Climate and Prehistoric Migration*

Peter Huybers Marco Tabellini Charles A. Taylor Francesco Toti

March 2026

Abstract

What factors drove human migration before modern states, markets, and borders? We develop a framework of ecological sorting in which climate-specific subsistence knowledge depreciates with ecological distance. To test this, we use ancient DNA identity-by-descent segments to construct bilateral migration flows across Western Eurasia over the last 10,000 years. Four findings emerge. First, migration flows decline with differences in growing degree days, precipitation, and soil characteristics between origins and destinations. Second, the dimensions of climate that bind vary across subsistence systems: farmers exhibit strong thermal and soil matching, while pastoralists match most strongly on precipitation—consistent with differential ecological constraints and limiting factors. Third, migration responds adaptively to climate change: historical warming increases farmer expansion while cooling increases pastoral expansion, and these responses recover the broad temporal pattern of major archaeological migration episodes. Fourth, genetic flow predicts subsequent convergence in destination vegetation toward migrants' ecological profiles, suggesting that migration contributed to the demic diffusion of farming. Ecological sorting and climate adaptation thus operated as a fundamental driver of human movement over the Holocene.

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 and Michael Cleary for excellent research assistance. All remaining errors are ours.

1 Introduction

Throughout most of human history, migration decisions were thought to be shaped primarily by ecological conditions rather than by wage differentials, state borders, or formal institutions. Individuals accumulated detailed knowledge tied to specific environmental niches—how to farm a particular growing season, manage livestock under specific forage cycles, or exploit local wild resources. This knowledge constituted a form of human capital whose value depreciated with ecological distance: a farmer who understood a cool-temperate growing season could not simply redeploy that knowledge in a subtropical regime, just as a pastoralist adapted to arid grasslands faced challenges managing herds through a humid continental winter (Henrich, 2004; Boyd et al., 2011). Climate-specific human capital thus generates ecological sorting: populations should migrate disproportionately toward environments that resemble those to which their subsistence knowledge is adapted.

This paper shows that the dimension of environmental matching that shaped prehistoric migration depended on subsistence technology—and provides the first comprehensive evidence linking climate similarity to population movements over 10,000 years of human history. We develop a simple ecological sorting framework, inspired by Roy (1951), in which individuals choose locations that yield the highest returns net of moving costs. We test these predictions using a novel migration dataset derived from genetic connections between ancient individuals.

Building on Ringbauer et al. (2024)'s method for detecting IBD segments even in low-coverage ancient DNA, we construct a bilateral panel of migration flows across Western Eurasia. We discretize space into 1° gridcells and time into 500-year windows, and then link these genetically inferred flows to high-resolution measures of ecological conditions at origins and destinations, including growing-season heat accumulation, precipitation, and soil characteristics.

We document four main findings. First, we use a gravity model of bilateral flows (Anderson and Van Wincoop, 2003) to show that migration flows decline sharply with ecological distance between origin and destination. We account for geographic distance as well as various measures of environmental distance including growing-season heat accumulation, precipitation, and soil characteristics. We find that a one-standard-deviation increase in growing degree day distance (about 700 degree-days, or roughly a 3.5°C shift in mean temperature over a 200-day growing season) between gridcells reduces migration flows by 13%. A one-standard-deviation increase in precipitation distance (about 300 mm per year) reduces flows by 18%. A one-standard-deviation increase in soil composite distance reduces flows by 27%. These magnitudes are equivalent to increasing geographic distance by 430–950

kilometers, underscoring that ecological mismatch along a single climate dimension carries the same penalty as the travel cost associated with hundreds of kilometers physical distance (Table 2).

We include origin and destination gridcell fixed effects to absorb all time-invariant location characteristics (i.e., nonrandomness in archaeological sampling) and period fixed effects to account for aggregate changes in sample availability and migration intensity over time. The results also survive origin-by-era and destination-by-era interacted fixed effects that absorb all time-varying push and pull factors; additional environmental controls including UV-B radiation, elevation, and coastal proximity; terrain-adjusted travel-cost distance that accounts for topography and water barriers. We also replicate the analysis using an individual-level specification of over five million pairwise observations as an alternative to gridcell aggregation (Section 5).

Our second finding involves heterogeneity in climate matching by subsistence technology. Farmers, whose crop development depends on accumulated thermal time—a characteristic heat sum required for maturity that constrains where a crop variety can be grown (Guedes and Butler, 2014; d'Alpoim Guedes and Bocinsky, 2018)—show the strongest matching along growing-season heat and soil characteristics, and null sensitivity to precipitation. Pastoralists, whose forage base—grassland productivity—is governed primarily by precipitation (Sala et al., 1988; McGuirk and Nunn, 2025), show the opposite pattern: strong precipitation matching but null thermal sensitivity. Foragers—a small share of the sample—show no statistically significant matching on any individual dimension (Table 3).

The distance coefficients further differentiate the groups. Each additional 100 km of geographic separation between two gridcells reduces pastoral migration flows by only 4%, compared with 18% for farmers and 32% for foragers. This ordering reflects the mobility of productive assets: pastoral wealth is livestock, which walks to new pasture, while farming requires land-intensive fixed investments—cleared fields, soil improvements, stored seed—that raise the cost of relocation. Foragers, operating at the smallest spatial scales, face the steepest penalty of all. We validate these patterns using named archaeological culture groups: steppe pastoral cultures (e.g., Yamnaya) show near-zero geographic distance sensitivity and null GDD matching, while farming cultures show a steep distance penalty and localized thermal matching (Figure 3).

It is worth noting that this pattern—in which the dimension of matching that binds depends on subsistence technology—provides a source of identification beyond the geographic fixed effects. Temperature generally varies along a latitudinal gradient, precipitation along coast-to-interior gradients, and soil characteristics reflect bedrock geology rather than latitude—so no single geographic confounder can explain matching along all three dimen-

sions simultaneously. The fact that each subsistence system matches most tightly on the ecological dimension it cannot technologically buffer suggests a mechanism rooted in the transferability of ecological knowledge, not omitted bilateral confounders.

Third, climate change differentially mobilized subsistence groups. Warming at the origin increased farmer expansion (0.86, $p < 0.01$), consistent with expanding the cereal-viable zone, while cooling increased pastoral expansion (-1.32, $p < 0.01$), consistent with grassland expansion favoring herding (Table 4). When we apply these estimated coefficients to realized Holocene temperature shocks, the model recovers the broad temporal pattern of major archaeological expansions from climate data alone (Figure 4): farmer migration peaks during the Holocene Climatic Optimum (∼5500–4500 BCE) when the LBK and Funnel Beaker cultures spread across Europe, and pastoral migration spikes during Neoglacial cooling after 3000 BCE when the Yamnaya, Sintashta, and Scythian expansions occurred. That a parsimonious climate model reproduces the temporal pattern of subsistence-specific migration episodes documented in the archaeological record lends further credibility to the ecological sorting mechanism.

Moreover, migration systematically offset local climate shifts in both directions: populations exposed to warming relocated toward cooler destinations, and those exposed to cooling relocated toward warmer ones. Migration compensated for roughly one-third of the local temperature shift (Table 5; Figure 5). The response is asymmetric: warming shocks produce roughly two-and-a-half times the directional sorting response of cooling shocks.

Fourth, if migrants carry climate-specific skills, they should reshape the ecology of their destinations. We find that locations receiving more migrants from ecologically different origins experience convergence in grassland and tree pollen profiles ($p < 0.01$) and marginally significant convergence in Cerealia-type pollen ($p < 0.10$). For cereals, the reverse pattern does not hold—the coefficient on prior destination vegetation change interacted with genetic flow is significantly negative, inconsistent with reverse causality. These results suggest that migration facilitated the demic diffusion of agricultural practices rather than the diffusion of ideas alone (Table 6).

It is also worth noting that the destination temperature controls in the diffusion regressions reinforce the niche expansion interpretation: warming increases cereal and tree pollen while decreasing grassland cover, consistent with the well-documented Holocene pattern of forest and agricultural expansion during warm periods at the expense of open grassland (Marcott et al., 2013). These ecological shifts are precisely those that would expand the viable range for farming while contracting pastoral rangeland, linking the diffusion results back to the differential mobilization documented above.

Our findings resonate with a long tradition in anthropology, archaeology, and paleoe-

cology emphasizing the role of seasonal climate structure in shaping human mobility. A central insight of anthropology is that mobility strategies were organized around seasonal patterns—where food could be produced, stored, or accessed, and how predictable those conditions were across space (Lee and DeVore, 1968). Quantitative studies of crop dispersal show that thermal thresholds and growing-season length constrained where early farmers could move, shaping migration routes and settlement patterns (Guedes and Butler, 2014; Castiello et al., 2025).

Despite these insights, much of this literature infers mobility indirectly from settlement locations, crop assemblages, or environmental reconstructions rather than from observed movements of people themselves. We provide a population-level test by leveraging recent advances in ancient DNA analysis (Reich, 2018; Ringbauer et al., 2024) to construct bilateral migration flows from IBD segments in ancient genomes. To our knowledge, this is the first study in economics to use IBD segments as a quantitative measure of bilateral migration flows. The implied displacement rates independently recover canonical estimates from the archaeological literature—approximately 1 km per year for the Neolithic expansion (Ammerman and Cavalli-Sforza, 1971) and 4.2 km per year for the Yamnaya steppe expansion (Racimo et al., 2020)—providing strong external validation of the measure (Section 4.3).

Our results contribute to the economics literature on ecological constraints and long-run development. Michalopoulos (2012) shows that land quality and terrain shape ethnolinguistic differentiation by limiting movement. Fenske (2014) documents how ecological diversity promoted trade and state formation. Ashraf and Michalopoulos (2015), Giuliano and Nunn (2021), and Matranga (2024) emphasize how climatic variability, stability, and seasonality shaped agricultural diffusion, cultural persistence, and the transition from nomadism to sedentarism. We complement this work by showing that ecological constraints shaped not only where populations settled, but how they moved between settlements, and that the relevant dimension of climate depended on subsistence technology.

Our evidence that migration served as a margin of climate adaptation further complements studies linking environmental change, population dynamics, and long-run development (Galor and Moav, 2002; Ashraf and Galor, 2013; Galor and Özak, 2016), as well as research in cultural evolution stressing the slow accumulation of locally adapted knowledge (Henrich, 2004; Boyd et al., 2011; Nunn, 2021). Bazzi et al. (2016) provide causal evidence for this mechanism in a modern setting: in Indonesia's Transmigration Program, farmers assigned to agroclimatically dissimilar locations experienced large productivity losses, while those sent to similar environments performed substantially better. Obolensky et al. (2024) document systematic climate matching in contemporary migration.

We extend these insights to a setting that strips away markets, borders, and amenity

channels, and that provides variation in subsistence technology—enabling us to show that the dimension of climate matching depends on production strategy, a test that is unavailable in any modern dataset. Our evidence on ecological diffusion connects to Diamond (1997)'s hypothesis that the east–west orientation of Eurasia facilitated the spread of agriculture by limiting climatic barriers, historical evidence that 19th-century US farmers migrated along east–west corridors toward familiar climatic zones (Steckel, 1983), and Nicolò et al. (2026), who show that European colonial success depended on environmental similarity to the colonizers' homelands.

The prehistoric setting offers three advantages for isolating the ecological mechanism. First, it strips away modern labor markets, borders, and formal institutions, allowing us to observe ecological sorting without the confounding influence of wage differentials or immigration policy. Second, prehistory provides variation across coexisting production systems—foragers, pastoralists, and farmers occupying the same landscape—that no modern dataset can replicate. This variation is what allows us to test whether the dimension of matching scales with ecological dependence. Third, the 10,000-year horizon spans major technological and climatic shifts, including the adoption of the wheel and sustained episodes of warming and cooling, enabling us to study how transport costs and climate change altered the role of ecological similarity in shaping mobility.

The remainder of the paper is organized as follows. Section 2 presents the conceptual framework. Section 3 describes the data. Section 4 presents descriptive evidence on the spatial and temporal patterns of migration, climate, and subsistence. Section 5 develops the empirical strategy and reports the main climate matching results, including subsistence heterogeneity and robustness checks. Section 8 examines migration as an adaptive response to long-run climate change, covering both climate niche expansion across subsistence types (Section 8.1) and adaptive sorting (Section 8.2). Section 9 tests for ecological diffusion. Section 10 concludes.

2 Conceptual Framework

Prehistoric migration was shaped not only by physical distance but by the ecological similarity between origins and destinations. Individuals who possessed detailed knowledge of local growing conditions, soil characteristics, and seasonal cycles could redeploy that knowledge most effectively in environments similar to those they left behind. We develop a simple framework that formalizes this idea, allowing the degree and dimension of climate specificity to vary across subsistence modes—farmers, pastoralists, and foragers—and generating testable predictions about the relationship between ecological matching, subsistence technology, and

migration.

2.1 Setup

Consider a landscape of locations indexed by their climate type $\boldsymbol{\theta} = (\theta_T, \theta_P, \theta_S)$, where θ_T represents thermal conditions (cumulative growing degree days), θ_P represents hydrological conditions (annual precipitation), and θ_S summarizes soil characteristics. An individual i born at a location with climate $\boldsymbol{\theta}_i$ inherits a stock of subsistence knowledge $k(\boldsymbol{\theta}_i)$ adapted to that environment. Following Becker (1964), we treat this knowledge as a productive asset whose returns depend on the match between what the individual knows and the environment in which it is deployed. Productivity when individual i occupies a location with climate $\boldsymbol{\theta}_j$ is

$$f(k(\boldsymbol{\theta}_i), \boldsymbol{\theta}_j) = A - \sum_{d \in \{T,P,S\}} c_d \cdot |\theta_{j,d} - \theta_{i,d}|^{\alpha_d} \tag{1}$$

where A is a common productivity level, $c_d > 0$ scales the cost of mismatch along dimension d, and $\alpha_d > 0$ governs the curvature of knowledge transferability. When α_d is large, even modest differences along dimension d generate steep productivity losses; when α_d is small, knowledge transfers more smoothly. Which dimensions bind depends on the subsistence technology in use.

The individual chooses a destination j to maximize net payoff, where $\tau(d_{ij})$ is a transport cost increasing in geographic distance:

$$\max_j \left\{ A - \sum_{d \in \{T,P,S\}} c_d \cdot |\theta_{j,d} - \theta_{i,d}|^{\alpha_d} - \tau(d_{ij}) \right\} \tag{2}$$

This embeds a Roy (1951)-style location choice in which individuals sort to destinations where their ecological knowledge is most productive, net of moving costs.[1]

2.2 Subsistence Heterogeneity

The critical parameters are the α_d values, which we allow to vary across subsistence modes. Farmers depend on accumulated thermal time: crop varieties, planting calendars, and pest management are calibrated to local growing-season conditions, making thermal mismatch

[1]The structure parallels Borjas (1987)'s application of the Roy model to immigrant self-selection and extends Bazzi et al. (2016), who model agroclimatic similarity as location-specific human capital in Indonesia's Transmigration Program. We abstract from within-origin heterogeneity—all individuals from a given location inherit the same knowledge stock—so the model is best understood as a location-choice framework rather than a full selection model. Appendix A provides a fuller discussion.

costly. But farming technologies—crop selection, water management—largely decouple productivity from local precipitation, concentrating vulnerability on the thermal and edaphic dimensions. Pastoralists face a different constraint: livestock tolerate a wider range of temperatures than any single crop, but forage availability is governed primarily by precipitation (Sala et al., 1988; McGuirk and Nunn, 2025). Foragers rely on broad ecological knowledge that may transfer more readily across climate gradients, making them the least constrained group along any single climate dimension. However, forager knowledge is largely landscape-indexed—organized around cognitive maps of specific plant patches, game trails, and water sources—rather than climate-indexed. This knowledge depreciates with any geographic displacement, not specifically with climate distance, implying that foragers should show weak climate matching but the steepest distance penalty of the three groups.

These differences imply a matrix of climate sensitivity rather than a scalar ordering:

$$\alpha_T^{\text{farmer}} \gg \alpha_T^{\text{pastor}} \approx \alpha_T^{\text{forager}} \tag{3}$$

$$\alpha_P^{\text{pastor}} > \alpha_P^{\text{farmer}} \approx \alpha_P^{\text{forager}} \tag{4}$$

The central contrast is thus between farmers and pastoralists: farmers occupy a niche that is narrow along the temperature axis but broad along the precipitation axis, while pastoralists occupy a niche that is broad along the temperature axis but dependent on moisture for grassland productivity. The soil dimension is partially ordered: farmers face a hard workability constraint—clay-heavy soils were un-farmable with Neolithic tillage technology, creating a sharp boundary at the loess frontier (White, 1962)—suggesting α_S^{farmer} is the largest. Pastoralists require suitable rangeland and foragers depend on local ecosystem quality, but the ordering between them is ambiguous:

$$\alpha_S^{\text{farmer}} > \alpha_S^{\text{pastor}} \gtrless \alpha_S^{\text{forager}} \tag{5}$$

2.3 Climate Change and Adaptive Migration

When climate evolves over time, the individual's knowledge stock becomes less well matched to the new local conditions. The observed migration response depends on the sensitivity to climate mismatch (α_d) relative to the cost of relocating (κ_s), which varies across livelihood strategies: farmers face high κ (fixed fields, soil improvements) and pastoralists face low κ (mobile livestock).

This distinction separates two margins. The first is *niche expansion*: climate change alters the geographic scope of each subsistence system. Warming expands the thermal range in which cereal cultivation is viable, opening new territory for farmer colonization; cooling

shifts the ecological frontier toward grassland, expanding the pastoral niche. Both groups should respond to temperature shocks, but in opposite directions. The second is *destination selection*: conditional on migrating, individuals should relocate toward destinations that offset the local climate shift, preserving their ecological niche. If migrants also carry subsistence knowledge to their destinations, migration should drive ecological convergence between connected locations—a prediction that distinguishes this framework from standard gravity models in which migration is a purely demographic process.

2.4 Testable Predictions

The framework generates the following predictions:

1. **Climate matching.** Migration flows decline with climate distance between origin and destination, holding geographic distance constant. (Section 6.1)

2. **Subsistence heterogeneity in multidimensional matching.** Farmers match most tightly on the thermal and edaphic dimensions ($\alpha_T^{\text{farmer}} \gg \alpha_T^{\text{pastor}}$), reflecting growing-season specificity. Pastoralists face the strongest constraint along the precipitation dimension ($\alpha_P^{\text{pastor}} > \alpha_P^{\text{farmer}}$), because consistent moisture is the binding input for forage and grassland productivity. Farmers should dominate on soil; the pastoral-forager soil ordering is an empirical question. The distance penalty should be steepest for foragers (landscape-indexed knowledge), intermediate for farmers (fixed agricultural investments), and weakest for pastoralists (mobile livestock). As transport costs fall—e.g., the adoption of wheeled transport around 3250 BCE (Anthony, 2007)—migrants gain access to a wider set of destinations, increasing the scope for climate-based sorting and strengthening observed matching. (Section 7)

3. **Adaptive migration.** Warming expands the cereal-viable zone, increasing farmer migration; cooling expands grassland, increasing pastoral migration. The two groups should respond in opposite directions. Beyond volume, migrants should sort toward destinations that offset local climate shifts. (Section 8)

4. **Ecological diffusion.** Migrants carry subsistence knowledge that transforms destination ecology, driving convergence in vegetation and land cover between connected locations. (Section 9)

We emphasize that this framework is intended as a guide to interpretation, not as a structural model to be estimated. The goal is to provide a unified lens connecting climate matching, subsistence heterogeneity, temporal dynamics, and ecological diffusion within a single economic logic. We now turn to the data.

