Climate and Prehistoric Migration*

Peter Huybers Marco Tabellini Charles A. Taylor Francesco Toti

January 2026

Abstract

How did climate shape human migration before modern states, markets, and borders? We study prehistoric migration over the past 10,000 years using ancient DNA to measure population movements directly. Building on recent advances in genetics, we use identity-by-descent links between ancient individuals to construct a panel of bilateral migration flows across Eurasia, analogous to migration datasets commonly used in economics but measured from genetic relatedness. We combine these flows with high-resolution climate data to study how environmental similarity shaped migration decisions over the very long run. We document three main findings. First, migration flows decline with climate and ecological distance, as measured by the difference in growing degree days, precipitation, and soil characteristics between origins and destinations. Second, these patterns are stronger in farming and pastoral societies than among foragers, pointing to subsistence-based mechanisms. Third, migration responds to climate change: populations exposed to warming migrate toward cooler destinations, suggesting that local adaptation was costly and that migration served as an important margin of adjustment. Together, the results provide new evidence on the ecological foundations of human mobility in deep history.

Keywords: Migration, ancient human DNA, prehistory, climate, ecology

JEL codes: N50, O13, Q54, Z13

*Huybers: Department of Earth and Planetary Science, Harvard University. Tabellini: Harvard Business School and NBER. Taylor: Harvard Kennedy School and NBER. Toti: University of Michigan, Ann Arbor. We thank Gabrielle Lévy for excellent research assistance. All remaining errors are ours.

1 Introduction

For most of human history, migration was not a response to wage differentials, institutions, or political borders, but to the spatial distribution of ecological opportunities and risks. Hunter-gatherers, early farmers, and pastoralists all depended on local ecological conditions, though the specific constraints differed: farmers were bound to thermal regimes governing crop growth, while foragers and pastoralists were more sensitive to water availability and pasture conditions. A central insight of anthropology, paleoecology, and archaeology is that mobility strategies were organized around seasonal patterns: where food could be produced, stored, or accessed during the growing season, and how predictable those conditions were from one place to the next (Lee and DeVore, 1968). In mid- and high-latitude environments, where the length and intensity of the warm season constrains biological productivity, moving into a region with an unfamiliar thermal regime could threaten subsistence itself.

Quantitative studies of crop dispersal show that thermal thresholds and growing-season length constrained where early farmers could move, shaping migration routes, settlement patterns, and even the adoption of mixed subsistence strategies (Guedes and Butler, 2014; Castiello et al., 2025). Other work emphasizes that climate influences migration indirectly, through its effects on vegetation and land cover (Racimo et al., 2020). At the same time, paleoclimate evidence shows that major migration episodes often occurred under very different mean climate states, complicating deterministic narratives and pointing instead to the importance of seasonal structure (Tierney et al., 2017). Despite these insights, much of this literature remains necessarily indirect: it infers mobility from settlement locations, crop assemblages, or environmental reconstructions, rather than from observed movements of people themselves. As a result, we still lack a population-level test of whether humans systematically moved along corridors of seasonal climate familiarity over the long run.

This paper brings new evidence to bear on this question using ancient human DNA that tracks population ancestry across space and time. By observing where the ancestors of individuals appear and reappear across Western Eurasia over the past 10,000 years, we can measure migration flows rather than infer them. This is made possible by recent advances in population genetics (Reich, 2018). In particular, long identity-by-descent (IBD) segments in ancient DNA reveal recent genealogical links between individuals who lived hundreds or even thousands of kilometers apart. Because recombination rapidly breaks down these segments over generations, their length provides a tight bound on the recency of shared ancestry. Ringbauer et al. (2024) introduce an algorithm (ancIBD) that reliably detects IBD segments even in low-coverage ancient genomes, allowing biological relatedness to be traced at time horizons of centuries across thousands of individuals.

We use these pairwise genealogical links to construct a panel of migration flows across Western Eurasia, broadly defined to match the spatial coverage of existing ancient DNA datasets, from 8000 BCE onwards. We discretize space into geographic cells of $1^\circ \times 1^\circ$ latitude–longitude and time into 500-year windows, and define migration between two locations as the count of individual pairs sharing IBD segments linking individuals buried in different places but close in time. Aggregating these links yields a migration matrix that captures the intensity of population movement across space and time. We then combine this genetically derived migration matrix with high-resolution modern climate data and paleoclimate reconstructions to test whether prehistoric migration was systematically shaped by similarity in local climate and ecological factors.

We adopt a gravity framework commonly used in the international trade literature (Anderson and Van Wincoop, 2003), and express bilateral migration as a function of physical distance. Building on recent work that documents the importance of climate matching for migration in the modern era (Obolensky et al., 2024), we augment the baseline gravity specification by introducing climate distance between origin and destination. We consider three main proxies for climate: temperature, precipitation, and soil.

Our preferred measure of temperature distance is the absolute difference in cumulative growing-degree days (GDD) above 10°C, which captures biologically relevant heat accumulation for plant growth and animal productivity (Schlenker and Roberts, 2009; Butler and Huybers, 2013). Unlike seasonal or annual temperature averages, degree days summarize both the intensity and duration of warmth during the growing season, and are therefore closely aligned with subsistence constraints emphasized in anthropology and archaeology. Paleoclimate evidence further shows that growing-season heat accumulation can evolve independently of mean summer temperatures over the Holocene, making degree days a more appropriate measure of climatic similarity in deep history (Mourshed, 2012; Marsicek et al., 2018).

We measure precipitation using average annual rainfall, which captures broad differences in water availability relevant for agriculture and pasture. To account for soil conditions, we construct a composite index summarizing key characteristics that shape agricultural suitability—organic carbon content, bulk density, pH, clay content, and water-holding capacity. As with temperature, we include absolute differences in precipitation and soil characteristics between origin and destination to control for other dimensions of environmental similarity that may influence migration decisions.

We find that climate distance reduces migration along all three ecological dimensions. Migration flows decline sharply with differences in cumulative GDD above 10°C between origin and destination, while temperature differences outside the growing season—proxied

by winter temperature—play a limited role. Precipitation distance also consistently reduces migration. Soil characteristics emerge as another important dimension of ecological matching: differences in soil quality between origin and destination significantly reduce migration flows, consistent with the idea that agricultural knowledge and practices were tied to specific edaphic conditions.

A natural concern is the spatial correlation of climate: nearby locations tend to have similar environmental conditions, and migrants tend to move short distances. To address this, we flexibly control for geographic distance by including both distance and distance squared. This reduces the magnitude of the estimated temperature-distance coefficients, but leaves their precision largely unchanged. By contrast, the coefficient on precipitation distance is essentially unaffected by the inclusion of geographic distance, suggesting that differences in water availability operated as an independent ecological constraint on migration rather than simply proxying for spatial proximity. Beyond bilateral frictions, we further account for location-specific factors that may influence migration. Our preferred specification includes origin fixed effects, destination fixed effects, and era fixed effects at 500-year intervals. Era fixed effects absorb aggregate shifts in migration intensity over time as well as broad changes in climatic conditions, such as periods of sustained warming or cooling. Origin and destination fixed effects account for persistent differences across locations in population size, mobility propensity, and data coverage.

According to our preferred specification, increasing the distance in GDD by 1,000 reduces migration flows by 17%. For reference, a difference of 1,000 GDD is equivalent to a 5°C higher mean daily temperature sustained over a standard 200-day growing season—that is, roughly 5°C more warmth on each day above the 10°C threshold. From an agronomic perspective, this represents a substantial climatic shift, comparable to the difference between a cool-temperate cropping system such as wheat (which requires on the order of 1,500 GDD) and a warm-temperate system such as maize (which requires closer to 2,500 GDD). Quantitatively, this effect is also comparable in magnitude to increasing geographic distance by roughly 250 kilometers. Differences in precipitation and soil characteristics also have meaningful effects. This underscores the importance of edaphic conditions for agricultural viability and suggests that knowledge and practices tied to specific soil environments were difficult to transfer across locations.

We subject these findings to an extensive battery of robustness checks. First, the results are robust to interacting origin and destination fixed effects with era fixed effects, which absorb time-varying push and pull factors such as changes in local population size, subsistence systems, environmental conditions, or sampling intensity. Second, the baseline effect is robust to restricting the sample to moves of at least 100 km, and to estimating the model

separately for horizontal (east–west) and vertical (north–south) movements. Notably, horizontal moves exhibit patterns similar to those in the full sample, while vertical moves—which are far fewer in number—show weaker but directionally consistent effects. Third, our results are robust to replacing straight-line geographic distance with a travel-cost measure that accounts for terrain difficulty and water barriers. Specifically, we compute least-cost paths using a transition matrix derived from slope (from elevation data) and water body coverage, where steep terrain and water crossings increase effective travel distance. This adjustment matters because simple geographic distance may substantially understate the difficulty of crossing mountain ranges or bodies of water that shaped prehistoric mobility corridors.

In addition, the results hold when using alternative dependent variables, including counts of IBD segments above 12 cM and 16 cM thresholds as well as the cumulative length of shared segments. This robustness indicates that climate matching operates similarly for both very recent genealogical links (longer segments) and more distant relationships (shorter segments). We also obtain similar results in individual-level specifications that measure, for each ancestor, the total amount of shared genetic material with individuals located in other grid cells within the same 500-year period. Moreover, the results are unchanged when controlling for additional environmental distances—such as ruggedness, elevation, and diurnal temperature range—beyond our baseline climate measures. Finally, the precision of the estimates is unaffected when using Conley (1999)-adjusted standard errors or when clustering at larger spatial units.

These patterns are consistent with the idea that migration costs in prehistory were driven primarily by the specificity of subsistence technologies. Success in pre-industrial societies relied on detailed, locally accumulated knowledge of biotic cycles—when to plant, when to herd, and what to forage. These cycles are governed in large part by accumulated thermal time rather than by mean temperature alone. The results align with anthropological accounts emphasizing seasonal resource scheduling (Lee and DeVore, 1968) and with archaeological evidence showing that the spread of farming and herding was constrained by growing-season length and thermal thresholds (Guedes and Butler, 2014; Castiello et al., 2025). More broadly, the findings highlight the importance of climate-specific human capital as a mechanism shaping migration decisions (Steckel, 1983; Obolensky et al., 2024): individuals possess skills and knowledge adapted to particular climatic regimes, whose transferability declines with climate distance (e.g., pastoralists from warm climates may struggle to keep livestock alive in freezing conditions, while cold-climate pastoralists face severe thermal stress and parasite loads in hot summers).

Consistent with this interpretation, we find that the response to GDD distance is stronger among populations with greater reliance on farming and pastoralism, as compared to pre-

dominantly foraging societies. In contrast, we find no corresponding heterogeneity along the winter temperature dimension. This pattern reinforces the view that growing-season conditions were the primary constraint shaping migration decisions, particularly in societies whose subsistence depended directly on crops or managed livestock. For foragers, whose livelihoods were less tightly linked to fixed growing seasons, climate matching along the summer dimension appears to have been less salient.

One advantage of our setting is that it spans nearly 10,000 years, allowing us to examine how the relationship between climate similarity and migration evolves over time. We find that climate matching is weaker during periods of warming. One interpretation is that warming relaxes ecological constraints on subsistence, expanding the range of environments in which crops, livestock, and vegetation regimes can be sustained. As growing seasons lengthen and thermal constraints ease, the costs of moving into climatically unfamiliar areas fall, reducing the penalty associated with climate mismatch. In this sense, climate matching matters most when environments are near subsistence thresholds, whereas periods of warming attenuate these constraints by making a broader set of destinations viable. This pattern is consistent with archaeological evidence that phases of climatic amelioration are associated with range expansions, ecological experimentation, and reduced sensitivity to local growing-season conditions (Richerson et al., 2001; Dow et al., 2009; Olsson, 2024).

The long time span of our setting, together with substantial climate change over the Holocene (Marcott et al., 2013), allows us to ask whether and how individuals adapted to changing climatic conditions. In particular, we examine whether migration served as a mechanism to offset local climate change. We compare the climate experienced by individuals' ancestors at their origin with the climate at their eventual destination, focusing on pairs of related individuals observed within 500 years. We then relate changes in temperature over time at the origin to changes in temperature over space induced by migration.[1] Intuitively, if individuals respond to local warming by relocating toward places that match their ancestral ecological niche, then increases in average temperature at the origin should be associated with migration toward cooler destinations.

We find evidence that migration acted as an adaptive response to climate change. When origins experience warmer summers, individuals systematically migrate toward destinations with cooler summer temperatures, partially offsetting local warming. These patterns mirror our earlier findings on bilateral migration flows and highlight the central role of growing-season conditions in shaping mobility decisions. These results imply that local adaptation

[1]Because paleoclimate reconstructions do not permit reliable measurement of within-year temperature variation or growing-degree days at fine temporal resolution, this analysis uses average annual temperature as a coarser proxy for climatic conditions.

to climate change was costly or constrained. If adaptation through changes in crops, technologies, or practices were easy or immediate, individuals would have had little incentive to relocate in response to gradual warming. Instead, the systematic migration responses we document indicate that migration served as a key margin of adjustment when local subsistence systems were slow to adapt. This interpretation is consistent with evidence that preindustrial subsistence depended tightly on growing-season conditions and that adapting to new thermal regimes required substantial changes in crops, knowledge, or social organization (Richerson et al., 2001; Ashraf and Michalopoulos, 2015).

Our paper is related to the literature that studies how environmental conditions and subsistence systems shape long-run economic and social outcomes. Michalopoulos (2012) shows that variation in land quality and terrain fosters ethnolinguistic differentiation by limiting population movement. Fenske (2014) documents that ecological diversity promotes trade across zones and supports the emergence of centralized states in pre-colonial Africa. Ashraf and Michalopoulos (2015) link climatic variability to the diffusion of agriculture, while Giuliano and Nunn (2021) show that environmental stability across generations fosters cultural persistence. Matranga (2024) highlights the role of climatic seasonality in the transition from mobility to sedentarism and the invention of agriculture.

A primary contribution of this paper is to introduce to the literature an approach and dataset that captures prehistorical migration flows. Building on seminal work in genetics (Reich, 2018; Ringbauer et al., 2024), we derive population-level bilateral migration flows from ancient DNA and analyze prehistoric mobility using the same empirical framework—gravity-style flows, distance frictions, and origin–destination structure—that economists routinely apply to modern migration. In doing so, we provide evidence on how populations adjusted to environmental constraints and climate change in deep history, and document systematic patterns linking migration to seasonality, subsistence systems, and long-run climatic change. These findings complement theories that emphasize environmental forces, population dynamics, and adaptation in shaping long-run development (Galor and Moav, 2002; Ashraf and Galor, 2013; Galor and Özak, 2016). Our evidence on migration responses to climate change also complements work in cultural evolution that highlights the slow, cumulative nature of local adaptation and the costs of acquiring ecologically specific subsistence knowledge, particularly in seasonal environments (Henrich, 2004; Boyd et al., 2011; Nunn, 2021).

