migration and the formation of the early english gene pool. *Nature*, pages 1–8, 21 September 2022. ISSN 0028-0836. doi: 10.1038/s41586-022-05247-2.

12. Iosif Lazaridis, Nick Patterson, David Anthony, Leonid Vyazov, Romain Fournier, Harald Ringbauer, Iñigo Olalde, Alexander A. Khokhlov, Egor P. Kitov, Natalia I. Shishlina, Sorin C. Ailincăi, Danila S. Agapov, Sergey A. Agapov, Elena Batieva, Baitanayev Bauyrzhan, Zsolt Bereczki, Alexandra Buzhilova, Piya Changmai, Andrey A. Chizhevsky, Ion Ciobanu, Mihai Constantinescu, Marietta Csányi, János Dani, Peter K. Dashkovskiy, Sándor Évinger, Anatoly Faifert, Pavel Flegontov, Alin Frînculeasa, Mădălina N. Frînculeasa, Tamás Hajdu, Tom Higham, Paweł Jarosz, Pavol Jelínek, Valeri I. Khartanovich, Eduard N. Kirginekov, Viktória Kiss, Alexandera Kitova, Alexeiy V. Kiyashko, Jovan Koledin, Arkady Korolev, Pavel Kosintsev, Gabriella Kulcsár, Pavel Kuznetsov, Rabadan Magomedov, Aslan M. Mamedov, Eszter Melis, Vyacheslav Moiseyev, Erika Molnár, Janet Monge, Octav Negrea, Nadezhda A. Nikolaeva, Mario Novak, Maria Ochir-Goryaeva, György Pálfi, Sergiu Popovici, Marina P. Rykun, Tatyana M. Savenkova, Vladimir P. Semibratov, Nikolai N. Seregin, Alena Šefčáková, Raikhan S. Mussayeva, Irina Shingiray, Vladimir N. Shirokov, Angela Simalcsik, Kendra Sirak, Konstantin N. Solodovnikov, Judit Tárnoki, Alexey A. Tishkin, Viktor Trifonov, Sergey Vasilyev, Ali Akbari, Esther S. Brielle, Kim Callan, Francesca Candilio, Olivia Cheronet, Elizabeth Curtis, Olga Flegontova, Lora Iliev, Aisling Kearns, Denise Keating, Ann Marie Lawson, Matthew Mah, Adam Micco, Megan Michel, Jonas Oppenheimer, Lijun Qiu, J. Noah Workman, Fatma Zalzala, Anna Szécsényi-Nagy, Pier Francesco Palamara, Swapan Mallick, Nadin Rohland, Ron Pinhasi, and David Reich. The genetic origin of the indo-europeans. *Nature*, 639(8053):132–142, February 2025. ISSN 1476-4687. doi: 10.1038/s41586-024-08531-5.

13. Alexis Simon and Graham Coop. The contribution of gene flow, selection, and genetic drift to five thousand years of human allele frequency change. *Proceedings of the National Academy of Sciences*, 121(9):e2312377121, 2024.

14. Fernando Racimo, Jessie Woodbridge, Ralph M. Fyfe, Martin Sikora, Karl-Göran Sjögren, Kristian Kristiansen, and Marc Vander Linden. The spatiotemporal spread of human migrations during the european holocene. *Proceedings of the National Academy of Sciences*, 117(16):8989–9000, 2020. doi: 10.1073/pnas.1920051117.

15. Joseph Marcus, Wooseok Ha, Rina Foygel Barber, and John Novembre. Fast and flexible estimation of effective migration surfaces. *eLife*, 10:e61927, jul 2021. ISSN 2050-084X. doi: 10.7554/eLife.61927.

16. Clemens Schmid and Stephan Schiffels. Estimating human mobility in holocene western eurasia with large-scale ancient genomic data. *Proceedings of the National Academy of Sciences*, 120(9):e2218375120, 2023. doi: 10.1073/pnas.2218375120.

17. Rasa A Muktupavela, Martin Petr, Laure Ségurel, Thorfinn Korneliussen, John Novembre, and Fernando Racimo. Modeling the spatiotemporal spread of beneficial alleles using ancient genomes. *eLife*, 11:e73767, dec 2022. ISSN 2050-084X. doi: 10.7554/eLife.73767.

18. Takashi Okada, Giulio Isacchini, QinQin Yu, and Oskar Hallatschek. Uncovering heterogeneous intercommunity disease transmission from neutral allele frequency time series. *Proceedings of the National Academy of Sciences*, 122(48):e2500663122, 2025.

19. Daniel B Weissman. Rapid evolution reveals long-range spread of sars-cov-2. *Proceedings of the National Academy of Sciences*, 123(6):e2533093123, 2026.

20. QinQin Yu, Joao A Ascensao, Takashi Okada, COVID-19 Genomics UK (COG-UK) Consortium, Olivia Boyd, Erik Volz, and Oskar Hallatschek. Lineage frequency time series reveal elevated levels of genetic drift in sars-cov-2 transmission in england. *Plos Pathogens*, 20 (4):e1012090, 2024.

21. David Reich, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price, and Lalji Singh. Reconstructing indian population history. *Nature*, 461(7263):489–494, Sep 2009. ISSN 1476-4687. doi: 10.1038/nature08365.

22. Nick Patterson, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. Ancient Admixture in Human History. *Genetics*, 192(3):1065–1093, 11 2012. ISSN 1943-2631. doi: 10.1534/genetics.112.145037.

23. Montgomery Slatkin. Statistical methods for analyzing ancient dna from hominins. *Current Opinion in Genetics & Development*, 41:72–76, 2016. ISSN 0959-437X. doi: https://doi.org/10.1016/j.gde.2016.08.004. Genetics of human origin.

24. Benjamin M Peter. Admixture, Population Structure, and F-Statistics. *Genetics*, 202(4): 1485–1501, 02 2016. ISSN 1943-2631. doi: 10.1534/genetics.115.183913.

25. Benjamin M. Peter. A geometric relationship of f2, f3 and f4-statistics with principal component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377 (1852):20200413, 2022. doi: 10.1098/rstb.2020.0413.

26. FMTA Busing and R Van Der Leeden. Delete-m jackknife for unequal m. *Statistics and Computing*, 9(1):3–8, January 1999.

27. Franz Baumdicker, Gertjan Bisschop, Daniel Goldstein, Graham Gower, Aaron P Ragsdale, Georgia Tsambos, Sha Zhu, Bjarki Eldon, E Castedo Ellerman, Jared G Galloway, Ariella L Gladstein, Gregor Gorjanc, Bing Guo, Ben Jeffery, Warren W Kretzschumar, Konrad Lohse, Michael Matschiner, Dominic Nelson, Nathaniel S Pope, Consuelo D Quinto-Cortés, Murillo F Rodrigues, Kumar Saunack, Thibaut Sellinger, Kevin Thornton, Hugo van Kemenade, Anthony W Wohns, Yan Wong, Simon Gravel, Andrew D Kern, Jere Koskela, Peter L Ralph, and Jerome Kelleher. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3):iyab229, 2022.

28. Jerome Kelleher, Alison M Etheridge, and Gilean McVean. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS computational biology*, 12(5): e1004842, 2016.

29. Jeffrey R Adrion, Christopher B Cole, Noah Dukler, Jared G Galloway, Ariella L Gladstein, Graham Gower, Christopher C Kyriazis, Aaron P Ragsdale, Georgia Tsambos, Franz Baumdicker, Jedidiah Carlson, Reed A Cartwright, Arun Durvasula, Ilan Gronau, Bernard Y Kim, Patrick McKenzie, Philipp W Messer, Ekaterina Noskova, Diego Ortega-Del Vecchyo, Fernando Racimo, Travis J Struck, Simon Gravel, Ryan N Gutenkunst, Kirk E Lohmueller, Peter L Ralph, Daniel R Schrider, Adam Siepel, Jerome Kelleher, and Andrew D Kern. A community-maintained standard library of population genetic models. *eLife*, 9:e54967, jun 2020. ISSN 2050-084X. doi: 10.7554/eLife.54967.

30. M. Elise Lauterbur, Mariz Izabel A. Cavassim, Ariella L. Gladstein, Graham Gower, Nathaniel S. Pope, Georgia Tsambos, Jeffrey Adrion, Saurabh Belsare, Arjun Biddanda, Victoria Caudill, Jean Cury, Ignacio Echevarria, Benjamin C. Haller, Ahmed R. Hasan, Xin Huang, Leonardo Nicola Martin Iasi, Ekaterina Noskova, Jana Obšteter, Vitor Antonio Corrêa Pavinato, Alice Pearson, David Peede, Manolo F. Perez, Murillo F. Rodrigues, Chris C. R. Smith, Jeffrey P. Spence, Anastasia Teterina, Silas Tittes, Per Unneberg, Juan Manuel Vazquez, Ryan K. Waples, Anthony Wilder Wohns, Yan Wong, Franz Baumdicker, Reed A. Cartwright, Gregor Gorjanc, Ryan N. Gutenkunst, Jerome Kelleher, Andrew D. Kern, Aaron P. Ragsdale, Peter L. Ralph, Daniel R. Schrider, and Ilan Gronau. Expanding the stdpopsim species catalog, and lessons learned for realistic genome simulations. *bioRxiv*, 2022. doi: 10.1101/2022.10.29.514266.

31. Jean-Pierre Bocquet-Appel, Stephan Naji, Marc Vander Linden, and Janusz K. Kozlowski. Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. *Journal of Archaeological Science*, 36(3):807–820, 2009. doi: 10.1016/j.jas.2008.11.004.

32. Marc Vander Linden and Fabio Silva. Dispersals as demographic processes: testing and

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

describing the spread of the neolithic in the balkans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1816):20200231, 2021. doi: 10.1098/rstb.2020.0231.

