Climate and Prehistoric Migration*

Peter Huybers Marco Tabellini Charles A. Taylor Francesco Toti

March 2025

Abstract

What is the relationship between climate and migration over the very long run? We merge gridded paleoclimate data with a recently-assembled dataset from ancient human DNA extracted from archaeological sites that predicts the location of an individual's ancestral population. This allows us to trace population-level mobility in Western Eurasia over the past 10,000 years. We find that climate similarity predicts migration flows. This relationship is robust to controlling for distance and bio-physical features, and does not merely reflect the spatial correlation of climate. The effect of climate similarity on migration is strongest between 3300BCE and 2000BCE—a period coinciding with large Bronze Age population movements into Western and Central Europe. We also find preliminary evidence of population movements as an adaptive response to long-run climate change, and that climate similarity influenced the spread of agriculture between regions.

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

1 Introduction

How did climate patterns shape prehistoric human migration? This paper investigates whether climate similarity between regions influenced where early humans moved and, as a result, how innovation spread. Motivated by long-standing questions about the role of geography on long-run development, we empirically examine the hypothesis that populations migrate and exchange technologies along climatically-similar corridors (Crosby, 1986). For example, the relatively uniform east–west transects of Eurasia are thought to have enabled the diffusion of people, crops, and inventions, whereas the north–south orientation of the Americas and Africa—with its stark climatic gradients—slowed the spread of innovation (Diamond, 1997). Understanding the extent to which "climate matching" influenced early migration can shed light on why certain societies experienced faster adoption of agriculture and technology, and why cultural evolution diverged across continents.

Recent research suggests that climate matching plays a major role in shaping migration patterns in the modern era. Obolensky et al. (2024) find that climate matching has been a consistent factor driving internal and international migration in the US from the 19$^{\text{th}}$ century to the present day. They also find that there are consequences of climate mismatch in the form of negative health and economic outcomes. Climate-specific human capital is a prominent mechanism driving climate matching (see also Michalopoulos, 2012) and the negative effects of mismatch. That is, individuals have a set of climate specific skills and knowledge whose transferability declines with climate distance (e.g., pastoralists from a warm climate will not know how to keep their livestock alive in a blizzard).

Our paper extends this framework into prehistory, examining whether climate similarity shaped migration in an era where mobility was likely dictated more by environmental constraints than institutional factors. We confirm that prehistoric human populations exhibited the same tendency to move into climatically-familiar regions. This suggests that climate-driven migration patterns may have deeper evolutionary and behavioral roots, reinforcing the idea that environmental familiarity is an important factor in human success.

To get at these conclusions, we leverage recently-available ancient DNA data and paleoclimate reconstructions. First, we find that climate similarity is a strong predictor of prehistoric migration patterns over the course of multiple centuries. That is, migration flows occur at higher rates between regions with more similar climates, where climate is measured in terms of temperature and rainfall regimes. This result holds after controlling for physical distance, geographic features, and various fixed effects.[1] Quantitatively, our estimates imply

[1] Ramachandran and Rosenberg (2011) use DNA data from modern populations to test whether continental axes explain observed genetic variation. In line with Diamond (1997), they find that distance in latitude explains more genetic variation than longitude across Eurasia, suggesting climatic factors at play.

that a 7.5°C reduction in temperature distance (or, the sample mean) increases bilateral migration flows 4.3%.

Next, we show that climate similarity had meaningful consequences for technology diffusion. For example, the initial transmission of agriculture from the Fertile Crescent into Europe was largely confined to areas around the Mediterranean with relatively similar climates—while it took much longer for crop domestication to spread into central and northern Europe. Our analysis confirms that climate distance is a robust predictor of the timing of the uptake of agriculture between regions.[2] This suggests that innovations were disseminated through migration networks shaped by geography and climate rather than arising in isolation or via the gradual spatial diffusion of ideas across groups.

In terms of economic history, our findings offer empirical support for the drivers of prehistoric migration flows along climate corridors that enabled knowledge to accumulate and diffuse. This contributes to a broader narrative that the"deep roots" of comparative development are intertwined with climate and migration—complementing other determinants like genetic diversity (Ashraf and Galor, 2013) and early state formation (Allen et al., 2023). Our results also imply that the climate-migration link is not just a modern phenomenon or a response to recent climate change, but has been a factor shaping human movement and cultural exchange for millennia.

2 Data and Sample Construction

2.1 Genetic Migration

To study ancient migration patterns, we rely on paleo-genetic data to infer a possible distribution of moves over time. Schmid and Schiffels (2023) use ancient DNA samples from burial sites in Europe and near Asia to predict a plausible distribution of genetic similarity over time and space, which, in turn, is used to predict past movements of individuals. This allows us to obtain a panel of individuals with information on their location at the time of death (i.e. where their remains were found), the likely year of death (from carbon dating), some demographic characteristics, such as age or gender, and the predicted locations of each individuals' ancestry 334, 667, and 1,000 years before their death. The sample consists of 3,138 individual ancient DNA samples from the Allen Ancient DNA Resource collected in Europe and near Asia.[3]

[2]The spread of agriculture has also been linked to climate volatility and seasonality (Ashraf and Michalopoulos, 2015; Matranga, 2024).

[3]The authors use Multidimensional Scaling (MDS), a dimensionality reduction method that projects complex genetic distance relationships from SNPs (Single Nucleotide Polymorphisms) into a lower-dimensional

Figure 1 plots the count of observations (i.e., individuals for whom we have information on genetic migration flows) during each 1,000 year period between 8,000BCE and 0 based on year of death, with a large share of observations dating to the Bronze Age between 3500BCE and 1500BCE. Figure 2 maps the locations of the ancestry and descendants' locations of individuals across two major periods: 8,000-3,300BCE; and 3,300-0BCE.

2.2 Climate

Modern climate Modern climate datasets offer the advantage of higher spatial and within-year resolution. As such, we first proxy a location's climate with TerraClimate (Abatzoglou et al., 2018). We compute average annual temperature and precipitation for each $1^\circ \times 1^\circ$ grid cell over the period between 1960 and 2000. Figure 3 shows the average temperature over 1960-2000 for the grid cells in our sample.

Paleoclimate To exploit time variation in climate, we use reconstructions of paleoclimate from two distinct sources: a marine-based one (Osman et al., 2021) and terrestrial-based one (Mauri et al., 2015).

For the first, we use NOAA/WDS Paleoclimatology's gridded reconstruction of annual mean surface air temperature in 200 year time steps since the last glacial maximum (Osman et al., 2021). Estimates are derived from marine geochemical proxies of sea surface temperature including oxygen isotopes in carbonate (d18Oc), Mg/Ca ratios, TEX86 (Tetraether Index), and UK'37 (Alkenone Unsaturation Index).[4] These proxies are combined with simulations from the Community Earth System Model (iCESM1.2 and iCESM1.3). The spatial resolution is 1.9° x 2.5° (latitude x longitude) and we aggregate it to two major periods of interest, namely 8,000 to 3,300BCE and 3,300 to 0CE.

In the bottom panel of Figure 3, we plot the temperature at an example point in time (4,950BCE), while in Figure 4 we show how average global temperature has evolved over time according to this reconstruction, along with its standard deviation. In terms of regional heterogeneity in trends, we define regions according to Table A3. In Figure 5, we show the

space. They then use a Gaussian Process Regression (GPR), an interpolation method that models data points as outcomes of a multivariate normal distribution (i.e., kriging). To define how the spatial and temporal relationships between points are modeled, the authors use an anisotropic Gaussian kernel to allow different behaviors in different directions.

[4]Oxygen isotopes in foraminifera shells (δ18Oc) reflect past sea surface temperatures and ice volume, with lower δ18O values indicating warmer seas. The magnesium-to-calcium (Mg/Ca) ratio in these shells also increases with rising temperatures. TEX86, based on lipids produced by marine archaea, provides another temperature proxy and is useful when foraminifera data are limited. UK'37, derived from alkenones produced by coccolithophores, tracks sea surface temperatures by measuring the degree of unsaturation in these molecules, which varies with water temperature.

evolution of temperature over across the five regions outlined above. The left panel shows the evolution of average temperature in each time point for each region, while the right panel shows the evolution of temperature over time, demeaned by the average temperature in each period. In line with the scientific understanding, northern latitude locations (Nordic Europe) experienced relatively greater warming from 8,000 to 4,000BCE, a period categorized by overall warming.

