ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 144/26

Climate and Prehistoric Migration

Peter Huybers, Marco Tabellini, Charles A. Taylor, Francesco Toti

www.rfberlin.com

MAY 2026

Climate and Prehistoric Migration

Authors

Peter Huybers, Marco Tabellini, Charles A. Taylor, Francesco Toti

Reference

JEL Codes: N50, O13, Q54, Z13

Keywords: migration, ancient DNA, prehistory, climate, ecology

Recommended Citation: Peter Huybers, Marco Tabellini, Charles A. Taylor, Francesco Toti (2026): Climate and Prehistoric Migration. RFBerlin Discussion Paper No. 144/26

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Climate and Prehistoric Migration*

Peter Huybers Marco Tabellini Charles A. Taylor Francesco Toti

May 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. We document four main findings. 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, periods of warming increase farmer expansion while cooling increases pastoral expansion in patterns that recover known archaeological migration episodes. Migration also acts as a margin of climate adaptation: populations exposed to temperature change move to destinations that partly offset the shift. Fourth, genetic flow predicts subsequent convergence in destination vegetation toward migrants' ecological profiles, consistent with migration shaping landscape change and the demic diffusion of subsistence practices.

Keywords: migration, ancient 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 Joe Henrich, Elias Papaioannou, and David Reich for useful comments. We especially thank members of the Reich Lab at Harvard and Harald Ringbauer for sharing data. Michael Cleary, Francesco Farsetti, Gabrielle Lévy, and Elena Verzellesi provided excellent research assistance. All remaining errors are ours.

1 Introduction

Throughout most of human history, migration decisions were 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 within 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 tropical regime, just as a pastoralist adapted to arid grasslands faced challenges managing herds through a humid continental winter (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 takes this idea to the data. We study whether ecological similarity shaped human mobility over the past 10,000 years and whether migration served as an adaptive response to changing local conditions. To guide the analysis, 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 across Western Eurasia (Ringbauer et al., 2024). When two people share an identity-by-descent (IBD) DNA segment, they both inherited it from a recent common ancestor. If they are buried in different locations, someone in their shared lineage must have moved between those places. The volume of such genealogical links between two locations therefore provides a direct measure of multi-generational population flow.

The scale of these implied prehistoric movements is striking. Among the 2.3% of candidate pairs in our sample that share an 8 cM IBD segment—indicating a shared ancestor within roughly 10–15 generations—the average ancestor-descendant pair is separated by 1,550 km. This is comparable to relocating from Berlin to Moscow in just a few hundred years. Some individual corridors are even more striking. A single Yamnaya individual buried at Samara (∼3200 BCE) has documented descendants 2,600 km east in southern Siberia and 2,200 km west in Central Europe within a few centuries—a documented common ancestor for both the eastward and westward Bronze Age steppe expansions. Three millennia later, IBD links still connect Avar-period Central Europe to medieval Mongolia across more than 6,500 km. Figure 1 illustrates these individual-level connections. While migrations occur in all directions, often zigzagging, the data recover the prevailing flows of major Holocene expansions, including the Neolithic farming expansion across Europe and the Yamnaya steppe

expansion. The implied rates of movement in our dataset align with existing estimates from the archaeological literature (Ammerman and Cavalli-Sforza, 1971; Racimo et al., 2020).

We aggregate this matrix of individual genetic connections to construct a bilateral panel of migration flows, discretizing space into $1^\circ \times 1^\circ$ gridcells and time into 500-year windows. We link this 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, using a gravity model of bilateral flows (Anderson and Van Wincoop, 2003), we show that migration declines sharply with geographic and ecological distance between origin and destination. A one-standard-deviation increase in growing degree days (GDD), precipitation, and soil composition distance reduces migration flows by 20%, 21%, and 16%, respectively. Each of these effects is equivalent to traveling an additional several hundred kilometers—a sizable effect relative to the 1,550 km average distance between genetically-linked pairs in our sample. The degree day magnitude is in line with Obolensky et al. (2024), who find that a 1°C increase in temperature distance reduces U.S. internal and international migration by 2%–27% over 1850–2020.

Our preferred specification incorporates a terrain-dependent travel cost, because basic geographic distance alone cannot represent the actual difficulty of movement. This measure incorporates variation in mountains, rivers, and coastlines, thereby capturing the effective cost of moving between locations in pre-modern environments. The specification further includes origin and destination gridcell fixed effects to absorb all time-invariant location characteristics—such as non-randomness in archaeological sampling—and period fixed effects to account for aggregate changes in sample availability and migration intensity over 500 year eras.

The results are robust to including origin-by-era and destination-by-era fixed effects that absorb all time-varying push and pull factors, as well as to controlling for additional environmental distances like UV-B radiation, elevation, and coastal proximity. We further replicate the analysis using an individual-level specification with over three million pairwise observations. This addresses concerns that granular variation in sparsely populated origin–destination cells may generate spurious correlations (Dingel and Tintelnot, 2026). We present additional robustness checks showing that our findings are stable across specifications and samples. The coefficients on precipitation and soil distance are consistently negative and precisely estimated; the coefficient on growing degree days is more sensitive to specification, reflecting its correlation with broader latitudinal gradients, which we discuss below.

Our second finding involves heterogeneity in the dimension of matching by subsistence technology. In terms of ecological distance, farmers, whose crop development depends on

accumulated thermal time (d'Alpoim Guedes and Butler, 2014; d'Alpoim Guedes and Bocinsky, 2018), match most strongly on growing-season heat and soil characteristics, with weaker sensitivity to precipitation. Pastoralists, whose forage base is governed more by precipitation regimes (Sala et al., 1988; McGuirk and Nunn, 2025), show the opposite pattern: precipitation matching is strongest, while thermal matching is weaker. Forager estimates are more sensitive to specification due to their smaller sample, but the distinctive feature of forager groups is the steepness of their distance penalty across ecological dimensions, consistent with landscape-indexed rather than climate-indexed knowledge. The fact that each group matches most closely along the ecological dimension its production technology is least capable of buffering is in line with the model's prediction. Moreover, no single geographic confounder can explain matching along all three dimensions simultaneously given that temperature largely varies along a latitudinal gradient, precipitation along coast-to-interior gradients, and soil with geology rather than broad geographic gradients.

The travel cost coefficients further differentiate the subsistence groups. Each additional 100 km of travel, equivalent to about 80 hours of walking in our terrain-adjusted measure, reduces pastoral migration by 2%, compared with 7% for farmers and 13% for foragers—reflecting differences in what each group stands to lose by moving. Pastoral wealth is livestock, which walks to new pasture. Farming requires land-intensive fixed investments—cleared fields, soil improvements, storage—that raise the cost of relocation. Foragers, with their landscape-specific knowledge, face the highest travel cost penalty, as they also do with ecological distance. We validate these patterns with specific archaeological culture groups: steppe pastoralists (e.g., Yamnaya) show little sensitivity to travel cost or thermal matching when migrating, whereas farming cultures incur a larger travel cost penalty and show stronger thermal matching.

Our third finding is that population movements respond to climate change. Warming expands the cereal-viable zone, increasing migration among farmers and foragers as the set of climatically suitable destinations broadens. Cooling shifts ecological conditions in favor of grassland and grazing, increasing migration among pastoralists as the steppe-viable zone expands. The "climate pulse model" (Olsson, 2024) anticipates the warming-driven dispersal of farmers and foragers; our pastoral result indicates that the same logic operates for grazing systems but in the opposite temperature direction.

Applying these estimated responses to realized Holocene temperature variation, our model recovers the broad temporal pattern of major expansions from the archaeological record: farmer migration peaks (5500–4500 BCE) during the Holocene Climatic Optimum, when the Neolithic LBK expansion carried farming across central Europe, while pastoral migration increases during Neoglacial cooling after 3000 BCE, coinciding with the Yamnaya,

Sintashta, and Scythian expansions. That a parsimonious climate-based model is consistent with these well-documented episodes lends further credibility to the ecological sorting mechanism.

Beyond these niche-expansion effects, migration also operates as a margin of adaptation in response to climate change over time. We document that populations exposed to warming systematically relocate toward cooler destinations, while those exposed to cooling move toward warmer ones, partially offsetting local climate variation. Quantitatively, a 1°C origin warming is associated with descendants settling approximately 0.4°C cooler than their ancestors—compensating for 40% of the local temperature change and indicating that relocation played a meaningful role in adjustment over the course of centuries.

Finally, we examine whether migration reshaped local ecological conditions by linking bilateral genetic flows to subsequent changes in destination vegetation, as measured by fossil pollen records. Although this analysis lacks random assignment of migrants to destinations and should be interpreted cautiously, we find that the vegetation in destinations receiving many migrants from a given origin converged with that of the origin. This result holds after controlling for origin, destination, and period fixed effects as well as changes in local temperature. These patterns align with the demic diffusion of farming practices and with migrants bringing climate-adapted expertise that transforms the ecology of the places they settle. Our analysis also supports the niche expansion interpretation by recovering the well-documented Holocene pattern in which warming promotes tree cover and agricultural expansion at the expense of open grassland (Davis et al., 2003; Marcott et al., 2013; Giesecke et al., 2014).

Our results resonate with a long tradition in anthropology, archaeology, and paleoecology emphasizing the role of ecological conditions in shaping human mobility. A central insight of this literature is that subsistence strategies were organized around local environmental constraints—what could be produced, where, and under what climatic conditions (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 (d'Alpoim 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 the social sciences to use IBD segments as a quantitative measure of bilateral migration flows. As noted

above, our implied displacement rates independently recover canonical estimates from the archaeological literature, providing strong external validation of the measure.

Our findings also bear on how populations adapt to their environments over the long run. When movement is systematically directed along corridors of ecological similarity, local environments act not only as constraints on mobility but as selection environments shaping the distribution of traits across space. In such settings, adaptation operates through the accumulation and transmission of local knowledge, as emphasized in the cultural evolution literature (Henrich, 2004; Boyd et al., 2011; Nunn, 2021). Persistent ecological sorting may therefore reinforce cultural and, potentially, biological adaptation to local conditions (Mathieson et al., 2015; Akbari et al., 2026), contributing to the long-run divergence of populations across environments.

A growing literature in economic history relies on novel data sources—such as trade records, transportation networks, and the spatial diffusion of economic objects—to recover patterns of economic interaction in data-scarce historical environments. Using shipment records from Assyrian merchants, Barjamovic et al. (2019) estimate bilateral trade flows and the location of ancient cities. Boehm and Chaney (2024) use the circulation of coins to infer trade and reconstruct regional economic dynamics in Late Antiquity, while Flückiger et al. (2022) and Flückiger et al. (2024) show that transport networks and trade corridors shaped both contemporaneous exchange and the long-run evolution of economic integration and urbanization. Hornbeck et al. (2024) exploit the location of ancient ports to study the spatial propagation of economic activity over the long run. These papers, like ours, use indirect proxies to recover spatial interactions and their persistence in settings where direct data are unavailable. We add to this literature by focusing on migration rather than trade and showing that ecological similarity—not just trade costs or network structure—was a key determinant of spatial interaction in pre-modern environments.

More broadly, 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. Relatedly, Spolaore and Wacziarg (2009) show that genetic distance between populations—a proxy for the time since populations shared a common ancestor—predicts barriers to the cross-country diffusion of development. We complement these works 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; Waldinger, 2022), as well as work showing how mobility and inter-group interactions shape conflict and the scope of trust (McGuirk and Nunn, 2025; Le Rossignol and Lowes, 2026). It also relates to the importance of skill–environment matching for economic outcomes and migration decisions (Bazzi et al., 2016; Obolensky et al., 2024). We extend these insights to a setting in which the institutional and market forces that shape modern migration—wage differentials, policy constraints, and formal borders—operate weakly or are absent, allowing the ecological margin to be isolated more cleanly than in modern data. Prehistory also provides variation across coexisting production systems—foragers, pastoralists, and farmers occupying the same landscape—that no modern dataset can replicate.

In addition, the 10,000-year timeframe of our study encompasses prolonged periods of warming and cooling that allow us to examine how climate change and ecology jointly shaped human mobility. 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. It is also consistent with historical patterns of climate-based sorting, ranging from 19th-century U.S. farmers seeking familiar growing conditions (Steckel, 1983) to European colonial success depending on environmental similarity between origin and destination (Crosby, 1986; Nicolò et al., 2026).

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. Section 5 outlines the empirical approach. Section 6 presents the results: climate matching, subsistence heterogeneity, adaptive migration, and ecological diffusion. Section 7 concludes.

2 Conceptual Framework

Prehistoric migration was shaped not only by physical distance but also 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. The model generates testable predictions about the relationship between ecological matching, subsistence technology, and migration, which we test in the empirical analysis below.

Setup. The idea that populations sort across environments based on accumulated ecological knowledge relates to the ecological concept of a "niche." In ecology, a species' niche is defined as the set of environmental conditions under which it can persist (Hutchinson, 1957). Building on the cultural niche framework developed in evolutionary anthropology and niche construction theory (Odling-Smee et al., 2003; Wiens and Graham, 2005; Boyd et al., 2011), we treat the relevant niche for prehistoric human populations as depending on subsistence technologies that determine which environmental dimensions are most costly to mismatch. Migration along corridors of environmental similarity can then be understood as a mechanism through which populations track these niches across space.

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 born at origin o with climate $\boldsymbol{\theta}_o$ inherits a stock of subsistence knowledge $h(\boldsymbol{\theta}_o)$ 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. This formulation is related to Bazzi et al. (2016)'s modelling of agroclimatic similarity as location-specific human capital.

Productivity when an individual from origin o occupies a destination with climate $\boldsymbol{\theta}_d$ is

$$f(h(\boldsymbol{\theta}_o), \boldsymbol{\theta}_d) = A - \sum_{k \in \{T,P,S\}} c_k \cdot |\theta_{d,k} - \theta_{o,k}|^{\alpha_k} \tag{1}$$

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

The individual chooses a destination d to maximize net payoff, where $\tau(\text{dist}_{od})$ is a transport cost increasing in geographic distance:

$$\max_d \left\{ A - \sum_{k \in \{T,P,S\}} c_k \cdot |\theta_{d,k} - \theta_{o,k}|^{\alpha_k} - \tau(\text{dist}_{od}) \right\} \tag{2}$$

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]

[1]For simplicity, 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

2.1 Heterogeneity by Subsistence Mode

The critical parameters are the α_k values, which we allow to vary across subsistence modes, and the transport cost τ, which varies by group's fixed investments. We treat subsistence type as a characteristic of populations, not individuals, reflecting the dominant mode of production at each origin. Three ex-ante considerations pin down the key orderings.

First, farmers depend on the accumulated thermal regime: crop varieties, planting calendars, and pest management are calibrated to narrow growing-season windows (d'Alpoim Guedes and Butler, 2014; d'Alpoim Guedes and Bocinsky, 2018), making thermal mismatch costly. Farming technologies—crop selection, stored seed, and water management—can help decouple productivity from local precipitation, concentrating vulnerability on the thermal and soil dimensions. Second, pastoralists' forage base is governed primarily by precipitation regime: grassland productivity is moisture-limited in the mid-latitudes (Sala et al., 1988; McGuirk and Nunn, 2025), while livestock tolerate a wider range of temperatures than any single crop. Third, soil characteristics constrain farmers more than pastoralists. Specific crops often require certain soil profiles in terms of drainage, acidity, nutrient content, etc. Early Neolithic farming concentrated on loess—which is typically fertile, well-drained, and lightly forested—with the LBK distribution mapping closely onto its extent (Bogucki, 1988; Shennan, 2018), while pastoral systems were less beholden to soil-tillage constraints.

These considerations imply parallel contrasts between farmers and pastoralists in *ecological-mismatch costs* (α_k):

$$\alpha_T^{\text{farmer}} > \alpha_T^{\text{pastor}}, \qquad \alpha_P^{\text{pastor}} > \alpha_P^{\text{farmer}}, \qquad \alpha_S^{\text{farmer}} > \alpha_S^{\text{pastor}}. \tag{3}$$

Farmers face binding constraints along the thermal (α_T) and soil (α_S) dimensions, while pastoralists face a binding constraint along the precipitation dimension (α_P). The soil prediction is less commonly discussed in the prior literature than the temperature and precipitation predictions, but follows from the same logic: Pastoralists, who do not till, are largely indifferent to soil workability, whereas farmers depend on it and on soil characteristics that determine which crops can be grown.[2]

Subsistence groups also differ in their sensitivity to transport costs. The farmer–pastoralist contrast reflects fixed investment differences: farmers' productive capital—cleared land, soil improvements, stored seed—must be abandoned upon relocation, while pastoralists' wealth is livestock that walks to new pasture, lowering their relocation cost. Forager knowl-

full selection model.

[2]It is worth noting that soil characteristics do influence grass species composition and forage productivity, factors which are relevant to pastoralists.

edge is less well described by the α_k framework: rather than being calibrated to climatic productivity, it is largely landscape-indexed—organized around cognitive maps of specific plant patches, game trails, and water sources—and depreciates with any geographic displacement rather than specifically with climate distance.[3] These considerations yield *transport costs* (τ):

$$\tau^{\text{forager}} > \tau^{\text{farmer}} > \tau^{\text{pastor}}. \tag{4}$$

Together, this discussion suggests that both the relevant dimension of ecological matching and the sensitivity to geographic distance vary systematically across subsistence systems—predictions we test empirically below.

2.2 Climate Change and Adaptive Migration

When climate changes, the individual's knowledge stock becomes less well matched to local conditions. Suppose the climate at origin o shifts by $\Delta\boldsymbol{\theta}_o$. The individual can stay and bear the increased mismatch, earning payoff

$$V^{\text{stay}} = A - \sum_k c_k \cdot |\Delta\theta_{o,k}|^{\alpha_k} \tag{5}$$

or migrate to a destination d at cost $\tau(\text{dist}_{od})$, as in equation (2). Migration occurs when the best available destination yields a higher payoff than staying. The resulting threshold depends on the ratio of climate sensitivity to relocation cost: groups with high α_k (strong mismatch penalties) and low τ (mobile assets) are mobilized by smaller climate shifts. Farmers face high relocation costs but steep thermal sensitivity; pastoralists face low relocation costs and broader thermal tolerance but greater precipitation sensitivity. The migration response to climate change therefore varies across subsistence type in both magnitude and direction.

This variation generates two empirically distinguishable 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, creating 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

[3]The extent to which foragers match on individual ecological dimensions remains an empirical question—their knowledge may transfer across similar biomes (generating matching through proxies such as growing season heat and soil characteristics) or it may be specific to a landscape that encompasses several biomes in the context of seasonal migrations.

niche.

2.3 Testable Predictions

The framework described above generates the following predictions:

1. Ecological matching. Migration flows decline with climate and soil distance between origin and destination, holding geographic distance constant.

2. Subsistence heterogeneity in multidimensional matching. Farmers match more tightly than pastoralists on the thermal and soil dimensions ($\alpha_T^{\text{farmer}} > \alpha_T^{\text{pastor}}$, $\alpha_S^{\text{farmer}} > \alpha_S^{\text{pastor}}$), reflecting growing-season specificity and soil workability constraints. Pastoralists match relatively more on precipitation ($\alpha_P^{\text{pastor}} > \alpha_P^{\text{farmer}}$), because moisture regime is a major factor in forage productivity. In terms of travel costs, the geographic distance penalty is steepest for foragers (landscape-indexed knowledge), intermediate for farmers (fixed agricultural investments), and weakest for pastoralists (mobile livestock): $\tau^{\text{forager}} > \tau^{\text{farmer}} > \tau^{\text{pastor}}$.

3. Adaptive migration: niche expansion and destination selection. Warming expands the cereal-viable zone and increases farmer migration, whereas cooling expands grasslands and increases pastoral migration. The two groups should therefore respond to temperature shocks in opposite directions. A related prediction is that migrants in general should sort toward destinations that offset local climate shifts. That is, warming at the origin induces movement toward cooler destinations, while cooling induces movement toward warmer ones.

4. Ecological diffusion. Migrants carry subsistence practices—cereal cultivation, animal husbandry, forest clearance—that reshape the destination's managed landscape. Locations receiving migrants from cereal-cultivating origins should subsequently exhibit higher cereal pollen; locations receiving migrants from pastoral origins should exhibit shifts in the grassland-forest balance consistent with expanded grazing. The mechanism operates through transmitted human practices rather than through changes in the underlying climate, and is bounded by what the destination's environment permits.

We test these four predictions in Section 6, where we provide evidence on climate matching (Section 6.1), subsistence heterogeneity (Section 6.2), adaptive migration (Section 6.3), and ecological diffusion (Section 6.4). We emphasize that this framework is intended as a guide to interpretation—providing a unified lens that connects the empirical evidence—rather than as a structural model to be estimated.

3 Data

This section describes the data. We combine information on migration flows inferred from ancient DNA with high-resolution measures of local environmental conditions—including temperature, precipitation, and soil characteristics—as well as information on subsistence systems and vegetation via fossil pollen records. Table A1 provides details on the construction and sources of all variables used in the analysis.

3.1 Ancient Human DNA as Migration Proxy

Prehistoric migration patterns are difficult to study because direct records of individual movements do not exist. To overcome this limitation, we leverage ancient DNA (aDNA) data that 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—approximately 10–15 generations. 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 of at least 8 cM.[4] This

[4]We focus on this threshold because segments of this length can be reliably detected in low-coverage ancient DNA and capture recent genealogical connections over a few hundred years (Ringbauer et al., 2024).

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 within a few generations rather than the long-run ancestry turnovers documented for the Neolithic and steppe expansions (Lazaridis et al., 2014; Allentoft et al., 2015), we restrict attention to pairs of individuals who lived within 500 years of one another. To focus on inter-generational connections rather than contemporaneous kinship, we exclude pairs separated by fewer than 100 years. We also exclude pairs buried fewer than 50 km apart, as these likely reflect local kinship or co-burial rather than migration between ecologically distinct locations. Finally, we restrict the sample to individuals dated to after 8000 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).[5]

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. Figure A1 characterizes the resulting sample: only 2.3% of the 3.1 million candidate ancestor-descendant pairs share an IBD segment of at least 8 cM, and these IBD-positive pairs span an average geographic distance of 1,550 km and time gap of 283 years. As documented by Ringbauer et al. (2024), individuals sharing close biological links can span continents within a few centuries.

For example, a single Yamnaya individual buried at Samara on the Russian steppe (3213 BCE) has documented descendants in two opposite directions: an Afanasievo descendant 2,600 km east in southern Siberia three centuries later, sharing 140 cM of IBD—indicating a recent common ancestor within a few generations—and a Corded Ware descendant 2,200 km west in Central Europe four centuries later, sharing 55 cM of IBD. The same individual is thus a documented common ancestor for both the eastward Bronze Age steppe expansion and the westward demic spread that brought steppe ancestry into Europe. Closer to the present, a Kushnarenkovo individual from the southern Urals (850 CE) shares 27 cM of IBD with a Magyar descendant in Hungary 130 years later, 2,800 km west, consistent with the

The 8 cM threshold also yields substantially more observations than stricter cutoffs—71,290 IBD-linked pairs at 8 cM, vs. 15,456 at 12 cM and 4,681 at 16 cM, a 5–15× reduction in sample size—while the underlying distance and time-gap distributions are broadly similar across thresholds (mean distance 1,551 / 1,407 / 1,214 km; mean time gap 283 / 267 / 260 years at 8 / 12 / 16 cM, respectively; Figure A1). We also replicate the analysis using alternative thresholds.

[5]We convert ancient DNA radiocarbon dates from years before present (BP, with present = 1950 CE) to BCE/CE as BCE = BP − 1950. Date cutoffs in the text are rounded to the nearest 1,000 years (e.g., the 10,000 BP sample cutoff is reported as 8000 BCE).

historically attested late 9th-century Magyar migration into the Pannonian Basin. Most of the longest-distance corridors in our sample involve pastoralist populations with some connection to the Eurasian steppe; Table A2 lists additional examples.

We then 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 gridcell pair in a given period, 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—excluding within-cell pairs, which likely reflect local kinship rather than migration and for which all climate distances are zero by construction—but within the same time window.[6]

Table 1 presents summary statistics for the gridcell pair panel. The unit of observation is an origin–destination gridcell pair in a 500-year era, yielding approximately 530,000 observations. Panel A reports alternative measures of migration intensity constructed from IBD segments. Our baseline dependent variable—the count of IBD-linked pairs at the 8 cM threshold—has a mean of 0.13 and a median of zero, reflecting the sparsity of the migration matrix (95.8% of cells record zero flow). Alternative measures of genetic flow display similar distributions.

3.2 Climate

We use three sources of climate information. Modern reanalysis fields measure spatial gradients in temperature (ERA5) and precipitation (TerraClimate); we use these to construct bilateral climate distances between origin and destination, the central regressor in our main analyses (Sections 5–6). Paleoclimate reconstructions from the Last Glacial Maximum Reanalysis (LGMR) measure temporal change in temperature at fixed locations; we use these to construct origin-specific climate shocks for the adaptation analysis (Section 6.3). The choice of field varies with the identifying signal: modern fields are the lower-error estimate of spatial gradients at any era we care about, while paleoclimate fields are the only source of temporal variation across the Holocene.

Modern climate fields. We measure temperature using ERA5 (Hersbach et al., 2020), a global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts that provides hourly fields at 9 km native resolution from 1940 onward. Hourly temperatures allow direct integration of growing degree days as a daily sum of degrees above a threshold; we use cumulative growing degree days above 10°C (GDD_{10}) as our baseline

[6]For robustness, we also run our analysis using non-aggregated, individual-level pairwise genetic flows.

thermal measure.[7] We compute average annual values for each $1^\circ \times 1^\circ$ gridcell over 1960–2000.[8]

We measure precipitation using TerraClimate (Abatzoglou et al., 2018), a high-resolution (≈4 km) global dataset that interpolates station-based monthly precipitation onto the WorldClim climatological baseline. TerraClimate is anchored to direct gauge observations and has substantially smaller measurement error for precipitation than reanalysis products that estimate rainfall as a model-derived flux.[9] We compute average annual precipitation for each $1^\circ \times 1^\circ$ gridcell over the same 1960–2000 window as the temperature data.

Figure A2 illustrates the spatial distribution of growing degree days (Panel A) and precipitation (Panel B) across the study region. GDD declines sharply with latitude, reflecting the north–south temperature gradient, while precipitation varies primarily along east–west and coast-to-interior patterns. These distinct spatial structures indicate that different climate dimensions capture separate aspects of the ecological environment, generating independent sources of variation in environmental similarity across locations.

Paleoclimate fields. For the adaptation analysis (Section 6.3) we require temporal change in temperature at origin locations across the Holocene. No modern field provides this, so we use a paleoclimate reconstruction. We use Last Glacial Maximum Reanalysis (LGMR) from Osman et al. (2021), a paleoclimate data assimilation product that combines marine geochemical proxies with the iCESM model prior developed in Tierney et al. (2020). LGMR provides 200-year mean temperatures on a $2.5^\circ \times 1.9^\circ$ grid spanning the Last Glacial Maximum to the present, with gridcell level 1-sigma uncertainty of 1–2°C.

We use LGMR rather than the alternative pollen-based reconstructions of Mauri et al.

[7] 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 GDD_{10} as a general measure of growing-season heat accumulation that captures thermal gradients relevant across subsistence systems; results are robust to alternative thresholds.

[8] For each 1° gridcell, climate and soil values are area means over the underlying high-resolution pixels (ERA5 ≈9 km, TerraClimate ≈4 km, SoilGrids ≈250 m), with water-masked pixels excluded. This ensures coverage on island and coastal gridcells (e.g., Sardinia, Crete, Iceland, the Hebrides) where the cell centroid pixel alone is often water-masked. For the handful of 1° cells whose entire footprint is water-masked under one or more sensors—most notably the submerged Doggerland Mesolithic site at $(53^\circ\text{N}, 2.5^\circ\text{E})$, but also small-island archaeological sites at Ibiza, Crete, the Brittany tidal islets, and Öland—we fill the missing value from the nearest 1° cell with finite coverage for that variable. This nearest-neighbor fallback is applied per variable and propagates into both the gridcell panel and the individual-level data; it affects 5 cells and approximately 25 individuals (<0.4% of the canonical sample), with each receiving a mainland-neighbor's modern climate value typically 50–200 km away.

[9] When both TerraClimate and ERA5 precipitation distance are included simultaneously in our main gravity specification, the TerraClimate coefficient remains negative, highly significant, and larger in magnitude, while the ERA5 coefficient flips sign relative to its standalone estimate and becomes only marginally significant; all other coefficients are unchanged. We interpret this as TerraClimate's station-anchored precipitation absorbing the bilateral signal, with ERA5's residual capturing reanalysis biases relative to the gauge network.

(2015) and Davis et al. (2003), or the multi-method synthesis of Kaufman et al. (2020), for three reasons. First, the data-assimilation framework integrates spatially sparse proxies with a physically consistent model prior, giving full-field coverage where pollen networks are sparse. Second, the iCESM prior carries an isotope-enabled hydrological cycle that allows internal consistency between temperature and precipitation fields. Third, LGMR's 200-year temporal resolution matches the adaptation analysis's identification window (Section 6.3), whereas pollen-based reconstructions are typically resolved at the millennial scale. We rely on modern data for precipitation throughout the analysis: Holocene precipitation reconstructions remain more uncertain and contested than temperature reconstructions, with disagreement even on the sign of mid-Holocene precipitation change in many regions (Smerdon et al., 2017; Herzschuh et al., 2023).

The reconstruction itself carries substantial uncertainty. Beyond the gridcell level error, different reconstructions disagree on the magnitude and even the direction of long-run Holocene temperature trends—the "Holocene temperature conundrum." Proxy-based stacks suggest mid-Holocene warmth followed by Neoglacial cooling (Marcott et al., 2013; Kaufman et al., 2020), while data-assimilation reconstructions favor model-driven warming through the Holocene (Osman et al., 2021; Erb et al., 2022). This conundrum is not fully resolved in the paleoclimate literature (Bova et al., 2021; Essel et al., 2024). The disagreement concerns *long-term trends* in Holocene mean temperature, not the high-frequency within-region anomalies that identify our adaptation analysis. Because we identify the adaptation regression from 400-year centered differences in local temperature at fixed origin locations, our estimates are less exposed to long-run reconstruction disagreement than analyses identified from absolute Holocene mean trends.