3 Data

This section describes the main datasets used in the analysis. Table 1 presents summary statistics for the gridcell-pair panel. The unit of observation is an origin–destination–era triple, yielding approximately 879,000 observations. The dependent variable—IBD pair count at the 8 cM threshold—has a mean of 0.18, reflecting the predominance of zeros in the migration matrix.

3.1 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 8 cM. 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 $1^\circ \times 1^\circ$ 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 8 cM or longer, linking individuals buried in different cells but within the same time window.

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

3.2 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 temperature using ERA5, a global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing consistent gridded fields at hourly resolution (Hersbach et al., 2020). We use the daily ERA5 data to construct cumulative growing-degree days above 10°C (GDD10), which summarizes biologically relevant heat accumulation for plant growth and animal productivity.[2] For precipitation, we use TerraClimate (Abatzoglou et al., 2018), a high-resolution (≈4 km) global dataset that interpolates observed station measurements using climatologically aided interpolation. We prefer TerraClimate over ERA5 for precipitation because the observation-based product has less measurement error for rainfall—when both are included simultaneously, TerraClimate dominates while ERA5 precipitation becomes insignificant, with all other coefficients unchanged. We compute average annual values for each $1^\circ \times 1^\circ$ gridcell over the period 1960–2000.

Paleoclimate. We complement modern climate data with paleoclimate reconstructions

[2]The standard base temperature for wheat and barley—the foundational Neolithic cereals—is 0–5°C, while a 10°C base is more common for warm-season crops. We use GDD10 as a general measure of growing-season heat accumulation that captures thermal gradients relevant across subsistence systems.

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). The resulting gridded temperature data have a spatial resolution of $1.9° \times 2.5°$, which we aggregate to the $1° \times 1°$ grid used to measure migration flows.

Reconstructions of paleoclimate 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.

3.3 Soil, Topography, and Other Environmental Data

Soil characteristics. We obtain gridded soil data—including clay content, soil organic carbon, pH, water-holding capacity, and bulk density—from SoilGrids (Hengl, 2018b; Hengl and Wheeler, 2018; Hengl and Gupta, 2019; Hengl, 2018a) at 250-meter resolution, aggregated to our analysis grid. For the baseline specification, we construct a composite soil distance measure as the Euclidean distance between standardized clay content and soil organic carbon. We select these two components because they are the most geologically determined soil properties—clay content is nearly uncorrelated with precipitation distance ($r = 0.09$) and is temporally stable over Holocene timescales, while organic carbon captures fertility differences visible to prehistoric cultivators. Alternative soil measures, including a PCA-based index and individual component distances, are examined in robustness (Table A2).

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 and subsistence classification. For the heterogeneity analysis, we classify each ancient DNA population by subsistence type—farmer, pastoralist, or forager—using a two-step procedure. First, each of the approximately 1,800 unique populations in the Ringbauer et al. (2024) dataset is assigned to one of 50 archaeological culture groups based on keyword

matching of population labels (e.g., "LBK" → Early European Farmers; "Yamnaya" → Early Steppe Expansion) with a country-and-date fallback for unmatched populations. Second, each population is assigned continuous subsistence shares (proportion farmer, pastoralist, forager, summing to one) based on the archaeological and ethnographic record for its culture, region, and time period. These shares were cross-validated against a rule-based classifier and manually curated. The categorical subsistence type is the mode with the largest share.

As a robustness check, we also construct a gridcell-level classification using the HYDE 3.2 database (Klein Goldewijk et al., 2017), which assigns subsistence type based on the relative intensity of cropland versus grazing in each cell and period. Both approaches yield qualitatively similar results (Section 7). Figure A2 displays the evolution of subsistence composition over time under both classifications: foragers dominate the sample before approximately 5000 BCE, after which farmers become the majority, with pastoralists emerging as a persistent share from approximately 3000 BCE onward.

Geographic distance. Geographic distance between gridcell 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.

Vegetation reconstructions. For the ecological diffusion analysis (Section 9), we use the REVEALS vegetation reconstruction from Fyfe et al. (2021), which provides bias-corrected estimates of land-cover composition at 1-degree resolution in 500-year intervals across Europe. The REVEALS model (Sugita, 2007) converts raw pollen counts into vegetation cover estimates by correcting for taxon-specific differences in pollen productivity and dispersal distance, yielding more accurate land-cover reconstructions than uncorrected pollen percentages (Githumbi et al., 2022). We use three land-cover variables: arable land (Cerealia-type pollen), grassland, and tree cover. The gridded REVEALS estimates are interpolated to each analysis cell via inverse-distance weighting from nearby cells within 500 km. Spatial coverage is dense in Scandinavia and northwestern Europe but sparser in the Mediterranean and eastern regions, a limitation we discuss in Section 9.

4 Descriptive Evidence

4.1 The Spatial Distribution of Ancient DNA Remains

The sample includes 6,065 individuals from Ringbauer et al. (2024), with dates of death ranging from 7550 BCE to 1823 CE. Because ancient DNA degrades faster under heat and moisture (Reich, 2018), cooler, drier locations in northern and central Europe are overrep-

resented. Table A3 in Appendix B documents these preservation correlates systematically. This spatial selection does not threaten identification: we exploit variation across origin–destination pairs, and origin and destination fixed effects absorb persistent differences in sampling intensity across locations.

4.2 Climate

Key climate variables exhibit pronounced spatial gradients across the study region: growing degree days decline sharply with latitude, while precipitation varies substantially across longitudinal bands (Figure A1). Crucially, the Holocene was marked by substantial temperature fluctuations over time (Marcott et al., 2013), with alternating phases of warming and cooling (Figure A5). Paleoclimate reconstructions from marine proxy data (Osman et al., 2021) reveal that these temperature changes were spatially heterogeneous—within the same latitude bands, some regions warmed while others cooled (Figure A6). This combination of temporal variation and spatial heterogeneity generates the identifying variation we exploit below.

4.3 Migration

Figure 1 visualizes migration flows across Western Eurasia in four periods from 5000 to 2000 BCE, displaying links between locations where genetically related individuals are observed.[3] Genetic flow declines steeply with geographic distance, with most connections concentrated below 2,000 km (Figure A11 in the Appendix). A notable pattern is the prevalence of east–west movement, consistent with migrants sorting toward climatically similar environments—a pattern documented in other historical and modern contexts (Steckel, 1983; Diamond, 1997; Obolensky et al., 2024; Ringbauer et al., 2024).

Cultural groups. Ancient DNA data reveal rich variation in the spatial distribution of cultural groups over time. Groups are classified based on Ringbauer et al. (2024)'s population classification, with some categories homogenized into tighter groups. Figure A10 plots the locations of major groups, including early farming societies, steppe-associated populations, and later Bronze and Iron Age groups. Early farming groups are initially spatially concentrated, while steppe-associated groups exhibit much wider dispersion; over time, groups increasingly

[3]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.

overlap, reflecting repeated episodes of migration and admixture (Lazaridis et al., 2014; Allentoft et al., 2015). Figure A7 plots migration flows separately by cultural group, revealing substantial heterogeneity in mobility across cultural contexts. These classifications also provide an alternative basis for assigning subsistence type to individuals, which we exploit in the heterogeneity analysis (Section 7).

Mobility by subsistence type. The geographic scale of migration varies sharply across subsistence strategies. Table A4 reports IBD-weighted mean migration distances: pastoralists travel roughly three times farther than foragers (2,182 km vs. 694 km) and twice as far as farmers (1,071 km), with agro-pastoralists in between (1,561 km). This ordering is confirmed by the geographic distance coefficient in the gravity model (Section 7): the distance penalty is roughly ten times weaker for pastoralists (-0.375) than for foragers (-3.91). As Figure A2 shows, the composition of the sample shifts dramatically over time—from forager-dominated to farmer-dominated—so aggregate patterns reflect a changing mixture of subsistence-specific behaviors.

Figure 2 documents a structural break in the geographic scale of migration. Prior to approximately 3250 BCE, the average distance between genetically related individuals remained stable at 500–900 km. After this date, distances increased sharply, peaking above 2,000 km around 2500 BCE before declining to approximately 1,300 km by the Roman period. The timing closely coincides with the adoption of the wheel and wagon in the Pontic-Caspian steppe (Anthony, 2007). The lower panel confirms this is not a sample size artifact: while the number of observed remains increases around 3000 BCE, it subsequently declines to pre-3000 BCE levels, whereas migration distances remain persistently elevated.

Implied migration speed. The ratio of geographic distance to temporal separation between related individuals yields an implied displacement rate. Figure A3 plots this measure by era. Pre-4000 BCE rates average approximately 1.1 km per year, matching canonical estimates of Neolithic demic diffusion (Ammerman and Cavalli-Sforza, 1971; Pinhasi et al., 2005; Racimo et al., 2020). A sharp acceleration around 3000 BCE yields rates of approximately 4.0 km per year, aligning with independent estimates of the Yamnaya expansion at 4.2 km per year (Racimo et al., 2020). This correspondence validates the IBD measure as a quantitative proxy for prehistoric population movements. Decomposing by subsistence type (Figure A8), farmers consistently move more slowly than pastoralists across all eras, with the aggregate acceleration around 3000 BCE reflecting both a technology effect (the wheel) and a composition effect (the rising share of pastoralists).

5 Empirical Strategy

Our empirical analysis builds on the origin–destination migration panel constructed from ancient DNA data, as described in Section 3.1. 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:

$$M_{odt} = \exp[\alpha_o + \gamma_d + \delta_t + \beta \text{Dist}_{od}^{\text{Climate}} + \theta \text{Dist}_{od}^{\text{Physical}}]\epsilon_{odt} \tag{6}$$

M_{odt} is a measure of migration intensity between the origin and the destination in a given period. We use, as a proxy for migration intensity, genetic connectedness implied by 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.[4] In our baseline specification, 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

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

segments to contribute more than one), and restriction to longer segment thresholds of 12 and 16 cM.

$\text{Dist}_{od}^{Climate}$ captures the absolute difference in environmental conditions between the two locations along three dimensions. We select these dimensions based on a systematic analysis of correlations among candidate variables (Appendix Figure A4). Temperature-related distances—GDD, summer temperature, seasonal range, and mean temperature—are highly correlated with one another and with latitude distance, forming a cluster from which only one representative should be included. We select cumulative growing degree days (GDD) above 10°C as the most agriculturally relevant thermal measure. Winter temperature distance is strongly correlated with geographic distance ($r = 0.63$) and becomes statistically insignificant when GDD is included, suggesting it primarily captures geographic proximity rather than independent climate matching. Annual precipitation provides an orthogonal moisture dimension ($r = 0.22$ with GDD distance).

For soils, we construct a Euclidean distance from two standardized components: clay content and soil organic carbon. Clay content is determined primarily by geological substrate and is nearly uncorrelated with precipitation distance ($r = 0.09$), providing a genuinely climate-independent dimension of environmental similarity. Soil organic carbon captures fertility differences visible to prehistoric cultivators as soil darkness. Together, the composite has low correlation with both precipitation ($r = 0.15$) and geographic distance ($r = 0.11$).

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 across destination locations that differ in climatic similarity, holding constant origin-specific, destination-specific, and period-specific factors. Standard errors are clustered at the cell pair 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 is itself correlated with local climatic conditions. The inclusion of origin and destina-

tion 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 (Section 7) 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.

A powerful source of reassurance comes from the multidimensional structure of our results. A single bilateral confounder—such as shared expansion corridors or geographic proximity along a latitudinal gradient—would generate matching on all environmental dimensions simultaneously. Instead, we find that the *dimension* of environmental matching that binds depends on subsistence technology: farming populations match most strongly on growing-season heat (GDD), while pastoral populations match most strongly on precipitation (Section 7). This crossover pattern is difficult to reconcile with omitted bilateral confounders, which would predict uniform matching across all groups and dimensions. The subsistence-specific pattern is, however, a direct prediction of a framework in which migrants carry location-specific ecological knowledge whose value depreciates along different environmental gradients depending on subsistence strategy.

6 Main Results

6.1 Dimensions of Climate and Ecological Matching

Table 2 presents results from equation (6). All columns include the same four environmental distance measures—GDD, precipitation, soil composite, and geographic distance—with progressively more demanding fixed-effects structures. Column 1 includes no fixed effects; column 2 adds era fixed effects; column 3 adds origin and destination fixed effects; column 4 (our preferred specification) includes all three. Columns 5 and 6 further tighten identification with origin-by-era and destination-by-era interacted fixed effects.

All four environmental dimensions enter with negative and statistically significant coefficients across all specifications, indicating that migration flows systematically decline with ecological distance. As expected, the GDD coefficient attenuates substantially when origin and destination fixed effects are introduced (from -0.855 to -0.201), reflecting the absorption of persistent differences across locations. Despite this stringent fixed-effects structure, the estimate remains precisely estimated in the preferred specification (-0.200, $p < 0.01$). Precipitation and soil composite coefficients are notably stable across columns, and all four

variables survive even the fully interacted specification in column 6.

In the preferred specification, a difference of 1,000 GDD above 10°C—slightly above the sample mean, equivalent to sustaining a 5°C higher mean temperature over a 200-day growing season—is associated with a reduction in migration of about 18%. A one-standard-deviation increase in precipitation distance (about 300 mm per year) reduces flows by roughly 18%. Soil composite distance has a coefficient larger in magnitude than GDD (-0.352), indicating that edaphic similarity represents an independent and economically important margin of ecological matching. These magnitudes are comparable to those of increasing geographic distance by roughly 600 kilometers, underscoring that ecological similarity shapes mobility at least as strongly as substantial physical separation.

7 Subsistence Heterogeneity

Subsistence heterogeneity. 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 depend on consistent moisture for forage. We test this prediction by estimating the baseline specification separately for each subsistence type, classifying origin populations using archaeological population labels as described in Section 3.

Table 3 presents the results, with all regressors standardized within each subsample for cross-variable comparability. The central finding is a crossover pattern in which the dimension of ecological matching that binds depends on subsistence technology. Farmers exhibit the strongest GDD matching (-0.183, $p < 0.01$) and strong soil matching (-0.225, $p < 0.01$), but no significant precipitation matching (-0.008). Pastoralists show the opposite pattern: null GDD matching (-0.003, insignificant) but the strongest precipitation matching (-0.223, $p < 0.01$), with weaker soil effects (-0.096, $p < 0.05$). Foragers—a small subsample (26,397 observations)—show no statistically significant matching on any dimension, though coefficient signs are negative for GDD and soil.

The farmer–pastoralist contrast is the paper's most distinctive finding. Farmers match on the thermal and edaphic dimensions most relevant to crop production: growing-season heat determines what can be grown, and soil workability determines where it can be planted. Pastoralists match on the moisture dimension most relevant to livestock: precipitation determines forage availability and grassland productivity. Neither group matches significantly on the other's binding dimension, consistent with the framework's prediction that subsistence

technology determines which ecological axis is hardest to buffer.

The distance coefficients reinforce this interpretation. Pastoralists show the weakest distance penalty (-0.375), farmers a steep penalty (-1.96), and foragers the most localized movement (-3.91). This ordering maps onto relocation costs: mobile herding permits continental-range movement, while settled farming and local foraging constrain migration to shorter distances. Table A5 in the Appendix confirms these patterns using continuous share interactions in the pooled sample, with forager as the reference group. The interaction of farmer share with GDD distance is large and negative (-1.17, $p < 0.01$), confirming that farmers match more strongly on temperature than foragers. Table A6 replicates the split-sample analysis at the individual level with qualitatively similar results.

Culture-group validation. As a complementary approach, we examine heterogeneity directly across named archaeological culture groups. Figure 3 reports GDD matching and geographic distance coefficients estimated separately for 14 major culture groups. The results validate the subsistence-based patterns at a finer resolution: steppe pastoral cultures (Yamnaya, Corded Ware, Steppe Turkic) exhibit near-zero distance sensitivity and null GDD matching, consistent with continental-range mobility and thermal flexibility. Farming cultures (Early European Farmers, Copper Age) show a steep distance penalty, consistent with localized, corridor-bound expansion. Hunter-gatherer groups (Western and Eastern European hunter-gatherers) are the most spatially constrained.

Precipitation and drought. The concentration of precipitation matching among pastoralists—rather than foragers or farmers—aligns with archaeological evidence on the role of moisture in pastoral economies. The 4.2 ka megadrought event (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), and is associated with increased pastoral mobility as farming became unviable in drying regions. More broadly, the persistent effect of precipitation distance in our results, and its concentration among pastoralists, suggests that water availability was the binding ecological constraint for herding societies—complementing the thermal constraints that bound farmers.

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 A7). 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).

Robustness Checks. In this section, we assess the robustness of the baseline results by varying both the measurement of climate similarity and the inclusion of additional environmental controls. Table A8 examines sensitivity to additional environmental distances beyond our baseline climate measures, progressively adding controls for UV-B radiation, diurnal temperature range, elevation, coastal distance, latitude, and longitude differences. Across all specifications, the baseline coefficients remain stable, indicating that the results are robust to controlling for these additional environmental dimensions.

We further assess robustness along several dimensions. First, Table A9 confirms that results are consistent across all alternative measures of migration intensity described above. Second, we replicate the main results using individual-level data in Table A10, 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, Table A11 stratifies the sample by distance bands (100–500 km, 500–1000 km, etc.), confirming that climate matching operates across a range of geographic scales rather than being driven by particular distance ranges. Fourth, Table A2 decomposes the soil composite 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. Fifth, Table A12 confirms that results hold when restricting the sample to horizontal (east–west) or vertical (north–south) moves and when excluding short-distance pairs ($\geq$100 km), ruling out the possibility that climate matching is driven by a particular directional corridor or by local-scale genetic structure. Sixth, Table A13 shows that results are robust to more demanding fixed-effects structures—including origin$\times$era

and destination×era fixed effects that absorb all time-varying push and pull factors—and to broader spatial clustering at the 2.5-degree level, which accounts for spatial correlation in the error term beyond the baseline 1-degree pair clustering. Seventh, the subsistence heterogeneity results are validated at a finer resolution using named archaeological culture groups (Figure 3), as discussed in Section 7. The culture-group-level patterns—steppe pastoral cultures showing null GDD matching and near-zero distance sensitivity, farming cultures showing a steep distance penalty—confirm that the subsistence-based ordering is not an artifact of the classification methodology. Eighth, Table A14 shows that results are robust to alternative distance specifications, including quadratic, log, and least-cost-path distance on foot that accounts for topography and terrain ruggedness (Appendix Figure A12). We note that subsistence classification is itself endogenous to location—populations in certain environments are more likely to be classified as farmers or pastoralists. Origin fixed effects absorb level differences in subsistence composition across locations, and the split-sample approach estimates climate responses within each subsistence group conditional on its spatial distribution.