33. Mehmet Ozdogan. Reconsidering the early neolithic of anatolia. recent recoveries, some excerpts and generalities. *ANTHROPOLOGIE*, 2022.
34. Pontus Skoglund, Helena Malmström, Maanasa Raghavan, Jan Storå, Per Hall, Eske Willerslev, M. Thomas P. Gilbert, Anders Götherström, and Mattias Jakobsson. Origins and genetic legacy of neolithic farmers and hunter-gatherers in europe. *Science*, 336(6080): 466–469, 2012. doi: 10.1126/science.1216304.
35. Iosif Lazaridis, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H. Sudmant, Joshua G. Schraiber, Sergi Castellano, Mark Lipson, Bonnie Berger, Christos Economou, Ruth Bollongino, Qiaomei Fu, Kirsten I. Bos, Susanne Nordenfelt, Heng Li, Cesare de Filippo, Kay Prüfer, Susanna Sawyer, Cosimo Posth, Wolfgang Haak, Fredrik Hallgren, Elin Fornander, Nadin Rohland, Dominique Delsate, Michael Francken, Jean-Michel Guinet, Joachim Wahl, George Ayodo, Hamza A. Babiker, Graciela Bailliet, Elena Balanovska, Oleg Balanovsky, Ramiro Barrantes, Gabriel Bedoya, Haim Ben-Ami, Judit Bene, Fouad Berrada, Claudio M. Bravi, Francesca Brisighelli, George B. J. Busby, Francesco Cali, Mikhail Churnosov, David E. C. Cole, Daniel Corach, Larissa Damba, George van Driem, Stanislav Dryomov, Jean-Michel Dugoujon, Sardana A. Fedorova, Irene Gallego Romero, Marina Gubina, Michael Hammer, Brenna M. Henn, Tor Hervig, Ugur Hodoglugil, Aashish R. Jha, Sena Karachanak-Yankova, Rita Khusainova, Elza Khusnutdinova, Rick Kittles, Toomas Kivisild, William Klitz, Vaidutis Kučinskas, Alena Kushniarevich, Leila Laredj, Sergey Litvinov, Theologos Loukidis, Robert W. Mahley, Béla Melegh, Ene Metspalu, Julio Molina, Joanna Mountain, Klemetti Näkkäläjärvi, Desislava Nesheva, Thomas Nyambo, Ludmila Osipova, Jüri Parik, Fedor Platonov, Olga Posukh, Valentino Romano, Francisco Rothhammer, Igor Rudan, Ruslan Ruizbakiev, Hovhannes Sahakyan, Antti Sajantila, Antonio Salas, Elena B. Starikovskaya, Ayele Tarekegn, Draga Toncheva, Shahlo Turdikulova, Ingrida Uktveryte, Olga Utevska, René Vasquez, Mercedes Villena, Mikhail Voevoda, Cheryl A. Winkler, Levon Yepiskoposyan, Pierre Zalloua, Tatijana Zemunik, Alan Cooper, Cristian Capelli, Mark G. Thomas, Andres Ruiz-Linares, Sarah A. Tishkoff, Lalji Singh, Kumarasamy Thangaraj, Richard Villems, David Comas, Rem Sukernik, Mait Metspalu, Matthias Meyer, Evan E. Eichler, Joachim Burger, Montgomery Slatkin, Svante Pääbo, Janet Kelso, David Reich, and Johannes Krause. Ancient human genomes suggest three ancestral populations for present-day europeans. *Nature*, 513(7518):409–413, Sep 2014. ISSN 1476-4687. doi: 10.1038/nature13673.
36. Wolfgang Haak, Oleg Balanovsky, Juan J. Sanchez, Sergey Koshel, Valery Zaporozhchenko, Christina J. Adler, Clio S. I. Der Sarkissian, Guido Brandt, Carolin Schwarz, Nicole Nicklisch, Veit Dresely, Barbara Fritsch, Elena Balanovska, Richard Villems, Harald Meller, Kurt W. Alt, Alan Cooper, and the Genographic Consortium. Ancient dna from european early neolithic farmers reveals their near eastern affinities. *PLOS Biology*, 8(11):1–16, 11 2010. doi: 10.1371/journal.pbio.1000536.
37. Dilek Koptekin, Ayça Aydoğan, Cansu Karamurat, N. Ezgi Altınışık, Kıvılcım Başak Vural, D. Deniz Kazancı, Ayça Küçükakdağ Doğu, Damla Kaptan, Hasan Can Gemici, Eren Yüncü, Hannah M. Moots, Gülsün Umurtak, Refik Duru, Erkan Fidan, Özlem undefinedevik, Burçin Erdoğu, Taner Korkut, Christopher J. Knüsel, Scott Haddow, Clark Spencer Larsen, Rana Özbal, Fokke Gerritsen, Eylem Özdoğan, Ali Akbaba, Uygar Ozan Usanmaz, Yasin Cemre Derici, Mine Uçmazoğlu, Flora Jay, Mehmet Özdoğan, Anders Götherström, Yılmaz Selim Erdal, Anna-Sapfo Malaspinas, undefinediğdem Atakuman, Füsun Özer, and Mehmet Somel. Out-of-anatolia: Cultural and genetic interactions during the neolithic expansion in the aegean. *Science*, 388(6754), June 2025. ISSN 1095-9203. doi: 10.1126/science.adr3326.
38. Swapan Mallick, Adam Micco, Matthew Mah, Harald Ringbauer, Iosif Lazaridis, Iñigo Olalde, Nick Patterson, and David Reich. The allen ancient dna resource (aadr) a curated compendium of ancient human genomes. *Scientific Data*, 11(1), February 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-03031-7.
39. Swapan Mallick and David Reich. The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes, 2021.
40. Jean Guilaine. The neolithization of mediterranean europe: Mobility and interactions from the near east to the iberian peninsula. In *The Oxford Handbook of Neolithic Europe*. Oxford University Press, 03 2015. ISBN 9780199545841. doi: 10.1093/oxfordhb/9780199545841.013.064.
41. Johannes Müller. Movement of plants, animals, ideas, and people in south-east europe. In *The Oxford Handbook of Neolithic Europe*. Oxford University Press, 03 2015. ISBN 9780199545841. doi: 10.1093/oxfordhb/9780199545841.013.063.
42. Caroline Malone. The neolithic in mediterranean europe. In *The Oxford Handbook of Neolithic Europe*. Oxford University Press, 03 2015. ISBN 9780199545841. doi: 10.1093/oxfordhb/9780199545841.013.069.
43. Martin Furholt. Massive migrations? the impact of recent adna studies on our view of third millennium europe. *European journal of archaeology*, 21(2):159–191, 2018.
44. Lara M. Cassidy, Rui Martiniano, Eileen M. Murphy, Matthew D. Teasdale, James Mallory, Barrie Hartwell, and Daniel G. Bradley. Neolithic and bronze age migration to ireland and establishment of the insular atlantic genome. *Proceedings of the National Academy of Sciences*, 113(2):368–373, December 2015. ISSN 1091-6490. doi: 10.1073/pnas.1518445113.
45. Iñigo Olalde, Selina Brace, Morten E. Allentoft, Ian Armit, Kristian Kristiansen, Thomas Booth, Nadin Rohland, Swapan Mallick, Anna Szécsényi-Nagy, Alissa Mittnik, Eveline Altena, Mark Lipson, Iosif Lazaridis, Thomas K. Harper, Nick Patterson, Nasreen Broomandkhoshbacht, Yoan Diekmann, Zuzana Faltyskova, Daniel Fernandes, Matthew Ferry, Eadaoin Harney, Peter de Knijff, Megan Michel, Jonas Oppenheimer, Kristin Stewardson, Alistair Barclay, Kurt Werner Alt, Corina Liesau, Patricia Ríos, Concepción Blasco, Jorge Vega Miguel, Roberto Menduiña García, Azucena Avilés Fernández, Eszter Bánffy, Maria Bernabò-Brea, David Billoin, Clive Bonsall, Laura Bonsall, Tim Allen, Lindsey Büster, Sophie Carver, Laura Castells Navarro, Oliver E. Craig, Gordon T. Cook, Barry Cunliffe, Anthony Denaire, Kirsten Egging Dinwiddy, Natasha Dodwell, Michal Ernée, Christopher Evans, Milan Kuchařík, Joan Francès Farré, Chris Fowler, Michiel Gazenbeek, Rafael Garrido Pena, María Haber-Uriarte, Elżbieta Haduch, Gill Hey, Nick Jowett, Timothy Knowles, Ken Massy, Saskia Pfrengle, Philippe Lefranc, Olivier Lemercier, Arnaud Lefebvre, César Heras Martínez, Virginia Galera Olmo, Ana Bastida Ramírez, Joaquín Lomba Maurandi, Tona Majó, Jacqueline I. McKinley, Kathleen McSweeney, Balázs Gusztáv Mende, Alessandra Modi, Gabriella Kulcsár, Viktória Kiss, András Czene, Róbert Patay, Anna Endrődi, Kitti Köhler, Tamás Hajdu, Tamás Szeniczey, János Dani, Zsolt Bernert, Maya Hoole, Olivia Cheronet, Denise Keating, Petr Velemínský, Miroslav Dobeš, Francesca Candilio, Fraser Brown, Raúl Flores Fernández, Ana-Mercedes Herrero-Corral, Sebastiano Tusa, Emiliano Carnieri, Luigi Lentini, Antonella Valenti, Alessandro Zanini, Clive Waddington, Germán Delibes, Elisa Guerra-Doce, Benjamin Neil, Marcus Brittain, Mike Luke, Richard Mortimer, Jocelyne Desideri, Marie Besse, Günter Brücken, Mirosław Furmanek, Agata Hałuszko, Maksym Mackiewicz, Artur Rapiński, Stephany Leach, Ignacio Soriano, Katina T. Lillios, João Luís Cardoso, Michael Parker Pearson, Piotr Włodarczak, T. Douglas Price, Pilar Prieto, Pierre-Jérôme Rey, Roberto Risch, Manuel A. Rojo Guerra, Aurore Schmitt, Joël Serralongue, Ana Maria Silva, Václav Smrčka, Luc Vergnaud, João Zilhão, David Caramelli, Thomas Higham, Mark G. Thomas, Douglas J. Kennett, Harry Fokkens, Volker Heyd, Alison Sheridan, Karl-Göran Sjögren, Philipp W. Stockhammer, Johannes Krause, Ron Pinhasi, Wolfgang Haak, Ian Barnes, Carles Lalueza-Fox, and David Reich. The beaker phenomenon and the genomic transformation of northwest europe. *Nature*, 555 (7695):190–196, February 2018. ISSN 1476-4687. doi: 10.1038/nature25738.
46. Tina Saupe, Francesco Montinaro, Cinzia Scaggion, Nicola Carrara, Toomas Kivisild, Eugenia D'Atanasio, Ruoyun Hui, Anu Solnik, Ophélie Lebrasseur, Greger Larson, Luca Alessandri, Ilenia Arienzo, Flavio De Angelis, Mario Federico Rolfo, Robin Skeates, Letizia Silvestri, Jessica Beckett, Sahra Talamo, Andrea Dolfini, Monica Miari, Mait Metspalu, Stefano Benazzi, Cristian Capelli, Luca Pagani, and Christiana L. Scheib. Ancient genomes reveal structural shifts after the arrival of steppe-related ancestry in the italian peninsula. *Current Biology*, 31(12):2576–2591.e12, June 2021. ISSN 0960-9822. doi: 10.1016/j.cub.2021.04.022.
47. Daniel M. Fernandes, Alissa Mittnik, Iñigo Olalde, Iosif Lazaridis, Olivia Cheronet, Nadin Rohland, Swapan Mallick, Rebecca Bernardos, Nasreen Broomandkhoshbacht, Jens Carlsson, Brendan J. Culleton, Matthew Ferry, Beatriz Gamarra, Martina Lari, Matthew Mah, Megan Michel, Alessandra Modi, Mario Novak, Jonas Oppenheimer, Kendra A. Sirak, Kristin Stewardson, Kirsten Mandl, Constanze Schattke, Kadir T. Özdoğan, Michaela Lucci, Gabriella Gasperetti, Francesca Candilio, Gianfranca Salis, Stefania Vai, Edgard Camarós, Carla Calò, Giulio Catalano, Marián Cueto, Vincenza Forgia, Marina Lozano, Elisabetta Marini, Margherita Micheletti, Roberto M. Miccichè, Maria R. Palombo, Damià Ramis, Vittoria Schimmenti, Pau Sureda, Luís Teira, Maria Teschler-Nicola, Douglas J. Kennett, Carles Lalueza-Fox, Nick Patterson, Luca Sineo, Alfredo Coppa, David Caramelli, Ron Pinhasi, and David Reich. The spread of steppe and iranian-related ancestry in the islands of the western mediterranean. *Nature Ecology & Evolution*, 4(3):334–345, February 2020. ISSN 2397-334X. doi: 10.1038/s41559-020-1102-0.
48. Joseph H. Marcus, Cosimo Posth, Harald Ringbauer, Luca Lai, Robin Skeates, Carlo Sidore, Jessica Beckett, Anja Furtwängler, Anna Olivieri, Charleston W. K. Chiang, Hussein Al-Asadi, Kushal Dey, Tyler A. Joseph, Chi-Chun Liu, Clio Der Sarkissian, Rita Radzevičiūtė, Megan Michel, Maria Giuseppina Gradoli, Patrizia Marongiu, Salvatore Rubino, Vittorio Mazzarello, Daniela Rovina, Alessandra La Fragola, Rita Maria Serra, Pasquale Bandiera, Raffaella Bianucci, Elisa Pompianu, Clizia Murgia, Michele Guirguis, Rosana Pla Orquin, Noreen Tuross, Peter van Dommelen, Wolfgang Haak, David Reich, David Schlessinger, Francesco Cucca, Johannes Krause, and John Novembre. Genetic history from the middle neolithic to present on the mediterranean island of sardinia. *Nature Communications*, 11 (1), February 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-14523-6.
49. Iñigo Olalde, Swapan Mallick, Nick Patterson, Nadin Rohland, Vanessa Villalba-Mouco, Marina Silva, Katharina Dulias, Ceiridwen J. Edwards, Francesca Gandini, Maria Pala, Pedro Soares, Manuel Ferrando-Bernal, Nicole Adamski, Nasreen Broomandkhoshbacht, Olivia Cheronet, Brendan J. Culleton, Daniel Fernandes, Ann Marie Lawson, Matthew Mah, Jonas Oppenheimer, Kristin Stewardson, Zhao Zhang, Juan Manuel Jiménez Arenas, Isidro Jorge Toro Moyano, Domingo C. Salazar-García, Pere Castanyer, Marta Santos, Joaquim Tremoleda, Marina Lozano, Pablo García Borja, Javier Fernández-Eraso, José Antonio Mujika-Alustiza, Cecilio Barroso, Francisco J. Bermúdez, Enrique Viguera Mínguez, Josep Burch, Neus Coromina, David Vivó, Artur Cebrià, Josep Maria Fullola, Oreto García-Puchol, Juan Ignacio Morales, F. Xavier Oms, Tona Majó, Josep Maria Vergès, Antònia Díaz-Carvajal, Imma Ollich-Castanyer, F. Javier López-Cachero, Ana Maria Silva, Carmen Alonso-Fernández, Germán Delibes de Castro, Javier Jiménez Echevarría, Adolfo Moreno-Márquez, Guillermo Pascual Berlanga, Pablo Ramos-García, José Ramos-Muñoz, Eduardo Vijande Vila, Gustau Aguilella Arzo, Ángel Esparza Arroyo, Katina T. Lillios, Jennifer Mack, Javier Velasco-Vázquez, Anna Waterman, Luis Benítez de Lugo Enrich, María Benito Sánchez, Bibiana Agustí, Ferran Codina, Gabriel de Prado, Almudena Estalrrich, Álvaro Fernández Flores, Clive Finlayson, Geraldine Finlayson, Stewart Finlayson, Francisco Giles-Guzmán, Antonio Rosas, Virginia Barciela González, Gabriel García Atiénzar, Mauro S. Hernández Pérez, Armando Llanos, Yolanda Carrión Marco, Isabel Collado Beneyto, David López-Serrano, Mario Sanz Tormo, António C. Valera, Concepción Blasco, Corina Liesau, Patricia Ríos, Joan Daura, María Jesús de Pedro Michó, Agustín A. Diez-Castillo, Raúl Flores Fernández, Joan Francès Farré, Rafael Garrido-Pena, Victor S. Gonçalves, Elisa Guerra-Doce, Ana Mercedes Herrero-Corral, Joaquim Juan-Cabanilles, Daniel López-Reyes, Sarah B. McClure, Marta Merino Pérez, Arturo Oliver Foix, Montserrat Sanz Borràs, Ana Catarina Sousa, Julio Manuel Vidal Encinas, Douglas J. Kennett, Martin B. Richards, Kurt Werner Alt, Wolfgang Haak, Ron Pinhasi, Carles Lalueza-Fox, and David Reich. The genomic history of the iberian peninsula over the past 8000 years. *Science*, 363(6432):1230–1234, March 2019. ISSN 1095-9203. doi: 10.1126/science.aav4040.
50. Rui Martiniano, Lara M. Cassidy, Ros Ó'Maoldúin, Russell McLaughlin, Nuno M. Silva, Licinio Manco, Daniel Fidalgo, Tania Pereira, Maria J. Coelho, Miguel Serra, Joachim Burger, Rui Parreira, Elena Moran, Antonio C. Valera, Eduardo Porfirio, Rui Boaventura, Ana M. Silva, and Daniel G. Bradley. The population genomics of archaeological transition in west iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLOS Genetics*, 13(7):e1006852, July 2017. ISSN 1553-7404. doi: 10.1371/journal.pgen.1006852.
51. Hannes Schroeder, Ashot Margaryan, Marzena Szmyt, Bertrand Theulot, Piotr Włodarczak, Simon Rasmussen, Shyam Gopalakrishnan, Anita Szczepanek, Tomasz Konopka, Theis Z. T. Jensen, Barbara Witkowska, Stanisław Wilk, Marcin M. Przybyła, Łukasz Pospieszny,