Figure A3 summarizes the temporal and spatial variation in reconstructed temperature over the Holocene used in our analysis. Panel A shows substantial fluctuations in absolute temperature, with alternating phases of warming and cooling and a pronounced rise into the Holocene Climatic Optimum (roughly 6000–3000 BCE) followed by Neoglacial cooling (Marcott et al., 2013). Panels C–F show that these shifts were spatially heterogeneous: even within similar latitude bands, some regions experienced warming while others cooled. Despite this temporal and regional variability in absolute conditions, bilateral gradients between any two locations remained largely preserved across the Holocene, as we discuss next.

Modern fields as a measure of past bilateral gradients. The bilateral matching analysis uses modern climate fields rather than paleoclimate reconstructions to measure spatial gradients between origins and destinations. Two observations support this choice.

First, our identification relies on *bilateral differences* between origin and destination, not on absolute levels. Holocene climate change is dominated by orbital forcing, which

is spatially coherent at the continental scale: when two locations both warm during the Holocene Climatic Optimum, the bilateral gradient between them is preserved even as their absolute temperatures shift (Kaufman et al., 2020; Osman et al., 2021). Even the regional heterogeneity visible in Panels C–F of Figure A3 is small relative to the cross-sectional bilateral spread our analysis exploits.

We quantify this directly in Appendix B.1 by computing the within-LGMR temporal evolution of bilateral gradients—the change in bilateral temperature distance between any era and the modern bin, holding the reconstruction fixed so that LGMR's own spatial biases cancel in the difference. The standard deviation of this within-reconstruction temporal residual is approximately 0.53°C, capturing 99.6% of the variance in bilateral paleoclimate temperature distances. Under the classical measurement error benchmark, the implied attenuation of our temperature distance coefficients is approximately 0.4%, biasing our estimates toward zero by an economically negligible amount.[10]

Second, even in the modern period, ERA5 is the more accurate estimate of spatial gradients than LGMR. ERA5 is anchored to billions of station and satellite observations and resolves spatial structure at 9 km native resolution; LGMR's $2.5° \times 1.9°$ native resolution smooths out spatial structure that ERA5 captures, and its 1–2°C gridcell level uncertainty is present at every era including 0 ka. Substituting LGMR for ERA5 in the bilateral analysis would therefore introduce additional bilateral measurement error of order 3°C at every era, large relative to the within-LGMR temporal evolution that substitution would in principle correct.

The same argument applies with greater force to the growing degree days measure specifically: LGMR provides only 200-year annual means, so computing GDD from LGMR would require imposing a model-derived seasonal cycle drawn from the iCESM prior. The paleoclimate literature has flagged seasonal cycle bias as a substantial source of error in Holocene reconstructions (Bova et al., 2021), and using LGMR-derived GDD would inherit this bias in addition to the spatial-resolution and gridcell uncertainty problems. Modern reanalysis is therefore the lower-error choice for the bilateral analysis on the spatial dimension we quantify here, with the seasonal cycle bias flagged by Bova et al. (2021) reinforcing this conclusion for the GDD measure specifically.

Climate distance. We select the climate dimensions based on a systematic analysis of correlations among candidate bilateral distance measures (Figure A4; for the corresponding per gridcell level correlations, see Figure A5). Temperature-related measures—growing degree

[10]The within-LGMR residual is mildly correlated with the bilateral paleoclimate gradient itself ($r \approx -0.44$ in the production sample), so classical measurement error is an approximation rather than an exact characterization. Even sizable departures from classical assumptions cannot meaningfully change the magnitude given how small the implied bias is.

days, summer temperature, and annual mean temperature—are highly correlated with one another and with latitude distance, forming a cluster from which only one representative should be included. We use cumulative GDD_{10} as our baseline thermal measure because it captures biologically relevant heat accumulation for production rather than ambient temperature alone. By contrast, winter temperature distance is strongly correlated with geographic distance ($r = 0.64$) and becomes statistically insignificant when GDD is included, suggesting that it primarily reflects geographic proximity rather than an independent dimension of climate matching. Annual precipitation provides a largely orthogonal moisture dimension ($r = 0.17$ with GDD distance), capturing variation in water availability that is distinct from thermal conditions. GDD distance is also largely orthogonal to geographic distance itself ($r = 0.20$), indicating that thermal matching captures more than spatial proximity.

3.3 Soil

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. Panels C and D of Figure A2 illustrate the spatial distribution of clay content and soil organic carbon. In contrast to temperature and precipitation, soil characteristics do not follow simple latitudinal or longitudinal gradients. Instead, they exhibit patchy spatial patterns that reflect underlying geological formations and local soil processes. This distinct structure implies that soil characteristics capture a largely independent dimension of environmental similarity, complementing the climatic variation documented in Panels A and B. We note, however, that modern soil characteristics partly reflect millennia of land management—tilling, planting choices, and deforestation—particularly for properties such as soil organic carbon. Geologically determined features (e.g., clay content, parent material) are more stable.

Soil distance. To capture non-climatic environmental constraints, we construct a measure of soil distance based on the Euclidean distance between clay content and soil organic carbon. We focus on these two components because they are the most geologically determined soil properties: clay content is largely driven by underlying substrate and is nearly uncorrelated with precipitation distance ($r = 0.09$), while soil organic carbon captures fertility differences observable to prehistoric cultivators. Together, the composite provides a dimension of environmental similarity that is largely orthogonal to climatic conditions, with modest correlation with thermal distance ($r = 0.33$ with GDD, reflecting partial latitudinal co-variation) and precipitation ($r = 0.29$), and low correlation with geographic distance ($r = 0.12$). Re-

sults are robust to using a PCA-based index or each individual soil component—clay, organic carbon, pH, water-holding capacity, and bulk density—in place of the composite.

3.4 Additional Environmental Characteristics

We also construct additional environmental controls used in robustness checks. We measure diurnal temperature range—the difference between daily maximum and minimum temperatures—from ERA5, which captures temperature variability that is relevant for crop stress, frost risk, and livestock management. We construct elevation distance using data from the Shuttle Radar Topography Mission (SRTM), defined as the absolute difference in mean elevation between origin and destination gridcells.

We include ultraviolet (UV-B) radiation distance, constructed from the glUV global dataset (Beckmann et al., 2014), which provides annual mean erythemal UV-B radiation at 15 arc-minute resolution derived from NASA Aura-OMI satellite measurements. UV-B radiation is determined primarily by solar zenith angle and varies almost monotonically with latitude ($r = 0.77$ with latitude distance), providing a measure of the north–south solar gradient that is moderately correlated with GDD distance ($r = 0.43$, reflecting the shared latitudinal component) and largely orthogonal to precipitation ($r = 0.11$). Including UV-B distance therefore serves both as a robustness check on the latitudinal component of GDD matching and as a test for an independent biological channel of climate-specific adaptation (Jablonski and Chaplin, 2000; Chaplin, 2004; Mathieson et al., 2015; Akbari et al., 2026).

3.5 Subsistence Modes

For the heterogeneity analysis (Sections 6.2 and 6.3), we classify each ancient DNA population by subsistence type—farmer, pastoralist, or forager—using a three-step procedure. First, each of the approximately 1,800 unique populations in the Ringbauer et al. (2024) dataset is assigned to one of approximately 50 archaeological culture groups. The primary assignment uses keyword matching of population labels against a curated dictionary of archaeological terms (e.g., "LBK" → Early European Farmers; "Yamnaya" → Early Steppe Expansion; "Corded Ware" → Corded Ware). Populations not matched by keywords are assigned via a country-and-date fallback that maps each country×date combination to the most likely culture group based on the archaeological record. A small number of remaining unmatched populations are assigned manually.

Next, each population is classified by subsistence type and assigned continuous subsistence shares (proportion farmer, pastoralist, forager, summing to one). This classification uses a large language model with reasoning ability (OpenAI o4-mini) prompted with each

population's name, geographic location, and date, and asked to assign subsistence shares based on the archaeological and ethnographic literature. The model's classifications were cross-validated against an independent rule-based classifier and manually curated; the two approaches agreed on the dominant subsistence type for over 95% of populations. The categorical subsistence type is the mode with the largest share. The culture group assignments and subsistence shares are then merged into a unified crosswalk that links each population to both its archaeological culture and its subsistence profile.

Figure A6 displays the evolution of subsistence composition over time: 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.

3.6 Vegetation and Land Use

For the ecological diffusion analysis (Section 6.4), we use the REVEALS vegetation reconstruction from Fyfe et al. (2021), which provides bias-corrected estimates of land-cover composition at 1° 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 pollen cores within 250 km, separately for each 500-year era; cells whose nearest pollen observation in a given era falls beyond that radius receive missing values for that era.

Figure A7 maps the spatial distribution of mean cereal pollen and the tree–grass balance across our study window, showing that coverage is dense in Scandinavia and northwestern Europe but sparser in the Mediterranean and eastern regions—a limitation we discuss in Section 6.4. Figure A8 complements this by showing changes in vegetation across consecutive periods, documenting the progressive expansion of cereal cultivation and regionally heterogeneous shifts in land cover that motivate the diffusion analysis below.

3.7 Travel Cost and Geographic Distance

We measure geographic distance using a terrain-adjusted least-cost-path travel time from the Özak (2018) Human Mobility Index (HMI), which estimates maximum sustainable walking speed at each location based on slope, terrain type, temperature, and humidity, with premodern seafaring on navigable seas. This allows movement across both land and maritime

routes in a way that reflects the constraints faced by pre-modern populations. To construct bilateral travel costs, we invert the HMI to obtain a cost surface (hours per kilometer) and compute the least-cost path between gridcell centroids, accumulating travel time along the path of lowest resistance.[11]

The resulting measure—reported in thousands of hours of walking—captures the cumulative cost imposed by terrain and natural barriers along optimal routes. For context, the median pair in our sample is separated by about 2,600 km of Haversine distance and 2,040 hours of travel time, implying an average pace of roughly 0.8 hours per kilometer. While travel time is strongly correlated with geodesic distance ($r = 0.91$), it better reflects the presence of barriers such as mountains, deserts, and water bodies. Figure A9 illustrates this difference by plotting travel-time isochrones from two distinct origins: Central Europe (Lyon) in Panel A and the Pontic Steppe in Panel B. The shape of the isochrones highlights pronounced asymmetries in travel costs, showing how terrain-adjusted distances diverge from standard geographic (Haversine) distance.

We verify that the results are robust to replacing HMI travel time with Haversine ("as-the-crow-flies") great-circle distance between gridcell centroids, using linear, quadratic, and logarithmic functional forms.

3.8 Summary of Bilateral Distance Measures

Panel B of Table 1 summarizes the main distance measures. Figure 2 provides a complementary visualization, mapping the distance between each gridcell and a reference location in central France (Lyon; 46°N, 5°E) for growing degree days (Panel A), precipitation (Panel B), and soil characteristics (Panel C). Temperature and precipitation distances exhibit substantial dispersion across origin–destination pairs, reflecting pronounced spatial heterogeneity. Soil distance likewise displays considerable variation and is only weakly correlated with climatic distances, consistent with its geological origins. The figure also shows that the three distance measures exhibit distinct spatial patterns, indicating that they capture separate dimensions of environmental similarity.

[11] A small number of gridcell centroids fall on water pixels in the HMI raster (typically coastal cells whose 1° centroid sits offshore); these are snapped to the nearest land pixel before computing the cost surface. The few remaining pairs in disconnected components of the HMI graph—chiefly Iceland and the Canary Islands—are imputed using a linear fit of travel time on Haversine distance estimated on the finite pairs ($\beta \approx 0.73$ hours/km, $R^2 = 0.94$). Together these adjustments affect less than 1% of the sample.

4 Descriptive Evidence

4.1 The Spatial Distribution of Ancient DNA Remains

The sample includes 7,032 individuals from Ringbauer et al. (2024), grouped into 500-year eras with midpoints spanning 7550 BCE to 1450 CE (9,500–500 BP). As is well known in the aDNA literature, the spatial distribution of recovered samples reflects both historical population patterns and differential preservation conditions. Because DNA degrades faster under heat and moisture (Reich, 2018), cooler and drier regions—particularly in northern and central Europe—are overrepresented in the data. Table A3 documents these patterns systematically. We consider $1^\circ \times 1^\circ$ gridcells across Western Eurasia and regress an indicator for the presence of at least one observed individual in a gridcell on the same environmental variables used in the main gravity specification.[12] Columns 1–3 sequentially add controls; column 4 adds $2.5^\circ \times 2.5^\circ$ gridcell fixed effects.

The probability of collecting viable ancient DNA is negatively associated with cumulative heat exposure and precipitation, and positively associated with clay content—consistent with established taphonomic processes whereby higher temperatures and moisture accelerate chemical degradation (Allentoft et al., 2012; Kistler et al., 2017), while clay-rich soils help stabilize genetic material (Lindahl, 1993). The coefficient on soil organic carbon is negative, consistent with more acidic, organic-rich soils accelerating DNA hydrolysis. The coefficients on clay and soil organic carbon remain stable when including $2.5^\circ \times 2.5^\circ$ gridcell fixed effects (column 4), while the GDD and precipitation coefficients are less precisely estimated, consistent with fixed effects absorbing the broad geographic gradients along which these variables vary.

While these correlations are informative for interpreting the data, they do not pose a threat to our empirical strategy. As discussed in Section 5, our analysis includes origin, destination, and era fixed effects and exploits variation across origin–destination pairs in environmental similarity. As a result, identification does not rely on cross-sectional differences in sample density, and time-invariant differences in preservation and sampling intensity—and their correlates—are absorbed by the fixed effects.

4.2 Ancient Migration Trends

Stylized facts. Figure 1 visualizes migration flows across Western Eurasia in three 1,000-year periods spanning 4500 to 1500 BCE, displaying links between locations where genetically

[12] Standard errors are clustered at the $2.5^\circ \times 2.5^\circ$ gridcell level to allow for spatial correlation across neighboring cells.

related individuals are observed.[13] Figure A10 shows the joint structure of geographic distance and time gap across IBD-positive pairs. Three patterns emerge. First, migration flows display a clear east–west orientation, consistent with individuals moving 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).

Second, genetic flow declines steeply with geographic and ecological distance, with most of the decline occurring over the first 1,500–2,000 km of geographic distance and similarly steep drop-offs along each ecological dimension. Figure A11 plots the mean IBD segment count as a function of four bilateral distance measures—geographic, GDD, precipitation, and soil composite—splitting each curve by the origin's subsistence type. Flows fall sharply at short distances across every dimension, consistent with migration being predominantly local to regional in scale and motivating the multiplicative gravity form we estimate below. The curves also reveal subsistence heterogeneity that we document formally in Section 6.2: farmers show the steepest decay with GDD distance, while pastoralists show the steepest decay with precipitation distance, consistent with the ecological constraints governing crop production and grazing, respectively. This heterogeneity—visible in the raw data—is one of the key empirical regularities the gravity framework will quantify.

Third, the geographic scale of migration changes markedly over time. Figure 3 documents a structural break around 3000 BCE: prior to this date, the average distance between genetically related individuals remains stable at 600–1,200 km, whereas afterward it increases sharply, peaking near 2,000 km around 2500 BCE before declining to approximately 1,300 km by the Roman period. The timing coincides closely with the adoption of the wheel and wagon in the Pontic-Caspian steppe (Anthony, 2007), though we interpret this date as a convenient marker for a broader cluster of transport innovations associated with the Secondary Products Revolution (Sherratt, 1981). These include dairying (Wilkin et al., 2021), ox-drawn traction, horseback riding (Librado et al., 2021), and wheeled transport, all emerging between approximately 4000 and 3000 BCE. The acceleration we document likely reflects this full bundle rather than any single technology. Panel B shows that this pattern is not driven by changes in sample size: although the number of observed individuals increases around 3000 BCE, it subsequently remains stable for several millennia, while migration distances remain persistently elevated.

[13] To reduce visual clutter, we only show pairs of individuals that share at least 16 centimorgans (cM) of identity-by-descent segments. As noted, our analysis uses a threshold of 8 cM—the minimum length at which the method in Ringbauer et al. (2024) reliably identifies IBD. Figure A1 compares pair counts and the underlying distance and time-gap distributions across the 8, 12, and 16 cM thresholds. As discussed below, our results are robust to alternative IBD length cutoffs. All subsequent figures and the regression analysis use the baseline 8 cM threshold unless otherwise noted.

Implied migration speed. The ratio of geographic distance to temporal separation between related individuals yields an implied displacement rate. Figure 4 plots the median rate by era and origin subsistence type. Pre-4000 BCE, median rates are around 2 km per year, in line with 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 median rates of 5–6 km per year, with pastoral-origin pairs reaching above 7 km/yr during the Yamnaya/Afanasievo era, exceeding earlier archaeological estimates of the Yamnaya expansion at approximately 4 km per year (Racimo et al., 2020). This correspondence validates the IBD measure as a quantitative proxy for prehistoric population movements.

Mobility by subsistence type. In the analysis below, we distinguish three main subsistence modes: foragers, pastoralists, and farmers (with agro-pastoralists as an intermediate category). Migration patterns differ markedly across these groups. Figure 4 shows that farmers consistently exhibit lower implied migration speeds than pastoralists across all eras. The aggregate acceleration in migration around 3000 BCE reflects both a technology effect—the adoption of the wheel—and a composition effect driven by the rising share of pastoralists (Figure A6). These patterns are consistent with the estimates presented in Section 6.2, where the implied distance penalty is substantially smaller for pastoralists than for foragers. Table A4 reports the underlying mean migration distance, time gap, and median speed by subsistence type, pooled across eras.

Culture 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 taxonomy, with some categories consolidated into more homogeneous groupings. Figure A12a 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 overlap, reflecting repeated episodes of migration and admixture (Lazaridis et al., 2014; Allentoft et al., 2015). Figure A12b 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 Section 6.2.

5 Empirical Strategy

5.1 Estimating Equation

Our empirical analysis builds on the origin–destination migration panel constructed from aDNA data, as described in Section 3. The unit of observation is a pair of locations in a given (500-year) period. For each period, we consider the set of gridcells 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.

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 ecological 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:

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

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 6–10 generations ago. Ringbauer et al. (2024) identify 8 cM as the minimum segment length at which the ancIBD method reliably detects IBD in ancient DNA. For this reason, 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 using longer segment thresholds of 12 and 16 cM.

$\text{Dist}_{od}^{Ecology}$ captures differences in environmental conditions between origin and destination along three dimensions: temperature, precipitation, and soils. As described in Section 3, we measure thermal distance using differences in cumulative growing degree days (GDD),

precipitation distance using differences in average annual precipitation, and soil distance using a composite index based on clay content and soil organic carbon. These measures are selected to capture distinct ecological constraints: temperature and precipitation span the main climatic gradients, while soil characteristics provide a largely orthogonal, geologically determined dimension of environmental similarity (see also Figure 2).

To address concerns related to the spatial correlation of environmental conditions and migration, we control for the measure of travel costs described in Section 3.7. The baseline specification also includes origin fixed effects (α_o), destination fixed effects (α_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 ecological similarity, holding constant origin-specific, destination-specific, and period-specific factors. The identifying assumption is that, conditional on these fixed effects and travel costs, ecological distance between origin and destination is uncorrelated with unobserved bilateral determinants of migration flows. Standard errors are clustered at the cell pair level.

5.2 Threats to Identification

There are four remaining concerns. First, the preservation and discovery of ancient DNA is itself correlated with local climatic and environmental conditions (see also Section 4.1). This raises two related issues. The first is sample selection: locations with unfavorable preservation conditions may be underrepresented in the data, limiting external validity and potentially biasing estimates if selection is correlated with ecological distance across origin–destination pairs. The second is measurement: conditional on being observed, locations with better preservation may yield more individuals and thus more observed genetic links, artificially inflating measured migration flows. We mitigate these concerns by including origin and destination fixed effects, which absorb all time-invariant differences in preservation across locations. As a result, identification relies on within-sample comparisons across destinations for a given origin, rather than on cross-sectional differences in sample density.

Second, because environmental conditions are spatially autocorrelated, ecological distance may be correlated with geographic proximity and with other bilateral characteristics that independently shape migration, such as geographic barriers, travel routes, or cultural linkages. Since these factors operate at the origin–destination level, they are not fully absorbed by location fixed effects. To mitigate this concern, our baseline specification controls for terrain-adjusted travel costs, which capture the fact that similarity in these frictions de-

clines with geographic separation between origin and destination. To further assess robustness, we use alternative measures of distance—including linear, quadratic, and logarithmic transformations of geographic distance—and control for additional bilateral characteristics, such as elevation, coastal access, and latitude. Moreover, the multidimensional structure of our results—where different subsistence groups respond to different ecological dimensions—provides further reassurance that the estimated effects are unlikely to be driven by a single omitted bilateral confounder.

Third, migration flows may be influenced by origin- or destination-specific shocks that affect migration independently of ecological similarity. These shocks pose a problem if they are correlated with ecological distance and generate systematic differences in migration flows across destinations that are not driven by ecological matching. We address this concern by showing that results are robust to including origin-by-period and destination-by-period fixed effects, which absorb all time-varying push and pull factors at the origin and destination.

Finally, the sparsity of the migration matrix implies that many origin–destination cells contain few observed links, raising concerns about bias from granular variation in thin cells (Dingel and Tintelnot, 2026). To address this concern, we perform a permutation test that randomly reshuffles climate and soil distances across origin–destination pairs within each era and re-estimates the baseline specification. We then compare the resulting distribution of placebo coefficients to the baseline estimates, and show that our results lie well outside the range implied by random assignment.

6 Results

This section presents the main empirical results, and tests the predictions of the framework in Section 2. We proceed in four steps, each corresponding to one of the model's predictions. First, we document climate matching in migration, showing that mobility declines with ecological distance between origin and destination (Section 6.1). Second, we examine heterogeneity across subsistence systems, demonstrating that the relevant dimension of ecological matching depends on production strategy (Section 6.2). Third, we study migration as a margin of climate adaptation, documenting that populations respond to temperature shocks by relocating in ways that offset local changes, and that these responses vary systematically across subsistence systems (Section 6.3). Finally, using interpolated pollen records from sediment cores, we show that genetic flow predicts subsequent convergence in destination vegetation toward the pollen profiles of migrants' origins, consistent with the demic diffusion of agricultural and pastoral practices (Section 6.4).

6.1 Climate Matching and Migration

6.1.1 Baseline Estimates

Table 2 presents estimates of equation (6). All columns include the same three ecological distance measures—GDD, precipitation, and soil composite—along with geographic distance, while progressively introducing more demanding fixed-effects structures. Column 1 reports a parsimonious specification without fixed effects. All distance measures enter with negative and statistically significant coefficients, indicating that migration flows decline with both ecological and geographic distance. Adding era fixed effects in column 2 modestly shifts the precipitation and soil coefficients while leaving GDD and travel time essentially unchanged; the signs and significance of all four distance measures are preserved, indicating that aggregate time trends are not the principal driver of the results. In column 3, the inclusion of origin and destination fixed effects leads to a substantial attenuation of the GDD coefficient (from -1.10 to -0.318), consistent with the absorption of persistent differences in climate and migration intensity across locations. Despite this more demanding specification, all coefficients remain negative and statistically significant.

Column 4 presents our preferred specification, which includes origin, destination, and era fixed effects. The coefficients remain stable relative to column 3, confirming that the results are not driven by time variation in aggregate migration intensity. In this specification, a one-standard-deviation increase in GDD distance—approximately 700 degree-days, corresponding to roughly a 3.5°C difference in mean temperature over a 200-day growing season—is associated with a reduction in migration of about 20%. Precipitation distance has a comparable effect: a one-standard-deviation increase is associated with a 21% reduction in migration. Soil distance is somewhat smaller: a one-standard-deviation increase is associated with a 16% reduction. The implied effects are each equivalent to several hundred kilometers of additional travel, indicating that ecological differences shape migration at least as strongly as substantial physical separation. For temperature, these estimates are also consistent with evidence from modern settings: Obolensky et al. (2024) find that a 1°C increase in temperature distance reduces internal and international migration in the United States over the period 1850–2020 by 2%–27%, placing our estimate well within the range observed in contemporary data.

Columns 5 and 6 introduce origin-by-era and destination-by-era fixed effects, which absorb all origin and destination push and pull shocks. The estimated coefficients remain negative and statistically significant across all environmental dimensions, confirming that the results are not driven by time-varying changes in origin or destination characteristics. A key feature is that ecological matching emerges consistently across three environmen-

tally distinct dimensions—temperature (latitudinal), precipitation (coast-to-interior), and soil (geological)—so no single omitted geographic factor is likely to generate matching along all three simultaneously. The subsistence-specific crossover documented in Section 6.2 further strengthens this argument.

The estimates line up closely with the model's first prediction that migration responds to the full bundle of ecological frictions rather than to a single climatic index (Section 2). In the model, movement is easier when origins and destinations are similar along the dimensions that matter for production; correspondingly, the negative coefficients on GDD, precipitation, and soil distance indicate that environmentally similar places are more tightly linked in the migration network. The fact that soil distance enters with a magnitude comparable to GDD and precipitation distances suggests that ecological matching was not confined to temperature or moisture, but also operated through locally embedded, less easily transferable features of the production environment.

6.1.2 Summary of Robustness Checks

In this section, we briefly summarize the robustness checks, which are presented in detail in Appendix B.2. First, we augment the baseline specification with additional environmental controls—including ultraviolet (UV) radiation distance, diurnal temperature range distance, elevation distance, indicators for coastal-to-non-coastal movements, and latitude and longitude distance—and show that results remain stable (Table B2). Second, we replace the composite soil index with the first principal component of a broader set of soil characteristics and, alternatively, include these characteristics one at a time (Table B3). Third, we examine the role of spatial structure by separately considering horizontal (east–west) and vertical (north–south) moves and by restricting the sample to pairs of locations separated by at least 100, 200, and 300 km (Table B4). Fourth, we verify that the results are robust to alternative specifications of geographic distance: replacing the baseline HMI travel time with Haversine distance in linear, quadratic, or log form yields coefficients on precipitation and soil distance that are similar in magnitude and significance (Table B5). Fifth, we replicate the analysis using more stringent thresholds of 12 cM and 16 cM to define migration events, capturing more recent genealogical connections (Table B6).

Sixth, we show that the results are robust to winsorizing extreme observations, and omitting each 1,000-year period in turn (Figure B1). Seventh, we estimate an individual-level specification based on more than 3 million pairwise links (Table B7). Eighth, we assess robustness to alternative clustering structures for standard errors (Table B8). Finally, to address concerns that granular variation in sparsely populated origin–destination cells may generate spurious correlations (Dingel and Tintelnot, 2026), we perform a permutation test

that reshuffles ecological distances within each era and shows that our baseline estimates lie well outside the distribution obtained under random assignment (Figure B2).

Overall, the results are robust across a wide range of specifications and alternative samples. The coefficients on precipitation and soil distance are consistently negative and precisely estimated. By contrast, the estimates for GDD are more sensitive to specification choices: they attenuate and become statistically insignificant when controlling for highly correlated measures of solar exposure and when adopting alternative distance specifications or stricter definitions of migration.

6.2 Subsistence Heterogeneity

6.2.1 Main Findings

Having established that migration declines with ecological distance, we now turn to the second prediction of the framework outlined in Section 2: the dimension of climate matching should vary across subsistence types—farmers, pastoralists, and foragers. We also consider agro-pastoralist populations that combine farming and herding activities—as an intermediate category. We classify origin populations into subsistence groups using the archaeological culture group classification described in Section 3, and estimate the baseline specification separately for each group. Table 3 presents the results. Column 1 replicates the preferred specification from Table 2 in unstandardized units for comparability. Columns 2–6 standardize all regressors within each subsample (mean zero, unit standard deviation) to facilitate cross-variable comparability. Column 2 pools the full sample; columns 3–6 report estimates separately for farmers, agro-pastoralists, pastoralists, and foragers.

The main takeaway is that the dimension of ecological matching that binds depends on subsistence technology. Farmers exhibit strong GDD matching (-0.44) and substantial soil matching (-0.20), but null precipitation matching ($+0.03$, not statistically distinguishable from zero). Pastoralists show the opposite pattern: strong precipitation matching (-0.37) but null GDD (-0.03) and soil (-0.02) matching. The crossover is consistent with the framework's prediction that binding constraints differ across production systems—thermal and soil conditions for crop production, moisture for forage—rather than implying that any group is necessarily unresponsive to its non-binding margin. Agro-pastoralists—who combine farming and herding activities—also exhibit strong GDD matching (-0.46), with a smaller negative coefficient on soil (-0.10) and an insignificant precipitation coefficient, consistent with their joint reliance on crops and livestock. Foragers—the smallest subsample (10,801 origin–destination–era observations)—show significant matching on GDD (-0.35) but smaller and statistically insignificant matching on soil (-0.11) and precipitation (-0.13); what most dis-

tinguishes foragers is their steeper travel-cost penalty (discussed below), consistent with the idea that forager knowledge depreciates rapidly with geographic displacement.

The farmer–pastoralist contrast is informative. Farmers match on the thermal and soil 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.[14] Neither group matches significantly on the other's binding dimension. These patterns line up with the model's second prediction: that subsistence technology determines which ecological axis is hardest to buffer. The travel-cost coefficients reinforce this interpretation: pastoralists show the weakest travel-time penalty (-0.36), reflecting the mobility of livestock; farmers a steeper penalty (-1.6), reflecting fixed agricultural investments; and foragers the most localized movement (-3.0), consistent with landscape-indexed knowledge that depreciates rapidly with any geographic displacement.

These results also strengthen the credibility of the empirical design. Temperature, precipitation, and soil characteristics exhibit distinct spatial gradients—temperature varies primarily with latitude, precipitation along coast-to-interior patterns, and soil properties with underlying geology. As a result, it is difficult to rationalize the observed pattern with a single omitted geographic confounder. Moreover, any such confounder would need not only to generate matching along multiple ecological dimensions, but also to do so in a way that differs systematically across subsistence groups.