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

The matching results establish *which* environmental dimensions bind for each subsistence group: farmers are most constrained by growing-season heat and soil conditions, while pastoralists are most constrained by moisture. We now test *what happens* when climate shifts along these dimensions. The Holocene was marked by distinct episodes of warming and cooling that shifted local growing-season conditions across Western Eurasia (Marcott et al., 2013). This question has two distinct margins. The first is *migration propensity*: does climate change differentially mobilize particular subsistence groups, expanding the geographic range of some livelihood strategies while contracting others? The second is *destination selection*: conditional on a climate shock, do migrants redirect toward destinations that offset the change? The first margin concerns the volume of flows; the second concerns their direction.

8.1 Climate Niche Expansion: Who Moves More?

The matching results establish which environmental dimensions bind for each subsistence group. We now test whether climate change differentially mobilizes these groups—not where migrants go, but which groups move more when climate shifts. Here we examine how the overall *volume* of migration responds to origin climate shocks, allowing the response to vary

across subsistence types.

Table 4 presents the results, estimating the baseline gravity specification separately for each subsistence type with the addition of the origin temperature shock ($\Delta T_{i,t}$). The results reveal a striking asymmetry: warming at the origin increases farmer migration (coefficient 0.86, $p < 0.01$) but decreases pastoral migration (-1.32, $p < 0.01$), with foragers showing no significant response. Table A16 in the Appendix confirms this pattern using continuous subsistence share interactions in the pooled sample.

These differential responses are consistent with climate shifts altering the comparative advantage of different subsistence strategies: warming expands the viable range for cereal cultivation, mobilizing farmer migration, while cooling shifts the ecological frontier in favor of grassland, favoring pastoral expansion. This interpretation resonates with the archaeological record. The Neolithic agricultural expansion across Europe broadly coincided with Holocene warming, which increased growing-degree days and expanded the cereal-viable zone (Richerson et al., 2001). Conversely, major steppe pastoral expansions—including the Yamnaya (3300–2500 BCE) and Sintashta/Andronovo (2100–1500 BCE)—occurred during periods of Neoglacial cooling and aridification (Anthony, 2007). Our results provide a quantitative mechanism for this pattern, consistent with Ricardian models of evolving comparative advantage under climate change (Costinot et al., 2016) and with evidence that reduced rainfall drives pastoral expansion into agricultural zones in contemporary settings (McGuirk and Nunn, 2025). The correspondence is not universal—some pastoral expansions occurred under warm conditions (Dong et al., 2023), and technological innovations such as dairying likely complemented climate (Wilkin et al., 2021)—but the broad pattern aligns with the differential climate sensitivities estimated here.

Figure 4 synthesizes these results by computing predicted migration responses over the Holocene. For each 200-year interval, we apply the estimated subsistence-specific coefficients to the realized Osman et al. (2021) temperature shocks at each origin gridcell, weighting by local subsistence shares.[5] Panel D plots the resulting predicted migration response. Farmer migration is predicted to peak during the Holocene Climatic Optimum (5500–4500 BCE), precisely when the LBK and Funnel Beaker expansions occurred (Panel C). Share-weighting implies that this peak reflects both the within-group climate sensitivity and the changing composition of subsistence types: farmer migration peaks not only because warming was strongest, but also because the share of farmers within the population was rising rapidly during this period. Pastoral migration is predicted to spike during Neoglacial

[5]We use coefficients from the split-sample PPML (Table 4) rather than the pooled interaction model because the split sample allows subsistence-group-specific fixed effects, avoiding the imposition of a common fixed-effect structure across regions with fundamentally different baseline migration patterns. Both approaches agree on signs and significance.

cooling episodes after 3000 BCE, coinciding with the Yamnaya, Sintashta, and Scythian expansions. The model thus recovers the broad temporal pattern of subsistence-specific migration from climate data alone.

8.2 Adaptive Sorting: Where Do Migrants Go?

Beyond differential mobilization, migration was also directionally responsive to climate shocks. We turn to the individual-level dataset, exploiting variation across ancestor–descendant pairs rather than aggregated gridcell flows. Figure 5 previews the result: binning pairs by the magnitude of climate change at the origin reveals a clear negative relationship—populations exposed to warming relocated toward cooler destinations, and vice versa.

For each pair, we construct two measures. The first captures *temperature change over time* at the origin: using Osman et al. (2021) reconstructions, we compute the temperature difference over a 400-year window centered on the ancestor's date, where positive values indicate warming. This window approximately spans the temporal gap between ancestor and descendant (100–500 years), capturing the climate trend experienced during the migration period. The second captures *temperature change over space*: the modern TerraClimate temperature difference between the descendant's and ancestor's locations (Abatzoglou et al., 2018), where negative values indicate the descendant settled somewhere cooler. If migration serves as an adaptive response, populations experiencing warming should relocate toward cooler destinations, implying $\beta < 0$ in:

$$\text{Temp Change Over Space}_{od} = \beta \text{ Temp Change Over Time}_{o} + \mathbf{X}'_{od}\gamma + \alpha_o + \alpha_d + \delta_t + \varepsilon_{od} \quad (7)$$

Appendix Table A15 presents individual-level OLS results. In the preferred specification with full fixed effects (column 3), $\hat{\beta} = -0.30$ (SE $= 0.09$): a 1°C increase in temperature at the origin over 400 years is associated with descendants settling in locations approximately 0.30°C cooler than their ancestors' origin. This implies that migration offset roughly one-third of the local warming experienced at the origin.

We complement these individual-level results with a gridded PPML specification that tests adaptive sorting within the gravity framework. We interact the origin temperature shock with the modern spatial temperature difference between destination and origin:

$$E[\text{Flow}_{ij,t}] = \exp\big(\alpha_i + \alpha_j + \delta_t + \beta_1 \Delta T_{i,t} + \beta_2 \Delta T_{i,t} \times (T_j - T_i) + \mathbf{X}'_{ij}\gamma\big) \quad (8)$$

A negative $\hat{\beta}_2$ indicates that warming at origin shifts flows toward cooler destinations. Table 5 reports our results. The interaction is negative and significant across all columns, ranging

from -0.063 in Column 1 to -0.138 in Column 3 ($p < 0.01$), and is stable to the inclusion of climate distance controls and further fixed effects. Column 4 splits the origin shock into warming and cooling components. The warming interaction is large and significant (-0.21, SE $= 0.06$), while the cooling interaction is smaller and marginally significant (-0.08, SE $= 0.05$, $p < 0.10$), consistent with warming shocks generating stronger adaptive responses.

The evidence thus far establishes that migration was ecologically directed (Section 5) and responsive to climate shocks—both in the subsistence groups mobilized and in the direction of flows. A remaining question is whether migrants carried their subsistence practices with them, reshaping the ecology of their destinations.

9 Ecological Diffusion of Agricultural Practices

If migrants carry subsistence-specific knowledge, they may reshape the ecology of their destinations. We test this using the REVEALS bias-corrected vegetation reconstruction (Section 3), examining whether genetic flow predicts subsequent changes in three land-cover variables—Cerealia-type pollen, grassland, and tree cover—at destination locations.

Table 6 examines whether genetic flow predicts destination pollen *change* as a function of the origin-destination difference. We estimate:

$$\Delta\text{Pollen}_{dt} = \beta_1 \text{Flow}_{od} + \beta_2 \text{Diff}_{odt} + \beta_3 (\text{Flow} \times \text{Diff})_{odt} + \beta_4 (\text{Distance} \times \text{Diff})_{od} + \mathbf{X}'_{od}\gamma + \alpha_o + \alpha_d + \delta_t + \varepsilon_{odt} \tag{9}$$

where ΔPollen_{dt} is the change in destination pollen from t to $t + 500$, Diff_{odt} is origin minus destination pollen at time t, and $\mathbf{X}_{od}$ includes climate and geography controls as well as contemporaneous temperature change at the destination from the Osman et al. (2021) paleoclimate reconstruction, ensuring that the results are not driven by climate-induced vegetation change. The key coefficient β_3 tests whether migration accelerates convergence: a positive value indicates that destinations receiving more migrants from origins with different vegetation change *toward* those origins' profiles. We include Distance $\times$ Diff to control for proximity-driven spatial spillovers, isolating the migration channel.

For all three pollen types, the coefficient on Flow$\times$Diff is positive and significant for grassland and tree cover ($p < 0.01$), and positive and marginally significant for cereals ($p < 0.10$) (Table 6, columns 1, 3, and 5). Destinations that receive more migrants from ecologically different origins subsequently shift *toward* those origins' vegetation profiles—beyond what geographic proximity alone would predict. This is consistent with demic diffusion: migrants physically carried agricultural knowledge and practices that reshaped the landscapes where they settled. The destination temperature change control is itself highly significant and car-

ries intuitive signs: warming is associated with increased cereal pollen and tree cover but decreased grassland, consistent with warmer conditions favoring both agriculture and forest expansion at the expense of open grassland (Marcott et al., 2013).

A natural concern is reverse causality: that migrants were attracted to destinations already converging toward their origin's ecological profile. We address this by controlling for the destination's prior pollen trajectory and its interaction with genetic flow (even-numbered columns). If reverse causality drives the result, the interaction of flow with prior destination change should be positive—migrants would flow toward destinations already on a convergent path. For cereals, this coefficient is significantly *negative*, directly inconsistent with reverse causality. The main diffusion effect actually strengthens after adding this control.

10 Conclusion

This paper provides the first population-level evidence on how climate similarity shaped human migration over the long run. Using identity-by-descent segments in ancient DNA to construct bilateral migration flows across Western Eurasia from 8000 BCE onward, we document that prehistoric populations moved systematically along corridors of ecological familiarity. Migration declined sharply with differences in growing degree days, precipitation, and soil characteristics between origin and destination—dimensions that together define the ecological niche relevant for subsistence. These patterns survive extensive controls for geographic distance, origin and destination fixed effects, and alternative specifications. We further show that migration responded adaptively to climate change—warming mobilized farmers while cooling mobilized pastoralists, and populations relocated toward destinations that offset local temperature shifts. Applying the estimated subsistence-specific coefficients to realized Holocene temperature shocks, the model recovers the broad temporal pattern of major archaeological migration episodes from climate data alone. We also find that genetic flow predicted subsequent changes in destination vegetation, consistent with the demic diffusion of agricultural practices. Four distinct findings—climate matching, subsistence heterogeneity, adaptive response, and ecological diffusion—emerge from one empirical framework and one dataset, and together paint a coherent picture of ecologically directed mobility in deep history.

The heterogeneity across subsistence systems is the paper's most distinctive finding. Farming populations exhibit strong climate matching along the GDD and soil dimensions—the ecological axes most relevant to crop production. Pastoralists show null thermal matching but strong precipitation matching, consistent with dependence on moisture for forage and grassland productivity. This crossover cannot be replicated in any modern dataset, because

contemporary populations do not vary in the degree to which their livelihoods depend on specific ecological dimensions. The pattern pins down the mechanism underlying climate matching: it is not a preference for familiar environments or a byproduct of distance decay, but the specificity of ecological knowledge. Farmers possessed detailed understanding of when to plant, which varieties to cultivate, and how to manage soils under particular thermal conditions—knowledge that depreciated sharply with climate distance. Pastoralists, whose herds required consistent moisture and suitable rangeland, matched on the precipitation dimension that determined forage availability. Foragers—whose landscape-indexed knowledge of local plant patches, game trails, and water sources depreciated with any geographic displacement—showed no significant climate matching but the steepest distance penalty of the three groups. The distance coefficients themselves reinforce the mechanism: each additional 100 km reduces pastoral flows by only 4%, compared with 18% for farmers and 32% for foragers, reflecting the mobility of productive assets. This complementary pattern across climate dimensions, subsistence types, and distance sensitivities is difficult to reconcile with explanations other than climate-specific human capital.

Our results also connect to a broader literature on the long-run origins of economic and cultural differentiation. Michalopoulos (2012) shows that ecological diversity fosters ethnolinguistic fragmentation by limiting population mixing, and Galor and Özak (2016) trace modern differences in time preference to the agricultural conditions of ancestral homelands. The ecological sorting we document—in which farming populations clustered in climatically similar regions over millennia—may be one origin of the spatial patterns of genetic and cultural distance that predict modern economic outcomes. If populations systematically sorted into ecological niches and accumulated locally adapted knowledge over thousands of years, the resulting population-climate match is itself a form of intangible capital that shapes comparative development. Climate change threatens to erode this match: as thermal regimes shift faster than populations can relocate or adapt, the accumulated correspondence between people and places built over millennia begins to break down (Obolensky et al., 2024). Understanding the origins of this correspondence—and the conditions under which it was maintained or disrupted—is relevant for assessing the long-run consequences of contemporary climate change.

Several limitations should be noted. First, paleoclimate reconstructions do not permit reliable measurement of growing-degree days at fine temporal resolution, which is why our adaptation analysis relies on annual mean temperature as a coarser proxy. Extending these results to GDD-based measures will require improved paleoclimate products that resolve within-year temperature variation. Second, our analysis is restricted to Western Eurasia, reflecting the current geographic coverage of ancient DNA datasets; as sequencing efforts

expand to other continents, it will be important to test whether these patterns generalize. Third, the adaptation coefficient—which we estimate at approximately -0.30 in the preferred individual-level specification, implying that migration offset roughly one-third of local warming—is confirmed by the gridded PPML analysis, where origin warming shocks significantly redirect bilateral flows toward cooler destinations, with warming shocks producing roughly two-and-a-half times the adaptive response of cooling shocks. Both estimates may partly reflect selection. We observe individuals whose genes survived to appear in the archaeological record, and successful migrants who relocated to suitable environments may be overrepresented relative to those who stayed and experienced declining fitness. Disentangling adaptive migration from differential survival remains an important challenge. Finally, the 500-year temporal resolution of our panel limits our ability to identify the precise mechanisms through which climate signals translated into migration decisions at shorter horizons.

For 10,000 years, humans moved along corridors of ecological familiarity. The knowledge accumulated in one climate zone—when to plant, what to grow, how to manage soils, how to survive winter—was a form of human capital that shaped the geography of populations. As growing seasons lengthened and agriculture spread, this ecological sorting intensified, binding farming populations ever more tightly to the thermal regimes they knew. The ancient DNA record preserves a trace of these movements, allowing us to observe directly what earlier generations of scholars could only infer from settlement patterns and crop assemblages. Understanding these deep patterns of ecologically directed mobility illuminates both the historical roots of today's population distribution and the likely human response to future climate change.

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.

Ammerman, Albert J. and L. Luca Cavalli-Sforza (1971) "Measuring the Rate of Spread of Early Farming in Europe," *Man*, 6 (4), 674–688.

——— (1984) *The Neolithic Transition and the Genetics of Populations in Europe*, Princeton, NJ: Princeton University Press.

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.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong (2016) "Skill Transferability, Migration, and Development: Evidence from a Quasi-Experiment," *American Economic Review*, 106 (9), 2658–2698.

Becker, Gary S. (1964) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, New York: Columbia University Press.

Borjas, George J. (1987) "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77 (4), 531–553.

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.

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.

Costinot, Arnaud, Dave Donaldson, and Cory Smith (2016) "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," *Journal of Political Economy*, 124 (1), 205–248.

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

Dong, Guang-Hui et al. (2023) "Prehistoric population expansion in Central Asia promoted by the Altai Holocene Climatic Optimum," *Nature Communications*, 14, 3701.

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.

Fyfe, Ralph M., Esther Githumbi, Anna-Kari Trondmann et al. (2021) "A full Holocene record of transient gridded vegetation cover in Europe," 10.1594/PANGAEA.937075, Dataset.

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.

Githumbi, Esther, Ralph Fyfe, Marie-José Gaillard et al. (2022) "European pollen-based REVEALS land-cover reconstructions for the Holocene: methodology, mapping and potentials," *Earth System Science Data*, 14 (4), 1581–1619, 10.5194/essd-14-1581-2022.

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

d'Alpoim Guedes, Jade and R. Kyle Bocinsky (2018) "Climate change stimulated agricultural innovation and exchange across Asia," *Science Advances*, 4 (10), eaar4491, 10.1126/sciadv.aar4491.

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.

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.

Hutchinson, G. Evelyn (1957) "Concluding Remarks," *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415–427, 10.1101/SQB.1957.022.01.039.

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.

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

McGuirk, Eoin and Nathan Nunn (2025) "Transhumant Pastoralism, Climate Change, and Conflict in Africa," *Review of Economic Studies*, 92 (1), 404–441.

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

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.

Odling-Smee, F. John, Kevin N. Laland, and Marcus W. Feldman (2003) *Niche Construction: The Neglected Process in Evolution*, Princeton, NJ: Princeton University Press.

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.

Pinhasi, Ron, Joaquim Fort, and Albert J. Ammerman (2005) "Tracing the Origin and Spread of Agriculture in Europe," *PLoS Biology*, 3 (12), e410, 10.1371/journal.pbio.0030410.

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.

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.

Roy, A. D. (1951) "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3 (2), 135–146.

Sala, Osvaldo E., William J. Parton, Linda A. Joyce, and William K. Lauenroth (1988) "Primary production of the central grassland region of the United States," *Ecology*, 69 (1), 40–45, 10.2307/1943158.

Silva, JMC Santos and Silvana Tenreyro (2006) "The log of gravity," *The Review of Economics and statistics*, 641–658.

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.

Sugita, Shinya (2007) "Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition," *The Holocene*, 17 (2), 229–241, 10.1177/0959683607075837.

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

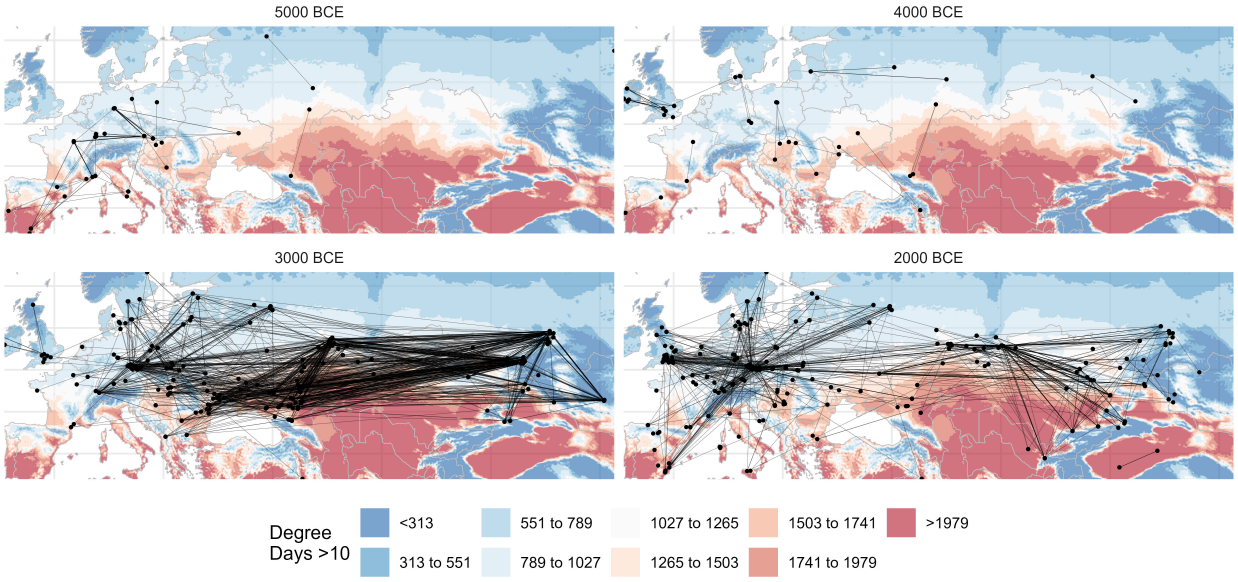
White, Lynn (1962) *Medieval Technology and Social Change*: Oxford University Press.

