

Department of Economics

Ancestral Composition Distance

Mehmet Burak Akren, Eren Arbatlı, Sabri Arhan Ertan

Working Paper No. 16, 2026

Department Economics

Durham University Business School

Mill Hill Lane

Durham DH1 3LB, UK

Tel: +44 (0)191 3345200

<https://www.durham.ac.uk/business/about/departments/economics/>

© Durham University, 2026

Ancestral Composition Distance

Mehmet Burak Akren*, Eren Arbatli†, Sabri Arhan Ertan‡

Abstract

This paper introduces *Ancestral Composition Distance* (ACD); a bilateral measure of how similarly two countries assemble their populations from a common set of ancestral clusters. Each cluster contains countries whose native populations were culturally close around 1500. Across a global sample of country dyads (involving 163 countries), greater ancestral composition distance is systematically associated with greater contemporary cultural distance and human capital gap. More ancestrally distant countries also exhibit larger differences in income, human development, institutional characteristics, and foreign-policy positions, while trading less and exchanging fewer migrants. These relationships survive an extensive set of geographic, historical, institutional, genetic, linguistic, and religious controls, indicating that *ACD* captures a dimension of bilateral relatedness not summarized by standard measures. Decomposing *ACD* into differences already present in 1500 and differences that emerged due to post-Columbian population movements, we show that pairs whose ancestral compositions converged after 1500 are culturally closer today, conditional on their initial ancestral separation. All these results are robust to alternative ways to group countries into pre-1500 cultural clusters. We view *ACD* both as a useful measure for understanding contemporary cultural differences as well as a historically determined bilateral resistance term that complements conventional measures of geography, shared history, and population relatedness.

JEL Classification: F22, Z13, N30, O15, F14.

Keywords: ancestral distance; cultural distance; historical persistence, World Migration Matrix; population composition

*University of Duisburg-Essen, Department of Economics, 45141 Essen, Germany. mehmet.akren@vwl.uni-due.de

†Durham University, Department of Economics, DH1-3LB Durham, UK. eren.arbatli@durham.ac.uk

‡Boğaziçi University, Department of International Trade, 34342 Istanbul, Turkey. arhan.ertan@bogazici.edu.tr

1 Introduction

Five centuries of migration have left the world’s populations assembled in strikingly different proportions from the same set of pre-Columbian origins observed around 1500s. Brazil descends from European, Sub-Saharan African, and indigenous American ancestors in a mixture that has no counterpart anywhere in the Old World; the United States and Australia each draw on several ancestral spheres in proportions of their own; Japan and Ethiopia draw on essentially one. Neighbours on a map may descend from populations that an ocean separated in 1500, and countries on opposite sides of the world may share most of their ancestral heritage.

Since cultural traits and behavioural dispositions are transmitted across generations, similarities in ancestral composition may continue to shape interactions between societies long after the original populations have mixed, migrated, and dispersed. The extent to which countries with similar ancestral compositions are also culturally closer today informs us about the long-run persistence of culturally transmitted traits.

Departing from this simple observation, we measure the similarity in the ways two countries assembled their populations from a common set of pre-1500 origins, partly as a legacy of the history of world populations in the pre-Columbian period and partly as a result of the post-1500 population movements. We then evaluate whether this relational ancestral similarity predicts how far apart two countries stand in culture and human capital, in income and institutions, how much they trade, how much migration happens between them and how aligned their foreign policy preferences are.

To answer these questions, we introduce a novel dyadic metric that we call *Ancestral Composition Distance* (ACD). Its construction involves three main steps:

1. **Define Pre-1500 Cultural Clusters (Ancestral Spheres):** We first group pre-Columbian origin populations into culturally coherent historical spheres whose member populations were close to one another around 1500 CE and clearly separated from those of other spheres. We construct ACD under two alternative sets of cultural country clusters. Our primary clustering method partitions 1500 origin populations into ten ancestral spheres using Partitioning Around Medoids (PAM) applied to the year-1500 pairwise genetic distance matrix (F_{ST}) of Spolaore and Wacziarg (2018). Following the historical persistence literature (Spolaore and Wacziarg, 2009; Becker et al., 2020), genetic distance in 1500 serves as an objective proxy for general cultural and genealogical separation prior to post-Columbian migrations (detailed in Section 3.1). To confirm that our results do not hinge on the use of genetic data or the clustering algorithm we applied, we also construct an independent, non-algorithmic classification that aggregates ancestry shares into nine broad historical-cultural spheres.¹ Derived from historical, institutional, religious, and linguistic literature

¹These spheres are Western Christian, Eastern Christian, Islamic/Arab, Turkic/Persianate, East

without utilizing genetic data at any stage, this alternative measure yields very similar empirical patterns (detailed in Appendix C.2).