6.2.2 Heterogeneity by Culture Groups

As a complementary approach, we examine heterogeneity directly across named archaeological culture groups. The 11 groups comprise five farming cultures (Early European Farmers, Bell Beaker, Aegean Bronze Age, Roman, and Germanic—including Viking), five pastoral cultures (Yamnaya, Corded Ware, Indo-Iranian, Iron-Age Nomads, and Steppe Turkic), and one forager group (Hunter-Gatherers, pooling WHG and EHG). Table A5 lists the groups along with their corresponding time periods, subsistence classifications, brief descriptions, and key references. Figure 5 reports coefficients on travel costs (Panel A), GDD distance (Panel B), precipitation distance (Panel C), and soil distance (Panel D) from the baseline gravity specification, estimated separately for each group. The full-sample coefficient is shown at the top of each panel in dark grey, while culture-specific estimates are color-coded

[14] The concentration of precipitation matching among pastoralists 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 less viable in drying regions.

by subsistence mode—light blue for farmers, green for foragers, and yellow for pastoralists. The results validate the subsistence-based patterns at a finer resolution across all four ecological dimensions.

Panel A confirms that the travel-cost penalty varies by subsistence type. Steppe pastoral cultures (Yamnaya, Corded Ware, Indo-Iranian, Steppe Turkic) exhibit near-zero distance sensitivity, consistent with continental-range herding mobility. Farming cultures show steep distance penalties, with Early European Farmers, Roman, and Germanic groups all clustering between -0.5 and -1.0. Hunter-gatherers are the most spatially constrained.

Panel B shows the thermal matching crossover: farming cultures display negative GDD coefficients (stronger thermal matching), while pastoral cultures cluster around zero (null thermal matching). Panel C reveals the mirror image for precipitation: pastoral cultures—especially Steppe Turkic, Indo-Iranian, and Corded Ware—show strong precipitation matching, while most farmer groups are near zero. Notably, Roman-era populations show weaker precipitation matching than other farming cultures, potentially reflecting the development of irrigation infrastructure that partially decoupled agricultural productivity from local rainfall.

Panel D confirms that soil matching is concentrated among farming cultures, consistent with the workability constraints emphasized in the framework.

6.2.3 Robustness Checks

In this section, we validate the split-sample findings from Table 3 in two ways. The corresponding tables and a more detailed discussion of these results are provided in Appendix B.3.

Individual-level specification. First, Table B9 reproduces the subsistence heterogeneity at the individual-pair level. Across more than 3 million pairwise links, farmer migration responds to GDD but not to precipitation (-0.50 vs. an insignificant 0.02), while pastoralist migration responds to precipitation but not to GDD (-1.25 vs. an insignificant -0.05). Both dimensions of climate matching map onto distinct subsistence technologies, consistent with the gridded results.

Interacted specification. Second, Table B10 confirms the split-sample findings from Table 3 using a pooled specification with interactions between environmental distances and subsistence shares (with foragers as the omitted category). The results closely mirror those in Table 3: GDD matching is strongest among farmers and weakest for pastoralists, while precipitation matching is concentrated among pastoralists and negligible for farmers. In the joint specification, the same pattern holds across dimensions, with some attenuation for soil effects, which we interpret with caution given their sensitivity relative to the split-sample estimates.

6.3 Adaptation: Migration in Response to Climate Change

We now turn to the third prediction of the framework: how populations respond when local environmental conditions change. Our long time horizon spans periods of substantial climatic variation across both time and space (Figure A3), allowing us to study migration as a margin of climate adaptation. If individuals possess climate-specific knowledge and are adapted to particular ecological conditions, changes in local climate should reduce the value of that knowledge, creating incentives to relocate toward environments that better match their existing capabilities. Concretely, populations exposed to warming (cooling) should move toward cooler (warmer) locations. The framework further predicts that both the magnitude and direction of this response vary across subsistence groups, reflecting differences in environmental sensitivity across production systems and in relocation costs associated with the mobility of productive assets. We now test these predictions.

6.3.1 Climate Niche Expansion

We begin by examining how the overall level of migration responds to climate shocks and how this response varies across subsistence types. To do so, we re-estimate the baseline gravity specification in equation (6) by adding an origin-specific temperature change, ΔT_{ot}, computed as the reconstructed temperature 200 years after period t minus the temperature 200 years before, spanning a 400-year window centered on t (positive values indicate warming).[15] Table 4 reports the results. Column 1 presents estimates for the full sample of origin–destination–period cells that can be classified by subsistence type, while columns 2–5 report results separately for farmers, agro-pastoralists, pastoralists, and foragers.

On average, warming at the origin increases migration, consistent with the "climate pulse model" (Olsson, 2024), which posits that periods of warming expand the set of viable locations and promote population dispersal, while cooling leads to contraction and retreat into more favorable environments. The remaining columns reveal a notable asymmetry across subsistence systems. Warming at the origin increases migration among farmers (coefficient 0.89) and foragers (1.82), while it decreases migration among pastoralists (-2.34); the estimate for agro-pastoralists is negative but statistically indistinguishable from zero. Table A6 shows a qualitatively consistent pattern using continuous subsistence share interactions in the pooled sample.[16]

[15] As discussed in Section 3, the underlying paleoclimate data are available at a coarser spatial resolution ($2.5° \times 1.9°$). For this reason, each temperature change is assigned to all $1° \times 1°$ gridcells within the corresponding paleoclimate cell.

[16] Columns 2–4 each interact origin temperature change with one subsistence share at a time. Column 5 includes farmer and pastoral interactions jointly, with forager as the omitted reference group. Agro-pastoralists are not separately estimated; they enter through their farmer and pastoral shares.

Figure 6 shows these patterns in the raw individual-pair data: the unconditional bin-scatter of migration volume against origin temperature change (Panel A) yields a sharply negative slope for pastoralists (−0.213, SE 0.012) and a near-zero slope for farmers (0.010, SE 0.005), consistent with pastoralists having relatively more descendant connections during cooling and farmers during warming. Conditioning on origin and era fixed effects yields a similar pattern (Panel B): the slope remains negative for pastoralists, though attenuated, and becomes positive and statistically significant for farmers, indicating that these responses are not driven by cross-sectional differences across locations. These conditional patterns align closely with the regression estimates in Table 4, which likewise exploit within-origin and within-era variation.

This asymmetry is consistent with the framework in Section 2, and reflects how climate shocks alter the relative viability of subsistence strategies: warming expands the agricultural frontier, increasing the returns to crop production and inducing farmer migration, whereas cooling shifts ecological conditions in favor of grassland and pastoral production. The negative pastoral response to warming is consistent with both this direct ecological mechanism and a competitive-displacement mechanism in which farmer expansion under warming pushes pastoralists out of intermediate land; our framework and data do not separately identify these channels. The positive forager response to warming is consistent with warmer conditions expanding biological productivity and the habitable range (Olsson, 2024): foragers, who are not tied to specific crops or pasture, can shift activity into newly-habitable territory as the cold-tolerance margin retreats, and warmer conditions raise the resource base on which mobile populations depend. The forager coefficient is also less precisely identified than the farmer or pastoral coefficients given the smaller subsample (10,801 origin–destination–era observations).

While some pastoral expansions occurred under warm conditions and technological innovations such as dairying likely complemented climatic factors (Wilkin et al., 2021; Dong et al., 2023), the broad pattern is consistent 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).

Figure 7 translates the estimated climate responses into predicted migration dynamics

over the Holocene. Panels A and B summarize the magnitude and spatial extent of climate shocks over time, plotting, respectively, net temperature changes at 200-year intervals, averaged across origin gridcells in the sample, and the share of gridcells experiencing warming versus cooling in each period.[17] For each 200-year interval, we apply the estimated subsistence-specific coefficients to the reconstructed temperature shocks at each origin gridcell (Osman et al., 2021), weighting by local subsistence shares.[18] We plot the resulting predicted migration response implied by the model in Panel D.

The predicted patterns closely match the historical record. Farmer migration peaks between 5500 and 4500 BCE, coinciding with the LBK expansion across central Europe shown in Panel C, which marks the timing of major archaeological migration episodes by subsistence type. This peak reflects both stronger warming and a rising share of farmers in the population, highlighting the joint role of climate shocks and subsistence composition. By contrast, pastoral migration is predicted to increase during Neoglacial cooling episodes after 3000 BCE, consistent with the timing of the steppe pastoral expansions discussed above. Overall, the model recovers the broad temporal pattern of subsistence-specific migration using climate variation alone.

6.3.2 Adaptive Sorting

If individuals value local climatic conditions—or are adapted to them through climate-specific knowledge—then changes in the local environment should induce movement toward more favorable conditions. The conceptual framework in Section 2 predicts that populations exposed to warming (cooling) at the origin will relocate toward cooler (warmer) environments, partially offsetting local climate shocks.

We begin by presenting reduced-form evidence at the individual level, using variation across ancestor–descendant pairs. Because the sample conditions on observed IBD links, this analysis captures the *directional* margin of adaptation—where descendants settle, conditional on a documented migration—complementing the gridded PPML below, which captures the joint extensive-and-directional margin. A related caveat is that maladaptive migrants who left fewer descendants are likely underrepresented in the aDNA record. Figure 8 plots binned averages of ancestor–descendant pairs, relating temperature change at the origin to

[17] The patterns in Figure 7 differ from Figure A3 in what they emphasize: Figure A3 shows absolute temperature levels over time and spatial patterns of change across consecutive 1,000-year windows, while Panels A and B of Figure 7 focus on the sample-weighted climate shocks actually experienced by the origin cells in our aDNA data.

[18] 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.

the temperature difference between destination and origin locations. Panel A reports the unconditional relationship; Panel B residualizes both axes on origin and era fixed effects, showing that the same negative pattern obtains within origin–era cells. The negative and statistically significant slope in both panels indicates that populations exposed to warming at the origin tend to relocate toward cooler destinations, while those exposed to cooling move toward warmer ones. This pattern is consistent with migration as a margin of climate adaptation. Appendix B.4 corroborates this evidence using individual-level OLS regressions; a 1°C warming at the origin is associated with relocation to destinations between roughly 0.4°C and 1.1°C cooler, depending on the specification, with our preferred estimate—which adds destination fixed effects and climate-distance controls—at the lower end (Table B11). Figure B3 provides suggestive evidence of an analogous precipitation-adaptation pattern using δ^{18}O of precipitation from the Osman et al. (2021) reanalysis as a moisture-margin proxy; we present this as suggestive rather than as a coefficient estimate because $\delta^{18}O_p$ integrates temperature, precipitation amount, moisture-source, and seasonality channels that cannot be cleanly separated.

Motivated by this evidence, we test adaptive sorting more systematically by augmenting the gravity framework in equation (6) to allow migration flows to respond to the interaction between climate shocks at the origin and the temperature difference between origin and destination:

$$M_{odt} = \exp[\alpha_o + \alpha_d + \delta_t + \beta_1 \Delta\text{Temp}_{ot} + \beta_2 \Delta\text{Temp}_{ot} \times \Delta\text{Temp}_{od} + \mathbf{X}'_{od}\gamma]\epsilon_{odt} \tag{7}$$

where all variables are defined as in previous specifications. The key term is the interaction between the origin-specific temperature shock, ΔTemp_{ot}, and the spatial temperature difference between destination and origin, $\Delta\text{Temp}_{od} = (T_d - T_o)$. This interaction captures whether climate shocks at the origin induce systematic reallocation toward destinations with more favorable temperature conditions.[19]

We report results in Table 5. Column 1 includes the origin temperature shock, its interaction with the temperature difference between origin and destination, travel costs, as well as era, origin, and destination fixed effects. The interaction coefficient is negative and statistically significant, indicating that warming at the origin shifts migration flows toward cooler destinations. Columns 2 through 4 sequentially add climate distance controls—GDD distance in column 2, precipitation distance in column 3, and soil composite distance in column 4. The interaction coefficient remains negative and statistically significant at each step. This suggests the adaptive sorting finding is not mechanically explained by any of the

[19]Standard errors are clustered at the gridcell pair level.

observed climate distance covariates. To interpret the preferred specification in column 4: for each 1°C of origin warming, a destination that is 1°C warmer than the origin receives approximately 7% fewer migrants relative to a destination at the same temperature. Column 5 shows that the interaction is virtually unchanged when we replace the 1° origin and destination fixed effects with fixed effects defined at the 2.5° paleoclimate cell (2.5° × 1.9°, the native resolution of the shock), ruling out the concern that the 1° gridding interacts mechanically with the shock's coarser spatial resolution. Table B12 reports the interaction coefficient across three alternative clustering schemes (pair, origin, and origin × era) for both the 1° and 2.5° paleoclimate-cell specifications. At the 1° resolution, the result is significant under pair-clustering but not under coarser cluster definitions. At the 2.5° paleoclimate-cell resolution—the shock's native spatial scale—the magnitude is larger (-0.104) and the result survives both pair- and origin-level clustering, becoming statistically insignificant only under the most aggressive origin × era clustering.

6.4 Ecological Diffusion

The fourth prediction of the framework in Section 2 is that migration can reshape the ecology of destination locations through the diffusion of subsistence-specific knowledge. If migrants carry such knowledge, they may alter local land use and vegetation at their destinations. We test this hypothesis using REVEALS bias-corrected vegetation reconstructions (see also Section 3.6), examining whether genetic flow predicts subsequent changes in land cover at the destination. We focus on three outcomes—Cerealia-type pollen, grassland, and tree cover—and ask whether migration leads to convergence in vegetation toward the ecological characteristics of origin locations. Over our study period, land use changes substantially: cereal cultivation expands (the share of cells with positive cereal change rises from 35% to 53%), while the tree–grass balance shifts in regionally heterogeneous ways (Figure A8), providing empirical motivation for the diffusion test.

Using the gridded origin–destination–period dataset, we estimate regressions of the form:

$$\Delta\text{Pollen}_{dt+1} = \alpha_o + \alpha_d + \delta_t + \beta_1 M_{odt} + \beta_2 \Delta\text{Pollen}_{odt} + \beta_3 (M_{odt} \times \Delta\text{Pollen}_{odt}) + \mathbf{X}'_{odt}\gamma + \varepsilon_{odt} \quad (8)$$

where $\Delta\text{Pollen}_{dt+1}$ denotes the change in vegetation at the destination between periods t and t+500, M_{odt} is the number of genetically-linked pairs (≥ 8 cM) between origin and destination in the preceding period ($t - 500$ to t), and $\Delta\text{Pollen}_{odt}$ is the difference in vegetation between origin and destination at time t. The use of lagged migration flows ensures that movements precede changes in destination vegetation, mitigating concerns of reverse causality whereby ecological changes at the destination could attract migrants.

The model also includes origin (α_o), destination (α_d), and era (δ_t) fixed effects, absorbing all time-invariant differences across locations and common temporal shocks. We further control for a vector of bilateral characteristics ($\mathbf{X}_{odt}$), including ecological distances, travel costs, as well as contemporaneous temperature changes at the destination based on Osman et al. (2021) reconstructions. In addition, we interact the distance measures with the initial vegetation gap between origin and destination. These interactions flexibly capture proximity-driven spatial spillovers in vegetation, allowing us to isolate the contribution of migration flows to ecological change.[20]

This analysis should be interpreted as suggestive for two reasons. First, the spatial coverage of the underlying pollen data is uneven: dense in Scandinavia and northwestern Europe but substantially sparser in the Mediterranean and eastern regions (see Figure A7 for the spatial extent). Because pollen values are interpolated only when at least one pollen core is within 250 km in a given era, gridcell era observations without nearby pollen records are excluded from the regression. The resulting sample is substantially smaller than the baseline gridded gravity panel, so the estimated relationships primarily reflect areas with better data availability, limiting generalization to less-sampled regions. Second, migration flows are not randomly assigned across origin–destination pairs. Unobserved factors—such as pre-existing trade routes, cultural connections, or unmeasured environmental similarities—may jointly influence both migration patterns and vegetation outcomes. While the inclusion of fixed effects and controls mitigates these concerns, the results should be viewed as consistent with ecological diffusion rather than as definitive causal evidence.

With these caveats in mind, we present results in Table 6. Columns 1, 3, and 5 report the baseline specification for cereal, grassland, and tree pollen respectively; columns 2, 4, and 6 augment the baseline with lagged changes in vegetation at the destination and their interaction with migration, as a check against reverse causation or differential pre-trends. For all three outcomes, the coefficient on the interaction between lagged migration and the pollen gap is positive and statistically significant in the baseline specifications (columns 1, 3, and 5), indicating that destinations receiving more migrants from ecologically different origins subsequently shift toward those origins' vegetation profiles. Once we add controls for the destination's own prior vegetation change (columns 2, 4, and 6), the migration–pollen-gap interaction attenuates by roughly 40–60% but remains statistically significant—at the 1% level for cereals and at the 5% level for grassland and tree cover. The prior-vegetation-change interaction is negative for cereals and positive for grassland and tree cover, suggesting that some of the residual attenuation in the latter may reflect selection of migrants into already-converging destinations.

[20] We cluster standard errors at the gridcell pair level.

Note that this pattern operates over and above proximity-driven diffusion. The interaction between travel costs and the pollen gap is positive and statistically significant for grassland and tree cover in the baseline specifications (columns 3 and 5), capturing spatial spillovers in vegetation across nearby locations. The persistence of the migration effect after controlling for these channels suggests that migrants themselves contribute to ecological change, consistent with demic diffusion: the transmission of subsistence-specific knowledge and practices through population movement. The destination temperature change control is itself highly significant and exhibits the expected signs: warming is associated with increases in cereal pollen and tree cover, and declines in grassland, consistent with warmer conditions favoring both agricultural and forest expansion at the expense of open grassland (Marcott et al., 2013).

7 Conclusion

For most of human history, migration unfolded along corridors of ecological familiarity. The knowledge accumulated in one climate zone—when to plant, what to grow, how to manage soils, how to endure seasonal scarcity—constituted a form of human capital whose value was intrinsically local and depreciated with ecological distance. In this setting, the geography of populations was shaped by the distribution of climates and the portability of knowledge across them.

This paper develops and tests a framework of ecological sorting in which climate-specific human capital governs migration decisions. The framework yields four predictions linking migration flows to environmental similarity, subsistence technology, climate change, and ecological transformation. To evaluate these predictions, we construct a novel measure of bilateral migration flows using identity-by-descent segments from aDNA (Ringbauer et al., 2024), covering Western Eurasia over the past 10,000 years. We combine these flows with high-resolution data on climate, soils, and paleoclimate reconstructions. This allows us to examine how migration responds to spatial and temporal variation in ecological conditions and how migrants themselves reshape local environments.

We find that migration declines with ecological distance: populations disproportionately moved toward destinations with similar temperature regimes, precipitation patterns, and soil characteristics. Moreover, we document that the relevant dimension of similarity depends on subsistence technology. Farmers match strongly on thermal and soil conditions, while pastoralists match primarily on precipitation, consistent with the ecological constraints governing crop production and grazing. This heterogeneity is important not only for interpretation but also for identification: the fact that different groups respond to distinct and

orthogonal environmental dimensions assuages concerns that the results may be driven by a single omitted confounder.

We next show that migration responds to climate change. Overall, populations exposed to warming experience increased migration on average, while those exposed to cooling migrate less. These responses vary across subsistence systems: warming increases migration among farmers and foragers, as the cereal-viable zone expands and viable habitats broaden, whereas cooling increases migration among pastoralists, as ecological conditions shift in favor of grazing. The migration responses implied by our framework align with the archaeological record: mapping estimated climate responses into predicted flows over time, the model recovers the timing and direction of major prehistoric migration episodes, including the Neolithic expansion of farming and the later spread of steppe pastoralists. Migration also operates as a margin of climate adaptation that allows populations to track shifting ecological niches over time: populations exposed to environmental change relocate toward destinations that offset local shocks, partially preserving the match between their inherited knowledge and the environment. Finally, we provide evidence that migration reshaped the environments into which populations moved. Locations receiving migrants from ecologically different origins experienced systematic changes in vegetation, converging toward the ecological profiles of those origins.

Several limitations should be noted. First, 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. Second, our analysis relies on modern climate data to measure ecological similarity across locations. Appendix B.1 quantifies the bias from this choice: within-LGMR bilateral gradients evolved by only 0.53°C across the Holocene against an 8.67°C cross-sectional spread, implying a classical measurement error attenuation of approximately 0.4%. Long-run climatic shifts at individual locations are larger, but the bilateral identification used in our matching analysis cancels common-mode temporal variation. Third, paleoclimate reconstructions do not permit reliable measurement of growing degree days at fine temporal resolution—they provide annual means with model-derived seasonal cycles (Bova et al., 2021)—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. Tree-ring proxies, which preferentially record growing-season conditions, are a natural starting point for such products, though they currently lack the spatial coverage and within-season resolution required to reconstruct GDD directly. Fourth, the adaptation analysis 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 evidence on ecological diffusion should be interpreted with caution, as we lack exogenous variation in exposure to migration from origins with different land-use profiles.

Our findings open several directions for future research. One is to connect ecological sorting in prehistory to the long-run distribution of economic activity, cultural traits, and human capital. Another is to examine how technological change alters the portability of knowledge and the strength of ecological constraints. More broadly, understanding how populations historically adapted to environmental change through migration can inform how societies may respond 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.

Akbari, Ali, Annabel Perry, Alison R Barton et al. (2026) "Ancient DNA reveals pervasive directional selection across West Eurasia," *Nature*, 1–10.

Allentoft, Morten E., Matthew Collins, David Harker et al. (2012) "The Half-Life of DNA in Bone: Measuring Decay Kinetics in 158 Dated Fossils," *Proceedings of the Royal Society B: Biological Sciences*, 279 (1748), 4724–4733.

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.

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.

Bailey, Geoff and Penny Spikins (2008) *Mesolithic Europe*: Cambridge University Press.

Barjamovic, Gojko, Thomas Chaney, Kerem Coşar, and Ali Hortacsu (2019) "Trade, Merchants, and the Lost Cities of the Bronze Age," *Quarterly Journal of Economics*, 134 (3), 1455–1503.

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.

Beckmann, Michael, Tomáš Václavík, Anas M. Manceur, Lenka Šprátová, Holger von Trepte, Javier Alanis-Huerta, Emy Mattsson, and Dirk Zeuss (2014) "glUV: A Global UV-B Radiation Data Set for Macroecological Studies," *Methods in Ecology and Evolution*, 5, 372–383.

Boehm, Johannes and Thomas Chaney (2024) "Trade and the End of Antiquity," *Working Paper*.

Bogucki, Peter (1988) *Forest Farmers and Stockherders: Early Agriculture and Its Consequences in North-Central Europe*: Cambridge University Press.

Bova, Samantha, Yair Rosenthal, Zhengyu Liu, Shital P. Godad, and Mi Yan (2021) "Seasonal origin of the thermal maxima at the Holocene and the last interglacial," *Nature*, 589, 548–553, 10.1038/s41586-020-03155-x.

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.

Brink, Stefan and Neil Price (2008) *The Viking World*: Routledge.

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.

Chaplin, George (2004) "Geographic Distribution of Environmental Factors Influencing Human Skin Coloration," *American Journal of Physical Anthropology*, 125, 292–302.

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.

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

Davis, B A S, S Brewer, A C Stevenson, and J Guiot (2003) "The temperature of Europe during the Holocene reconstructed from pollen data," *Quaternary Science Review*, 22 (15-17), 1701–1716.

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

Dingel, Jonathan and Felix Tintelnot (2026) "Spatial Economics for Granular Settings," *Econometrica*, 94 (2), 407–464.

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

Erb, Michael P., Nicholas P. McKay, Nathan Steiger, Sylvia Dee, Cody Hancock, Ruza F. Ivanovic, Lauren J. Gregoire, and Paul Valdes (2022) "Reconstructing Holocene temperatures in time and space using paleoclimate data assimilation," *Climate of the Past*, 18 (12), 2599–2629, 10.5194/cp-18-2599-2022.

Essel, Hans, Jan Esper, Heinz Wanner, and Ulf Büntgen (2024) "Rethinking the Holocene temperature conundrum," *Climate Research*, 92, 61–64, 10.3354/cr01718.

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

Flückiger, Matthias, Erik Hornung, Mario Larch, Markus Ludwig, and Allard Mees (2022) "Roman Transport Network Connectivity and Economic Integration," *Review of Economic Studies*, 89 (2), 774–810.

Flückiger, Matthias, Mario Larch, Markus Ludwig, and Luigi Pascali (2024) "The Dawn of Civilization: Metal Trade and the Rise of Hierarchy," *Working Paper*.

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.

Giesecke, Thomas, Basil Davis, Simon Brewer et al. (2014) "Towards Mapping the Late Quaternary Vegetation Change of Europe," *Vegetation History and Archaeobotany*, 23 (1), 75–86, 10.1007/s00334-012-0390-y.

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.

Golden, Peter B. (1992) *An Introduction to the History of the Turkic Peoples*: Harrassowitz Verlag.

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.

d'Alpoim Guedes, Jade 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.

Halstead, Paul (2014) *Two Oxen Ahead: Pre-Mechanized Farming in the Mediterranean*: Wiley-Blackwell.

Hamerow, Helena (2002) *Early Medieval Settlements: The Archaeology of Rural Communities in North-West Europe 400–900*: Oxford University Press.

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.

Herzschuh, Ulrike, Thomas Böhmer, Chenzhi Li, Xianyong Cao, Raphaël Hébert, Anne Dallmeyer, Richard J. Telford, and Stefan Kruse (2023) "Regional pollen-based Holocene temperature and precipitation patterns depart from the Northern Hemisphere mean trends," *Climate of the Past*, 19 (7), 1481–1506, 10.5194/cp-19-1481-2023.

Hornbeck, Richard, Guy Michaels, and Ferdinand Rauch (2024) "Identifying Agglomeration Shadows: Long-run Evidence from Ancient Ports," *Working Paper*.

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

Jablonski, Nina G. and George Chaplin (2000) "The Evolution of Human Skin Coloration," *Journal of Human Evolution*, 39 (1), 57–106, 10.1006/jhev.2000.0403.

Kaufman, Darrell, Nicholas McKay, Cody Routson, Michael Erb, Christoph Dätwyler, Philipp S. Sommer, Oliver Heiri, and Basil Davis (2020) "Holocene global mean surface temperature, a multi-method reconstruction approach," *Scientific Data*, 7, 201, 10.1038/s41597-020-0530-7.

Khazanov, Anatoly M. (1994) *Nomads and the Outside World*: University of Wisconsin Press, 2nd edition.

Kistler, Logan, Roselyn Ware, Oliver Smith, Matthew Collins, and Robin G. Allaby (2017) "A New Model for Ancient DNA Decay Based on Paleogenomic Meta-Analysis," *Nucleic Acids Research*, 45 (11), 6310–6320.

Kristiansen, Kristian (1998) *Europe Before History*: Cambridge University Press.

Kuz'mina, Elena E. (2007) *The Origin of the Indo-Iranians*: Brill.

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.

Le Rossignol, Etienne and Sara Lowes (2026) "Mobility, Competition, and the Scope of Trust: Evidence from Transhumant Pastoralism," Working paper.

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

Librado, Pablo, Naveed Khan, Antoine Fages et al. (2021) "The Origins and Spread of Domestic Horses from the Western Eurasian Steppes," *Nature*, 598, 634–640.

Lindahl, Tomas (1993) "Instability and Decay of the Primary Structure of DNA," *Nature*, 362 (6422), 709–715.

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.

Mathieson, Iain, Iosif Lazaridis, Nadin Rohland et al. (2015) "Genome-Wide Patterns of Selection in 230 Ancient Eurasians," *Nature*, 528, 499–503.

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

Mauri, A, B A S Davis, P M Collins, and J O Kaplan (2015) "The climate of Europe during the Holocene: a gridded pollen-based reconstruction and its multi-proxy evaluation," *Quaternary Science Review*, 112, 109–127.

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) "Ecological Imperialism: Climate, European Overseas Settlement, and Long-Run Development," 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.

Olsson, Ola (2024) *Paleoeconomics*: Springer.

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.

Özak, Ömer (2018) "Distance to the pre-industrial technological frontier and economic development," *Journal of Economic Growth*, 23 (2), 175–221.

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.

Rolle, Renate (1989) *The World of the Scythians*: Batsford.

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.

Shennan, Stephen (2018) *The First Farmers of Europe: An Evolutionary Perspective*: Cambridge University Press.

Sherratt, Andrew (1981) "Plough and Pastoralism: Aspects of the Secondary Products Revolution," in Hodder, Ian, Glynn Isaac, and Norman Hammond eds. *Pattern of the Past: Studies in Honour of David Clarke*, 261–305: Cambridge University Press.

Sjögren, Karl-Göran, T. Douglas Price, and Kristian Kristiansen (2016) "Diet and Mobility in the Corded Ware of Central Europe," *PLOS ONE*, 11 (5), e0155083.

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, 10.5194/cp-13-1851-2017.

Spolaore, Enrico and Romain Wacziarg (2009) "The Diffusion of Development," *Quarterly Journal of Economics*, 124 (2), 469–529.

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.

Tierney, Jessica E., Jiang Zhu, Jonathan King, Steven B. Malevich, Gregory J. Hakim, and Christopher J. Poulsen (2020) "Glacial cooling and climate sensitivity revisited," *Nature*, 584 (7822), 569–573, 10.1038/s41586-020-2617-x, Provides the iCESM-based LGM data assimilation methodology subsequently extended through the Holocene by Osman et al. (2021).

Vander Linden, Marc (2024) *The Bell Beaker Phenomenon in Europe: A Harmony of Difference*: Cambridge University Press.

Waldinger, Maria (2022) "The Economic Effects of Long-Term Climate Change: Evidence from the Little Ice Age," *Journal of Political Economy*, 130 (9), 2275–2314.