The second source of paleo-climate reconstructions, which we use as an alternative for assessing robustness, is the European Holocene Pollen-based Gridded Climate Reconstruction (Mauri et al., 2015). It consists of gridded climate reconstruction for Europe for the last 12,000 years drawn from pollen data. The reconstruction is an update of Davis et al. (2003) using the same methodology, but with a greatly expanded fossil and surface-sample dataset. The time resolution of this dataset is 1,000 years, the spatial resolution is 1°× 1°, and the climate is measured in terms of anomalies relative to the pre-industrial period (1850CE), i.e., a temperature of -2°C is 2°C cooler in a given gridcell and time period relative to 1850CE.

Soil and topography We obtain information on elevation and soil characteristics like soil pH, soil water holding capacity, and organic carbon (Hengl, 2018; Hengl and Wheeler, 2018; Hengl and Gupta, 2019). In Figure 6, we plot the grid cell-level correlation between temperature and precipitation and each of these soil and topographic variables.

2.3 Agriculture

We use two distinct data sources for agricultural data. The first is the data collected and organized in Matranga (2024), in particular information on the earliest evidence of agriculture in Eurasia. We plot the spatial distribution of the archaeological sites in Europe and Middle East for which we have information on agricultural presence and its timing in Figure 7.

The second comes from historical land-use reconstructions from Klein Goldewijk et al. (2017), a gridded data product at 5 arc min spatial resolution (0.08°) that spans the Holocene in 1,000 year intervals. The HYDE 3.3 product incorporates information on agricultural adoption derived from radiocarbon archaeological data. We present summary statistics in Table 4, showing the type of agriculture present at the origin grid cell at the individual level. We code an observation as *Other* if there is no cropland or grazing at the origin grid cell, as *Cropland* if there is more cropland than grazing, and as *Grazing* if there is more grazing than cropland.

2.4 Sample Construction

We consider locations with latitudes between 29° and 65° and longitudes between -10° and 65°. We exclude from the analysis areas overlapping with ocean or large water bodies. Using the modern TerraClimate dataset, we end up with 2,442 grid cells with non-missing climate information, each containing a unique longitude-latitude combination. Our sample consists of all the possible pairs of these cells (i.e., a matrix of all grid cell combinations).

We then match each individual in the genetic migration dataset from Schmid and Schiffels (2023) to the location of the closest grid cell, for both their ancestral and descendant locations. We limit the sample to individuals who died in the period 8,000BCE - 0, as estimated from carbon dating. Throughout the paper, we consider as origin the cell corresponding to the (predicted) position of the individual's ancestry 667 years before the individual's death, and as destination the grid cell of death (i.e., where the remains were found).[5]

We use the spatially-granular modern TerraClimate data to compute climate distances. Our main regressors of interest are the absolute difference in mean annual temperature between the grid cell of ancestry location o and the cell of death (descendants' location) d, *Temperature Distance*, and the absolute difference in precipitation between cell o and cell d, *Precipitation Distance*:

$$\begin{aligned} \text{Temperature Distance}_{od} &= \mid \text{Temperature}_o - \text{Temperature}_d \mid \\ \text{Precipitation Distance}_{od} &= \mid \text{Precipitation}_o - \text{Precipitation}_d \mid \end{aligned} \tag{1}$$

We perform the same procedure to compute the o-d distances in terms of geographic distance, elevation, and soil characteristics. We present summary statistics in Table 1.

We note that climate distance can be defined in several ways beyond mean annual temperature and precipitation (e.g., min/max temperature, degree days, drought indices, seasonality measures). While we explore other climate statistics, for parsimony our baseline analysis focuses on temperature (and to a lesser extent precipitation). This choice is motivated in part by Obolensky et al. (2024), which finds that mean annual temperature is a strong and robust measure of climate that predicts migration in the modern US context. Further, climate statistics that incorporate additional information like within-year variation are often infeasible or unreliable in the context of the sparse inputs used for paleo climate reconstructions, as discussed below.

Paleo Climate We also employ paleo climate datasets for two reasons: first to account for that fact that modern climate is different than prehistoric climate, and second to allow

[5]The 667-year period is the suggested baseline value used in Schmid and Schiffels (2023).

us to exploit variation in climate over time during our prehistoric study period. When using the marine-based climate reconstruction, we start with 446 unique grid cells of 1.9° x 2.5° (latitude x longitude) and create a dataset with all possible combinations of cells containing the same set of information as above. With the pollen-based dataset we have 1,775 1° x 1° grid cells with non-missing climate information. Table A1 and Table A2 present summary statistics for each grid cell pair using marine proxies (Osman et al., 2021) and terrestrial proxies (Mauri et al., 2015).

Figure 2 maps the locations of the ancestry and descendants' locations of individuals across two major periods: 8,000-3,300BCE; and 3,300-0BCE. In Table A5, we show the number of individuals by direction of move (North-to-South, East-to-West, etc) across several periods of time. All move directions are well represented in all periods, except for North-to-South in the earliest years.

3 Climate Distance and Migration

3.1 Empirical Strategy

In our baseline analysis, we consider the entire period from 8,000BCE to 0. We count the number of migrants between any pair of cells and estimate a Pseudo Poisson Maximum Likelihood (PPML) of the form:[6]

$$M_{od} = \exp\{\beta_1 TempDist_{od} + \beta_2 PrecipDist_{od} + \beta_3 GeoDist_{od} + \beta_4 Soil_{od} + \alpha_d\}\varepsilon_{od} \qquad (2)$$

where M_{od} is the number of migrants over the entire period between origin grid cell o and origin cell d, $TempDist_{od}$ and $PrecipDist_{od}$ are temperature and precipitation distance defined above, and α_d refers to the descendants' location cell fixed effects. Standard errors are clustered at the descendants' location cell level.

When examining the relationship between climate distance and migration, a major concern is that climate is spatially correlated: areas with similar climates tend to be geographically close. Hence, our estimates from (2) may capture the role of geographic distance as opposed to climate distance. To address this concern, our baseline model controls for geographic distance between the ancestral location and the descendant one, $GeoDist_{od}$, in

[6]PPML models are widely used in international economics to estimate the effects of distance (typically, geographic distance) between countries on bilateral trade flows. An important advantage of this class of models is that they perform well when the data include a large number of "zeros" as is the case in our context. See Silva and Tenreyro (2006) for more details.

several ways. Our results are robust to controlling flexibly for geographic distance and accounting for the spatial correlation of climate in our standard error treatment. We discuss these and more robustness checks later in the paper.

Another potential concern is that migrants from places that were climatically dissimilar moved to areas that were attracting migrants from everywhere else—regardless of climate distance. For instance, over the course of centuries (which is the time span of migration in our context) people could move closer to major emerging urban centers, fresh-water basins, or areas with generally milder climate. A related issue is that our dataset may suffer from sample selection, such that the preservation and discovery of burial sites is related to a location's climate. In both cases, the inclusion of descendants' location cell fixed effects, α_d, should address these concerns. However, what we cannot rule out is that, after controlling for geographic distance and for descendants' location cell fixed effects, climate distance may be correlated with other variables that independently drive migration patterns. Later we seek to assuage this concern by examining the sensitivity of our estimates to the inclusion of different cell-pair-specific controls, such as soil characteristics and elevation.

3.2 Main Results

We present results from Equation (2) in Table 2. In column 1, we estimate a parsimonious regression that only includes climate distances. The coefficient on both temperature and precipitation distance is negative and precisely estimate. This indicates that migration flows are larger between origin-destination cells that have more similar climates. In column 2, we add geographic distance and geographic distance squared. The coefficient on temperature distance halves, but remains negative and statistically significant, while that on precipitation becomes positive (and precisely estimated). In column 3, we further include descendants' location cell fixed effects. Interestingly, although the coefficient on temperature distance remains similar—albeit smaller in absolute value—to that in column 2, the coefficient on precipitation distance becomes again negative.