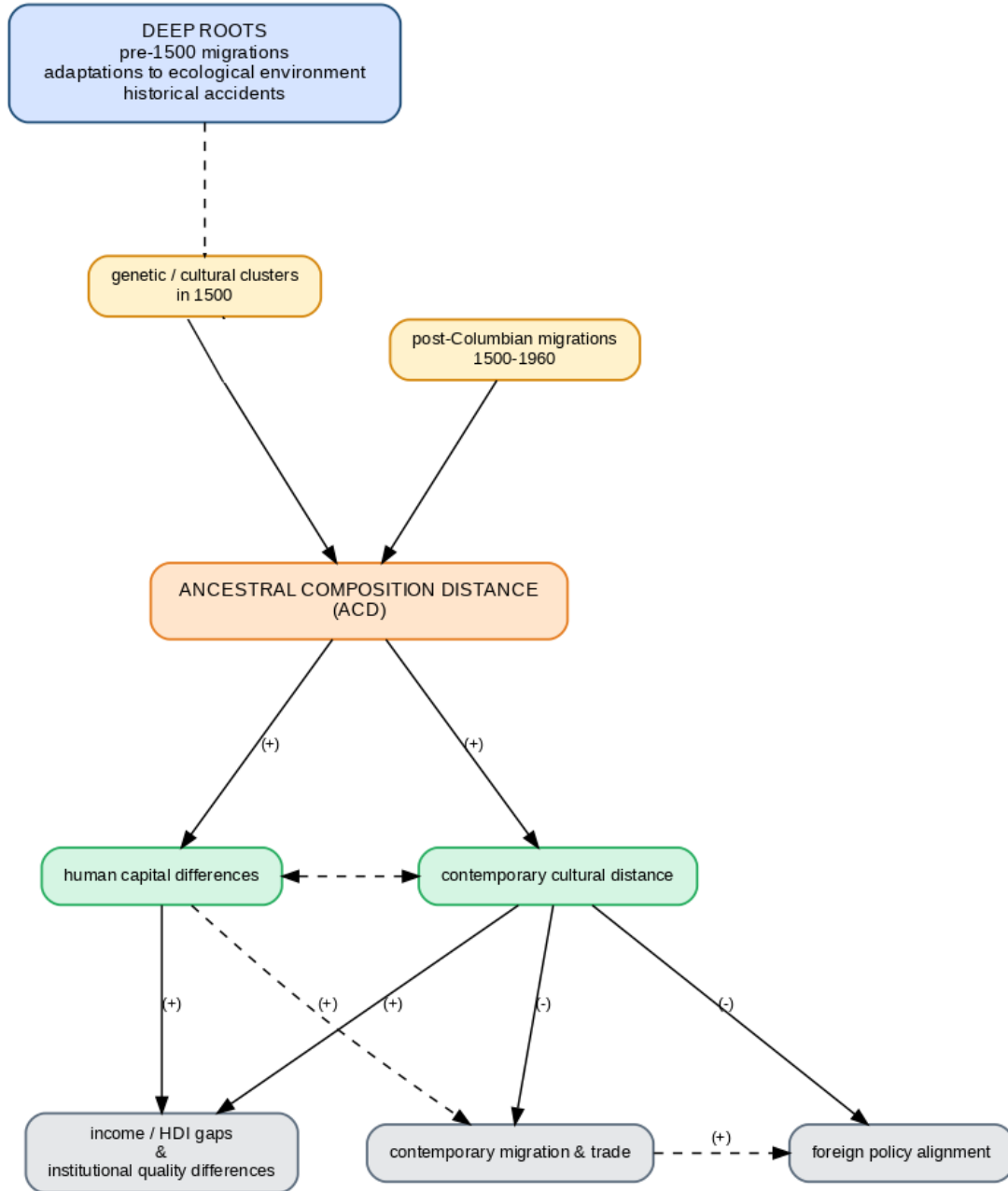
2. **Construct Country-level Ancestry Profiles:** For each modern nation-state, we compute a multi-dimensional demographic profile vector representing the shares of its population descending from each ancestral sphere. We extract these ancestral shares from the World Migration Matrix (WMM) of Putterman and Weil (2010) but using the 1960 population benchmark variant offered by Chanda et al. (2014). The use of 1960 benchmark –instead of the year 2000 benchmark– helps overcome reverse causation from the contemporary outcomes we study and partly mitigate endogeneity bias.
3. **Compute a Bilateral Composition Distance:** Finally, we evaluate the dissimilarity between country pairs’ ancestry profile vectors. Our headline formulation evaluates direct compositional overlap, i.e. the cosing distance between ancestry profiles identified in the previous step. We call this the *PAM10 Cosine Distance*. In this formulation, two ancestry profiles occupying different clusters are maximally distant however those clusters are related. To relax this restrictive assumption, we offer an alternative formulation of ACD (*WMM10 Hellinger–Wasserstein (HW) Distance*) which gives partial credit to ancestry in different clusters to the extent descendants from those clusters ended up in similar places.