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

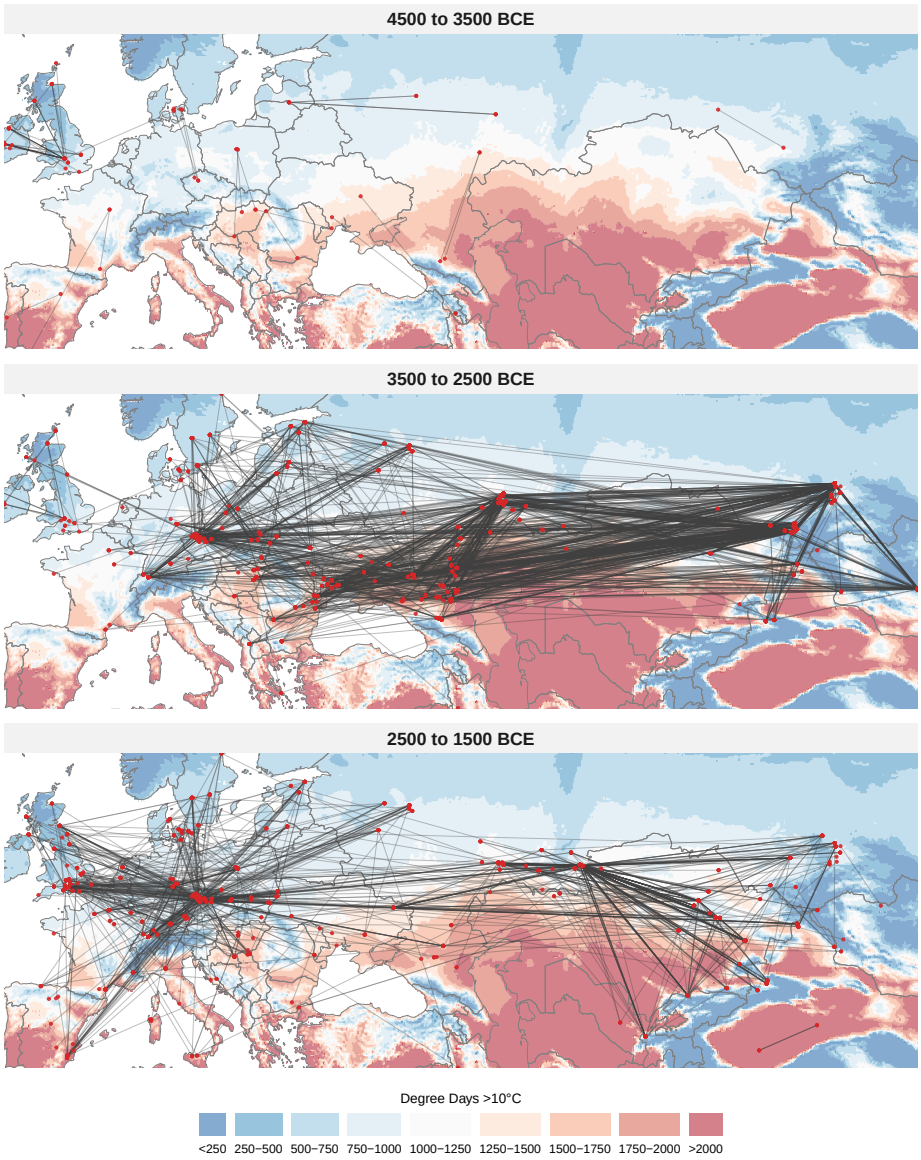
White, K. D. (1970) *Roman Farming*: Thames and Hudson.

Wiens, John J. and Catherine H. Graham (2005) "Niche conservatism: Integrating evolution, ecology, and conservation biology," *Annual Review of Ecology, Evolution, and Systematics*, 36, 519–539, 10.1146/annurev.ecolsys.36.102803.095431.

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

Figures

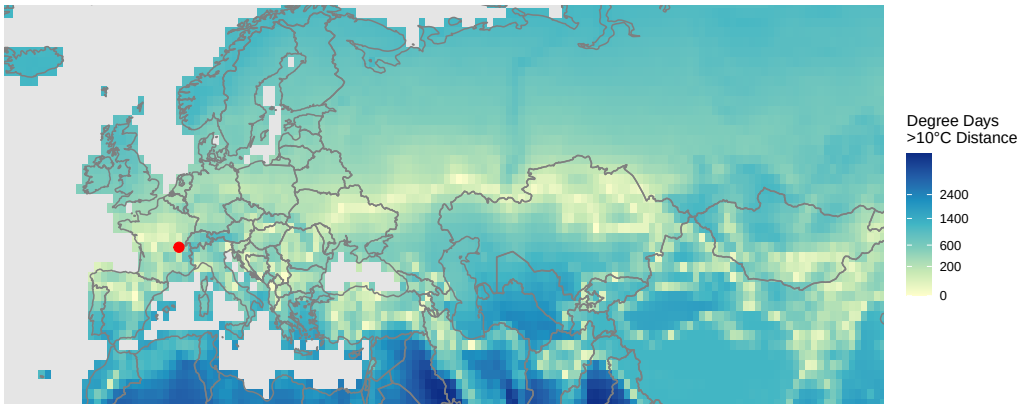
Figure 1: Flows of Individuals



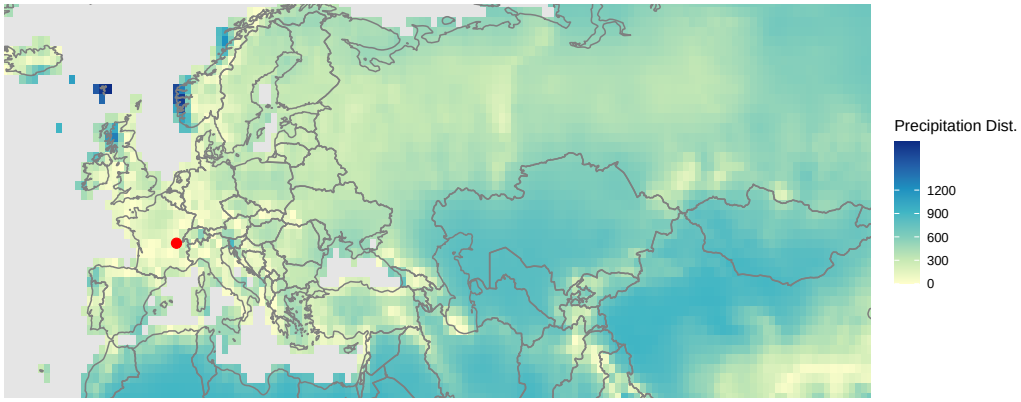
Notes: Each line connects a pair of individuals sharing at least one identity-by-descent (IBD) segment greater than 16 cM, indicating a common ancestor within approximately 6–10 generations. We use a 16 cM threshold here for visual clarity; the rest of the paper uses an 8 cM baseline (the minimum threshold at which Ringbauer et al. 2024's method reliably detects IBD), and results are robust to either threshold. The sample is restricted to pairs living at most 500 years apart and separated by at least 500 km. Panels show three consecutive 1,000-year periods covering the late Neolithic through the Bronze Age: 4500–3500 BCE (late LBK and Funnel Beaker farming horizons; pre-Yamnaya steppe), 3500–2500 BCE (Yamnaya steppe expansion, early Corded Ware, Afanasievo migration to Inner Asia), and 2500–1500 BCE (Bell Beaker climax, Sintashta, Andronovo). Red dots mark origin and destination burial locations; grey lines connect pairs. The background color denotes modern degree days above 10°C, binned in 250-degree-day intervals.

Figure 2: Ecological Distance from a Reference Location

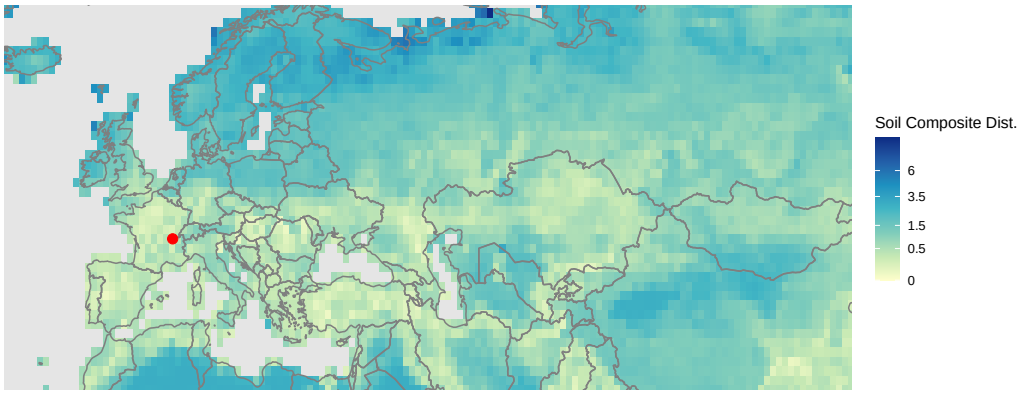
(a) Growing Degree Day Distance



(b) Precipitation Distance

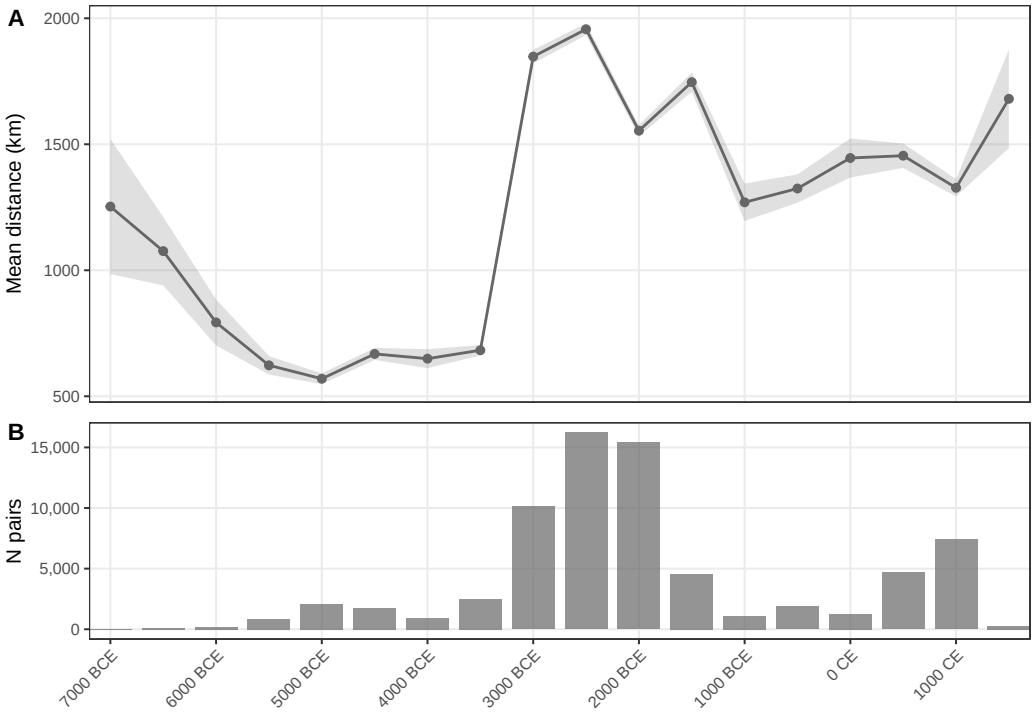


(c) Soil Composite Distance



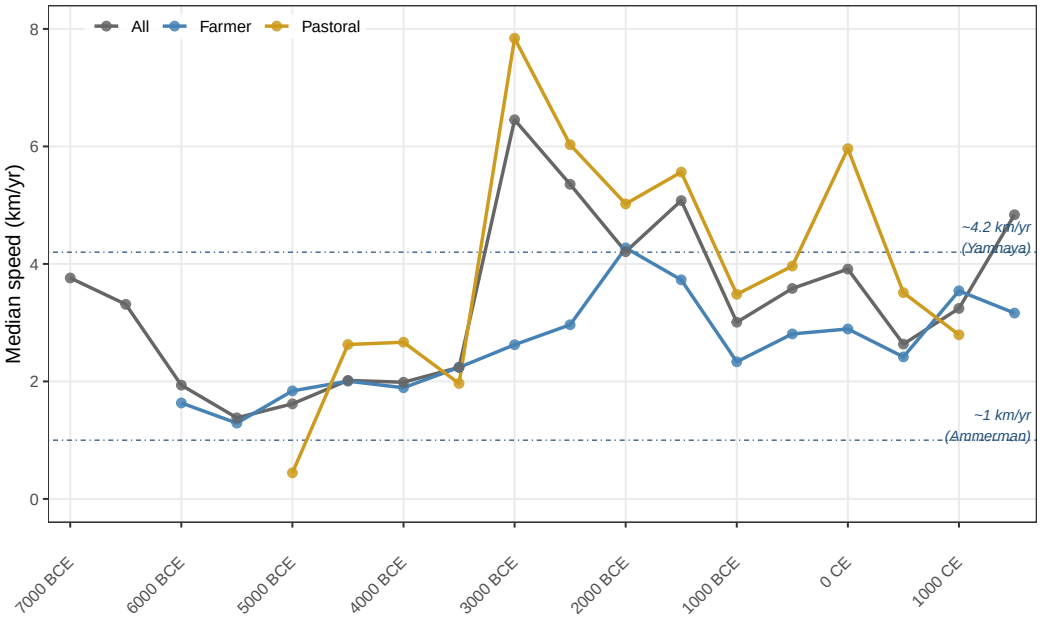
Notes: Spatial distribution of the three ecological distance measures from a reference location in central France (Lyon; 46°N, 5°E, marked in red). Panel A: absolute difference in cumulative growing degree days above 10°C. Panel B: absolute difference in mean annual precipitation. Panel C: Euclidean distance on clay content and soil organic carbon. The distinct spatial patterns demonstrate that the three dimensions capture largely orthogonal gradients: temperature varies primarily with latitude, precipitation with coast-to-interior position, and soil characteristics with underlying geology.

Figure 3: Migration Distance Over Time



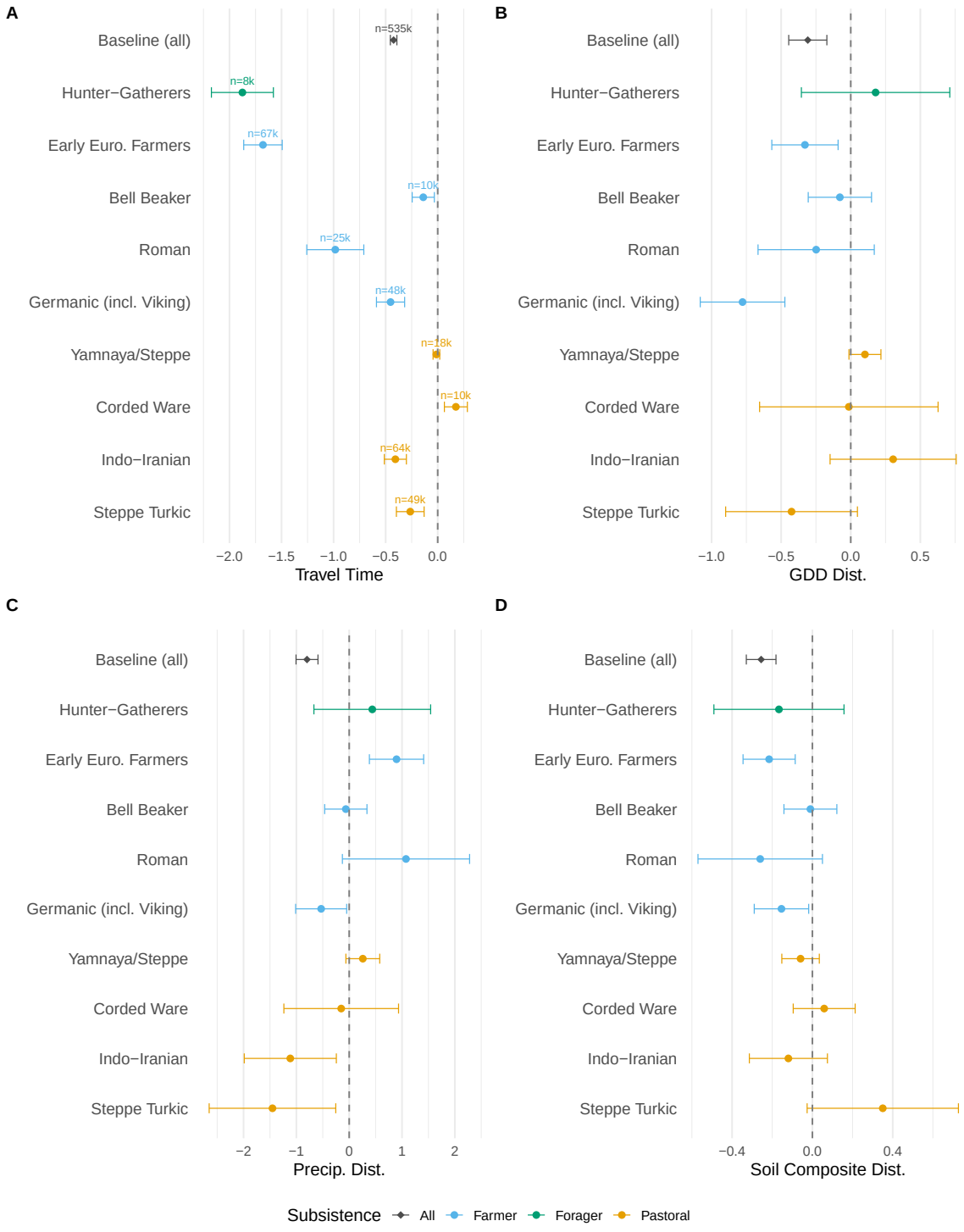
Notes: Both panels use the same sample of IBD-linked individual pairs (segment $\geq$ 8 cM, separated by 100–500 years) aggregated to 500-year eras. Panel A plots the mean geographic distance between genetically related individuals with 95% confidence bands. Panel B shows the number of pairs in each era; the rise around 3000 BCE indicates increased sample density rather than driving the distance trajectory in Panel A.

Figure 4: Implied Migration Speed by Era and Subsistence



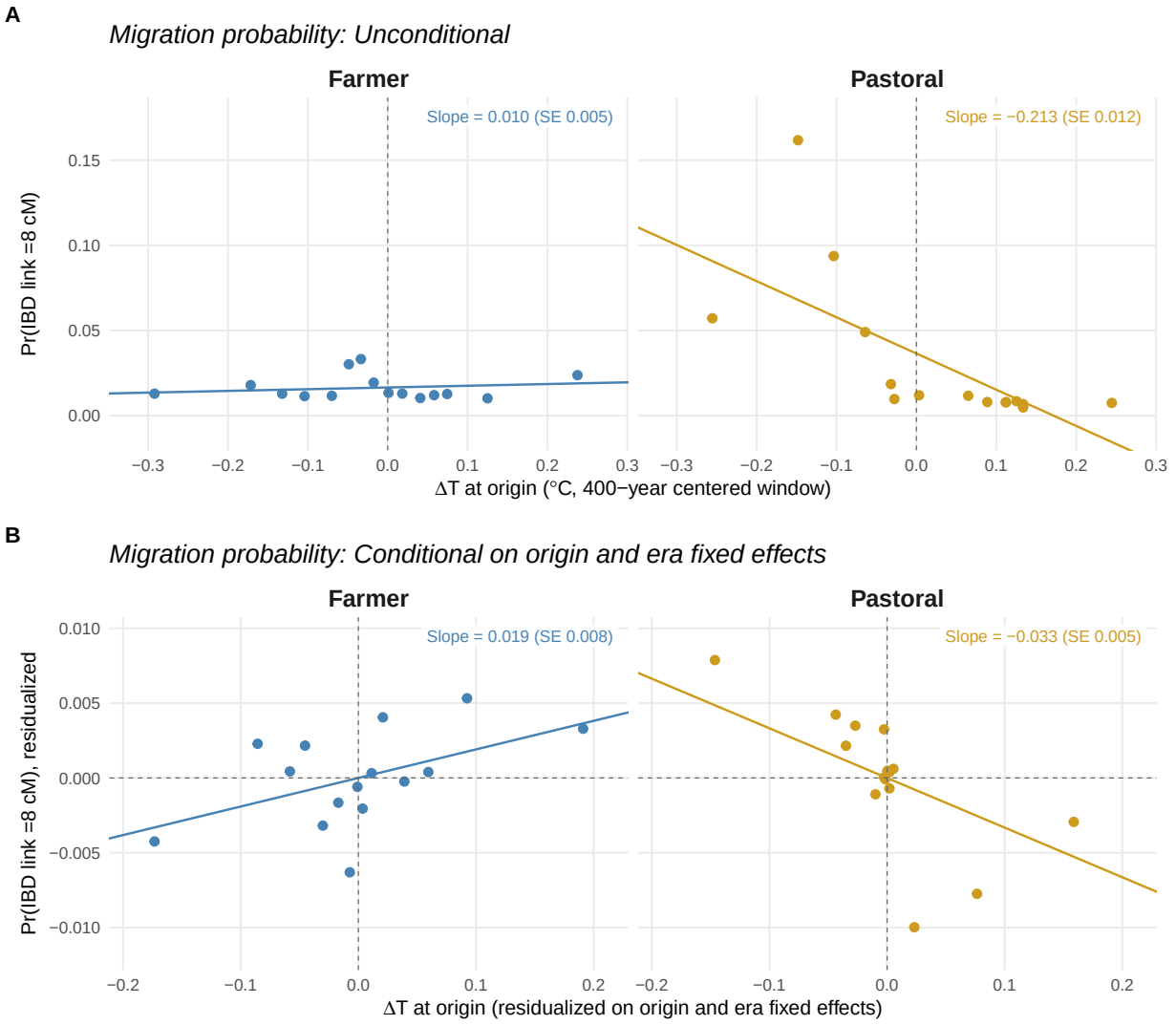
Notes: Median implied displacement rate (km/year) by 500-year era and origin subsistence type (Farmer, Pastoral, and the pooled All series), computed from the same IBD-linked individual pairs used in Figure 3. We use the median rather than the mean because the per-pair speed distribution has a heavy right tail. Horizontal reference lines mark canonical estimates from the literature: 1 km/yr for the Neolithic wave of advance (Ammerman and Cavalli-Sforza, 1971) and 4.2 km/yr for the Yamnaya expansion (Racimo et al., 2020). The wave-of-advance benchmarks measure 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.

Figure 5: Climate Matching and Distance Sensitivity by Culture Group



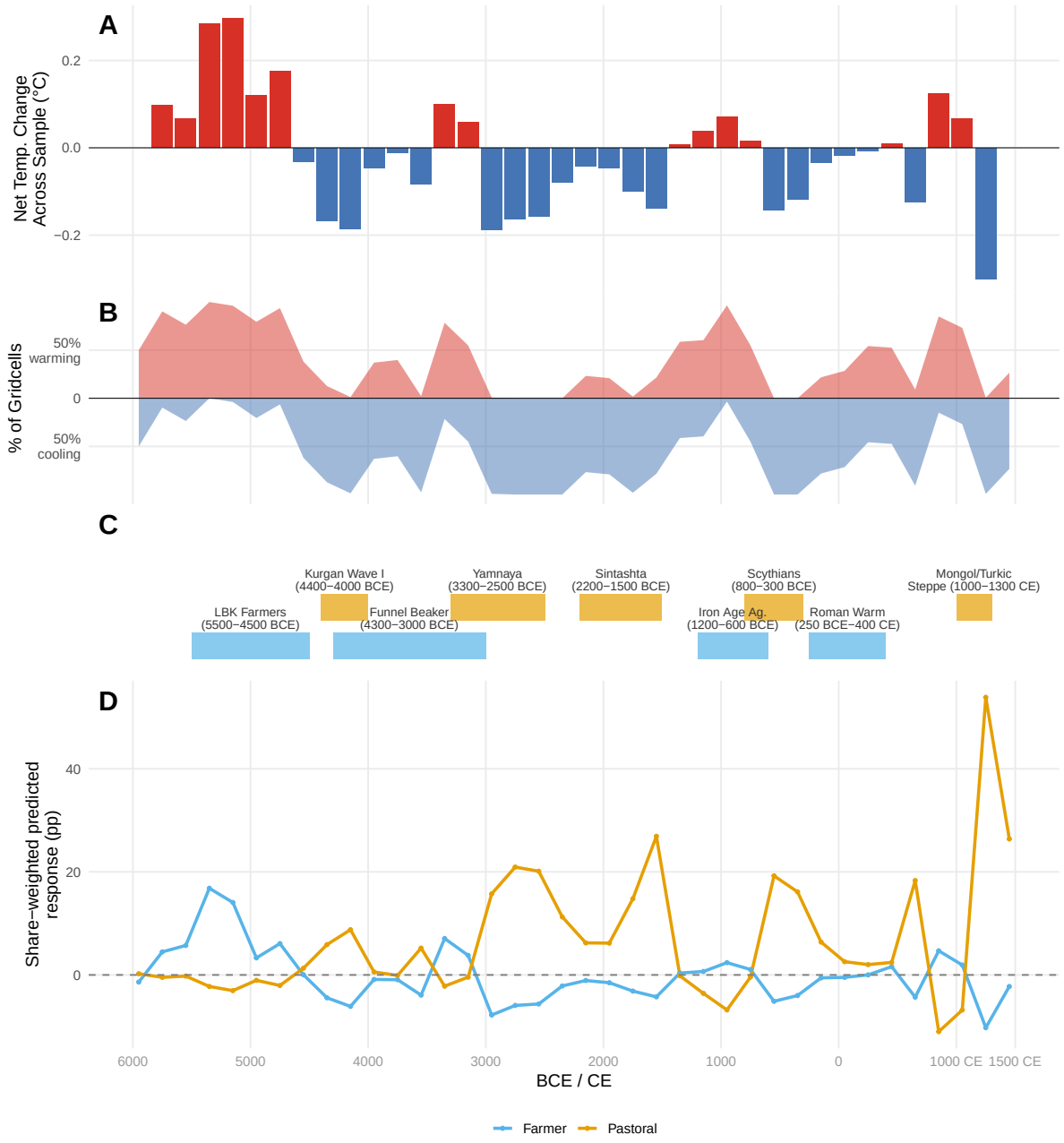
Notes: Coefficient plots from the baseline gravity specification (Equation 6) estimated separately for each of 11 named archaeological culture groups, colored by subsistence type: farmer (blue), pastoral (gold), forager (green). Panel A: geographic distance coefficient. Panel B: growing degree day (GDD) distance coefficient. Panel C: precipitation distance coefficient. Panel D: soil composite distance coefficient. Dashed line marks zero. Baseline (all) shown in grey.

Figure 6: Subsistence-Specific Migration Response to Origin Climate Change



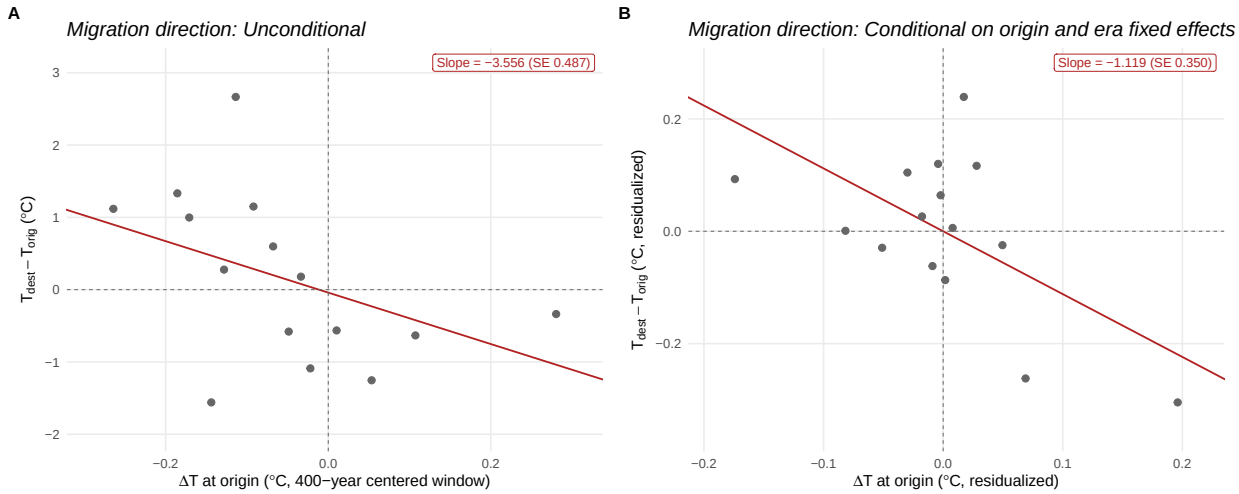
Notes: Individual-level data of $N = 2{,}867{,}964$ ancestor-descendant pairs (one observation per pair, including IBD-zero pairs), restricted to non-forager origins: approximately 60% farmer-origin and 40% pastoral-origin. Binscatters plot the probability that a pair shares any IBD segment of at least 8 cM (i.e., $\mathbf{1}\{count_{8cM} > 0\}$, the same outcome used in the individual gravity regressions in Table B7) against the 400-year temperature change at the origin (positive values indicate warming). Sample restricted to pairs separated by more than 50 km, with 100–500 year burial gaps, and modal origin gridcell subsistence of farmer or pastoral. Panel A reports the unconditional binscatter; Panel B residualizes both axes on origin gridcell and era fixed effects within subsistence (Frisch-Waugh-Lovell). Slope labels report linear fits to the underlying micro-data with standard errors clustered at the paleoclimate cell pair level.

Figure 7: 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: pastoral (gold) and farmer (blue).[21] Panel D plots the share-weighted predicted migration response (in percentage points) for farmer and pastoral origins. For each 200-year era and origin gridcell, the response is computed as $\text{share}_s \times (e^{\beta_s \cdot \text{tc400}} - 1) \times 100$, where share_s is the share of subsistence type s at the origin, tc400 is the 400-year centered temperature change, and β_s is the subsistence-specific coefficient from Table 4 (farmer: 0.89; pastoral: −2.34). Panel D is a mechanical projection of model coefficients onto observed climate shocks at sample origins; it represents the model's predicted relative response, not a direct estimate of historical migration flows.