Finally, our paper complements a recent literature that documents the role of climate similarity in shaping migration and development in historical and modern settings. Early historical evidence documents that 19th-century US farmers disproportionately migrated along east–west corridors toward regions with familiar climatic and agricultural conditions (Steckel, 1983). Obolensky et al. (2024) confirm this pattern, showing that migrants in

the United States between 1850 and 2020 systematically sorted toward climatically similar destinations. Nicolò et al. (2026) provide evidence in support of Alfred Crosby's hypothesis that Europeans thrived in the Neo-Europes because of environmental similarity between origin and destination (Crosby, 1986). More broadly, our findings speak to long-standing arguments that the east–west orientation of Eurasia facilitated the diffusion of people, crops, and technologies by limiting climatic barriers to movement (Diamond, 1997), a pattern also reflected in genetic evidence showing that latitude explains more human genetic variation than longitude (Ramachandran and Rosenberg, 2011). We contribute to this literature by pushing the analysis back into prehistory and by showing that climate similarity matters primarily along seasonal dimensions. In contrast to existing work, which focuses on mean annual temperature, our results highlight the central role of growing-season conditions in shaping migration decisions.

2 Data

This section describes the main datasets used in the analysis. Table 1 presents summary statistics for the main variables.

2.1 Inferring Migration from Ancient Human DNA

Prehistoric migration patterns are difficult to study because direct records of individual movements do not exist. To overcome this limitation, we rely on ancient DNA (aDNA) data, which make it possible to infer population movements indirectly from genealogical relationships observed across space and time. Our approach builds on recent advances in population genetics that allow researchers to identify pairs of individuals who share a recent common ancestor and to use the spatial separation of such pairs as evidence of past mobility (Reich, 2018).

The key object in our analysis is identity-by-descent (IBD). Two individuals share an IBD segment when they inherit the same stretch of DNA from a common ancestor without recombination. Because recombination breaks chromosomes into shorter segments in each generation, the length of an IBD segment—measured in centimorgans (cM)—provides information about the recency of shared ancestry, with longer segments indicating more recent genealogical links. In particular, segments of at least 8 cM correspond to shared ancestors within a relatively recent time horizon. In modern genetic data, IBD has been widely used to study recent population structure and migration (Palamara et al., 2012; Ralph and Coop, 2013). Until recently, however, its use in ancient DNA was limited by the low genomic

coverage of ancient samples.

Ringbauer et al. (2024) develop a method, ancIBD, that overcomes this limitation by reliably detecting IBD segments even in low-coverage aDNA. Applying this method to a large collection of ancient genomes from Western Eurasia, they identify thousands of pairwise genealogical links spanning distances of up to several hundred kilometers and time horizons of a few hundred years. We rely on their publicly available IBD calls and associated metadata as the foundation of our migration measure. We interpret a pair of related individuals buried in different locations as evidence that at least one movement occurred between those locations within the genealogical window implied by the IBD segment greater than 8cM. This interpretation does not require observing the migrant directly: even if both individuals are buried at their places of birth, the existence of a recent genealogical link across space implies migration by an intermediate ancestor. In this sense, IBD captures population movement rather than individual travel and is well suited to studying migration over long time horizons.

To focus on movements that are plausibly interpretable as migration rather than long-run population replacement, we restrict attention to pairs of individuals who lived within 1,000 years of one another. We further restrict the sample to individuals dated to after 8,000 BCE, a period for which both the density of aDNA samples and the quality of paleoclimate reconstructions (discussed below) improve substantially (Marcott et al., 2013; Reich, 2018). For each qualifying pair, we observe the burial locations of the two individuals, the time difference between them, and the total length of IBD segments they share. We aggregate these pairwise genealogical links to construct population-level migration flows at the gridcell level. We discretize space into $1^o \times 1^o$ latitude–longitude cells and time into 500-year windows. For each origin–destination–period triple, we define migration intensity as the count of individual pairs sharing at least one IBD segment of 8cM or longer, linking individuals buried in different cells but within the same time window.

For robustness, we also run our analysis using non-aggregated pair-wise genetic flows between individuals.

2.2 Climate

Modern. Our baseline analysis uses modern climate data, which offer substantially higher spatial resolution and richer within-year variation than currently available paleoclimate reconstructions. We measure local climate conditions using data from two sources. First, we use the ERA5, a global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing consistent, gridded fields of key variables like temperature and precipitation at hourly resolution from 1940 to present (Hers-

bach et al., 2020). We also use monthly climate data from TerraClimate (Abatzoglou et al., 2018), a widely used, high-resolution (≈4 km) global climate dataset constrained by a hydro-climate model that provides surface water and energy balance variables from 1958 onward. We compute average annual temperature and precipitation for each $1^o \times 1^o$ grid cell over the period 1960–2000, as well as summer (April–September) and winter (October–March) averages.[2] In addition, we use the daily ERA5 data to construct measures that capture growing-season conditions and seasonal exposure, including cumulative growing-degree days above 10°C—which summarize biologically relevant heat accumulation for plant growth and animal productivity.

Paleoclimate. We complement modern climate data with paleoclimate reconstructions both as a robustness check and to exploit temporal variation in climate over the Holocene. We use gridded reconstructions of annual mean surface air temperature from the NOAA World Data Service for Paleoclimatology, based on the Last Glacial Maximum Reanalysis (LGMR) product (Osman et al., 2021). This dataset provides temperature estimates at 200-year intervals and combines information from marine geochemical proxies—such as oxygen isotope ratios in carbonate ($\delta^{18}O_c$), Mg/Ca ratios, TEX_{86} (Tetraether Index), and UK'_{37} (Alkenone Unsaturation Index) with simulations from the Community Earth System Model (iCESM1.2 and iCESM1.3).[3] The resulting gridded temperature data have a spatial resolution of $1.9^o \times 2.5^o$, which we aggregate to the $1^o \times 1^o$ grid used to measure migration flows.

Paleoclimate reconstructions of temperature are highly correlated with modern temperature patterns and are generally regarded as reliable representations of long-run spatial and temporal variation in temperature. By contrast, precipitation reconstructions remain more uncertain and contested in the paleoclimate literature (Smerdon et al., 2017). For this reason, we only rely on modern data to measure precipitation throughout the analysis.

2.3 Soil, Topography, and Other Environmental Data

Soil characteristics. We obtain information on elevation and soil characteristics such as soil pH, soil water holding capacity, organic carbon, bulk density, and clay content from Hengl (2018b); Hengl and Wheeler (2018); Hengl and Gupta (2019); Hengl (2018a). We combine these soil variables into a composite soil quality index using principal component

[2] ERA5 precipitation is provided in meters per day; we convert to millimeters per month for comparability with TerraClimate.

[3] Oxygen isotope ratios in foraminifera shells ($\delta^{18}O_c$) reflect past sea surface temperatures and global ice volume, with lower values indicating warmer conditions. Mg/Ca ratios in these shells increase with temperature. TEX_{86}, based on lipids produced by marine archaea, provides an additional temperature proxy, particularly useful when carbonate records are sparse. UK'_{37}, derived from alkenones produced by coccolithophores, captures sea surface temperatures through temperature-sensitive variation in molecular unsaturation.

analysis, which captures the primary dimension of variation in edaphic conditions across Western Eurasia. In Figure A1, we plot the grid cell-level correlation between temperature and precipitation and each of these soil and topographic variables.

Diurnal temperature range. We also construct measures of diurnal temperature range—the difference between daily maximum and minimum temperatures—from ERA5 reanalysis data. This variable captures temperature volatility, which is relevant for crop stress, frost risk, and livestock management, and may represent a distinct dimension of climate matching beyond mean temperatures.

Land use. For the heterogeneity analysis, we obtain historical land use data from the HYDE 3.2 database (Klein Goldewijk et al., 2017), which provides gridded estimates of cropland and pasture area over the Holocene at 500-year intervals. We classify origin locations as predominantly forager, pastoral, or farming based on the relative intensity of cropland versus grazing land in each grid cell and time period.

Geographic distance. Geographic distance between grid cell centroids is computed using the Haversine formula, which accounts for the curvature of the Earth. We include both distance and distance squared in our specifications to allow for a flexible, non-linear relationship between migration flows and spatial separation.

3 Descriptive Evidence

3.1 The Spatial Correlates of Remains

We begin by mapping the ancient individuals from Ringbauer et al. (2024) that form the basis of our analysis. In total, the sample includes 6,065 individuals, with (approximate) dates of death ranging from 9500 BP (7550 BCE) to 127 BP (1823 CE).

A natural question is whether the spatial distribution of observed remains is systematically correlated with environmental or climatic characteristics. While our analysis below exploits variation across origin–destination cell pairs and controls for origin and destination fixed effects, it is nonetheless informative to document the geographic correlates of where remains are observed. To this end, we conduct a simple descriptive exercise. We define an indicator equal to one if at least one individual is observed in a given $1^\circ \times 1^\circ$ grid cell over the entire 7550 BCE–1823 CE period, and zero otherwise. We then regress this indicator on a set of environmental, climatic, and geographic variables measured at the grid-cell level.[4] Table 2 reports the results.

Column 1 relates the presence of remains to cumulative GDD above 10°C (expressed per

[4]Standard errors are clustered at the $2.5^\circ \times 2.5^\circ$ grid cell level.

1,000) and winter temperature (expressed in degrees Celsius). The coefficient on GDD is negative and statistically significant, indicating that remains are more likely to be observed in cooler parts of Western Eurasia. This pattern is consistent with existing evidence that ancient DNA preservation and recovery are more successful in cooler environments, where lower cumulative heat exposure slows DNA degradation (Reich, 2018). In contrast, winter temperature enters with a small but positive and statistically significant coefficient, suggesting that—conditional on cumulative heat exposure—remains are somewhat more likely to be observed in locations with milder winters. One possible interpretation is that winter cold plays a more limited role in long-run DNA degradation than cumulative thermal exposure during warmer months, while being correlated with human settlement and burial practices.

Columns 2–4 progressively introduce soil characteristics. Column 2 adds soil pH to the baseline specification, column 3 further adds soil clay content, and column 4 additionally controls for soil organic content. Soil pH and clay content are both positively and statistically significantly associated with the presence of remains. Higher clay content is consistent with reduced oxygen penetration and microbial activity, which can favor preservation. By contrast, soil organic content enters with a negative and statistically significant coefficient once included, suggesting that organic-rich soils—while often fertile—may accelerate biological decay processes that reduce the likelihood of long-run preservation.

Column 5 further augments the specification by adding precipitation. The coefficient on precipitation is negative and statistically significant, indicating that remains are less likely to be observed in wetter locations. This pattern may reflect preservation mechanisms—such as increased microbial activity or soil acidity in high-moisture environments—rather than lower historical population density.

Column 6 then jointly introduces diurnal temperature range and elevation. The coefficient on elevation is negative and statistically significant, indicating that remains are less likely to be observed at higher altitudes, consistent with both lower long-run population density and harsher environmental conditions affecting burial, preservation, and recovery. Diurnal temperature range enters with a small but positive and statistically significant coefficient. Conditional on precipitation and cumulative heat exposure, larger day–night temperature variation is characteristic of drier, more continental environments, which tend to inhibit microbial activity and organic decay. The positive coefficient on diurnal range is therefore consistent with improved long-run preservation in arid or semi-arid settings.

Finally, column 7 replicates the specification in column 6 while including fixed effects at the $2.5^\circ \times 2.5^\circ$ grid-cell level. This specification absorbs all time-invariant characteristics of each larger grid cell—including persistent differences in climate, topography, and archaeological sampling intensity—and therefore restricts identification to within-region vari-

ation. While several coefficients attenuate in magnitude, cumulative GDD, soil pH, soil clay content, and diurnal temperature range remain statistically significant, indicating that the correlations documented above are not driven solely by broad cross-regional environmental gradients.

Taken together, these patterns indicate that the spatial distribution of observed remains is correlated with several environmental characteristics that plausibly affect preservation and recovery. Importantly, however, these correlations are unlikely to drive our main results. The analysis below exploits variation in bilateral characteristics across origin–destination cell pairs, rather than variation in the presence of remains at individual locations, and includes origin and destination fixed effects that absorb all time-invariant differences across cells, including persistent environmental conditions and sampling intensity. As a result, identification comes from comparing migration flows across destinations that differ in climatic similarity for the same origin, and vice versa, holding constant location-specific factors. We return to this issue in Section 4 below when presenting the empirical strategy.

3.2 Climate

We next document the spatial distribution of key climate variables across the study region. Figure 4 displays two measures of modern climate at $1^\circ \times 1^\circ$ resolution. Panel A shows cumulative GDD above 10°C, derived from ERA5 reanalysis data, which captures biologically relevant heat accumulation during the growing season—a key determinant of agricultural productivity and vegetation patterns. Panel B shows mean monthly precipitation from ERA5. Both variables exhibit pronounced spatial gradients, with growing degree days declining sharply with latitude and precipitation varying substantially across longitudinal bands.

A defining feature of the Holocene is that it was marked by substantial fluctuations in average temperature over time (Marcott et al., 2013). Figure 6 illustrates these long-run dynamics by plotting reconstructed average temperature in the study region from 8,000 BCE to 0, highlighting alternating phases of warming and cooling. Figure 5 shows the spatial distribution of paleoclimate temperature reconstructions from marine proxy data (Osman et al., 2021). Panel A displays reconstructed temperature for an illustrative period (2950 BCE), while Panels B and C show how temperature changed over time: Panel B plots the change between 5000 and 4000 BCE, a period characterized by broad warming, while Panel C presents the corresponding change between 3500 and 2000 BCE, a period of net cooling.

The maps reveal pronounced spatial heterogeneity in temperature change: some regions experienced substantial warming, others cooled, and many saw relatively muted change over

the same period. Even within the same broad latitude bands, temperature shifts varied markedly across locations. This combination of long-run temporal change and fine-grained spatial heterogeneity generates meaningful variation in the climatic conditions faced by populations over time, which we exploit in the analysis below.

3.3 Migration

Figure 1 visualizes the ancient DNA derived migration flows across Western Eurasia in four panels—one for each time period, from 5000 BCE to 2000 BCE. Each panel displays links between locations where genetically related individuals are observed.[5] The figure reveals substantial variation in migration intensity across periods, as well as pronounced shifts in migration routes over millennia.

One pattern that stands out is the prevalence of "horizontal" movements, with migration occurring more frequently along east–west axes than along north–south ones. Such east–west mobility gradient has been documented in recent work using genealogical links in ancient DNA (Ringbauer et al., 2024). While this partly reflects the geography of the region under study, it is also consistent with the idea that migrants tended to move toward climatically similar environments. Such horizontal sorting has been documented in other historical and modern contexts (Steckel, 1983; Diamond, 1997; Obolensky et al., 2024), and our analysis below confirms that prehistoric migration was likewise shaped by climate similarity.

In addition to reconstructed migration flows, ancient DNA data allow us to observe rich variation in the spatial distribution of cultural groups over time. Groups are classified based on Ringbauer et al. (2024)'s classification by population, by homogenizing and rationalizing some groups into tighter categories. Figure A2 plots the locations of major cultural groups represented in the sample. Each panel corresponds to a distinct group, together with approximate date, including early farming societies, steppe-associated populations, and later Bronze and Iron Age groups. The figure highlights both continuity and change in the geographic footprint of cultural groups, with some remaining spatially concentrated over long periods and others spreading rapidly across large parts of Western Eurasia.