In column 4, we estimate a very stringent specification that, in addition to controlling for descendants' location cell fixed effects, replaces geographic distance and geographic distance square with the distance in latitude and longitude. The coefficient on temperature distance falls, possibly reflecting the fact controlling for latitude distance absorbs part of the variation in temperature distance, but remains statistically significant. The coefficient on precipitation distance, instead, is not affected in either magnitude or precision.

Finally, in column 5, which presents the baseline specification, we replicate column 3 by adding as control a vector of distance in soil characteristics (pH, water, carbon and elevation).

The coefficient on temperature distance remains negative and statistically significant, while that on precipitation distance drops to zero and becomes imprecisely estimated. According to our estimates, increasing temperature distance by 7.5° C (about the sample mean) reduces migration by 4.3%. This is remarkably similar to the magnitude obtained by Obolensky et al. (2024) for migration (domestic and international) in the US between 1850 and 2020. The fact that the coefficient on precipitation distance is less stable than that on temperature distance is also consistent with findings in Obolensky et al. (2024).

To assess the robustness of our results to possible mis-specification in the way we measure climate, we reproduce our main empirical analysis using the two paleo climate reconstructions described in Section 2 (marine-based proxies and pollen-based proxies). We present results in Table A4. In column 1, we replicate our baseline specification. In columns 2 and 3, we show that results are robust to using the pollen and the marine climate datasets, respectively. Columns 4 and 5 replicate column 3 using the marine dataset separately for two periods (8,000BCE to 3,300BCE; 3,300BCE to 0).

Lastly, our results might be biased due to the inclusion of ancestral and descendant locations that lacked any pre-historical settlements. As such, our coefficient of interest might be capturing omitted characteristics of a location that are correlated with its climate and have an effect on migration decisions per se. For instance, the presence of game animals in a location might be directly linked with its climate and might also affect the presence of individuals on its own. To provide evidence on the robustness of our results to this issue, we present Table A6 where we replicate the main analysis by restricting the sample to locations with at least an individual (either as their ancestry location or descendant one); we present results in column 2. Next, in columns 3 and 4, we consider population density data from Klein Goldewijk et al. (2017) and replicate the main analysis by considering two sub-samples with only cells where the average population density is non-zero. Results are generally consistent with the main results. Lastly, in columns 7 and 8, we replicate the main analysis by including as additional control the distance in population density between the two cells. Also in this case, the coefficient in the 3,300BCE-0CE period is negative and statistically significant.

4 Climate Change and Adaptation Analysis

Given the time span of our genetic migration measure (667 years), we can test whether populations responded to multi-century periods of warming or cooling with offsetting migration. In other words, did people move to marginally cooler places when the overall climate was warming?

Temperature Change over Time For each individual, we use the paleoclimate data to calculate the temperature change over time, which is defined as the difference between the temperature at time of death at the origin cell (o), and the temperature at the origin cell 1,000 years before death, which we call temperature at birth for simplicity. Formally, this quantity is defined as:

$$\text{Temp Change Over Time}_o = Temp_o^{Death} - Temp_o^{birth} \tag{3}$$

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

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

$$\text{Temp Change Over Space}_{od} = Temp_d^{TerraClimate} - Temp_o^{TerraClimate} \tag{4}$$

We also include a similarly defined measure of local elevation change to account for this margin of adaptation.

We then regress our temperature change over time on variations in temperature and elevation across space. Results are in Table 3. In Column 1, we regress *Temp Change over Time* on *Temp Change Over Space*. In Column 2, we add descendant location cell fixed effects. In Column 3, we add as a regressor *Elevation Change over Space*. Column 4 replicates Column 3 with descendant location cell fixed effects.

In each case, we find a negative coefficient on temperature change over space (i.e., people move to cooler places during periods of warming). This suggests a potential adaptive response to climate change via migration.

5 The Diffusion of Agriculture[7]

Climate matching during pre-history, and in particular at the time of the Neolithic Revolution, could influence agricultural adoption. To this end, we study whether climate similarity impacted the diffusion of agriculture. In a set up similar to Matranga (2024), we first examine whether the presence of agriculture has been recorded earlier in areas that were climatically more similar to the first (presumed) locations of the agricultural revolution, namely sites in

[7]This section is currently in progress.

modern day Turkey and Iraq. Focusing on a sample of 749 archaeological sites in Europe and Middle East, we estimate the following model by OLS:

$$\text{Year of Adoption}_c = \beta_1 \text{TempDist}_c + \beta_2 \text{PrecDist}_c + X_c + \varepsilon_c \tag{5}$$

where the dependent variable, *Year of Adoption*, namely the year of the earliest presence of agricultural practice in site c, is regressed on temperature and precipitation distance from either Iraq or Turkey cradle, and a vector of controls including geographic and altitude distance, location cell baseline characteristics (temperature, precipitation and altitude) and distance in climate seasonality.

We present results in Table 5. Irrespective of whether we define Turkey or Iraq as initial locations of the agricultural revolution, we find a consistently negative coefficient on climate distance. That is, locations that with a lower climate distance adopted agriculture earlier. These results hold when controlling for geographic distance and its quadratic as well as for climate at origin (column 1). They also hold when controlling for seasonality (column 2). Together with our previous results, this evidence is consistent with the notion that climate matching in migration may have facilitated the spread of agriculture across space during the agricultural revolution—especially in light of the importance of climate-specific skills in farming.

Next, we replicate our main analysis on the relationship between migration and climate similarity, adding information on historical agricultural land use from Klein Goldewijk et al. (2017). We present results in Table 6. In column 1, we replicate our baseline specification. In column 2, we restrict the sample to cell-pairs with cropland at the descendant location. In column 3, we restrict the analysis to pairs with grazing at descendant location. In both cases, the coefficient on temperature distance remains negative and statistically significant. Interestingly, the magnitude is greater when restricting the sample to cropland.

In column 5, we ask whether agriculture and grazing reduced the strength of climate matching in migration. To this end, we interact climate distance with dummies for the presence of agriculture and grazing at descendant locations.[8] Formally, our model becomes:

$$\begin{aligned} M_{od} = \exp\{&\text{ClimaDist}_{od} + X_{od} + 1[Grazing]_d + 1[Cropland]_d + \\ &+ (1[Grazing]_d + 1[Cropland]_d) \times \text{ClimaDist}_{od}\}\varepsilon_{od} \end{aligned} \tag{6}$$

Results in column 5 indicate that agriculture at destination altered the gradient of climate matching in migration. The positive coefficient on the interaction between temperature distance and the presence of cropland or grazing in column 5 suggests that agriculture may

[8] In all such specifications, we include dummies for presence of cropland and grazing at that location.

have lowered the extent of climate matching in migration.

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.

Allen, Robert C, Mattia C Bertazzini, and Leander Heldring (2023) "The economic origins of government," *American Economic Review*, 113 (10), 2507–2545.

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.

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.

Hengl, Tomislav (2018) "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.

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

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.

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

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

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-.

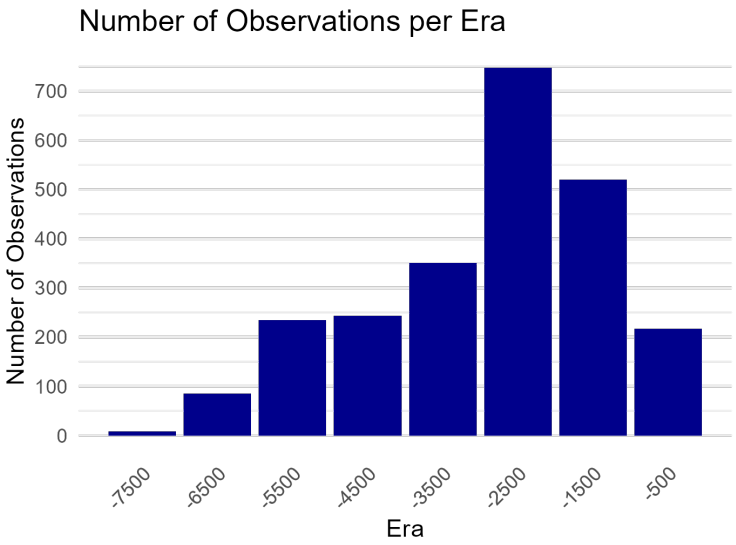
Ramachandran, Sohini and Noah A Rosenberg (2011) "A test of the influence of continental axes of orientation on patterns of human gene flow," *American journal of physical anthropology*, 146 (4), 515–529.