Wilkin, Shevan et al. (2021) "Dairying enabled Early Bronze Age Yamnaya steppe expansions," *Nature*, 598, 629–633.

Xu, Chi, Timothy A. Kohler, Timothy M. Lenton, Jens-Christian Svenning, and Marten Scheffer (2020) "Future of the Human Climate Niche," *Proceedings of the National Academy of Sciences*, 117 (21), 11350–11355.

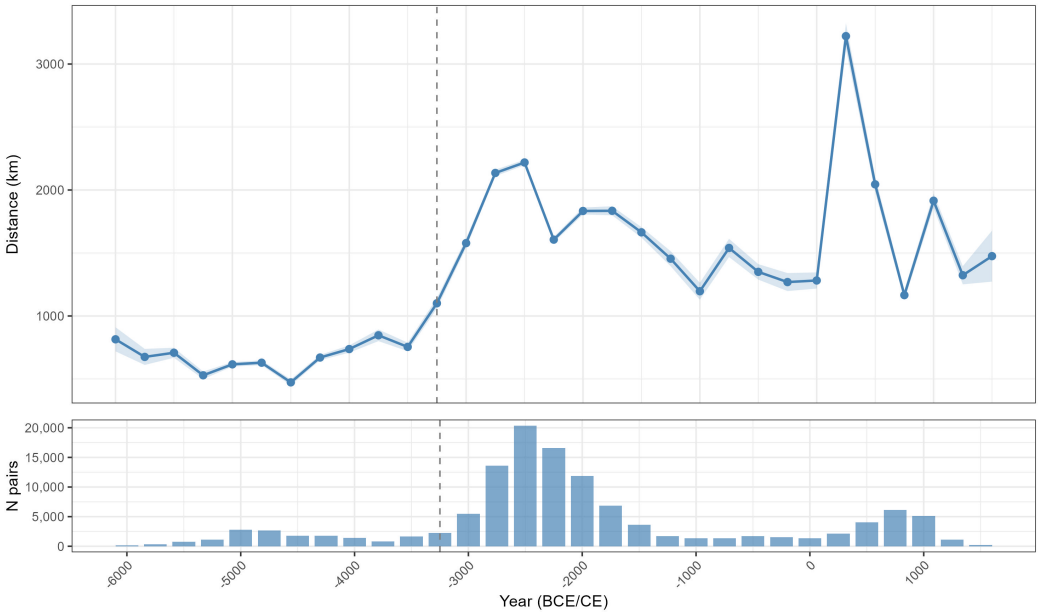
Figures

Figure 1: Flows of Individuals



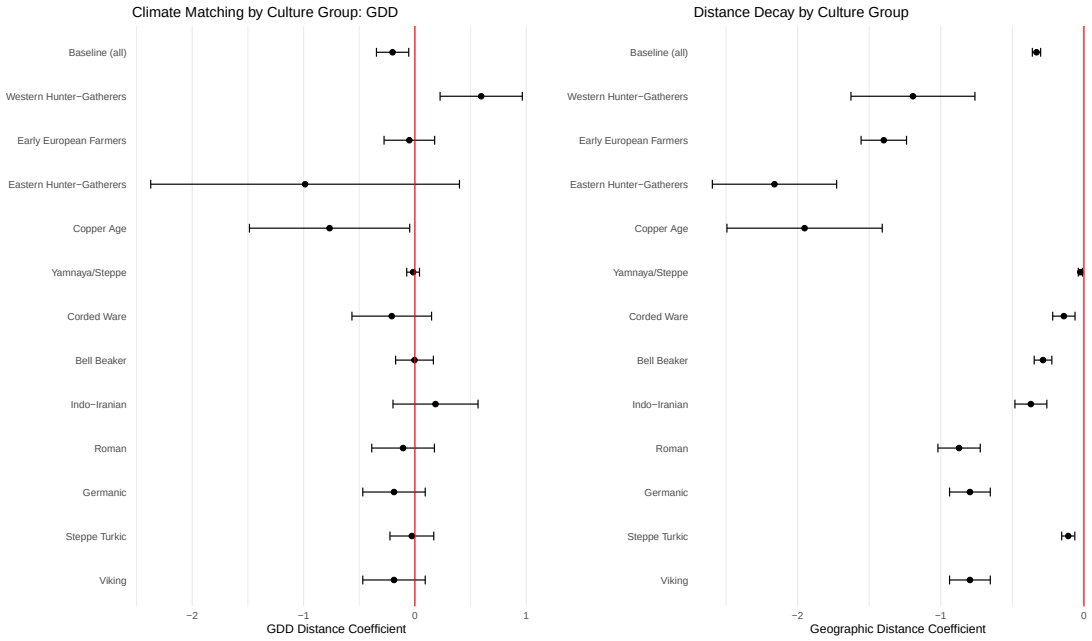
Notes: Each line connects a pair of individuals sharing at least one IBD segment greater than 16 cM, indicating a common ancestor within approximately 6–10 generations. The sample is restricted to pairs living at most 500 years apart and separated by at least 500 km. Panels show four 1,000-year periods (labeled by BCE midpoint).

Figure 2: Average Distance Between Related Individuals Over Time



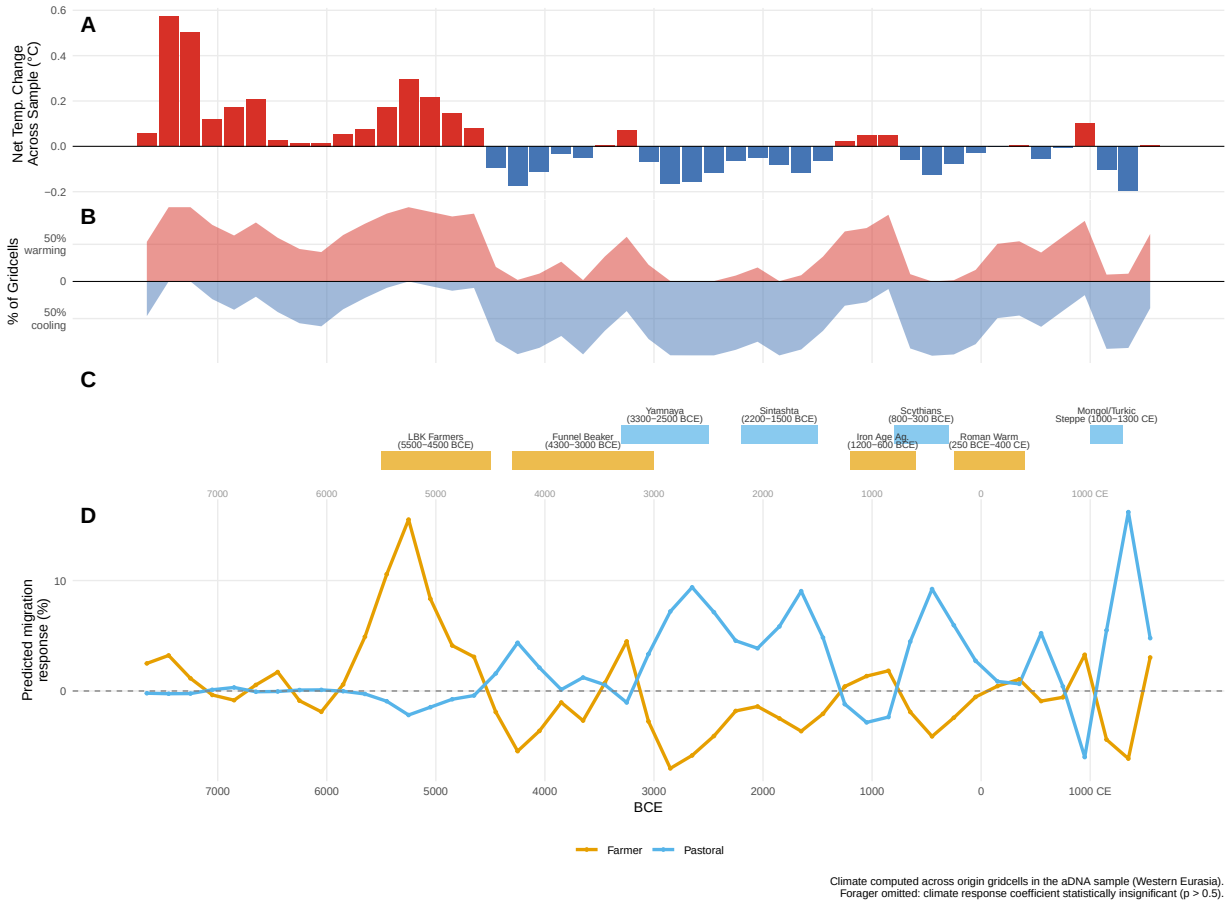
Notes: The figure plots the average geographic distance (in kilometers) between genetically related individuals over time, computed at 250-year intervals. Individuals are classified as related if they share at least one IBD segment of 8 cM or longer. The vertical dashed line marks approximately 3250 BCE, corresponding to the adoption of wheel and wagon technology in the Pontic-Caspian steppe region.

Figure 3: Climate Matching and Distance Sensitivity by Culture Group



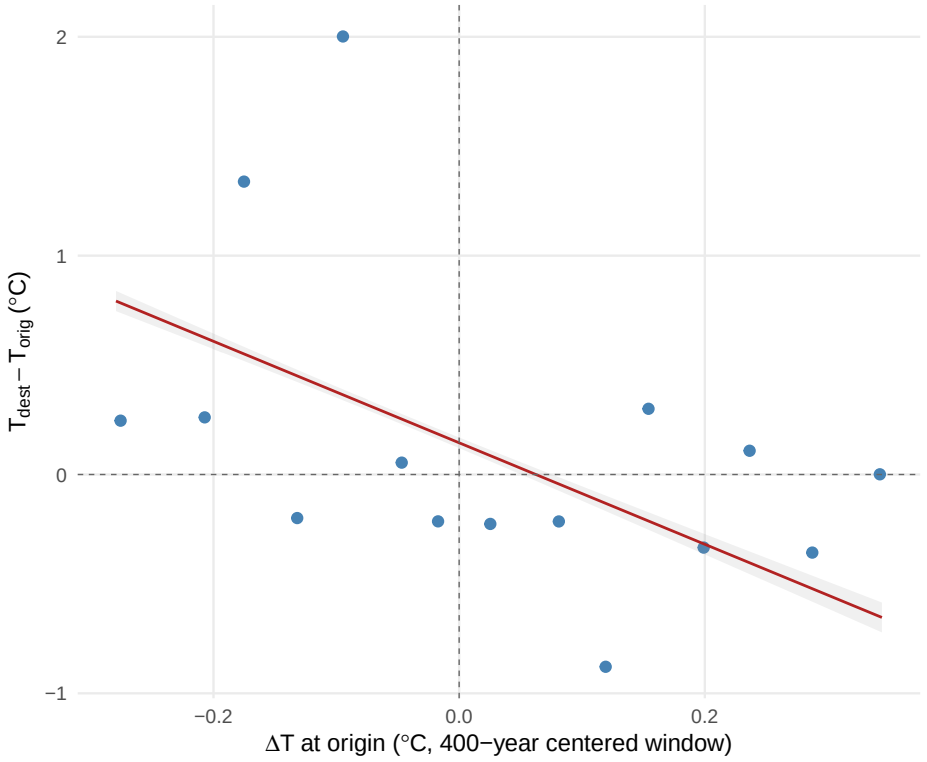
Notes: Coefficient plots from the baseline gravity specification estimated separately for each of 14 named archaeological culture groups. Left panel: GDD distance coefficient (more negative = stronger thermal matching). Right panel: geographic distance coefficient (more negative = steeper distance penalty). Dashed line shows the pooled baseline estimate. Culture groups ordered by geographic distance coefficient.

Figure 4: Predicted Migration Response by Subsistence Type Over the Holocene



Notes: Panel A shows net temperature change (200-year intervals) averaged across origin gridcells in the aDNA sample, from Osman et al. (2021) reconstructions. Panel B shows the share of gridcells experiencing warming (red) vs. cooling (blue). Panel C marks major archaeological migration events, colored by subsistence type: farmer (gold) and pastoral (blue). Panel D plots the predicted migration response, computed by applying the estimated subsistence-specific climate coefficients (farmer: 0.856; pastoral: −1.32) to realized temperature shocks at each origin, weighted by local subsistence shares. Forager omitted (coefficient statistically insignificant).

Figure 5: Adaptive Migration: Origin Climate Change vs. Spatial Direction of Movement



Notes: The figure plots the relationship between climate change at the origin (x-axis) and the spatial direction of migration in temperature space (y-axis). The x-axis measures the 400-year centered temperature change at the origin location using Osman et al. (2021) reconstructions ($T_{t-200} - T_{t+200}$, where t is the ancestor's date in years BP; positive values indicate warming). The y-axis measures the modern temperature difference between destination and origin ($T_{\text{dest}} - T_{\text{orig}}$). The 400-year climate window approximately spans the temporal gap between ancestor–descendant pairs (100–500 years), capturing the climate trend experienced during the migration period. Each point represents the IBD-weighted mean within one of 15 equal-width bins, with observations winsorized at the 2.5th and 97.5th percentiles. The red line shows a weighted linear fit to the underlying microdata. The negative slope indicates that populations exposed to warming systematically relocated toward cooler destinations, partially offsetting local temperature shifts.

Tables

Table 1: Summary Statistics

	Mean	SD	Min	P10	Median	P90	Max	N
IBD Count (8cM)	0.181	3.054	0.000	0.000	0.000	0.000	1076.000	879007
IBD Length (12cM)	0.079	1.868	0.000	0.000	0.000	0.000	951.922	879007
IBD Length (8cM)	0.273	2.889	0.000	0.000	0.000	0.000	966.465	879007
GDD Dist.	0.787	0.705	0.000	0.104	0.613	1.700	5.909	872565
Precipitation Dist. (m)	0.313	0.303	0.000	0.040	0.231	0.676	2.677	878870
Distance (km)	3.195	2.130	0.042	0.822	2.650	6.318	12.869	879007
Soil Dist.	1.034	0.888	0.000	0.142	0.797	2.270	7.824	872565
Elevation Dist.	0.742	0.958	0.000	0.041	0.383	1.789	5.096	878870
Summer Temp. Dist.	4.968	4.049	0.000	0.704	4.028	10.594	29.789	872565
Winter Temp. Dist.	9.733	7.490	0.000	1.303	8.068	20.514	54.427	872565

Notes: Unit of observation is a grid cell pair-era. IBD Count is the number of IBD segments $\geq$ 8cM shared between individuals in the origin and destination grid cells. Climate and geographic distances are absolute differences between origin and destination. GDD Dist. is in 1000s of degree-days. Distance is in 1000s of km.

Table 2: Environmental Distance and Prehistoric Migration

Dependent Variable:	Genetic Flow					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-0.855***	-0.830***	-0.201***	-0.200***	-0.173**	-0.122***
	(0.051)	(0.052)	(0.075)	(0.074)	(0.071)	(0.038)
Precip. Dist.	-0.701***	-0.596***	-0.670***	-0.667***	-0.623***	-0.798***
	(0.108)	(0.108)	(0.112)	(0.109)	(0.110)	(0.067)
Soil Composite Dist.	-0.208***	-0.246***	-0.351***	-0.352***	-0.354***	-0.329***
	(0.033)	(0.037)	(0.056)	(0.056)	(0.055)	(0.034)
Distance (km)	-0.425***	-0.406***	-0.327***	-0.331***	-0.347***	-0.355***
	(0.022)	(0.022)	(0.015)	(0.015)	(0.014)	(0.011)
Fixed-effects						
Era		Yes		Yes		
Orig. gridcell			Yes	Yes		
Dest. gridcell			Yes	Yes	Yes	
Orig. gridcell-Era					Yes	Yes
Dest. gridcell-Era						Yes
Fit statistics						
Observations	872,565	872,565	835,786	835,786	739,412	420,550
Pseudo R^2	0.11	0.18	0.45	0.50	0.61	0.75

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

Notes: PPML estimates. Dependent variable: IBD segment count ($\geq$ 8 cM) between 1-degree gridcell pairs. GDD Dist. = $|\text{GDD}_{10,o} - \text{GDD}_{10,d}|/1000$. Distance in thousands of km. Soil Composite Dist. is the Euclidean distance on standardized clay content and soil organic carbon — properties determined primarily by geological substrate rather than current climate. Column (4) is the preferred specification. Columns progressively add fixed effects. Standard errors clustered by gridcell pair. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Climate Matching by Subsistence Type: Split Sample

Dependent Variable:	Genetic Flow				
	All	Farmer	Agro-Past.	Pastoral	Forager
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
GDD Dist.	-0.142***	-0.183***	-0.496***	-0.003	-0.069
	(0.052)	(0.058)	(0.069)	(0.072)	(0.142)
Precip. Dist.	-0.203***	-0.008	-0.113**	-0.223***	0.054
	(0.033)	(0.039)	(0.045)	(0.066)	(0.184)
Soil Composite Dist.	-0.243***	-0.225***	-0.061*	-0.096**	-0.088
	(0.039)	(0.045)	(0.033)	(0.042)	(0.079)
Distance (km)	-0.705***	-1.96***	-0.796***	-0.375***	-3.91***
	(0.032)	(0.096)	(0.052)	(0.037)	(0.255)
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	820,528	393,221	56,254	252,618	26,397
Pseudo R^2	0.50	0.53	0.57	0.61	0.51
Regressors standardized	Yes	Yes	Yes	Yes	Yes

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

Notes: PPML estimates. Dependent variable: IBD segment count ($\geq$ 8 cM). All regressors standardized within each subsample (mean zero, unit SD) for cross-variable comparability. Each column restricts the sample to origin gridcell-eras classified as the indicated subsistence type (GPT-based archaeological classification). Agro-pastoral origins are classified as mixed farmer-pastoralist. Origin gridcell, destination gridcell, and era FE. Pair-clustered SEs. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Climate Niche Expansion by Subsistence Type

Dependent Variable:	Genetic Flow				
	All	Farmer	Agro-Past.	Pastoral	Forager
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Temp. Change (origin)	1.42***	0.856***	-0.975**	-1.32***	0.312
	(0.216)	(0.265)	(0.474)	(0.437)	(0.514)
Distance (km)	-0.324***	-0.888***	-0.400***	-0.190***	-1.59***
	(0.015)	(0.045)	(0.026)	(0.018)	(0.128)
Climate dist. controls	Yes	Yes	Yes	Yes	Yes
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	800,588	384,816	56,107	252,526	16,978
Pseudo R^2	0.51	0.54	0.57	0.61	0.52

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

Notes: PPML estimates. Dependent variable: IBD segment count ($\geq$ 8 cM). Temp. Change (origin) is the Osman et al. (2021) reconstructed temperature change (positive = warming). Each column restricts the sample to origin gridcell-eras classified as the indicated subsistence type (GPT-based archaeological classification). Distance in thousands of km. Climate distance controls: GDD, precipitation, and soil composite distance. Origin, destination, and era FE. Pair-clustered SEs. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Adaptive Migration: Origin Climate Shocks and Destination Temperature