Figure 1 outlines the conceptual framework we use to think about the mechanisms through which ancestral composition may be associated with contemporary outcomes. First, by construction, ACD has two distinct components: the genetic and cultural clusters that deep roots—pre-1500 migration, adaptation to local ecology, historical accident—had put in place by 1500, and the post-Columbian migrations of 1500–1960 that then determined how modern populations were assembled from those clusters. Second, we expect ACD to reach contemporary outcomes through two proximate channels—differences in culture and differences in human capital— which may themselves be interrelated. Third, these channels bear on three families of outcomes. For income, human development, and institutional quality the two channels push together, widening gaps. For trade and migration, cultural distance acts as a resistance term, raising the cost of transacting and settling across a pair of countries, while differences in human capital may work in opposing directions so the net prediction is therefore ambiguous. We connect foreign policy alignment to cultural distance alone, seeing no comparable reason for human-capital differences as such to directly affect countries’ disagreement at the United Nations.

Empirically, we document that countries assembled from more dissimilar ancestral origins are significantly more distant culturally today. The association holds on both

Asian, South–Southeast Asian, Pre-Columbian American, Sub-Saharan African, and Oceanian.

Figure 1: Conceptual Framework



Notes: Solid arrows denote the associations the paper argues for, with the expected sign in parentheses; dashed arrows denote relationships we treat as plausible but do not estimate. Deep roots and the post-Columbian migrations enter as the two sources of variation in ACD; culture and human capital are the proximate channels through which it is expected to operate; income and institutional gaps, bilateral exchange, and foreign-policy alignment are the contemporary outcomes.

alternative benchmarks we use for cultural distance; and they are close to one another in magnitude: 0.196 standard deviations on the behavioral measure of Obradovich et al. (2022), which covers nearly 60,000 interest dimensions from Facebook, and 0.221 on the World Values Survey measure. As outlined in Figure 1, the reach of the measure extends well beyond culture. Ancestrally distant pairs exhibit larger differences in income and human development, trade less with one another, exchange fewer migrants, vote further apart at the United Nations, and show modestly greater differences in governance quality. These magnitudes are considerably large but not implausible. Importantly, ACD remains statistically significant when included jointly with standard proxies –or dimensions– of cultural distance such as linguistic, religious and genetic distance.

The two components of ACD in Figure 1 raise a natural question: how much of the estimated relationships reflects differences already in place in 1500, and how much reflects what post-Columbian migrations did afterwards? By decomposing ACD into initial 1500 cluster separations and subsequent post-Columbian demographic movements, we demonstrate that the role of post-1500 population reshuffling remains clearly legible in modern cultural proximity: pairs whose ancestral compositions converged after 1500 are systematically closer culturally today, conditional on their initial 1500 separation. The cultural map we observe today is therefore not simply the legacy of where populations stood in 1500; the subsequent reshuffling is legible in it.

Existing measures of historical relatedness between countries typically focus either on weighted distances between ancestral populations (e.g., genetic distance) or on ancestry-adjusted historical or geographic characteristics. Yet, countries may be assembled from very different mixtures of ancestral populations even when they exhibit similar average historical characteristics. If cultural traits are transmitted vertically across generations, then similarities in ancestral composition may create enduring commonalities between contemporary societies that are not fully captured by existing measures of relatedness. In line with this proposition, ACD is not merely capturing the omitted role of genetic, linguistic or religious distances. Its correlation with 1500 genetic distance is 0.32 and with contemporary genetic distance 0.43. Its correlation with linguistic and religious distances are 0.46 and 0.25, respectively. When included alongside contemporary genetic, religious or linguistic distance, the explanatory power of ACD for contemporary cultural distance is somewhat reduced. However, its coefficient remains significant both economically and statistically.

The pre-1500 clusters we use in the construction of ACD are doing real work. If we assign origin countries to spheres at random and rebuild the whole ACD measure from scratch, the relationship vanishes: no random grouping comes anywhere near what the grouping based on 1500 genetic distances delivers. Yet, the particular clustering method we use is not driving the results. A classification of countries into nine historical-cultural spheres drawn from the historical literature –rather than from any algorithm, defined

by the major institutional, religious, and linguistic formations in place around 1500 and using no genetic information at any stage, reproduces the empirical patterns we document. Clusters built from geodesic or linguistic rather than genetic distance and specifications that vary the number of clusters also lead to similar results.