Schmid, Clemens and Stephan Schiffels (2023) "Estimating human mobility in Holocene Western Eurasia with large-scale ancient genomic data," *Proceedings of the National Academy of Sciences*, 120 (9), e2218375120, 10.1073/pnas.2218375120.

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

Figures

Figure 1: Number of Migrants per 1,000 Years Era

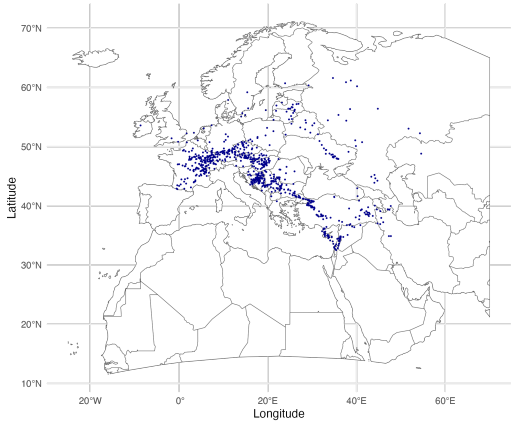


Notes: The figure plots the number of migrants present in the Schmid and Schiffels (2023) paleo-genetic dataset for each 1,000 year period between 8,000BCE and 0.

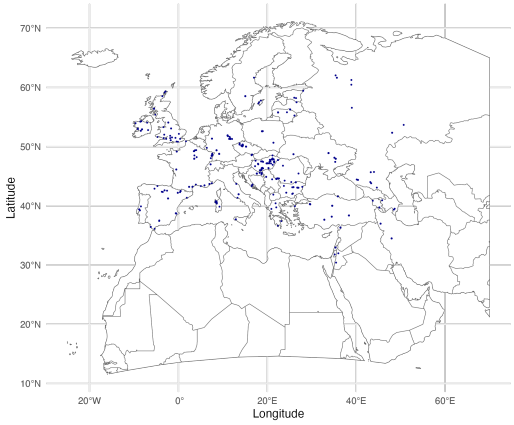
Figure 2: Ancestral and Descendant Locations across Time Periods

(a) 8,000BCE – 3,300BCE

Ancestry

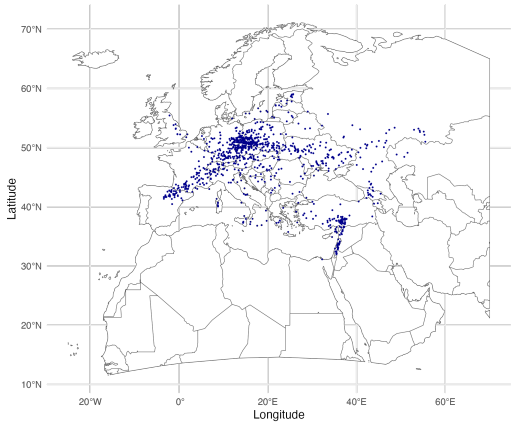


Descendant



(b) 3,300BCE – 0CE

Ancestry



Descendant

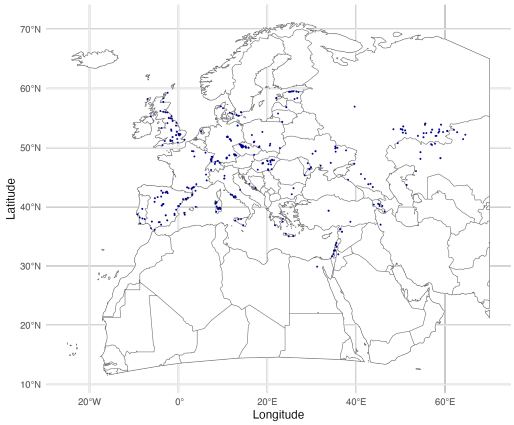
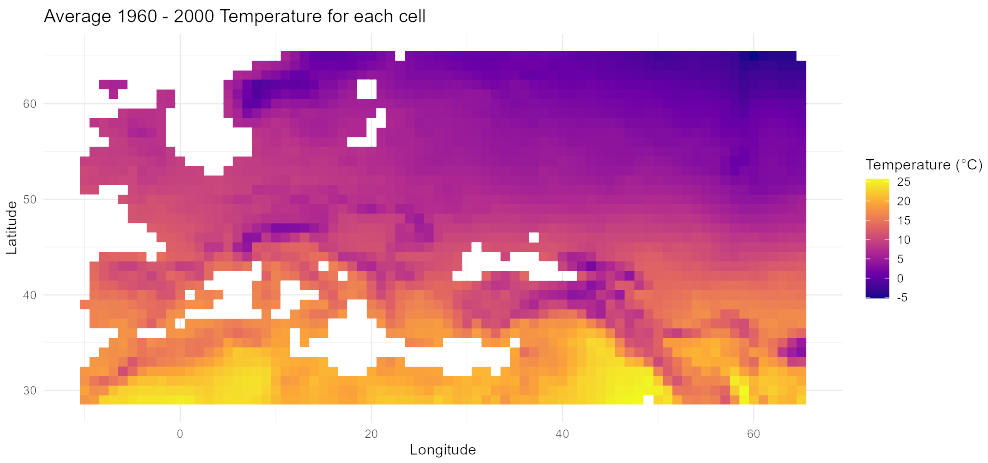


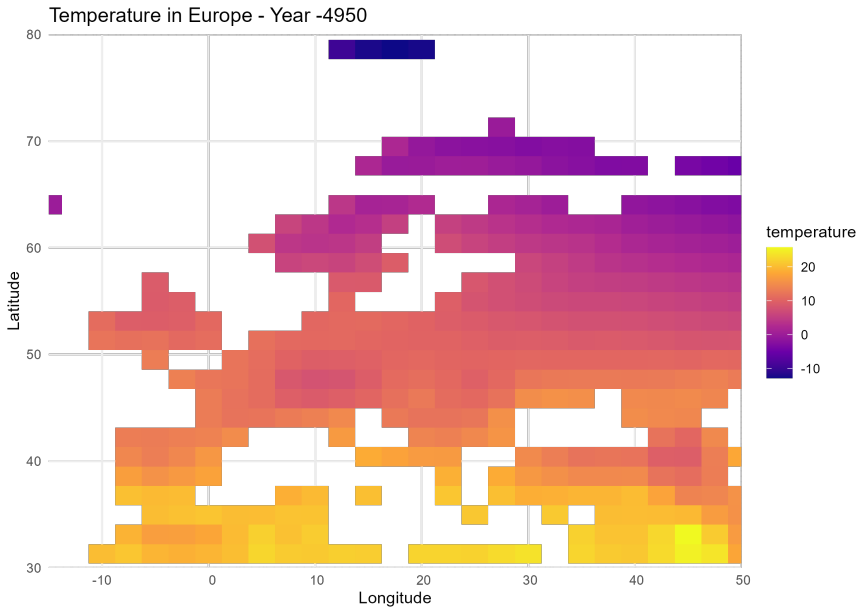
Figure 3: Measures of Average Temperature

Panel A: TerraClimate, 1960-2000



Notes: The figure plots the spatial distribution of average temperature at 1°resolution using temperature data from Abatzoglou et al. (2018) averaged over 1960-2000.