Dependent Variable:	Genetic Flow			
Model:	(1)	(2)	(3)	(4)
Variables				
Temp. Change (origin)	0.547***	1.58***	1.59***	
	(0.175)	(0.213)	(0.212)	
Temp. Change × Temp. Diff.	-0.063***	-0.130***	-0.138***	
	(0.019)	(0.023)	(0.023)	
Warming (origin)				0.059
				(0.516)
Warming × Temp. Diff.				-0.213***
				(0.057)
Cooling (origin)				2.52***
				(0.356)
Cooling × Temp. Diff.				-0.084*
				(0.050)
Distance (km)	-0.390***	-0.395***	-0.326***	-0.327***
	(0.015)	(0.015)	(0.015)	(0.015)
Climate dist. controls	No	No	Yes	Yes
Fixed-effects				
Orig. gridcell	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes
Era		Yes	Yes	Yes
Fit statistics				
Observations	815,332	815,332	815,332	815,332
Pseudo R^2	0.45	0.50	0.51	0.51

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

Notes: PPML estimates. Dependent variable: IBD segment count ($\geq$ 8 cM) between 1-degree gridcell pairs. All columns include origin and destination fixed effects. Temp. Change (origin) is the Osman et al. (2021) reconstructed temperature change over a centered 400-year window. Temp. Diff. is the modern ERA5 temperature difference (destination minus origin). A negative interaction indicates warming at origin shifts flows toward cooler destinations. Column (4) splits the shock into warming ($\max(\Delta T, 0)$) and cooling ($\min(\Delta T, 0)$) components. Climate distance controls: GDD distance, precipitation distance, and soil composite distance. Standard errors clustered by gridcell pair. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Ecological Diffusion of Subsistence Practices

	Δ Cereal Pollen		Δ Grassland Pollen		Δ Tree Pollen	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Genetic Flow (o–d, t-500 to t)	-0.064***	-0.130***	0.416	0.420***	-0.327	-0.356***
	(0.021)	(0.026)	(0.296)	(0.134)	(0.284)	(0.129)
Pollen Dist. ($o - d$, t)	21.0***	24.8***	8.16***	11.8***	8.75***	12.0***
	(0.380)	(0.453)	(0.133)	(0.155)	(0.141)	(0.165)
Genetic Flow × Pollen Dist.	0.099*	0.114**	0.063***	0.027***	0.066***	0.032***
	(0.056)	(0.050)	(0.016)	(0.009)	(0.017)	(0.010)
Distance × Pollen Dist.	0.173*	0.040	0.124***	0.043*	0.108***	0.032
	(0.096)	(0.090)	(0.021)	(0.023)	(0.022)	(0.024)
Pollen Change (d, t-500 to t)		11.4***		21.1***		19.7***
		(0.536)		(0.302)		(0.278)
Genetic Flow × Pollen Change		-0.201***		0.215***		0.173***
		(0.071)		(0.051)		(0.046)
Temp. Change (dest.)	4.48***	7.82***	-77.3***	-67.2***	96.7***	91.5***
	(0.443)	(0.556)	(2.47)	(3.32)	(2.51)	(3.43)
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	331,007	210,437	331,007	210,437	331,007	210,437
R^2	0.563	0.581	0.461	0.539	0.518	0.583
Climate & geography controls	Yes	Yes	Yes	Yes	Yes	Yes

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

Notes: The dependent variable is the change in destination pollen from t to $t + 500$, scaled ×100. "Cereal" is Cerealia-type pollen from the REVEALS vegetation reconstruction. Pollen Dist. = Origin pollen minus Destination pollen at time t. A positive coefficient on Genetic Flow × Pollen Dist. indicates that genetic flow drives destination vegetation toward the origin's profile beyond what geographic proximity predicts. Even-numbered columns add the destination's prior pollen change (t-500 to t) and its interaction with Genetic Flow as a reverse-causality control; if migrants were attracted to destinations already converging, Genetic Flow × Prior Change should be positive. Climate and geography controls (GDD distance, precipitation distance, soil composite distance, geographic distance and its square, and destination temperature change from Osman et al. (2021)) included in all columns. Origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by origin–destination pair.

Appendix A - Additional Tables and Figures

Appendix B discusses the spatial coverage and preservation characteristics of the ancient DNA dataset used in this paper.

The following appendix tables document the robustness of the main results:

- Table A1: Implied migration speed by era
- Table A2: Soil component decomposition
- Table A3: Environmental correlates of ancient DNA remains (Appendix B)
- Table A4: Migration distance by subsistence type
- Table A5: Subsistence heterogeneity: continuous share interactions (gridded)
- Table A6: Subsistence heterogeneity (individual)
- Table A7: 1000-year period analysis (gridded)
- Table A8: Additional environmental controls (diurnal range, elevation, ruggedness, coastal distance)
- Table A9: Alternative dependent variables (mean/sum segment lengths, 8 cM/12cM)
- Table A10: Individual-level main results
- Table A11: Distance band stratification (gridded)
- Table A12: Sample robustness (horizontal, vertical, $\geq$100km)
- Table A13: Fixed effects and inference robustness (interacted FE, broader spatial clustering)
- Table A14: Alternative distance specifications (linear, quadratic, log, travel cost)
- Table A15: Individual-level climate adaptation (OLS)
- Table A16: Subsistence heterogeneity in adaptation (continuous interactions)
- Table A17: Main results at 2.5-degree resolution
- Table A18: Alternative climate measures (modern, marine proxy, Beyer, pair FE)
- Table A19: Ecological diffusion excluding pairs within 500km

A Extended Conceptual Framework

This appendix provides additional discussion of the conceptual framework presented in Section 2. We elaborate on the ecological foundations, the connection to existing models in economics, and the mechanisms underlying each testable prediction.

A.1 Ecological Foundations

The idea that populations occupy climate niches defined by their accumulated ecological knowledge connects to a long tradition in ecology. Hutchinson (1957) defined the fundamental niche as the set of environmental conditions under which a species can persist; we adapt this concept to human populations, where the niche is endogenous—shaped by the subsistence technology in use. Farmers occupy narrow thermal niches because their crop varieties, planting calendars, and pest management practices are calibrated to local growing-season conditions. The dimensions along which other subsistence groups are most constrained—and therefore along which matching should be strongest—is an empirical question that we address in Section 7. Migration along corridors of environmental similarity is the mechanism through which populations track their niches across space.

The key departure from ecological models is that human niche breadth is not fixed by genetics but is modified by subsistence technology—a form of niche construction (Odling-Smee et al., 2003) that endogenously reshapes the dimensions along which ecological matching operates. Recent work estimates that a substantial share of the global population may be pushed outside the historical human climate niche by the end of this century (Xu et al., 2020), underscoring the relevance of understanding the mechanisms by which populations have tracked their niches in the past.

A.2 Connection to the Roy Model and Location Choice

The location-choice formulation in Equation 2 embeds a mechanism in the spirit of Roy (1951): individuals select into locations where their particular bundle of climate-specific knowledge yields the highest returns, net of moving costs. The structure parallels Borjas (1987)'s application of the Roy model to immigrant self-selection, where workers sort across countries based on the returns to their skills, and extends it to a setting in which the relevant "skill" is ecological rather than occupational. We abstract from within-origin heterogeneity—all individuals from a given location inherit the same knowledge stock—so the model is best understood as a location-choice framework rather than a full selection model. The key friction is not a border or a wage differential but the transferability of subsistence knowledge across climatic gradients.

Our approach also relates to Bazzi et al. (2016), who model agroclimatic similarity as a measure of location-specific human capital in the context of Indonesia's Transmigration Program. We extend this framework by allowing mismatch penalties to differ across climate dimensions and subsistence technologies. The framework connects naturally to the concept of climate-specific human capital developed in Obolensky et al. (2024) and to the broader literature on Ricardian comparative advantage in agriculture (Costinot et al., 2016). Just as modern farmers specialize in crops suited to local conditions, generating patterns of trade that reflect underlying climatic endowments, prehistoric populations accumulated knowledge stocks whose value was tied to specific ecological niches. Migration in this framework is a form of spatial reallocation governed by the portability of those knowledge stocks.

A.3 Subsistence Heterogeneity: Extended Discussion

Farmers. Farming depends directly on accumulated thermal time: which crops to plant, when to sow, when to harvest, and how to manage pests are all tightly linked to local growing-season conditions. A farmer who has spent generations calibrating planting schedules to a 1,500 degree day environment faces substantial losses if transplanted to a 2,500 degree day environment, where different crops, different pest pressures, and different seasonal rhythms prevail. At the same time, farming populations developed technologies—crop selection, manuring, and water management—that partially decouple productivity from local precipitation. The effect is to reduce α_P relative to its value absent technology, concentrating the remaining vulnerability on the thermal and edaphic dimensions: thermal conditions determine what can be grown, while soil workability—clay content, organic matter, drainage—determines where it can be planted. Neither constraint is easily buffered by Neolithic-era techniques.

Pastoralists. Livestock tolerate a wider range of temperatures than any single crop variety, relaxing the thermal binding constraint. But forage availability—grassland productivity—is governed primarily by precipitation (Sala et al., 1988; McGuirk and Nunn, 2025), and pastoral production systems operate at spatial scales too large for soil amendment to be feasible. The result is that pastoralists are most exposed along the moisture dimension, with secondary sensitivity to soil conditions that influence rangeland carrying capacity.

Foragers. Foragers rely on broad ecological knowledge—awareness of edible plants, game patterns, water sources—that may transfer more readily across both thermal and hydrological gradients. Their subsistence strategies are diverse and opportunistic, drawing on a wide portfolio of resources rather than specializing in a narrow ecological niche. This generalism may reduce foragers' sensitivity to any single environmental dimension, making them the least constrained group overall.

A.4 Transport Costs and the Scope of Matching

The interaction between climate distance and transport costs shapes the equilibrium pattern of migration in an important way. When transport costs τ are high, individuals can only reach nearby destinations, and observed migration flows are concentrated at short distances regardless of climate matching. As transport costs fall—through technological innovations such as the domestication of pack animals or the adoption of wheeled vehicles—migrants gain access to a wider set of destinations. With more destinations to choose among, the scope for climate-based sorting increases: individuals can bypass nearby but climatically dissimilar locations in favor of more distant destinations that better match their ecological niche.

This mechanism generates a prediction about how the strength of observed climate matching should evolve over time. If a reduction in transport costs expands the effective choice set, then conditional on geographic distance, climate similarity should become a stronger predictor of migration after the cost reduction. In our empirical setting, the most prominent candidate for such a shift is the adoption of wheel and wagon technology around 3250 BCE

(Anthony, 2007), which is associated with a dramatic expansion in the geographic reach of migration (Figure 2).

A.5 Knowledge Accumulation and Relocation Costs

The observed migration response to climate change is governed not by α_T alone but by the ratio α_T/κ_s—the sensitivity to climate mismatch relative to the cost of relocating. The low relocation costs of pastoralists—whose herds walk to new pasture and whose social organization is built around seasonal movement—may amplify the pastoral response relative to farmers, who face higher costs from fixed agricultural investments.

This prediction relates to a broader point emphasized in the cultural evolution literature: that the accumulation of locally adapted knowledge is slow and costly (Henrich, 2004; Boyd et al., 2011). The knowledge required for successful subsistence in a particular environment reflects generations of trial and error, and cannot be easily recreated in a new climatic regime. Migration is therefore not merely a response to reduced productivity at the origin; it is a strategy for preserving the value of an inherited knowledge stock when the local environment shifts beneath it. But the availability of this strategy depends on relocation costs, which vary with the mobility of productive assets and the degree of fixed investment in the landscape.

A.6 Ecological Diffusion

If individuals bring climate-specific knowledge $k(\boldsymbol{\theta}_i)$ to their destination, and if that knowledge shapes the ecology of the destination—through the introduction of crops, livestock, or land management practices—then migration should drive ecological convergence between origins and destinations. Specifically, destinations that receive migrants from ecologically different origins should experience changes in vegetation, crop assemblages, or land cover that move the destination's ecological profile toward that of the origin.

This prediction distinguishes our framework from models in which migration is a purely demographic process. In a standard gravity model of migration (Anderson and Van Wincoop, 2003; Silva and Tenreyro, 2006), bilateral flows depend on distance and on origin and destination characteristics, but migrants do not transform the destinations they enter. In our framework, because migrants carry ecologically productive knowledge, they reshape the biotic environment at the destination. This mechanism is closely related to the concept of demic diffusion in archaeology and population genetics (Ammerman and Cavalli-Sforza, 1984), in which the spread of agriculture is driven by the movement of farming populations rather than the diffusion of ideas alone.

Table A1: Implied Migration Speed by Era

Era	N	Dist. (km)	Time gap (yr)	Speed (km/yr)	km/generation
Panel A: By era					
7,000 BCE	85	475	475	0.75	19
6,000 BCE	979	601	591	1.18	30
5,000 BCE	7,720	523	489	1.05	26
4,000 BCE	5,807	552	500	1.15	29
3,000 BCE	32,179	1,649	390	4.11	103
2,000 BCE	47,885	1,337	475	2.99	75
1,000 BCE	6,606	1,160	593	2.14	53
1 BCE/CE	7,773	1,206	560	2.64	66
1,000 CE	14,824	825	300	2.21	55
2,000 CE	44	1,325	365	3.34	84
All	123,902	1,155	442	2.72	68
Panel B: Literature benchmarks					
Neolithic demic diffusion				1.00	25
Neolithic expansion, aDNA				1.80	45
Yamnaya expansion, aDNA				4.20	105

Notes: Each row in Panel A reports statistics for all IBD-linked pairs (sharing at least one segment $\geq$8 cM) assigned to the given era. Eras are defined by rounding the midpoint date of each pair to the nearest millennium. Time gaps are restricted to 100–1,000 years. Speed is computed as distance divided by time gap; "km per generation" assumes a 25-year generation. Panel B reports population-level wave-of-advance estimates from the literature for comparison. The wave-of-advance measures the speed at which a population frontier expands, which is conceptually distinct from the net displacement of genetic material between IBD-linked pairs; the close correspondence nonetheless suggests that IBD-inferred flows reflect the same underlying population movements. Sources: Ammerman and Cavalli-Sforza (1971); Racimo et al. (2020).

B Dataset: Spatial Coverage and Preservation

The ancient DNA dataset used in this paper is derived from Ringbauer et al. (2024), who apply the ancIBD algorithm to a large collection of ancient genomes from Western Eurasia. As with any archaeological dataset, the spatial distribution of observed remains reflects a combination of historical population patterns and taphonomic processes—the physical and chemical conditions governing whether biological material survives over millennia. Understanding these patterns is important for interpreting the data and for researchers in economics and social science who may be less familiar with the characteristics of ancient DNA samples.

Table A3 reports results from regressing an indicator for the presence of at least one observed individual in a $1° \times 1°$ gridcell on environmental and climatic characteristics.[6] Several patterns emerge that are consistent with known preservation mechanisms in the ancient DNA literature.

Temperature and heat exposure. Cumulative GDD above 10°C is negatively associated with the presence of remains (column 1), indicating that samples are more likely to survive and be recovered in cooler environments. This reflects the well-documented relationship between thermal exposure and DNA degradation: higher cumulative heat accelerates the hydrolysis and oxidation of nucleotides, reducing the probability that recoverable ancient DNA persists over thousands of years (Reich, 2018). Conditional on cumulative heat, winter temperature enters with a small positive coefficient, suggesting that mild winters—which may correlate with human settlement density and burial practices—do not independently harm preservation.

Soil characteristics. Soil pH and clay content are both positively associated with remains (columns 2–4). Alkaline soils slow the acid-driven hydrolysis of DNA, while clay minerals adsorb and stabilize organic molecules, reducing microbial access. Soil organic content, by contrast, enters negatively: organic-rich soils support greater microbial activity, which accelerates biological decay. These patterns are broadly consistent with findings in the archaeogenomics literature on the determinants of DNA survival in burial contexts.

Moisture and aridity. Precipitation is negatively associated with the presence of remains (column 5), consistent with the role of moisture in promoting microbial activity and chemical degradation. Diurnal temperature range—a proxy for aridity and continentality—enters positively (column 6), reinforcing the pattern that drier, more continental environments favor long-run preservation.

Elevation. Remains are less likely to be observed at higher altitudes (column 6), consistent with both lower historical population density in mountainous terrain and the logistical challenges of archaeological recovery at elevation.

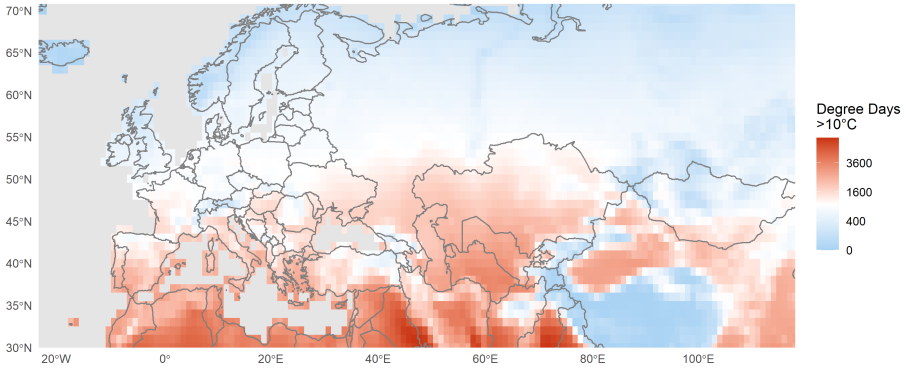
[6] Standard errors are clustered at the $2.5° \times 2.5°$ gridcell level.

Within-region variation. Column 7 adds fixed effects at the $2.5° \times 2.5°$ level, restricting identification to within-region variation. Cumulative GDD, soil pH, clay content, and diurnal temperature range remain significant, indicating that the preservation correlates documented above operate at fine spatial scales and are not driven solely by broad regional gradients.

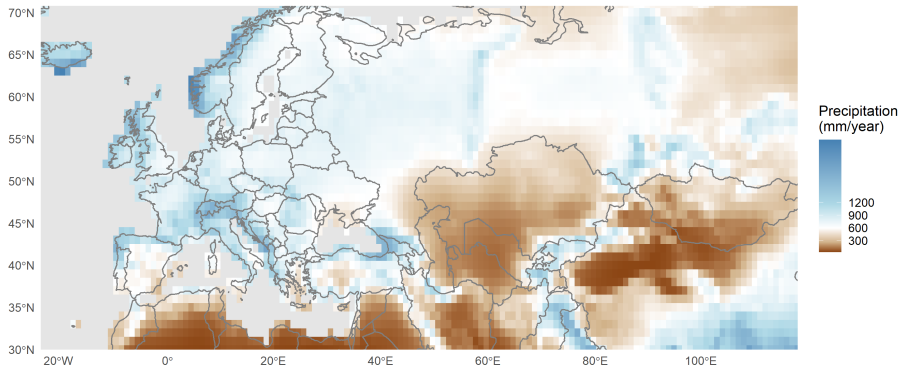
These patterns have two implications for our analysis. First, they underscore that the ancient DNA sample is not a random draw from the historical population: cooler, drier, clay-rich locations are systematically overrepresented. Second, and more importantly for identification, the correlates of preservation are location-specific characteristics that are fully absorbed by the origin and destination fixed effects in our bilateral migration regressions. Our identification exploits variation across destination locations for the same origin, comparing migration flows to climatically similar versus dissimilar destinations—a margin that is orthogonal to the level of DNA preservation at any single location. Moreover, our dependent variable measures bilateral genetic connections between pairs of locations, not the mere presence of remains at a single location. Preservation-driven selection into the sample affects which locations appear in the data, but conditional on a location being observed, the pattern of bilateral flows across destinations is the variation we exploit.

