When Roots Diverge: Genes, Culture and Economic Development

Arcangelo Dimico*

Very Preliminary Version - Not for Circulation

Abstract

I evaluate the role that divergent cultural and biological histories have on long-run economic development. I exploit mismatches between genetic and linguistic histories across 397 ethnolinguistic groups worldwide to separate the effects of biological ancestry from those of culturally transmitted traits. Combining data on genetic distance, language families, folklore traditions, and nighttime lights, I show that cultural similarity strongly predicts cooperation, trust, prosocial norms, institutions, and economic development, while genetic distance primarily affects time preferences and risk attitudes. Results suggest that culture and biological history both matter for development, but they appear to operate through different channels.

JEL Classification: O11.

Keywords: Biological History, Cultural History, Geography, Cooperation, Preferences, Persistence.

---

[*] Arcangelo Dimico: Queen's University Belfast, GLO, IZA, CEPH, QUCEH, and HHDI. Email: a.dimico@qub.ac.uk.

1 Introduction

Why do some differences in economic development persist for centuries? A large part of the literature argues that current outcomes reflect not only contemporary policies and institutions, but also traits, beliefs, and constraints transmitted across generations. Two explanations have received particular attention. One emphasizes demographic history and biological ancestry. The other emphasizes culture: socially transmitted beliefs, norms, and practices that shape behavior within groups. Both approaches reject the view that development can be understood solely through contemporary incentives. However, they differ sharply in their interpretation of the most relevant deep roots - whether these are best understood in terms of biological endowments and demographic history, or in terms of cumulative culture and institutionalized norms. This distinction is sometimes blurred because population movements shape not only the distribution of genes, but also the diffusion, differentiation, and persistence of languages, norms, and social practices. As a result, patterns of cultural and linguistic diversification frequently mirror patterns of human genetic differentiation (Cavalli-Sforza et al. 1994). This approach has also been a dominant view in the economic literature on persistence with a large number of studies which has focused either on culture or genes in isolation, or combined the two into a single index (i.e., Becker et al. (2020)).

Yet history also produced many cases in which biological and cultural histories diverged. Groups may adopt new languages without large-scale population replacement. Migrants may mix genetically while retaining cultural traditions. Conquest, trade, and religious change may alter language and norms without fully changing ancestry. Indeed, recent evidence by Barbieri et al. (2022), shows that, although genetic and linguistic histories often evolve together, a substantial share of populations exhibits divergence between the two. Using data on 397 ethnolinguistic groups worldwide, they show that roughly one fifth of groups display a mismatch between genetic and linguistic histories, reflecting processes such as language replacement, cultural diffusion, and population mixing. These divergences are useful because they create variation with which to study the separate empirical roles of demographic and cultural histories on economic outcomes.

I combine data from Barbieri et al. (2022) with information on the distribution of languages across countries from Giuliano & Nunn (2018), folklore traditions from Michalopoulos & Xue (2021), and nighttime light data from NOAA to study the independent influence of genetic and linguistic histories on economic behavior and economic development. Following Becker et al. (2020), I focus on pairwise distances across populations to estimate whether groups that are genetically and culturally closer/further have smaller/larger differences in outcomes. I use F_{ST} as a measure of genetic distance. This measure captures the extent to which populations have diverged genetically through long-run demographic separation. Cultural relatedness, in line with Barbieri et al. (2022) and Becker et al. (2020), is proxied using language-family classifications, as groups speaking languages in the same family are more likely to share deeper cultural ancestry. Economic outcomes related to cooperation, innovation, investment and market expansion, together with preferences associated with such outcomes are proxied using differences in oral traditions from Michalopoulos & Xue (2021), which, as they show, closely track well-known geographic and social attributes across groups. I, finally, look at the macro-level implications (i.e., difference in incomes) to establish a link between culture vs genes, micro-level behaviour/decisions, and economic development.

Exploiting divergences in cultural and demographic histories across the 397 groups in Barbieri et al. (2022), and controlling for geographical distance between groups to avoid confounding population and cultural histories with physical distance, I find that groups speaking languages in the same family, and hence culturally closer, tend to have much more similar levels of cooperation. The estimated difference in cooperation is about 0.15 standard deviations smaller for culturally related pairs. By contrast, F_{ST} has little statistical relationship with cooperation in the baseline specification. This different effect that cultural relatedness and genetic distance exert on cooperation is largely explained by the different impact that the two have on preferences and beliefs associated with cooperation. Groups speaking languages in the same family have smaller differences in universal altruism, intrinsic giving, prosociality, reciprocity, trust, and norms related to reputation and punishment. Genetic distance, on the other hand, significantly affects only universal altruism and beliefs related to trust.

Because culturally similar groups tend to have similar levels of cooperation, they also have smaller differences in oral traditions related institutions, while genetic distance does not exert any significant effect on the latter. This result is consistent with the literature that has focused on cultural transmission, cooperation, and the functioning of institutions, such as Tabellini (2008*b*, 2010), and with the idea that humans adapt through socially transmitted rules, beliefs, and institutions that evolve much faster than genes. Such cumulative cultural evolution generates large-scale cooperation and fosters institutions (Bowles & Gintis 2003, Henrich 2004, Henrich et al. 2005, Henrich & Henrich 2006, Chudek & Henrich 2011, Chudek et al. 2013, Richerson & Henrich 2012, Henrich 2018, 2021, Savitskiy et al. 2026).

While cooperation and institution seem to be associated primarily to culture, differences in patience, risk aversion, gambling, and forward-looking behavior are associated with both genetic distance and linguistic relatedness. Groups that are genetically and culturally closer also exhibit smaller differences in oral traditions related to market expansion, investment, and innovation, which is in line with the joint effect that genes and culture exert on time preferences and risk attitudes. These patterns, therefore, suggest that demographic and cultural histories may both be relevant for traits linked to intertemporal choice, risk, and market activity.

Finally, I relate these divergent demographic and cultural histories to nighttime lights over the period 2003-2012 from NOAA which I use as a proxy for differences in income in levels. Pairs of groups that are genetically farther apart tend to display larger differences in nighttime lights. Pairs whose languages belong to the same family tend to display smaller differences. A one-standard-deviation increase in genetic distance is associated with an increase in light differences of about 0.1 standard deviations, while shared language-family membership is associated with a reduction of about 0.18 standard deviations. Interestingly, geographic distance itself exerts no significant effect on differences in income, which rules out concerns related to spatial correlation in the lights data.

These results are consistent with the view that both demographic and cultural histories are related to contemporary development, though the evidence does not by itself identify a single causal mechanism. I find that both processes have affected long-term differences in economic development. However, the mechanisms through which they affect development are distinct. Culture shapes cooperation and institutions most directly. Genes and culture jointly affect other traits - such as time preferences and risk attitudes - that also influence market expansion and long-run development.

Despite the empirical methodology closely resembles approaches in the literature, results do not imply a definitive causal relationship and should be read as a decomposition of historical associations rather than as a claim that genes or culture can be randomly assigned. The design controls for geographic distance and country fixed effects, and later adds a broad set of geographic and agro-climatic controls. These controls reduce several important confounds, but the estimates still capture conditional relationships between deep historical relatedness and observed outcomes. I therefore considers these results as evidence on channels and patterns and not as definitive causal effects.

Relative to the existing literature, this paper contributes along three margins. First, it separates two historical variables (i.e., genetic distance and linguistic relatedness) that are often considered in isolation or combined together. Becker et al. (2020), for example, use a composite ancestral-distance index to study the deep roots of preferences. I instead use cases in which genetic and linguistic histories diverge to study whether the two dimensions are associated with different outcomes, and which one predominantly affect economic outcomes. The paper also differs from the literature related to Ashraf & Galor (2013*a*, 2018), Arbatli et al. (2020), Galor et al. (2024), Galor, Klemp & Wainstock (2025), who study how prehistoric out-of-Africa migration shaped genetic diversity and, through it, comparative development, cultural fragmentation, conflict, and contemporary cultural diversity. While those papers show that deep population history is an important determinant of economic and cultural variation, I show that the non-alignment of genes and culture is itself empirically informative and can be used to separate their influence. Finally, compared with the literature that focuses on F_{ST} and uses genetic distance as a reduced-form measure of separation in vertically transmitted characteristics, such as Spolaore & Wacziarg (2009), this paper opens that black box by distinguishing the role of biological ancestry from that of cultural relatedness.

More broadly, the paper sits at the intersection of the literatures on genes and development and on culture and development, but differs from both. The literature on genes and development has shown that genetic diversity, genetic distance, and human dispersal are associated with income differences, trade, conflict, and cultural fragmentation. The culture-and-development literature, in turn, shows that values, beliefs, trust, and norms persist over time and affect development and institutional quality. These two traditions, however, have mostly proceeded in parallel: one emphasizes deep population history and biological ancestry, while the other emphasizes culturally transmitted values and institutions. The contribution of this paper is to provide a unified empirical framework, grounded in gene–culture coevolution, that studies both at once and separates them empirically. The resulting picture is nuanced: genes and culture jointly shape broad moral dispositions and market-related traditions, but culture plays the larger overall role in cooperation, especially where enforcement, prosocial norms, and institutions are concerned.

2 Literature Review

The paper relates to a broad and growing literature on the deep historical determinants of economic development. Ashraf & Galor (2013*b*) argue that the prehistoric dispersal of humans out of Africa generated systematic variation in genetic diversity and that this variation had long-run consequences for comparative economic development. Their argument builds on the serial founder effect, whereby successive migrations reduced within-group diversity as human populations moved farther from East Africa. In related work, Ashraf & Galor (2013*a*) show that genetic diversity is also associated with ethnic and cultural fragmentation, suggesting that ancient demographic processes affected not only productivity but also the internal social organization of populations. Galor & Moav (2002) stress evolutionary mechanisms in the transition to growth, and more recently, Galor et al. (2024), Galor, Klemp & Wainstock (2025), Galor, Özak & Sarid (2025) have extended the deep-roots approach to the origins of cultural diversity and the dispersal of productive traits. In this literature, gene-related or population-history related variables have been linked to economic performance associated with immigrant genetic diversity in the United States (Ager & Brueckner 2018), and to conflict or fragmentation associated with deep-root diversity patterns (Ashraf & Galor 2013*a*, Arbatli et al. 2020). A few studies focus on genetic distance rather than genetic diversity. Spolaore & Wacziarg (2009) consider genetic distance as a reduced-form proxy for the time elapsed since populations shared common ancestors, and therefore as a measure of barriers to the diffusion of development. Bove & Gokmen (2017) and Gokmen (2017) evaluate the impact of genetic and civilizational similarity on trade.