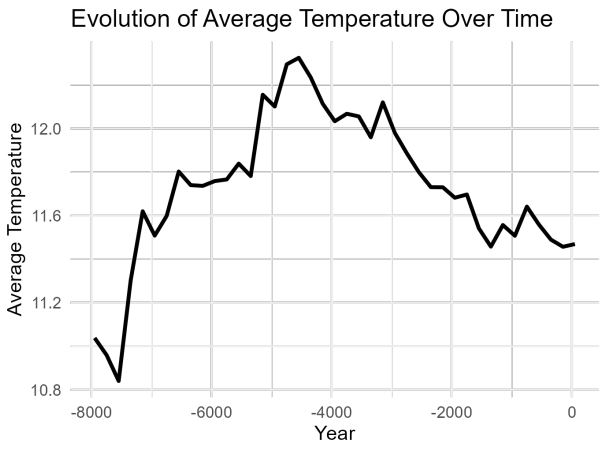
Panel A: Marine Proxy Paleoclimate Reconstructions



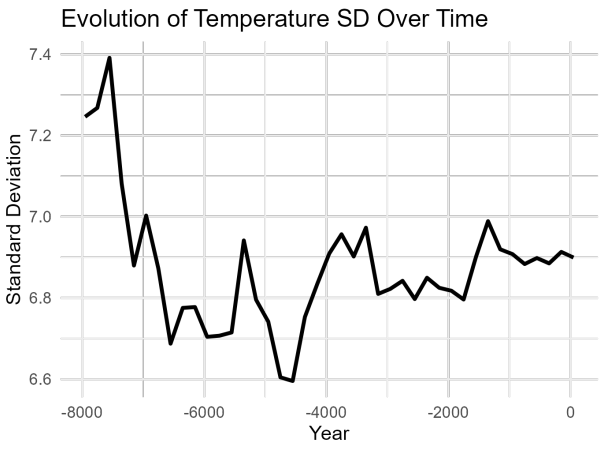
Notes: The figure plots the reconstructed average temperature in 4950BCE from the marine-based paleoclimate dataset (Osman et al., 2021).

Figure 4: Evolution of Temperature over Time

(a) Temperature - Mean



(b) Temperature - SD



(c) Temperature - Mean

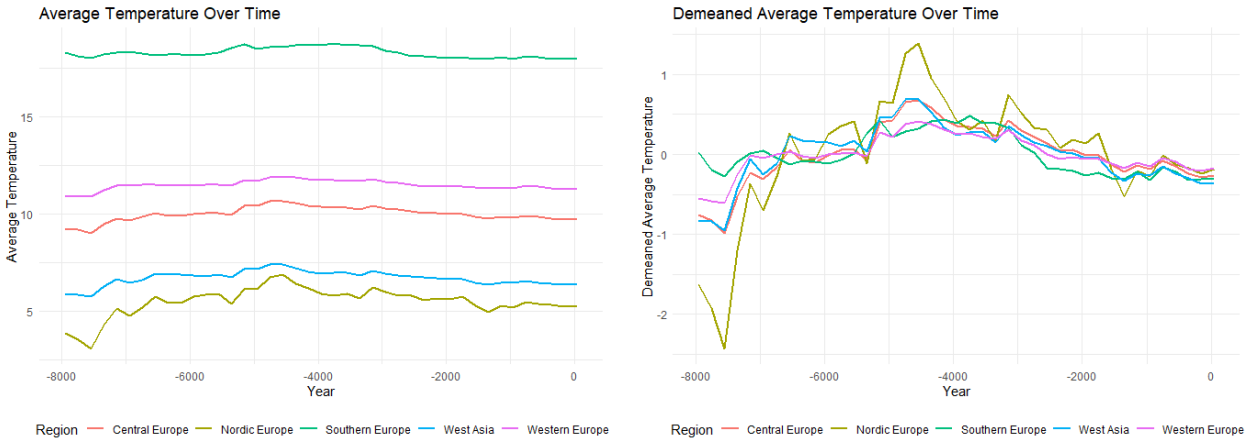
(d) Temperature - SD



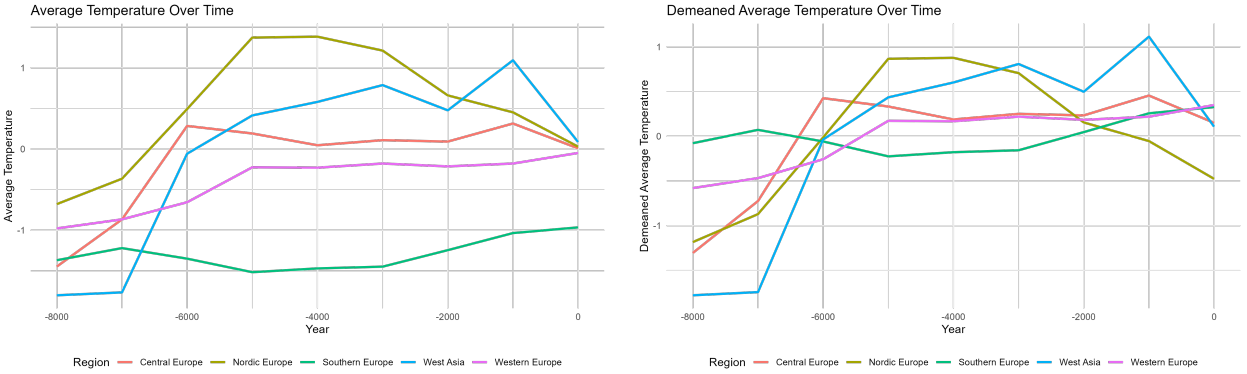
Notes: The top figures present the evolution of average global temperature over our period of interest (8,000BCE–0) along with its standard deviation using the marine-based proxies paleoclimate dataset (Osman et al., 2021). The bottom row presents the evolution of average global temperature (8,000BCE–0) along with its standard deviation for both paleoclimate reconstructions (marine-based in blue and pollen-based in green). The left-axis corresponds to the marine-based reconstruction while the right-axis corresponds to the pollen-based anomaly measure.