Figure 8: Adaptive Migration: Origin Climate Change vs. Spatial Direction of Movement



Notes: Individual-level data of $N = 71{,}290$ ancestor-descendant pairs (one observation per IBD-positive pair sharing $\geq$8 cM of segments, separated by more than 50 km, with 100–500 year burial gaps). Both panels plot the relationship between climate change at the origin (x-axis) and the spatial direction of migration in temperature space (y-axis). The x-axis measures the 400-year centered temperature change at the origin location using Osman et al. (2021) reconstructions ($T_{t-200} - T_{t+200}$, where t is the ancestor's date in years BCE (negative for CE); positive values indicate warming). The y-axis measures the modern temperature difference between destination and origin ($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. Panel A reports the unconditional binscatter; Panel B residualizes both axes on origin gridcell and era fixed effects (Frisch-Waugh-Lovell), showing the within-origin-era association. Each point represents the unweighted mean within one of 15 equal-N (ntile) bins of the x-axis variable. The red line shows an OLS fit to the underlying microdata, reporting the bivariate slope with standard errors clustered at the paleoclimate cell pair level. Table B11 reports the regression estimate after also adding destination fixed effects and a saturated set of controls. The negative slope in both panels indicates that populations exposed to warming systematically relocated toward cooler destinations, partially offsetting local temperature shifts.

Tables

Table 1: Summary Statistics of Origin-Destination Gridcells (N = 531,137)

	Mean	SD	Min	P10	Median	P90	Max
Panel A: Migration Measures							
Any IBD ($\geq$ 8 cM)	0.04	0.20	0.00	0.00	0.00	0.00	1.00
IBD Count (8cM)	0.13	1.86	0.00	0.00	0.00	0.00	443.00
IBD Count (12cM)	0.04	0.71	0.00	0.00	0.00	0.00	121.00
IBD Count (16cM)	0.01	0.27	0.00	0.00	0.00	0.00	62.00
Panel B: Bilateral Distances							
GDD Dist. (1000 degree-days)	0.79	0.71	0.00	0.10	0.61	1.70	5.91
Temp. Mean Dist. (°C)	6.06	4.80	0.00	0.81	5.05	12.82	39.57
Precip. Dist. (m/yr)	0.31	0.30	0.00	0.04	0.23	0.67	2.64
Soil Composite Dist.	1.08	0.69	0.00	0.29	0.94	2.09	4.35
Distance (1000 km)	3.17	2.13	0.04	0.80	2.61	6.30	12.81
Travel Time (HMI, 1000 hrs)	2.38	1.61	0.03	0.53	2.04	4.73	10.41
Panel C: Origin Subsistence Shares							
Farmer Share (origin)	0.47	0.31	0.00	0.03	0.60	0.82	1.00
Pastoral Share (origin)	0.40	0.29	0.00	0.05	0.30	0.88	1.00
Forager Share (origin)	0.12	0.20	0.00	0.00	0.10	0.20	1.00

Notes: Unit of observation is a $1° \times 1°$ gridcell pair-era, restricted to the sample in columns 1–2 of Table 2 (pairs with GDD distance available). The sample covers Western Eurasia and spans 764 unique origin gridcells, 768 destination gridcells, and 19 500-year eras. Any IBD is a binary indicator equal to 1 if the OD-era cell has at least one pair sharing an IBD segment $\geq$ 8 cM; its mean reports the share of OD-era cells with positive flow. IBD Count is the number of individual pairs sharing IBD segments at the indicated cM threshold; 95.8% of observations have zero IBD count (8cM). GDD Dist. is cumulative growing degree days above 10°C (ERA5). Temp. Mean Dist. is the absolute difference in annual mean temperature (ERA5, °C). Precip. Dist. is the absolute difference in annual precipitation (TerraClimate, m/yr). Soil Composite Dist. is Euclidean distance on clay and organic carbon (SoilGrids). Travel Time (HMI) is least-cost walking time between gridcells in thousands of hours, from Özak (2018). Subsistence shares (farmer/pastoralist/forager, summing to one) at the origin gridcell are based on the archaeological culture group classification described in the data section.

Table 2: Environmental Distance and Prehistoric Migration

Dependent Variable:	Genetic Flow					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
GDD Dist.	-1.10***	-1.06***	-0.318***	-0.309***	-0.271***	-0.272***
	(0.052)	(0.052)	(0.071)	(0.070)	(0.062)	(0.047)
Precip. Dist.	-0.898***	-0.798***	-0.803***	-0.799***	-0.844***	-0.946***
	(0.126)	(0.127)	(0.112)	(0.106)	(0.101)	(0.076)
Soil Composite Dist.	-0.145***	-0.180***	-0.246***	-0.255***	-0.273***	-0.210***
	(0.029)	(0.030)	(0.039)	(0.038)	(0.034)	(0.029)
Travel Time	-0.431***	-0.404***	-0.410***	-0.424***	-0.443***	-0.378***
	(0.023)	(0.023)	(0.016)	(0.016)	(0.015)	(0.013)
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	531,137	531,137	478,034	478,034	415,014	228,296
Pseudo R^2	0.10	0.17	0.44	0.48	0.57	0.69
Dep. var. mean		0.134	0.149	0.149	0.172	0.312

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments ≥ 8 cM between individuals in an origin and destination 1° gridcell in a 500-year era. GDD Dist. is the absolute difference in cumulative growing degree days (base 10°C) between origin and destination, scaled by 1,000. Precip. Dist. is the absolute difference in annual precipitation. 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. Travel Time is HMI least-cost-path hours (thousands) between origin and destination. Column 4 is the preferred specification; columns progressively add fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Climate Matching by Subsistence Type: Split Sample

Dependent Variable:	Genetic Flow					
	Baseline	All (Std.)	Farmer	Agro-Past.	Pastoral	Forager
Model:	(1)	(2)	(3)	(4)	(5)	(6)
GDD Dist.	-0.309***	-0.218***	-0.437***	-0.462***	-0.032	-0.345**
	(0.070)	(0.049)	(0.059)	(0.081)	(0.058)	(0.162)
Precip. Dist.	-0.799***	-0.239***	0.034	0.083	-0.371***	-0.134
	(0.106)	(0.032)	(0.047)	(0.062)	(0.050)	(0.204)
Soil Composite Dist.	-0.255***	-0.177***	-0.204***	-0.097**	-0.024	-0.110
	(0.038)	(0.026)	(0.036)	(0.041)	(0.036)	(0.087)
Travel Time	-0.424***	-0.682***	-1.63***	-0.750***	-0.355***	-3.01***
	(0.016)	(0.026)	(0.076)	(0.060)	(0.026)	(0.279)
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	478,034	478,034	228,730	30,196	137,622	10,801
Pseudo R^2	0.48	0.48	0.50	0.63	0.57	0.49
Regressors standardized	No	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	0.149	0.149	0.131	0.226	0.243	0.084
Mean distance (km)	1,717	1,717	1,335	1,666	2,139	1,135

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments ≥ 8 cM between 1° gridcell pairs in a 500-year era. Column 1 replicates the preferred specification from Table 2 in unstandardized units; columns 2–6 restrict the sample to origins classified into the indicated subsistence type by archaeological culture group (agro-pastoral origins classified as mixed farmer-pastoralist) and standardize all regressors within each subsample (mean zero, unit SD) for cross-variable comparability. Mean distance is computed conditional on positive genetic flow. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Temperature Change and Migration

Dependent Variable:	Genetic Flow				
	All	Farmer	Agro-Past.	Pastoral	Forager
Model:	(1)	(2)	(3)	(4)	(5)
Temp. Change (origin)	0.937***	0.888***	-0.551	-2.34***	1.82***
	(0.191)	(0.310)	(0.706)	(0.480)	(0.509)
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	478,034	228,730	30,196	137,622	10,801
Pseudo R^2	0.48	0.50	0.63	0.57	0.50
Dep. var. mean	0.149	0.131	0.226	0.243	0.084

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments ≥ 8 cM between 1° gridcell pairs in a 500-year era. Temp. Change (origin) is the Osman et al. (2021) reconstructed temperature change over a centered 400-year window (positive = warming). Each column restricts the sample to origin gridcell-eras classified as the indicated subsistence type (archaeological culture group classification). All specifications include the full set of bilateral gravity controls from Table 2 (GDD distance, precipitation distance, soil composite distance, and HMI travel time) and origin gridcell, destination gridcell, and era fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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

Dependent Variable:	Genetic Flow				
Model:	(1)	(2)	(3)	(4)	(5)
Temp. Change (origin)	0.969***	0.998***	1.00***	1.01***	0.979***
	(0.194)	(0.194)	(0.194)	(0.195)	(0.237)
Temp. Change × Temp. Diff.	-0.065**	-0.068**	-0.067**	-0.070***	-0.104***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
GDD Dist.		-0.528***	-0.487***	-0.310***	-0.251***
		(0.055)	(0.056)	(0.070)	(0.087)
Precip. Dist.			-0.818***	-0.800***	-0.673***
			(0.109)	(0.106)	(0.136)
Soil Composite Dist.				-0.257***	-0.320***
				(0.038)	(0.043)
Travel Time	-0.502***	-0.473***	-0.436***	-0.426***	-0.420***
	(0.016)	(0.015)	(0.016)	(0.016)	(0.021)
Gridcell resolution	1°	1°	1°	1°	2.5°
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	478,034	478,034	478,034	478,034	495,898
Pseudo R^2	0.48	0.48	0.48	0.48	0.39
Dep. var. mean	0.149	0.149	0.149	0.149	0.144

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments ≥ 8 cM between 1° gridcell pairs in a 500-year era. Temp. Change (origin) is the Osman et al. (2021) reconstructed temperature change over a centered 400-year window; Temp. Diff. is the modern ERA5 destination-minus-origin temperature difference. A negative interaction indicates warming at origin shifts flows toward cooler destinations. Columns 1–5 use 1° origin and destination gridcells, and era fixed effects; columns 2–4 sequentially introduce GDD distance, precipitation distance, and soil composite distance as controls. Column 5 repeats the saturated specification with fixed effects defined at the 2.5° paleoclimate cell (the native resolution of the Osman et al. (2021) reconstruction, exact dimensions 1.9° latitude by 2.5° longitude). Standard errors, reported in parentheses, are clustered at the gridcell pair level (columns 1–4) or 2.5° cell-pair level (column 5). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

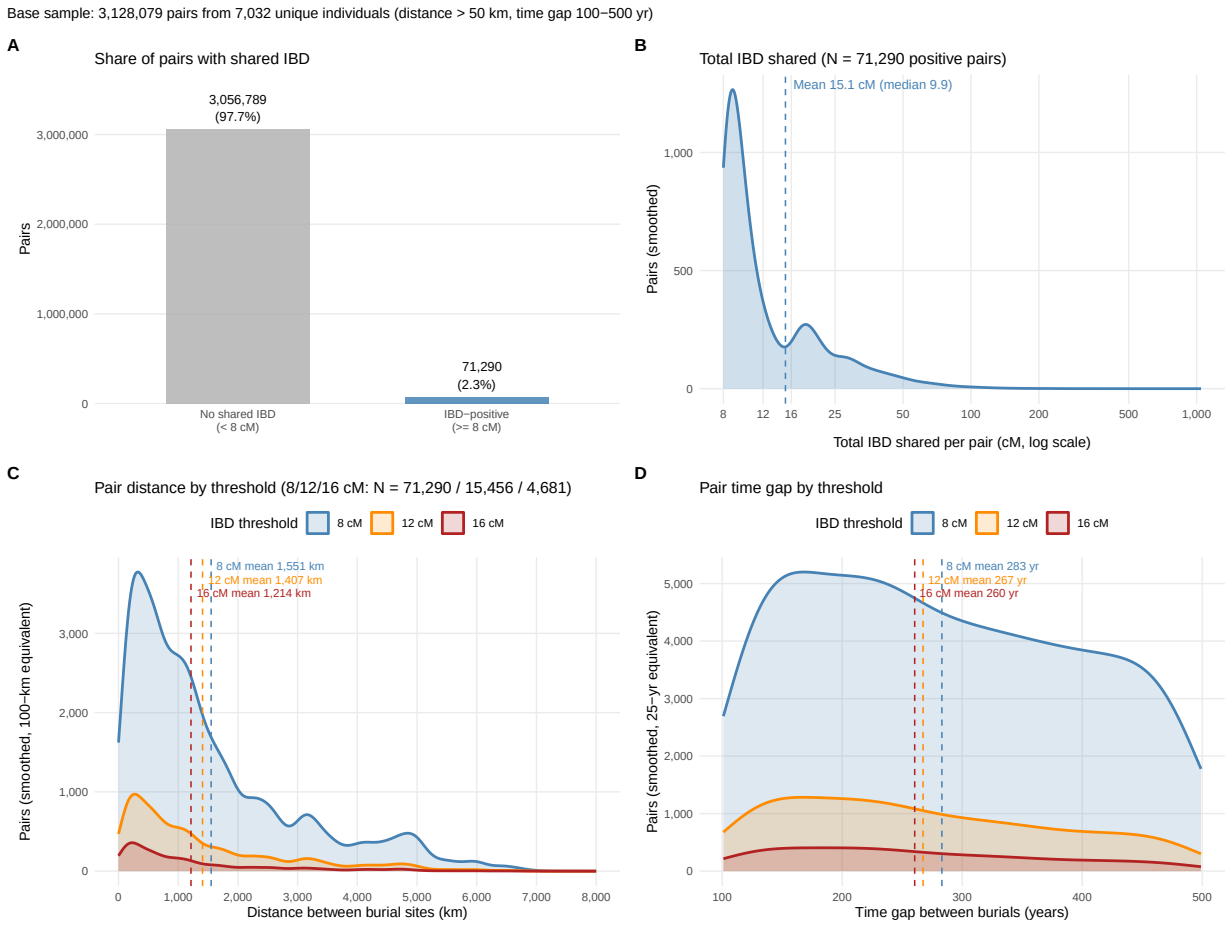
Table 6: Ecological Diffusion

	Δ Cereal Pollen		Δ Grassland Pollen		Δ Tree Pollen	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Pollen Dist. (o–d, t)	19.8***	21.3***	9.32***	11.1***	10.0***	11.3***
	(0.372)	(0.484)	(0.171)	(0.227)	(0.182)	(0.236)
Genetic Flow (o–d, t-500 to t)	-0.198***	-0.273***	0.937***	0.541**	-0.648**	-0.298
	(0.024)	(0.052)	(0.301)	(0.223)	(0.317)	(0.211)
Genetic Flow × Pollen Dist.	0.460***	0.264***	0.069***	0.029**	0.068***	0.029**
	(0.092)	(0.088)	(0.019)	(0.013)	(0.019)	(0.013)
Travel Time × Pollen Dist.	0.152	0.160	0.109***	-0.011	0.091***	-0.025
	(0.163)	(0.156)	(0.030)	(0.039)	(0.031)	(0.041)
Bilateral gravity controls	Yes	Yes	Yes	Yes	Yes	Yes
Reverse-causality controls		Yes		Yes		Yes
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	209,254	105,394	209,254	105,394	209,254	105,394
R^2	0.615	0.618	0.501	0.538	0.560	0.584

Notes: OLS estimates. The dependent variable in each pair of columns is the 500-year change in destination land cover (cereal, grassland, or tree pollen share). The interaction Genetic Flow × Pollen Dist. is the coefficient of interest: it tests whether migration from ecologically different origins predicts subsequent convergence toward the origin's pollen profile. All specifications include a Travel Time × Pollen Distance interaction, a control for temperature change at the destination, the full set of bilateral gravity controls from Table 2 (GDD distance, precipitation distance, soil composite distance, HMI travel time, and squared HMI travel time), and origin, destination, and era fixed effects. 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. The estimated interaction is significantly negative for cereals (inconsistent with reverse causality) and positive for grassland and tree cover (consistent with migrants selecting into ecologically converging destinations). Standard errors, reported in parentheses, are clustered at the gridcell pair level.

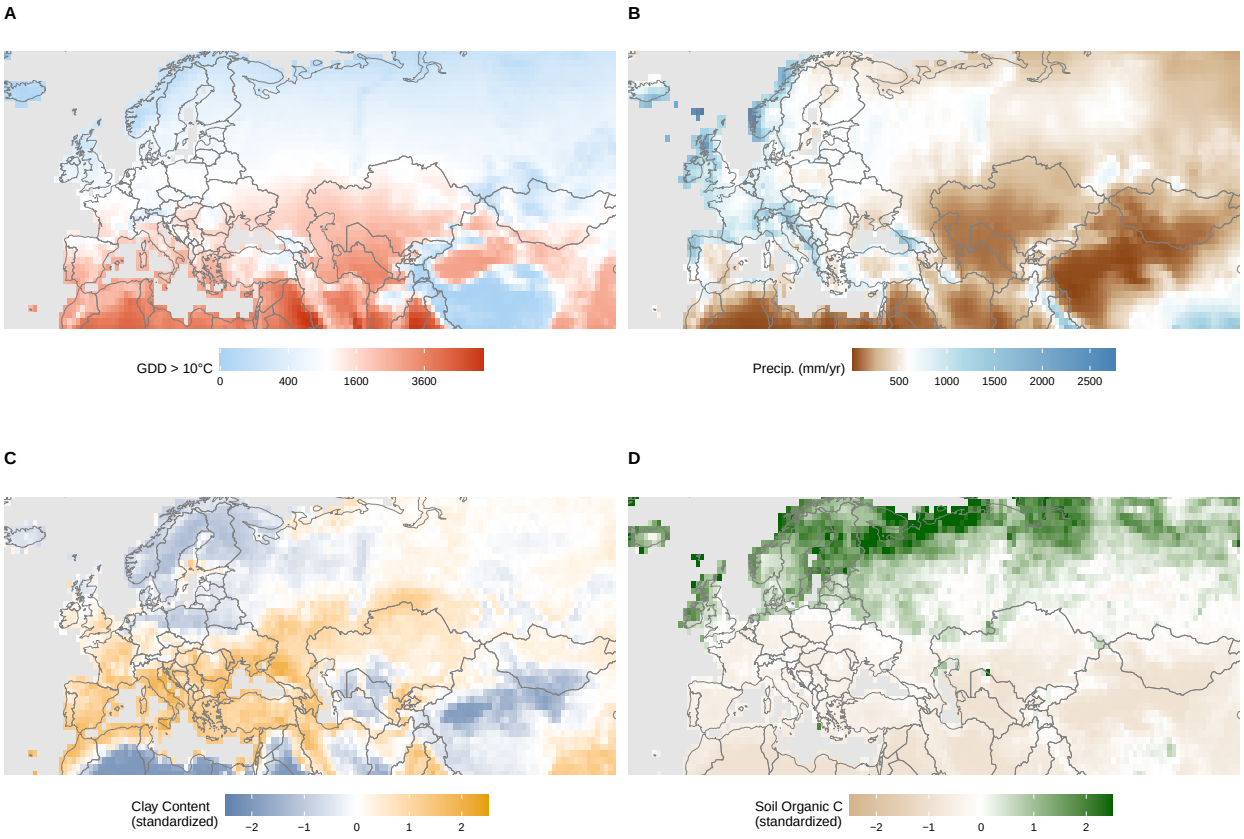
A Additional Tables and Figures

Figure A1: Geographic and Genetic Scale of Ancestor-Descendant Pairs



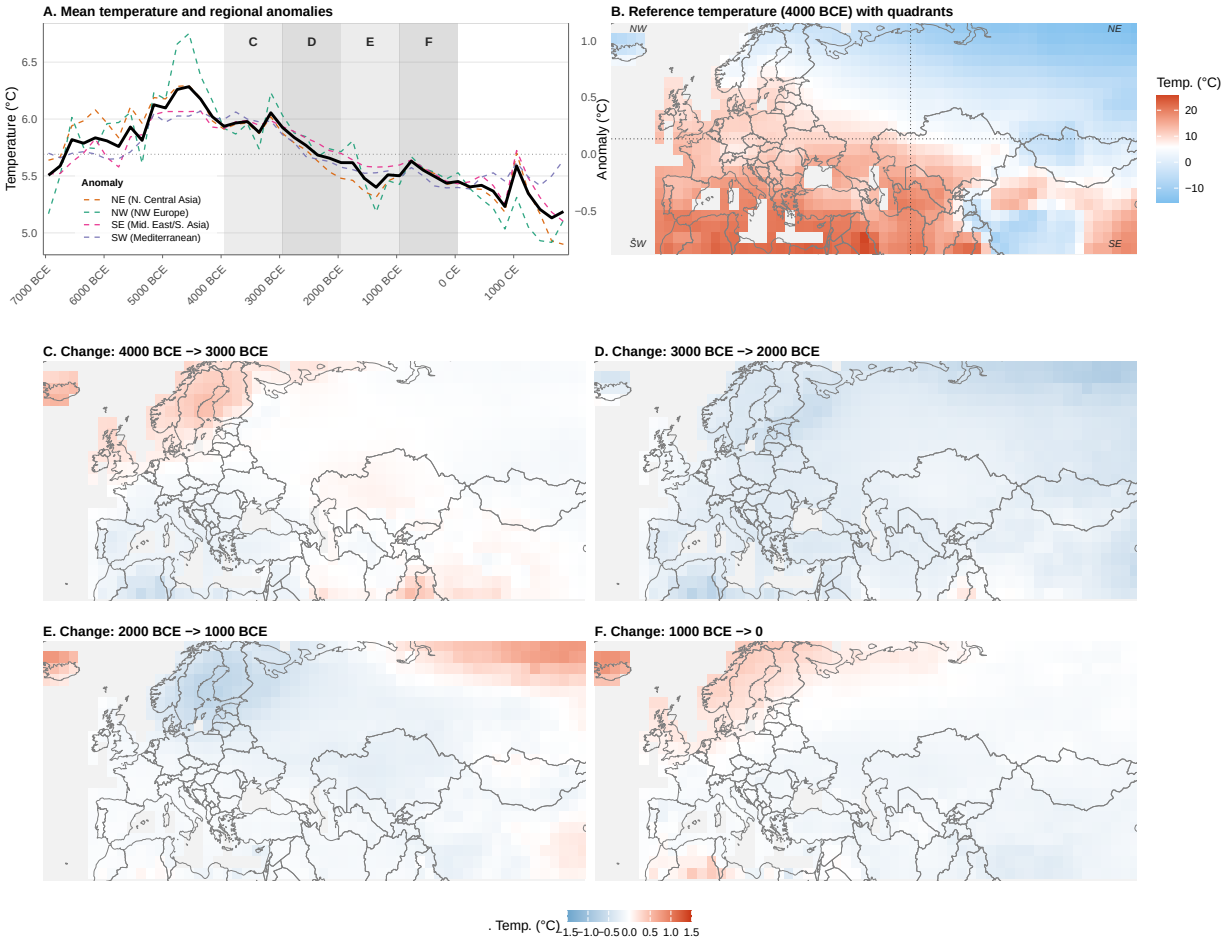
Notes: Sample comprises 3,128,079 candidate pairs constructed from 7,032 unique aDNA individuals (Ringbauer et al., 2024), restricted to pairs separated by more than 50 km and a time gap of 100–500 years. The regression sample (Table B7) is smaller (3,050,584 pairs) due to gridcells with missing covariate data. Panel A: share of base pairs sharing at least one IBD segment $\geq$ 8 cM (2.3%, N = 71,290) versus pairs with no detected IBD (97.7%). Panel B: distribution of total IBD shared (cM, log scale) among the 71,290 IBD-positive pairs; vertical lines mark the 8 cM detection floor, the median (9.9 cM), and the mean (15.1 cM). Panel C: kernel density of geographic distance between paired burial locations at IBD thresholds of 8, 12, and 16 cM (means: 1,551 / 1,407 / 1,214 km). Panel D: kernel density of the ancestor-descendant time gap at the same thresholds (means: 283 / 267 / 260 yr).

Figure A2: Modern Climate and Soil Variables



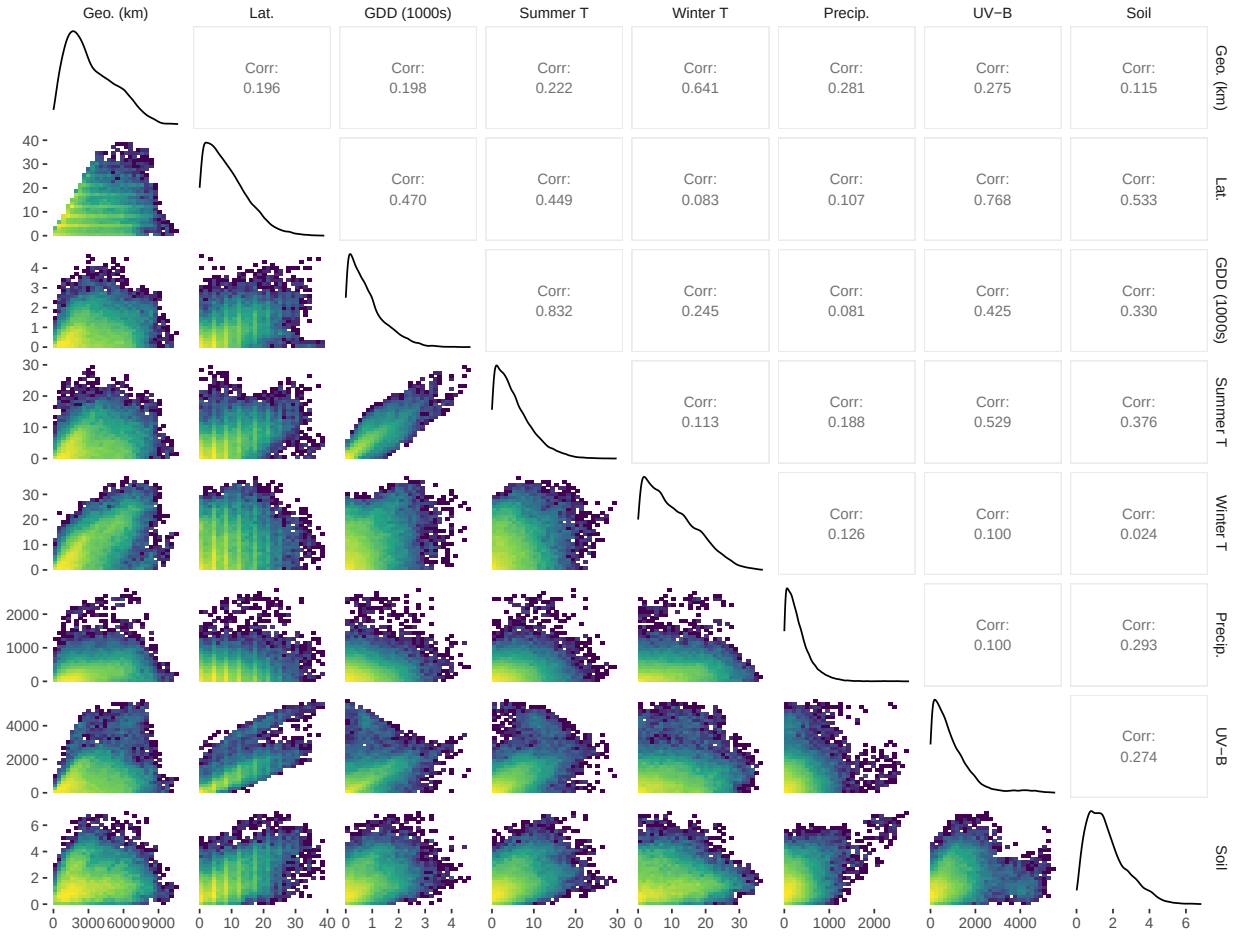
Notes: Spatial distribution of the four ecological variables used in the baseline specification, at 1° resolution. Panel A: cumulative growing degree days above 10°C (ERA5 reanalysis). Panel B: mean annual precipitation (TerraClimate, mm/yr). Panel C: clay content (SoilGrids, standardized). Panel D: soil organic carbon (SoilGrids, standardized). GDD and precipitation capture the main climatic gradients; clay and organic carbon provide a largely orthogonal, geologically determined dimension of environmental similarity.

Figure A3: Paleoclimate Variation: Temperature over Time and Space



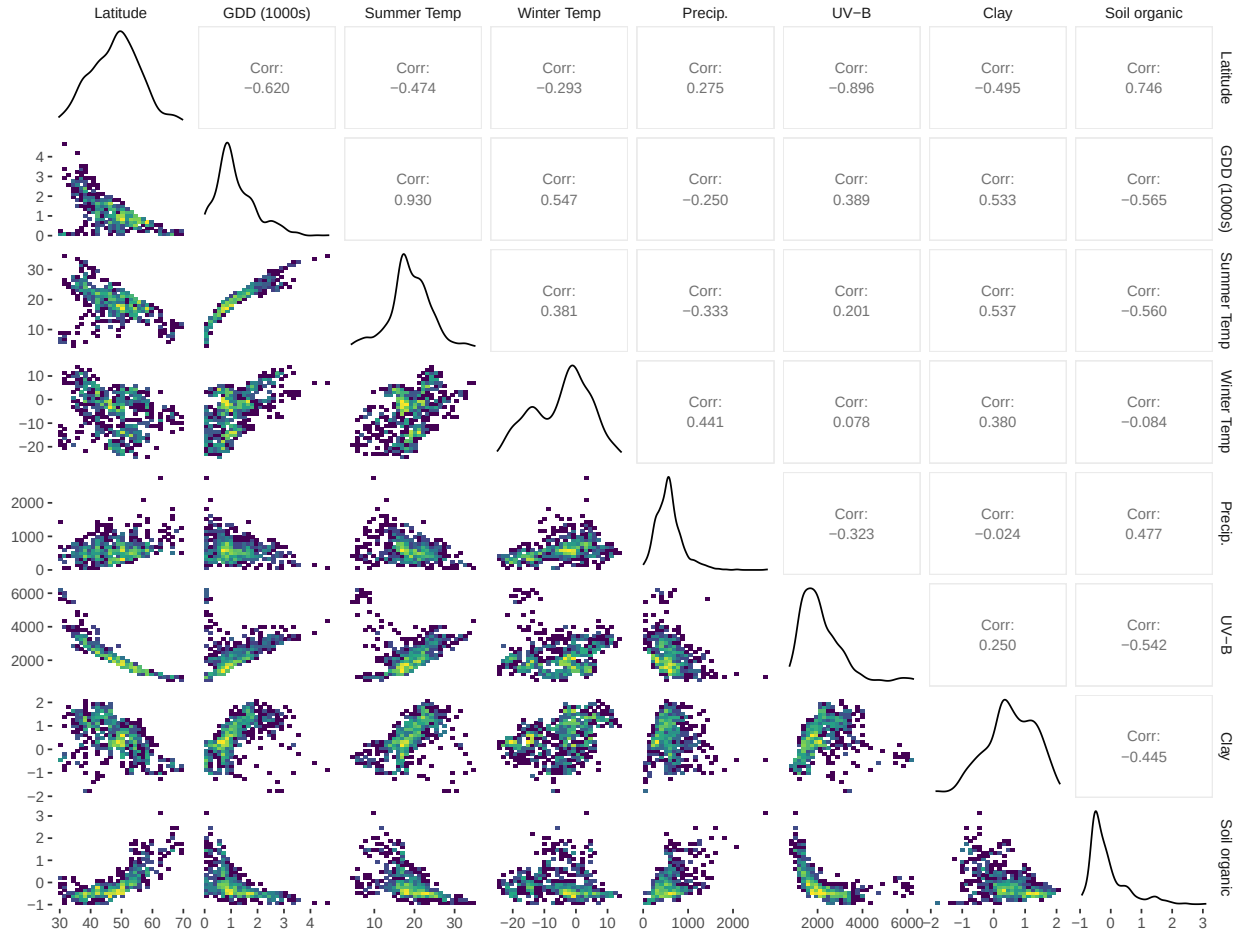
Notes: Panel A plots reconstructed temperature over time at 200-year resolution. The colored dashed lines show each of four spatial quadrants defined by the area's lon/lat midpoint (demeaned, right axis). The solid black line is the unweighted mean of the four quadrant series (left axis). Grey shaded bands mark the four consecutive 1,000-year windows shown in Panels C through F. Panel B displays the spatial distribution of reconstructed temperature at 4000 BCE, with dotted lines showing the quadrant boundaries. Panels C–F show the change in temperature across four consecutive millennial windows from 4000 BCE to 0. Temperature reconstructions are from Osman et al. (2021) at approximately 2.5° × 1.9° resolution.