Several broad patterns are apparent. Early farming groups are initially clustered in relatively narrow regions, while later steppe-associated groups exhibit much wider geographic dispersion. Over time, cultural groups increasingly overlap in space, reflecting repeated episodes of migration, interaction, and admixture. These patterns are in line with exist-

[5]For visualization purposes, we define two individuals as genetically related only if they share at least 16 centimorgans (cM) of identity-by-descent segments. In the main analysis, we use a lower threshold of 8 cM—the minimum length at which the method in Ringbauer et al. (2024) reliably identifies IBD; as discussed below, our results are robust to alternative IBD length cutoffs. The higher threshold used in this figure serves solely to reduce visual clutter.

ing archaeological and genetic accounts emphasizing large-scale population movements and cultural diffusion in prehistoric Eurasia (Lazaridis et al., 2014; Allentoft et al., 2015), and underscore that prehistoric mobility was closely intertwined with the expansion and interaction of distinct cultural groups. Figure 2 builds on this information by plotting migration flows separately for each major cultural group. The figure reveals substantial heterogeneity in mobility across cultural contexts, both in the scale of movement and in its spatial orientation.

A particularly striking example is provided by the Yamnaya and related early steppe groups, originating in the Pontic–Caspian steppe during the late Neolithic and early Bronze Age, and widely regarded as central to the spread of Indo-European languages (Reich, 2018). Migration flows associated with the Yamnaya exhibit a pronounced east–west orientation, spanning large distances across the Eurasian steppe belt. This horizontal pattern is consistent with archaeological and genetic evidence documenting long-distance movements along ecologically similar steppe environments (Anthony, 2007; Haak et al., 2015) and with recent ancient DNA evidence showing extensive east–west genealogical connectivity across the Eurasian steppe (Ringbauer et al., 2024). More broadly, the figure highlights that the spatial structure of migration differs across cultural groups. Farming societies tend to display more localized movements, while steppe-associated groups exhibit wider-ranging and more directional mobility.

Figure 3 documents a striking temporal pattern in the geographic scale of migration. The upper panel of the figure plots the average distance between genetically related individuals over time, computed at 250-year intervals from 6000 BCE to the present. Prior to approximately 3250 BCE, the average distance between related individuals remained relatively stable, fluctuating between 500 and 900 kilometers. After this date, migration distances increased dramatically, peaking at over 2,000 kilometers around 2500 BCE before gradually declining toward approximately 1,300 kilometers by the Roman period. The vertical dashed line marks the approximate date of wheel and wagon adoption in the Pontic-Caspian steppe region (Anthony, 2007). The timing of the increase in migration distance closely coincides with this technological innovation, suggesting that wheeled transport substantially expanded the geographic reach of population movements.

The lower panel of Figure 3 plots the number of observed remains in the Ringbauer et al. (2024) dataset over the same period. While the number of remains increases around 3000 BCE, it subsequently declines toward pre-3000 BCE levels by roughly 1500 BCE. In contrast, average migration distances remain persistently elevated well beyond this point. This divergence indicates that the increase in migration distance is unlikely to be driven mechanically by changes in sample size or data coverage. Instead, it points to a genuine

expansion in the scale of mobility, consistent with technological changes that reduced the cost of long-distance travel.

4 Climate Matching and Migration

4.1 Empirical Strategy

Our empirical analysis builds on the origin–destination migration panel constructed from ancient DNA data, as described in Section 2.1. The unit of observation is a pair of locations in a given (500-year) period. For each period, we consider the set of grid cells that contain at least one observed individual in that period or in the immediately subsequent period. Restricting attention to this set ensures that origins and destinations correspond to locations that were plausibly occupied and connected at that point in time, while excluding regions for which migration flows are not meaningfully defined. We then construct all origin–destination pairs among these cells, yielding a panel of bilateral migration flows that includes both positive flows and imputed zeros for pairs of locations between which no genealogical links are observed.

To study how environmental similarity shapes migration, we estimate gravity-style models of migration, building on the trade literature (Anderson and Van Wincoop, 2003). In this framework, bilateral migration flows depend on origin-specific and destination-specific factors that vary over time, as well as on bilateral frictions that are constant within a period. Following Obolensky et al. (2024), we augment the standard gravity specification with measures of climate distance between origin and destination, alongside physical distance.

Because many origin-destination pairs record zero flows, we estimate all regressions using Poisson Pseudo-Maximum Likelihood (PPML). Formally, for origin o, destination d, and period t, we estimate specifications of the form:

$$M_{odt} = \exp[\alpha_o + \gamma_d + \delta_t + \beta \text{Dist}_{od}^{\text{Climate}} + \theta \text{Dist}_{od}^{\text{Physical}}]\epsilon_{odt} \tag{1}$$

M_{odt} is a measure of migration intensity between the origin and the destination in a given period. We measure genetic connectedness using identity-by-descent (IBD) segments—stretches of DNA shared by two individuals due to common ancestry. Longer segments indicate more recent shared ancestors: an 8 cM segment corresponds to a common ancestor approximately 10–15 generations ago, a 12 cM segment to approximately 8–12 generations ago, and a 16 cM segment to approximately 6–9 generations ago.[6] In our baseline specifi-

[6] Ringbauer et al. (2024) identify 8 cM as the minimum segment length at which the ancIBD method reliably detects IBD in ancient DNA.

cation, M_{odt} is the count of individual pairs sharing at least one IBD segment of 8 cM or longer—each genetically linked pair represents one migration event connecting the two locations. We show robustness to alternative measures, including total length of all qualifying segments (which captures both the number of connections and their intensity), the count of qualifying segments (allowing close relatives sharing multiple segments to contribute more than one), and restriction to longer segment thresholds of 12 and 16 cM.

$\text{Dist}_{od}^{Climate}$ is the absolute difference in climatic conditions between the two locations. We consider three main dimensions: temperature, precipitation, and soil. for temperature, our preferred definition is the distance in cumulative growing-degree days (GDD) above 10°C, which summarizes biologically relevant heat accumulation for plant growth and animal productivity. Degree days aggregate daily temperatures exceeding a meaningful physiological threshold, so that warmer days contribute more while colder days contribute little or nothing. As a result, this measure captures both the intensity and duration of warmth during the growing season and is closely aligned with subsistence constraints emphasized in anthropology and archaeology, as well as with paleoclimate evidence showing that growing-season heat accumulation can evolve independently of seasonal or mean temperatures over the Holocene (Mourshed, 2012; Marsicek et al., 2018).[7] We additionally control for winter temperature distance to account for climatic differences outside the growing season that may affect survival, storage, and pastoral practices, and to ensure that the estimated effect of growing-season heat accumulation is not confounded by cold-season exposure.

We measure precipitation using average annual rainfall (in millimeters per month), which captures broad differences in water availability relevant for agriculture, pasture productivity, and settlement viability. To account for soil conditions, we construct a composite index summarizing key characteristics that shape agricultural suitability—organic carbon content, bulk density, pH, clay content, and water-holding capacity. As with temperature, we include absolute differences in precipitation and soil characteristics between origin and destination in order to control for additional dimensions of environmental similarity that may independently influence migration costs and the transferability of subsistence practices.

To address concerns related to the spatial correlation of climate and migration, we flexibly control for geographic distance by including both physical distance between o and d and its square ($\text{Dist}_{od}^{Physical}$). The baseline specification also includes origin fixed effects (α_o), destination fixed effects (γ_d), and 500-year period fixed effects (δ_t). Origin and destination fixed effects absorb persistent differences across locations in population size, mobility propensity, and data coverage, while period fixed effects capture aggregate changes in migration intensity over time. Identification therefore comes from variation in migration flows

[7] We also show robustness to alternative climate measures, such as growing-season temperature.

across destination locations that differ in climatic similarity, holding constant origin-specific, destination-specific, and period-specific factors. Standard errors are clustered at the origin–destination level.

A remaining concern is that migrants from places that were climatically dissimilar moved to areas that were attracting migrants from everywhere else—regardless of climate distance. For example, over the multi-century horizons relevant in our setting, individuals may have moved toward emerging population centers, major freshwater basins, or regions with generally milder conditions that attracted migrants from many origins. A further issue is potential sample selection: the preservation and discovery of burial sites may itself be correlated with local climatic conditions. The inclusion of origin and destination fixed effects mitigates these concerns by absorbing all time-invariant characteristics of ancestral and descendant locations. However, we cannot rule out the possibility that, even after conditioning on geographic distance and destination fixed effects, climate distance is correlated with other bilateral factors that independently influence migration. We therefore further assess robustness by examining the sensitivity of our estimates to the inclusion of additional cell-pair controls and to more demanding fixed-effects structures, including origin-by-period and destination-by-period fixed effects.

4.2 Main Results

Table 3 presents results from equation (1). Column 1 reports a parsimonious specification that includes only distance in cumulative growing-degree days (GDD) above 10°C and winter temperature distance. Both coefficients are negative and statistically significant, indicating that migration flows decline as differences in both growing-season heat and winter temperatures between origin and destination increase.

Column 2 augments this specification by adding precipitation distance. The coefficients on GDD and winter temperature distance remain negative, of similar magnitude, and precisely estimated, suggesting that the relationship between thermal similarity and migration is not driven by other coarse dimensions of climate. Precipitation distance also enters with a negative and highly significant coefficient, indicating that migrants were less likely to move between locations with dissimilar rainfall regimes.

Column 3 further adds soil distance. The coefficient on GDD attenuates modestly in absolute value but remains large and precisely estimated; the coefficients on winter temperature and precipitation distance are similarly stable. Soil distance itself enters negatively and with high statistical significance, suggesting that edaphic similarity represents an additional and independent margin of ecological matching.

Column 4 further controls for geographic distance by including both distance and distance squared. As expected, the coefficient on GDD distance attenuates—from -0.872 to -0.693—but remains precisely estimated, indicating that climatic similarity matters over and above simple spatial proximity. In contrast, the coefficient on winter temperature distance declines sharply in absolute value and becomes statistically insignificant, suggesting that once growing-season heat accumulation and geographic distance are accounted for, winter temperatures play a limited independent role in shaping migration flows. The coefficients on precipitation and soil distance become somewhat smaller in magnitude but remain precisely estimated.

Columns 5 through 7 progressively introduce era fixed effects, destination fixed effects, and origin fixed effects. We treat column 7 as our preferred specification Relative to specifications without fixed effects, the coefficient on GDD distance is substantially smaller in magnitude (-0.185), reflecting the absorption of persistent differences across locations in population size, mobility propensity, and sampling intensity. Despite this stringent fixed-effects structure, though, the estimate remains highly precise. The coefficients on precipitation and soil distance are stable and remain statistically significant, indicating that ecological matching along these dimensions operates even within origin–destination comparisons net of all location-specific factors.

In the preferred specification, a difference of 1,000 GDD above 10°C—approximately the sample mean—is associated with a reduction in migration of about 17%. As a benchmark, a difference of 1,000 GDD is equivalent to sustaining a 5°C higher mean daily temperature over a 200-day growing season. From an agronomic perspective, this represents a major change in subsistence possibilities, comparable to the climatic gap between cool-temperate cropping systems such as wheat and warmer temperate systems such as maize, which require on the order of 2,500 GDD.

Precipitation distance also has economically meaningful effects. In the preferred specification, a one–standard deviation increase in precipitation distance—approximately 29 millimeters of average monthly rainfall—is associated with a reduction in migration flows of about 14%. This difference corresponds roughly to the contrast between semi-arid steppe environments and moderately humid temperate regions, or between Mediterranean climates with pronounced summer dryness and more evenly distributed rainfall regimes. Such differences affect water availability for crops and pasture, the reliability of surface water sources, and the seasonal timing of agricultural and pastoral activities.

Beyond climate, soil characteristics represent another dimension of ecological matching that consistently predicts migration flows. The coefficient on soil index distance is negative and highly significant, with a magnitude (-0.262) comparable to that of GDD and precip-

itation distance. This pattern is consistent with the importance of edaphic knowledge for prehistoric subsistence. Early farmers developed detailed, location-specific understanding of soil conditions—such as which soils retained moisture, which drained well, and which required amendment with ash or manure—and this knowledge was difficult to transfer to unfamiliar substrates. Soil pH, organic content, and water-holding capacity vary substantially across Western Eurasia and shape which crops can be cultivated and what yields can be expected. Unlike climate, soil characteristics are highly stable over Holocene timescales, so modern soil data reliably capture the conditions these populations faced. The persistent significance of soil matching, even after controlling for climate and geographic distance, suggests that ecological familiarity extended beyond atmospheric conditions to the substrate itself.

4.3 Robustness Checks

In this section, we assess the robustness of the baseline results by varying both the measurement of climate similarity and the stringency of the fixed-effects structure. Columns 6 and 7 of Table 3 examine sensitivity to alternative climate measures. In column 6, we replace growing-degree-day distance with average summer temperature distance. The estimated coefficient remains negative and statistically significant, indicating that the results are not driven by the specific construction of degree days and continue to point to the importance of growing-season conditions more broadly. In column 7, we further augment the baseline specification by controlling for differences in diurnal temperature range, capturing variation in temperature volatility and extremes; the coefficient on growing-degree-day distance is largely unchanged.

Column 8 introduces origin-by-period and destination-by-period fixed effects, allowing push and pull factors at each location to vary flexibly over time. With these fixed effects, identification comes from within-period differences across destinations that face the same origin, and symmetrically, from within-period differences across origins that face the same destination. In other words, the estimates compare migration from a given origin in a given period toward destinations that differ in climatic similarity, net of any time-varying changes in the overall attractiveness of origins or destinations. Even under this demanding specification, the estimated effect of climate distance remains similar in magnitude.

We further assess robustness along several dimensions. First, Table A3 confirms that results are consistent across all alternative measures of migration intensity described above; Table A7 presents analogous tests at the individual level, varying both the centiMorgan threshold (8cM vs. 12cM) and the measure of genetic flow (segment count vs. total length).

Second, we replicate the main results using individual-level data in Table A1, leveraging variation across pairs of individuals rather than aggregated grid cells. The consistency of results across levels of aggregation suggests that the patterns are not artifacts of spatial binning. Third, Tables A4 and A6 stratify the sample by distance bands (100–500 km, 500–1000 km, etc.) using gridded and individual-level flows, respectively, confirming that climate matching operates across a range of geographic scales and levels of aggregation rather than being driven by particular distance ranges. Fourth, Table A5 decomposes the soil index into its constituent components (pH, water-holding capacity, organic content, clay content, and bulk density), showing that organic content and clay content are the primary drivers of the soil matching effect.

4.4 Mechanisms

Land use. If climate matching reflects the specificity of subsistence technologies, we would expect the effect to vary systematically across populations with different modes of production. Farming societies depend heavily on accumulated thermal time for crop growth, making climate-specific knowledge particularly valuable. Pastoralists, while still reliant on seasonal patterns for grazing and animal husbandry, maintain greater mobility and flexibility. Foragers, by contrast, rely on broader ecological knowledge that may transfer more easily across climatic gradients.