Figure 5: Temperature Evolution across Regions

(a) Marine climate dataset

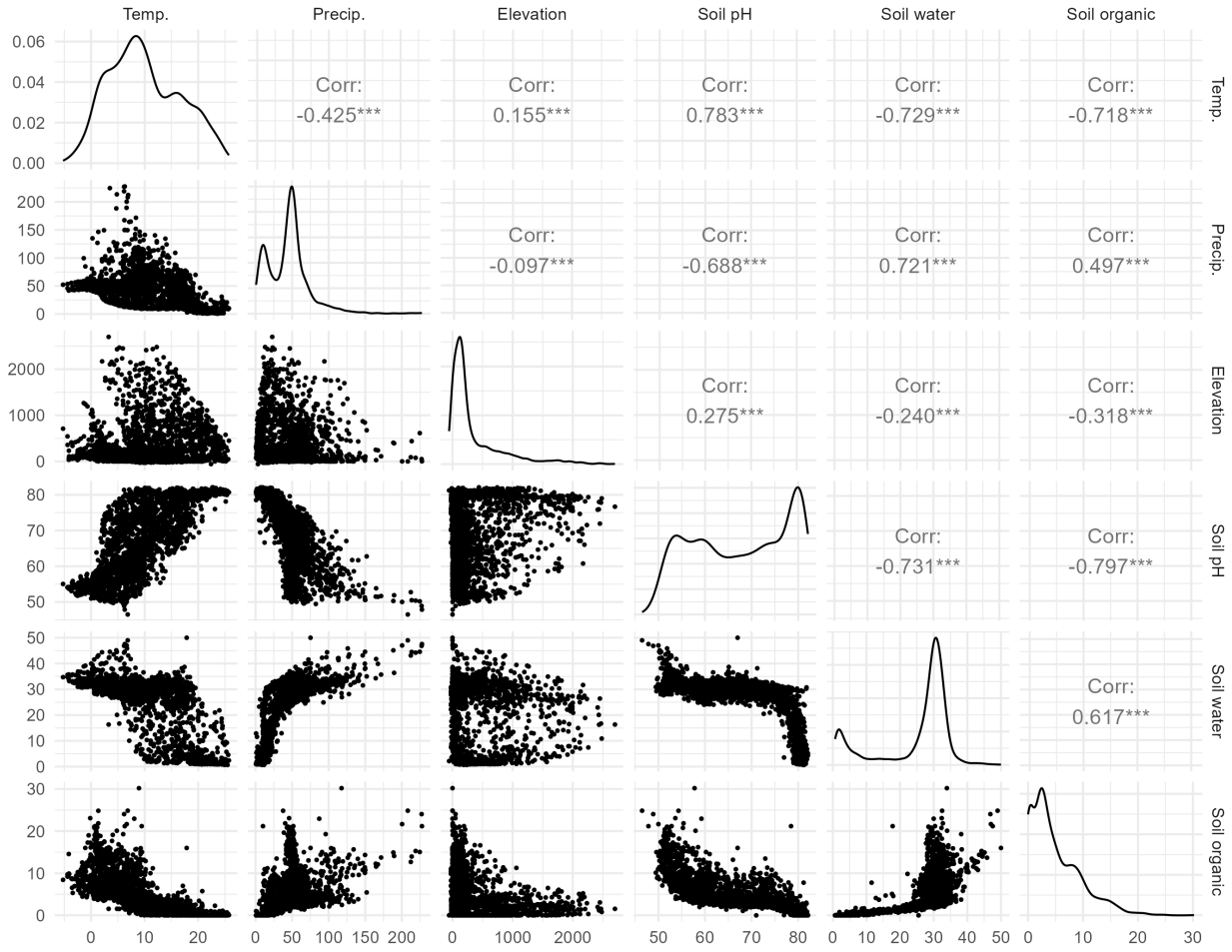


(b) Pollen climate dataset, anomalies



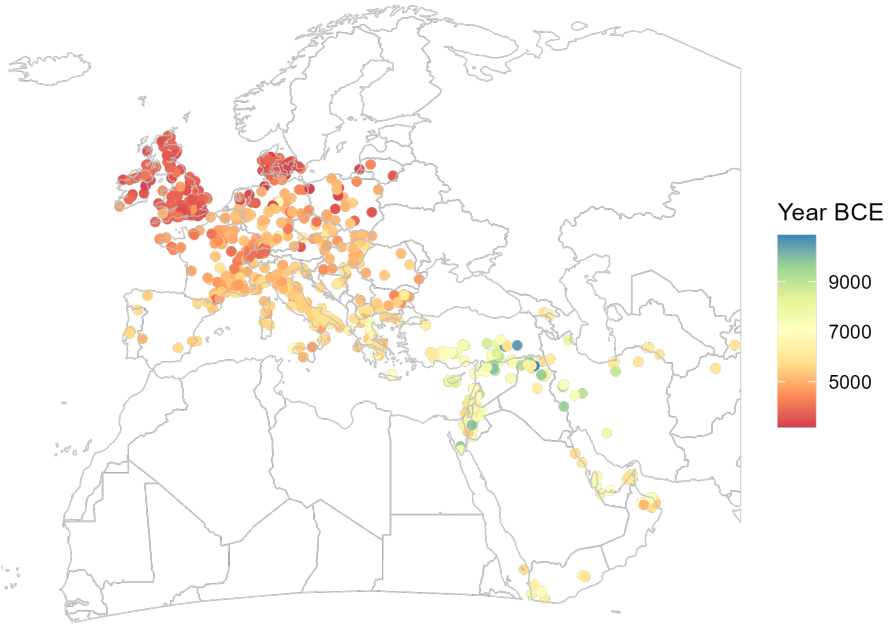
Notes: The figure plots the evolution of temperature across our period of interest, 8,000BCE - 0, for the five main European regions defined in Table A3 using reconstructions from marine-based proxies (panel a), pollen-based temperature anomalies (panel b). The right column shows demeaned temperature over time.

Figure 6: Correlation matrix between climate and soil variables



Notes: This figure plots the correlation plots between temperature, precipitation, elevation and the soil variables (soil pH, water holding capacity, organic carbon content, elevation) in each 1° x 1° cell of our analysis. Climate data come from Abatzoglou et al. (2018) and soil data from Hengl (2018); Hengl and Gupta (2019); Hengl and Wheeler (2018).

Figure 7: Year of Evidence of Agriculture



Notes: The figure presents the spatial distribution of the the 750 archaeological sites used for the analysis on the year of adoption of agriculture. Data are from Matranga (2024).

Tables

Table 1: Summary Statistics

	Mean	SD	Min	P10	Median	P90	Max	N
N Migrants, x 100,000	40.77	2827.60	0.00	0.00	0.00	0.00	1,200,000	5,040,665
Main Regressors								
Temperature Distance	7.64	5.60	0.00	1.17	6.55	15.84	30.07	5,040,665
Precipitation Distance	28.50	25.36	0.00	3.05	23.80	57.13	229.31	5,040,665
Geographic Distance	2,500.05	1,271.47	47.05	895.43	2,389.04	4,254.21	7,162.32	5,040,665
Soil and Topography								
pH Distance	11.75	8.49	0.00	1.30	10.34	24.49	35.60	5,001,430
Water Capacity Distance	11.22	10.99	0.00	0.74	5.58	28.55	49.33	5,001,430
Organic Distance	4.73	4.05	0.00	0.52	3.64	10.38	30.20	5,001,430
Elevation Distance	437.92	485.43	0.00	31.62	232.16	1,145.89	2,757.15	5,040,665

Notes: The table shows summary statistics on the sample of all possible pairs of ancestral and descendant locations cells. The number of Migrants has been multiplied by 100,000 for readability. Climate and soil variables have been computed based on Abatzoglou et al. (2018) and Hengl (2018); Hengl and Gupta (2019); Hengl and Wheeler (2018), respectively.

Table 2: Effect of Climate Similarity on Migration - 8,000BCE-0

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.379***	-0.190***	-0.138***	-0.063***	-0.108***
	(0.015)	(0.017)	(0.016)	(0.019)	(0.017)
Precipitation Distance	-0.019***	0.006***	-0.010***	-0.011***	-0.0006
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Observations	5,040,665	5,040,665	710,331	710,331	704,802
Pseudo R^2	0.088	0.148	0.278	0.282	0.288
Descendant Location FE			Yes	Yes	Yes

Notes: The table shows PPML regressions where the dependent variable is the number of migrants between a pair of ancestral and descendant locations cells. The sample consists of all possible pairs of ancestral locations cells and descendant location cells, and it includes all individuals who died over the period 8,000 BCE - 0. *Temperature Distance* (resp. *Precipitation Distance*) is the absolute value of the difference in average temperature (precipitation) between the two cells and is computed using Abatzoglou et al. (2018) data. In columns (2), (3) and (5) we include as additional controls the distance between the two cells, while in (4) the longitude and latitude distance. Additionally, in column (5), we include the distance in elevation, soil pH, water holding capacity, and organic carbon. Standard errors, reported in parentheses, are clustered at the cell pair level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table 3: Regression Results - Temporal Climate Change

Dep. var.:	Temp Change Over Time			
	(1)	(2)	(3)	(4)
Temperature Change Over Space	-0.009^{***}	-0.013^{***}	-0.015^{***}	-0.029^{***}
	(0.003)	(0.003)	(0.004)	(0.006)
Elevation Change Over Space			-0.0001^{***}	-0.0001^{***}
			(0.00003)	(0.00004)
Constant	-0.062^{***}		-0.063^{***}	
	(0.017)		(0.016)	
Descendant Location FE		Yes		Yes
Observations	2,396	2,396	2,396	2,396
Adjusted R^2	0.017	0.457	0.027	0.476

Notes: Each observation in this dataset corresponds to an individual who died between 8,000BC and 0. Our dependent variable is *Temp Change over Time* defined in Equation 3. *Temp Change over space* and *Elevation Change over space* are computed according to Equation 4 using the marine-proxy climate dataset. Column 2 and 4 include destination cell FE. Standard errors are clustered at the descendant location cell level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Summary Statistics at the Individual Level, by Agriculture Status