Figure A4: Pairwise correlations between bilateral distance measures



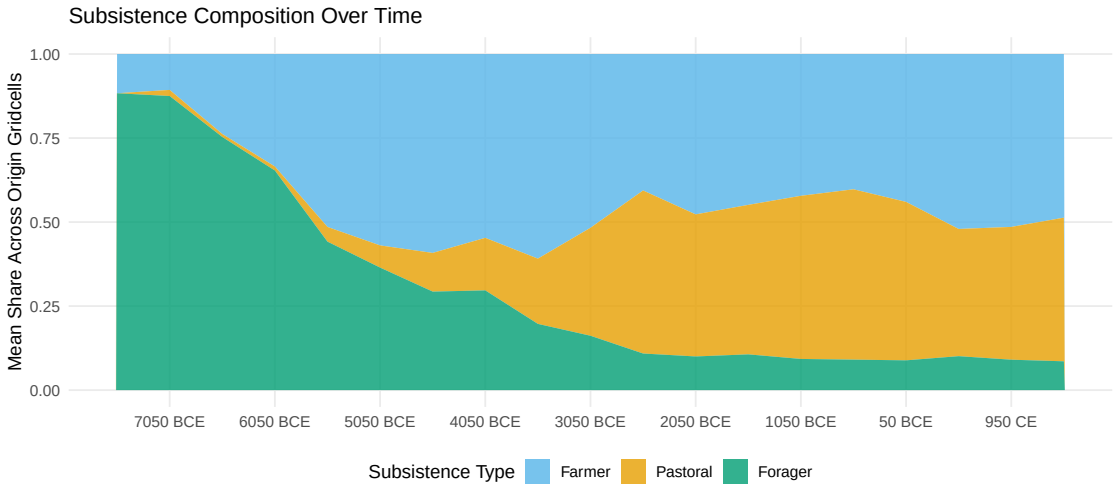
Notes: This figure plots pairwise correlations between bilateral distance measures across unordered pairs of gridcells appearing as origin or destination in our analytical sample (729 gridcells, ≈265,000 pairs; subsampled to 30,000 for the lower-triangle density visualization). Variables: Haversine geographic distance (km), latitude distance, GDD distance, summer temperature (July) distance, winter temperature (January) distance, precipitation distance, UV-B radiation distance, and soil composite distance (Euclidean over clay content and soil organic carbon). Diagonals show single variable distributions within the sample, lower triangle shows binned two-dimensional density (viridis palette, log-scale shading); upper triangle reports Pearson correlations. Temperature and GDD from ERA5 reanalysis (Hersbach et al., 2020); precipitation from TerraClimate (Abatzoglou et al., 2018); UV-B from glUV (Beckmann et al., 2014); soil data from SoilGrids (Hengl, 2018b; Hengl and Gupta, 2019; Hengl and Wheeler, 2018).

Figure A5: Pairwise correlations between climate and soil variables (per-cell levels)



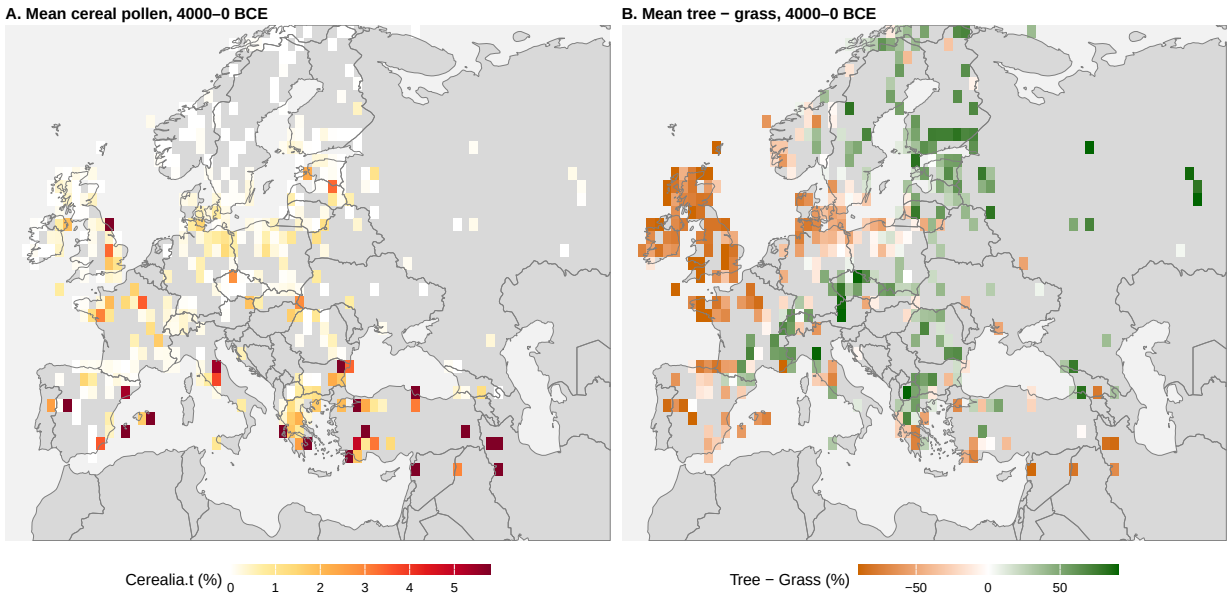
Notes: This figure plots pairwise correlations between per-cell environmental variables across all gridcells appearing as origin or destination in our analytical sample (729 gridcells). Variables: latitude, cumulative growing degree days above 10°C (GDD), summer temperature (July), winter temperature (January), precipitation, UV-B radiation, clay content, and soil organic carbon. Diagonals show single variable distributions within the sample, lower triangle shows binned two-dimensional density (viridis palette, log-scale shading); upper triangle reports Pearson correlations. Temperature and GDD from ERA5 reanalysis (Hersbach et al., 2020); precipitation from TerraClimate (Abatzoglou et al., 2018); UV-B from glUV (Beckmann et al., 2014); soil data from SoilGrids (Hengl, 2018b; Hengl and Gupta, 2019; Hengl and Wheeler, 2018). This per-cell view complements the bilateral distance correlations reported in Figure A4, which the main analysis relies on.

Figure A6: Subsistence Composition Over Time



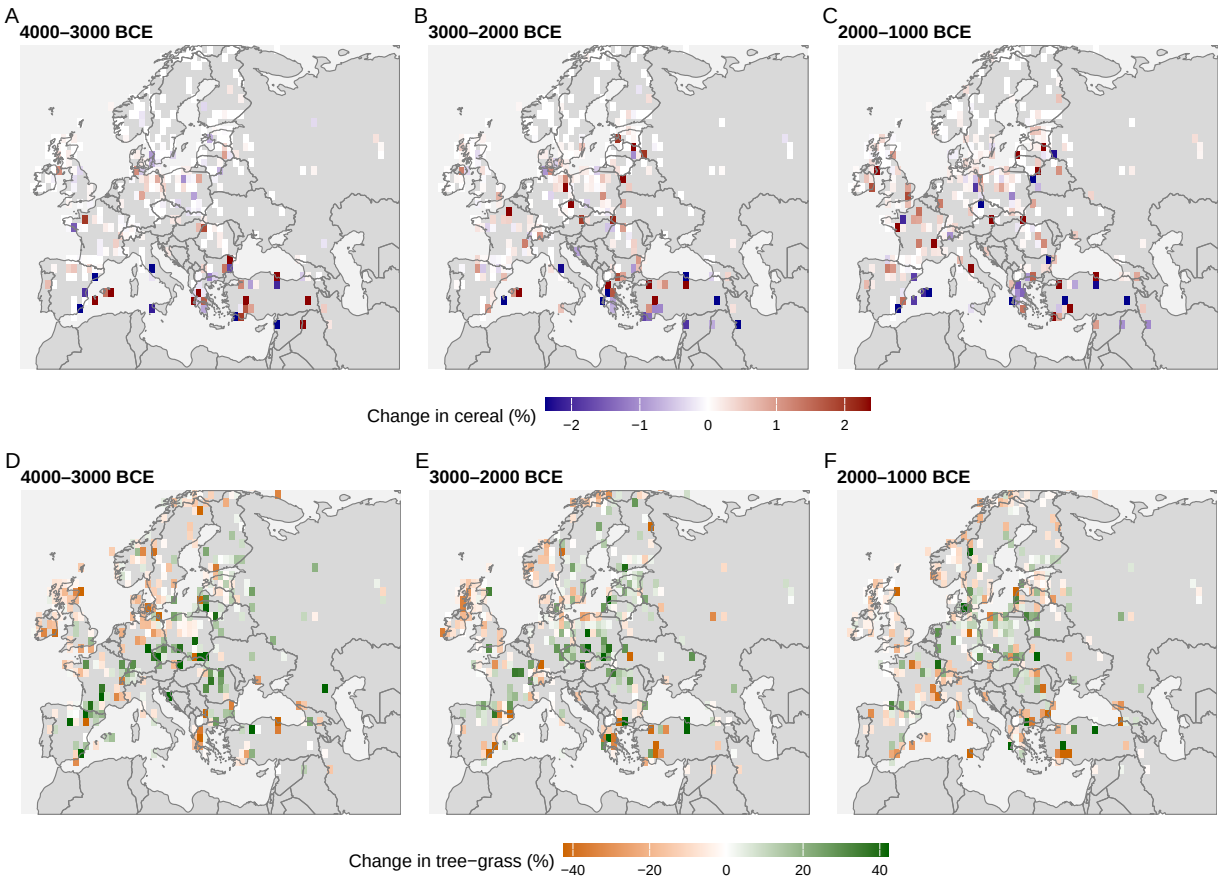
Notes: Stacked area plot showing the mean subsistence share across origin gridcells by 500-year era, using the archaeological culture group classification (continuous shares summing to one). The figure shows a transition from forager-dominated to farmer-dominated samples around 5000 BCE, with pastoralists emerging as a persistent share from approximately 3000 BCE.

Figure A7: Mean Cereal and Tree–Grass Pollen, 4000 BCE–0



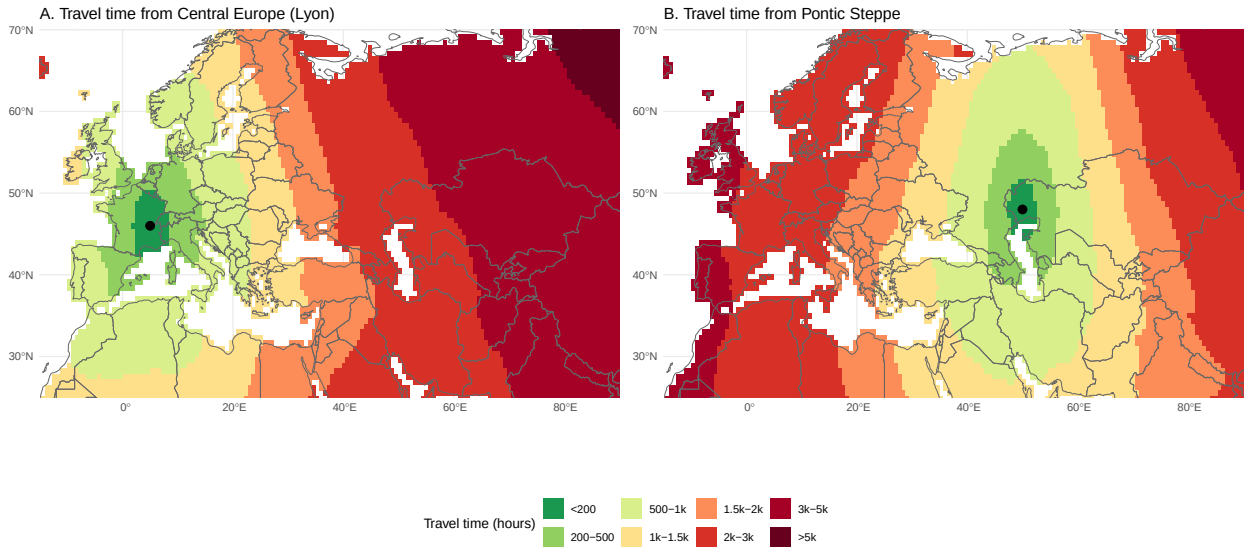
Notes: Spatial distribution of REVEALS bias-corrected pollen records across Western Eurasia, pooled across the four 1,000-year windows from 4000 BCE to 0 (390 1°×1° gridcells with pollen observations). Panel A: mean Cerealia-type pollen share (%); white = no cereal pollen detected, deeper red = higher cereal presence. Panel B: mean (Tree − Grass) pollen share (%); orange = grassland-dominated, green = forest-dominated, white = balanced. Light grey land cells indicate land surface with no pollen records in our sample (which is sparser in the Mediterranean and Eastern Europe). Source: Fyfe et al. (2021).

Figure A8: Vegetation Change Across Consecutive 1,000-Year Windows



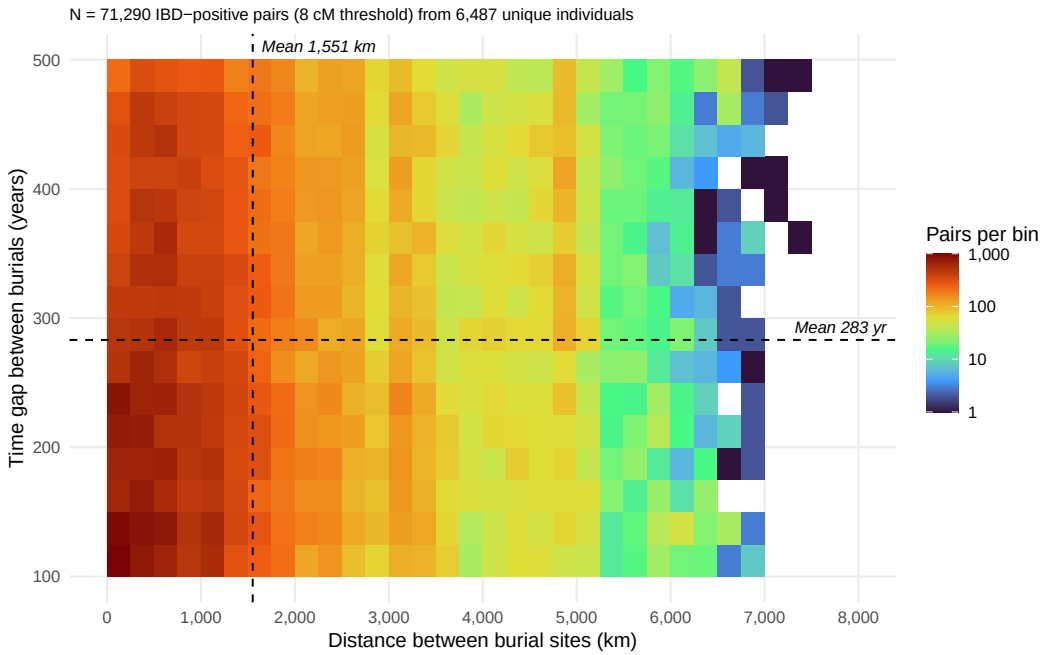
Notes: Change in pollen composition between adjacent 1,000-year era windows for the 336 gridcells with valid pollen records in all four windows. Top row: change in Cerealia-type pollen share (percentage points); blue = decline, red = expansion. Bottom row: change in (Tree − Grass) pollen share (percentage points); orange = grassland gain, green = forest gain. Each column reports the change between the indicated 1,000-year windows: 4000–3000 BCE compared with 3000–2000 BCE; 3000–2000 BCE compared with 2000–1000 BCE; and 2000–1000 BCE compared with 1000–0 BCE. The share of cells with positive cereal change rises monotonically across the three transitions (34.8%, 44.6%, 53.0%), consistent with the late-Holocene broadening of cereal cultivation that motivates the diffusion test in Section 6.4. Source: Fyfe et al. (2021).

Figure A9: Travel Time from the Human Mobility Index



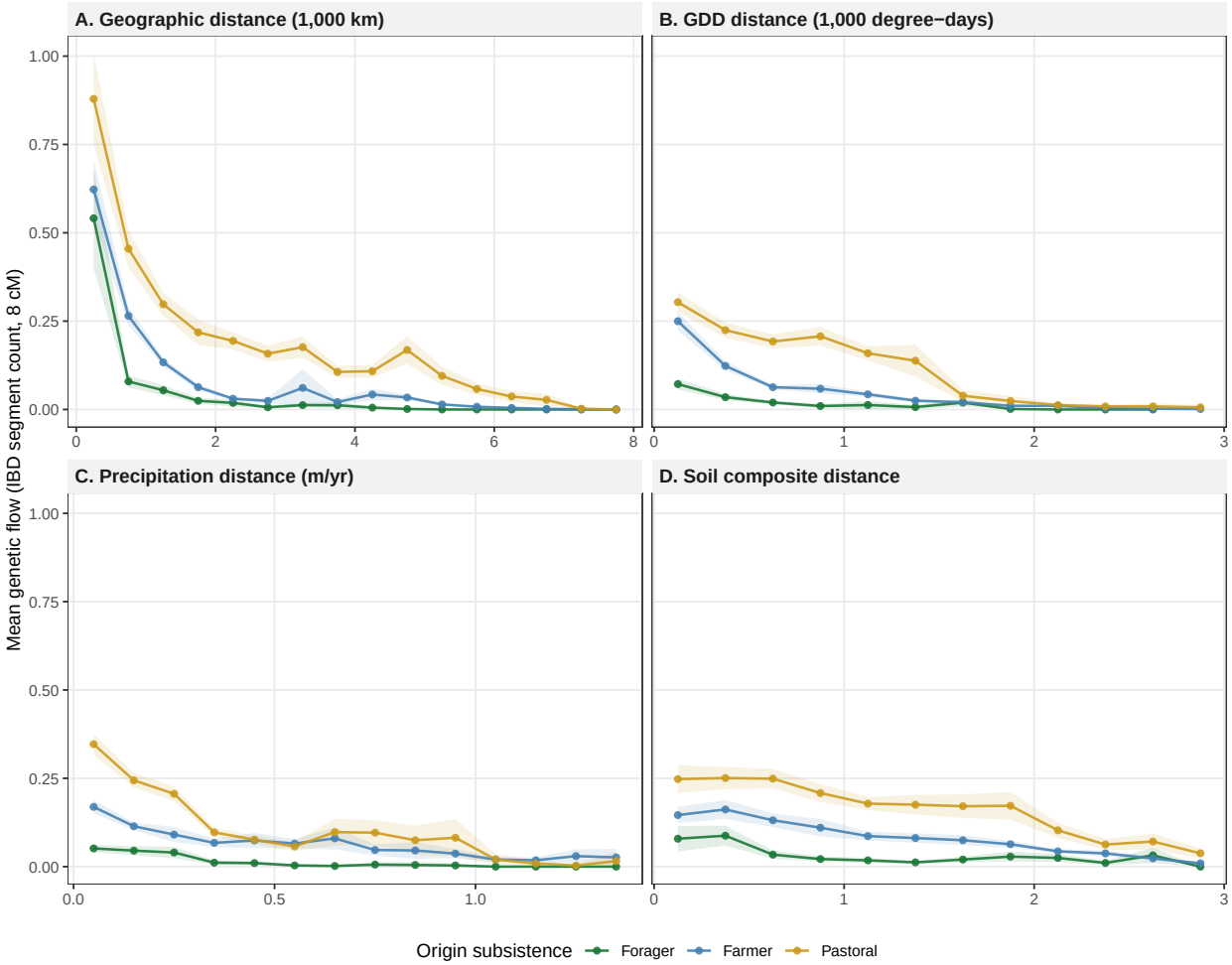
Notes: Each panel shows accumulated least-cost-path travel time (in hours of walking) from an illustrative origin (black dot), computed from the Human Mobility Index (Özak, 2018). The HMI estimates maximum sustainable walking speed at each location based on slope, terrain type, temperature, and humidity, incorporating pre-modern seafaring on historically navigable seas. Panel A originates from Central Europe (Lyon); Panel B from the Pontic Steppe. The asymmetry of the isochrones demonstrates that straight-line geographic distance can substantially misrepresent the actual frictions faced by prehistoric populations.

Figure A10: Joint Distribution of Geographic Distance and Time Gap, IBD-Positive Pairs



Notes: Two-dimensional binned scatter of geographic distance (km) and ancestor-descendant time gap (yr) for the 71,290 individual pairs sharing at least one IBD segment $\geq$ 8 cM, restricted to pairs separated by more than 50 km and 100–500 years. Bins are 250 km $\times$ 25 yr; shading denotes pair counts on a log scale. Reference lines mark sample means (1,550 km; 283 yr). The positive co-movement of distance and time gap is consistent with diffusive multi-generational dispersal.

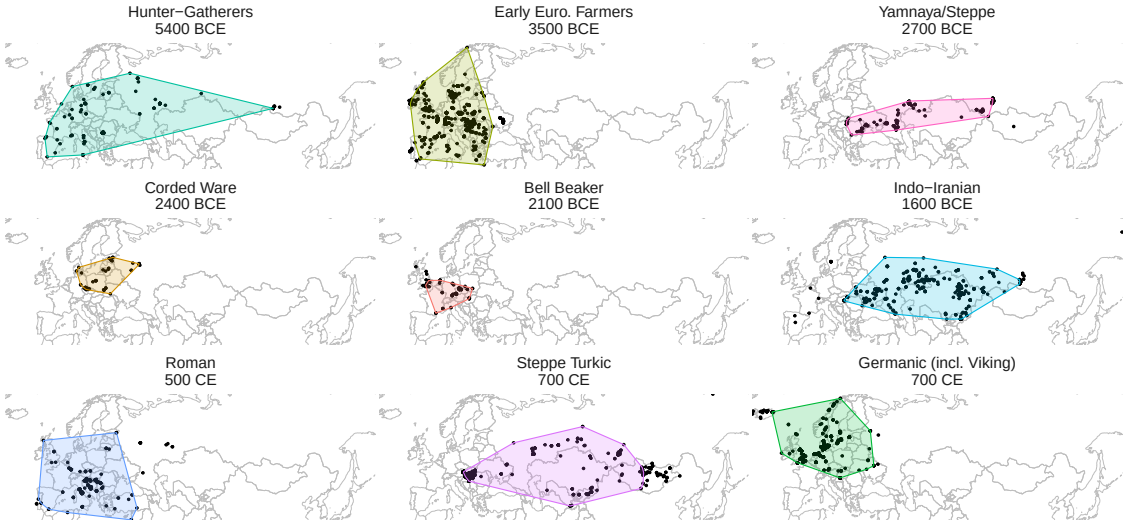
Figure A11: Genetic Flow by Ecological Distance, by Origin Subsistence



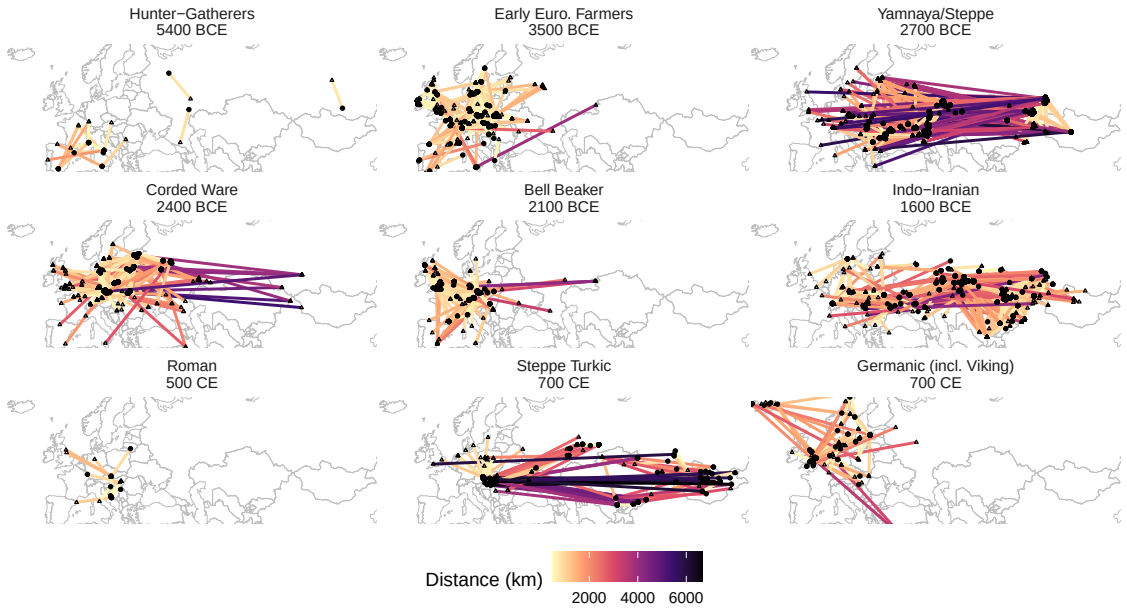
Notes: Each panel plots mean genetic flow (IBD segment count $\geq$ 8 cM) between 1-degree gridcell pairs against a measure of bilateral distance, split by the origin gridcell's subsistence classification (modal within each era). Sample is restricted to non-self pairs with origin era covering 8000 BCE to 1450 CE. Points are bin midpoints; ribbons show 95% confidence intervals. Bins with fewer than 100 observations are dropped. The panels parallel the bilateral-distance controls in the gravity specification (Table 2): geographic distance, GDD distance, precipitation distance, and soil composite distance. The subsistence contrast is most pronounced in GDD distance (farmers decay fastest) and precipitation distance (pastoralists decay fastest), consistent with the heterogeneity documented in Table 3.

Figure A12: Major Culture Groups

(a) Locations



(b) Movements



Notes: As in Figure 1, Panel B uses a 16 cM IBD threshold for visual clarity (rather than the 8 cM baseline used in the rest of the paper) and restricts to pairs living at most 500 years apart and separated by at least 500 km. Panel A shows the spatial distribution of ancient DNA individuals assigned to the indicated culture group: black dots mark burial sites, and shaded polygons are 95% convex hulls of site locations (computed after trimming observations beyond the 95th distance percentile from each group's centroid). Panel B plots IBD-inferred movements for the same groups, with origin (solid circle) and destination (triangle) of each linked pair connected by a line colored by geographic distance. Subpanel titles report the unweighted mean calendar date of individuals in each culture group, rounded to the nearest century; subpanels are ordered from oldest to most recent. The 11 culture groups shown are those used in Figure 5.

Table A1: Variable Definitions

Variable	Description	Source
Genetic Flow	Number of pairs of individuals with at least 1 segment greater than 8 cM between the two cells within an era (500 years period). Robustness specifications use the same construction with stricter thresholds (12 cM and 16 cM). We assign as origin (resp. destination) location the one of the older (resp. younger) of the two.	Ringbauer et al. (2024)
Climate		
GDD Distance	Absolute difference in cumulative growing-degree days above 10^{o}C (in thousands of degree-days) between the two locations.	Hersbach et al. (2020)
Temperature Distance	Absolute difference in mean annual temperature (^{o}C) between the two locations (used for summary statistics only; not a regressor in any output table).	Hersbach et al. (2020)
Diurnal Temp. Range Distance	Absolute difference in the daily diurnal temperature range (max minus min, ^{o}C) between the two locations.	Hersbach et al. (2020)
UV Distance	Absolute difference in ultraviolet (UV-B) radiation between the two locations (J/m^2/day, scaled by 1/100).	Beckmann et al. (2014)
Precipitation Dist.	Absolute difference in mean annual precipitation (mm/year, scaled by 1/1,000) between the two locations.	Abatzoglou et al. (2018)
Climate Change		
Temp. Diff. (space)	Mean annual temperature at the descendant's location minus the ancestor's location (^{o}C).	Hersbach et al. (2020)
Temp. Change (400yr)	Reconstructed temperature 200 years after period t minus the temperature 200 years before, at the origin gridcell (^{o}C); a 400-year window centered on t, with intermediate values linearly interpolated from the LGMR time series.	Osman et al. (2021)
Precip. Diff. (space)	Mean annual precipitation at the descendant's location minus the ancestor's location (mm/year).	Abatzoglou et al. (2018)
Soil		
Soil Composite Dist.	Euclidean distance over clay content and soil organic carbon between the two locations.	Hengl (2018b,a); Hengl and Gupta (2019); Hengl and Wheeler (2018)
Additional soil charact.	Distance in the (singularly taken) clay, organic carbon, water content, pH, bulk, and first principal component of these variables.	Hengl (2018b,a); Hengl and Gupta (2019); Hengl and Wheeler (2018)
Subsistence		
Subsistence type	Categorical assignment of each origin population to farmer, pastoralist, forager, or agro-pastoralist, based on archaeological culture group and a large language model classification cross-validated against a rule-based classifier.	Ringbauer et al. (2024); OpenAI o4-mini
Subsistence shares	Continuous shares (farmer, pastoralist, forager) summing to one, derived from the same classification procedure. Used in the pooled interaction specifications.	Ringbauer et al. (2024); OpenAI o4-mini
Geography		
Travel Cost	Least-cost-path walking time between gridcell centroids (in thousands of hours), derived from the Human Mobility Index cost surface.	Özak (2018)
Haversine Distance	Great-circle distance between gridcell centroids (in thousands of km). Used as a robustness alternative to travel cost in linear, quadratic, and logarithmic forms.	—
Latitude / Longitude Distance	Absolute difference in latitude and longitude (degrees) between the two cells. Used as additional spatial controls in robustness specifications.	—
Coastal Indicator	Equals 1 if origin and destination differ in coastal status (within 20 km of coast); 0 otherwise.	Özak (2018)

Continued on next page

Table A1 – continued from previous page

Variable	Description	Source
Elevation	Absolute difference in mean elevation (in thousands of meters) between the two cells.	NASA Shuttle Radar Topography Mission
Pollen		
Dest. Pollen Change	Change in vegetation at the destination between periods t and t+500 in the arable land (Cerealia-type pollen), grassland, and tree cover.	Fyfe et al. (2021)
Pollen Distance	Difference in vegetation between origin and destination at time t in the arable land (Cerealia-type pollen), grassland, and tree cover.	Fyfe et al. (2021)

Note: GDD is constructed from daily ERA5 temperatures; precipitation is constructed from monthly TerraClimate aggregates. All climate variables are averaged over 1960–2000 within each $1° \times 1°$ gridcell unless otherwise specified.