We test this prediction by interacting climate distance with indicators of land use at the origin location. Using historical data on cropland and grazing intensity from the HYDE database (Klein Goldewijk et al., 2017), we classify origin locations as predominantly forager, pastoral, or farming based on the relative intensity of cropland versus grazing. Table 6 presents results from both interaction specifications and subsample analyses.

The results reveal striking heterogeneity across subsistence types. Column 1 reports the baseline specification, while columns 2 and 3 add interactions between climate distance and land use type. The GDD $\times$ Farmer interaction is large and negative (-0.73 to -0.85), indicating that farmers show substantially stronger climate matching than the forager reference group. The GDD $\times$ Pastoral interaction is also negative but smaller in magnitude (-0.44 to -0.55). Columns 4–6 present subsample analyses: foragers show no significant response to GDD distance (coefficient: -0.03), pastoralists show a small and statistically insignificant effect (-0.04), while farmers exhibit strong and precisely estimated climate matching (-0.24).

These patterns are consistent with the interpretation that climate-specific human capital was most binding for agricultural populations. Farmers possessed detailed knowledge of

local growing seasons, planting times, and crop varieties that was costly to redeploy in climatically unfamiliar environments. Pastoralists, while still affected by seasonal patterns, maintained greater flexibility through mobility and the adaptability of livestock. Foragers, whose subsistence drew on diverse ecological resources, appear to have been least constrained by climate similarity in their migration decisions. Table A8 in the Appendix replicates this analysis using individual-level IBD segment counts, with qualitatively similar results for both GDD and precipitation interactions.

Precipitation and water availability. The results for precipitation distance reveal a complementary pattern that further illuminates the ecological foundations of climate matching. While growing-degree-day matching is strongest among farmers, precipitation matching shows the *opposite* heterogeneity: foragers and pastoralists exhibit strong and significant responses to precipitation distance (-0.009 and -0.007, respectively), while farmers show no significant effect (column 3 of Table 6). This pattern is consistent with the distinct water dependencies of different subsistence systems. Foragers relied on predictable water sources that concentrated game and supported plant resources; pastoralists required reliable water and pasture for livestock. Both groups possessed detailed knowledge of seasonal water availability—knowledge that transferred poorly to unfamiliar precipitation regimes. Farmers, by contrast, may have adapted to variable rainfall through irrigation, water storage, or drought-tolerant crop selection, reducing their sensitivity to precipitation matching in migration decisions.

This interpretation aligns with archaeological evidence on the devastating role of drought in prehistoric societies. The 4.2 ka megadrought (c. 2200 BCE) triggered widespread societal disruptions across the Mediterranean and West Asia, including the collapse of the Akkadian Empire and Egypt's Old Kingdom (Weiss, 2017). Paleoclimate evidence documents a 30–50% reduction in precipitation delivered by Mediterranean westerlies during this period, with cascading effects on settlement patterns and population movements. The later Bronze Age collapse (c. 1200 BCE) has similarly been linked to prolonged drought conditions in the Eastern Mediterranean (Kaniewski et al., 2013). More broadly, the persistent and stable effect of precipitation distance in our results—together with its concentration among foragers and pastoralists—suggests that water availability was a distinct axis of ecological constraint, complementing the thermal constraints captured by growing-degree days.

Ecological diffusion: Pollen evidence. If migration facilitated the spread of subsistence practices, we would expect genetic flow to predict subsequent changes in vegetation at destination locations. We test this directly using palynological data from sediment cores

across Western Eurasia. Specifically, we measure pollen abundance for cereals and grassland taxa at origin and destination grid cells, interpolating from nearby pollen cores using inverse-distance weighting (restricting to cells within 500 km of a core to ensure reliable interpolation). We then regress the pollen distance between origin (at time t) and destination (at time $t + 500$ years) on genetic flow, controlling for climate distance, geographic distance, and origin, destination, and era fixed effects.

Table 7 presents the results. Columns 1–2 examine cereals pollen, while columns 3–4 examine grassland. In both cases, genetic flow is associated with significantly smaller pollen distance at $t + 500$, indicating that destinations receiving more migrants from a given origin came to resemble that origin's vegetation profile over the subsequent 500 years. For cereals, the coefficient is -0.0013 (SE $= 0.0004$), attenuating slightly to -0.0010 when controlling for local pollen change at the destination. For grassland, the effect is substantially larger (-0.037, SE $= 0.011$) and similarly robust to controls. These patterns are consistent with migration driving the diffusion of agricultural practices: as populations moved, they brought crops and livestock management techniques that reshaped local vegetation. The stronger effect for grassland may reflect the spread of pastoral systems, which transform landscapes through grazing pressure and the introduction of fodder species.

Warming vs cooling. A key advantage of our setting is that it spans nearly 10,000 years, a period marked by substantial and persistent climate change over the Holocene. Paleoclimate reconstructions document multi-century episodes of both warming and cooling, with pronounced regional heterogeneity in timing and magnitude (Marcott et al., 2013). As a result, migration decisions unfolded under very different aggregate climatic regimes, ranging from periods in which growing-season conditions became progressively more favorable to periods in which ecological constraints tightened. This long-run temporal variation allows us to move beyond average effects, and examine whether the role of climate similarity in shaping migration differed systematically across climatic regimes.

We exploit this variation by estimating the baseline specification separately for each 1000-year period, allowing the GDD coefficient to vary over time (Table A10; Table A11 replicates this analysis at the individual level). The results reveal a notable temporal pattern: climate matching emerges around 3500 BCE and peaks around 2500 BCE—a period that coincides with the widespread adoption of agriculture across much of Western Eurasia. Earlier periods (pre-3500 BCE) show weaker or even positive GDD coefficients, suggesting that climate matching was less important during the predominantly foraging era. This pattern is consistent with the idea that warming relaxes growing-season constraints by expanding the set of ecologically viable destinations, thereby reducing the penalty associated

with climate mismatch (Richerson et al., 2001; Dow et al., 2009).

5 Adaptation: Migration in Response to Long-Run Climate Change

As noted above, the period we study was characterized by distinct episodes of warming and cooling over the Holocene. These long-run changes imply that populations were repeatedly exposed to shifts in local climatic conditions rather than to a stable environment. We now examine whether migration operated as an adaptive response to climate change. Specifically, we test whether individuals responded to changes in the climate of their ancestral locations by relocating toward destinations whose climatic conditions partially offset those changes.

We return to the individual level dataset (so we move away from the cell-pair dataset). For each ancestor–descendant pair, we first compute the difference between the paleo-climate reconstructed temperature at the ancestor's location evaluated at the time of the descendant and the temperature at the same location at the time of the ancestor. Positive values indicate periods of local warming, while negative values indicate cooling. We then measure temperature change over space induced by migration by comparing the climatic conditions at the descendant's location to those at the ancestral location, using high-resolution modern climate data. This spatial temperature difference captures whether migration moved individuals toward warmer or cooler environments relative to their place of origin.

Given the granularity over time of our data, we can test whether individuals responded to multi-century periods of warming or cooling with offsetting migration. In other words, did people move to marginally cooler places when the overall climate was warming?

Temperature Change over Time For each pair of individuals (ancestor and descendant), we use Paleo-climate data reconstructions (Osman et al., 2021) to calculate the temperature change over time, which we define as the difference between the temperature at time of the descendant at the ancestor's location (o), and the temperature at the ancestor's location at the time of the ancestor. Formally, this quantity is defined as:

$$\text{Temp Change Over Time}_o = Temp_o^{Descendant} - Temp_o^{Ancestor} \tag{2}$$

Thus, a positive value denotes a period of warming (i.e., the temperature has increased in a given location o).

Temperature Change over Space For each pair of individuals, we define the temperature change over space as the temperature difference between the ancestral location cell (o) and the descendant location cell (d), relying on spatially-granular modern TerraClimate data (Abatzoglou et al., 2018).

$$\text{Temp Change Over Space}_{od} = Temp_d^{TerraClimate} - Temp_o^{TerraClimate} \tag{3}$$

We then regress temperature change over time on variations in temperature, precipitation and elevation across space:

$$\text{Temp Change Over Space}_{od} = \beta \text{Temp Change Over Time}_o + X_{od} + \varepsilon_{od} \tag{4}$$

We report results in Table 4. In Column 1, we regress *Temp Change over Time* on *Temp Change Over Space*. In Column 3, we add ancestor and descendant location cell, and time fixed effects. In Column 5, we add as regressors *Precipitation Change over Space*, on top of a quadratic in distance. Column 2, 4 and 6 replicate – respectively – Columns 1, 3, and 5 by weighting observations by the length of segments greater than 8 cM. In all specifications, we cluster standard errors at the ancestor by descendant location cell by 500-year period level.

In each case, we find a negative effect of temperature change over time on the temperature gradient of moves. In other words, our findings indicate that people move to cooler places during periods of warming. This suggests a potential adaptive response to climate change via migration, whereby individuals offset changes in their own "ancestral" temperature over time by moving towards areas that are climatically more similar to their initial climate.

[STOP HERE – WORK IN PROGRESS]

References

Abatzoglou, John T, Solomon Z Dobrowski, Sean A Parks, and Katherine C Hegewisch (2018) "TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015," *Scientific data*, 5 (1), 1–12.

Allentoft, Morten E, Martin Sikora, Karl-Göran Sjögren et al. (2015) "Population genomics of bronze age Eurasia," *Nature*, 522 (7555), 167–172.

Anderson, James E and Eric Van Wincoop (2003) "Gravity with gravitas: A solution to the border puzzle," *American economic review*, 93 (1), 170–192.

Anthony, David W. (2007) *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*, Princeton, NJ: Princeton University Press.

Ashraf, Quamrul and Oded Galor (2013) "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103 (1), 1—-46.

Ashraf, Quamrul and Stelios Michalopoulos (2015) "Climatic Fluctuations and the Diffusion of Agriculture," *The Review of Economics and Statistics*, 97 (3), 589–609.

Boyd, Robert, Peter J Richerson, and Joseph Henrich (2011) "The cultural niche: Why social learning is essential for human adaptation," *Proceedings of the National Academy of Sciences*, 108 (supplement_2), 10918–10925.

Butler, Ethan E. and Peter Huybers (2013) "Adaptation of US maize to temperature variations," *Nature Climate Change*, 3 (1), 68–72, 10.1038/nclimate1668.

Castiello, Maria Elena, Emmanuele Russo, Héctor Martínez-Grau, Ana Jesus, Georgina Prats, and Ferran Antolín (2025) "Understanding the Spread of Agriculture in the Western Mediterranean (6th–3rd Millennia BC) with Machine Learning Tools," *Nature Communications*, 16, 678, 10.1038/s41467-024-55541-y.

Conley, Timothy G (1999) "GMM estimation with cross sectional dependence," *Journal of econometrics*, 92 (1), 1–45.

Crosby, Alfred W (1986) *Ecological Imperialism: The Biological Expansion of Europe, 900-1900*: Cambridge University Press.

Diamond, Jared (1997) *Guns, Germs, and Steel: The Fates of Human Societies*: W. W. Norton.

Dow, Gregory K., Clyde G. Reed, and Nancy Olewiler (2009) "Climate Reversals and the Transition to Agriculture," *Journal of Economic Growth*, 14 (1), 27–53, 10.1007/s10887-009-9038-x.

Fenske, James (2014) "Ecology, trade, and states in pre-colonial Africa," *Journal of the European Economic Association*, 12 (3), 612–640.

Galor, Oded and Omer Moav (2002) "Natural Selection and the Origin of Economic Growth," *The Quarterly Journal of Economics*, 117 (4), 1133–1191, 10.1162/003355302320935007.

Galor, Oded and Ömer Özak (2016) "The Agricultural Origins of Time Preference," *American Economic Review*, 106 (10), 3064–3103, 10.1257/aer.20150020.

Giuliano, Paola and Nathan Nunn (2021) "Understanding cultural persistence and change," *The Review of Economic Studies*, 88 (4), 1541–1581.

Guedes, Jade D'Alpoim and Ethan E. Butler (2014) "Modeling Constraints on the Spread of Agriculture to Southwest China with Thermal Niche Models," *Quaternary International*, 349, 29–41, 10.1016/j.quaint.2014.08.003.

Haak, Wolfgang, Iosif Lazaridis, Nick Patterson et al. (2015) "Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe," *Nature*, 522 (7555), 207–211, 10.1038/nature14317.

Hengl, Tomislav (2018a) "Soil bulk density (fine earth) 10 x kg / m-cubic at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.2)," 10.5281/zenodo.2525665.

——— (2018b) "Soil pH in H2O at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.2)," 10.5281/zenodo.2525664.

Hengl, Tomislav and Surya Gupta (2019) "Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.1)," 10.5281/zenodo.2784001.

Hengl, Tomislav and Ichsani Wheeler (2018) "Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.2)," 10.5281/zenodo.2525553.

Henrich, Joseph (2004) "Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses—the Tasmanian case," *American antiquity*, 69 (2), 197–214.

Hersbach, Hans, Bill Bell, Paul Berrisford et al. (2020) "The ERA5 global reanalysis," *Quarterly journal of the royal meteorological society*, 146 (730), 1999–2049.

Kaniewski, David, Elise Van Campo, Joel Guiot, Sabine Le Burel, Thierry Otto, and Cecile Baeteman (2013) "Environmental roots of the Late Bronze Age crisis," *PloS one*, 8 (8), e71004.

Klein Goldewijk, K., A. Beusen, J. Doelman, and E. Stehfest (2017) "Anthropogenic land use estimates for the Holocene – HYDE 3.2," *Earth System Science Data*, 9 (2), 927–953.

Lazaridis, Iosif, Nick Patterson, Alissa Mittnik et al. (2014) "Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans," *Nature*, 513 (7518), 409–413, 10.1038/nature13673.

Lee, Richard B. and Irven DeVore eds. (1968) *Man the Hunter*, Chicago: Aldine de Gruyter.

Marcott, Shaun A., Jeremy D. Shakun, Peter U. Clark, and Alan C. Mix (2013) "A Reconstruction of Regional and Global Temperature for the Past 11,300 Years," *Science*, 339 (6124), 1198–1201, 10.1126/science.1228026.

Marsicek, Jeremiah, Bryan N. Shuman, Patrick J. Bartlein, Sarah L. Shafer, and Simon Brewer (2018) "Reconciling Divergent Trends and Millennial Variations in Holocene Temperatures," *Nature*, 554 (7690), 92–96, 10.1038/nature25464.

Matranga, Andrea (2024) "The Ant and the Grasshopper: Seasonality and the Invention of Agriculture," *Quarterly Journal of Economics*, 139 (3), 1467–1504.

Michalopoulos, Stelios (2012) "The origins of ethnolinguistic diversity," *American Economic Review*, 102 (4), 1508–1539.

Mourshed, Monjur (2012) "Relationship between Annual Mean Temperature and Degree-Days," *Energy and Buildings*, 54, 418–425, 10.1016/j.enbuild.2012.07.024.