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Karl-Göran Sjögren, Zdzislaw Belka, Jesper Olsen, Kristian Kristiansen, Eske Willerslev, Karin M. Frei, Martin Sikora, Niels N. Johannsen, and Morten E. Allentoft. Unraveling ancestry, kinship, and violence in a late neolithic mass grave. *Proceedings of the National Academy of Sciences*, 116(22):10705–10710, May 2019. ISSN 1091-6490. doi: 10.1073/pnas.1820210116.

52. Anna Linderholm, Gülşah Merve Kılınç, Anita Szczepanek, Piotr Włodarczak, Paweł Jarosz, Zdzislaw Belka, Jolanta Dopieralska, Karolina Werens, Jacek Górski, Mirosław Mazurek, Monika Hozer, Małgorzata Rybicka, Mikołaj Ostrowski, Jolanta Bagińska, Wiesław Koman, Ricardo Rodríguez-Varela, Jan Storå, Anders Götherström, and Maja Krzewińska. Corded ware cultural complexity uncovered using genomic and isotopic analysis from south-eastern poland. *Scientific Reports*, 10(1), April 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-63138-w.
53. Luka Papac, Michal Ernée, Miroslav Dobeš, Michaela Langová, Adam B. Rohrlach, Franziska Aron, Gunnar U. Neumann, Maria A. Spyrou, Nadin Rohland, Petr Velemínský, Martin Kuna, Hana Brzobohatá, Brendan Culleton, David Daněček, Alžběta Danielisová, Miluše Dobisíková, Josef Hložek, Douglas J. Kennett, Jana Klementová, Michal Kostka, Petr Krištuf, Milan Kuchařík, Jana Kuljavceva Hlavová, Petr Limburský, Drahomíra Malyková, Lucia Mattiello, Monika Pecinovská, Katarína Petriščáková, Erika Průchová, Petra Stránská, Lubor Smejtek, Jaroslav Špaček, Radka Šumberová, Ondřej Švejcar, Martin Trefný, Miloš Vávra, Jan Kolář, Volker Heyd, Johannes Krause, Ron Pinhasi, David Reich, Stephan Schiffels, and Wolfgang Haak. Dynamic changes in genomic and social structures in third millennium bce central europe. *Science Advances*, 7(35), August 2021. ISSN 2375-2548. doi: 10.1126/sciadv.abi6941.
54. Samantha Brunel, E. Andrew Bennett, Laurent Cardin, Damien Garraud, Hélène Barrand Emam, Alexandre Beylier, Bruno Boulestin, Fanny Chenal, Elsa Ciesielski, Fabien Convertini, Bernard Dedet, Stéphanie Desbrosse-Degobertiere, Sophie Desenne, Jerôme Dubouloz, Henri Duday, Gilles Escalon, Véronique Fabre, Eric Gailledrat, Muriel Gandelin, Yves Gleize, Sébastien Goepfert, Jean Guilaine, Lamys Hachem, Michael Ilett, François Lambach, Florent Maziere, Bertrand Perrin, Suzanne Plouin, Estelle Pinard, Ivan Praud, Isabelle Richard, Vincent Riquier, Réjane Roure, Benoit Sendra, Corinne Thevenet, Sandrine Thiol, Elisabeth Vauquelin, Luc Vergnaud, Thierry Grange, Eva-Maria Geigl, and Melanie Pruvost. Ancient genomes from present-day france unveil 7, 000 years of its demographic history. *Proceedings of the National Academy of Sciences*, 117(23):12791–12798, May 2020. ISSN 1091-6490. doi: 10.1073/pnas.1918034117.
55. Anja Furtwängler, A. B. Rohrlach, Thiseas C. Lamnidis, Luka Papac, Gunnar U. Neumann, Inga Siebke, Ella Reiter, Noah Steuri, Jürgen Hald, Anthony Denaire, Bernadette Schnitzler, Joachim Wahl, Marianne Ramstein, Verena J. Schuenemann, Philipp W. Stockhammer, Albert Hafner, Sandra Lösch, Wolfgang Haak, Stephan Schiffels, and Johannes Krause. Ancient genomes reveal social and genetic structure of late neolithic switzerland. *Nature Communications*, 11(1), April 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15560-x.
56. Andaine Seguin-Orlando, Richard Donat, Clio Der Sarkissian, John Southon, Catherine Thèves, Claire Manen, Yaramila Tchérémissinoff, Eric Crubézy, Beth Shapiro, Jean-François Deleuze, Love Dalén, Jean Guilaine, and Ludovic Orlando. Heterogeneous hunter-gatherer and steppe-related ancestries in late neolithic and bell beaker genomes from present-day france. *Current Biology*, 31(5):1072–1083.e10, March 2021. ISSN 0960-9822. doi: 10.1016/j.cub.2020.12.015.
57. Helena Malmström, Torsten Günther, Emma M. Svensson, Anna Juras, Magdalena Fraser, Arielle R. Munters, Łukasz Pospieszny, Mari Tõrv, Jonathan Lindström, Anders Götherström, Jan Storå, and Mattias Jakobsson. The genomic ancestry of the scandinavian battle axe culture people and their relation to the broader corded ware horizon. *Proceedings of the Royal Society B: Biological Sciences*, 286(1912):20191528, October 2019. ISSN 1471-2954. doi: 10.1098/rspb.2019.1528.
58. Anne Friis-Holm Egfjord, Ashot Margaryan, Anders Fischer, Karl-Göran Sjögren, T. Douglas Price, Niels N. Johannsen, Poul Otto Nielsen, Lasse Sørensen, Eske Willerslev, Rune Iversen, Martin Sikora, Kristian Kristiansen, and Morten E. Allentoft. Genomic steppe ancestry in skeletons from the neolithic single grave culture in denmark. *PLOS ONE*, 16(1): e0244872, January 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0244872.
59. Sandra Penske, Adam B. Rohrlach, Ainash Childebayeva, Guido Gnecchi-Ruscone, Clemens Schmid, Maria A. Spyrou, Gunnar U. Neumann, Nadezhda Atanassova, Katrin Beutler, Kamen Boyadzhiev, Yavor Boyadzhiev, Igor Bruyako, Alexander Chohadzhiev, Blagoje Govedarica, Mehmet Karaucak, Raiko Krauss, Maleen Leppek, Igor Manzura, Karen Privat, Shawn Ross, Vladimir Slavchev, Adéla Sobotkova, Meda Toderaş, Todor Valchev, Harald Ringbauer, Philipp W. Stockhammer, Svend Hansen, Johannes Krause, and Wolfgang Haak. Early contact between late farming and pastoralist societies in southeastern europe. *Nature*, 620(7973):358–365, July 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06334-8.
60. Iain Mathieson, Songül Alpaslan-Roodenberg, Cosimo Posth, Anna Szécsényi-Nagy, Nadin Rohland, Swapan Mallick, Iñigo Olalde, Nasreen Broomandkhoshbacht, Francesca Candilio, Olivia Cheronet, Daniel Fernandes, Matthew Ferry, Beatriz Gamarra, Gloria González Fortes, Wolfgang Haak, Eadaoin Harney, Eppie Jones, Denise Keating, Ben Krause-Kyora, Isil Kucukkalipci, Megan Michel, Alissa Mittnik, Kathrin Nägele, Mario Novak, Jonas Oppenheimer, Nick Patterson, Saskia Pfrengle, Kendra Sirak, Kristin Stewardson, Stefania Vai, Stefan Alexandrov, Kurt W. Alt, Radian Andreescu, Dragana Antonović, Abigail Ash, Nadezhda Atanassova, Krum Bacvarov, Mende Balázs Gusztáv, Hervé Bocherens, Michael Bolus, Adina Boroneanţ, Yavor Boyadzhiev, Alicja Budnik, Josip Burmaz, Stefan Chohadzhiev, Nicholas J. Conard, Richard Cottiaux, Maja Čuka, Christophe Cupillard, Dorothée G. Drucker, Nedko Elenski, Michael Francken, Borislava Galabova, Georgi Ganetsovski, Bernard Gély, Tamás Hajdu, Veneta Handzhyiska, Katerina Harvati, Thomas Higham, Stanislav Iliev, Ivor Janković, Ivor Karavanić, Douglas J. Kennett, Darko Komšo, Alexandra Kozak, Damian Labuda, Martina Lari, Catalin Lazar, Maleen Leppek, Krassimir Leshtakov, Domenico Lo Vetro, Dženi Los, Ivaylo Lozanov, Maria Malina, Fabio Martini, Kath McSweeney, Harald Meller, Marko Menđušić, Pavel Mirea, Vyacheslav Moiseyev, Vanya Petrova, T. Douglas Price, Angela Simalcsik, Luca Sineo, Mario Šlaus, Vladimir Slavchev, Petar Stanev, Andrej Starović, Tamás Szeniczey, Sahra Talamo, Maria Teschler-Nicola, Corinne Thevenet, Ivan Valchev, Frédérique Valentin, Sergey Vasilyev, Fanica Veljanovska, Svetlana Venelinova, Elizaveta Veselovskaya, Bence Viola, Cristian Virag, Joško Zaninović, Steve Zäuner, Philipp W. Stockhammer, Giulio Catalano, Raiko Krauß, David Caramelli, Gunita Zariņa, Bisserka Gaydarska, Malcolm Lillie, Alexey G. Nikitin, Inna Potekhina, Anastasia Papathanasiou, Dušan Borić, Clive Bonsall, Johannes Krause, Ron Pinhasi, and David Reich. The genomic history of southeastern europe. *Nature*, 555(7695):197–203, February 2018. ISSN 1476-4687. doi: 10.1038/nature25778.
61. Iosif Lazaridis, Songül Alpaslan-Roodenberg, Ayşe Acar, Ayşen Açıkkol, Anagnostis Agelarakis, Levon Aghikyan, Uğur Akyüz, Desislava Andreeva, Gojko Andrijašević, Dragana Antonović, Ian Armit, Alper Atmaca, Pavel Avetisyan, Ahmet İhsan Aytek, Krum Bacvarov, Ruben Badalyan, Stefan Bakardzhiev, Jacqueline Balen, Lorenc Bejko, Rebecca Bernardos, Andreas Bertsatos, Hanifi Biber, Ahmet Bilir, Mario Bodružić, Michelle Bonogofsky, Clive Bonsall, Dušan Borić, Nikola Borovinić, Guillermo Bravo Morante, Katharina Buttinger, Kim Callan, Francesca Candilio, Mario Carić, Olivia Cheronet, Stefan Chohadzhiev, Maria-Eleni Chovalopoulou, Stella Chryssoulaki, Ion Ciobanu, Natalija Čondić, Mihai Constantinescu, Emanuela Cristiani, Brendan J. Culleton, Elizabeth Curtis, Jack Davis, Ruben Davtyan, Tatiana I. Demcenco, Valentin Dergachev, Zafer Derin, Sylvia Deskaj, Seda Devejyan, Vojislav Djordjević, Kellie Sara Duffett Carlson, Laurie R. Eccles, Nedko Elenski, Atilla Engin, Nihat Erdoğan, Sabiha Erir-Pazarcı, Daniel M. Fernandes, Matthew Ferry, Suzanne Freilich, Alin Frînculeasa, Michael L. Galaty, Beatriz Gamarra, Boris Gasparyan, Bisserka Gaydarska, Elif Genç, Timur Gültekin, Serkan Gündüz, Tamás Hajdu, Volker Heyd, Suren Hobosyan, Nelli Hovhannisyan, Iliya Iliev, Lora Iliev, Stanislav Iliev, İlkay İvgin, Ivor Janković, Lence Jovanova, Panagiotis Karkanas, Berna Kavaz-Kındığılı, Esra Hilal Kaya, Denise Keating, Douglas J. Kennett, Seda Deniz Kesici, Anahit Khudaverdyan, Krisztián Kiss, Sinan Kılıç, Paul Klostermann, Sinem Kostak Boca Negra Valdes, Saša Kovačević, Marta Krenz-Niedbała, Maja Krznarić Škrivanko, Rovena Kurti, Pasko Kuzman, Ann Marie Lawson, Catalin Lazar, Krassimir Leshtakov, Thomas E. Levy, Ioannis Liritzis, Kirsi O. Lorentz, Sylwia Łukasik, Matthew Mah, Swapan Mallick, Kirsten Mandl, Kristine Martirosyan-Olshansky, Roger Matthews, Wendy Matthews, Kathleen McSweeney, Varduhi Melikyan, Adam Micco, Megan Michel, Lidija Milašinović, Alissa Mittnik, Janet M. Monge, Georgi Nekhrizov, Rebecca Nicholls, Alexey G. Nikitin, Vassil Nikolov, Mario Novak, Iñigo Olalde, Jonas Oppenheimer, Anna Osterholtz, Celal Özdemir, Kadir Toykan Özdoğan, Nurettin Öztürk, Nikos Papadimitriou, Niki Papakonstantinou, Anastasia Papathanasiou, Lujana Paraman, Evgeny G. Paskary, Nick Patterson, Ilian Petrakiev, Levon Petrosyan, Vanya Petrova, Anna Philippa-Touchais, Ashot Piliposyan, Nada Pocuca Kuzman, Hrvoje Potrebica, Bianca Preda-Bălănică, Zrinka Premužić, T. Douglas Price, Lijun Qiu, Siniša Radović, Kamal Raeuf Aziz, Petra Rajić Šikanjić, Kamal Rasheed Raheem, Sergei Razumov, Amy Richardson, Jacob Roodenberg, Rudenc Ruka, Victoria Russeva, Mustafa Şahin, Ayşegül Şarbak, Emre Savaş, Constanze Schattke, Lynne Schepartz, Tayfun Selçuk, Ayla Sevim-Erol, Michel Shamoon-Pour, Henry M. Shephard, Athanasios Sideris, Angela Simalcsik, Hakob Simonyan, Vitalij Sinika, Kendra Sirak, Ghenadie Sirbu, Mario Šlaus, Andrei Soficaru, Bilal Söğüt, Arkadiusz Sołtysiak, undefinedilem Sönmez-Sözer, Maria Stathi, Martin Steskal, Kristin Stewardson, Sharon Stocker, Fadime Suata-Alpaslan, Alexander Suvorov, Anna Szécsényi-Nagy, Tamás Szeniczey, Nikolai Telnov, Strahil Temov, Nadezhda Todorova, Ulsi Tota, Gilles Touchais, Sevi Triantaphyllou, Atila Türker, Marina Ugarković, Todor Valchev, Fanica Veljanovska, Zlatko Videvski, Cristian Virag, Anna Wagner, Sam Walsh, Piotr Włodarczak, J. Noah Workman, Aram Yardumian, Evgenii Yarovoy, Alper Yener Yavuz, Hakan Yılmaz, Fatma Zalzala, Anna Zettl, Zhao Zhang, Rafet undefinedavuşoğlu, Nadin Rohland, Ron Pinhasi, and David Reich. The genetic history of the southern arc: A bridge between west asia and europe. *Science*, 377(6609), August 2022. ISSN 1095-9203. doi: 10.1126/science.abm4247.
62. Suzanne Freilich, Harald Ringbauer, Dženi Los, Mario Novak, Dinko Tresić Pavičić, Stephan Schiffels, and Ron Pinhasi. Reconstructing genetic histories and social organisation in neolithic and bronze age croatia. *Scientific Reports*, 11(1), August 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-94932-9.
63. Dániel Gerber, Bea Szeifert, Orsolya Székely, Balázs Egyed, Balázs Gyuris, Julia I Giblin, Anikó Horváth, Kitti Köhler, Gabriella Kulcsár, Ágnes Kustár, István Major, Mihály Molnár, László Palcsu, Vajk Szeverényi, Szilvia Fábián, Balázs Gusztáv Mende, Mária Bondár, Eszter Ari, Viktória Kiss, and Anna Szécsényi-Nagy. Interdisciplinary analyses of bronze age communities from western hungary reveal complex population histories. *Molecular Biology and Evolution*, 40(9), August 2023. ISSN 1537-1719. doi: 10.1093/molbev/msad182.
64. Eirini Skourtanioti, Yilmaz S. Erdal, Marcella Frangipane, Francesca Balossi Restelli, K. Aslıhan Yener, Frances Pinnock, Paolo Matthiae, Rana Özbal, Ulf-Dietrich Schoop, Farhad Guliyev, Tufan Akhundov, Bertille Lyonnet, Emily L. Hammer, Selin E. Nugent, Marta Burri, Gunnar U. Neumann, Sandra Penske, Tara Ingman, Murat Akar, Rula Shafiq, Giulio Palumbi, Stefanie Eisenmann, Marta D'Andrea, Adam B. Rohrlach, Christina Warinner, Choongwon Jeong, Philipp W. Stockhammer, Wolfgang Haak, and Johannes Krause. Genomic history of neolithic to bronze age anatolia, northern levant, and southern caucasus. *Cell*, 181(5):1158–1175.e28, May 2020. ISSN 0092-8674. doi: 10.1016/j.cell.2020.04.044.
65. Dilek Koptekin, Eren Yüncü, Ricardo Rodríguez-Varela, N. Ezgi Altınışık, Nikolaos Psonis, Natalia Kashuba, Sevgi Yorulmaz, Robert George, Duygu Deniz Kazancı, Damla Kaptan, Kanat Gürün, Kıvılcım Başak Vural, Hasan Can Gemici, Despoina Vassou, Evangelia Daskalaki, Cansu Karamurat, Vendela K. Lagerholm, Ömür Dilek Erdal, Emrah Kırdök, Aurelio Marangoni, Andreas Schachner, Handan Üstündağ, Ramaz Shengelia, Liana Bitadze, Mikheil Elashvili, Eleni Stravopodi, Mihriban Özbaşaran, Güneş Duru, Argyro Nafplioti, C. Brian Rose, Tuğba Gencer, Gareth Darbyshire, Alexander Gavashelishvili, Konstantine Pitskhelauri, Özlem undefinedevik, Osman Vuruşkan, Nina Kyparissi-Apostolika, Ali Metin Büyükkarakaya, Umay Oğuzhanoğlu, Sevinç Günel, Eugenia Tabakaki, Akper Aliev, Anar Ibrahimov, Vaqif Shadlinski, Adamantios Sampson, Gülşah Merve Kılınç, undefinediğdem Atakuman, Alexandros Stamatakis, Nikos Poulakakis, Yılmaz Selim Erdal, Pavlos Pavlidis, Jan Storå, Füsun Özer, Anders Götherström, and Mehmet Somel. Spatial and temporal heterogeneity in human mobility patterns in holocene southwest asia and the east mediterranean. *Current Biology*, 33(1):41–57.e15, January 2023. ISSN 0960-9822. doi: 10.1016/j.cub.2022.11.034.
66. Lily Agranat-Tamir, Shamam Waldman, Mario A.S. Martin, David Gokhman, Nadav Mishol, Tzilla Eshel, Olivia Cheronet, Nadin Rohland, Swapan Mallick, Nicole Adamski, Ann Marie Lawson, Matthew Mah, Megan Michel, Jonas Oppenheimer, Kristin Stewardson, Francesca Candilio, Denise Keating, Beatriz Gamarra, Shay Tzur, Mario Novak, Rachel Kalisher,