The literature on culture as a determinant of development is also tightly related to this paper. Guiso et al. (2006) define culture as inherited beliefs and values transmitted across generations and argue that such traits affect economic outcomes in systematic ways. Contributions by Fernández (2008, 2011), Giavazzi et al. (2019), and Giuliano & Nunn (2021) further consolidate this perspective, showing how persistent cultural traits help explain variation in labour supply, education, family structure, trust, and political preferences. Within this literature, trust and cooperation occupy a central role. Algan & Cahuc (2010) show that inherited trust matters for growth, while François & Zabojnik (2005) stress the relationship between trust, social capital, and development. Tabellini (2008*a*, 2010) link values and the scope of cooperation to institutional performance and regional development in Europe. Bisin & Verdier (2000, 2001) develop a theoretical framework to study cultural transmission. In these models, preferences, identities, and values persist because parents and communities deliberately socialize the next generation. Bisin et al. (2004) focus on cooperation itself as a culturally transmitted trait. This theoretical framework has been especially useful for understanding why norms related to family, trust, gender, and civic conduct persist even when formal institutions or economic conditions change. Examples include Figlio et al. (2019) on long-term orientation and education, Anderson (2025) and Becker (2025) on gender-biased and sexual norms, and Grosjean (2014) and Nisbett (1993) on cultures of honour and violence. Chen (2013) and Galor & Özak (2016) show that future orientation is affected by a structural feature of languages and historical agricultural productivity, respectively. Posch et al. (2026) use surnames in the United States as a proxy for cultural diversity to estimate the impact of culture on innovation. The common implication is that cultural traits are not epiphenomenal; they are durable mechanisms through which history shapes present-day economic outcomes.

A related literature studies the origins of cooperation through cultural evolution and gene-culture coevolution Henrich (2004), Henrich & Henrich (2006), Henrich (2021), Chudek & Henrich (2011), Chudek et al. (2015, 2013). They argue that large-scale cooperation emerged because socially transmitted norms, punishment systems, reputational mechanisms, and institutions altered the environment in which human interaction took place. Culture thereby became cumulative: once norms and institutions could be socially transmitted and improved across generations, societies became capable of sustaining forms of cooperation far beyond kinship-based interaction. Bowles & Gintis (2003), Richerson & Henrich (2012), Bhui et al. (2019), and Savitskiy et al. (2026) reinforce this view that culture is not only persistent but also adaptive, allowing groups to solve collective-action problems and expand exchange. This literature has important implications for economics because suggests that different behavioural traits may be transmitted through different channels. Generalized morality, reciprocity, and trust may reflect both inherited predispositions and culturally shaped expectations, while norms of punishment, reputation, and institutionalized collective action are more likely to depend on social learning and community enforcement. The empirical literature on economic preferences is consistent with this interpretation. Becker et al. (2020) trace cross-population variation in preferences back to ancient migration patterns, suggesting deep historical origins, while Enke (2019) links kinship structures to cooperation and moral systems. Similarly, Cao et al. (2021) show how herding and conflict help generate cultures of honour. Together, these contributions suggest that many traits relevant for development - trust, patience, risk-taking, and moral universalism - have deep roots, but they do not always distinguish whether persistence operates through biological inheritance, cultural transmission, or the interaction between the two.

The paper is also related to the literature that consider the relationship between cultural and biological histories directly. Cavalli-Sforza et al. (1994) document the historical geography of human genes and highlight the extent to which migration and interaction shaped genetic patterns. More recently, Barbieri et al. (2022) show that genetic and linguistic histories often evolve together, but not always. A substantial minority of groups exhibit mismatches between biological ancestry and language history due to language replacement, cultural diffusion, and population mixing. Related evidence on linguistic and cultural distance appears in Desmet et al. (2009) and Felbermayr & Toubal (2010), while broader work on ancestral characteristics and population flows is provided by Giuliano & Nunn (2018) and Putterman & Weil (2010). This literature opens the possibility of using divergence between cultural and demographic histories as an identification strategy.

Finally, the paper also speaks to work linking cooperation, conflict, and institutions to long-run development. Montalvo & Reynal-Querol (2005), Moscona et al. (2020), and Guarnieri & Tur-Prats (2023) show how social organization and cultural distance affect conflict outcomes. Voigtländer & Voth (2012) demonstrate the persistence of historically rooted hostility and violence. Greif et al. (2025) and Mokyr (2014) place culture at the centre of comparative institutional development, arguing that beliefs and values shape the evolution of prosperity over the long run.

3 Background

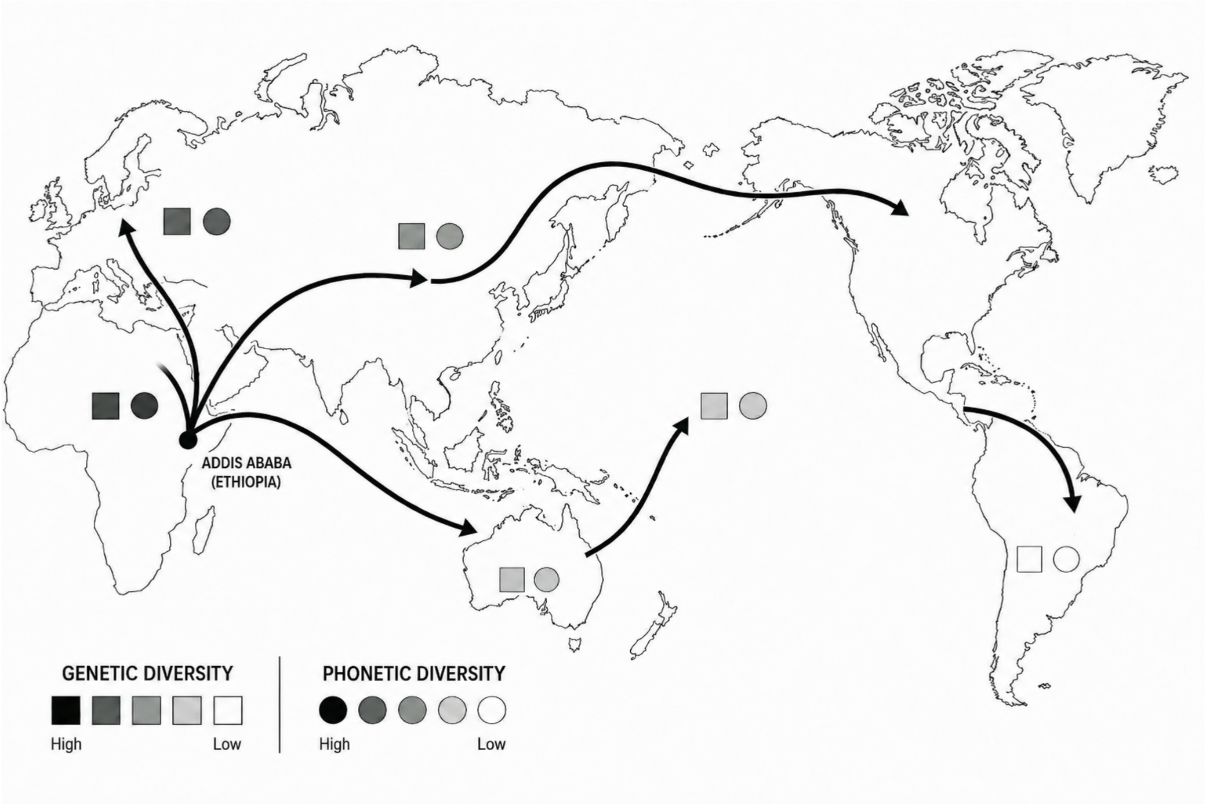
A large number of studies in human population genetics have shown that contemporary genetic variation is strongly associated with the prehistoric expansion of modern humans out of Africa. Migrating populations repeatedly split off from ancestral groups as humans expanded across the globe. Each split carried only a subset of the genetic variation present in the source population. Successive founder events reduced the stock of variation carried by migrating populations. As a result, within-population genetic diversity, captured by genetic heterozygosity, tends to decline with geographic distance from East Africa (Prugnolle et al. 2005, Ramachandran et al. 2005). The same demographic processes also generate genetic distance between populations, which is generally measured using Wright (1951)'s fixation index, F_{ST}. F_{ST} summarizes the share of total genetic variation attributable to differences in allele frequencies across populations and captures deep historical separation between populations. Low values of F_{ST} indicate that two populations are genetically similar, while higher values indicate greater differentiation.

Human populations, however, have rarely evolved in isolation. Migration, conquest, trade, intermarriage, and frontier settlement have repeatedly brought together groups with different demographic histories. This genetic mixture can affect within-population genetic diversity, reduce genetic distance between previously separated groups, and weaken the correspondence between geography, ancestry, and culture. Different cultural practices, in turn, alter the social environment in which individuals act, influencing selection pressures over long time horizons (Boyd & Richerson 1985).

Language diversity plays a central role in this interaction. Languages are transmitted culturally, but they often track population histories because they are typically inherited within groups over time. As a result, populations that share recent common ancestry tend to be both genetically and linguistically similar. Atkinson (2011), for example, argues that phenotypic diversity declines with distance from Africa, as predicted by a serial founder. This alignment between cultural and genetic diversity is shown in Figure 1 which shows how successive founder events decreased the phonetic and genetic pool of a population.

However, cultural traits, including language, can spread horizontally through mechanisms such as conquest, trade, religious conversion, or institutional change, without proportional genetic replacement. Conversely, populations may experience substantial genetic admixture while maintaining linguistic continuity. Consistent with this, Barbieri et al. (2022) document that although genetic and linguistic histories are often aligned, a

Figure 1: Genetic and Linguistic Serial Founder Effect



Notes: Serial founder effect map illustrating how genetic and phenotypic diversity declines the further a population migrates from its ancestral origin. The map is a readaptation of Atkinson (2011) substantial share of populations exhibit mismatches between the two. These mismatches reflect historical episodes in which cultural transmission, such as language replacement or diffusion, and biological transmission, such as ancestry and genetic mixing, followed different trajectories. Figure 2 shows the probability of mismatches across 313 out 397 groups in Barbieri et al. (2022). Under the linguistic serial founder hypothesis, the average genetic distance (FST) within language families should be much smaller than the average FST across families. Hence, the difference between average FST between families within a given region (i.e., continent) and average FST within families within the same region should be relatively high. The figure shows that there are several groups for which FST and language distance are, in probability, aligned (darker colours). However, there is a considerable number of groups for which the two seem to be misaligned.