Nicolò, Alessandra, Marco Tabellini, and Charles A. Taylor (2026) "Climates of Conquest: Ecological Imperialism through Climate Matching," Working paper.

Nunn, Nathan (2021) "History as evolution," in *The handbook of historical economics*, 41–91: Elsevier.

Obolensky, Marguerite, Marco Tabellini, and Charles Taylor (2024) "Migration, Climate Similarity, and the Consequences of Climate Mismatch,"Technical report, National Bureau of Economic Research.

Olsson, Ola (2024) "Paleoeconomics," *Springer Books*.

Osman, Matthew B., Jessica E. Tierney, Jiang Zhu, Robert Tardif, Gregory J. Hakim, Jonathan King, and Christopher J. Poulsen (2021) "Globally resolved surface temperatures since the Last Glacial Maximum," *Nature*, 599 (7884), 239–244, 10.1038/s41586-021-03984-4.

Palamara, Pier Francesco, Todd Lencz, Ariel Darvasi, and Itsik Pe'er (2012) "Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History," *American Journal of Human Genetics*, 91 (5), 809–822, 10.1016/j.ajhg.2012.08.030.

Racimo, Fernando, Jessie Woodbridge, Ralph M. Fyfe, Martin Sikora, Karl-Göran Sjögren, Kristian Kristiansen, and Marc Vander Linden (2020) "The Spatiotemporal Spread of Human Migrations during the European Holocene," *Proceedings of the National Academy of Sciences*, 117 (16), 8989–9000, 10.1073/pnas.1920051117.

Ralph, Peter and Graham Coop (2013) "The Geography of Recent Genetic Ancestry across Europe," *PLOS Biology*, 11 (5), e1001555, 10.1371/journal.pbio.1001555.

Ramachandran, Sohini and Noah A Rosenberg (2011) "A test of the influence of continental axes of orientation on patterns of human gene flow," *American journal of physical anthropology*, 146 (4), 515–529.

Reich, David (2018) *Who we are and how we got here: Ancient DNA and the new science of the human past*: Oxford University Press.

Richerson, Peter J., Robert Boyd, and Robert L. Bettinger (2001) "Was Agriculture Impossible during the Pleistocene but Mandatory during the Holocene? A Climate Change Hypothesis," *American Antiquity*, 66 (3), 387–411, 10.2307/2694241.

Ringbauer, Harald, Yilei Huang, Ali Akbari, Swapan Mallick, Iñigo Olalde, Nick Patterson, and David Reich (2024) "Accurate detection of identity-by-descent segments in human ancient DNA," *Nat. Genet.*, 56 (1), 143–151.

Schlenker, Wolfram and Michael J. Roberts (2009) "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change," *Proceedings of the National Academy of Sciences*, 106 (37), 15594–15598, 10.1073/pnas.0906865106.

Smerdon, Jason E, Jürg Luterbacher, Steven J Phipps et al. (2017) "Comparing proxy and model estimates of hydroclimate variability and change over the Common Era," *Climate of the Past*, 13 (12), 1851–1900.

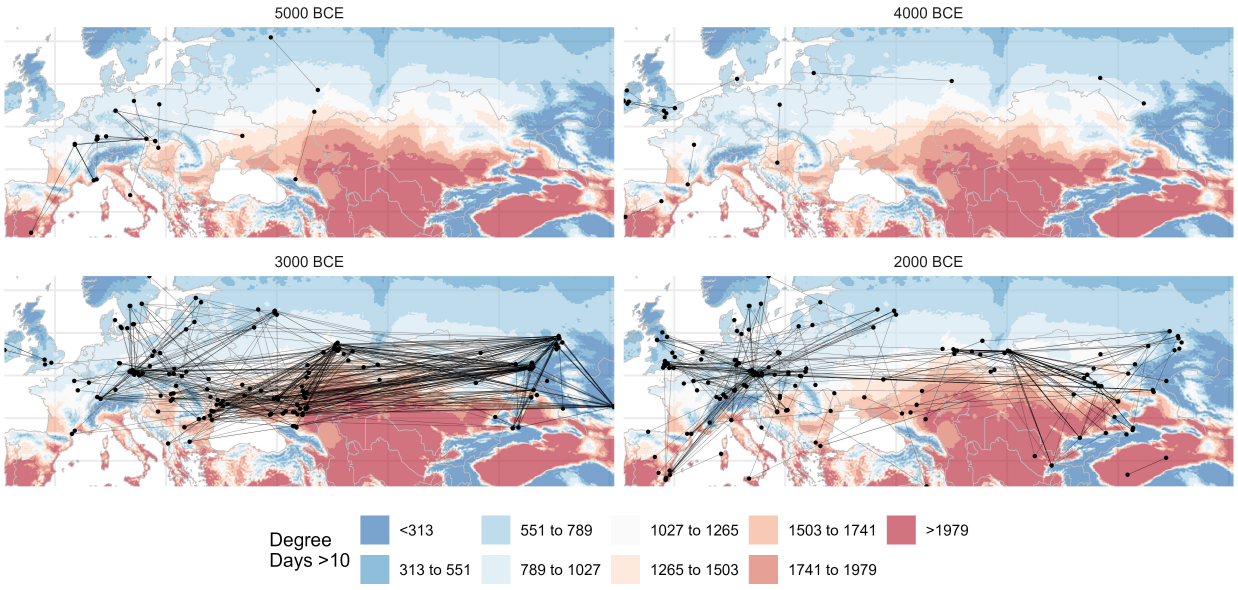
Steckel, Richard H (1983) "The economic foundations of East-West migration during the 19th century," *Explorations in Economic History*, 20 (1), 14–36.

Tierney, Jessica E., Peter B. deMenocal, and Paul D. Zander (2017) "A Climatic Context for the Out-of-Africa Migration," *Geology*, 45 (11), 1023–1026, 10.1130/G39457.1.

Weiss, Harvey (2017) "4.2 ka BP megadrought and the Akkadian collapse," *Megadrought and Collapse: From Early Agriculture to Angkor*, 93–160.

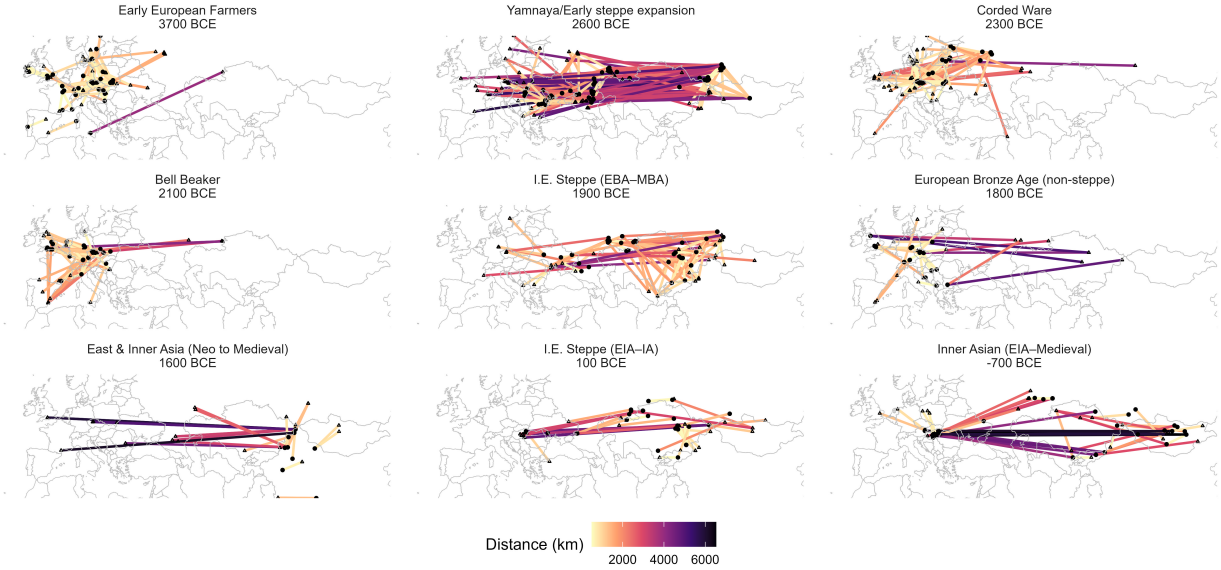
Figures

Figure 1: Flows of Individuals



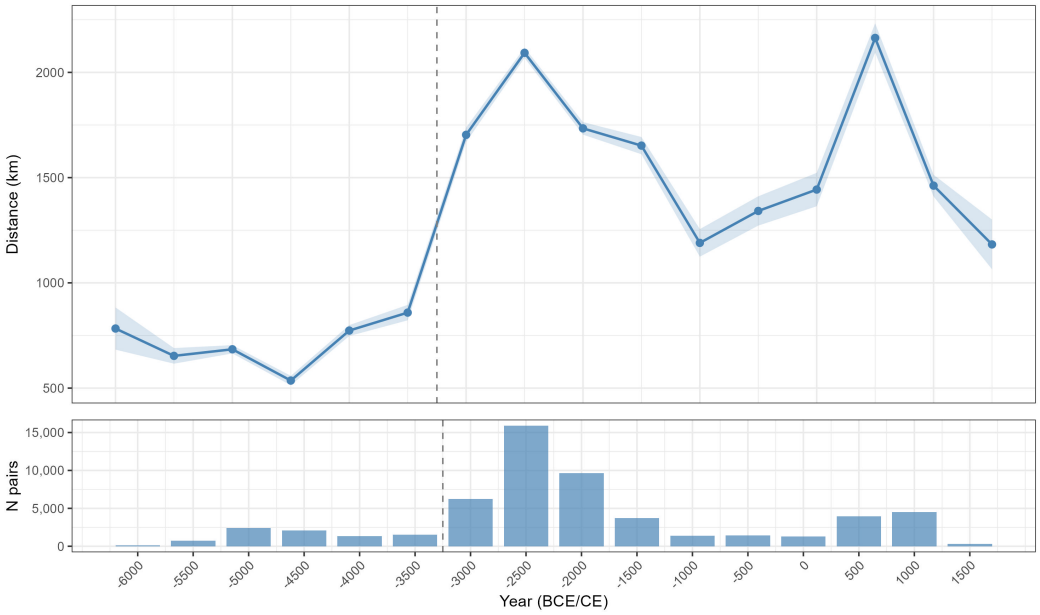
Notes: The figure plots reconstructed flows of individuals in 4 different periods.

Figure 2: Movements of Major Culture Groups



Notes: The maps show movements of individuals for major cultural groups in our sample. The sample is further restricted to include only pairs of individuals sharing at least one segment greater than 16cM, living at most 500 years apart of each other and whose (geographic) distance is of at least 500km. I.E. = Indo-European.

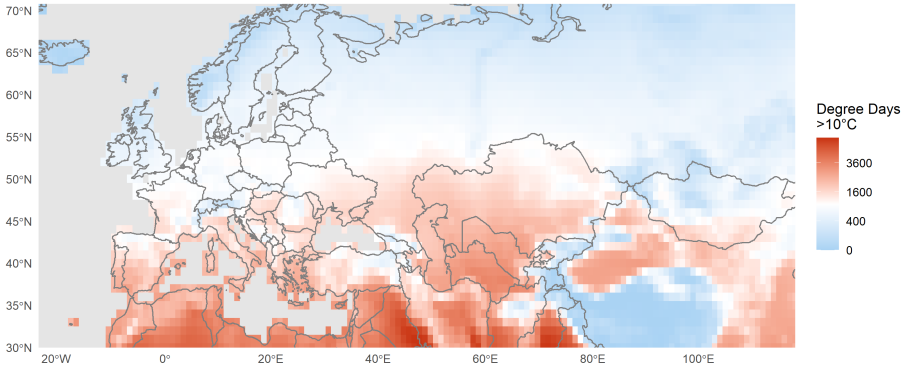
Figure 3: Average Distance Between Related Individuals Over Time



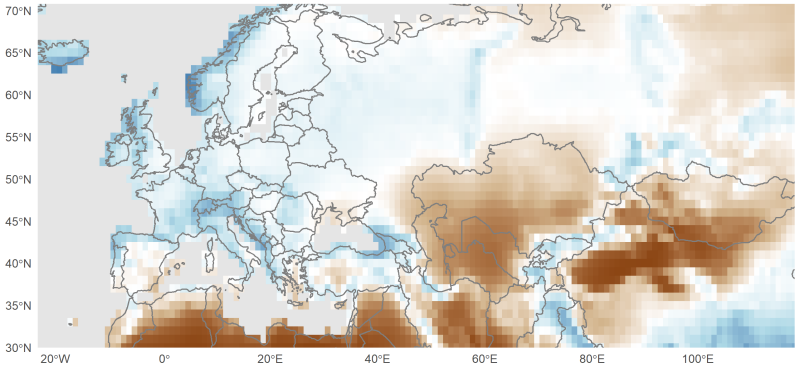
Notes: The figure plots the average geographic distance (in kilometers) between genetically related individuals over time, computed at 500-year intervals. Individuals are classified as related if they share at least one IBD segment of 8 cM or longer. The vertical dashed line marks approximately 3250 BCE, corresponding to the adoption of wheel and wagon technology in the Pontic-Caspian steppe region.

Figure 4: Modern Climate Variables

(a) Growing Degree Days above 10°C



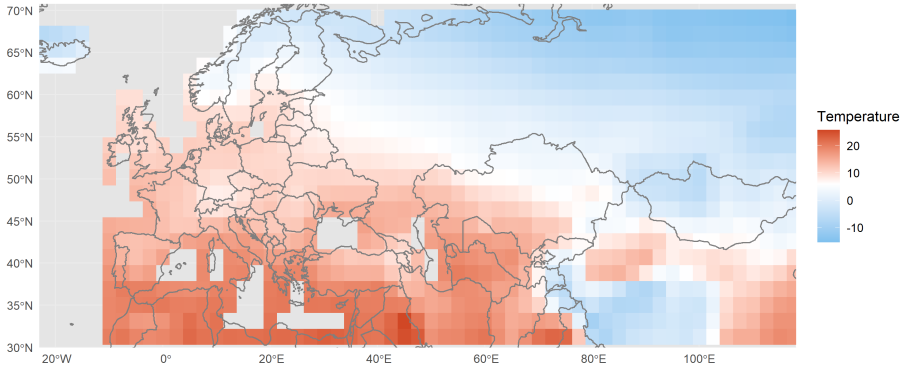
(b) Precipitation



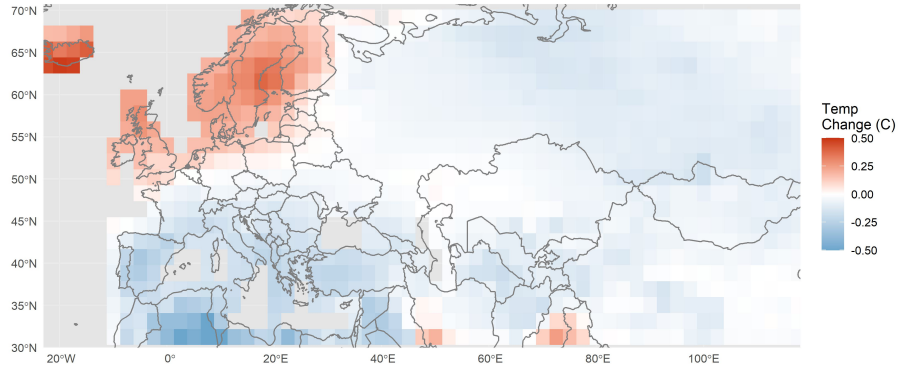
Notes: The figure plots the spatial distribution of modern climate variables at 1° resolution. Panel A shows cumulative growing degree days above 10°C, derived from ERA5 reanalysis data, which captures biologically relevant heat accumulation during the growing season. Panel B shows mean monthly precipitation from ERA5 reanalysis data.