Figure A1: Modern Climate Variables

(a) Growing Degree Days above 10°C

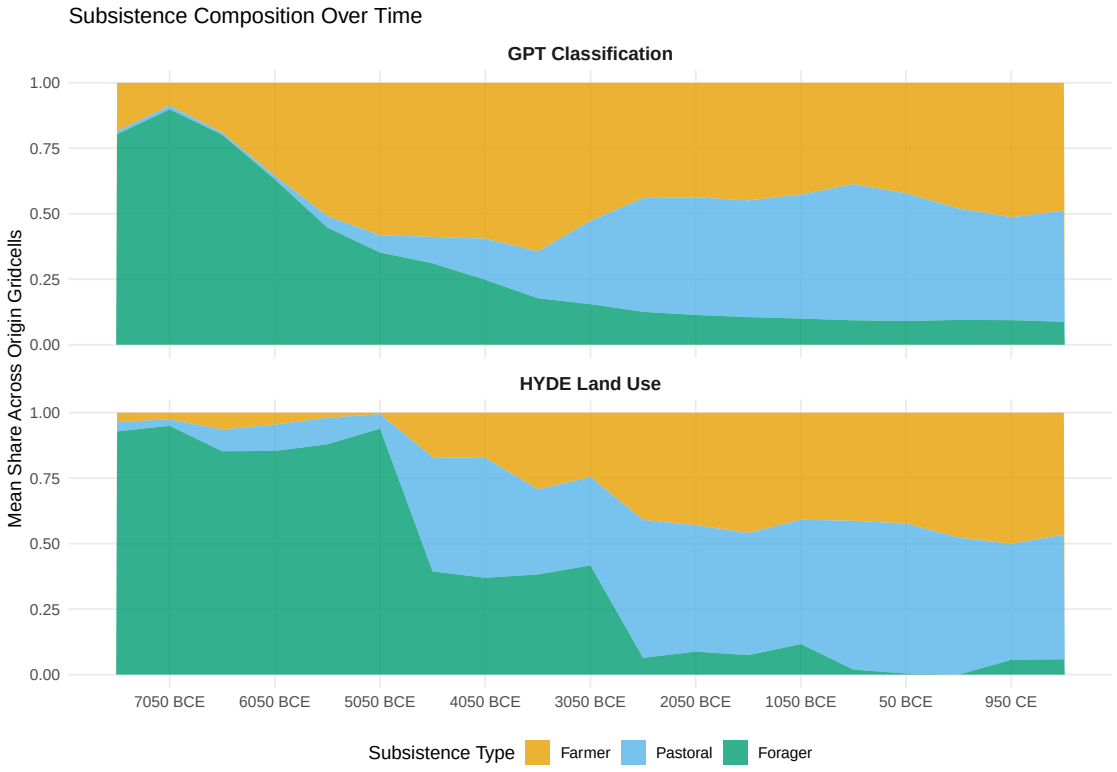


(b) Precipitation



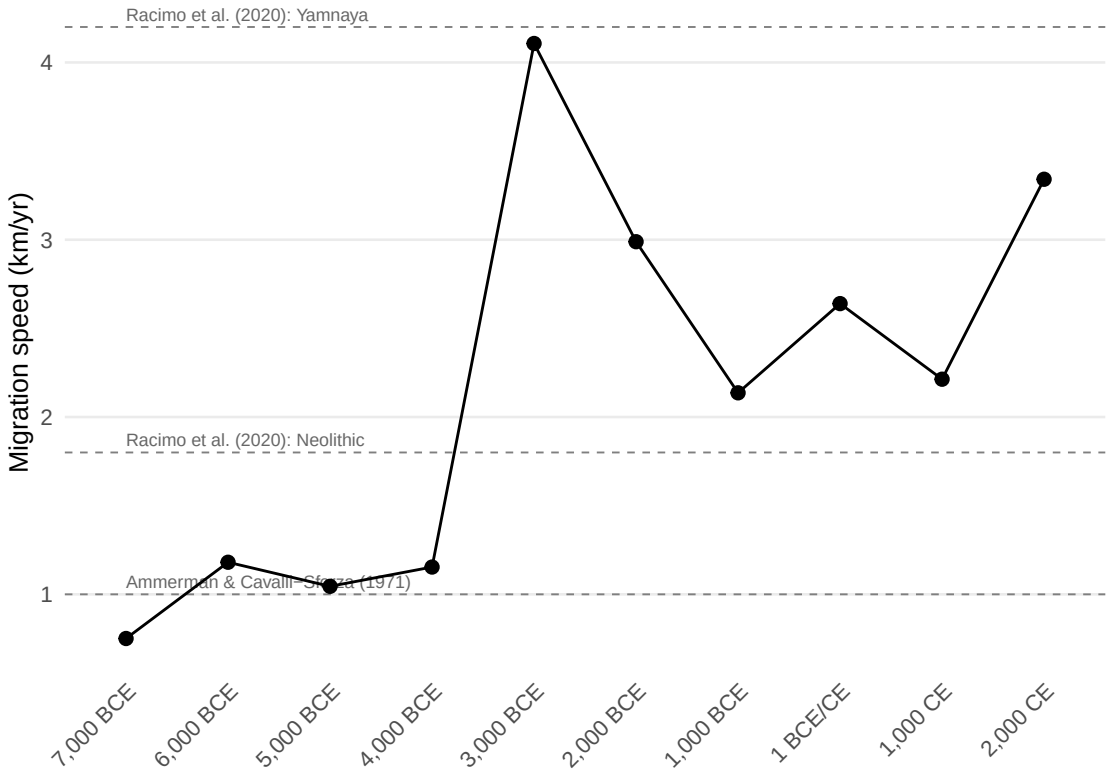
Notes: The figure plots the spatial distribution of modern climate variables at 1° resolution. Panel A shows cumulative growing degree days above 10°C, derived from ERA5 reanalysis data, which captures biologically relevant heat accumulation during the growing season. Panel B shows mean annual precipitation.

Figure A2: Subsistence Composition Over Time



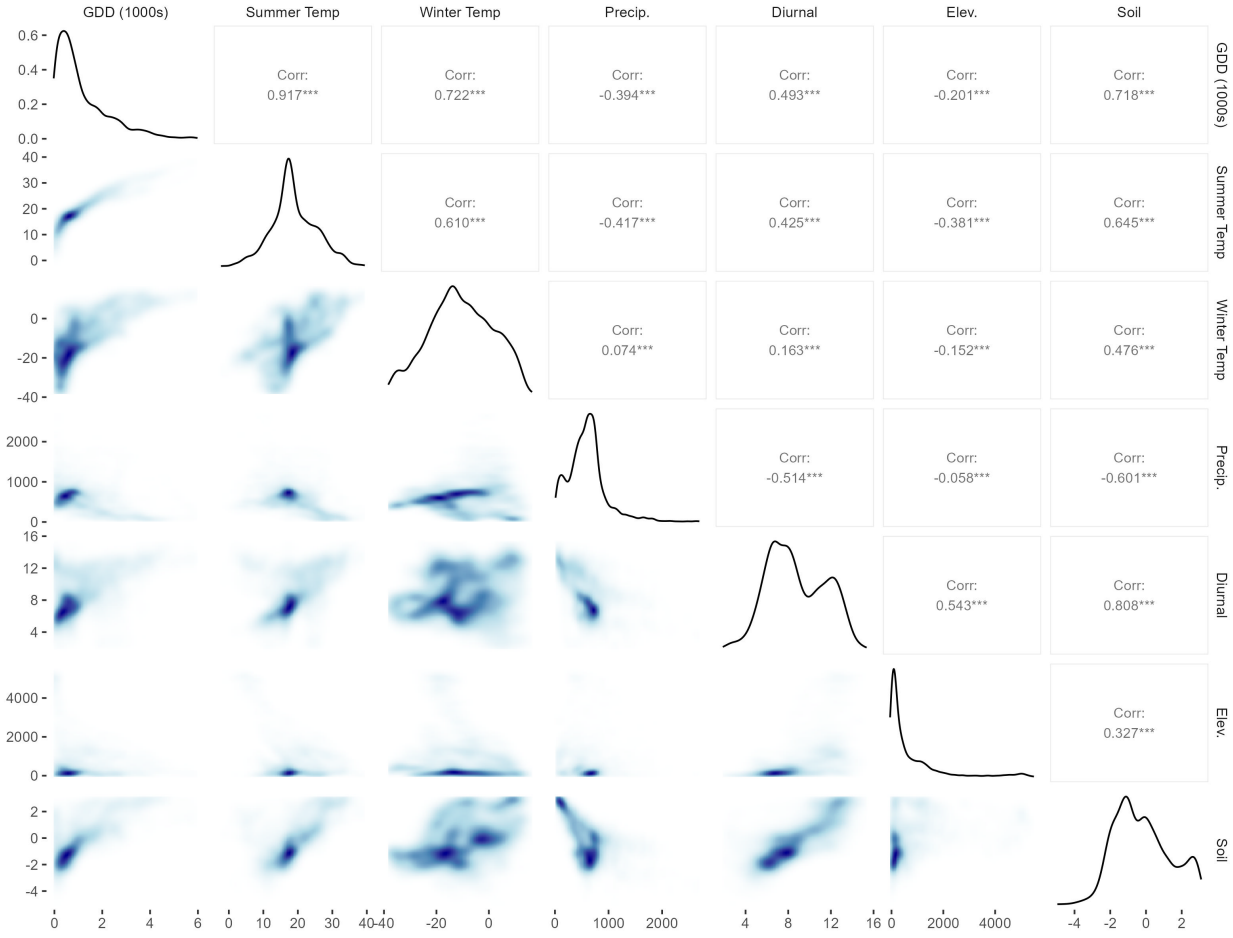
Notes: Stacked area plots showing the mean subsistence share across origin gridcells by 500-year era. Upper panel uses the archaeological population-level classification (continuous shares summing to one). Lower panel uses HYDE 3.2 gridded cropland and grazing data. Both classifications show a transition from forager-dominated to farmer-dominated samples around 5000 BCE, with pastoralists emerging as a persistent share from approximately 3000 BCE.

Figure A3: Implied Migration Speed by Era



Notes: Each point reports the median ratio of geographic distance to temporal separation for all IBD-linked pairs (sharing at least one segment $\geq$8 cM) assigned to the given era, restricted to pairs separated by 100–1,000 years. Dashed lines indicate population-level wave-of-advance estimates from the literature. Sources: Ammerman and Cavalli-Sforza (1971); Pinhasi et al. (2005); Racimo et al. (2020).

Figure A4: 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 distance measure. Temperature and GDD from ERA5 reanalysis (Hersbach et al., 2020); precipitation from TerraClimate (Abatzoglou et al., 2018); soil data from SoilGrids (Hengl, 2018b; Hengl and Gupta, 2019; Hengl and Wheeler, 2018).

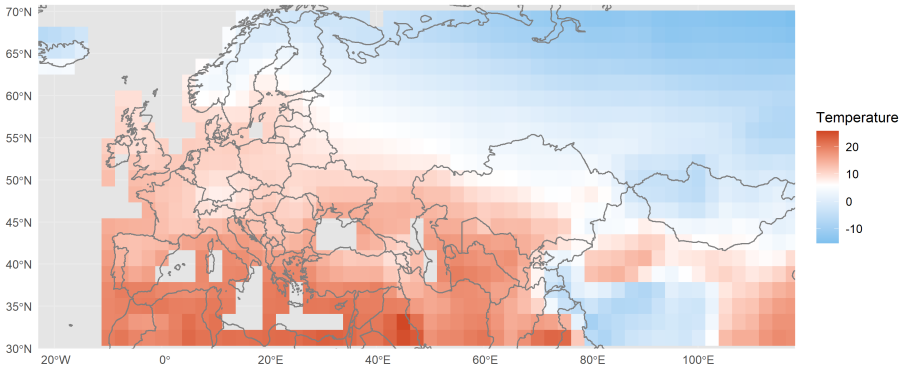
Figure A5: Evolution of Average Temperature over Time



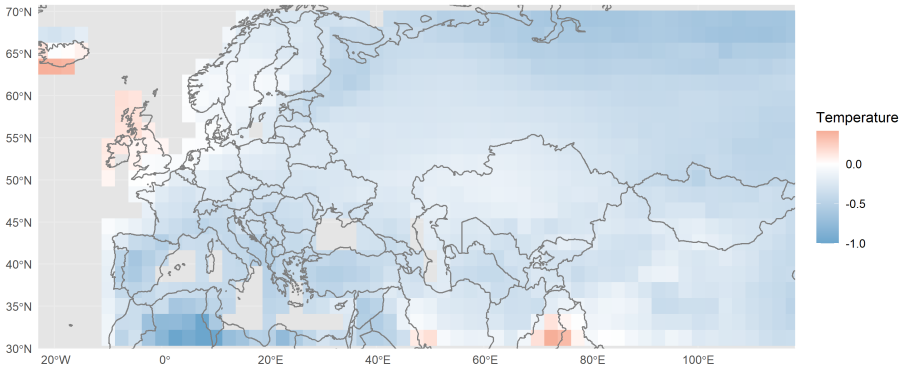
Notes: The figure plots the evolution of (reconstructed) temperature for the study region over time. Temperature reconstructions are from Osman et al. (2021).

Figure A6: Paleoclimate Temperature Reconstructions

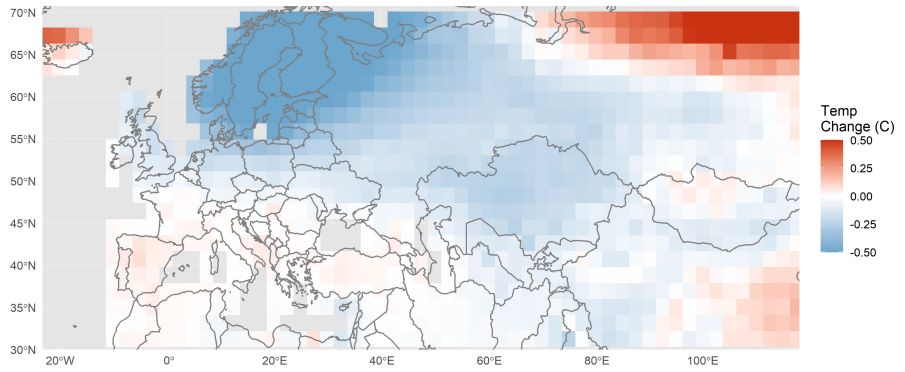
(a) Reconstructed Temperature, 3000 BCE



(b) Temperature Change: 3650 to 2250 BCE (warming)

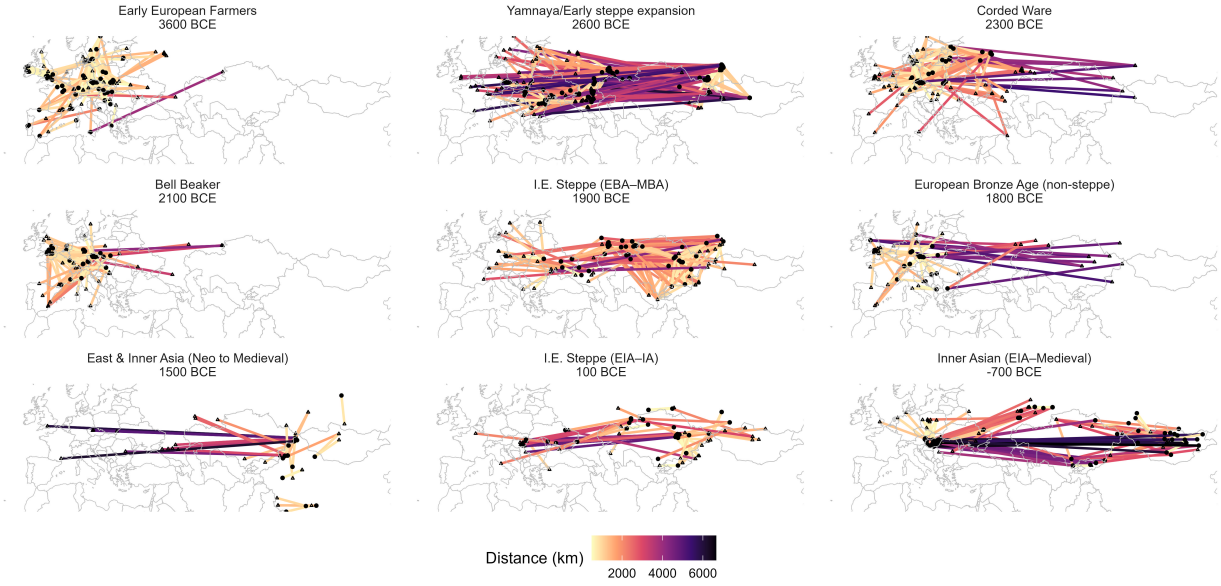


(c) Temperature Change: 1800 to 1200 BCE



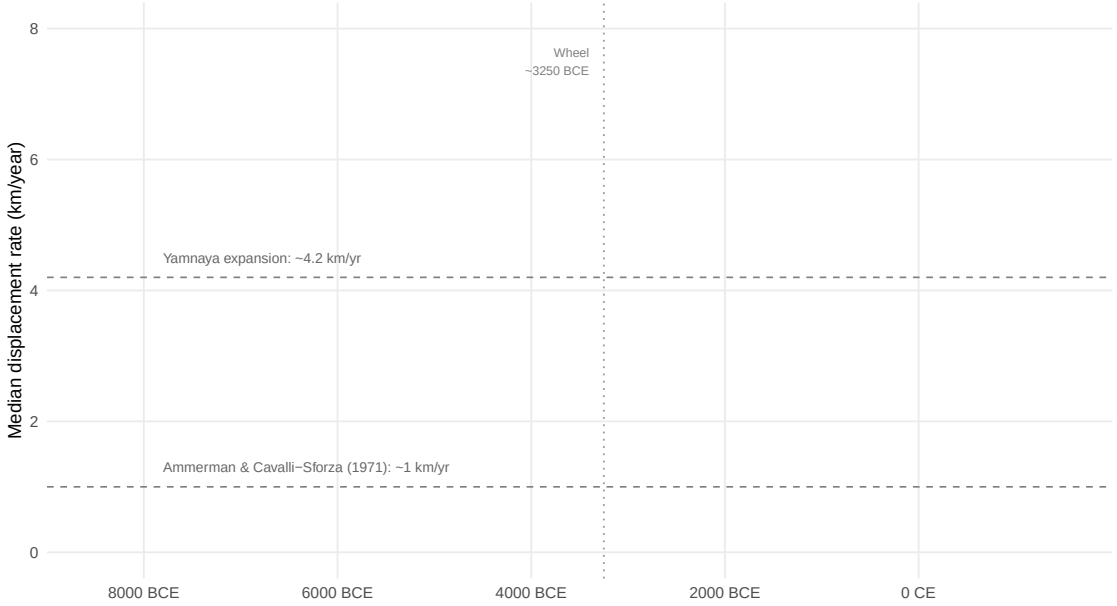
Notes: Panel A shows reconstructed average temperature in 3000 BCE from (Osman et al., 2021) at 2.5° resolution. Panels B and C plot temperature change over multi-century intervals.