Figure 2: Probability Distribution of Mismatches

Notes: Gradients of black show the probability of mismatches considering the median FST between language families in a give region and the median FST within language families. Darker colours show a lower probability of mismatch.

This distinction is also important for understanding the mechanisms that generate long-run differences in development. Genetic heterozygosity and F_{ST} capture dimensions of demographic history and population separation, while language families capture cultural history and inherited norms. Because the two are correlated but imperfectly aligned, mismatches between genetic and linguistic histories provide useful variation for separating biological ancestry from cultural transmission. Populations that are genetically close but linguistically distant, or linguistically close but genetically distant, allow to distinguish whether observed differences in economic development are more closely associated with deep demographic history or with culturally transmitted similarity. This difference also has practical implications for economic development, given that culture changes at a much faster pace than genes, with implications for persistence and possibility to move to different equilibria.

4 Data and Empirical Model

4.1 FST and Language Families

Data on genetic diversity and language families are collected from Barbieri et al. (2022). They provide data for 397 populations typed for the Human Origins Array (Affymetrix), grouped into 53 language families and 295 languages, distributed across continents. For each pair in the dataset, they compute the F_{ST}, which they then use to generate the median F_{ST} between populations in their macro-geographic region and the median F_{ST} between populations within a radius of 1,000 km. Mismatches are identified by looking at the proportion of genetically close and linguistically unrelated pairs and by comparing F_{ST} distributions within and between language families.

Using population names and Glottocodes for each ethnolinguistic group in Barbieri et al. (2022), I match this source with the spatial distribution of linguistic groups from Ethnologue 16 provided by Giuliano & Nunn (2018). Figure B1 in Appendix B shows the distribution of these 397 groups across the globe, and Figure B2 reports the number of neighbouring groups within 1,000 km speaking languages within different families. The latter shows substantial cultural heterogeneity across continents. Even in Europe, which is generally considered much more culturally homogeneous, there is considerable cultural heterogeneity. Hungarians represent a typical example. They are genetically similar to neighbouring groups ($F_{ST} = 0.0017$), but, different from other European groups they speak a Uralic rather than an Indo-European language, denoting substantial cultural difference. Figures B3 and B4 show the distribution of groups that are genetically close to unrelated language families and the probability of mismatch, in percentiles, between cultural and biological histories, respectively. As argued by Barbieri et al. (2022), despite a tendency for cultural and biological histories to follow similar patterns, there is a substantial number of groups - around 20% according to Barbieri et al. (2022) - for which roots tend to diverge.

4.2 Preferences, Beliefs, and Economic Outcomes

I complement data from Barbieri et al. (2022) and Giuliano & Nunn (2018) with oral traditions provided by Michalopoulos & Xue (2021). Folkloric traditions, through the transmission of narratives, encode normative structures that can be used to recover unobserved historical traits and, hence, mapped into economic preference domains. The data is based on narratives recorded in ethnolinguistic communities and reported in Berezkin (2015) who classify thousands of motifs across 958 societies. These motifs are coded and linked to the Ethnographic Atlas by Michalopoulos & Xue (2021) and can therefore be matched easily with data in Giuliano & Nunn (2018).

Michalopoulos & Xue (2021) show that motifs reflect known geographic and social attributes and that these traditions are strongly associated with current traits related to trust, risk aversion, patience, reciprocity, and other contemporary traits, which they proxy using current surveys such as the WVS and GPS. This means that I can use information in the Folklore dataset to generate proxies for ancestral preferences, beliefs, and other economic outcomes for each population in Barbieri et al. (2022). These motifs represent stories less affected by modernization and, hence, traits which reflect ancestral characteristics transmitted over generations. Hence these traits, unlike preferences and beliefs collected from current surveys at a country level, are likely to be less volatile because they do not respond to economic and political conditions, and have a lower exposure to genetic mixture and cultural amalgamation,

To map relevant motif-related terms into preferences and beliefs, I use a fully AI-assisted procedure. I start by feeding into AI all motif-related concepts in the Folklore dataset - more than 9,000. I then ask it to generate variables by choosing terms relevant to each trait. For example, to generate a proxy for cooperation, I ask AI to choose relevant terms which can be associated with cooperation in public-good games. From the provided lists, I choose the recommended core subset of terms. I then ask AI to group these preferences related to cooperation into: 1) pure/universal altruism; 2) warm glow/intrinsic giving; 3) reciprocity; 4) generalized trust; and 5) social norms related to reputation, punishment, and signalling, to understand the mechanisms through which genes and culture affect cooperation. I repeat the same procedure to generate proxies for time preferences and risk aversion, and behaviours related to those preferences, such as gambling and forward-looking behaviour. Finally, I also generate proxies for economic and institutional outcomes such as markets, institutions, collective action, investment, and innovation. For example, to generate a proxy for institutions, I ask AI to consider the definition in North (1991) who defined institutions as *"the rules of the game in a society - humanly devised constraints that shape human interaction, incentives, and economic outcomes"*, and to choose terms accordingly. Appendix A reports queries and links to the search, the subset of terms obtained from the search, and a short explanation of how variables are generated.

I also generate a measure of income per capita at the group level to consider the final impact of genes and culture on economic development. This measure relies on the strong association between GDP per capita and nighttime-light data (Henderson et al. 2012). From NOAA's National Geophysical Data Center, I download nighttime-light data from 2003 to 2012.[1] I overlay nighttime-light data onto linguistic groups from Giuliano & Nunn (2018) and generate the average level of economic activity for each group over the period 2003–2012.

4.3 Geographical Controls

Galor, Özak & Sarid (2025) argue that the coevolution of cultural and linguistic traits reflects variation in geographical characteristics that were conducive to higher natural returns to agricultural investment, while Galor & Özak (2016) relate the origins of time preferences to pre-industrial agro-climatic characteristics. I overlay rasters related to the caloric suitability index and plow potential obtained from them[2] onto language groups from Giuliano & Nunn (2018), and construct measures of average caloric suitability, and positive and negative plow-suitable crops before and after Columbus for each linguistic group. Data on average temperature, rainfall, and altitude are from WorldClim,[3] which again is overlaid onto the distribution of linguistic groups to obtain measures of average temperature and precipitation at the group level. Natural Earth[4] is the source for the distribution of water bodies and urban places, which I use to generate distance measures for each group. I also generate the centroid distance of each group from the shore and an index of ruggedness using data altitude.

4.4 Empirical Model

From an empirical point of view, I estimate the same model as in Becker et al. (2020). Given that F_{ST} measures the distance between groups, I focus on differences between groups. It should be noted that pairs are symmetric in this context, since F_{ST} and all other differences in outcomes between group i and group j are exactly the same as those between group j and group i. Hence, keeping both pairs (i.e., i,j and j,i) would only

---

[1] Version 4 DMSP-OLS Nighttime Lights Time Series.

[2] These rasters are available from Ömer Özak at: `https://ozak.github.io/Caloric-Suitability-Index/`.

[3] `https://www.worldclim.org/data/bioclim.html`.

[4] `https://www.naturalearthdata.com/`.

inflate the degrees of freedom. For this reason, I focus on differences between group i and all the other $N-i$ groups in the dataset[5].

I estimate variations of the following model:

$$|Y_{i,c_1} - Y_{j,c_2}| \;=\; \gamma FST_{ij} \,+\, \lambda \text{LangDistance}_{ij} \,+\, \theta \text{GeoDistance}_{ij} \,+\, \rho_{i,c_1} \,+\, \psi_{j,c_2} \,+\, \varepsilon_{ij}.$$

In the model above, $|Y_{i,c_1} - Y_{j,c_2}|$ measures the absolute difference in outcome variables between group i in country c_1 and group j in country c_2. For example, for cooperation, $|Y_{i,c_1} - Y_{j,c_2}|$ is the sum of all terms related to cooperation for group i, normalized by the total number of motifs for group i, minus the sum of all terms related to cooperation for group j, normalized by the total number of motifs for group j. Given the symmetry, I take the absolute value. FST_{ij} measures the genetic distance between groups i and j. LangDistance_{ij} captures the cultural distance between groups, which, as in Barbieri et al. (2022), I proxy using a dummy for whether languages belong to the same language family. GeoDistance_{ij} represents the physical distance between groups and is used to avoid confounding geography, because of the out-of-Africa process, with genetic and cultural differences. Finally, ρ_{i,c_1} and ψ_{j,c_2} represent country fixed effects for country c_1 and country c_2, respectively, to account for national institutions that can affect outcomes.

There are three major differences compared with Becker et al. (2020). The first, and most important, is that the model above decomposes the effect of culture from the effect of genes by exploiting mismatches between cultural and biological histories at the ethnolinguistic-group level, while Becker et al. (2020) uses a composite index aimed at capturing temporal distance between countries. Second, the model controls for the geographical distance between groups, given that groups farther from each other are more likely to have larger genetic distance and greater cultural diversity. Third, given that culture and other traits are likely to be transmitted across groups independently of whether they are within the same national boundaries, I allow for spatial correlation in the error term and use a Conley spatial HAC robust standard errors with a 1,000 km threshold to allow for transmission across groups that are relatively far from each other. Dependence within the dyad is excluded given that I only exploit differences in one direction due to symmetry.

---

[5] This means that differences between j and N-j are not considered because of the symmetry.

5 Results

5.1 Cooperation and Mechanisms Sustaining Cooperation

I start by evaluating the impact of culture and genes on cooperation, which has been central to the debate on the coevolution of culture and genes. Bowles & Gintis (2003), Henrich (2004), and Henrich (2021), for example, highlight how cumulative cultural evolution, norm psychology, and culturally transmitted institutions generated large-scale cooperation. Tabellini (2010) relates interpersonal trust to linguistic features and the genetic distance between two populations.