Figure 5: Paleoclimate Temperature Reconstructions

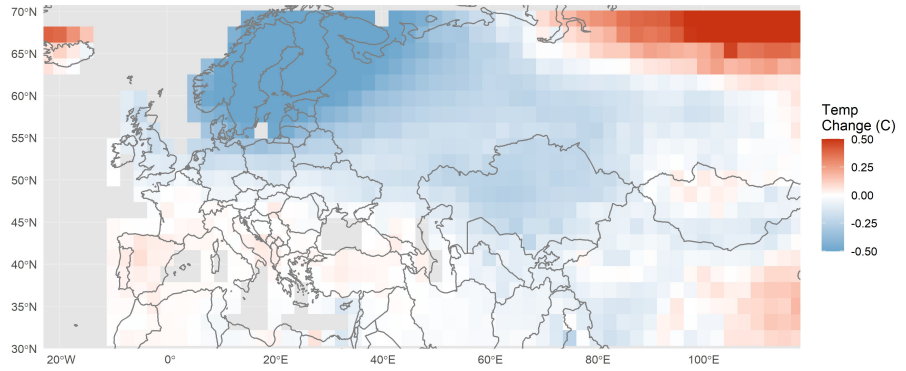
(a) Reconstructed Temperature, 3000 BCE



(b) Temperature Change: 3600 to 3000 BCE

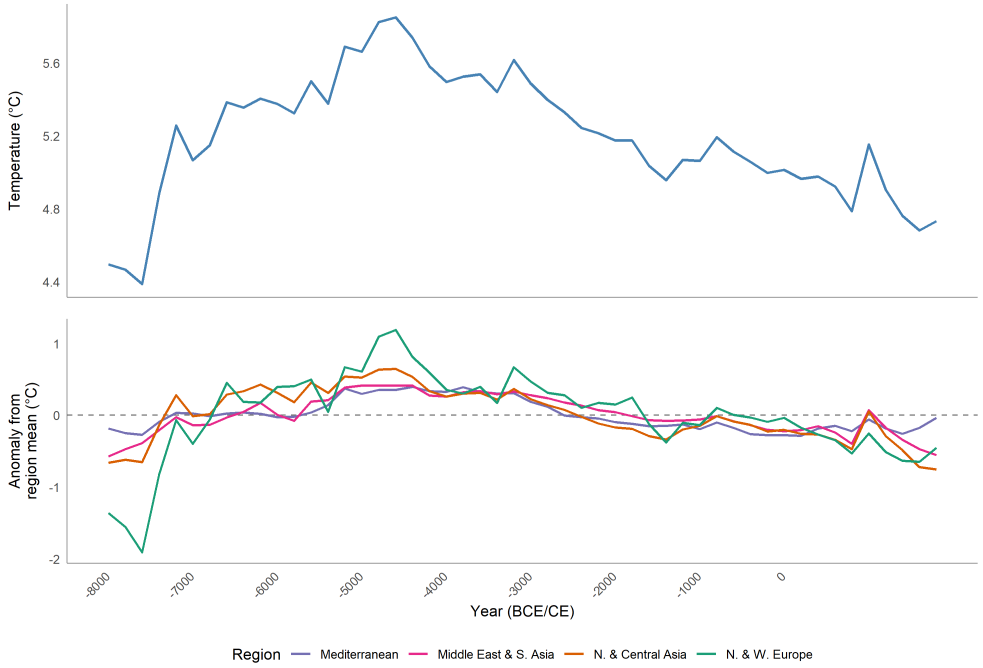


(c) Temperature Change: 1800 to 1200 BCE



Notes: Panel A shows reconstructed average temperature in 3000 BCE from the marine-based paleoclimate dataset (Osman et al., 2021) at 2.5° resolution. Panels B and C plot the change in temperature over 600-year intervals, defined as the difference between the later and earlier period.

Figure 6: Evolution of Average Temperature over Time



Notes: The figure plots the evolution of (reconstructed) temperature for our sample region of interest over time. Temperature reconstructions are from Osman et al. (2021).

Tables

Table 1: Summary Statistics

	Mean	SD	Min	P10	Median	P90	Max
IBD Count (8cM)	0.084	1.527	0.000	0.000	0.000	0.000	430.000
IBD Length (12cM)	0.039	1.508	0.000	0.000	0.000	0.000	951.922
IBD Length (8cM)	0.129	2.112	0.000	0.000	0.000	0.000	966.465
GDD Dist.	0.779	0.695	0.000	0.104	0.610	1.680	5.909
Precipitation Dist.	27.522	28.655	0.000	3.470	19.934	59.382	354.906
Distance (000 km)	3.187	2.126	0.042	0.821	2.641	6.306	12.869
Soil Index Dist.	1.123	0.956	0.000	0.137	0.823	2.633	5.404
Elevation Dist.	0.743	0.959	0.000	0.041	0.383	1.790	5.096
Summer Temp. Dist.	4.941	4.020	0.000	0.701	4.012	10.529	29.789
Winter Temp. Dist.	9.722	7.483	0.000	1.299	8.059	20.490	54.427

Notes: Unit of observation is a grid cell pair-era. IBD Count is the number of IBD segments $\geq$ 8cM sh

Table 2: Presence of Archeological Sites across Climate Areas

Dependent Variable:	1[at least one remain]							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDD	-0.091***	-0.110***	-0.100***	-0.102***	-0.108***	-0.108***	-0.137***	-0.056***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.019)
Winter Temp.	0.010***	0.011***	0.009***	0.009***	0.011***	0.011***	0.011***	-0.001
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0006)	(0.0006)	(0.005)
Soil pH		0.004***	0.005***	0.004***	0.0006	0.0005	0.002***	0.009***
		(0.0004)	(0.0004)	(0.0005)	(0.0007)	(0.0008)	(0.0008)	(0.003)
Soil Clay			0.005***	0.004***	0.006***	0.006***	0.006***	0.007***
			(0.0004)	(0.0004)	(0.0005)	(0.0006)	(0.0006)	(0.002)
Soil Organic				-0.004***	-0.007***	-0.007***	-0.006***	-0.003
				(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Precipitation					-30.5***	-30.6***	-12.1***	5.99
					(3.78)	(3.76)	(3.90)	(11.8)
Diurnal Range						0.0002	0.016***	0.015**
						(0.003)	(0.003)	(0.007)
Elevation/1000							-0.057***	-0.052*
							(0.006)	(0.030)
Observations	6,069	6,048	6,048	6,048	6,048	6,048	6,048	6,048
R^2	0.081	0.095	0.110	0.112	0.119	0.119	0.132	0.431
2.5 Grid FEs								Yes

Notes: OLS estimates. Robust standard errors in parentheses; clustered at the 2.5 grid cell level in the last column. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Climate Matching: Main Results

Dependent Variable:	Genetic Flow				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
GDD Dist.	-1.16***	-1.05***	-0.872***	-0.693***	-0.185***
	(0.054)	(0.054)	(0.054)	(0.055)	(0.065)
Winter Temp. Dist.	-0.099***	-0.093***	-0.091***	-0.007	0.004
	(0.005)	(0.005)	(0.005)	(0.004)	(0.006)
Precipitation Dist.		-0.013***	-0.011***	-0.007***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
Soil Index Dist.			-0.428***	-0.294***	-0.262***
			(0.029)	(0.029)	(0.035)
Distance (000 km)				-0.764***	-0.786***
				(0.052)	(0.048)
Distance-sq				0.058***	0.069***
				(0.006)	(0.007)
Fixed-effects					
Orig. gridcell					Yes
Dest. gridcell					Yes
Era					Yes
Fit statistics					
Observations	868,609	868,609	868,609	868,609	797,605
Pseudo R^2	0.070	0.076	0.086	0.106	0.463

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. Dependent variable is count of IBD segments $\geq$ 8cM between grid cell pairs. Column 5 includes origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Climate Adaptation (Individual)

Dependent Variable:	Temp. Diff. (space)			
Model:	(1)	(2)	(3)	(4)
Variables				
Temp. Change (400yr)	-1.45***	-1.52***	-0.332***	-0.334**
	(0.355)	(0.349)	(0.128)	(0.148)
Distance (000 km)		-0.424	-0.015	-0.005
		(0.281)	(0.025)	(0.023)
Distance-sq		-0.013	-0.002	-0.004
		(0.049)	(0.005)	(0.005)
Precip. Diff. (space)				-0.031***
				(0.003)
Fixed-effects				
Orig. gridcell	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes
Dest. gridcell			Yes	Yes
Fit statistics				
Observations	31,943	31,943	31,916	31,916
Adjusted R^2	0.451	0.468	0.950	0.954

Clustered (geo_pair_ordered_2p5) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: OLS estimates. Dependent variable is temperature difference between origin and destination. Uses 2.5-degree gridcells for the fixed effects. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Climate Matching: Sample and Outcome Robustness

Dependent Variable:	Genetic Flow			
	Baseline	≥100km	Horizontal	Vertical
Model:	(1)	(2)	(3)	(4)
Variables				
GDD Dist.	-0.136**	-0.089	-0.137*	-0.192*
	(0.063)	(0.057)	(0.073)	(0.103)
Winter Temp. Dist.	-0.002	-0.002	0.002	-0.004
	(0.006)	(0.006)	(0.006)	(0.017)
Precipitation Dist.	-0.005***	-0.004***	-0.004***	0.0009
	(0.001)	(0.001)	(0.001)	(0.003)
Distance (000 km)	-0.736***	-0.693***	-0.662***	-1.39***
	(0.049)	(0.050)	(0.056)	(0.214)
Distance-sq	0.070***	0.063***	0.057***	0.155**
	(0.007)	(0.007)	(0.008)	(0.065)
Soil Index Dist.	-0.220***	-0.191***	-0.200***	-0.188***
	(0.036)	(0.032)	(0.041)	(0.060)
Elevation Dist.	-0.127**	-0.134**	-0.123**	-0.362**
	(0.063)	(0.066)	(0.060)	(0.143)
Diurnal Range Dist.	-0.077***	-0.074***	-0.076***	0.018
	(0.016)	(0.016)	(0.017)	(0.040)
Fixed-effects				
Orig. gridcell	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes
Fit statistics				
Observations	797,605	792,659	646,885	90,594
Pseudo R^2	0.463	0.459	0.486	0.440

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Climate Matching: Land Use Heterogeneity

Dependent Variable:	Genetic Flow					
		Baseline		Foragers	Pastoral	Farmers
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-0.185***	0.322*	0.357**	0.317*	-0.157*	-0.246**
	(0.065)	(0.176)	(0.177)	(0.188)	(0.082)	(0.114)
Winter Temp. Dist.	0.004	0.037***	0.037***	-0.049***	0.032***	0.049***
	(0.006)	(0.008)	(0.008)	(0.018)	(0.009)	(0.010)
Precipitation Dist.	-0.005***	-0.006***	-0.010**	-0.011***	-0.005***	-0.003
	(0.001)	(0.002)	(0.004)	(0.003)	(0.001)	(0.003)
Soil Index Dist.	-0.262***	-0.262***	-0.260***	-0.371***	-0.221***	-0.178***
	(0.035)	(0.048)	(0.048)	(0.080)	(0.039)	(0.059)
Distance (000 km)	-0.786***	-1.09***	-1.09***	-1.07***	-0.928***	-1.34***
	(0.048)	(0.066)	(0.066)	(0.157)	(0.058)	(0.092)
Distance-sq	0.069***	0.086***	0.086***	0.064***	0.063***	0.104***
	(0.007)	(0.009)	(0.009)	(0.020)	(0.009)	(0.012)
Pastoral		-0.683***	-0.768***			
		(0.234)	(0.250)			
Farmer		-0.905***	-0.947***			
		(0.181)	(0.196)			
GDD Dist. × Pastoral		-0.500***	-0.559***			
		(0.183)	(0.180)			
GDD Dist. × Farmer		-0.741***	-0.775***			
		(0.224)	(0.219)			
Precip. Dist. × Pastoral			0.006			
			(0.004)			
Precip. Dist. × Farmer			0.004			
			(0.005)			
Fixed-effects						
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	797,605	340,809	340,809	38,191	119,028	130,806
Pseudo R^2	0.463	0.509	0.510	0.683	0.503	0.606

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. Land use types based on HYDE 3.2 cropland and grazing shares at origin. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by grid cell pair in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Ecological Diffusion