Status	N	Avg Temp Dist	SD Temp Dist	Avg Geo Dist	SD Geo Dist	Avg Grazing	Avg Cropland
Cropland	653	0.51	3.45	775.53	676.23	0.46	1.95
Grazing	1239	0.78	3.34	789.40	553.42	0.92	0.06
Other	508	0.32	3.25	844.26	731.37	0.00	0.00

Table 5: Dissemination of Agriculture by Distance

Dep. var.:	Year of Adoption BCE			
	From Turkey cradle		From Iraq cradle	
	(1)	(2)	(3)	(4)
Temp. Distance	-69.513***	-82.940***	-112.580***	-114.144***
	(19.569)	(18.771)	(30.798)	(31.868)
Precip. Distance	1.006***	0.957***	0.983***	0.994***
	(0.282)	(0.271)	(0.367)	(0.380)
Altitude Distance	0.010	0.062	0.560	0.565
	(0.216)	(0.223)	(0.503)	(0.482)
Geo Distance	-1.086***	-0.895***	-0.785***	-0.739**
	(0.185)	(0.233)	(0.299)	(0.335)
Geo Distance-sq	0.0001	0.0001	0.00001	0.00002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Temp. (at dest.)	67.130***	50.612**	−11.610	-11.834
	(18.201)	(19.637)	(22.840)	(23.789)
Precip. (at dest.)	-0.415	-0.367	-0.528	-0.520
	(0.277)	(0.266)	(0.346)	(0.339)
Altitude (at dest.)	0.321***	0.372***	−0.185	-0.176
	(0.122)	(0.119)	(0.479)	(0.464)
Temp. Seasonality Dist.		-1,020.987***		-41.155
		(386.298)		(210.720)
Precip. Seasonalty Dist.		167.509		106.629
		(210.615)		(235.165)
Temp. Seas. (at dest.)		-1,001.708***		-22.983
		(379.615)		(200.564)
Precip. Seas. (at dest.)		295.332		135.317
		(183.899)		(237.762)
Observations	749	749	749	749
R^2	0.701	0.706	0.697	0.698
Adjusted R^2	0.698	0.701	0.694	0.693

Notes: The table shows OLS regressions where the dependent variable is year of earliest adoption of agriculture in a given archaeological site. The sample consists of 749 different archaeological sites from Matranga (2024). *Temperature Distance* (resp. *Precipitation Distance*) is the absolute value of the difference in average temperature (precipitation) between the site and (presumed) origin point of agriculture, either Turkey or Iraq. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of Climate Similarity on Migration: Role of Agriculture

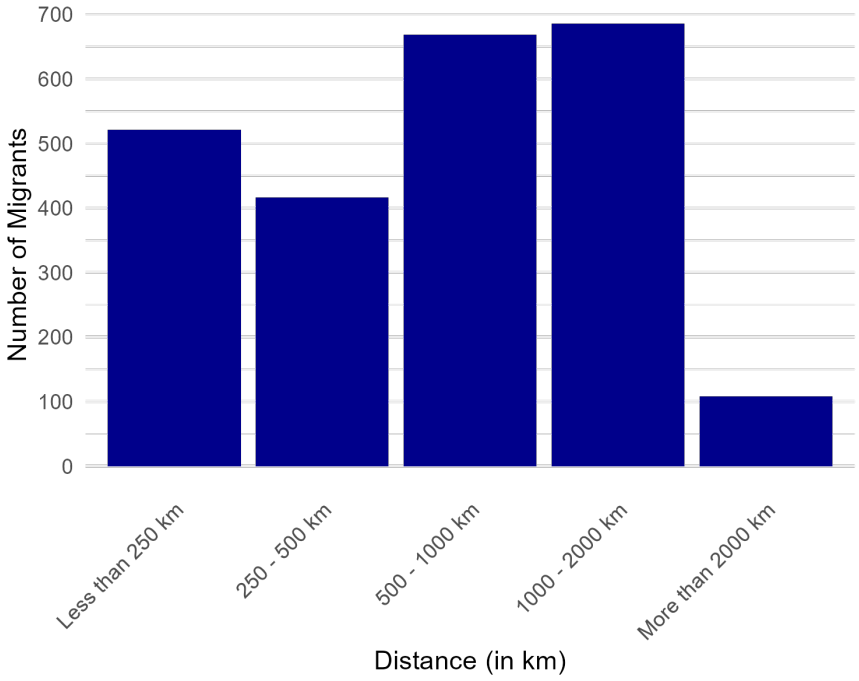
Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Panel A: Descendant Locations Fixed Effects					
Temperature Distance	-0.108***	-0.109***	-0.101***	-0.375***	-0.342***
	(0.017)	(0.031)	(0.020)	(0.102)	(0.064)
Precipitation Distance	-0.0006	-0.007*	-0.0006	0.009	0.011**
	(0.002)	(0.004)	(0.003)	(0.006)	(0.005)
Soil pH Distance	0.012	0.012	0.001	0.109***	0.012
	(0.008)	(0.014)	(0.010)	(0.032)	(0.008)
Water Hold. Capac. Distance	-0.117***	-0.086***	-0.144***	-0.138**	-0.115***
	(0.010)	(0.015)	(0.012)	(0.063)	(0.010)
Organic Carbon Distance	-0.106***	-0.215***	-0.047***	-0.152***	-0.106***
	(0.012)	(0.021)	(0.015)	(0.048)	(0.012)
Temp. Distance × 1[Cropland at Dest.]					0.200***
					(0.069)
Temp. Distance × 1[Grazing at Dest.]					0.267***
					(0.066)
Precip. Distance × 1[Cropland at Dest.]					-0.019***
					(0.006)
Precip. Distance × 1[Grazing at Dest.]					-0.010*
					(0.006)
Observations	704,802	273,686	389,942	41,174	704,802
Pseudo R^2	0.28812	0.34636	0.26307	0.30101	0.28922
Panel B: No Fixed Effects					
Temperature Distance	-0.122***	-0.146***	-0.081***	-0.390***	-0.406***
	(0.019)	(0.035)	(0.023)	(0.082)	(0.061)
Precipitation Distance	0.015***	0.003	0.021***	0.032***	0.033***
	(0.001)	(0.003)	(0.002)	(0.010)	(0.003)
Soil pH Distance	0.017***	0.041***	-0.015**	0.149***	0.014**
	(0.006)	(0.010)	(0.007)	(0.026)	(0.006)
Water Hold. Capac. Distance	-0.147***	-0.135***	-0.141***	-0.260***	-0.142***
	(0.008)	(0.014)	(0.009)	(0.062)	(0.008)
Organic Carbon Distance	-0.148***	-0.289***	-0.086***	-0.286***	-0.151***
	(0.016)	(0.026)	(0.019)	(0.080)	(0.016)
Temp. Distance × 1[Cropland at dest.]					0.237***
					(0.068)
Temp. Distance × 1[Grazing at dest.]					0.339***
					(0.063)
Precip. Distance × 1[Cropland at dest.]					-0.030***
					(0.004)
Precip. Distance × 1[Grazing at dest.]					-0.012***
					(0.003)
Observations	5,001,430	2,189,488	2,332,386	479,556	5,001,430
Pseudo R^2	0.16731	0.19208	0.15345	0.29330	0.17119
Sample	All	Cropland	Grazing	Others	All

Notes: The table shows PPML regressions where the dependent variable is the number of migrants between a pair of ancestral and descendant locations cells. *Temperature Distance* (resp. *Precipitation Distance*) is the absolute value of the difference in average temperature (precipitation) between the two cells and is computed using Abatzoglou et al. (2018) data. Column (1) replicates column (5) of Table 2; column (2) replicates it using only observations where the main mode of agriculture at the descendant location cell is cropland, Column (3) where the main mode of agriculture at the descendant location cell is grazing and Column (4) where the main mode of agriculture at the descendant location cell is neither of these. In column (5), we interact temperature and precipitation distance with dummies for the presence of cropland and grazing at destination, on top of including dummies for the presence of cropland and grazing at the descendant location. All regressions include a quadratic in geographic distance, on top of distance in elevation, soil pH, water holding capacity, and organic carbon; in panel A, we include descendant location cell FE. Standard errors, reported in parentheses, are clustered at the cell pair level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