Table 1, Panel A, reports results controlling only for country fixed effects. In Panel B, I add controls for the number of groups within 1,000 km from group i and group j, as a measure of fractionalization, and the lowest percentile mismatch considering the F_{ST} distribution within the continent, as a measure of regional genetic relatedness across unrelated languages. Panel C also controls for geographic characteristics for the two groups, such as the caloric suitability index before and after Columbus, negative and positive plow-suitable crops before and after Columbus, temperature, rainfall, elevation, ruggedness, distance from the coast, distance from water bodies, distance from urban places, latitude, and longitude.

In Model 1, Panel A, the dummy for same language family is significant at the 1 percent level. Speaking a language belonging to the same family decreases the difference in the extent of oral traditions associated with cooperation by almost 0.16 standard deviations. A one-standard-deviation increase in physical distance increases the difference in cooperation by approximately 0.37 standard deviations. Genetic distance, on the other hand, does not exert any significant effect on the difference in cooperation between groups. Adding controls, in Panels B and C, does not substantially change the magnitude or significance of the effects. For example, in Panel C, the effect of language is to reduce the difference in cooperation by 0.14 standard deviations. A one-standard-deviation increase in F_{ST} corresponds to a 0.03 standard-deviation change in Panel C and a 0.07 standard-deviation change in Panel A, but both effects are insignificant.

In the following models, I look at the effect on prosocial behaviour and preferences, beliefs, and norms generally associated with prosociality and cooperation. Groups speaking languages within the same family have smaller differences in terms of prosociality, pure or universal altruistic preferences, intrinsic giving or warm glow, reciprocity, trust, and norms related to reputation, punishment, and signalling. This effect persists even after controlling for fractionalization, probability of mismatch, and geography. Geographical distance also has a robust effect on each of the six dimensions considered. Genetic distance, on the other hand, has a significant impact on pure altruism, intrinsic giving, and generalized trust in Panel A. However, the effect on warm glow is not robust to the inclusion of control variables.

These results are consistent with Bowles & Gintis (2003), Henrich (2004), and Henrich (2021), who consider cooperation as a result of cultural adaptation rather than a consequence of demographic and biological histories. This difference in the effect of genes versus culture is attributable to the effect that culture exerts on prosocial behaviour and the motivations that sustain such behaviour.

5.2 Time Preferences and Risk Aversion

Time preferences and risk aversion have been central to the persistence literature. Chen et al. (2006) study the decision-making of capuchin monkeys using laboratory experiments and find that, like humans, capuchin monkeys also display reference-dependent choices and loss aversion, suggesting that decision-making biases may be the result of innate characteristics rather than learned or transmitted behaviours. Galor & Savitskiy (2018) attribute differences in loss aversion to climatic shocks. *Toward an understanding of the development of time preferences: Evidence from field experiments* (2019) find differences in time preferences across races and no correlation between the preferences of parents and children, excluding intergenerational transmission of preferences and pointing to racial or innate differences. Galor & Özak (2016) attribute differences in time preferences to pre-industrial agro-climatic differences, while Galor et al. (2020) and Chen (2013) study the impact of language on time preferences, human capital, and related behaviours.

Table 2 reports the effects of geography, cultural distance, and genetic distance on patience, used as a proxy for time preferences, and risk aversion. In Models 3 and 4, I replace the dependent variables with alternative measures of risk aversion and time preferences represented by gambling and forward-looking behaviour. All these dependent variables are AI-generated using data from the Folklore dataset and therefore represent ancestral characteristics orally transmitted across generations.

Culture, as proxied by the dummy for languages within the same family, and geography both exert significant effects on measures associated with time preferences, such as patience and forward-looking behaviour, and risk aversion, such as risk aversion and gambling. Populations speaking languages within the same family have smaller differences in patience, with the difference decreasing by 0.28 standard deviations, and risk aversion, with the difference decreasing by 0.23 standard deviations. Groups farther from each other have significantly larger differences in both time preferences and risk aversion. These effects barely change once the full set of controls is included in the model, as shown

Table 1: Cooperation and Preferences

Panel A

	(1) Cooperation	(2) Prosociality	(3) Altruism	(4) Warm Glow	(5) Reciprocity	(6) Trust	(7) Reputation
Same Language Fam.	-0.155***	-0.112**	-0.136***	-0.148***	-0.127**	-0.0966***	-0.334***
	(0.0444)	(0.0522)	(0.0500)	(0.0336)	(0.0530)	(0.0363)	(0.0637)
Geographic Distance	0.365***	0.208***	0.125***	0.171***	0.198***	0.227***	0.287***
	(0.0330)	(0.0316)	(0.0318)	(0.0311)	(0.0324)	(0.0303)	(0.0322)
Genetic Distance - FST	0.0725	0.0126	0.109***	0.0874**	0.0587	0.170***	0.0910*
	(0.0476)	(0.0390)	(0.0404)	(0.0386)	(0.0494)	(0.0415)	(0.0527)
Observation	51680	51680	51680	51680	51680	51680	51680

Panel B

	(1) Cooperation	(2) Prosociality	(3) Altruism	(4) Warm Glow	(5) Reciprocity	(6) Trust	(7) Reputation
Same Language Fam.	-0.156***	-0.121**	-0.131**	-0.152***	-0.143***	-0.0988***	-0.334***
	(0.0436)	(0.0539)	(0.0532)	(0.0336)	(0.0526)	(0.0373)	(0.0632)
Geographic Distance	0.402***	0.213***	0.137***	0.183***	0.234***	0.248***	0.316***
	(0.0346)	(0.0330)	(0.0354)	(0.0306)	(0.0346)	(0.0320)	(0.0322)
Genetic Distance - FST	0.0109	0.00561	0.118**	0.0588	-0.00434	0.129***	0.0354
	(0.0521)	(0.0437)	(0.0468)	(0.0438)	(0.0514)	(0.0444)	(0.0519)
Observation	48827	48827	48827	48827	48827	48827	48827

Panel C

	(1) Cooperation	(2) Prosociality	(3) Altruism	(4) Warm Glow	(5) Reciprocity	(6) Trust	(7) Reputation
Same Language Fam.	-0.147***	-0.114**	-0.114**	-0.143***	-0.154***	-0.113***	-0.327***
	(0.0432)	(0.0457)	(0.0554)	(0.0306)	(0.0493)	(0.0339)	(0.0632)
Geographic Distance	0.390***	0.201***	0.125***	0.197***	0.221***	0.247***	0.317***
	(0.0351)	(0.0270)	(0.0310)	(0.0315)	(0.0354)	(0.0345)	(0.0324)
Genetic Distance - FST	0.0354	0.0303	0.139***	0.0395	0.0257	0.137***	0.0321
	(0.0553)	(0.0384)	(0.0444)	(0.0412)	(0.0522)	(0.0484)	(0.0542)
Observation	48827	48827	48827	48827	48827	48827	48827

Notes: Dependent variables are standardized differences between group i and group j. Panel A includes only country fixed effects for country i and country j. Panel B adds the number of neighbouring groups within 1,000 km, as a measure of regional fractionalization, and the lowest percentile mismatch, as a measure of the distribution of regional mismatches. Panel C adds the caloric suitability index before and after Columbus, negative and positive plow suitability before and after Columbus, average temperature, rainfall, altitude, ruggedness, distance from the coast, distance from water bodies, distance from the closest urban place, latitude, and longitude. Controls are included for both group i and group j. Conley HAC standard errors use a 1,000 km threshold. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

in Panel C.

F_{ST} in Panel A has a significant effect only on proxies for time preferences, while the effects on risk aversion and gambling are barely significant. The effects on proxies for risk aversion become significant once the full set of controls is included, as shown in Panel C, suggesting a conditional effect of genetic distance that depends on the surrounding environment and highlighting the interplay between geography and biological history. A one-standard-deviation increase in F_{ST} is associated with approximately 0.15 standard deviations in risk aversion and 0.29 standard deviations in patience.

Table 2: Time Preferences and Risk Aversion

Panel A

	(1) Risk Aversion	(2) Patience	(3) Gambling	(4) Forward-Looking
Same Language Fam.	-0.284***	-0.227***	-0.196***	-0.244***
	(0.0848)	(0.0657)	(0.0543)	(0.0621)
Geographical Distance	0.221***	0.0794***	0.0929***	0.101***
	(0.0331)	(0.0262)	(0.0326)	(0.0212)
Genetical Distance - FST	0.119*	0.228***	-0.00908	0.155***
	(0.0672)	(0.0612)	(0.0225)	(0.0525)
Observation	51680	51680	51680	51680

Panel B

	(1) Risk Aversion	(2) Patience	(3) Gambling	(4) Forward-Looking
Same Language Fam.	-0.262***	-0.203***	-0.184***	-0.225***
	(0.0838)	(0.0657)	(0.0532)	(0.0623)
Geographical Distance	0.230***	0.0552*	0.0800***	0.0823***
	(0.0343)	(0.0284)	(0.0295)	(0.0245)
Genetical Distance - FST	0.123*	0.283***	0.0389**	0.203***
	(0.0663)	(0.0558)	(0.0186)	(0.0519)
Observation	48827	48827	48827	48827

Panel C

	(1) Risk Aversion	(2) Patience	(3) Gambling	(4) Forward-Looking
Same Language Fam.	-0.283***	-0.199***	-0.162***	-0.229***
	(0.0870)	(0.0695)	(0.0447)	(0.0661)
Geographical Distance	0.221***	0.0463	0.0705**	0.0777***
	(0.0295)	(0.0287)	(0.0298)	(0.0255)
Genetical Distance - FST	0.151**	0.298***	0.0583***	0.214***
	(0.0710)	(0.0555)	(0.0199)	(0.0532)
Observation	48827	48827	48827	48827

Notes: Dependent variables are standardized differences between group i and group j. Panel A includes only country fixed effects for country i and country j. Panel B adds the number of neighbouring groups within 1,000 km, as a measure of regional fractionalization, and the lowest percentile mismatch, as a measure of the distribution of regional mismatches. Panel C adds the caloric suitability index before and after Columbus, negative and positive plow suitability before and after Columbus, average temperature, rainfall, altitude, ruggedness, distance from the coast, distance from water bodies, distance from the closest urban place, latitude, and longitude. Controls are included for both group i and group j. Conley HAC standard errors use a 1,000 km threshold. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.3 Markets and Institutions

Table 3 reports implications for markets, institutions, and decisions related to investment in human and physical capital and innovation. As usual, Panel A reports results with only country fixed effects for groups i and j; Panel B includes controls related to regional fractionalization and the distribution of mismatches within the region; and Panel C includes the full set of controls.