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Shlomit Bechar, Vered Eshed, Douglas J. Kennett, Marina Faerman, Naama Yahalom-Mack, Janet M. Monge, Yehuda Govrin, Yigal Erel, Benjamin Yakir, Ron Pinhasi, Shai Carmi, Israel Finkelstein, Liran Carmel, and David Reich. The genomic history of the bronze age southern levant. *Cell*, 181(5):1146–1157.e11, May 2020. ISSN 0092-8674. doi: 10.1016/j.cell.2020.04.024.

67. Mathias Currat, Manuel Ruedi, Rémy J Petit, and Laurent Excoffier. The hidden side of invasions: massive introgression by local genes. *Evolution*, 62(8):1908–1920, 2008.
68. Lei Huang, Thiseas C Lamnidis, and Stephan Schiffels. High resolution analysis of population structure using rare variants. *bioRxiv*, page 2025.07.18.665597, 23 July 2025. doi: 10.1101/2025.07.18.665597.
69. David H Alexander, John Novembre, and Kenneth Lange. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, 19(9):1655–1664, July 2009.
70. Alkes L. Price, Michael E. Weale, Nick Patterson, Simon R. Myers, Anna C. Need, Kevin V. Shianna, Dongliang Ge, Jerome I. Rotter, Esther Torres, Kent D. Taylor, David B. Goldstein, and David Reich. Long-range ld can confound genome scans in admixed populations. *The American Journal of Human Genetics*, 83(1):132–135, July 2008. ISSN 0002-9297. doi: 10.1016/j.ajhg.2008.06.005.
71. Carl A Anderson, Fredrik H Pettersson, Geraldine M Clarke, Lon R Cardon, Andrew P Morris, and Krina T Zondervan. Data quality control in genetic case-control association studies. *Nature Protocols*, 5(9):1564–1573, August 2010. ISSN 1750-2799. doi: 10.1038/nprot.2010.116.

Isacchini *et al.* | Migration networks bioRχiv | 13

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Information

Inferring Migration Networks with Time-Lagged F_2 Statistics

Giulio Isacchini,Takashi Okada, Clemens Schmid, Divyaratan Popli, Harald Ringbauer, Benjamin M Peter, Stephan Schiffels, and Oskar Hallatschek

Wright-Fisher Simulations

Our Wright-Fisher simulations between N_D metapopulations are defined by the following steps:

- **Initialization:** once at beginning. Initially, allele frequency distributions are drawn via

$$f_0 = 10^a \text{ with } a \sim U(l, -1) \qquad \textbf{(S.1)}$$

  where $U(l,-1)$ represents a uniform distribution in the interval $[l,-1]$ with l a parameter. We generate N_{traj} distinct lineages for each population i we sample

$$f_0^i \sim f_0 + \sigma b \sqrt{f_0} \text{ with } b \sim N(0,1) \qquad \textbf{(S.2)}$$

  where σ is another parameter that modulates the initial genetic distance between the metapopulations.

- **Generation:** for T times. At each new generation t, we first determine the updated allele frequencies by calculating

$$\hat{f}_t^i = \sum_j A_{ij} f_{t-1}^j \qquad \textbf{(S.3)}$$

  where the matrix $\mathbf{A}_{ij}$ dictates how individuals are redistributed among metapopulations and f_t values are constrained in the $[0,1]$ interval. We then draw N_{pop} samples via a binomial distribution

$$f_t^i \sim \text{Bin}(N_{\text{pop}}, \hat{f}_t^i)/N_{\text{pop}} \qquad \textbf{(S.4)}$$

  for each metapopulations and lineage. Observations are performed via an additional binomial sampling $Bin(N_s, f_t^i)$ at each generation, which depends on the parameter N_s.

Benchmark of inference method on simulated data

This section aims to evaluate and compare the inference accuracy of our F2-based estimator (Eq.8 in the main text) with that of a straightforward estimator derived from least squares regression (Eq.4 in the main text).
We evaluate the effectiveness of our methods using Wright-Fisher simulations across several metapopulations. Our objective is to investigate how various simulation parameters affect the performance of the methods.