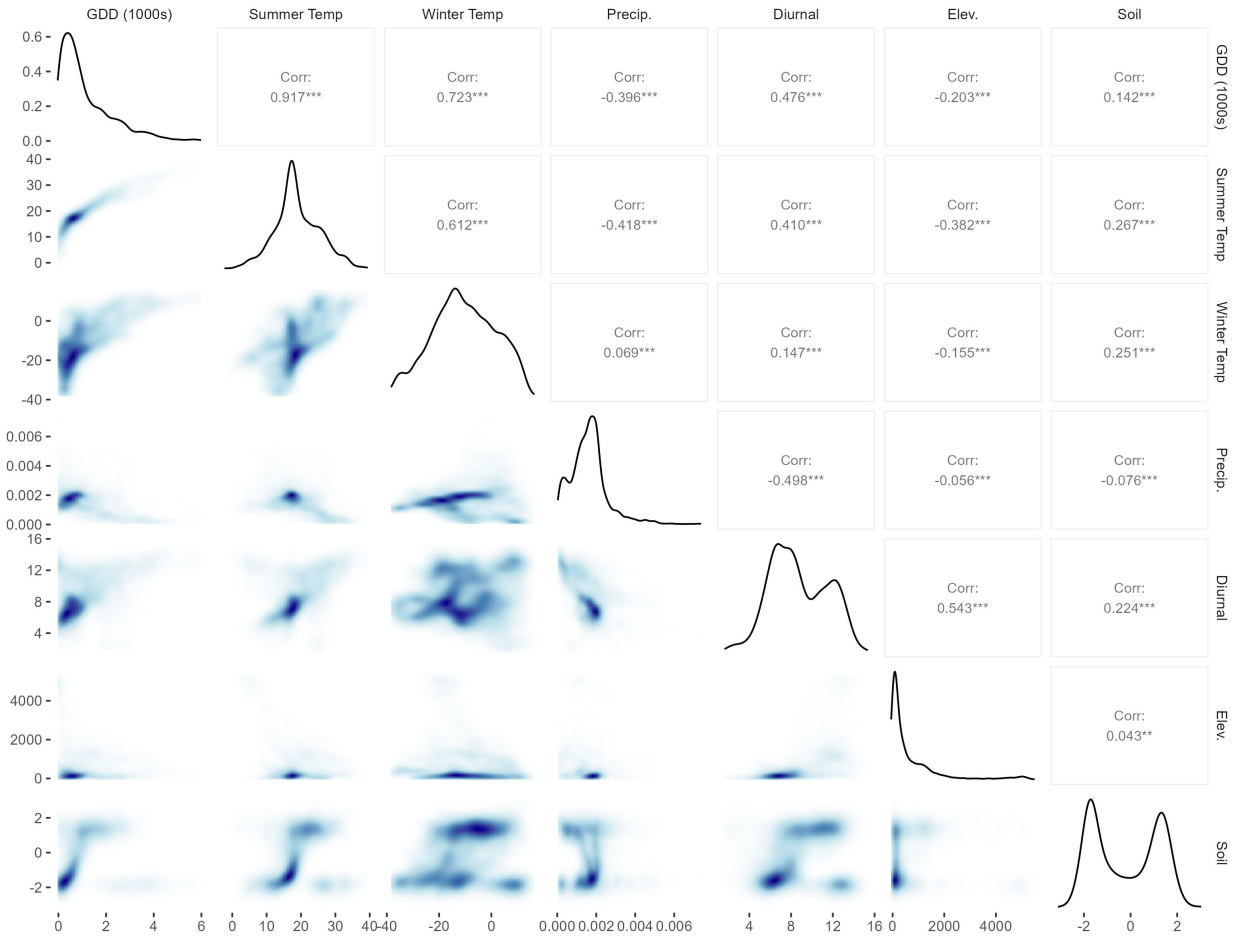
Dependent Variables:	Cereals Pollen Dist.		Grassland Pollen Dist.	
Model:	(1)	(2)	(3)	(4)
Variables				
Genetic Flow	-0.0012***	-0.0008**	-0.0375***	-0.0367***
	(0.0004)	(0.0004)	(0.0112)	(0.0110)
Winter Temp. Dist.	-0.0220***	-0.0223***	0.8306***	0.8303***
	(0.0011)	(0.0011)	(0.0120)	(0.0120)
Precipitation Dist.	0.0035***	0.0035***	-0.0051	-0.0050
	(0.0002)	(0.0002)	(0.0031)	(0.0031)
Distance (000 km)	322.4***	319.7***	7,330.5***	7,349.8***
	(10.80)	(10.84)	(166.8)	(166.9)
Distance-sq	-24,722.4***	-24,037.8***	-1,100,820.6***	-1,104,802.2***
	(2,028.6)	(2,034.9)	(36,264.5)	(36,301.0)
Soil Index Dist.	-0.0407***	-0.0411***	-1.549***	-1.556***
	(0.0056)	(0.0056)	(0.0609)	(0.0610)
Elevation Dist.	0.1986***	0.2013***	-1.614***	-1.607***
	(0.0355)	(0.0356)	(0.1862)	(0.1861)
Diurnal Range Dist.	0.0049	0.0049	0.8619***	0.8608***
	(0.0031)	(0.0031)	(0.0425)	(0.0425)
Dest. Pollen Change		0.1752***		-0.1052***
		(0.0061)		(0.0074)
Fixed-effects				
Orig. gridcell	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes
Fit statistics				
Observations	433,782	433,619	433,782	433,619
R^2	0.875	0.876	0.522	0.522

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: DV: Pollen distance between origin (at t) and destination (at t+500). Inverse distance weighted interpolation from pollen samples. Sample restricted to cells within 500km of pollen core. Dest. Pollen Change controls for local change at destination. Clustered SEs by grid pair.

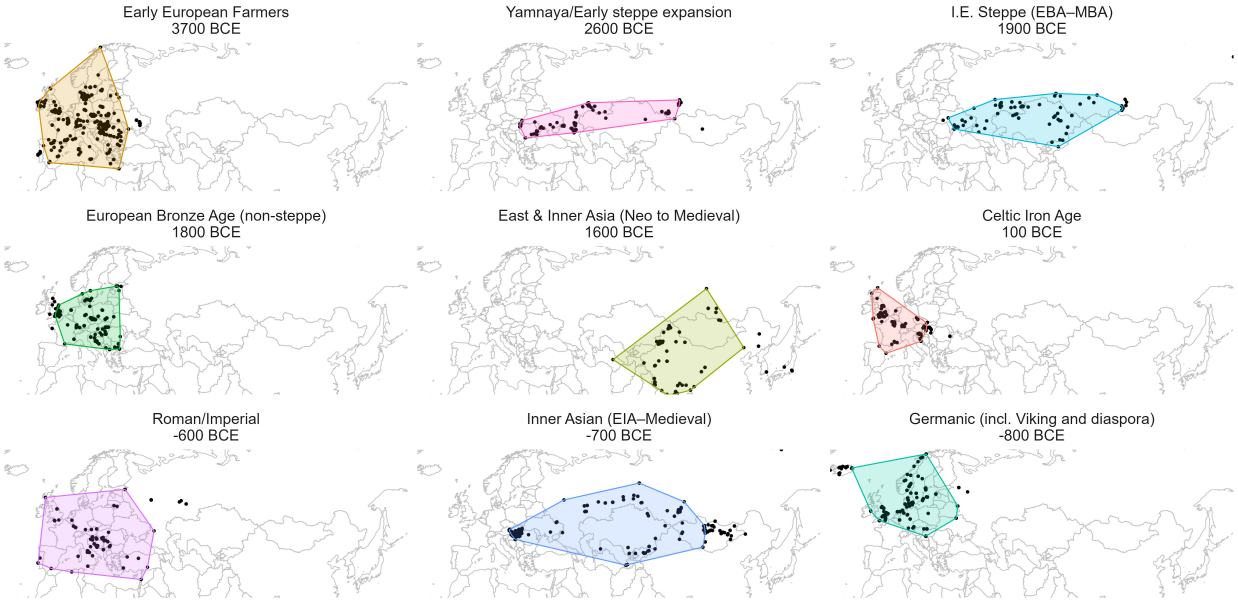
Appendix A - Additional Tables and Figures

Figure A1: Correlation matrix between climate and soil variables



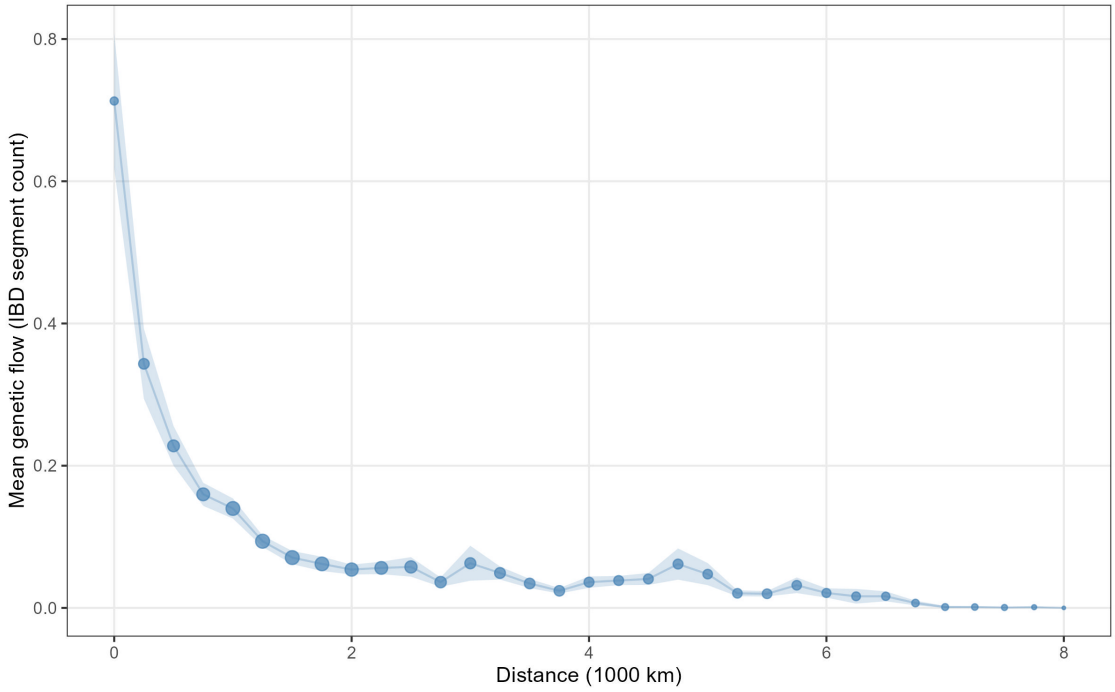
Notes: This figure plots the pairwise correlations between the main environmental variables used in our analysis, measured at 1° x 1° resolution. Variables include: cumulative growing degree days above 10°C (GDD), summer temperature (July), winter temperature (January), precipitation, diurnal temperature range, elevation, and a composite soil quality index (first principal component of soil pH, water holding capacity, organic carbon content, bulk density, and clay content). Climate data come from ERA5 reanalysis (Hersbach et al., 2020); soil data from Hengl (2018b); Hengl and Gupta (2019); Hengl and Wheeler (2018).

Figure A2: Location of Major Culture Groups



Notes: The maps plot, for each major culture group in our sample, the area where archaeological sites related to that culture are found.

Figure A3: Genetic Flow by Geographic Distance

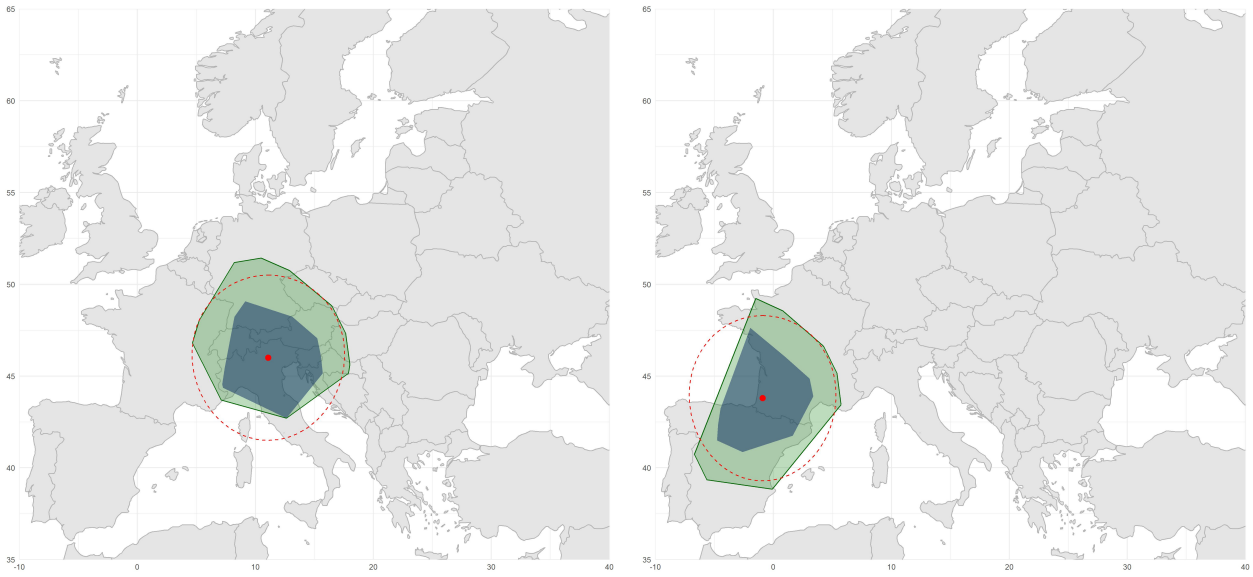


Notes: The figure shows mean genetic flow (IBD segment count $\geq$ 8cM) between grid cell pairs by geographic distance. Points represent 250km distance bins; point size reflects sample size. Shaded ribbon shows 95% confidence interval.

Figure A4: Travel Costs

(a) "Regular" Geography

(b) "Irregular" Geography



Notes: The figures plot spatial differences in travel costs for two random locations in Europe. The blue circle represents an area within 500km from the origin point. The dark blue shaded area is reachable within the by-foot travel cost equivalent of 500 km, while that in dark green the area reachable by the by-horse travel equivalent.

Table A1: Climate Matching: Main Results (Individual)

Dependent Variable:	Baseline (Any 8cM)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-0.806***	-0.564***	-0.275***	-0.263***	-0.183***	-0.323***
	(0.049)	(0.045)	(0.042)	(0.042)	(0.039)	(0.038)
Winter Temp. Dist.		-0.075***	0.010**	0.009**	0.003	0.008**
		(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Precipitation Dist.		-0.013***	-0.008***	-0.009***	-0.008***	-0.009***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Distance (000 km)			-0.759***	-0.752***	-0.719***	
			(0.039)	(0.038)	(0.038)	
Distance-sq			0.066***	0.065***	0.069***	
			(0.006)	(0.006)	(0.006)	
Soil Index Dist.				-0.079***	-0.072***	-0.077***
				(0.014)	(0.013)	(0.015)
Elevation Dist.					-0.028***	-0.022***
					(0.004)	(0.004)
Diurnal Range Dist.					-0.038***	-0.032***
					(0.011)	(0.011)
Travel Cost (horse-adj)						-0.003***
						(0.0001)
Fixed-effects						
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	5,503,125	5,503,125	5,503,125	5,503,125	5,503,125	5,503,125
Pseudo R^2	0.239	0.254	0.264	0.264	0.265	0.264

Clustered (geo_pair_ordered) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
Notes: Poisson pseudo-maximum likelihood estimates. Dependent variable is count of IBD segments $\geq$ 8cM between individual pairs. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2: Climate Matching and Migration (2.5° Grid)

Dependent Variable:	Genetic Flow				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
GDD Dist.	-0.389***	-0.205***	-0.125*	-0.391***	-0.263**
	(0.075)	(0.072)	(0.074)	(0.111)	(0.119)
Winter Temp. Dist.		-0.102***	-0.028***	0.0002	-0.010
		(0.007)	(0.005)	(0.006)	(0.006)
Precip. Dist.		-0.019***	-0.008***	-0.006***	-0.006***
		(0.002)	(0.002)	(0.002)	(0.002)
Distance (000 km)			-0.901***	-0.765***	-0.622***
			(0.077)	(0.068)	(0.059)
Distance-sq			0.068***	0.068***	0.057***
			(0.008)	(0.009)	(0.008)
Soil Index Dist.					-0.610***
					(0.117)
Elevation Dist.					0.140**
					(0.064)
Diurnal Range Dist.					-0.123***
					(0.020)
Fixed-effects					
Orig. gridcell				Yes	Yes
Dest. gridcell				Yes	Yes
Era				Yes	Yes
Fit statistics					
Observations	362,947	362,947	362,947	335,007	333,598
Pseudo R^2	0.003	0.069	0.108	0.481	0.487

Clustered (grid-cell pair) standard errors in parentheses
*Signif. Codes: ***: p<0.01, **: p<0.05, *: p<0.1*

Notes: Genetic Flow = sum of IBD segment lengths (>8cM). Coarser grid (2.5°) as robustness check.