Both F_{ST} and the dummy for languages within the same family have significant effects on the intensity of oral traditions related to collective action, markets, investment, and innovation. Groups that are culturally closer have smaller differences in each outcome variable, and groups that are more genetically distant exhibit larger differences. For example, speaking a language from the same family reduces differences in collective action by 0.16 standard deviations in Panel A, while a one-standard-deviation increase in F_{ST} increases the difference by 0.09 standard deviations. Including the full set of controls, in Panel C, makes little difference: the dummy for languages within the same family decreases the difference in collective action by 0.16 standard deviations, and a one-standard-deviation increase in F_{ST} increases the difference by 0.08 standard deviations.

The only significant difference is for institutions, for which the proxy for cultural similarity has a robust and significant effect, while F_{ST} does not exert any significant effect either without or with the inclusion of additional controls.

This set of results is in line with previous results and the more general economic literature. Cooperation has often been considered an essential element of functioning institutions (Bowles & Gintis 2003), consistent with the significant effect of culture on cooperation and the insignificant effect of F_{ST} above. Cooperation, universal altruism, and generalized trust are considered essential for the existence of markets (Tabellini 2008*b*), and the significant effects of culture and genes on markets reflect their impact on these dimensions. Similarly, time preferences and risk aversion represent building blocks in investment and innovation decisions.

Table 3: Economic Outcomes

Panel A

	(1) Collective-action	(2) Market	(3) Institutions	(4) Investment	(5) Innovation
Same Language Fam.	-0.164***	-0.310***	-0.212***	-0.387***	-0.181***
	(0.0405)	(0.0819)	(0.0640)	(0.0691)	(0.0397)
Geographical Distance	0.215***	0.164***	0.339***	0.108***	0.0581***
	(0.0340)	(0.0334)	(0.0466)	(0.0170)	(0.0215)
Genetical Distance - FST	0.0938**	0.280***	0.0650	0.134***	0.0601**
	(0.0404)	(0.0713)	(0.0558)	(0.0507)	(0.0263)
Observation	51680	51680	51680	51680	51680

Panel B

	(1) Collective-action	(2) Market	(3) Institutions	(4) Investment	(5) Innovation
Same Language Fam.	-0.155***	-0.292***	-0.211***	-0.377***	-0.179***
	(0.0420)	(0.0844)	(0.0664)	(0.0710)	(0.0405)
Geographical Distance	0.213***	0.151***	0.375***	0.101***	0.0456**
	(0.0328)	(0.0336)	(0.0462)	(0.0189)	(0.0202)
Genetical Distance - FST	0.0880**	0.315***	-0.00216	0.178***	0.0899***
	(0.0393)	(0.0746)	(0.0527)	(0.0596)	(0.0236)
Observation	48827	48827	48827	48827	48827

Panel C

	(1) Collective-action	(2) Market	(3) Institutions	(4) Investment	(5) Innovation
Same Language Fam.	-0.155***	-0.301***	-0.207***	-0.362***	-0.194***
	(0.0407)	(0.0949)	(0.0673)	(0.0734)	(0.0351)
Geographical Distance	0.215***	0.152***	0.382***	0.0960***	0.0513***
	(0.0343)	(0.0319)	(0.0468)	(0.0196)	(0.0192)
Genetical Distance - FST	0.0838**	0.321***	-0.0152	0.196***	0.0812***
	(0.0415)	(0.0814)	(0.0542)	(0.0640)	(0.0240)
Observation	48827	48827	48827	48827	48827

Notes: Dependent variables are standardized differences between group i and group j. Panel A includes only country fixed effects for country i and country j. Panel B adds the number of neighbouring groups within 1,000 km, as a measure of regional fractionalization, and the lowest percentile mismatch, as a measure of the distribution of regional mismatches. Panel C adds the caloric suitability index before and

5.4 Economic Development

Previous sections have shown that culture and genes have significant effects on specific preferences, beliefs, and behaviours, which then translate into essential features of a functioning economic system. To estimate the wider implications for economic development, I look at the impact of each on levels of GDP, proxied by nighttime-light data averaged over 2003–2012. Control variables are entered gradually. Model 1 shows the impact of cultural, geographic, and genetic distance, controlling only for country fixed effects. Model 2 enters proxies for regional fractionalization and the distribution of mismatches. Finally, Model 3 includes the full set of controls.

Both genes and culture have a significant effect on the level of development, and this effect is robust to the inclusion of additional controls. Culturally similar groups have differences in levels of GDP that are about 0.18 standard deviations smaller, while a one-standard-deviation increase in F_{ST} increases differences in levels of development by almost 0.09 standard deviations. Interestingly, geographic distance does not exert a robust effect on differences in levels of development, which should reduce concerns about spatial correlation in lights. If lights data were spatially correlated and neighbouring groups were more likely to be culturally and genetically closer, then it would be possible to overestimate the effect. However, this issue can be excluded, given that geographic distance is not associated with differences in economic development, which otherwise should have dominated the other effects.

Table 4: Levels of Development

	(1) Mean Lights	(2) Mean Lights	(3) Mean Lights
Same Language Fam.	-0.188***	-0.174***	-0.177***
	(0.0386)	(0.0393)	(0.0399)
Geographical Distance	0.0289**	0.000451	0.00950
	(0.0123)	(0.0116)	(0.0118)
Genetical Distance	0.0412***	0.104***	0.0914***
	(0.0137)	(0.0175)	(0.0151)
Observation	71630	68264	68264
Country i FE	Yes	Yes	Yes
Country j FE	Yes	Yes	Yes

Notes: The dependent variable is the standardized difference between group i and group j. Model 1 includes only country fixed effects for country i and country j. Model 2 adds the number of neighbouring groups within 1,000 km, as a measure of regional fractionalization, and the lowest percentile mismatch, as a measure of the distribution of regional mismatches. Model 3 adds the caloric suitability index before and after Columbus, negative and positive plow suitability before and after Columbus, average temperature, rainfall, altitude, ruggedness, distance from the coast, distance from water bodies, distance from the closest urban place, latitude, and longitude. Controls are included for both group i and group j. Conley HAC standard errors use a 1,000 km threshold. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

6 Robustness Checks

6.1 Using Language Genealogical Distance

Cultural distance in the previous section is captured using a dummy for languages within the same family, to capture inherited traditions that are likely to be transmitted through generations despite small variations in languages over time across culturally similar groups. However, it is also possible that cultural transmission decays as the distance between languages increases. As a result, two languages that are closer in the linguistic tree may be culturally closer than languages that are farther from each other.

To test this hypothesis, I replace the dummy for languages within the same family with a measure of genealogical distance between languages. Genealogical distance considers the distance between languages in the linguistic tree and whether they belong to the same family.

Beyond the conceptual difference—that is, culture persists versus culture decays—there are practical advantages and disadvantages to using genealogical distance compared with a dummy. A dummy represents a clear-cut measure that is unlikely to be affected by measurement error and noise in the deeper branches of the classification, and is therefore less prone to attenuation bias. However, it discards substantial variation. Genealogical distance, on the other hand, captures much more variation and the idea that cultural transmission decays with distance between languages. However, it relies more heavily on whether linguistic trees are truly representative of the distance between languages[6] and on the assumptions made to generate the measure.

Table C1 - C4 in Appendix C show results using this alternative measure. Table C1 reports results related to cooperation and associated preferences, beliefs, and mechanisms. Table C2 shows results for time preferences and risk aversion, while Table C3 reports results for economic outcomes. Differences in levels of development are reported in Table C4. There are a few small general differences in the effect of culture when using genealogical distance, which are consistent with attenuation bias. For example, culture has a significant effect on cooperation in Panel A, but the effect on prosociality, pure altruism, and norms related to reputation and punishment is barely significant. Results in Panels B and C, on the other hand, confirm the statistically significant impact of culture on all dependent variables considered. No significant differences emerge in the associated effect of F_{ST}.

Hence, despite marginal differences, the two measures provide a similar picture of their estimated impact on preferences, behaviours, and economic outcomes.

6.2 Double Clustering

As mentioned in the previous section, unlike Becker et al. (2020), I use Conley HAC robust standard errors to capture the idea that cultural transmission and the original population movements that generated genetic differences are not bounded by national borders. Table C5-C8 in Appendix C reports results using double-clustered standard errors at the country level, similar to Becker (2025).

Again, there are minor differences when using double-clustered standard errors, but these do not substantially change the overall picture. Culture exerts a significant effect on cooperation, preferences, and economic outcomes. The only difference is related to the insignificant effect of culture on prosociality and reciprocity in Panel A, while all other effects are confirmed. The effect of genetic distance, however, seems much weaker, despite F_{ST} retaining a significant effect on levels of development. More specifically, F_{ST} exerts

---

[6] For instance, two different languages can be closer or farther from each other, but they are always represented by new branches in the linguistic tree. Thus, new branches do not fully capture linguistic distance.

a significant effect on generalized altruism and trust in Panel A, consistent with previous results. However, the effect on time preferences and risk aversion in Panel B is barely significant, which is then reflected in a relatively marginal effect on collective action, institutions, investment, and innovation. The effect on markets, however, is retained.

References

Ager, P. & Brueckner, M. (2018), 'Immigrants' genes: Genetic diversity and economic development in the united states', *Economic Inquiry* **56**, 1149–1164.

Algan, Y. & Cahuc, P. (2010), 'Inherited trust and growth', *American Economic Review* **100**(5), 2060–2092.

Anderson, S. (2025), Gender biased norms, CEPR Discussion Paper 20211, CEPR Press, Paris and London.

Arbatli, C. E., Ashraf, Q. H., Galor, O. & Klemp, M. (2020), 'Diversity and conflict', *Econometrica* **88**(2), 727–797.

Ashraf, Q. H. & Galor, O. (2013*a*), 'Genetic diversity and the origins of cultural fragmentation', *American Economic Review* **103**(3), 528–533.