Key Parameters for Comparative Analysis To evaluate the complexity of the task, the following parameters are varied in the benchmark:

- l defines the range of initial frequency magnitudes; at decreasing frequencies, the sampling noise increases and the complexity of the inference as well.
- σ influences the initial genetic distance among metapopulations; larger genetic distances lead to easier inference of $\mathbf{A}_{ij}$.
- N_s determines the extent of sampling from the metapopulation, which is related to the level of observational noise; the higher the observational noise, the harder the inference.
- T represents the entire duration of the mixing process; a longer observation period provides more extractable information.
- N_D refers to the number of metapopulations; Since a squared matrix is being inferred, the number of parameters increases quadratically with N_D.

Evaluation The default parameters of the simulation are: $\sigma = 0.1$, $T = 5$, $l = -1$, $N_D = 5$, $N_{\text{traj}} = 5000$, $N_{\text{pop}} = 1000$ and $N_S = 100$. We compute the mean squared error between the ground truth values $\mathbf{A}_T$ and the inferred values from the least square method $\mathbf{A}_{LS}$ and the F2-based method $\mathbf{A}_{F2}$ on multiple independent simulations at varying parameters. The results of the analysis are depicted in Fig. S3 . The $F2$ method consistently outperforms the naive LS method across all parameter regimes.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Zooming into the third millennium BC.

In this section, we demonstrate the ability of our method to integrate information from more time points to infer migratory events across extended periods of time. We focus on the third millennium BCE, a period where many studies have uncovered substantial migration of populations from the steppe regions north and east of the Black sea into both Western and Eastern Europe (8, 9), as also reflected in our extensive analysis of Fig. 4.

We conducted an analysis of samples from predefined geographic locations dating back to -3300 and -1800 BCE, employing time intervals of 300-year bins. These samples consist of a subset of the samples analyzed for Fig. 4, where we exclude Iberia, Italy, and Scandinavia, as their total average input genetic flux within that time period is only $\sim 7\%$ for those populations. We initially quantify gene flow between each pair of consecutive time bins (Fig. S8A). Similarly to the analysis in the previous section, we observe here that the two Central European metapopulations experience significant gene flow from the Western Pontic Steppe in the first half of the observed period. The Britain and Ireland metapopulation experiences substantial gene flow from the two Central European populations in the second half of the observed period. The Western Pontic Steppe primarily acts as a source of gene flow.

We seek then to integrate information across all time bins and infer a coarse-grained migration matrix that summarizes the average gene flow over the period of interest. This can be achieved by adjusting the minimization equation in 10, averaging over all time points, and inferring a single migration matrix for all times. We additionally implement a simple constraint on neighbouring edges during inference in order to avoid serial passaging (A->B->C) mimicking long-range direct connections (A->C), see Fig. S8B.

In Fig. S8C, we present the inferred migration matrix $\mathbf{A}_{ij}$ resulting from this procedure, with populations ordered from West to East. We observe that the values in the upper triangular part are significantly higher than those in the lower triangular part, emphasizing a prevalent east-to-west migration trend. Moreover, the strongest elements are typically on the diagonal, consistent with the idea that the main source of genetic composition for a population in the next time step is the population itself at the previous time. The exceptions are Central Europe West and the British Isles, both of which have the highest migration coefficients from Eastern Central Europe and Western Central Europe respectively.

As a robustness test, if we further merge the two Central European populations into a single group and repeat the analysis, we still observe a pronounced east-to-west genetic flux, see Fig. S9.

In Fig. S8D we depict the strongest off-diagonal elements of the matrix (thresholded at 0.1 of the average value) to offer a clearer visual representation of these population interactions on the map. This representation provides enhanced clarity regarding the aforementioned points. In Fig. S8E-F we present a visualization of the matrix's spectral decomposition.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

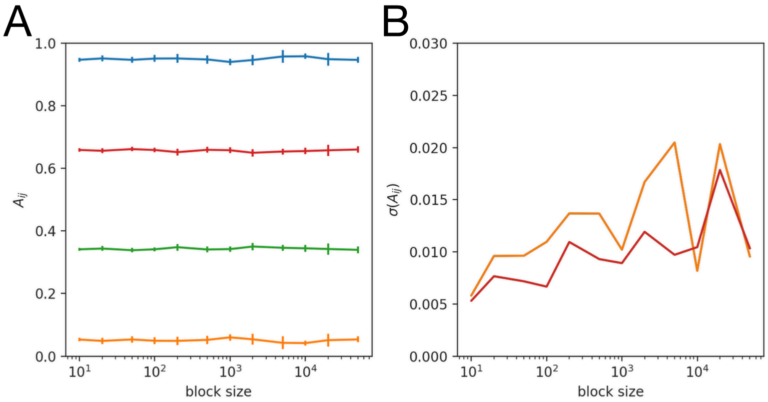


Fig. S1. **(A)** Mean and standard deviation of migration matrix elements computed via resampling of genetic blocks of increasing size, each color identifies a different element of $\mathbf{A}$. The evaluation is done on the same samples of Fig. 3 at the 4th time bin. **(B)** Standard deviation in the estimation of $\mathbf{A}$ as a function of block size. We find a limited dependency on block size.

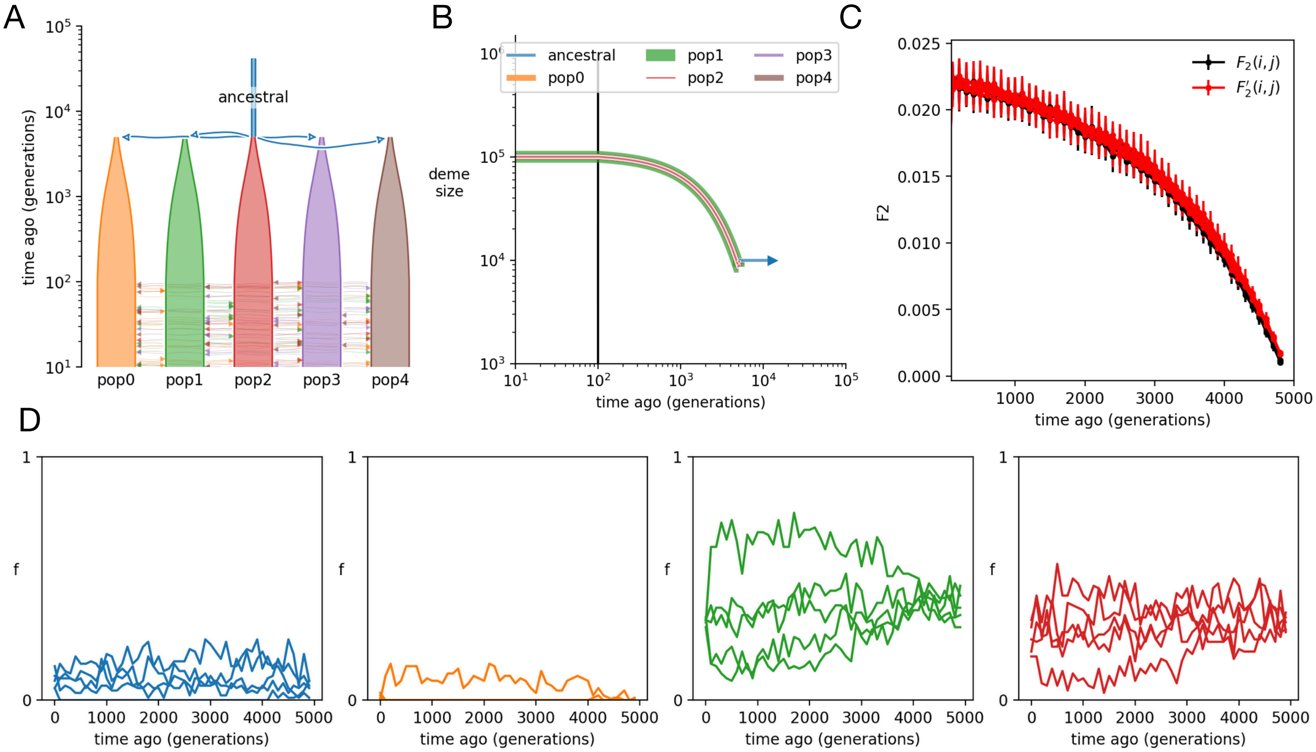


Fig. S2. **Establishment of genetic diversity across populations in the simulations.** **(A-B)** We simulate a similar scenario to that of Fig. 2 but sample across the whole period of isolation and expansion of the populations. **(C)** During this period the genetic divergence between populations monotonically increases. **(D)** Examples of lineage trajectories over the sampled period.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

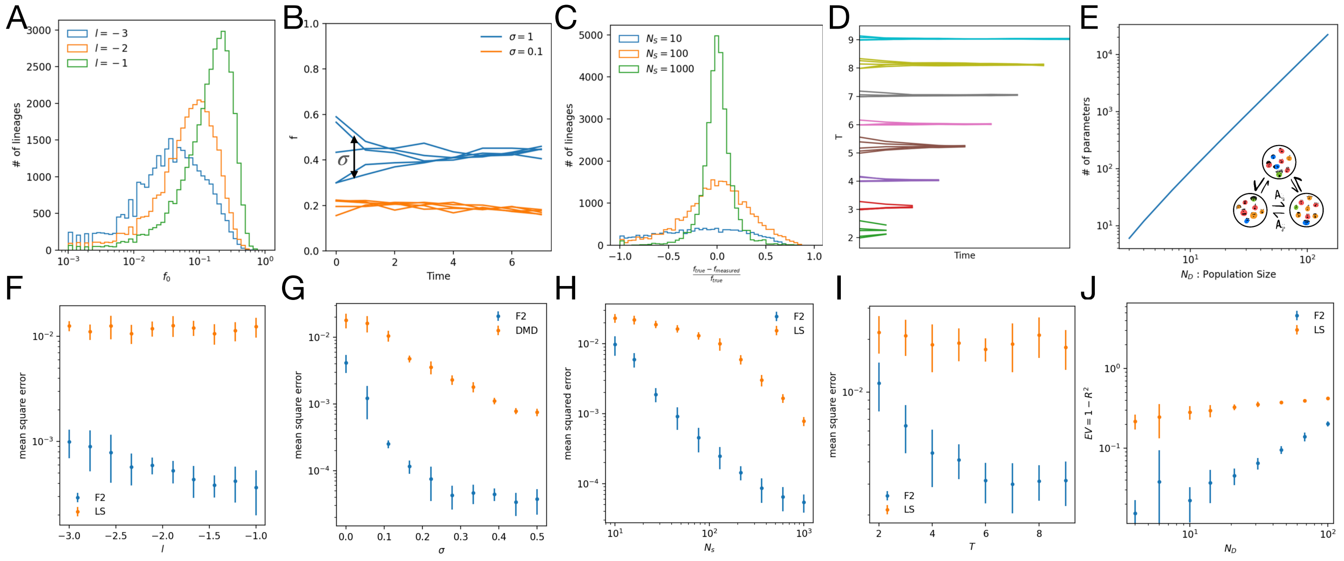


Fig. S3. Benchmark on simulated data of our F2-based method and least squares estimation (LS) We perform Wright Fisher simulations of multiple metapopulations where at each generation $\mathbf{A}_{ij}$ is used to reassign individuals between populations, see SI text for more details. **(A)** Distribution of initial allele frequencies f_0^i for different values of l and **(B)** depiction of lineage frequency trajectories as a function of the variance in the initial allele frequencies across demes σ. **(C)** Distribution of error induced by sampling noise as a function of N_S and **(D)** examples of different lineage trajectories for increasing T. **(E)** Given N_D populations, $(N_D - 1)^2$ parameters need to be inferred. **(F)** As expected, for bigger values of l, which induce a larger fraction of more common alleles, the inference is more accurate. **(G)** Our inference relies on detecting changes in allele frequencies through time. Thus, if σ is bigger, these differences are more significant and the inference is easier. **(H)** The more individuals are sampled at each generation, i.e. lower observational noise, the better the accuracy of the inference. **(I)**The method can integrate information across multiple time steps. We find an increase in accuracy as a function of the number of time steps until saturation. **(J)** As the number of parameters to be inferred increases, the accuracy is degraded. In this last benchmark we introduce increasingly high sparsity in the matrix when increasing the number of metapopulations to have a reasonable comparison across different parameter regimes. We estimate the error via the fraction of unexplained variance of non-zero matrix elements $EV = 1 - R^2$.