Table A2: Selected Long-Distance Ancestor-Descendant Corridors

Ancestor	Anc. ID	Anc. year	Descendant	Desc. ID	Desc. year	Dist. (km)	Time (yr)	Speed (km/yr)	IBD (cM)
Italy, Epigravettian	UZZ069.AG	6678 BCE	Serbia (Iron Gates), Mesolithic	I4877.AG	6555 BCE	1,053	123	8.6	47.4
Austria, LBK	I24903.AG	5150 BCE	Ukraine, Neolithic	I3719.AG	4881 BCE	1,267	269	4.7	32.7
Russia (Samara), Yamnaya	I6733.AG	3213 BCE	Czechia, Corded Ware	VLI007.AG	2789 BCE	2,172	424	5.1	54.8
Russia (Samara), Yamnaya	I6733.AG	3213 BCE	Russia, Afanasievo	I13812.AG	2900 BCE	2,586	313	8.3	139.8
Russia, Afanasievo	I26089.AG	3132 BCE	Moldova, Yamnaya	I7847.AG	2961 BCE	4,178	171	24.4	85.4
Russia (Don), Yamnaya	I24086.AG	2929 BCE	Russia (Upper Yenisei), Afanasievo	I32497.TW	2800 BCE	3,544	129	27.5	70.6
Russia, Afanasievo	I13813.AG	2900 BCE	Czechia, Corded Ware	TRM003.AG	2670 BCE	4,657	230	20.2	66.0
Russia, Afanasievo	I25160.AG	2700 BCE	England, Bell Beaker	I5512.AG	2300 BCE	5,556	400	13.9	19.7
Russia, Afanasievo	I25159.AG	2700 BCE	Greece, Sarakenos	G37.SG	2400 BCE	5,012	300	16.7	19.7
England, Early Bronze Age	I20997.AG	2125 BCE	Kazakhstan, Alakul	I6789.AG	1782 BCE	5,048	343	14.7	26.6
Russia (S. Urals), Sintashta	I1064.AG	1789 BCE	Portugal, Middle Bronze Age	TV3831.SG	1600 BCE	5,474	189	29.0	10.1
Moldova, Cimmerian	cim359.SG	925 BCE	Kazakhstan, Tasmola	KSH002.AG	607 BCE	3,370	318	10.6	19.6
Russia (S. Urals), Early Sarmatian	MJ-43.SG	362 BCE	Mongolia (Arkhangai), Xiongnu	NAI001.AG	50 BCE	3,035	312	9.7	22.4
China (Altai/Rouran), Xianbei	MGS-M7R.SG	150 CE	Hungary, Early Avar	KUP003.AG	625 CE	7,036	475	14.8	10.9
Mongolia (Sukhbaatar), Xiongnu	TAV005.AG	650 CE	Hungary, Late Avar	KFJ049.AG	762 CE	6,755	112	60.3	13.1
Austria (Vienna Basin), Avar	LEO098.AG	725 CE	Mongolia, Zubu	GD2-2.AG	932 CE	6,529	207	31.5	20.0
Russia (S. Urals), Kushnarenkovo	I19117.AG	850 CE	Hungary, Magyar	K3-6.SG	980 CE	2,793	130	21.5	26.7
Hungary, Magyar	K2-61$_d$.*SG*	925 CE	Russia (Volga-Urals), Chiyalik	I19081.AG	1085 CE	2,254	160	14.1	56.3

Notes: Each row reports a curated origin-destination population corridor selected to illustrate distinct long-distance migration episodes; the table is not an algorithmic top-N. The displayed individual pair is the highest-IBD ancestor-descendant link within the population cluster (the pair with the strongest evidence of recent shared ancestry). Sample restricted to pairs separated by more than 50 km and time gap 100–500 years. Specimen IDs and population labels follow Ringbauer et al. (2024). Cluster size (number of individual ancestor-descendant pairs in the cluster sharing at least one IBD segment ≥8 cM), by row: Italy, Epigravettian → Serbia (Iron Gates), Mesolithic: 11 pairs; Austria, LBK → Ukraine, Neolithic: 15 pairs; Russia (Samara), Yamnaya → Czechia, Corded Ware: 168 pairs; Russia (Samara), Yamnaya → Russia, Afanasievo: 236 pairs; Russia, Afanasievo → Moldova, Yamnaya: 48 pairs; Russia (Don), Yamnaya → Russia (Upper Yenisei), Afanasievo: 14 pairs; Russia, Afanasievo → Czechia, Corded Ware: 269 pairs; Russia, Afanasievo → England, Bell Beaker: 25 pairs; Russia, Afanasievo → Greece, Sarakenos: 6 pairs; England, Early Bronze Age → Kazakhstan, Alakul: 14 pairs; Russia (S. Urals), Sintashta → Portugal, Middle Bronze Age: 5 pairs; Moldova, Cimmerian → Kazakhstan, Tasmola: 7 pairs; Russia (S. Urals), Early Sarmatian → Mongolia (Arkhangai), Xiongnu: 5 pairs; China (Altai/Rouran), Xianbei → Hungary, Early Avar: 3 pairs; Mongolia (Sukhbaatar), Xiongnu → Hungary, Late Avar: 5 pairs; Austria (Vienna Basin), Avar → Mongolia, Zubu: 11 pairs; Russia (S. Urals), Kushnarenkovo → Hungary, Magyar: 2 pairs; Hungary, Magyar → Russia (Volga-Urals), Chiyalik: 35 pairs.

Table A3: Presence of Archaeological Sites across Climate Areas

Dependent Variable:	1[at least one obs.]			
Model:	(1)	(2)	(3)	(4)
Deg. Days $\geq 10^oC$	-0.027***	-0.025***	-0.039***	-0.016
	(0.002)	(0.002)	(0.003)	(0.010)
Precipitation		0.055***	0.018	-0.124***
		(0.012)	(0.017)	(0.038)
Clay			0.045***	0.092***
			(0.007)	(0.014)
Soil Organic			-0.057***	-0.022
			(0.007)	(0.017)
Fixed-effects				
2.5 gridcell				Yes
Fit statistics				
Observations	6,071	6,071	6,052	6,029
R^2	0.023	0.028	0.062	0.438
Within R^2				0.012

Notes: OLS estimates. Dependent variable: indicator for at least one ancient DNA individual observed in a 1-degree gridcell. Variables parallel those used in the main gravity specification. Standard errors clustered at the 2.5-degree level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Migration Distance, Time, and Speed by Subsistence Type

Subsistence	N pairs	Mean dist. (km)	Mean time gap (yr)	Median speed (km/yr)
All	71,290	1,551	283	4.04
Farmer	28,503	1,100	282	2.86
Forager	1,025	905	295	2.29
Agro-Pastoral	6,993	1,431	290	4.19
Pastoral	34,769	1,964	282	5.65

Notes: Each row reports summary statistics for IBD-linked ancestor-descendant pairs (segment $\geq$ 8 cM) by origin subsistence type, pooled across 500-year eras. Sample restricted to pairs separated by more than 50 km, with time gaps of 100–500 years, and origin era younger than 8000 BCE. The "All" row pools across subsistence types. Subsistence is classified at the origin using archaeological culture group labels, with agro-pastoral origins grouped as mixed farmer-pastoralist; pairs with unclassified subsistence are dropped. Distance is the great-circle distance between origin and destination gridcells. Speed is the per-pair ratio of distance to time gap; the median is reported because the per-pair speed distribution has a heavy right tail.

Table A5: Culture Group Classifications

Culture Group	Period	Type	Subsistence Basis	Key Reference
Forager				
Western Hunter-Gatherers (WHG)	14,000–5500 BCE	Forager	Hunting, fishing, wild plant gathering (Western and Southern Europe)	Bailey and Spikins (2008)
Eastern Hunter-Gatherers (EHG)	12,000–5500 BCE	Forager	Hunting, fishing, wild plant gathering (Eastern European and Russian forest–steppe)	Bailey and Spikins (2008)
Farmer				
Early Euro. Farmers (LBK)	6500–4000 BCE	Farmer	Wheat, barley, cattle/pig husbandry	Bogucki (1988); Shennan (2018)
Bell Beaker	2800–1800 BCE	Farmer	Cereal cultivation, animal husbandry	Vander Linden (2024)
Aegean Bronze Age	3000–1100 BCE	Farmer	Mediterranean polyculture (wheat, olive, vine)	Halstead (2014)
Roman/Imperial	27 BCE–476 CE	Farmer	Intensive cereal agriculture, viticulture	White (1970)
Germanic (incl. Viking)	400–1100 CE	Farmer	Mixed cereal/livestock on dispersed farmsteads	Hamerow (2002); Brink and Price (2008)
Pastoral				
Yamnaya	3300–2500 BCE	Pastoral	Mobile cattle/sheep pastoralism, dairy, wagons	Anthony (2007)
Corded Ware	2900–2300 BCE	Pastoral	Pastoral-oriented with some cereal production	Kristiansen (1998); Sjögren et al. (2016)
Indo-Iranian (Sintashta)	2100–900 BCE	Pastoral	Stockbreeding (cattle, sheep, horse), metallurgy	Kuz'mina (2007); Anthony (2007)
Iron-Age Nomads	900 BCE–400 CE	Pastoral	Nomadic horse pastoralism, transhumance	Rolle (1989); Khazanov (1994)
Steppe Turkic (Avar/Hun)	370–900 CE	Pastoral	Nomadic herding (horse, cattle, sheep), dairy	Golden (1992); Khazanov (1994)

Notes: Culture groups correspond to those in Figure 5. Subsistence classifications are based on the dominant mode of production documented in the archaeological literature. Populations are assigned to culture groups using keyword matching of population labels from Ringbauer et al. (2024), with subsistence shares estimated from archaeological and ethnographic sources (see Section 3). WHG and EHG are genetically and geographically distinct hunter-gatherer ancestries that we pool into a single "Hunter-Gatherers" group in the regression analysis to ensure adequate sample size.

Table A6: Adaptive Migration by Subsistence Type

Dependent Variable:	Genetic Flow				
	Pooled	Farmer	Pastoral	Forager	Joint
Model:	(1)	(2)	(3)	(4)	(5)
Temp. Change (origin)	0.937***	-0.780**	2.68***	0.978***	0.878***
	(0.191)	(0.331)	(0.283)	(0.240)	(0.277)
Temp. Change × Farmer Share		5.15***			3.63***
		(0.659)			(0.557)
Temp. Change × Pastoral Share			-4.22***		-2.86***
			(0.607)		(0.528)
Temp. Change × Forager Share				-0.137	
				(0.336)	
Farmer Share		-1.58***			-0.478***
		(0.149)			(0.174)
Pastoral Share			2.42***		1.96***
			(0.189)		(0.229)
Forager Share				-0.066	
				(0.147)	
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	478,034	478,034	478,034	478,034	478,034
Pseudo R^2	0.48	0.49	0.49	0.48	0.49

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments ≥ 8 cM between 1° gridcell pairs in a 500-year era. Temp. Change (origin) is the Osman et al. (2021) reconstructed temperature change over a centered 400-year window (positive = warming). Subsistence shares are continuous measures at the origin, based on archaeological culture group classification. 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. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Robustness Checks

B.1 Modern Climate to Measure Past Bilateral Climate Distances

The bilateral matching analysis uses modern climate fields (ERA5 for temperature, TerraClimate for precipitation) to measure spatial gradients between origins and destinations, while paleoclimate reconstructions (LGMR) are used for temporal change at fixed origin locations in the adaptation analysis. This appendix quantifies the bias from using modern fields as a proxy for past bilateral temperature gradients. We decompose the implied measurement error into two components: a small within-LGMR temporal evolution that genuinely represents proxy bias, and a much larger LGMR-vs-ERA5 disagreement at the modern bin that reflects LGMR's spatial error against an observation-anchored field rather than proxy bias.

Decomposition. For an origin-destination pair (o, d) at era t, the bilateral temperature distance using modern climatology is $\Delta T_{od}^{\text{ERA5}} = T_o^{\text{ERA5}} - T_d^{\text{ERA5}}$, while the corresponding paleoclimate quantity is $\Delta T_{od,t}^{\text{paleo}} = T_{o,t}^{\text{paleo}} - T_{d,t}^{\text{paleo}}$ taken from the LGMR reconstruction (Osman et al., 2021). The proxy residual

$$\varepsilon_{od,t} \;=\; \Delta T_{od,t}^{\text{paleo}} - \Delta T_{od}^{\text{ERA5}}$$

admits an exact decomposition by adding and subtracting the LGMR's own modern bin $\Delta T_{od,0}^{\text{paleo}}$:

$$\varepsilon_{od,t} \;=\; \underbrace{\left(\Delta T_{od,t}^{\text{paleo}} - \Delta T_{od,0}^{\text{paleo}}\right)}_{\varepsilon^*_{od,t}\text{: within-LGMR temporal evolution}} \;+\; \underbrace{\left(\Delta T_{od,0}^{\text{paleo}} - \Delta T_{od}^{\text{ERA5}}\right)}_{\delta_{od}\text{: LGMR-vs-ERA5 modern-bin disagreement}} .$$

The first term, $\varepsilon^*_{od,t}$, measures the actual temporal evolution of bilateral gradients across the Holocene within a single reconstruction; LGMR's own spatial biases cancel in the difference because both terms come from LGMR. The second term, δ_{od}, measures the disagreement between LGMR and ERA5 at the modern bin and is constant across eras for any given pair. Only the first term represents proxy bias from substituting modern for paleo; the second term reflects disagreement between two estimates of modern climate—ERA5, which is anchored to direct station and satellite observations, and LGMR's modern bin, which is not directly observed but inferred from sparse marine geochemical proxies and the iCESM model prior. The decomposition itself is symmetric in the two fields and does not adjudicate which is closer to truth at the modern bin; we treat ERA5 as the more accurate estimate based on its observational anchoring and resolution, as discussed below.

Within-LGMR temporal evolution. Across 531,137 origin-destination-era observations

covering 19 500-year eras of the Holocene—the same sample as Table 2—the pooled standard deviation of $\varepsilon^*_{od,t}$ is 0.53°C against a pooled bilateral cross-sectional spread $\mathrm{SD}(\Delta T^{\text{paleo}})$ of 8.67°C, capturing 99.62% of the variance in bilateral paleoclimate temperature distances within LGMR. Table B1 reports the decomposition by 1,000-year interval: the within-LGMR temporal evolution grows with age, ranging from 0.32°C in the most recent millennium to 0.73°C in the deepest densely-populated bin (8,000–9,000 BP); the 9,000–10,000 BP bin reaches 1.10°C but contains only 427 observations. The variance share captured exceeds 99% in every interval up to 9,000 BP. With pooled residual variance $\mathrm{Var}(\varepsilon^*) \approx 0.28\ ^\circ\mathrm{C}^2$ and bilateral cross-sectional variance $\mathrm{Var}(\Delta T^{\text{paleo}}) \approx 75\ ^\circ\mathrm{C}^2$, the classical measurement error attenuation factor under the assumption of unstructured measurement error is

$$\frac{\mathrm{Var}(\Delta T^{\text{paleo}})}{\mathrm{Var}(\Delta T^{\text{paleo}}) + \mathrm{Var}(\varepsilon^*)} \approx 0.996,$$

biasing our temperature distance coefficients toward zero by approximately 0.4%. The within-LGMR residual is mildly correlated with the bilateral paleoclimate gradient itself ($r \approx -0.44$ in the production sample), so classical measurement error is an approximation; even sizable departures from classical assumptions cannot meaningfully change the magnitude given how small the implied bias is. This is negligible relative to the standard errors on the growing degree day coefficient in Table 2 and runs against finding effects.

LGMR paleoclimate vs. ERA5 modern reanalysis at the modern bin. The second term in the decomposition, δ_{od}, is the disagreement between LGMR's modern bin (0–200 BP) and ERA5 at the same locations. Across the same 531,137 origin-destination-era observations the standard deviation of δ_{od} is 3.3°C. Three sources contribute. First, LGMR's $2.5^\circ \times 1.9^\circ$ native resolution smooths out spatial structure that ERA5 resolves at 9 km, particularly across mountain ranges, coastal gradients, and continental interiors. Second, LGMR's 1–2°C gridcell level uncertainty (Osman et al., 2021) is present at every era, including the modern bin. Third, data-assimilation reconstructions inherit the spatial covariance structure of their model prior; for LGMR, that prior is the iCESM ensemble used in the LGM data assimilation of Tierney et al. (2020), which propagates information from sparse marine proxies to data-poor continental interiors through model-derived covariances rather than direct observation.

At the modern bin, ERA5 is therefore the more reliable estimate of the spatial gradient. Substituting LGMR for ERA5 in the bilateral analysis would introduce additional bilateral measurement error of order 3°C at every era, which is large relative to the 0.32–0.73°C within-LGMR temporal evolution that substitution would in principle correct. The arithmetic favors ERA5.

B.2 Robustness of Climate Matching Results

Additional environmental controls. Our baseline specification already includes a rich set of controls—geographic distance, precipitation, temperature (GDD), and soil distance—together with origin, destination, and era fixed effects. Nonetheless, it is possible that other environmental factors correlated with ecological distance may independently shape migration patterns. Some of these factors, such as latitude and solar exposure, may themselves be economically relevant and consistent with the framework in Section 2. Table B2 assesses the robustness of our results to the inclusion of such additional environmental controls.

Column 1 replicates the baseline specification (column 4 of Table 2). Columns 2 through 4 sequentially add distances in diurnal temperature range, elevation, and an indicator for coastal-to-non-coastal movements. Each enters with a negative and statistically significant coefficient, indicating that these additional environmental margins also shape mobility; the baseline coefficients on precipitation, soil, and travel time remain stable, with only modest attenuation in the GDD coefficient.

Column 5 adds ultraviolet (UV-B) radiation distance, constructed from global satellite-derived UV-B data (Beckmann et al., 2014). The coefficient is negative and highly significant: a one-standard-deviation increase in UV-B distance—approximately 900 J/m^2/day, equivalent to the difference between Scandinavia and the central Mediterranean—is associated with a 54% reduction in bilateral genetic flow. The coefficient on GDD becomes small and statistically insignificant, while precipitation distance remains stable. This pattern reflects the correlation structure of the environmental variables: UV-B distance correlates strongly with latitude distance ($r = 0.77$) and moderately with GDD distance ($r = 0.43$), but is nearly orthogonal to precipitation distance ($r = 0.11$). UV-B thus captures the north-south solar and thermal gradient more precisely than GDD—which is confounded by continentality, ocean currents, and altitude—and absorbs the latitudinal component of the GDD signal. The diurnal range and elevation coefficients attenuate to near zero and become statistically insignificant while the coastal indicator strengthens (from -0.155 to -0.235). With latitude and longitude distance controls (column 6), the UV-B coefficient remains significant but its standard error nearly doubles, suggesting that much of its explanatory power operates through the latitude gradient, with the remainder capturing residual variation in the solar environment.

The strength of UV-B matching can be viewed through two complementary interpretations. First, UV-B may simply be a cleaner proxy for some latent north-south ecological gradient than GDD, with no independent biological channel. Or more interestingly, UV-B matching may reflect biological adaptation to the solar environment: UV-B radiation is the strongest environmental predictor of human skin pigmentation globally (Jablonski and

Chaplin, 2000; Chaplin, 2004), and ancient DNA evidence documents active selection on pigmentation loci (SLC24A5, SLC45A2) in Western Eurasia during our study period (Mathieson et al., 2015; Akbari et al., 2026). Under this interpretation, populations migrating to UV-dissimilar environments would face fitness costs from vitamin D deficiency in one direction or folate photolysis in the other. We cannot empirically distinguish these channels, but both are consistent with the broader finding that ecological similarity shaped prehistoric migration.

Column 6 adds latitude and longitude distance. Both enter with negative but statistically insignificant coefficients, indicating that once the UV-B gradient is controlled for, broad spatial orientation adds little additional explanatory power.

Overall, these results confirm that the relationship between ecological distance and migration is robust to the inclusion of a wide range of additional environmental controls. The only notable exception is GDD, whose coefficient becomes unstable when controlling for UV exposure, reflecting the high correlation between these measures.

Soil decomposition. Our baseline specification uses a composite soil distance defined as the Euclidean distance between clay content and soil organic carbon (see Section 3.3). In column 2 of Table B3, we replace this composite index with the first principal component of a broad set of soil characteristics (organic carbon, bulk density, pH, water content, and clay content); the results are similar to the baseline (reported in column 1). Columns 3–7 then replace the composite index with distances in individual soil characteristics, introduced one at a time. Across all specifications, the coefficients on GDD, precipitation, and geographic distance remain negative, statistically significant, and similar in magnitude to the baseline. The individual soil measures also enter with negative and statistically significant coefficients, indicating that migration flows decline with differences in both geologically determined properties (such as clay content) and more directly productivity-related characteristics (such as soil organic carbon and water-holding capacity).

Column 8 includes all soil characteristics simultaneously. All coefficients remain negative, except for that on soil bulk, and in most cases statistically significant. The coefficients on precipitation and geographic distance remain stable, while the coefficient on GDD attenuates modestly but remains statistically significant.

Direction of move and distance bands. Our baseline specification controls for geographic distance, but the spatial structure of the study region may still shape feasible migration paths. In particular, coastlines and geographic barriers can generate effective east–west corridors, potentially confounding climate matching with movement along constrained routes. Table B4 addresses these concerns. Columns 2 and 3 restrict the sample to horizontal (east–west) and vertical (north–south) moves, respectively. The coefficients on

GDD, soil, and geographic distance remain negative and broadly stable across both subsamples. By contrast, the precipitation coefficient becomes smaller and statistically insignificant when focusing on vertical moves, consistent with the fact that precipitation varies primarily along east–west gradients and exhibits less variation in the north–south direction.

Columns 4–6 further examine robustness across distance bands by restricting the sample to pairs of locations separated by at least 100, 200, and 300 km, respectively. Climate matching effects are present across these distance thresholds, with negative and statistically significant coefficients for most environmental dimensions. The coefficients on precipitation, soil, and geographic distance remain stable in both magnitude and significance as the minimum distance increases. By contrast, the GDD coefficient attenuates and becomes less precisely estimated when restricting the sample to pairs beyond 200 and 300 km, consistent with the idea that thermal gradients are more tightly identified at shorter distances, while longer-distance comparisons span broader climatic ranges with less precise temperature matching.

Alternative measures of geographic distance. A related concern is that spatial correlation in climate may not be fully captured by the travel cost measure, as in our baseline specification. If estimated walking-time distance is misspecified, it may imperfectly proxy for underlying spatial gradients—such as latitude or coast-to-interior variation—potentially confounding the estimated effects of ecological distance. At the same time, such confounding is unlikely to explain the full set of results, as the different environmental dimensions we consider—temperature, precipitation, and soil—exhibit distinct spatial patterns and gradients, as documented in Section 3.2.

Table B5 addresses this concern by varying the functional form and measurement of geographic distance. Columns 2, 3, and 4 replace the HMI index with linear, quadratic, and logarithmic transformations of Haversine distance, respectively. The coefficients on precipitation, soil, and geographic distance remain negative, statistically significant, and similar in magnitude to the baseline. By contrast, the coefficient on GDD attenuates and becomes statistically insignificant, consistent with the fact that thermal gradients are strongly correlated with latitude and therefore more sensitive to alternative parameterizations of distance that capture broad spatial trends.

Alternative migration measures. In the baseline specification, we measure migration as the number of individual pairs sharing at least 8 cM of identity-by-descent (IBD) segments. A potential concern is that this threshold may capture relatively distant genealogical links, thus including movements that are less directly tied to recent migration. Table B6 addresses this issue by replicating the analysis using more stringent thresholds of 12 cM (column 2) and 16 cM (column 3), which correspond to more recent genealogical connections (see Sec-

tion 5). As Appendix Figure A1 documents, the underlying pair-level distance and time-gap distributions track closely across the three thresholds, so the qualitative results should not depend on the threshold choice. The coefficient on GDD remains negative and similar in magnitude to the baseline (reported in column 1), but becomes statistically insignificant under the stricter thresholds. By contrast, the coefficients on precipitation, soil, and geographic distance remain negative, statistically significant, and, if anything, larger in magnitude.

Winsorizing extreme observations and excluding 1,000 year periods. In Figure B1, we assess robustness to outlier observations and to specific historical episodes that may drive our results. First, we replicate the baseline specification while omitting observations from each 1,000-year period in turn, to ensure that no single episode drives the estimates. Next, we re-estimate the baseline while winsorizing the top 1% of the climate distance measures and, separately, the outcome variable. Reassuringly, the coefficients on precipitation, soil, and travel-time distance remain virtually unchanged relative to the baseline (reported at the top of each panel) across specifications. By contrast, the GDD coefficient is somewhat more sensitive to the exclusion of specific periods, consistent with patterns documented elsewhere in the robustness analysis.

Individual-level specification. The baseline analysis aggregates IBD segments to origin-destination-period cells. As an alternative, Table B7 estimates the gravity specification at the level of all the possible individual ancestor-descendant pairs, with the dependent variable an indicator for whether each pair shares at least one IBD segment of 8 cM or longer. The specification uses the same climate, soil, and travel-time controls as the baseline, with origin gridcell, destination gridcell, and era fixed effects. Standard errors are clustered at the gridcell pair level. With over 3 million pairwise observations, the climate-matching coefficients are precisely estimated and consistent with the gridded baseline: GDD, precipitation, soil, and travel-time distance all enter with negative and significant coefficients.

Alternative standard errors. We also assess robustness to alternative clustering structures. Table B8 reports the results. Column 2 implements multiway clustering on origin and destination gridcells, and column 3 clusters standard errors at the 1° origin. The coefficient on GDD distance remains negative and statistically significant under both clustering specifications. The coefficients on precipitation, soil, and travel time distance are stable in magnitude and remain significant at the 1% level across all specifications.

Granular variation and inference: Permutation test. As noted in Section 5 and following Dingel and Tintelnot (2026), the sparsity of the migration matrix raises concerns that idiosyncratic noise in thin origin–destination cells could drive the baseline estimates. To assess this, we perform a simulation exercise in which we randomly reshuffle temperature,

precipitation, and soil distances across origin–destination pairs within each era 1,000 times, while holding geographic distance constant, and re-estimate the baseline specification (column 4 of Table 2) in each iteration. We then compare the distribution of placebo coefficients to the baseline estimate and compute empirical p-values.

Figure B2, Panel A, reports the results. For all coefficients, the baseline estimates lie well outside the distribution of placebo estimates, indicating that the estimated effects are unlikely to be driven by noise or random variation in migration flows. Panel B repeats the exercise while restricting the reshuffled climate distances to be drawn only from the distribution of distances among cells within a ($10° \times 30°$) neighborhood. The results are unchanged: baseline estimates again lie well outside the placebo distribution.

B.3 Robustness of Subsistence Heterogeneity Results

Section 6.2 shows that ecological matching varies across subsistence types (Table 3); here we provide additional robustness using alternative specifications.

Individual-level specification. In Table B9, we replicate the analysis at the individual level using all possible ancestor–descendant pairs, following the approach in Table B7.[22] Reassuringly, the results are consistent with those reported in Table 3.

Interacted specification. Table B10 complements the split-sample analysis by estimating the same specification in the pooled sample, interacting environmental distance measures with the share of the origin population classified as farmer or pastoralist (with foragers as the omitted category). The results closely mirror those in Table 3.

Columns 2 and 3 focus on GDD interactions. The coefficient on farmer share $\times$ GDD distance is large, negative, and highly significant (-1.42), while the coefficient on pastoral share $\times$ GDD distance is large and positive ($+1.62$), implying that GDD matching is strongest for farming populations and weakest—or even reversed—for pastoralists, consistent with the idea that thermal constraints bind primarily for agricultural production. Columns 4 and 5 report analogous specifications for precipitation with one interaction at a time. Both the farmer and pastoral single-interaction precipitation coefficients are small and statistically insignificant, reflecting the limited identifying variation when interactions are considered in isolation.

Column 6 combines all interactions, including soil distance. Here the crossover pattern emerges sharply: the pastoral share $\times$ precipitation interaction is large, negative, and highly significant (-2.22), indicating strong precipitation matching among pastoralists, while the farmer share $\times$ precipitation interaction is also negative (-0.99) but smaller. For GDD,

[22] Standard errors are clustered at the gridcell-pair level.

the pastoral interaction remains positive and significant (+1.14); the farmer interaction is small and insignificant in this saturated specification, consistent with the split-sample pattern in which farmers match strongly on GDD relative to the omitted forager category but the pooled interaction loses statistical power. The interaction between soil distance and pastoralist share is positive and marginally significant, while the farmer interaction is positive but insignificant, indicating that soil matching is concentrated among foragers in this specification. This pattern differs from the split-sample estimates, where soil effects were stronger for farmers, and we therefore interpret these coefficients with caution. Taken together, however, the joint specification confirms that the dimension of ecological matching varies systematically with subsistence technology.