Ashraf, Q. H. & Galor, O. (2013*b*), 'The "out of africa" hypothesis, human genetic diversity, and comparative economic development', *American Economic Review* **103**(1), 1–46.

Ashraf, Q. H. & Galor, O. (2018), 'The macrogenoeconomics of comparative development', *Journal of Economic Literature* **56**(3), 1119–1155.

Atkinson, Q. D. (2011), 'Phonemic diversity supports a serial founder effect model of language expansion from africa', *Science* **332**(6027), 346–349.

Barbieri, C., Blasi, D. E., Arango-Isaza, E., Sotiropoulos, A. G., Hammarström, H., Wichmann, S., Greenhill, S. J., Gray, R. D., Forkel, R., Bickel, B. & Shimizu, K. K. (2022), 'A global analysis of matches and mismatches between human genetic and linguistic histories', *Proceedings of the National Academy of Sciences* **119**(47).

Becker, A. (2025), 'On the economic origins of concerns over women's chastity', *Review of Economic Studies* **92**(4), 2303–2329.

Becker, A., Enke, B. & Falk, A. (2020), Ancient origins of the global variation in economic preferences, *in* 'AEA Papers and Proceedings', Vol. 110, pp. 319–323.

Berezkin, Y. (2015), 'Folklore and mythology catalogue: Its lay-out and potential for research', *The Retrospective Methods Network Newsletter* **10**, 58.

Bhui, R., Chudek, M. & Henrich, J. (2019), 'How exploitation launched human cooperation', *Behavioral Ecology and Sociobiology* **73**(78), 1–14.

Bisin, A., Topa, G. & Verdier, T. (2004), 'Cooperation as a transmitted cultural trait', *Rationality and Society* **16**(4), 477–507.

Bisin, A. & Verdier, T. (2000), 'Beyond the melting pot: Cultural transmission, marriage, and the evolution of ethnic and religious traits', *Quarterly Journal of Economics* **115**(3), 955–988.

Bisin, A. & Verdier, T. (2001), 'The economics of cultural transmission and the evolution of preferences', *Journal of Economic Theory* **97**(2), 298–319.

Bove, V. & Gokmen, G. (2017), 'Genetic distance, trade, and the diffusion of development', *Journal of Development Economics* **127**, 251–268.

Bowles, S. & Gintis, H. (2003), The genetic and cultural evolution of cooperation, *in* P. Hammerstein, ed., 'Genetic and Cultural Evolution of Cooperation', MIT Press, Cambridge, MA, pp. 429–444.

Boyd, R. & Richerson, P. J. (1985), *Culture and the Evolutionary Process*, University of Chicago Press.

Cao, Y., Enke, B., Falk, A., Giuliano, P. & Nunn, N. (2021), Herding, armed conflict, and a culture of honor: Global evidence, NBER Working Paper 29250, National Bureau of Economic Research.

Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994), *The History and Geography of Human Genes*, Princeton University Press, Princeton.

Chen, M. K. (2013), 'The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets', *American Economic Review* **103**(2), 690–731.

Chen, M., Lakshminarayanan, V. & Santos, L. (2006), 'How basic are behavioral biases? evidence from capuchin monkey trading behavior', *Journal of Political Economy* **114**(3), 517–537.

Chudek, M. & Henrich, J. (2011), 'Culture-gene coevolution, norm-psychology and the emergence of human prosociality', *Trends in Cognitive Sciences* **15**(5), 218–226.

Chudek, M., Muthukrishna, M. & Henrich, J. (2015), Cultural evolution, *in* D. M. Buss, ed., 'The Handbook of Evolutionary Psychology', 2 edn, Vol. 2, John Wiley and Sons.

Chudek, M., Zhao, W. & Henrich, J. (2013), Culture-gene coevolution, large-scale cooperation and the shaping of human social psychology, *in* R. Joyce, K. Sterelny & B. Calcott, eds, 'Signaling, Commitment, and Emotion', MIT Press, Cambridge, pp. 425–458.

Desmet, K., Weber, S. & Ortuño-Ortín, I. (2009), 'Linguistic diversity and redistribution', *Journal of the European Economic Association* **7**(6), 1291–1318.

Enke, B. (2019), 'Kinship, cooperation, and the evolution of moral systems', *Quarterly Journal of Economics* **134**(2), 953–1019.

Felbermayr, G. J. & Toubal, F. (2010), 'Cultural proximity and trade', *European Economic Review* **54**(2), 279–293.

Fernández, R. (2008), Culture and economics, *in* S. N. Durlauf & L. E. Blume, eds, 'The New Palgrave Dictionary of Economics', 2 edn, Palgrave Macmillan, Basingstoke.

Fernández, R. (2011), Does culture matter?, *in* J. Benhabib, M. O. Jackson & A. Bisin, eds, 'Handbook of Social Economics', Vol. 1A, North-Holland, Amsterdam, pp. 481–510.

Figlio, D., Giuliano, P., Özek, U. & Sapienza, P. (2019), 'Long-term orientation and educational performance', *American Economic Journal: Economic Policy* **11**(4), 272–309.

François, P. & Zabojnik, J. (2005), 'Trust, social capital, and economic development', *Journal of the European Economic Association* **3**(1), 51–94.

Galor, O., Klemp, M. & Wainstock, D. C. (2024), Roots of cultural diversity, Nber working paper, National Bureau of Economic Research.

Galor, O., Klemp, M. & Wainstock, D. C. (2025), The prehistoric dispersal from africa and the compression of within-group productive traits. Working Paper, Brown University.

Galor, O. & Moav, O. (2002), 'Natural selection and the origin of economic growth', *Quarterly Journal of Economics* **117**(4), 1133–1191.

Galor, O. & Özak, Ö. (2016), 'The agricultural origins of time preference', *American Economic Review* **106**(10), 3064–3103.

Galor, O., Özak, Ö. & Sarid, A. (2020), 'Linguistic traits and human capital formation', *AEA Papers and Proceedings* **110**, 309–13.

Galor, O., Özak, Ö. & Sarid, A. (2025), 'Roots of cultural and linguistic traits*', *The Economic Journal* p. ueaf117.

Galor, O. & Savitskiy, V. (2018), Climatic roots of loss aversion, Technical Report 25273, National Bureau of Economic Research.

Giavazzi, F., Petkov, I. & Schiantarelli, F. (2019), ‘Culture: Persistence and evolution’, *Journal of Economic Growth* **24**, 117–154.

Giuliano, P. & Nunn, N. (2018), ‘Ancestral characteristics of modern populations’, *Economic History of Developing Regions* **33**(1), 1–17.

Giuliano, P. & Nunn, N. (2021), ‘Understanding cultural persistence and change’, *Review of Economic Studies* **88**(4), 1541–1581.

Gokmen, G. (2017), ‘Clash of civilizations and the impact of cultural differences on trade’, *Journal of Development Economics* **127**, 449–458.

Greif, A., Mokyr, J. & Tabellini, G. (2025), *Two Paths to Prosperity: Culture and Institutions in Europe and China, 1000–2000*, Princeton University Press.

Grosjean, P. (2014), ‘A history of violence: The culture of honor as a determinant of homicide in the us south’, *Journal of the European Economic Association* **12**(5), 1285–1316.

Guarnieri, E. & Tur-Prats, A. (2023), ‘Cultural distance and conflict-related sexual violence’, *Quarterly Journal of Economics* **138**(3), 1817–1861.

Guiso, L., Sapienza, P. & Zingales, L. (2006), ‘Does culture affect economic outcomes?’, *Journal of Economic Perspectives* **20**(2), 23–48.

Henderson, J. V., Storeygard, A. & Weil, D. N. (2012), ‘Measuring economic growth from outer space’, *American Economic Review* **102**(2), 994–1028.

Henrich, J. (2004), ‘Cultural group selection, coevolutionary processes and large-scale cooperation’, *Journal of Economic Behavior and Organization* **53**(1), 3–35.

Henrich, J. (2018), ‘Human cooperation: The hunter-gatherer puzzle’, *Current Biology* **28**(19), R1143–R1145.

Henrich, J. (2021), Game theory in cultural evolution, *in* ‘One Hundred Years of Game Theory: A Nobel Symposium’, Cambridge University Press, pp. 357–374.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R. et al. (2005), ‘Models of decision-making and the coevolution of social preferences’, *Behavioral and Brain Sciences* **28**(6), 838–855.

Henrich, J. & Henrich, N. (2006), ‘Culture, evolution and the puzzle of human cooperation’, *Cognitive Systems Research* **7**(2), 220–245.

Michalopoulos, S. & Xue, M. M. (2021), 'Folklore*', *The Quarterly Journal of Economics* **136**(4), 1993–2046.

Mokyr, J. (2014), Culture, institutions, and modern growth, *in* S. Galiani & I. Sened, eds, 'Institutions, Property Rights, and Economic Growth: The Legacy of Douglass North', Cambridge University Press, pp. 151–191.

Montalvo, J. G. & Reynal-Querol, M. (2005), 'Ethnic polarization, potential conflict, and civil wars', *American Economic Review* **95**(3), 796–816.

Moscona, J., Nunn, N. & Robinson, J. A. (2020), 'Segmentary lineage organization and conflict in sub-saharan africa', *Econometrica* **88**(5), 1999–2036.

Nisbett, R. E. (1993), 'Violence and u.s. regional culture', *American Psychologist* **48**(4), 441–449.

North, D. C. (1991), 'Institutions', *Journal of Economic Perspectives* **5**(1), 97–112.

Posch, M., Schulz, J. & Henrich, J. (2026), 'How cultural diversity drives innovation: Surnames and patents in us history', *Journal of Political Economy* **134**(4).

Prugnolle, F., Manica, A. & Balloux, F. (2005), 'Geography predicts neutral genetic diversity of human populations', *Current Biology* **15**(5), R159–R160.

Putterman, L. & Weil, D. N. (2010), 'Post-1500 population flows and the long-run determinants of economic growth and inequality', *Quarterly Journal of Economics* **125**(4), 1627–1682.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W. & Cavalli-Sforza, L. L. (2005), 'Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in africa', *Proceedings of the National Academy of Sciences* **102**(44), 15942–15947.

Richerson, P. & Henrich, J. (2012), 'Tribal social instincts and the cultural evolution of institutions to solve collective action problems', *Cliodynamics* **3**(1).