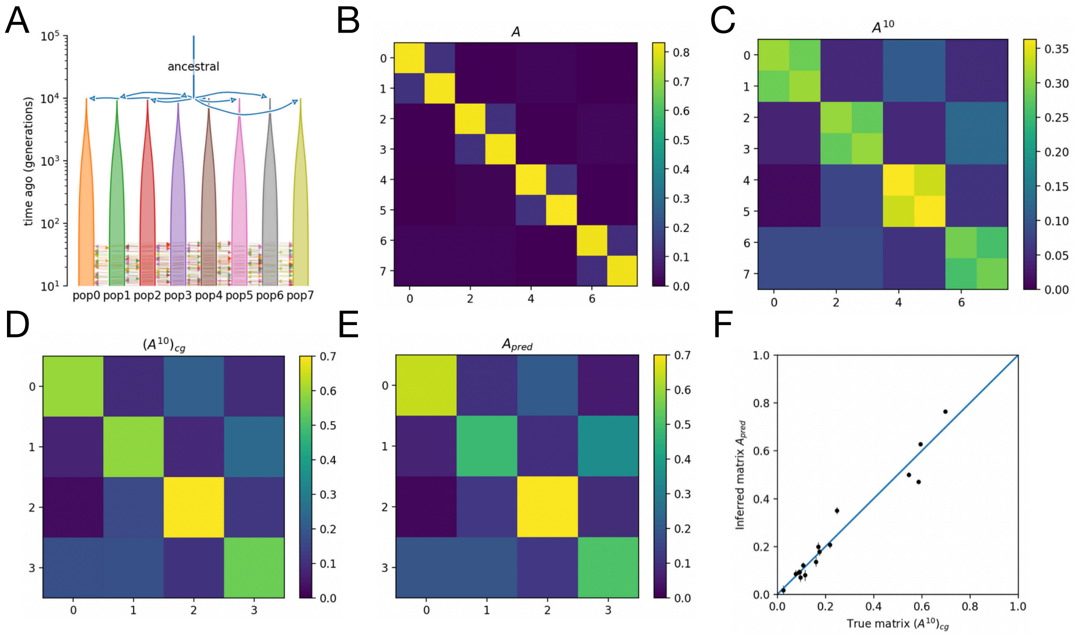


Fig. S4. Stability of inference when the meta-population has a substructure. **(A)** We repeat the simulation following the same parameter of Fig. 2 with 8 populations that are organized in pairs of more strongly connected demes. **(B)** Migration matrix used for the simulation **(C)** Since we resolve the dynamics every 10 generations, the time traces follow $\mathbf{A}^{10}$. **(D)** Coarse graining of the matrix is performed by: (I) a weighted average over the rows where the weights are proportional to the lineage steady-state distribution; (II) the sum of the elements of the resulting vector. **(E)** Inferred coarse grained matrix **(F)** Scatter plot of matrix elements of the true and inferred coarse grained matrices.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

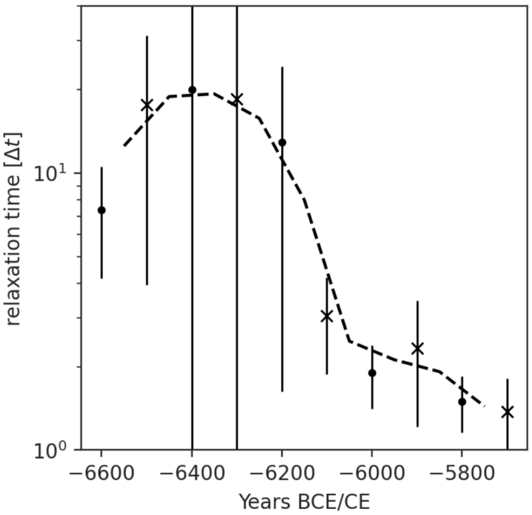


Fig. S5. Spectral decomposition and relaxation times. Estimation of relaxation time $t = (1-\lambda_2)^{-1}$ by eigenvalue decomposition λ_2 of the inferred migration matrices of Fig. 3. For each time slice, eigenvalues λ_i were computed via eigendecomposition of the migration matrix A. The dominant eigenvalue $\lambda_1 = 1$ reflects conservation of probability mass, while the subdominant eigenvalue λ_2 determines the asymptotic rate of convergence toward the stationary distribution. Error bars represent variability across bootstrap realizations of the inferred migration matrices.

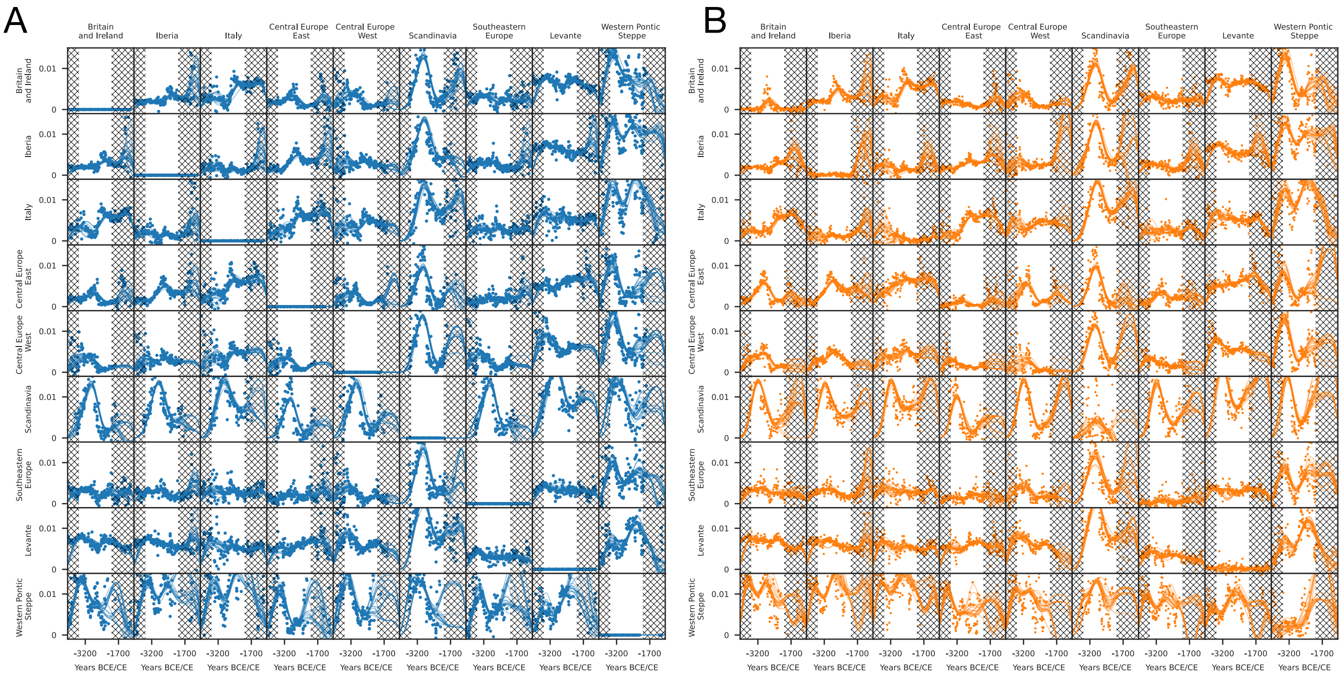


Fig. S6. Large Scale Analysis with Interpolation (A) Matrix of all pairwise comparisons of F statistics across time with interpolation lines. **(B)** Matrix of all pairwise comparisons of time-lagged F' statistics across time with interpolation lines. Crossed area refers to a region with lower confidence in our model. See Fig. 4 and Methods for more details.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

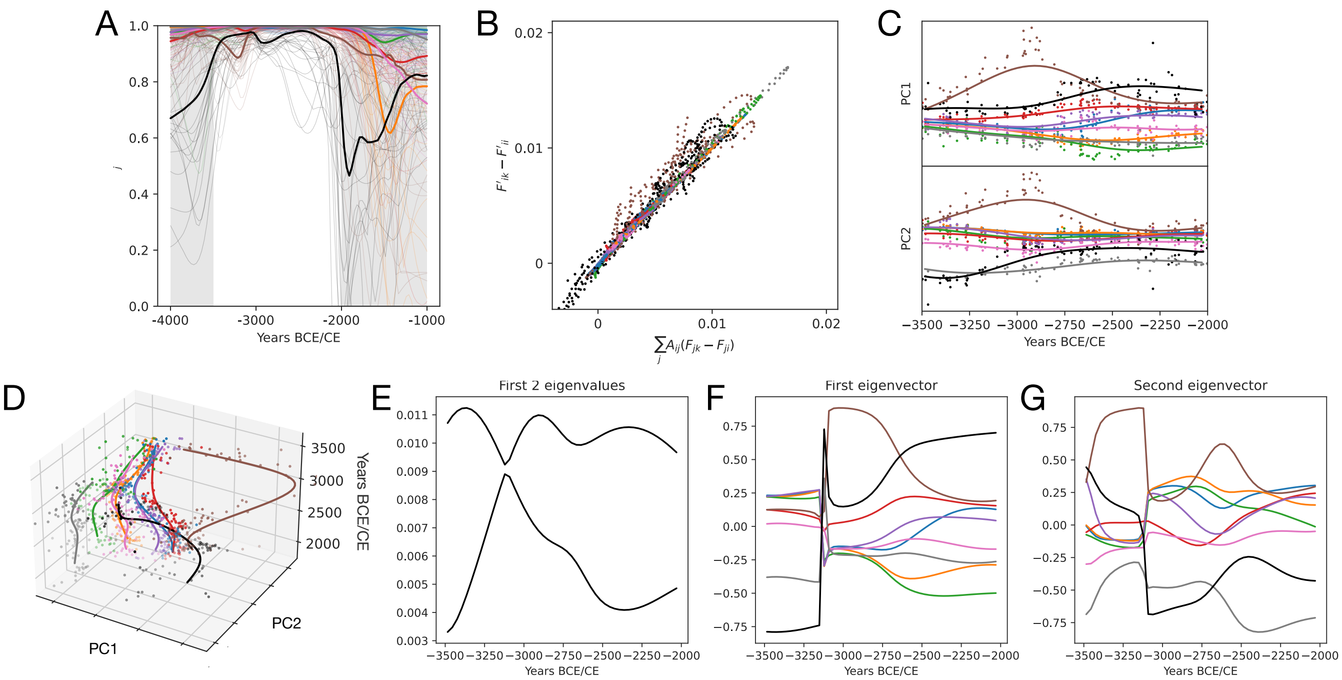
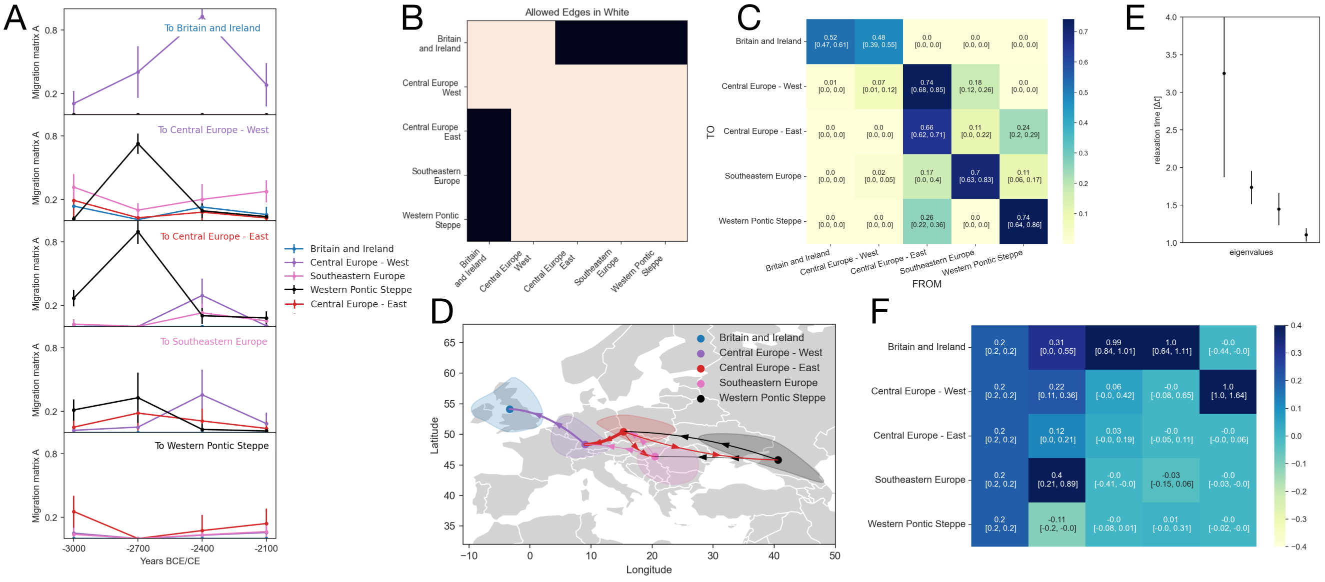


Fig. S7. Goodness-of-fit and Lower Dimensional Representation **(A)** Pearson correlation of left and right-hand side of Eq.2 for each row of $\mathbf{A}$ to check the consistency of the model. **(B)** Plot of left and right-hand side of eq.2 using the inferred migration matrix $\mathbf{A}$ within the period of high confidence (3500 - 2000 BCE). Western Pontic Steppe (black points) and Scandinavia (brown points) are the populations with lower confidence and higher errors. **(C-D)** Lower dimensional mapping of the F and F' matrices across time with average interpolation lines. The mapping is defined by the average first 2 eigenvectors of the centered F matrix across time $M = -0.5CF_2C$, see Methods for additional details. **(E)** First 2 eigenvalues of M across time. **(F)** Elements of the first eigenvector of M across time. **(G)** Elements of the second eigenvector of M across time.