Table A3: Climate Matching: Dependent Variable Robustness

Dependent Variables:	Genetic Flow	Mean Length (8cM)	Sum Lengths (8cM)	Mean Length (12cM)	Sum Lengths (12cM)
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
GDD Dist.	-0.136**	0.097	-0.110	0.162	-0.192
	(0.063)	(0.085)	(0.082)	(0.137)	(0.127)
Winter Temp. Dist.	-0.002	-0.003	0.0006	-0.002	0.007
	(0.006)	(0.007)	(0.007)	(0.011)	(0.012)
Precipitation Dist.	-0.005***	-0.011***	-0.007***	-0.014***	-0.009***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Distance (000 km)	-0.736***	-0.832***	-0.978***	-1.28***	-1.46***
	(0.049)	(0.051)	(0.061)	(0.078)	(0.089)
Distance-sq	0.070***	0.066***	0.094***	0.128***	0.155***
	(0.007)	(0.006)	(0.008)	(0.009)	(0.011)
Soil Index Dist.	-0.220***	-0.260***	-0.333***	-0.333***	-0.454***
	(0.036)	(0.032)	(0.070)	(0.057)	(0.082)
Elevation Dist.	-0.127**	0.236***	0.064	0.328***	0.252*
	(0.063)	(0.059)	(0.086)	(0.094)	(0.148)
Diurnal Range Dist.	-0.077***	-0.161***	-0.108***	-0.223***	-0.168***
	(0.016)	(0.026)	(0.023)	(0.040)	(0.038)
Fixed-effects					
Orig. gridcell	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	797,605	797,605	797,605	544,381	544,381
Pseudo R^2	0.46334	0.45056	0.53167	0.45427	0.51973

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Climate Matching: Distance Band Robustness

Dependent Variable:	Genetic Flow					
	Baseline	100-500km	500-1000km	1000-1500km	1500-2000km	>2000km
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-0.185***	-0.429	-0.378***	-0.271***	-0.0004	0.334***
	(0.065)	(0.341)	(0.117)	(0.081)	(0.088)	(0.066)
Winter Temp. Dist.	0.004	-0.033	-0.046**	-0.023**	-0.004	-0.021***
	(0.006)	(0.039)	(0.022)	(0.011)	(0.009)	(0.005)
Precipitation Dist.	-0.005***	0.002	-0.005	-0.001	-0.004**	-0.002
	(0.001)	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)
Soil Index Dist.	-0.262***	-0.273	-0.209***	-0.056	-0.047	-0.100***
	(0.035)	(0.201)	(0.070)	(0.056)	(0.042)	(0.034)
Distance (000 km)	-0.786***					
	(0.048)					
Distance-sq	0.069***					
	(0.007)					
Fixed-effects						
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	797,605	31,140	62,994	78,883	68,193	362,049
Pseudo R^2	0.463	0.438	0.413	0.424	0.463	0.552

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin gridcell, destination gridcell, and era fixed effects. Distance band columns exclude distance controls. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A5: Climate Matching: Soil Components

Dependent Variable:	genetic_movement						
	Index	pH	Water	Organic	Clay	Bulk	All Comp.
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
GDD Dist.	-0.185***	-0.064	-0.247***	-0.186***	-0.104	-0.234***	0.010
	(0.065)	(0.070)	(0.064)	(0.070)	(0.077)	(0.068)	(0.082)
Winter Temp. Dist.	0.004	0.001	0.007	0.006	0.006	0.008	-0.001
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Precip. Dist.	-0.005***	-0.003*	-0.005***	-0.005***	-0.006***	-0.005***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Soil Index Dist.	-0.262***						
	(0.035)						
Soil pH Dist.		-0.025***					-0.018***
		(0.006)					(0.007)
Soil Water Cap. Dist.			-0.013*				-0.009
			(0.008)				(0.009)
Soil Organic Dist.				-0.051***			-0.055***
				(0.009)			(0.010)
Soil Clay Dist.					-0.029***		-0.021***
					(0.006)		(0.006)
Soil Bulk Dens. Dist.						-0.003	0.011***
						(0.003)	(0.004)
Fixed-effects							
geo1	Yes	Yes	Yes	Yes	Yes	Yes	Yes
geo2	Yes	Yes	Yes	Yes	Yes	Yes	Yes
era500	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	797,605	797,605	797,605	797,605	797,605	797,605	797,605
Pseudo R^2	0.463	0.462	0.461	0.462	0.462	0.461	0.463

Clustered (geo_pair_ordered) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin gridcell, destination gridcell, and era

Table A6: Climate Matching: Distance Band Robustness (Individual)

Dependent Variable:	Baseline (Any 8cM)					
	Baseline	100-500km	500-1000km	1000-1500km	1500-2000km	>2000km
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-0.183***	-0.400**	-0.190**	-0.240***	-0.150*	0.165***
	(0.039)	(0.201)	(0.087)	(0.071)	(0.079)	(0.048)
Winter Temp. Dist.	0.003	-0.039*	-0.027**	-0.025***	-0.023**	-0.016***
	(0.004)	(0.022)	(0.012)	(0.009)	(0.010)	(0.005)
Precipitation Dist.	-0.008***	-0.009***	-0.003	-0.005***	-0.005**	-0.008***
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Distance (000 km)	-0.719***					
	(0.038)					
Distance-sq	0.069***					
	(0.006)					
Soil Index Dist.	-0.072***	0.014	-0.048*	-0.070***	-0.044*	-0.043**
	(0.013)	(0.029)	(0.027)	(0.020)	(0.026)	(0.017)
Elevation Dist.	-0.028***	-0.0008	-0.024***	-0.046***	-0.032***	0.008
	(0.004)	(0.020)	(0.009)	(0.009)	(0.010)	(0.006)
Diurnal Range Dist.	-0.038***	-0.058	-0.016	0.015	-0.023	-0.062***
	(0.011)	(0.045)	(0.025)	(0.023)	(0.028)	(0.016)
Fixed-effects						
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	5,503,125	535,752	782,168	877,038	616,545	2,038,106
Pseudo R^2	0.265	0.254	0.277	0.289	0.306	0.320

Clustered (geo_pair_ordered) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin gridcell, destination gridcell, and era fixed effects. Distance band columns exclude distance controls. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A7: Climate Matching: Dependent Variable Robustness (Individual)

Dependent Variables:	Baseline (Any 8cM)	Count 8cM	Sum Lengths 8cM	Any 12cM	Count 12cM	Sum Lengths 12cM
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-0.183***	-0.165***	-0.181***	-0.207***	-0.216**	-0.243**
	(0.039)	(0.056)	(0.065)	(0.077)	(0.094)	(0.102)
Winter Temp. Dist.	0.003	0.003	-0.002	0.008	0.002	-0.005
	(0.004)	(0.005)	(0.006)	(0.007)	(0.009)	(0.010)
Precipitation Dist.	-0.008***	-0.010***	-0.012***	-0.013***	-0.015***	-0.018***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Distance (000 km)	-0.719***	-0.905***	-1.06***	-1.15***	-1.41***	-1.59***
	(0.038)	(0.052)	(0.057)	(0.062)	(0.071)	(0.074)
Distance-sq	0.069***	0.088***	0.113***	0.120***	0.154***	0.183***
	(0.006)	(0.008)	(0.009)	(0.009)	(0.010)	(0.011)
Soil Index Dist.	-0.072***	-0.124***	-0.182***	-0.109***	-0.192***	-0.247***
	(0.013)	(0.023)	(0.023)	(0.028)	(0.028)	(0.030)
Elevation Dist.	-0.028***	-0.019***	-0.016***	-0.015***	-0.007	-0.004
	(0.004)	(0.004)	(0.005)	(0.006)	(0.006)	(0.007)
Diurnal Range Dist.	-0.038***	-0.053***	-0.066***	-0.075***	-0.104***	-0.126***
	(0.011)	(0.017)	(0.019)	(0.021)	(0.028)	(0.032)
Fixed-effects						
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	5,503,125	5,503,125	5,503,125	4,561,814	4,561,814	4,561,814
Pseudo R^2	0.265	0.345	0.430	0.300	0.385	0.506

Clustered (geo_pair_ordered) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A8: Climate Matching: Land Use Heterogeneity (Individual)

Dependent Variable:	Baseline (Any 8cM)					
		Baseline		Foragers	Pastoral	Farmers
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-0.183***	-0.306***	-0.279***	-0.189**	0.014	-0.191***
	(0.039)	(0.070)	(0.069)	(0.075)	(0.053)	(0.056)
Winter Temp. Dist.	0.003	0.002	0.002	-0.032***	0.013***	0.005
	(0.004)	(0.004)	(0.004)	(0.010)	(0.004)	(0.006)
Precipitation Dist.	-0.008***	-0.008***	-0.012***	-0.007***	-0.009***	-0.002
	(0.001)	(0.0010)	(0.002)	(0.002)	(0.001)	(0.002)
Distance (000 km)	-0.719***	-0.718***	-0.719***	-1.05***	-0.582***	-0.812***
	(0.038)	(0.038)	(0.037)	(0.069)	(0.043)	(0.061)
Distance-sq	0.069***	0.069***	0.069***	0.104***	0.047***	0.059***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.006)	(0.011)
Soil Index Dist.	-0.072***	-0.071***	-0.071***	-0.083***	-0.061***	-0.050***
	(0.013)	(0.013)	(0.013)	(0.020)	(0.016)	(0.016)
Elevation Dist.	-0.028***	-0.029***	-0.030***	0.032***	-0.053***	-0.002
	(0.004)	(0.004)	(0.004)	(0.011)	(0.003)	(0.010)
Diurnal Range Dist.	-0.038***	-0.036***	-0.036***	-0.024	-0.011	-0.084***
	(0.011)	(0.011)	(0.011)	(0.021)	(0.013)	(0.019)
Pastoral		-0.029	-0.069			
		(0.065)	(0.070)			
Farmer		-0.140**	-0.232***			
		(0.069)	(0.075)			
GDD Dist. × Pastoral		0.211***	0.193**			
		(0.079)	(0.078)			
GDD Dist. × Farmer		0.118	0.069			
		(0.088)	(0.086)			
Precip. Dist. × Pastoral			0.003*			
			(0.002)			
Precip. Dist. × Farmer			0.008***			
			(0.002)			
Fixed-effects						
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	5,503,125	5,494,179	5,494,179	641,713	1,853,608	2,802,297
Pseudo R^2	0.265	0.265	0.265	0.251	0.282	0.244

Clustered (geo_pair_ordered) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. Land use types based on HYDE 3.2 cropland and grazing shares at origin. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by grid cell pair in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A9: Climate Matching: Wheel Adoption Heterogeneity

Dependent Variable:	IBD Count (8cM)	
	Pre-3250 BCE	Post-3250 BCE
Model:	(1)	(2)
Variables		
GDD Dist.	0.2945*	-0.1804***
	(0.1569)	(0.0651)
Winter Temp. Dist.	-0.0067	0.0117**
	(0.0175)	(0.0051)
Precipitation Dist.	0.0036	-0.0049***
	(0.0038)	(0.0012)
Soil Index Dist.	-0.3227***	-0.2480***
	(0.0818)	(0.0356)
Distance (000 km)	-2.067***	-0.6988***
	(0.1638)	(0.0458)
Distance-sq	0.0984***	0.0561***
	(0.0355)	(0.0065)
Fixed-effects		
Orig. gridcell	Yes	Yes
Dest. gridcell	Yes	Yes
Era	Yes	Yes
Fit statistics		
Observations	69,990	702,354
Pseudo R^2	0.61796	0.49630

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. Dependent variable is count of IBD segments $\geq$ 8cM between grid cell pairs. Pre-3250 BCE corresponds to the period before widespread wheel and wagon adoption in the Pontic-Caspian steppe. Standard errors clustered by grid cell pair in parentheses.

Table A10: Climate Matching by 1000-Year Period

Dependent Variable:	IBD Count (8cM)							
	Pre-6000	5500	4500	3500	2500	1500	500	Post-500 CE
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
GDD Dist.	0.3280	0.5407**	0.3139*	0.4467***	-0.2155***	-0.1251*	-0.2418***	0.0107
	(0.2924)	(0.2144)	(0.1611)	(0.0970)	(0.0469)	(0.0731)	(0.0661)	(0.1027)
Winter Temp. Dist.	-0.0368	-0.0035	0.0231	0.0335***	0.0177***	-0.0196***	0.0042	0.0092
	(0.0319)	(0.0261)	(0.0214)	(0.0109)	(0.0063)	(0.0069)	(0.0075)	(0.0081)
Precipitation Dist.	0.0068	-0.0068	-0.0005	-0.0122***	-0.0025**	-0.0050***	-0.0053**	-0.0061***
	(0.0087)	(0.0054)	(0.0051)	(0.0024)	(0.0012)	(0.0014)	(0.0021)	(0.0022)
Fixed-effects								
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics								
Observations	3,063	14,129	15,210	49,904	102,224	77,491	91,423	171,743
Pseudo R^2	0.39870	0.74845	0.72208	0.56808	0.60315	0.52567	0.39286	0.57031

Clustered (Grid Pair) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin and destination gridcell fixed effects. Additional controls (not shown): distance, distance squared, soil quality, elevation, and diurnal temperature range. Standard errors clustered by grid cell pair in parentheses. First and last columns pool adjacent periods due to small samples. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Climate Matching by 1000-Year Period (Individual)

Dependent Variable:	Baseline (Any 8cM)							
	Pre-6000	5500	4500	3500	2500	1500	500	Post-500 CE
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
GDD Dist.	-0.2472	0.0495	-0.2981	-0.1245*	-0.2200***	-0.1574**	0.2671***	-0.2623**
	(0.2975)	(0.2110)	(0.2432)	(0.0635)	(0.0453)	(0.0775)	(0.0924)	(0.1031)
Winter Temp. Dist.	-0.1152***	-0.0846***	-0.0216	0.0395***	0.0040	0.0012	-0.0209**	-0.0060
	(0.0365)	(0.0259)	(0.0283)	(0.0066)	(0.0048)	(0.0080)	(0.0088)	(0.0091)
Precipitation Dist.	0.0004	-0.0015	0.0069	-0.0138***	-0.0008	-0.0078***	-0.0127***	-0.0135***
	(0.0077)	(0.0061)	(0.0055)	(0.0015)	(0.0013)	(0.0014)	(0.0021)	(0.0023)
Fixed-effects								
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics								
Observations	18,964	140,716	154,946	323,475	547,532	505,254	1,308,754	2,322,208
Pseudo R^2	0.25414	0.23846	0.25572	0.27292	0.17914	0.25554	0.20180	0.09061

Clustered (geo_pair_ordered) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. Dependent variable is count of IBD segments $\geq$ 8cM. All specifications include origin and destination gridcell fixed effects. Additional controls (not shown): distance, distance squared, soil quality, and elevation. Standard errors clustered by grid cell pair in parentheses. First and last columns pool adjacent periods due to small samples. *** p<0.01, ** p<0.05, * p<0.1