Savitskiy, V., Mu, R. & Henrich, J. (2026), 'Cultural evolution for economists'. Mimeo.

Spolaore, E. & Wacziarg, R. (2009), 'The diffusion of development', *Quarterly Journal of Economics* **124**(2), 469–529.

Tabellini, G. (2008*a*), 'Presidential address: Institutions and culture', *Journal of the European Economic Association* **6**(2–3), 255–294.

Tabellini, G. (2008*b*), 'The scope of cooperation: Values and incentives', *Quarterly Journal of Economics* **123**, 905–950.

Tabellini, G. (2010), 'Culture and institutions: Economic development in the regions of europe', *Journal of the European Economic Association* **8**(4), 677–716.

Toward an understanding of the development of time preferences: Evidence from field experiments (2019), *Journal of Public Economics* **177**, 104039.

Voigtländer, N. & Voth, H.-J. (2012), 'Persecution perpetuated: The medieval origins of anti-semitic violence in nazi germany', *Quarterly Journal of Economics* **127**(3), 1339–1392.

Wright, S. (1951), 'The genetical structure of populations', *Annals of Eugenics* **15**(1), 323–354.

Appendix A - Data

Queries to generate Preferences, Beliefs, and Economic Outcomes

To generate the dependent variables, I use a fully AI-assisted procedure. I begin by feeding all concept-related terms in the Folklore dataset into AI and then asking it to generate variables by selecting terms relevant to each dependent variable.

To generate a proxy for cooperation, I ask AI to *"choose all relevant terms in the Folklore dataset that can be associated with economic-theory-based preferences for cooperation in public-good games"*. From the resulting list, I select the recommended core subset of terms. I then ask AI to group these terms into: 1) pure altruism; 2) warm glow or intrinsic giving; 3) reciprocity; 4) social norms related to reputation, punishment, and signalling; and 5) generalized trust.

I repeat the same procedure for all other variables generated using the Folklore dataset. For example, to generate a proxy for institutions, I use the definition in **?** and ask AI to *"consider the following definition of institutions as the rules of the game in a society - humanly devised constraints that shape human interaction, incentives, and economic outcomes"*. To generate a proxy for markets, I ask AI to *"consider markets as institutions in which individuals or collective agents exchange goods and services. They usually use money as a medium of exchange, which leads to the formation of prices"*.

The entire script is available at: `https://chatgpt.com/share/69cc51f2-fb0c-838a-beb1-b3e8bec73543`.

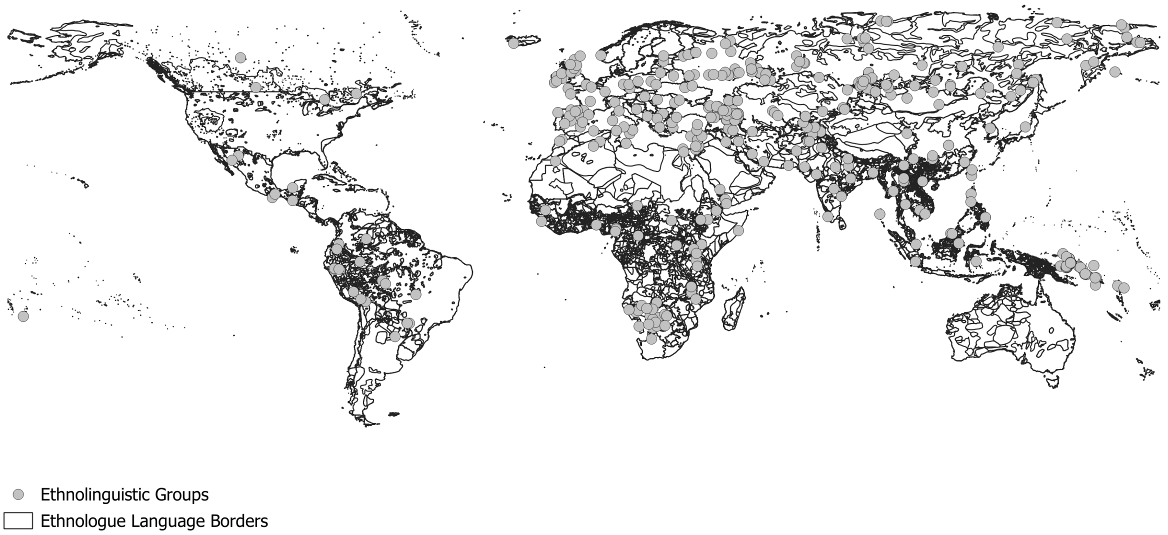
List of Motifs related concepts

Variable	Motifs related concepts
Cooperation	help_related support_related share_related together_related community_related group_related social_related trust_related fair_related equality_related responsibility_related obligation_related duty_related contribution_related volunteer_related donate_related charity_related participation_related collective_related union_related collaborate_related coordination_related agreement_related mutual_related reciprocal_related exchange_related justice_related punish_related punishment_related sanction_related reputation_related honor_related guilt_related shame_related protect_related service_related sacrifice_related loyalty_related peace_related compromise_related reconciliation_related religion_related faith_related ritual_related tribe_related clan_related village_related neighbor_related
Prosociality	help_related share_related support_related trust_related fair_related equality_related donate_related charity_related
Collective-action	community_related agreement_related mutual_related
Pure altruism	help_related aid_related care_related support_related protect_related rescue_related service_related sacrifice_related charity_related donate_related compassion_related sympathy_related kindness_related generous_related
Warm glow / intrinsic giving	gratitude_related love_related volunteer_related together_related belong_related community_related social_related friendship_related loyalty_related harmony_related faith_related ritual_related)
Reciprocity	reciprocal_related mutual_related exchange_related trade_related return_related repay_related reward_related favor_related agreement_related contract_related partnership_related collaborate_related coordination_related contribution_related
Social norms: reputation punishment signaling	reputation_related honor_related approval_related disapproval_related blame_related respect_related status_related punish_related punishment_related penalty_related sanction_related justice_related law_related rule_related enforcement_related obligation_related duty_related responsibility_related guilt_related shame_related signal_related display_related public_related ceremony_related pledge_related

Generalized trust	trust_related fair_related community_related social_related public_related common_related together_related shared_related peace_related welcome_related hospitality_related integrity_related
*Forward-looking behavior	future_related plan_related planning_related prepare_related preparation_related strategy_related goal_related anticipate_related predict_related forecast_related consider_related calculate_related evaluate_related estimate_related investment_related saving_related longterm_related
*Aversion to risk	risk_related safe_related safety_related danger_related fear_related afraid_related worry_related uncertainty_related caution_related careful_related warn_related prevent_related protect_related security_related avoid_related loss_related
Markets	market_related exchange_related trade_related buy_related sell_related money_related pay_related payment_related price_related cost_related value_related contract_related agreement_related deal_related
Institutions	rule_related law_related regulation_related obligation_related duty_related responsibility_related enforce_related punishment_related sanction_related justice_related compliance_related contract_related agreement_related government_related authority_related organization_related institutional_related
Patience	future_related wait_related later_related longterm_related plan_related planning_related prepare_related consider_related calculate_related evaluate_related investment_related saving_related
Gambling	gamble_related gambling_related bet_related lottery_related casino_related poker_related dice_related
Investment	investment_related invest_related capital_related saving_related fund_related funding_related financial_related profit_related return_related future_related longterm_related
Innovation	invent_related invention_related inventor_related create_related creative_related creativity_related discover_related discovery_related experiment_related improve_related adapt_related modify_related transform_related new_related breakthrough_related

Appendix B

Figure B1: Distribution of Groups

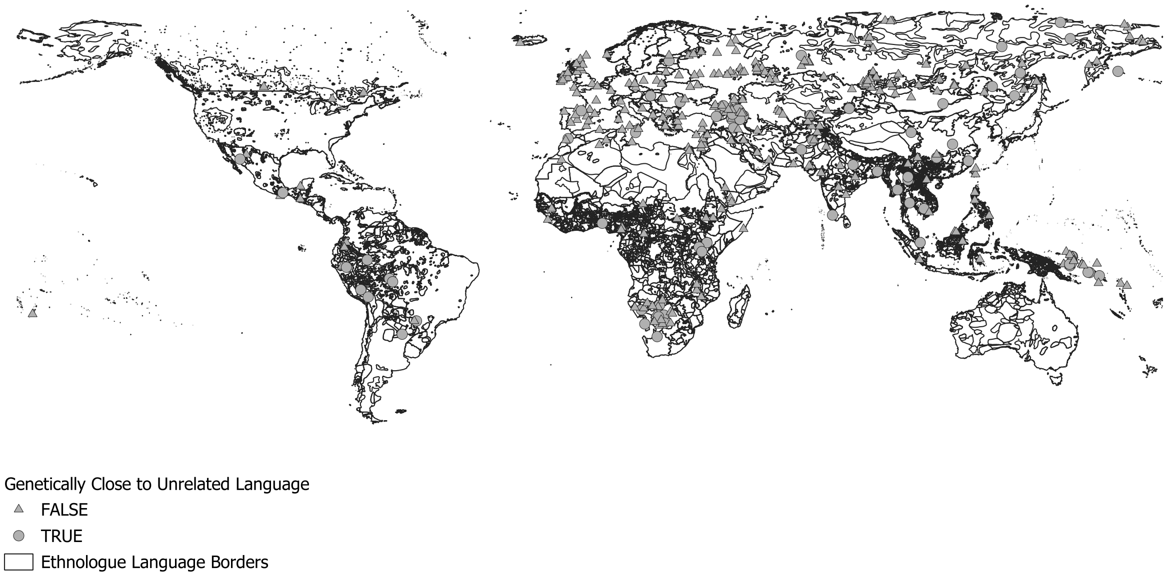


Notes: The figure shows the distribution of ethnolinguistic groups in the dataset, with dots representing centroid coordinates.