Figure A7: Movements of Major Culture Groups



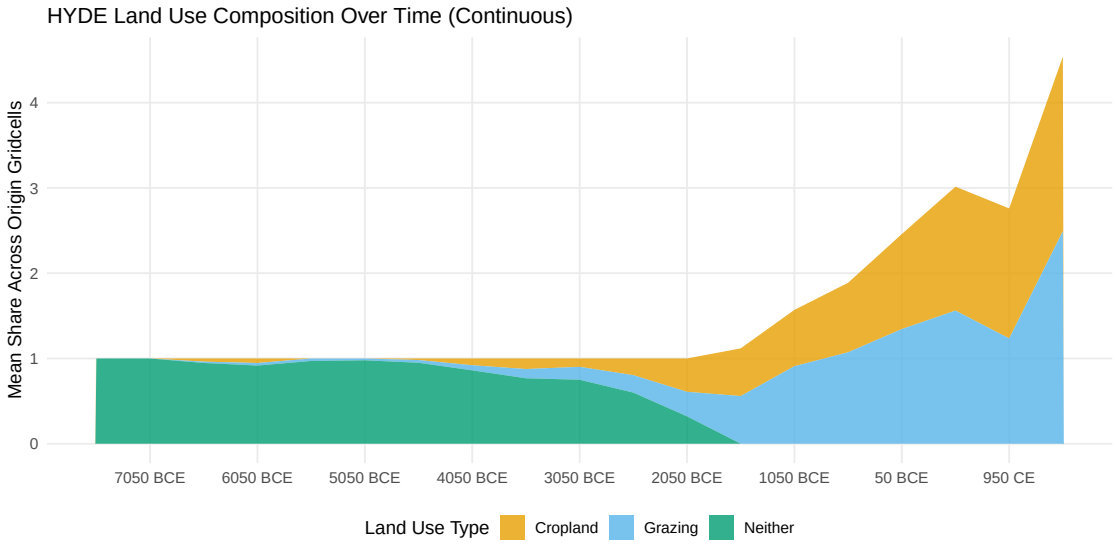
Notes: The maps show movements of individuals for major cultural groups. Sample restricted to pairs sharing at least one segment >16 cM, living at most 500 years apart, separated by at least 500km.

Figure A8: Implied Migration Speed by Subsistence Type



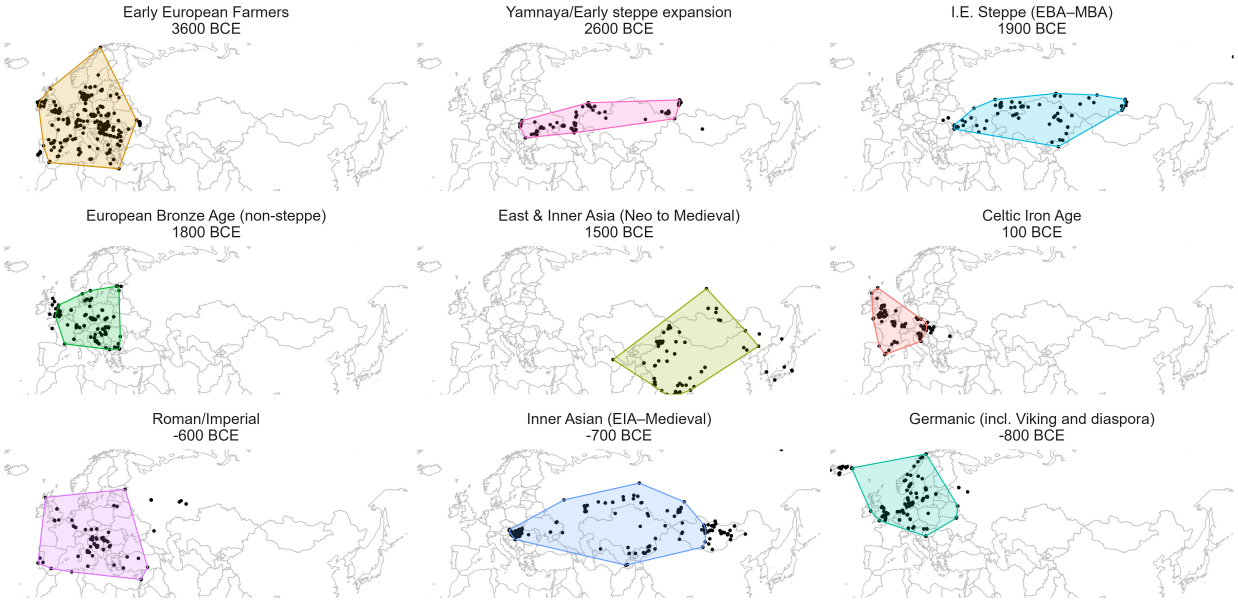
Notes: Median displacement rate (km/year) by era and subsistence type. Dashed lines indicate benchmark estimates from the literature.

Figure A9: HYDE Land Use Composition Over Time (Continuous)



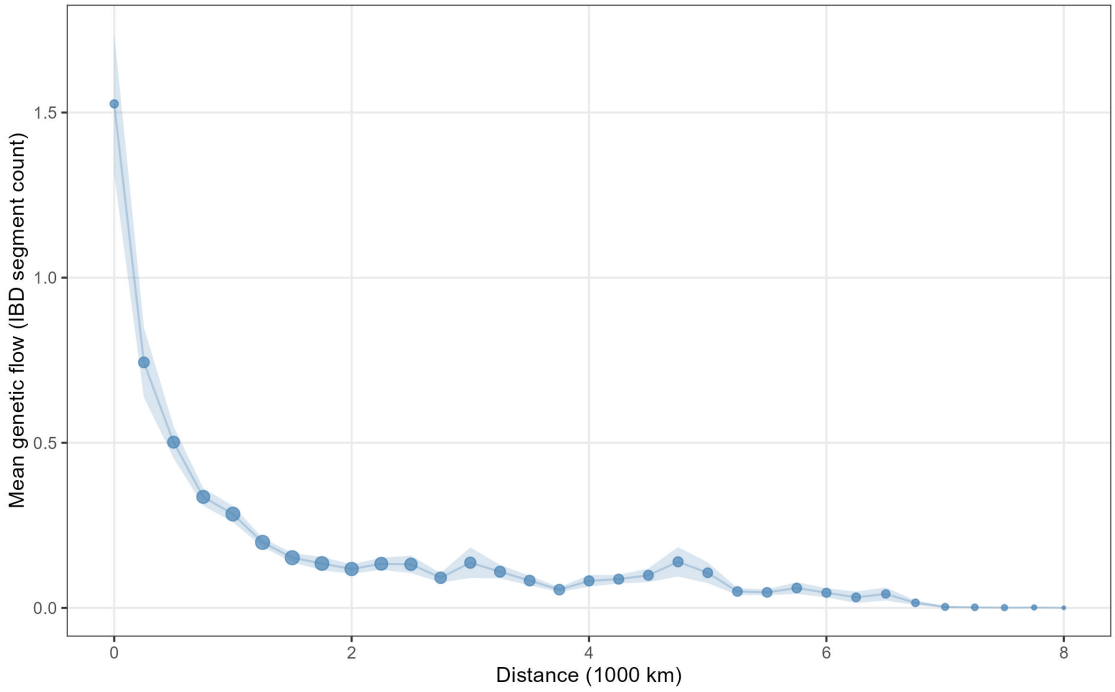
Notes: Mean continuous cropland and grazing shares from HYDE 3.2, averaged across origin gridcells by 500-year era. "Neither" captures gridcells with zero cropland and grazing (forager-type locations). Unlike the categorical classification in Figure A2, this figure uses the raw continuous HYDE land use intensities.

Figure A10: 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 A11: Genetic Flow by Geographic Distance

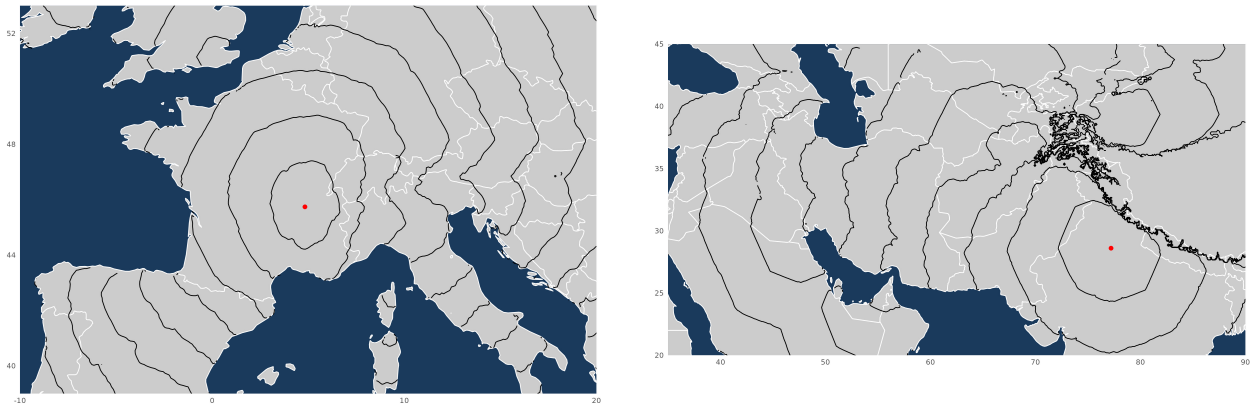


Notes: The figure shows mean genetic flow (IBD segment count $\geq$ 8 cM) between gridcell pairs by geographic distance. Points represent 250km distance bins; point size reflects sample size. Shaded ribbon shows 95% confidence interval.

Figure A12: Travel Cost Isochrones

(a) Central Europe

(b) South/Central Asia



Notes: Each panel shows least-cost-path travel isochrones from an illustrative origin (red dot), accounting for topography and terrain ruggedness. Contour lines represent equal travel cost from the origin. Panel A illustrates how the Alps distort travel distances relative to Haversine distance in central Europe. Panel B shows the effect of the Himalayas and surrounding mountain ranges in South/Central Asia. The asymmetry of the isochrones demonstrates that straight-line geographic distance can substantially misrepresent the actual frictions faced by prehistoric populations.

Figure A12 illustrates how terrain and water barriers cause effective travel costs to diverge from Haversine distance. To address the concern that straight-line geographic distance does not capture the actual frictions faced by prehistoric populations, Table A14 reports robustness to alternative distance specifications: linear, quadratic, log, and least-cost-path distance on foot that accounts for topography and terrain ruggedness. Climate matching coefficients are stable across all specifications, confirming that the results are not driven by the functional form of the distance control or by the distinction between straight-line and terrain-adjusted distance.

Table A2: Climate Matching: Soil Components

Dependent Variable:	Genetic Flow						
	Index	pH	Water	Organic	Clay	Bulk	All
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDD Dist.	-0.363***	-0.262***	-0.424***	-0.331***	-0.293***	-0.390***	-0.145*
	(0.055)	(0.063)	(0.058)	(0.060)	(0.070)	(0.058)	(0.075)
Winter Temp. Dist.	0.004	-0.0002	0.005	0.004	0.003	0.006	3.8×10^{-5}
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Precipitation Dist. (m)	-0.519***	-0.432***	-0.623***	-0.609***	-0.689***	-0.628***	-0.489***
	(0.118)	(0.137)	(0.116)	(0.111)	(0.113)	(0.113)	(0.124)
Soil Dist.	-0.153***						
	(0.049)						
Soil pH Dist.		-0.223***					-0.132*
		(0.083)					(0.079)
Soil Water Dist.			-0.235***				-0.052
			(0.083)				(0.091)
Soil Organic Dist.				-0.386***			-0.351***
				(0.054)			(0.050)
Soil Clay Dist.					-0.230***		-0.165***
					(0.059)		(0.055)
Soil Bulk Dist.						-0.113**	
						(0.056)	
Distance (km)	-0.362***	-0.350***	-0.359***	-0.349***	-0.351***	-0.367***	-0.333***
	(0.022)	(0.022)	(0.021)	(0.021)	(0.020)	(0.022)	(0.021)
Dep. var. mean	0.191	0.191	0.191	0.191	0.191	0.191	0.191
Fit statistics							
Observations	835,786	835,786	835,786	835,786	835,786	835,786	835,786
Pseudo R^2	0.498	0.499	0.498	0.500	0.499	0.498	0.500

Notes: Poisson pseudo-maximum likelihood estimates. All specifications include origin gridcell, destination gridcell, and era fixed effects. Column 1 uses the composite soil index (first principal component); columns 2–6 use individual soil characteristics. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Presence of Archaeological Sites across Climate Areas

Dependent Variable:	1[at least one obs.]					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
DDays ≥ 10	-0.026***	-0.024***	-0.038***	-0.039***	-0.052***	-0.045**
	(0.001)	(0.001)	(0.002)	(0.003)	(0.004)	(0.018)
Precipitation		4.82×10^{-5}***	5.81×10^{-6}	5.19×10^{-5}***	1.29×10^{-5}	-8.39×10^{-5}**
		(7.51×10^{-6})	(1.07×10^{-5})	(1.4×10^{-5})	(1.15×10^{-5})	(3.51×10^{-5})
Clay			0.049***	0.019**	0.045***	0.074***
			(0.005)	(0.008)	(0.005)	(0.014)
Soil Organic			-0.060***	-0.041***	-0.080***	-0.020
			(0.005)	(0.006)	(0.006)	(0.014)
Soil pH				0.070***		
				(0.010)		
Soil Water				0.052***		
				(0.012)		
Temp. Diurnal Range					0.003	0.019***
					(0.003)	(0.007)
Elevation					-0.039***	-0.048**
					(0.006)	(0.024)
Fixed-effects						
2.5 gridcell						Yes
Fit statistics						
Observations	6,069	6,067	6,048	6,048	6,048	6,022
R^2	0.022	0.026	0.065	0.072	0.074	0.429
Within R^2						0.010

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: OLS estimates. Dependent variable: indicator for at least one ancient DNA individual observed in a 1-degree gridcell. Robust standard errors in parentheses; clustered at the 2.5-degree level in the last column. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Migration Distance by Subsistence Type

Subsistence	N pairs	Mean (km)	Mean, IBD-wt	Median (km)	Median, IBD-wt	SD (km)
Farmer	3,228,117	2,158	1,071	1,532	736	1,887
Forager	126,676	3,199	694	2,507	373	2,282
Agro-Pastoral	348,685	2,940	1,561	2,577	1,237	1,895
Pastoral	1,907,416	2,802	2,182	2,136	1,843	2,132

Notes: Individual-level IBD pairs. Distance in km. IBD-weighted columns weight by IBD segment count (8cM). Subsistence classified by GPT-based archaeological classification at origin.

Table A5: Climate Matching Heterogeneity by Subsistence Type

Dependent Variable:	Genetic Flow					
Model:	Baseline (1)	GDD x Farm (2)	GDD x Past (3)	Prec x Farm (4)	Prec x Past (5)	Joint (6)
GDD Dist.	-0.201***	0.227**	-0.937***	-0.200***	-0.200***	0.031
	(0.074)	(0.095)	(0.114)	(0.074)	(0.074)	(0.347)
Precip. Dist.	-0.670***	-0.568***	-0.572***	-0.551***	-0.743***	1.27***
	(0.109)	(0.113)	(0.112)	(0.176)	(0.162)	(0.369)
Soil Composite Dist.	-0.351***	-0.368***	-0.344***	-0.368***	-0.352***	-0.953***
	(0.056)	(0.057)	(0.056)	(0.056)	(0.056)	(0.274)
Farmer Sh. × GDD Dist.		-1.17***				-0.803*
		(0.157)				(0.415)
Pastoral Sh. × GDD Dist.			1.34***			0.113
			(0.176)			(0.338)
Farmer Sh. × Precip. Dist.				-0.233		-1.28***
				(0.284)		(0.432)
Pastoral Sh. × Precip. Dist.					0.172	-2.29***
					(0.286)	(0.370)
Farmer Sh. × Soil Dist.						0.845***
						(0.285)
Pastoral Sh. × Soil Dist.						0.585**
						(0.275)
Farmer Share (origin)		-1.46***		-1.87***		-1.85***
		(0.179)		(0.178)		(0.570)
Pastoral Share (origin)			2.55***		3.06***	0.304
			(0.304)		(0.289)	(0.719)
Fit statistics						
Observations	820,528	820,528	820,528	820,528	820,528	820,528
Pseudo R^2	0.50	0.51	0.51	0.51	0.51	0.52

Notes: PPML estimates. Dependent variable: IBD segment count ($\geq$ 8 cM). Forager is the omitted reference group. Columns (2)–(3) interact GDD distance with farmer and pastoral shares. Columns (4)–(5) interact precipitation distance with farmer and pastoral shares. Column (6) interacts all climate and geography variables jointly with farmer and pastoral shares. Origin gridcell, destination gridcell, and era FE. Pair-clustered SEs. *** p<0.01, ** p<0.05, * p<0.1

Table A6: Climate Matching by Subsistence Type (Individual)

Dependent Variable:	Genetic Flow				
	All	Farmer	Agro-Past.	Pastoral	Forager
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
GDD Dist.	-0.396***	-0.247***	-0.495***	-0.115***	0.010
	(0.037)	(0.041)	(0.068)	(0.037)	(0.257)
Precip. Dist.	-0.600***	-0.213**	-0.389***	-0.607***	-0.193
	(0.088)	(0.096)	(0.139)	(0.099)	(0.447)
Soil Composite Dist.	-0.126***	-0.031	-0.026	-0.049	-0.261***
	(0.025)	(0.026)	(0.031)	(0.031)	(0.082)
Distance (000 km)	-0.325***	-0.997***	-0.466***	-0.140***	-1.62***
	(0.011)	(0.035)	(0.019)	(0.007)	(0.114)
Fixed-effects					
geo1	Yes	Yes	Yes	Yes	Yes
geo2	Yes	Yes	Yes	Yes	Yes
era500	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	5,571,429	3,091,518	309,948	1,837,271	94,196
Pseudo R^2	0.298	0.261	0.348	0.365	0.415

Notes: PPML estimates. Dependent variable: IBD segment count ($\geq$ 8 cM) between individual pairs. Each column restricts to origin individuals classified as the indicated subsistence type (GPT-based archaeological classification). Soil Composite Dist. is Euclidean on standardized clay and organic carbon. Origin, destination, and era FE. Pair-clustered SEs. *** p<0.01, ** p<0.05, * p<0.1

Table A7: 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.	-0.0949	0.9228***	0.3696**	0.3837***	-0.2260***	-0.1270*	-0.3526***	-0.3079***
	(0.2116)	(0.1845)	(0.1450)	(0.1214)	(0.0386)	(0.0651)	(0.0646)	(0.1012)
Precipitation Dist. (m)	0.2729	0.0012	-0.2470	-1.756***	0.0655	-0.2621**	-0.3009	-1.070***
	(0.7825)	(0.3984)	(0.3244)	(0.1785)	(0.0833)	(0.1286)	(0.2012)	(0.2205)
Dep. var. mean	0.1570	0.4760	0.3980	0.3660	0.6640	0.1620	0.0690	0.1380
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	4,281	16,519	18,179	54,811	110,946	89,756	110,889	197,843
Pseudo R^2	0.45613	0.75318	0.72216	0.62744	0.67472	0.55022	0.42207	0.59488