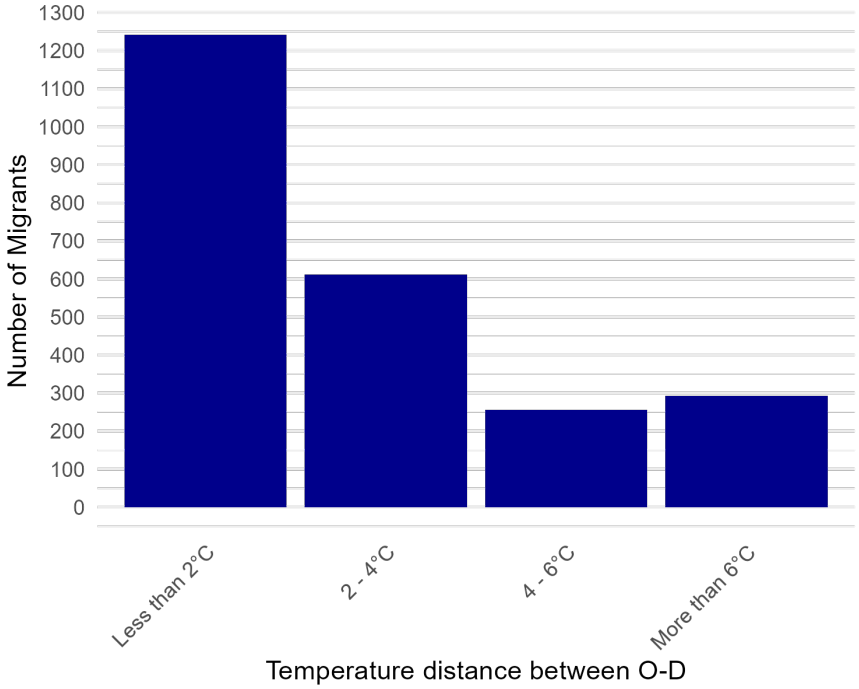
Appendix A - Additional Tables and Figures

Figure A1: Number of Migration Occurrences by Distance

Panel A: By Geographic Distance



Panel B: By Temperature Distance



Notes: Plot of the number of observations between ancestor and descendant location according to the geographic distance (Panel A) and temperature distance (Panel B). Temperature data associated with each gridcell comes from the marine paleodataset (Osman et al., 2021).

Table A1: Summary Statistics - Marine Paleo Climate dataset - Period 8,000BCE - 0

	Mean	SD	Min	P10	Median	P90	Max	N
N Migrants, x 100,000	2,196	31,854.70	0.00	0.00	0.00	0.00	2,600,000	96,858
Temperature Distance	7.90	5.68	0.00	1.23	6.96	16.21	31.02	96,858
Precipitation Distance	2.52	1.89	0.00	0.39	2.14	5.19	11.83	96,858
Geographic Distance	2,535.62	1,303.87	124.17	896.83	2,410.12	4,353.26	7,131.10	96,858

Notes: The table shows summary statistics on the sample of all possible pairs of ancestral and descendant locations cells for the sample using Marine-based paleoclimate reconstructions (Osman et al., 2021). The number of migrants has been multiplied by 100,000 for readability.

Table A2: Summary Statistics - NOAA Pollen Dataset - Period 8,000BCE - 0

	Mean	SD	Min	P10	Median	P90	Max	N
N Migrants, x 100,000	252.90	8,550.69	0.00	0.00	0.00	0.00	1,700,000	847,384
Temperature Distance	1.39	1.17	0.00	0.18	1.07	3.10	8.50	847,384
Precipitation Distance	3.40	2.84	0.00	0.48	2.66	7.39	19.26	847,384
Geographic Distance	2,139.71	1,019.80	47.05	804.96	2,082.68	3,550.71	5,583.32	847,384

Notes: The table shows summary statistics on the sample of all possible pairs of ancestral and descendant locations cells for the sample using pollen-based paleoclimate reconstructions (Mauri et al., 2015). The Number of Migrants has been multiplied by 100,000 for readability.

Table A3: Major European and Western Asia regions

Region	Longitude	Latitude
Western Europe	-5° to 10°	40° to 55°
Southern Europe	-10° to 30°	29° to 45°
Central Europe	10° to 40°	45° to 55°
Nordic Europe	0° to 40°	53° to 65°
West Asia	35° to 65°	40° to 65°

Table A4: Effect of Climate Similarity on Migration: Robustness to Paleo Climate

Dep. var.:	Number of Migrants				
	(1)	(2)	(3)	(4)	(5)
Temperature Distance	-0.108***	-0.179***	-0.074***	0.052	-0.152***
	(0.017)	(0.045)	(0.021)	(0.043)	(0.025)
Precipitation distance	-0.0006	0.018	-0.096	-0.191	0.043
	(0.002)	(0.015)	(0.071)	(0.156)	(0.076)
Observations	704,802	420,438	63,365	39,767	41,952
Pseudo R^2	0.288	0.295	0.370	0.356	0.359
Climate	Modern	Pollen	Marine	Marine	Marine
Period	All	All	All	8,000-3,300	3,300-0

Notes: The table shows PPML regressions where the dependent variable is the number of migrants between a pair of ancestral and descendant locations cells. *Temperature Distance* (resp. *Precipitation Distance*) is the absolute value of the difference in average temperature (precipitation) between the two cells and is computed using modern TerraClimate data. Column (1) replicates column (5) of Table 2; columns (2) replicates it using paleo climate reconstructions from Mauri et al. (2015), while (3) to (5) using reconstructions from Osman et al. (2021). All regressions include a quadratic in geographic distance, distance in elevation, soil pH, water holding capacity, and organic carbon, on top of destination cell fixed effects. Standard errors, reported in parentheses, are clustered at the cell pair level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A5: Number of Migrants by Direction of Move

Direction of Move	South to North	North to South	East to West	West to East
8,000BCE - 6,000BCE				
Number of migrants	44	9	32	20
6,000BCE - 3,000BCE				
Number of Migrants	226	214	263	174
3,000BCE - 0CE				
Number of Migrants	190	237	254	266

Table A6: Gravity Regressions - Robustness splits - Population density

Dep. var.:	Number of Migrants							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temperature Distance	-0.108***	-0.108***	0.262**	-0.074***	-0.122***	-0.098***	0.128	-0.073***
	(0.017)	(0.017)	(0.120)	(0.018)	(0.019)	(0.019)	(0.101)	(0.023)
Precipitation Distance	-0.0006	-0.0006	-0.018	-0.0005	0.015***	0.010***	-0.010	0.018***
	(0.002)	(0.002)	(0.023)	(0.003)	(0.001)	(0.002)	(0.015)	(0.002)
Population Density Distance							0.315***	-0.028
							(0.031)	(0.024)
Observations	704,802	704,802	48,440	404,474	5,001,430	2,634,788	4,997,008	5,000,830
Pseudo R^2	0.28812	0.28812	0.36405	0.30214	0.16731	0.17526	0.22286	0.16608
Period	8,000 BCE - 0	8,000 BCE - 0	8,000 BCE - 3,300BCE	3,300 BCE - 0	8,000 BCE - 0	8,000 BCE - 0	8,000 BCE - 3,300BCE	3,300 BCE - 0

Notes: The table replicates Table 2 by splitting the sample on pre-historical population density. Column (1) replicates column (5) of Table 2. In column (2), it restricts the sample to cells with individuals are present (either as ancestral or descendant location). Column (3) (resp. (4)) restricts the sample to cells with individuals and with population density is greater than 0 over the period 8,000BCE -3,300BCE (resp. 3,300BCE - 0). In columns (1) to (4) we include descendant location cell fixed effects; in columns (5) and (6) we replicate the specifications in (1) and (2) without these FEs. In Columns (7) and (8), we replicate column (5) of Table 2 by removing descendant location cell FEs and adding as a regressor the distance in population density. Standard errors, reported in parentheses, are clustered at the descendant location cell level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1