Figure B2: Distribution of Groups Speaking Languages of Different Families

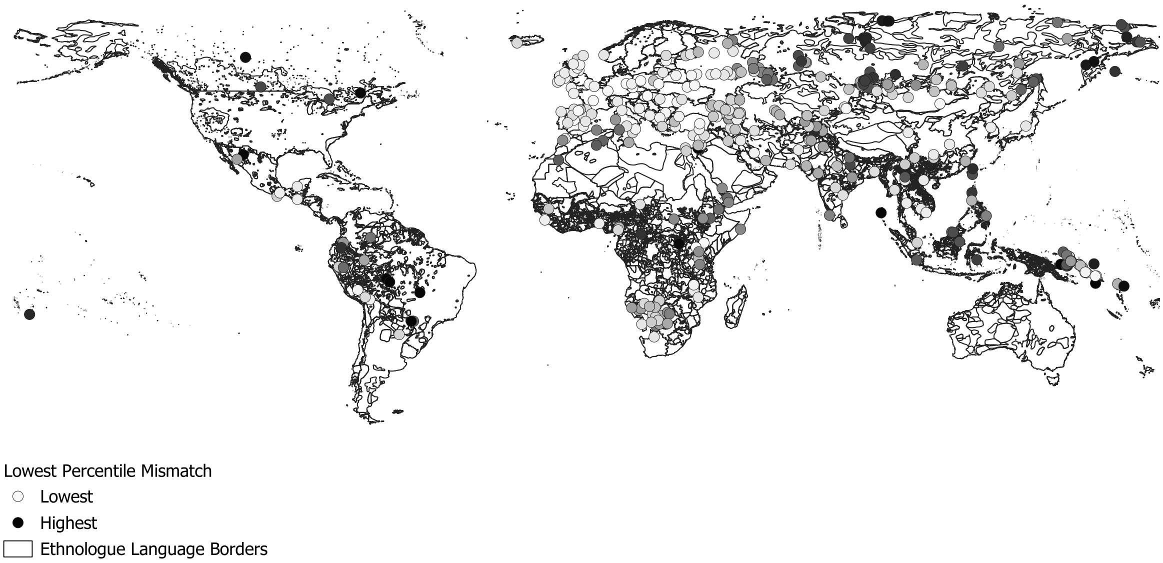
Notes: The figure shows the distribution of groups speaking languages of different families within a 1,000 km radius. Gradients of black represent an increasing number.

Figure B3: Groups Genetically Close to Unrelated Language Families



Notes: The figure shows the distribution of groups which are genetically close to groups speaking a language within a different language family.

Figure B4: Probability of Mismatch



Notes: The figure shows the distribution (in percentiles) of mismatches between cultural and biological histories.

Appendix C

Table C1 - Cooperation and Preferences Using Genealogical Distance

	(1) Cooperation	(2) Prosociality	(3) Altruism	(4) Warm Glow	(5) Reciprocity	(6) Trust	(7) Reputation
Genealogical Distance	0.371***	0.330*	0.107	0.653***	0.487***	1.023***	0.0172
	(0.111)	(0.176)	(0.152)	(0.139)	(0.136)	(0.121)	(0.155)
Geographical Distance	0.359***	0.204***	0.129***	0.155***	0.189***	0.194***	0.313***
	(0.0335)	(0.0325)	(0.0307)	(0.0291)	(0.0324)	(0.0265)	(0.0344)
Genetical Distance - FST	0.0885*	0.0206	0.119***	0.115***	0.0704	0.193***	0.118**
	(0.0472)	(0.0384)	(0.0378)	(0.0361)	(0.0488)	(0.0373)	(0.0509)
Observation	50402	50402	50402	50402	50402	50402	50402

Notes: Models include only country fixed effects for country i and country j. All dependent variables represent standardized differences between groups. Conley HAC standard errors use a 1,000 km threshold. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C2 - Risk Aversion and Patience Using Genealogical Distance

	(1) Risk Aversion	(2) Patience	(3) Gambling	(4) Forward Looking
Genealogical Distance	0.734***	0.347**	0.305***	0.648***
	(0.132)	(0.136)	(0.105)	(0.161)
Geographical Distance	0.212***	0.0808***	0.0944***	0.0902***
	(0.0310)	(0.0250)	(0.0329)	(0.0213)
Genetical Distance - FST	0.159**	0.258***	0.0154	0.189***
	(0.0636)	(0.0603)	(0.0203)	(0.0502)
Observation	50402	50402	50402	50402

Notes: Models include only country fixed effects for country i and country j. All dependent variables represent standardized differences between groups. Conley HAC standard errors use a 1,000 km threshold. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C3 - Other Outcomes Using Genealogical Distance

	(1) Collective-action	(2) Market	(3) Institutions	(4) Investment	(5) Innovation
Genealogical Distance	0.425**	0.477***	0.316**	0.598***	0.721***
	(0.193)	(0.163)	(0.157)	(0.139)	(0.134)
Geographical Distance	0.203***	0.165***	0.343***	0.111***	0.0458**
	(0.0309)	(0.0334)	(0.0460)	(0.0175)	(0.0200)
Genetical Distance - FST	0.115***	0.320***	0.0875*	0.174***	0.0882***
	(0.0371)	(0.0679)	(0.0520)	(0.0510)	(0.0246)
Observation	50402	50402	50402	50402	50402

Notes: Models include only country fixed effects for country i and country j. All dependent variables represent standardized differences between groups. Conley HAC standard errors use a 1,000 km threshold. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C4 - Income Differences Using Genealogical Distance

	(1) Mean Lights	(2) Mean Lights	(3) Mean Lights
Genealogical Distance	0.587***	0.542***	0.659***
	(0.102)	(0.102)	(0.125)
Geographical Distance	0.0223**	-0.00204	0.00139
	(0.0110)	(0.0114)	(0.0109)
Genetical Distance - FST	0.0751***	0.128***	0.117***
	(0.0171)	(0.0207)	(0.0185)
Observation	70124	66794	66794

Notes: Model 1 only includes country fixed effects for country i and country j. The number of neighbouring groups within 1,000 km, as a measure of regional fractionalization, and the lowest percentile mismatch, as a measure of the distribution of regional mismatches are added in Model 2. Model 3 also includes the caloric suitability index before and after Columbus, negative and positive plow suitability before and after Columbus, average temperature, rainfall, altitude, ruggedness, distance from the coast, distance from water bodies, distance from the closest urban place, latitude, and longitude. Controls are included for both group i and group j. The dependent variables is the standardized differences between groups. Conley HAC standard errors use a 1,000 km threshold. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C5 - Cooperation and Preference - Double Clustering

	(1) Cooperation	(2) Prosociality	(3) Altruism	(4) Warm Glow	(5) Reciprocity	(6) Trust	(7) Reputation
Same Language Fam.	-0.151**	-0.119	-0.135**	-0.147***	-0.132	-0.0942***	-0.331***
	(0.0717)	(0.0737)	(0.0579)	(0.0392)	(0.0862)	(0.0357)	(0.0815)
Geographical Distance	0.400***	0.216***	0.131***	0.188***	0.238***	0.242***	0.318***
	(0.0496)	(0.0433)	(0.0446)	(0.0424)	(0.0452)	(0.0405)	(0.0554)
Genetical Distance - FST	0.0246	0.00518	0.114**	0.0609	0.0119	0.147**	0.0394
	(0.0796)	(0.0591)	(0.0498)	(0.0759)	(0.0920)	(0.0560)	(0.0815)
Adj.R-squared	0.153	0.046	0.044	0.065	0.052	0.100	0.128
Observation	48827	48827	48827	48827	48827	48827	48827

Notes: Models include only country fixed effects for country i and country j. All dependent variables represent standardized differences between groups. Standard errors are double-clustered at country i and country j. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C6 - Risk Aversion and Patience - Double Clustering

	(1) Risk Aversion	(2) Patience	(3) Gambling	(4) Forward-Looking
Same Language Fam.	-0.265***	-0.147***	-0.190***	-0.230***
	(0.0936)	(0.0392)	(0.0576)	(0.0654)
Geographical Distance	0.226***	0.188***	0.0735**	0.0776***
	(0.0325)	(0.0424)	(0.0352)	(0.0211)
Genetical Distance - FST	0.122	0.0609	0.0292	0.200***
	(0.0947)	(0.0759)	(0.0184)	(0.0727)
Adj.R-squared	0.081	0.065	0.014	0.052
Observation	48827	48827	48827	48827

Notes: Models include only country fixed effects for country i and country j. All dependent variables represent standardized differences between groups. Standard errors are double-clustered at country i and country j. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C7 - Other Outcomes - Double Clustering

	(1) Collective-action	(2) Market	(3) Institutions	(4) Investment	(5) Innovation
Same Language Fam.	-0.150***	-0.296***	-0.213**	-0.379***	-0.182***
	(0.0238)	(0.110)	(0.0829)	(0.121)	(0.0485)
Geographical Distance	0.198***	0.149***	0.367***	0.100***	0.0529
	(0.0397)	(0.0357)	(0.0685)	(0.0243)	(0.0323)
Genetical Distance - FST	0.116	0.307***	0.00451	0.171*	0.0752*
	(0.0734)	(0.0964)	(0.0807)	(0.0921)	(0.0417)
Adj.R-squared	0.071	0.129	0.123	0.073	0.019
Observation	48827	48827	48827	48827	48827

Notes: Models include only country fixed effects for country i and country j. All dependent variables represent standardized differences between groups. Standard errors are double-clustered at country i and country j. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C8 - Income Differences - Double Clustering

	(1) Collective-action	(2) Market	(3) Institutions
Same Language Fam.	-0.164***	-0.155***	-0.162***
	(0.0510)	(0.0512)	(0.0518)
Geographical Distance	0.000496	-0.0103	-0.00822
	(0.0177)	(0.0136)	(0.0129)
Genetical Distance - FST	0.0657***	0.0931***	0.0908***
	(0.0229)	(0.0204)	(0.0206)
Adj.R-squared	0.013	0.028	0.067
Observation	48827	48827	48827

Notes: Model 1 only includes country fixed effects for country i and country j. The number of neighbouring groups within 1,000 km, as a measure of regional fractionalization, and the lowest percentile mismatch, as a measure of the distribution of regional mismatches are added in Model 2. Model 3 also includes the caloric suitability index before and after Columbus, negative and positive plow suitability before and after Columbus, average temperature, rainfall, altitude, ruggedness, distance from the coast, distance from water bodies, distance from the closest urban place, latitude, and longitude. Controls are included for both group i and group j. The dependent variables is the standardized differences between groups. Standard errors are double-clustered at country i and country j. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.