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): precipitation distance, soil composite distance, and geographic distance. 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 A8: Climate Matching: Additional Environmental Controls

Dependent Variable:	Genetic Flow							
Model:	Baseline (1)	Interacted FE (2)	+UV (3)	+Diurnal (4)	+Elevation (5)	+Coastal (6)	+Latitude (7)	+Longitude (8)
Variables								
GDD Dist.	-0.200***	-0.122***	0.063	0.061	0.056	0.056	0.055	0.041
	(0.074)	(0.038)	(0.073)	(0.073)	(0.072)	(0.072)	(0.071)	(0.072)
Precipitation Dist. (m)	-0.667***	-0.798***	-0.680***	-0.676***	-0.665***	-0.669***	-0.683***	-0.680***
	(0.109)	(0.067)	(0.104)	(0.104)	(0.105)	(0.104)	(0.104)	(0.103)
Soil Composite Dist.	-0.352***	-0.329***	-0.233***	-0.230***	-0.225***	-0.218***	-0.204***	-0.227***
	(0.056)	(0.034)	(0.054)	(0.055)	(0.055)	(0.056)	(0.054)	(0.053)
UV Distance			-0.0008***	-0.0008***	-0.0008***	-0.0008***	-0.0007***	-0.0007***
			(4.41×10^{-5})	(4.36×10^{-5})	(4.9×10^{-5})	(4.91×10^{-5})	(7.51×10^{-5})	(7.65×10^{-5})
Distance (km)	-0.331***	-0.355***	-0.310***	-0.308***	-0.312***	-0.316***	-0.315***	0.210**
	(0.015)	(0.011)	(0.014)	(0.015)	(0.016)	(0.016)	(0.015)	(0.102)
Diurnal Range Dist.				-0.007	-0.008	0.005	0.005	-0.004
				(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Elevation Dist.					0.072	0.073	0.061	0.059
					(0.049)	(0.049)	(0.050)	(0.050)
Coast to Non-coast Move						-0.174***	-0.170***	-0.164***
						(0.038)	(0.039)	(0.039)
Latitude Dist. (km)							-0.073	-0.300***
							(0.081)	(0.080)
Longitude Dist. (km)								-0.472***
								(0.086)
Dep. var. mean	0.191	0.379	0.191	0.191	0.191	0.191	0.191	0.191
Fixed-effects								
Orig. gridcell	Yes		Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes		Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes		Yes	Yes	Yes	Yes	Yes	Yes
Orig. gridcell-Era		Yes						
Dest. gridcell-Era		Yes						
Fit statistics								
Observations	835,786	420,550	835,786	835,786	835,786	835,786	835,786	835,786
Pseudo R^2	0.500	0.747	0.504	0.504	0.504	0.504	0.504	0.505

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 A9: Climate Matching: Specification Robustness

Dependent Variables:	Genetic Flow Baseline	Genetic Flow (12cM) 12cM Flow	Genetic Flow (16cM) 16cM Flow
Model:	(1)	(2)	(3)
Variables			
GDD Dist.	-0.200***	-0.169	-0.214
	(0.074)	(0.108)	(0.135)
Precipitation Dist. (m)	-0.667***	-1.20***	-1.45***
	(0.109)	(0.159)	(0.219)
Soil Composite Dist.	-0.352***	-0.588***	-0.806***
	(0.056)	(0.132)	(0.149)
Distance (km)	-0.331***	-0.468***	-0.628***
	(0.015)	(0.021)	(0.033)
Dep. var. mean	0.191	0.049	0.020
Fixed-effects			
Orig. gridcell	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes
Era	Yes	Yes	Yes
Fit statistics			
Observations	835,786	652,051	478,167
Pseudo R^2	0.500	0.471	0.446

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 A10: Environmental Distance and Migration (Individual)

Dependent Variable:	Genetic Flow					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
GDD Dist.	-1.07***	-0.902***	-0.430***	-0.396***	-0.339***	-0.336***
	(0.063)	(0.033)	(0.040)	(0.037)	(0.033)	(0.032)
Precip. Dist.	-0.558***	-0.581***	-0.549***	-0.600***	-0.630***	-0.758***
	(0.145)	(0.081)	(0.096)	(0.088)	(0.075)	(0.076)
Soil Composite Dist.	-0.175***	-0.039*	-0.114***	-0.126***	-0.122***	-0.116***
	(0.032)	(0.020)	(0.028)	(0.025)	(0.021)	(0.019)
Distance (000 km)	-0.125***	-0.223***	-0.313***	-0.325***	-0.341***	-0.354***
	(0.016)	(0.013)	(0.011)	(0.011)	(0.010)	(0.011)
Fixed-effects						
Era		Yes		Yes		
Orig. gridcell			Yes	Yes		
Dest. gridcell			Yes	Yes	Yes	
Orig. grid.-Era					Yes	Yes
Dest. grid.-Era						Yes
Fit statistics						
Observations	5,616,244	5,616,244	5,571,429	5,571,429	5,475,259	5,396,750
Pseudo R^2	0.041	0.196	0.267	0.298	0.316	0.327

Notes: PPML estimates. Dependent variable: IBD segment count ($\geq$ 8 cM) between individual pairs. Soil Composite Dist. is the Euclidean distance on standardized clay content and soil organic carbon. Column (4) is the preferred specification. Standard errors clustered by gridcell pair. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Climate Matching: Distance Band Robustness

Dependent Variable:	Genetic Flow				
	Baseline	100-500km	500-1000km	1000-1500km	>1500
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
GDD Dist.	-0.200***	-0.434	-0.476***	-0.154**	0.503***
	(0.074)	(0.308)	(0.104)	(0.072)	(0.062)
Precipitation Dist. (m)	-0.667***	0.875**	-0.462*	0.382**	-1.48***
	(0.109)	(0.411)	(0.259)	(0.164)	(0.101)
Soil Composite Dist.	-0.352***	-0.347***	-0.096	-0.197***	-0.067
	(0.056)	(0.120)	(0.062)	(0.050)	(0.042)
Distance (km)	-0.331***				
	(0.015)				
Dep. var. mean	0.191	1.01	0.459	0.279	0.110
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	835,786	34,921	69,426	90,364	558,030
Pseudo R^2	0.500	0.474	0.449	0.456	0.561

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 A12: Climate Matching: Sample Robustness

Dependent Variable:	Genetic Flow						
	Baseline	Two-Way Clustering	Horizontal	Vertical	≥100km	≥200km	≥300km
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
GDD Dist.	-0.200***	-0.200*	-0.268***	-0.167**	-0.139**	-0.075	-0.057
	(0.074)	(0.120)	(0.094)	(0.080)	(0.061)	(0.058)	(0.058)
Precipitation Dist. (m)	-0.667***	-0.667***	-0.661***	-0.004	-0.648***	-0.632***	-0.683***
	(0.109)	(0.251)	(0.142)	(0.148)	(0.110)	(0.117)	(0.121)
Soil Composite Dist.	-0.352***	-0.352***	-0.189**	-0.227***	-0.313***	-0.275***	-0.226***
	(0.056)	(0.073)	(0.078)	(0.056)	(0.039)	(0.039)	(0.040)
Distance (km)	-0.331***	-0.331***	-0.315***	-1.05***	-0.323***	-0.305***	-0.288***
	(0.015)	(0.043)	(0.018)	(0.083)	(0.015)	(0.015)	(0.015)
Dep. var. mean	0.191	0.191	0.184	0.249	0.185	0.174	0.167
Fixed-effects							
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	835,786	835,786	634,835	170,789	831,059	820,269	808,788
Pseudo R^2	0.500	0.500	0.528	0.477	0.498	0.500	0.502

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Poisson pseudo-maximum likelihood estimates. Genetic Flow = count of individual pairs sharing at least one IBD segment of 8 cM or longer. All specifications include origin gridcell, destination gridcell, and era fixed effects. Horizontal = moves where east-west distance (in km) exceeds north-south distance, accounting for longitude shrinkage at higher latitudes. Vertical = north-south distance exceeds east-west distance. Standard errors clustered by grid cell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A13: Inference Robustness: Alternative Standard Error Treatments

Dependent Variable:	Genetic Flow				
	Pair Cl.	Multi-way (1°)	Interact. FE	Origin Cl. (2.5°)	Multi-way (2.5°)
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
GDD Dist.	-0.200***	-0.200*	-0.122***	-0.200	-0.200
	(0.074)	(0.120)	(0.038)	(0.149)	(0.150)
Precipitation Dist. (m)	-0.667***	-0.667***	-0.798***	-0.667***	-0.667**
	(0.109)	(0.251)	(0.067)	(0.258)	(0.287)
Soil Composite Dist.	-0.352***	-0.352***	-0.329***	-0.352***	-0.352***
	(0.056)	(0.073)	(0.034)	(0.057)	(0.072)
Distance (km)	-0.331***	-0.331***	-0.355***	-0.331***	-0.331***
	(0.015)	(0.043)	(0.011)	(0.042)	(0.051)
Fixed-effects					
Orig. gridcell	Yes	Yes		Yes	Yes
Dest. gridcell	Yes	Yes		Yes	Yes
Era	Yes	Yes		Yes	Yes
Orig. gridcell-Era			Yes		
Dest. gridcell-Era			Yes		
Fit statistics					
Observations	835,786	835,786	420,550	835,786	835,786
Pseudo R^2	0.50	0.50	0.75	0.50	0.50

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: PPML estimates. All columns estimate the same preferred specification: GDD distance, precipitation distance (TerraClimate), soil composite distance (clay + organic C), and geographic distance. Column (1): pair-clustered SEs. Column (2): multi-way clustered on origin + destination. Column (3): origin×era and destination×era interacted FE with pair-clustered SEs. Column (4): clustered at the 2.5-degree origin level. Column (5): multi-way clustered at the 2.5-degree origin + destination level. *** p<0.01, ** p<0.05, * p<0.1

Table A14: Climate Matching: Distance Robustness

Dependent Variable:	Genetic Flow			
	Linear (Baseline)	Quadratic	Log	On-foot Travel Cost
Model:	(1)	(2)	(3)	(4)
Variables				
GDD Dist.	-0.201***	-0.104	-0.032	-0.184**
	(0.074)	(0.076)	(0.069)	(0.074)
Winter Temp. Dist.	0.004	0.002	-0.005	0.012**
	(0.005)	(0.005)	(0.005)	(0.005)
Soil Composite Dist.	-0.350***	-0.278***	-0.172***	-0.330***
	(0.056)	(0.052)	(0.045)	(0.052)
Precip. Dist.	-0.385***	-0.346***	-0.291***	-0.325***
	(0.091)	(0.090)	(0.089)	(0.093)
Distance (km)	-0.361***	-0.740***		
	(0.018)	(0.041)		
Distance (km)-sq		0.068***		
		(0.006)		
log(Distance)			-0.587***	
			(0.032)	
Travel Cost (Foot) Weighted 1000km				-0.351***
				(0.018)
Dep. var. mean	0.191	0.191	0.191	0.191
Fixed-effects				
Orig. gridcell	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes
Fit statistics				
Observations	835,786	835,786	835,786	835,786
Pseudo R^2	0.500	0.502	0.506	0.502

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. Column 1 uses linear distance as a control, Columns 2-3 use quadratic and natural log forms, respectively. Columns 4-5 use adjusted distances for horse and foot travel. *** p<0.01, ** p<0.05, * p<0.1

Table A15: Climate Adaptation (Individual)

Dependent Variable:	Temp. Diff. (space)			
Model:	(1)	(2)	(3)	(4)
Variables				
Temp. Change (400yr)	-1.24***	-1.29***	-0.298***	-0.253**
	(0.297)	(0.288)	(0.091)	(0.100)
Distance (000 km)		-0.538**	-0.026	-0.015
		(0.240)	(0.018)	(0.017)
Distance-sq		0.007	0.001	-0.0008
		(0.041)	(0.003)	(0.003)
Precip. Diff. (space)				-0.002***
				(0.0002)
Fixed-effects				
Orig. gridcell	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes
Dest. gridcell			Yes	Yes
Fit statistics				
Observations	69,991	69,991	69,970	69,970
Adjusted R^2	0.451	0.470	0.954	0.958

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

Table A16: Adaptive Migration by Subsistence Type

Dependent Variable:	Genetic Flow				
	Pooled	Farmer	Pastoral	Forager	Joint
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Temp. Change (origin)	1.42***	-0.731**	2.79***	1.80***	0.415*
	(0.216)	(0.286)	(0.293)	(0.255)	(0.247)
Temp. Change × Farmer Sh.		6.19***			4.95***
		(0.502)			(0.439)
Temp. Change × Pastoral Sh.			-3.62***		-1.85***
			(0.553)		(0.541)
Temp. Change × Forager Sh.				-1.27***	
				(0.321)	
Farmer Share (origin)		-2.00***			-0.726**
		(0.198)			(0.344)
Pastoral Share (origin)			2.92***		2.19***
			(0.315)		(0.499)
Forager Share (origin)				0.078	
				(0.261)	
Climate dist. controls	Yes	Yes	Yes	Yes	Yes
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	800,588	800,588	800,588	800,588	800,588
Pseudo R^2	0.51	0.52	0.52	0.51	0.52

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

Notes: PPML estimates. Dependent variable: IBD segment count (≥ 8 cM) between 1-degree gridcell pairs. Temp. Change (origin) is the Osman et al. (2021) reconstructed temperature change over a centered 400-year window (positive = warming). Subsistence shares are continuous GPT-classified measures at the origin, based on archaeological population names. Columns (2)–(4) each interact Temp. Change with one subsistence share. Column (5) includes farmer and pastoral interactions jointly with forager as the reference group; Temp. Change captures the forager response and the interactions measure deviations from it. Climate distance controls: GDD distance, precipitation distance, and soil composite distance. Origin gridcell, destination gridcell, and era fixed effects. Standard errors clustered by gridcell pair. *** p<0.01, ** p<0.05, * p<0.1

Table A17: Climate Matching: 2.5-Degree Grid Robustness

Dependent Variable:	Genetic Flow			
	No FE	Era	O+D	O+D+Era
Model:	(1)	(2)	(3)	(4)
Variables				
GDD Dist.	-0.975***	-0.946***	-0.202*	-0.207**
	(0.066)	(0.066)	(0.106)	(0.105)
Precipitation Dist. (m)	-0.372**	-0.370**	-0.433**	-0.426**
	(0.162)	(0.165)	(0.171)	(0.172)
Soil Composite Dist.	-0.074*	-0.105**	-0.297***	-0.294***
	(0.043)	(0.045)	(0.051)	(0.052)
Distance (km)	-0.464***	-0.444***	-0.320***	-0.326***
	(0.033)	(0.033)	(0.022)	(0.022)
Dep. var. mean		0.414	0.432	0.432
Fixed-effects				
Era		Yes		Yes
Orig. gridcell			Yes	Yes
Dest. gridcell			Yes	Yes
Fit statistics				
Observations	363,592	363,592	348,791	348,791
Pseudo R^2	0.13	0.21	0.45	0.52

Notes: PPML estimates at 2.5-degree gridcell resolution. Same specification as the 1-degree baseline (Table 2). Standard errors clustered by gridcell pair. *** p<0.01, ** p<0.05, * p<0.1

Table A18: Climate Matching: Alternative Climate Measures

Dependent Variable:	Genetic Flow					
	GDD (Modern)	Avg (Modern)	Summer (Modern)	Avg (Proxy)	Avg (Beyer)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
GDD Dist.	-0.199***					
	(0.074)					
Winter Temp. Dist.	0.003					
	(0.005)					
Temperature Dist.		0.044***		0.043***	0.056***	-0.799***
		(0.009)		(0.007)	(0.008)	(0.111)
Summer Temp. Dist.			-0.047***			
			(0.009)			
Precipitation Dist. (m)	-0.663***	-0.598***	-0.558***	-0.582***	-0.522***	-0.003***
	(0.110)	(0.113)	(0.111)	(0.114)	(0.119)	(0.0009)
Soil Composite Dist.	-0.351***	-0.457***	-0.331***	-0.461***	-0.455***	
	(0.056)	(0.047)	(0.052)	(0.046)	(0.050)	
Distance (km)	-0.341***	-0.391***	-0.330***	-0.397***	-0.412***	
	(0.020)	(0.018)	(0.014)	(0.018)	(0.020)	
Dep. var. mean	0.191	0.191	0.191	0.189	0.199	0.814
Fixed-effects						
Orig. gridcell	Yes	Yes	Yes	Yes	Yes	
Dest. gridcell	Yes	Yes	Yes	Yes	Yes	
Era	Yes	Yes	Yes	Yes	Yes	Yes
Grid Pair						Yes
Fit statistics						
Observations	835,786	835,786	835,786	841,964	782,186	190,852
Pseudo R^2	0.500	0.500	0.500	0.500	0.502	0.470

Notes: Poisson pseudo-maximum likelihood estimates. Dependent variable is count of IBD segments $\geq$ 8cM between gridcell pairs. Columns 1–3 use modern climate data (ERA5). Column 4 uses marine proxy paleoclimate reconstructions (Osman et al., 2021). Columns 5–6 use Beyer et al. modeled paleoclimate. Column 6 replaces additive origin and destination fixed effects with gridcell pair fixed effects, exploiting within-pair temporal variation in paleoclimate distance. Standard errors clustered by gridcell pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A19: Ecological Diffusion: Excluding Pairs Within 500km

Dependent Variables:	cereal_dest_change	Dest. Pollen Change (t to $t+500$, $\times 100$)		cereal_dest_change	Dest. Pollen Change (t to $t+500$, $\times 100$)	
	Δ Cereal	Δ Grassland	Δ Tree	Δ Cereal	Δ Grassland	Δ Tree
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Genetic Flow (o–d, t-500 to t)	-0.064***	0.416	-0.327	-0.164***	1.95***	-1.75***
	(0.021)	(0.296)	(0.284)	(0.024)	(0.397)	(0.387)
Pollen Dist. ($o-d$, t)	21.0***	8.16***	8.75***	21.0***	8.25***	8.89***
	(0.380)	(0.133)	(0.141)	(0.401)	(0.139)	(0.147)
Genetic Flow × Pollen Dist.	0.099*	0.063***	0.066***	0.187**	0.090***	0.093***
	(0.056)	(0.016)	(0.017)	(0.089)	(0.029)	(0.029)
Distance × Pollen Dist.	0.173*	0.124***	0.108***	0.140	0.090***	0.069***
	(0.096)	(0.021)	(0.022)	(0.104)	(0.022)	(0.022)
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	331,007	331,007	331,007	302,029	302,029	302,029
R^2	0.563	0.461	0.518	0.557	0.458	0.514
Sample	All pairs	All pairs	All pairs	≥500km	≥500km	≥500km
Climate & geography controls	Yes	Yes	Yes	Yes	Yes	Yes

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

Notes: See Table 6 for variable definitions. Cols 1–3 reproduce the baseline (all pairs). Cols 4–6 restrict to pairs ≥500km apart, excluding pairs within the pollen interpolation radius. Standard errors clustered by origin–destination pair.