TO \ FROM	Britain and Ireland	Central Europe - West	Central Europe - East	Southeastern Europe	Western Pontic Steppe
Britain and Ireland	0.52 [0.47, 0.61]	0.48 [0.39, 0.55]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
Central Europe - West	0.01 [0.0, 0.0]	0.07 [0.01, 0.12]	0.74 [0.68, 0.85]	0.18 [0.12, 0.26]	0.0 [0.0, 0.0]
Central Europe - East	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.66 [0.62, 0.71]	0.11 [0.0, 0.22]	0.24 [0.2, 0.29]
Southeastern Europe	0.0 [0.0, 0.0]	0.02 [0.0, 0.05]	0.17 [0.0, 0.4]	0.7 [0.63, 0.83]	0.11 [0.06, 0.17]
Western Pontic Steppe	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.26 [0.22, 0.36]	0.0 [0.0, 0.0]	0.74 [0.64, 0.86]

Britain and Ireland	0.2 [0.2, 0.2]	0.31 [0.0, 0.55]	0.99 [0.84, 1.01]	1.0 [0.64, 1.11]	-0.0 [-0.44, -0.0]
Central Europe - West	0.2 [0.2, 0.2]	0.22 [0.11, 0.36]	0.06 [-0.0, 0.42]	-0.0 [-0.08, 0.65]	1.0 [1.0, 1.64]
Central Europe - East	0.2 [0.2, 0.2]	0.12 [0.0, 0.21]	0.03 [-0.0, 0.19]	-0.0 [-0.05, 0.11]	-0.0 [-0.0, 0.06]
Southeastern Europe	0.2 [0.2, 0.2]	0.4 [0.21, 0.89]	-0.0 [-0.41, -0.0]	-0.03 [-0.15, 0.06]	-0.0 [-0.03, -0.0]
Western Pontic Steppe	0.2 [0.2, 0.2]	-0.11 [-0.2, -0.0]	-0.0 [-0.08, 0.01]	0.01 [-0.0, 0.31]	-0.0 [-0.02, -0.0]

Fig. S8. Onset and dynamics of the steppe migrations We seek to analyze the migration of Yamnaya herders from the Western Pontic Steppe into Europe, beginning around 3000 BCE. **(A)** Inferred migration over the entire time period, with each plot showing one row of the migration matrix $\mathbf{A}(t)$ at each time point. **(B)** Allowed edges during migration inference at the coarse-grained level. **(C)** Inferred migration after coarse-graining across time windows. For each matrix element, the mean value is shown along with the lower and upper quartiles. The pronounced upper-diagonal structure reflects westward movement, as populations are ordered from east to west in the matrix rows and columns. Confidence intervals correspond to the quartiles. **(D)** Visualization of the strongest migration links (values > 0.1) as directed arrows.**(E)** Relaxation times derived from the eigenvalues of the inferred matrix **(F)** Right eigenvectors from spectral decomposition of the inferred matrix.The first eigenvector identifies the steady state probability of the system and has the corresponding eigenvalue $\lambda = 1$.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

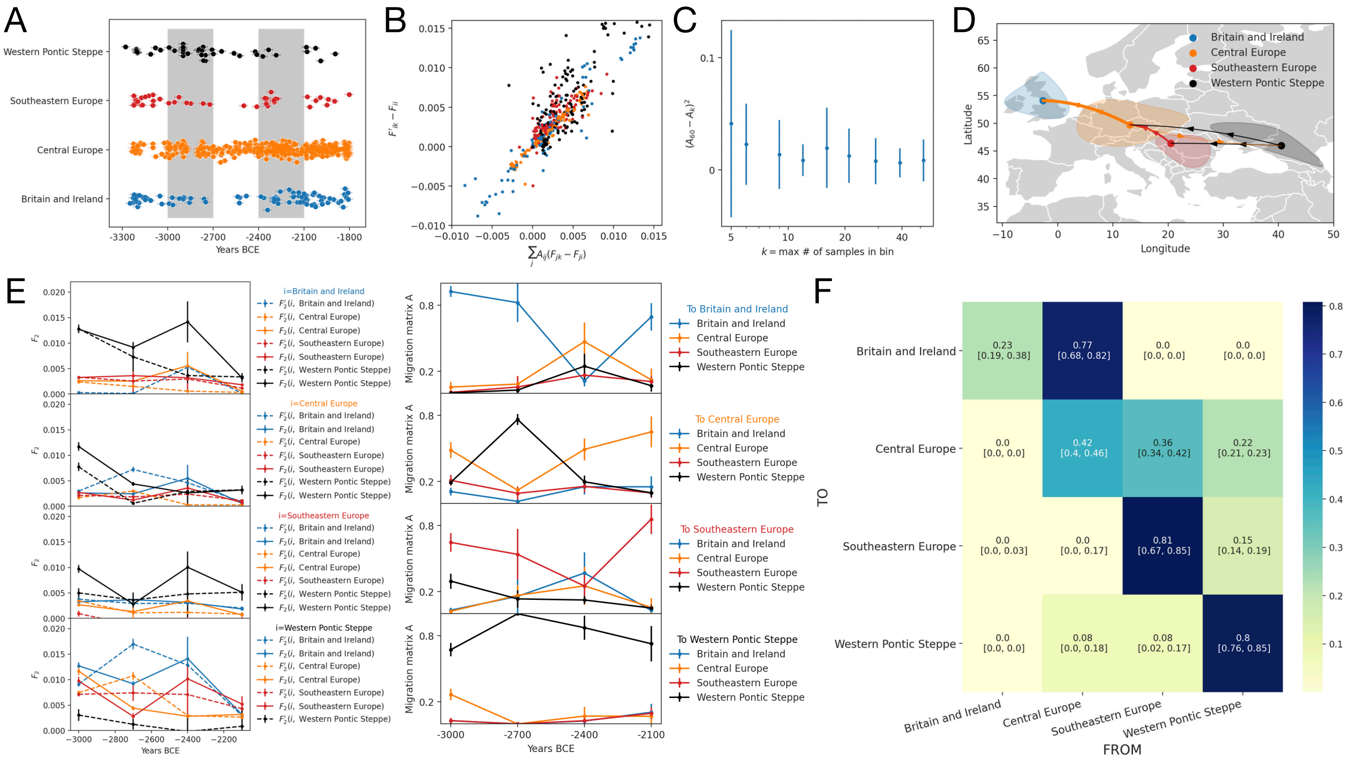


Fig. S9. Coarse graining of the two Central European clusters. Same analysis as in Fig.S8 but with "Central Europe - West" and "Central Europe - East" combined into one cluster. **(A)** Temporal organization of the samples analyzed. **(B)** Comparison of model prediction (the left side of Eq. 1) and observed time-lagged F2 statistics (the right side of Eq. 1) using the inferred matrix. **(C)** Stability of matrix inference with respect to subsampling of individuals in the Central Europe population within each bin. **(D)** Visualization of the strongest matrix elements define by the average matrix value bigger than 0.1. **(E)** Temporal variation of F_2 statistics (left) and inferred matrix elements (right) across the analyzed time period. **(F)** Inferred migration matrix obtained by integrating information across the whole time period.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

D Bibliography: AADR Dataset

[1] Cristina Gamba et al. “Genome flux and stasis in a five millennium transect of European prehistory”. In: *Nat Commun* 5.1 (Oct. 2014). DOI: `10.1038/ncomms6257`.

[2] Iñigo Olalde et al. “Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European”. In: *Nature* 507.7491 (Jan. 2014), pp. 225–228. DOI: `10.1038/nature12960`.

[3] Iosif Lazaridis et al. “Ancient human genomes suggest three ancestral populations for present-day Europeans”. In: *Nature* 513.7518 (Sept. 2014), pp. 409–413. DOI: `10.1038/nature13673`.

[4] Pontus Skoglund et al. “Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers”. In: *Science* 344.6185 (May 2014), pp. 747–750. DOI: `10.1126/science.1253448`.

[5] Eppie R. Jones et al. “Upper Palaeolithic genomes reveal deep roots of modern Eurasians”. In: *Nat Commun* 6.1 (Nov. 2015). DOI: `10.1038/ncomms9912`.

[6] Iñigo Olalde et al. “A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures”. In: *Mol Biol Evol* (Sept. 2015), msv181. DOI: `10.1093/molbev/msv181`.

[7] Iain Mathieson et al. “Genome-wide patterns of selection in 230 ancient Eurasians”. In: *Nature* 528.7583 (Nov. 2015), pp. 499–503. DOI: `10.1038/nature16152`.

[8] Lara M. Cassidy et al. “Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome”. In: *Proc. Natl. Acad. Sci. U.S.A.* 113.2 (Dec. 2015), pp. 368–373. DOI: `10.1073/pnas.1518445113`.

[9] Morten E. Allentoft et al. “Population genomics of Bronze Age Eurasia”. In: *Nature* 522.7555 (June 2015), pp. 167–172. DOI: `10.1038/nature14507`.

[10] Torsten Günther et al. “Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques”. In: *Proc. Natl. Acad. Sci. U.S.A.* 112.38 (Sept. 2015), pp. 11917–11922. DOI: `10.1073/pnas.1509851112`.

[11] Wolfgang Haak et al. “Massive migration from the steppe was a source for Indo-European languages in Europe”. In: *Nature* 522.7555 (Mar. 2015), pp. 207–211. DOI: `10.1038/nature14317`.

[12] Ayça Omrak et al. “Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool”. In: *Current Biology* 26.2 (Jan. 2016), pp. 270–275. DOI: `10.1016/j.cub.2015.12.019`.

[13] Farnaz Broushaki et al. “Early Neolithic genomes from the eastern Fertile Crescent”. In: *Science* 353.6298 (July 2016), pp. 499–503. DOI: `10.1126/science.aaf7943`.

[14] Gülşah Merve Kılınç et al. “The Demographic Development of the First Farmers in Anatolia”. In: *Current Biology* 26.19 (Oct. 2016), pp. 2659–2666. DOI: `10.1016/j.cub.2016.07.057`.

[15] Iosif Lazaridis et al. “Genomic insights into the origin of farming in the ancient Near East”. In: *Nature* 536.7617 (July 2016), pp. 419–424. DOI: `10.1038/nature19310`.

[16] Qiaomei Fu et al. “The genetic history of Ice Age Europe”. In: *Nature* 534.7606 (May 2016), pp. 200–205. DOI: `10.1038/nature17993`.

[17] Rui Martiniano et al. “Genomic signals of migration and continuity in Britain before the Anglo-Saxons”. In: *Nat Commun* 7.1 (Jan. 2016). DOI: `10.1038/ncomms10326`.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[18] Stephan Schiffels et al. “Iron Age and Anglo-Saxon genomes from East England reveal British migration history”. In: *Nat Commun* 7.1 (Jan. 2016). DOI: `10.1038/ncomms10408`.

[19] Zuzana Hofmanová et al. “Early farmers from across Europe directly descended from Neolithic Aegeans”. In: *Proc. Natl. Acad. Sci. U.S.A.* 113.25 (June 2016), pp. 6886–6891. DOI: `10.1073/pnas.1523951113`.

[20] Edwin C. M. van den Brink et al. “A Late Bronze Age II clay coffin from Tel Shaddud in the Central Jezreel Valley, Israel: context and historical implications”. In: *Levant* 49.2 (May 2017), pp. 105–135. DOI: `10.1080/00758914.2017.1368204`.

[21] Eppie R. Jones et al. “The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers”. In: *Current Biology* 27.4 (Feb. 2017), pp. 576–582. DOI: `10.1016/j.cub.2016.12.060`.

[22] Gloria González-Fortes et al. “Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin”. In: *Current Biology* 27.12 (June 2017), 1801–1810.e10. DOI: `10.1016/j.cub.2017.05.023`.

[23] Iosif Lazaridis et al. “Genetic origins of the Minoans and Mycenaeans”. In: *Nature* 548.7666 (Aug. 2017), pp. 214–218. DOI: `10.1038/nature23310`.

[24] Lehti Saag et al. “Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe”. In: *Current Biology* 27.14 (July 2017), 2185–2193.e6. DOI: `10.1016/j.cub.2017.06.022`.

[25] Marc Haber et al. “Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences”. In: *The American Journal of Human Genetics* 101.2 (Aug. 2017), pp. 274–282. DOI: `10.1016/j.ajhg.2017.06.013`.

[26] Mark Lipson et al. “Parallel palaeogenomic transects reveal complex genetic history of early European farmers”. In: *Nature* 551.7680 (Nov. 2017), pp. 368–372. DOI: `10.1038/nature24476`.