B.4 Robustness of Adaptive Sorting Results

Figure 8 (Section 6.3) documents a negative and statistically significant relationship between temperature changes at the origin and the temperature difference between destination and origin locations, indicating that individuals exposed to warming tend to relocate toward cooler destinations, and vice versa. Table B11 shows that this relationship is robust to estimating a more demanding individual-level specification that includes additional controls—HMI travel time, GDD distance, precipitation distance, soil composite distance, and contemporaneous precipitation differences—as well as origin, destination, and era fixed effects.

As discussed in Section 3, the paleoclimate data are available at a spatial resolution of $2.5° \times 1.9°$. Accordingly, in this analysis we define both origin and destination at this coarser level, rather than the $1° \times 1°$ grid used elsewhere in the paper. In the most demanding specification (adding destination fixed effects and climate-distance controls), the coefficient on temperature change is -0.358. Column 1 reports standard errors clustered at the gridcell pair level (defined at the $2.5° \times 1.9°$ resolution). Columns 2–4 replicate the specification using alternative clustering approaches: ancestor-level, paleoclimate cell $\times$ era, and paleoclimate-cell clustering, which groups all pairs sharing the same origin paleoclimate cell across eras. Reassuringly, statistical significance is preserved across all specifications.

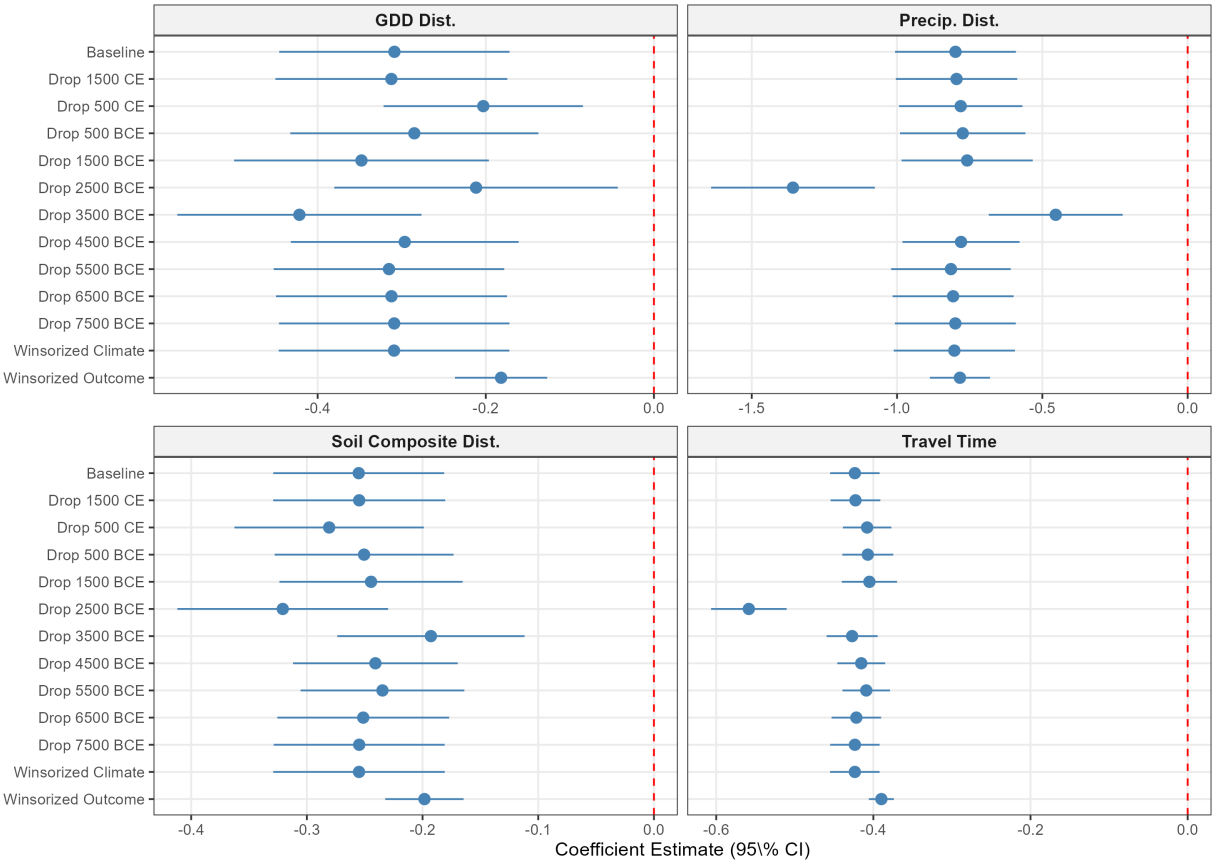
B.5 Suggestive Evidence of Precipitation Adaptation Using an Isotopic Hydroclimate Variable

Figure B3 replicates the adaptation binscatter (Figure 8) using $\Delta\delta^{18}O_p$ at origin (per mil) and the modern precipitation difference between destination and origin (TerraClimate, mm/yr) as the outcome. We caution that $\delta^{18}O_p$ in the Osman LGMR is iCESM model output

conditioned on SST-proxy assimilation, rather than directly constrained by $\delta^{18}O_p$ proxy records; it integrates temperature, precipitation amount, moisture-source, seasonality, and continental-recycling effects. The unconditional and FE-residualized panels show the same negative gradient observed in the temperature binscatter. The sign is consistent with an amount-effect interpretation, but the multiple channels embedded in $\delta^{18}O_p$ preclude a clean precipitation-amount inference from this figure alone.

B.6 Figures and Tables

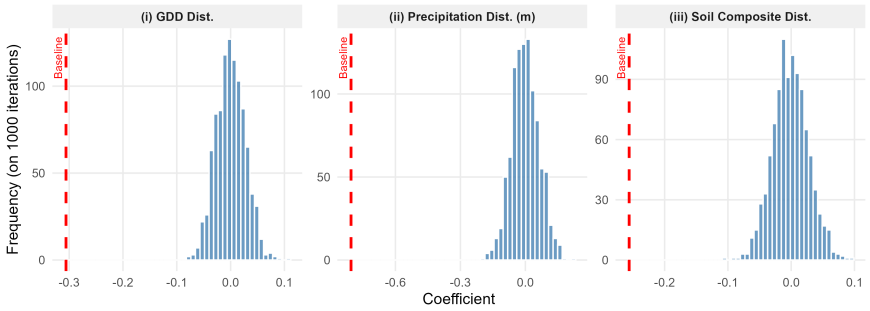
Figure B1: Robustness to excluding Time Periods



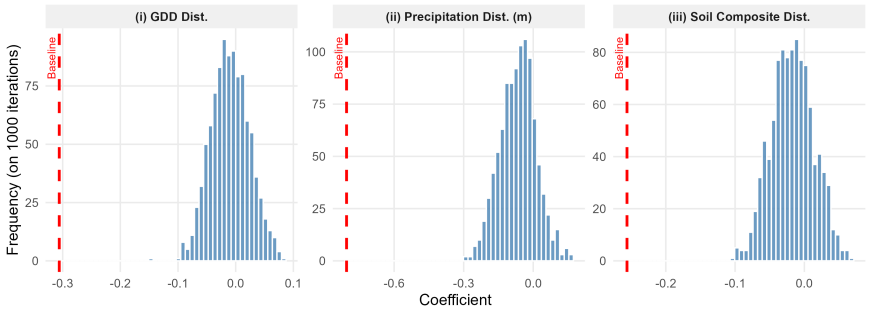
Notes: Each panel plots the coefficients, together with 95% confidence intervals, from PPML regressions that replicate the baseline model while dropping each 1,000-year period in turn (labeled by the bin's center in BCE/CE) and while winsorising the top 1% of the observations.

Figure B2: Alternative Inference on Migration Flows

(a) Unconstrained moves

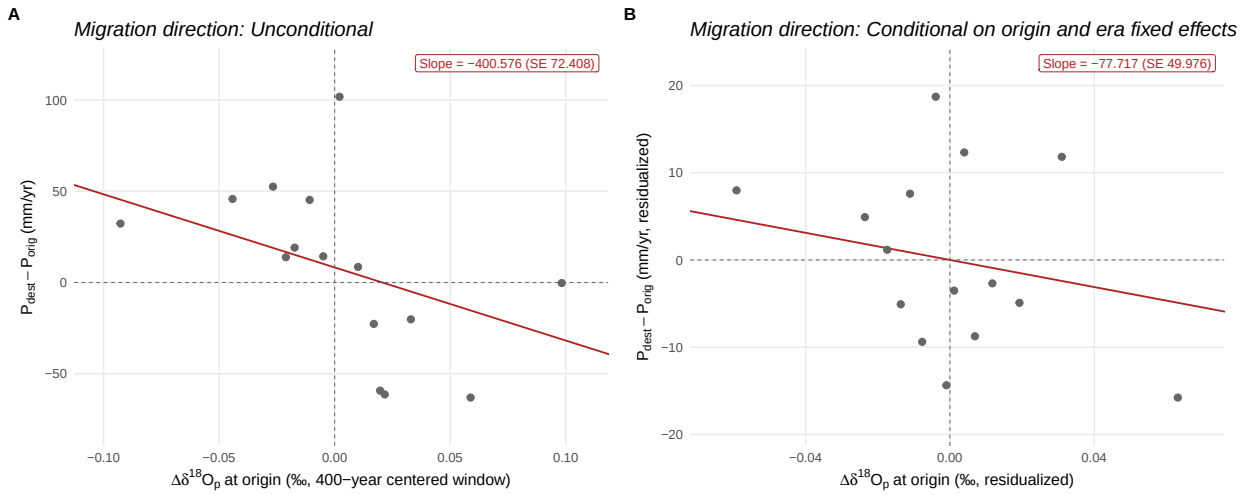


(b) Climate buffer-constrained moves



Notes: Each panel shows 3 figures plotting the distribution of coefficients from PPML regressions (blue) where we estimate the baseline model out of 1000 simulations with randomly assigned climate and soil distances, together with the estimated coefficient from Table 2 (dashed red line). In subfigure (a), we allow migrants to match with any climate, while in (b) we restrict moves to "buffers" of climate distances.

Figure B3: Adaptive Migration: Origin $\delta^{18}O_p$ Shock and Spatial Direction of Movement (Precipitation)



Notes: Individual-level data of $N = 71{,}290$ ancestor-descendant pairs (one observation per IBD-positive pair sharing $\geq$8 cM of segments, separated by more than 50 km, with 100–500 year burial gaps). The x-axis measures the 400-year centered change in $\delta^{18}O$ of precipitation at the origin from the Osman et al. (2021) reconstruction ($\delta^{18}O_{p,t-200} - \delta^{18}O_{p,t+200}$, where t is the ancestor's date; positive values indicate a shift toward less-negative isotopic composition). The y-axis is the modern precipitation difference between destination and origin gridcells (TerraClimate, mm/yr). $\delta^{18}O_p$ in the Osman LGMR is iCESM model output conditioned on SST-proxy assimilation rather than directly constrained by $\delta^{18}O_p$ proxy records; it integrates temperature, precipitation amount, moisture-source, seasonality, and continental-recycling effects. Panel A reports the unconditional binscatter; Panel B residualizes both axes on $2.5° \times 1.9°$ origin paleoclimate-cell and 500-year era fixed effects (Frisch-Waugh-Lovell). Each point is the mean within one of 15 equal-N bins of the x-axis variable. The line shows a linear fit to the underlying microdata, with standard errors clustered at the paleoclimate cell pair level.

Table B1: Within-LGMR (paleoclimate) variance decomposition across origin-destination pairs

Interval (yr BP)	$N_{\text{pair-era}}$	SD ΔT^{paleo} (°C)	SD ε^* (°C)	Variance captured (%)
0–1000	39,903	9.33	0.32	99.88
1000–2000	129,071	8.81	0.43	99.76
2000–3000	94,828	9.28	0.50	99.72
3000–4000	86,052	9.57	0.53	99.69
4000–5000	93,503	7.76	0.63	99.34
5000–6000	47,061	7.21	0.66	99.16
6000–7000	19,115	6.86	0.62	99.19
7000–8000	17,195	7.70	0.61	99.38
8000–9000	3,982	8.58	0.73	99.27
9000–10000	427	6.91	1.10	97.46
Pooled	531,137	8.67	0.53	99.62

Table B2: Climate Matching: Additional Environmental Controls

Dependent Variable:	Genetic Flow					
	Baseline	+Diurnal	+Elevation	+Coastal	+UV	+Lat/Lon
Model:	(1)	(2)	(3)	(4)	(5)	(6)
GDD Dist.	-0.309***	-0.286***	-0.263***	-0.268***	-0.020	-0.023
	(0.070)	(0.069)	(0.067)	(0.067)	(0.066)	(0.066)
Precip. Dist.	-0.799***	-0.735***	-0.753***	-0.762***	-0.753***	-0.712***
	(0.106)	(0.104)	(0.103)	(0.103)	(0.101)	(0.102)
Soil Composite Dist.	-0.255***	-0.191***	-0.196***	-0.192***	-0.098**	-0.111**
	(0.038)	(0.041)	(0.041)	(0.041)	(0.043)	(0.046)
Travel Time	-0.424***	-0.393***	-0.384***	-0.387***	-0.425***	-0.183**
	(0.016)	(0.018)	(0.018)	(0.018)	(0.018)	(0.079)
Diurnal Range Dist.		-0.097***	-0.087***	-0.073***	0.007	0.005
		(0.015)	(0.017)	(0.018)	(0.017)	(0.018)
Elevation Dist.			-0.153**	-0.154**	0.083	0.105*
			(0.067)	(0.068)	(0.052)	(0.055)
Coast to Non-coast Move				-0.155***	-0.235***	-0.237***
				(0.041)	(0.041)	(0.041)
UV-B Dist. (/100)					-0.086***	-0.086***
					(0.005)	(0.009)
Latitude Dist. (km)						-0.025
						(0.089)
Longitude Dist. (km)						-0.198***
						(0.060)
Dep. var. mean	0.149	0.149	0.149	0.149	0.149	0.149
Fit statistics						
Observations	478,034	478,034	478,034	478,034	478,034	478,034
Pseudo R^2	0.480	0.481	0.481	0.482	0.486	0.486

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B3: Climate Matching: Soil Components

Dependent Variable:	Genetic Flow							
	Base	Index	pH	Water	Organic	Clay	Bulk	All
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDD Dist.	-0.309***	-0.438***	-0.345***	-0.479***	-0.389***	-0.400***	-0.459***	-0.289***
	(0.070)	(0.056)	(0.060)	(0.056)	(0.059)	(0.068)	(0.059)	(0.068)
Precip. Dist.	-0.799***	-0.715***	-0.633***	-0.768***	-0.749***	-0.828***	-0.792***	-0.622***
	(0.106)	(0.110)	(0.117)	(0.108)	(0.105)	(0.109)	(0.107)	(0.117)
Soil Composite Dist.	-0.255***							
	(0.038)							
Soil Index (PCA) Dist.		-0.104***						
		(0.035)						
Soil pH Dist.			-0.184***					-0.138**
			(0.047)					(0.055)
Soil Water Dist.				-0.262***				-0.234**
				(0.093)				(0.103)
Soil Organic Dist.					-0.338***			-0.382***
					(0.045)			(0.049)
Soil Clay Dist.						-0.138***		-0.055
						(0.040)		(0.048)
Soil Bulk Dist.							-0.071*	0.164***
							(0.042)	(0.048)
Dep. var. mean	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149
Fit statistics								
Observations	478,034	478,034	478,034	478,034	478,034	478,034	478,034	478,034
Pseudo R^2	0.480	0.480	0.480	0.480	0.481	0.480	0.480	0.481

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. Column 2 uses a composite soil index (first principal component) including all the subsequent individual characteristics; columns 3–7 use individual soil characteristics. Column 8 includes all individual characteristics together. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B4: Climate Matching: Sample Robustness

Dependent Variable:	Genetic Flow					
	Baseline	Horizontal	Vertical	≥100km	≥200km	≥300km
Model:	(1)	(2)	(3)	(4)	(5)	(6)
GDD Dist.	-0.309***	-0.405***	-0.221**	-0.238***	-0.195***	-0.174***
	(0.070)	(0.087)	(0.094)	(0.058)	(0.057)	(0.057)
Precip. Dist.	-0.799***	-0.865***	-0.044	-0.783***	-0.797***	-0.851***
	(0.106)	(0.134)	(0.180)	(0.105)	(0.108)	(0.111)
Soil Composite Dist.	-0.255***	-0.061	-0.394***	-0.274***	-0.235***	-0.182***
	(0.038)	(0.048)	(0.053)	(0.036)	(0.035)	(0.036)
Travel Time	-0.424***	-0.429***	-1.14***	-0.422***	-0.411***	-0.397***
	(0.016)	(0.020)	(0.114)	(0.016)	(0.016)	(0.016)
Dep. var. mean	0.149	0.143	0.215	0.146	0.138	0.132
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	478,034	354,601	96,076	474,187	468,282	460,151
Pseudo R^2	0.480	0.501	0.481	0.479	0.482	0.482

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments ≥ 8 cM between 1° gridcell pairs in a 500-year era. 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, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B5: Climate Matching: Distance Robustness

Dependent Variable:	Genetic Flow			
	HMI (Baseline)	Haversine	Haversine Quad.	Log Haversine
Model:	(1)	(2)	(3)	(4)
GDD Dist.	-0.309***	-0.260***	-0.176**	-0.102
	(0.070)	(0.068)	(0.070)	(0.063)
Precip. Dist.	-0.799***	-0.738***	-0.736***	-0.862***
	(0.106)	(0.107)	(0.106)	(0.102)
Soil Composite Dist.	-0.255***	-0.216***	-0.154***	-0.045
	(0.038)	(0.038)	(0.038)	(0.040)
Travel Time	-0.424***			
	(0.016)			
Distance (km)		-0.384***	-0.687***	
		(0.015)	(0.043)	
Distance (km)-sq			0.055***	
			(0.006)	
log(Distance)				-0.583***
				(0.025)
Dep. var. mean	0.149	0.149	0.149	0.149
Fixed-effects				
Orig. gridcell	Yes	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes	Yes
Era	Yes	Yes	Yes	Yes
Fit statistics				
Observations	478,034	478,034	478,034	478,034
Pseudo R^2	0.480	0.481	0.483	0.484

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. Column 1 uses linear HMI least-cost-path travel time (in thousands of hours) from the Human Mobility Index (Özak, 2018) (baseline). Columns 2–4 use Haversine distance in linear, quadratic, and natural log forms as robustness checks. All specifications include origin gridcell, destination gridcell, and era fixed effects. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B6: Climate Matching: Specification Robustness

Dependent Variables:	Genetic Flow	Genetic Flow (12cM)	Genetic Flow (16cM)
	Baseline	12cM Flow	16cM Flow
Model:	(1)	(2)	(3)
GDD Dist.	-0.309***	-0.210**	-0.180
	(0.070)	(0.094)	(0.128)
Precip. Dist.	-0.799***	-1.51***	-1.88***
	(0.106)	(0.167)	(0.234)
Soil Composite Dist.	-0.255***	-0.360***	-0.451***
	(0.038)	(0.053)	(0.070)
Travel Time	-0.424***	-0.517***	-0.659***
	(0.016)	(0.022)	(0.031)
Dep. var. mean	0.149	0.046	0.022
Fixed-effects			
Orig. gridcell	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes
Era	Yes	Yes	Yes
Fit statistics			
Observations	478,034	334,880	217,129
Pseudo R^2	0.480	0.454	0.409

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments between 1° gridcell pairs, using linkage thresholds of at least 8cM (column 1), 12cM (column 2) and 16cM (column 3). All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B7: Environmental Distance and Migration (Individual)

Dependent Variable:	Genetic Flow					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
GDD Dist.	-1.09***	-1.02***	-0.249***	-0.234***	-0.223***	-0.211***
	(0.060)	(0.037)	(0.051)	(0.049)	(0.044)	(0.042)
Precip. Dist.	-0.493***	-0.570***	-0.921***	-0.881***	-0.924***	-0.929***
	(0.137)	(0.085)	(0.084)	(0.075)	(0.071)	(0.069)
Soil Composite Dist.	-0.185***	-0.079***	-0.323***	-0.300***	-0.289***	-0.279***
	(0.037)	(0.022)	(0.030)	(0.028)	(0.027)	(0.025)
Travel Time	-0.083***	-0.244***	-0.371***	-0.388***	-0.386***	-0.384***
	(0.019)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
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	3,128,079	3,128,079	3,050,584	3,050,584	2,986,138	2,927,429
Pseudo R^2	0.035	0.209	0.285	0.319	0.333	0.342

Notes: The sample consists of potentially linked pairs of ancient individuals in Western Eurasia over the last 10,000 years. The dependent variable is an indicator for whether the pair shares at least one IBD segment $\geq$ 8 cM. Soil Composite Dist. is the Euclidean distance on standardized clay content and soil organic carbon. Column 4 is the preferred specification. Standard errors, reported in parentheses, are clustered at the 1° gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B8: Inference Robustness: Alternative Standard Error Treatments

Dependent Variable:	Genetic Flow		
	Pair (1°)	Multi-way (1°)	Origin (1°)
Model:	(1)	(2)	(3)
GDD Dist.	-0.309***	-0.309**	-0.309***
	(0.070)	(0.143)	(0.112)
Precip. Dist.	-0.799***	-0.799***	-0.799***
	(0.106)	(0.256)	(0.205)
Soil Composite Dist.	-0.255***	-0.255***	-0.255***
	(0.038)	(0.066)	(0.052)
Travel Time	-0.424***	-0.424***	-0.424***
	(0.016)	(0.049)	(0.037)
Fixed-effects			
Orig. gridcell	Yes	Yes	Yes
Dest. gridcell	Yes	Yes	Yes
Era	Yes	Yes	Yes
Fit statistics			
Observations	478,034	478,034	478,034
Pseudo R^2	0.48	0.48	0.48

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. All columns estimate the same preferred specification: GDD distance, precipitation distance (TerraClimate), soil composite distance (clay + organic C), and HMI travel time. Column 1: pair-clustered SEs at 1-degree (primary). Column 2: multi-way clustered on origin + destination at 1-degree. Column 3: clustered at the 1-degree origin. All specifications include origin gridcell, destination gridcell, and era fixed effects. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B9: Climate Matching by Subsistence Type (Individual)

Dependent Variable:	Genetic Flow				
	All	Farmer	Agro-Past.	Pastoral	Forager
Model:	(1)	(2)	(3)	(4)	(5)
GDD Dist.	-0.234***	-0.504***	-0.515***	-0.050	-0.264
	(0.049)	(0.047)	(0.081)	(0.046)	(0.285)
Precip. Dist.	-0.881***	0.019	-0.294**	-1.25***	-0.385
	(0.075)	(0.084)	(0.140)	(0.098)	(0.493)
Soil Composite Dist.	-0.300***	-0.212***	-0.201***	-0.082***	-0.370***
	(0.028)	(0.032)	(0.039)	(0.030)	(0.121)
Travel Time	-0.388***	-1.02***	-0.578***	-0.154***	-1.68***
	(0.012)	(0.034)	(0.026)	(0.009)	(0.128)
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	3,050,584	1,627,535	166,058	1,073,556	36,246
Pseudo R^2	0.319	0.271	0.351	0.401	0.388

Notes: The dependent variable (Genetic Flow) is the count of IBD segments $\geq$ 8 cM between individual pairs. Each column restricts to origin individuals classified as the indicated subsistence type (archaeological culture group classification). Soil Composite Dist. is Euclidean on standardized clay and organic carbon. All specifications include origin gridcell, destination gridcell, and era fixed effects (defined at the 1° gridcell). Standard errors, reported in parentheses, are clustered at the 1° gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B10: Climate Matching Heterogeneity by Subsistence Type

Dependent Variable:	Genetic Flow					
	Baseline	GDD x Farm	GDD x Past	Prec x Farm	Prec x Past	Joint
Model:	(1)	(2)	(3)	(4)	(5)	(6)
GDD Dist.	-0.309***	0.221**	-1.19***	-0.309***	-0.309***	-0.960***
	(0.070)	(0.086)	(0.113)	(0.069)	(0.069)	(0.276)
Precip. Dist.	-0.799***	-0.690***	-0.694***	-0.717***	-0.842***	1.07**
	(0.106)	(0.103)	(0.103)	(0.168)	(0.181)	(0.432)
Soil Composite Dist.	-0.255***	-0.269***	-0.249***	-0.271***	-0.257***	-0.494***
	(0.038)	(0.037)	(0.037)	(0.037)	(0.037)	(0.157)
Farmer Share × GDD Dist.		-1.42***				-0.035
		(0.159)				(0.311)
Pastoral Share × GDD Dist.			1.62***			1.14***
			(0.161)			(0.301)
Farmer Share × Precip. Dist.				-0.179		-0.994**
				(0.302)		(0.490)
Pastoral Share × Precip. Dist.					0.077	-2.22***
					(0.304)	(0.478)
Farmer Share × Soil Dist.						0.240
						(0.183)
Pastoral Share × Soil Dist.						0.323*
						(0.172)
Farmer Share		-1.03***		-1.56***		-0.910***
		(0.155)		(0.147)		(0.281)
Pastoral Share			1.95***		2.59***	0.447
			(0.190)		(0.189)	(0.293)
Fit statistics						
Observations	478,034	478,034	478,034	478,034	478,034	478,034
Pseudo R^2	0.48	0.49	0.49	0.48	0.49	0.50

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. The dependent variable (Genetic Flow) is the count of IBD segments ≥ 8 cM between 1° gridcell pairs in a 500-year era. Farmer Share and Pastoral Share are origin-gridcell shares classified by archaeological culture group; forager is the omitted reference group. Columns 2–3 interact GDD distance with farmer and pastoral shares; columns 4–5 interact precipitation distance similarly; column 6 interacts all three climate distances (GDD, precipitation, soil composite) jointly with farmer and pastoral shares. All specifications include origin gridcell, destination gridcell, and era fixed effects. Standard errors, reported in parentheses, are clustered at the gridcell pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B11: Adaptive Migration: Origin Climate Shocks and Destination Temperature (Individual-Level)

Dependent Variable:	Temp. Diff. (dest − orig)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Temp. Change (origin)	-1.12***	-1.12***	-1.12***	-0.358***	-0.358***	-0.358***
	(0.351)	(0.241)	(0.331)	(0.089)	(0.112)	(0.139)
Era FE (500-yr)	Yes	Yes	Yes	Yes	Yes	Yes
Origin gridcell FE	2.5°	2.5°	2.5°	2.5°	2.5°	2.5°
Dest. gridcell FE	—	—	—	2.5°	2.5°	2.5°
Cluster SE	Pair	Ancestor	Cell × Era	Pair	Ancestor	Cell × Era
Num. clusters	9,500	5,665	1,192	9,500	5,665	1,192
Fit statistics						
Observations	71,269	71,269	71,269	71,251	71,251	71,251
Adjusted R^2	0.468	0.468	0.468	0.965	0.965	0.965

Notes: OLS estimates. Each observation is a pair of ancient individuals linked by at least one IBD segment $\geq$ 8 cM, whose dates differ by 100–500 years, both dated after 8000 BCE, and whose origin gridcells lie more than 50 km apart. The dependent variable is the modern temperature difference between destination and origin gridcells ($T_{\text{dest}} - T_{\text{orig}}$). Temp. Change (origin) is the 400-year centered paleoclimate change at the earlier individual's origin ($T_{t-200} - T_{t+200}$). Columns 1–3 use origin and era fixed effects defined at the 2.5° × 1.9° paleoclimate cell—the native spatial resolution of the paleoclimate reconstruction—and match the binscatter reported in Figure 8. Columns 4–6 add a destination paleoclimate-cell fixed effect to absorb time-invariant characteristics of where descendants ended up. Within each block, the three columns vary the clustering of standard errors: gridcell pair clustering (Cols. 1, 4); ancestor-level clustering, which groups all IBD-linked pairs sharing the same earlier individual (Cols. 2, 5); paleoclimate cell × 500-year era clustering (Cols. 3, 6), which matches the era fixed effect and fully spans the 400-year window over which the shock is computed. We refer to the 2.5° × 1.9° paleoclimate cell as 2.5° for brevity. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B12: Adaptive Migration: Robustness to Alternative Clustering

Dependent Variable:	Genetic Flow					
	1° Specification			2.5° Osman Specification		
	Pair (1°)	Origin (1°)	1° × Era	Pair (2.5°)	Origin (2.5°)	2.5° × Era
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Temp. Change (origin)	1.01***	1.01*	1.01	0.979***	0.979*	0.979
	(0.195)	(0.581)	(0.921)	(0.237)	(0.567)	(0.947)
Temp. Change × Temp. Diff.	-0.070***	-0.070	-0.070	-0.104***	-0.104**	-0.104
	(0.026)	(0.052)	(0.073)	(0.026)	(0.050)	(0.083)
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	478,034	478,034	478,034	495,898	495,898	495,898
Pseudo R^2	0.48	0.48	0.48	0.39	0.39	0.39

Notes: The sample includes origin–destination gridcell pairs by era across Western Eurasia over the past 10,000 years, restricted to cells with at least one IBD-linked pair. Columns 1–3 replicate the saturated specification of Table 5 column 4, with 1° origin, destination, and era fixed effects. Columns 4–6 replicate column 5 of Table 5, with origin and destination fixed effects defined at the 2.5° Osman LGMR paleoclimate cell (1.9° latitude by 2.5° longitude)—the native spatial resolution of the temperature shock ΔT_{ot} and the resolution used in the individual-level adaptation analysis (Table B11, Figure 8). Within each block, the three columns vary the clustering: pair-level (the paper's primary), origin-level, and two-way origin × 500-year era. All columns include the same controls (GDD distance, precipitation distance, soil composite distance, HMI travel time). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.