[27] Martin Sikora et al. “Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers”. In: *Science* 358.6363 (Oct. 2017), pp. 659–662. DOI: `10.1126/science.aao1807`.

[28] Martina Unterländer et al. “Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe”. In: *Nat Commun* 8.1 (Mar. 2017). DOI: `10.1038/ncomms14615`.

[29] Rui Martiniano et al. “The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods”. In: *PLoS Genet* 13.7 (July 2017). Ed. by Anna Di Rienzo, e1006852. DOI: `10.1371/journal.pgen.1006852`.

[30] Verena J. Schuenemann et al. “Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods”. In: *Nat Commun* 8.1 (May 2017). DOI: `10.1038/ncomms15694`.

[31] Éadaoin Harney et al. “Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation”. In: *Nat Commun* 9.1 (Aug. 2018). DOI: `10.1038/s41467-018-05649-9`.

[32] Alissa Mittnik et al. “The genetic prehistory of the Baltic Sea region”. In: *Nat Commun* 9.1 (Jan. 2018). DOI: `10.1038/s41467-018-02825-9`.

[33] Carlos Eduardo G. Amorim et al. “Understanding 6th-century barbarian social organization and migration through paleogenomics”. In: *Nat Commun* 9.1 (Sept. 2018). DOI: `10.1038/s41467-018-06024-4`.

[34] Cristina Valdiosera et al. “Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia”. In: *Proc. Natl. Acad. Sci. U.S.A.* 115.13 (Mar. 2018), pp. 3428–3433. DOI: `10.1073/pnas.1717762115`.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[35] D. M. Fernandes et al. “A genomic Neolithic time transect of hunter-farmer admixture in central Poland”. In: *Sci Rep* 8.1 (Oct. 2018). DOI: `10.1038/s41598-018-33067-w`.

[36] Iñigo Olalde et al. “The Beaker phenomenon and the genomic transformation of northwest Europe”. In: *Nature* 555.7695 (Feb. 2018), pp. 190–196. DOI: `10.1038/nature25738`.

[37] Iain Mathieson et al. “The genomic history of southeastern Europe”. In: *Nature* 555.7695 (Feb. 2018), pp. 197–203. DOI: `10.1038/nature25778`.

[38] Krishna R. Veeramah et al. “Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria”. In: *Proc. Natl. Acad. Sci. U.S.A.* 115.13 (Mar. 2018), pp. 3494–3499. DOI: `10.1073/pnas.1719880115`.

[39] Maja Krzewińska et al. “Genomic and Strontium Isotope Variation Reveal Immigration Patterns in a Viking Age Town”. In: *Current Biology* 28.17 (Sept. 2018), 2730–2738.e10. DOI: `10.1016/j.cub.2018.06.053`.

[40] Maja Krzewińska et al. “Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads”. In: *Sci. Adv.* 4.10 (Oct. 2018). DOI: `10.1126/sciadv.aat4457`.

[41] Niall O’Sullivan et al. “Ancient genome-wide analyses infer kinship structure in an Early Medieval Alemannic graveyard”. In: *Sci. Adv.* 4.9 (Sept. 2018). DOI: `10.1126/sciadv.aao1262`.

[42] Peter de Barros Damgaard et al. “137 ancient human genomes from across the Eurasian steppes”. In: *Nature* 557.7705 (May 2018), pp. 369–374. DOI: `10.1038/s41586-018-0094-2`.

[43] Peter de Barros Damgaard et al. “The first horse herders and the impact of early Bronze Age steppe expansions into Asia”. In: *Science* 360.6396 (June 2018). DOI: `10.1126/science.aar7711`.

[44] Pierre Zalloua et al. “Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza”. In: *Sci Rep* 8.1 (Dec. 2018). DOI: `10.1038/s41598-018-35667-y`.

[45] Thiseas C. Lamnidis et al. “Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe”. In: *Nat Commun* 9.1 (Nov. 2018). DOI: `10.1038/s41467-018-07483-5`.

[46] Torsten Günther et al. “Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation”. In: *PLoS Biol* 16.1 (Jan. 2018). Ed. by Nick Barton, e2003703. DOI: `10.1371/journal.pbio.2003703`.

[47] Alissa Mittnik et al. “Kinship-based social inequality in Bronze Age Europe”. In: *Science* 366.6466 (Nov. 2019), pp. 731–734. DOI: `10.1126/science.aax6219`.

[48] Christiana L. Scheib et al. “East Anglian early Neolithic monument burial linked to contemporary Megaliths”. In: *Annals of Human Biology* 46.2 (Feb. 2019), pp. 145–149. DOI: `10.1080/03014460.2019.1623912`.

[49] Chuan-Chao Wang et al. “Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions”. In: *Nat Commun* 10.1 (Feb. 2019). DOI: `10.1038/s41467-018-08220-8`.

[50] Federico Sánchez-Quinto et al. “Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society”. In: *Proc. Natl. Acad. Sci. U.S.A.* 116.19 (Apr. 2019), pp. 9469–9474. DOI: `10.1073/pnas.1818037116`.

[51] G. González-Fortes et al. “A western route of prehistoric human migration from Africa into the Iberian Peninsula”. In: *Proc. R. Soc. B.* 286.1895 (Jan. 2019), p. 20182288. DOI: `10.1098/rspb.2018.2288`.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[52] Hannes Schroeder et al. “Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave”. In: *Proc. Natl. Acad. Sci. U.S.A.* 116.22 (May 2019), pp. 10705–10710. DOI: `10.1073/pnas.1820210116`.

[53] Helena Malmström et al. “The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon”. In: *Proc. R. Soc. B.* 286.1912 (Oct. 2019), p. 20191528. DOI: `10.1098/rspb.2019.1528`.

[54] Iñigo Olalde et al. “The genomic history of the Iberian Peninsula over the past 8000 years”. In: *Science* 363.6432 (Mar. 2019), pp. 1230–1234. DOI: `10.1126/science.aav4040`.

[55] Lehti Saag et al. “The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers further East”. In: *Current Biology* 29.10 (May 2019), 1701–1711.e16. DOI: `10.1016/j.cub.2019.04.026`.

[56] Marc Haber et al. “A Transient Pulse of Genetic Admixture from the Crusaders in the Near East Identified from Ancient Genome Sequences”. In: *The American Journal of Human Genetics* 104.5 (May 2019), pp. 977–984. DOI: `10.1016/j.ajhg.2019.03.015`.

[57] Margaret L. Antonio et al. “Ancient Rome: A genetic crossroads of Europe and the Mediterranean”. In: *Science* 366.6466 (Nov. 2019), pp. 708–714. DOI: `10.1126/science.aay6826`.

[58] Mari Järve et al. “Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance”. In: *Current Biology* 29.14 (July 2019), 2430–2441.e10. DOI: `10.1016/j.cub.2019.06.019`.

[59] Martin Sikora et al. “The population history of northeastern Siberia since the Pleistocene”. In: *Nature* 570.7760 (June 2019), pp. 182–188. DOI: `10.1038/s41586-019-1279-z`.

[60] Michal Feldman et al. “Ancient DNA sheds light on the genetic origins of early Iron Age Philistines”. In: *Sci. Adv.* 5.7 (July 2019). DOI: `10.1126/sciadv.aax0061`.

[61] Michal Feldman et al. “Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia”. In: *Nat Commun* 10.1 (Mar. 2019). DOI: `10.1038/s41467-019-09209-7`.

[62] Selina Brace et al. “Ancient genomes indicate population replacement in Early Neolithic Britain”. In: *Nat Ecol Evol* 3.5 (Apr. 2019), pp. 765–771. DOI: `10.1038/s41559-019-0871-9`.

[63] Vagheesh M. Narasimhan et al. “The formation of human populations in South and Central Asia”. In: *Science* 365.6457 (Sept. 2019). DOI: `10.1126/science.aat7487`.

[64] Vanessa Villalba-Mouco et al. “Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula”. In: *Current Biology* 29.7 (Apr. 2019), 1169–1177.e7. DOI: `10.1016/j.cub.2019.02.006`.

[65] Alexandra Coutinho et al. “The Neolithic Pitted Ware culture foragers were culturally but not genetically influenced by the Battle Axe culture herders”. In: *Am J Phys Anthropol* 172.4 (June 2020), pp. 638–649. DOI: `10.1002/ajpa.24079`.

[66] Anja Furtwängler et al. “Comparison of target enrichment strategies for ancient pathogen DNA”. In: *BioTechniques* 69.6 (Dec. 2020), pp. 455–459. DOI: `10.2144/btn-2020-0100`.

[67] Anna Linderholm et al. “Corded Ware cultural complexity uncovered using genomic and isotopic analysis from south-eastern Poland”. In: *Sci Rep* 10.1 (Apr. 2020). DOI: `10.1038/s41598-020-63138-w`.

[68] Ashot Margaryan et al. “Population genomics of the Viking world”. In: *Nature* 585.7825 (Sept. 2020), pp. 390–396. DOI: `10.1038/s41586-020-2688-8`.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[69] Daniel M. Fernandes et al. "The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean". In: *Nat Ecol Evol* 4.3 (Feb. 2020), pp. 334–345. DOI: `10.1038/s41559-020-1102-0`.

[70] David Gokhman et al. "Differential DNA methylation of vocal and facial anatomy genes in modern humans". In: *Nat Commun* 11.1 (Mar. 2020). DOI: `10.1038/s41467-020-15020-6`.

[71] Eirini Skourtanioti et al. "Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus". In: *Cell* 181.5 (May 2020), 1158–1175.e28. DOI: `10.1016/j.cell.2020.04.044`.

[72] Joachim Burger et al. "Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 3,000 Years". In: *Current Biology* 30.21 (Nov. 2020), 4307–4315.e13. DOI: `10.1016/j.cub.2020.08.033`.

[73] Joseph H. Marcus et al. "Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia". In: *Nat Commun* 11.1 (Feb. 2020). DOI: `10.1038/s41467-020-14523-6`.

[74] Lara M. Cassidy et al. "A dynastic elite in monumental Neolithic society". In: *Nature* 582.7812 (June 2020), pp. 384–388. DOI: `10.1038/s41586-020-2378-6`.

[75] Lily Agranat-Tamir et al. "The Genomic History of the Bronze Age Southern Levant". In: *Cell* 181.5 (May 2020), 1146–1157.e11. DOI: `10.1016/j.cell.2020.04.024`.

[76] Maïté Rivollat et al. "Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers". In: *Sci. Adv.* 6.22 (May 2020). DOI: `10.1126/sciadv.aaz5344`.

[77] Marc Haber et al. "A Genetic History of the Near East from an aDNA Time Course Sampling Eight Points in the Past 4,000 Years". In: *The American Journal of Human Genetics* 107.1 (July 2020), pp. 149–157. DOI: `10.1016/j.ajhg.2020.05.008`.

[78] Samantha Brunel et al. "Ancient genomes from present-day France unveil 7,000 years of its demographic history". In: *Proc. Natl. Acad. Sci. U.S.A.* 117.23 (May 2020), pp. 12791–12798. DOI: `10.1073/pnas.1918034117`.

[79] Éadaoin Harney et al. "A minimally destructive protocol for DNA extraction from ancient teeth". In: *Genome Res.* 31.3 (Feb. 2021), pp. 472–483. DOI: `10.1101/gr.267534.120`.

[80] Andaine Seguin-Orlando et al. "Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France". In: *Current Biology* 31.5 (Mar. 2021), 1072–1083.e10. DOI: `10.1016/j.cub.2020.12.015`.

[81] Anne Friis-Holm Egfjord et al. "Genomic Steppe ancestry in skeletons from the Neolithic Single Grave Culture in Denmark". In: *PLoS ONE* 16.1 (Jan. 2021). Ed. by Peter F. Biehl, e0244872. DOI: `10.1371/journal.pone.0244872`.

[82] Guido Alberto Gnecchi-Ruscone et al. "Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians". In: *Sci. Adv.* 7.13 (Mar. 2021). DOI: `10.1126/sciadv.abe4414`.

[83] Lehti Saag et al. "Genetic ancestry changes in Stone to Bronze Age transition in the East European plain". In: *Sci. Adv.* 7.4 (Jan. 2021). DOI: `10.1126/sciadv.abd6535`.

[84] Luka Papac et al. "Dynamic changes in genomic and social structures in third millennium BCE central Europe". In: *Sci. Adv.* 7.35 (Aug. 2021). DOI: `10.1126/sciadv.abi6941`.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.12.710875; this version posted March 12, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[85] Mario Novak et al. "Genome-wide analysis of nearly all the victims of a 6200 year old massacre". In: *PLoS ONE* 16.3 (Mar. 2021). Ed. by Peter F. Biehl, e0247332. DOI: `10.1371/journal.pone.0247332`.

[86] Reyhan Yaka et al. "Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes". In: *Current Biology* 31.11 (June 2021), 2455–2468.e18. DOI: `10.1016/j.cub.2021.03.050`.