Unlike alternative proxies or predictors of cultural distance, ACD is predetermined vis-a-vis any post-1960 dyadic outcome. Ancestral shares were fixed by migrations that concluded before any outcome we study was realized. Yet, we are careful about the interpretation. Predetermination limits immediate reverse-causation concerns, while two-way fixed effects absorb additive country-level determinants. Neither feature delivers exogeneity with respect to unobserved bilateral history. We report associations, and we do not claim to have identified a causal effect.

The measure we propose can be used as a demography-based bilateral resistance term, available to any dyadic regression and predetermined with respect to post-1960 outcomes by construction. This has a practical implication for empirical work on bilateral flows: because ancestral composition distance predicts both cultural and institutional differences while plausibly affecting exchange through additional channels (e.g. human capital differences), omitting it inflates the estimated contribution of common culture and shared institutions in gravity specifications. Three qualifications deserve statement at the outset. First, when explaining bilateral income gaps via ACD, our empirical design does not separate the cultural from the human capital channel. The mediation evidence that differences in human capital and culture explain away the role of ACD is suggestive rather than causally identified. Second, the baseline cosine distance we employ treats all ten spheres as mutually equidistant, so that a European–Near Eastern gap in ancestry shares counts the same as a European–East Asian one, an assumption we relax in robustness checks using kernel-weighted and optimal-transport variants. Third, a substantial share of the identifying variation likely comes from the New World and from the legacy of the Transatlantic slave trade: restricting the sample to the Old World and additionally excluding Sub-Saharan Africa, the World Values Survey coefficient falls from 0.221 to 0.124 and remains significant at one percent, while the Facebook coefficient falls from 0.196 to 0.053. We read this as a feature of what the measure captures rather than as a confound—a measure of ancestral recomposition ought to have most of its variation where the recomposition actually occurred—but the reader should know that the Old World on its own supports a weaker version of the result than the full sample does.

Our closest antecedent is Spolaore and Wacziarg (2018), who rebuild the measurement of population relatedness on new genomic data on microsatellite variation (Pemberton et al., 2013) and show that genetic distance between two populations—and in particular a country’s genetic distance relative to the technological frontier—predicts differences in per capita income, in institutional quality, and in the adoption of agricultural, industrial, communications, and transport technologies. Their weighted genetic distance measure uses

the relative sizes of ethnolinguistic groups as probabilistic weights on pairwise inter-group genetic distances, capturing the expected genetic distance between two randomly drawn individuals, one from each country (Spolaore & Wacziarg, 2009). We use 1500 genetic distances only to partition origin countries into spheres, and then compare the relative sizes of those spheres across two populations. In our baseline analysis, we treat the spheres as distinct but do not exploit the genetic distances among them. The two measures are complements, and we show that ours retains predictive power with genetic distance included.

We differ from the deep roots tradition (Bockstette et al., 2002; Comin et al., 2010; Fulford et al., 2020; Giuliano & Nunn, 2018; Putterman & Weil, 2010) in making the ancestry weights themselves the object and remaining agnostic about the cultural content of any sphere—an agnosticism that is a feature rather than a gap, since the measure cannot be accused of selecting the characteristic that happens to work. In return, the paper offers the persistence literature (Giuliano & Nunn, 2021; Nunn, 2021) a novel measure and a corresponding result it has not previously had: post-Columbian migration did not merely transplant cultures across oceans but *rearranged bilateral cultural variation*, and the traces of that rearrangement remain legible five centuries later. Contemporary cultural distance is not reducible to the ancestral differences already in place in 1500: the convergence in ancestral composition that the intervening centuries produced carries its own association with cultural proximity, over and above the 1500 starting point.

The remainder of this paper is organized as follows: Section 2 reviews the related literature on historical persistence and bilateral relatedness. Section 3 details the construction of our Ancestral Composition Distance (ACD) measure, outlining the 1960 World Migration Matrix (WMM) population benchmark, the PAM10 genetic clustering protocol for pre-1500 origins, and the mathematical distance metrics (PAM10 Cosine Distance and WMM10 Hellinger–Wasserstein Distance). Section 4 describes the data, introducing our primary outcome measures—cultural distance, human capital gaps, income and HDI disparities, bilateral trade and migration flows, political alignment, and governance—as well as complementary distance terms. Section 5 presents our primary empirical findings across all outcome categories, evaluates ACD alongside alternative bilateral-resistance terms and complementary distance measures, and decomposes ACD into pre-1500 baseline separation and post-1500 demographic shifts. Section 6 conducts extensive robustness checks, incorporating controls for agro-climatic and ethnographic distance, alternative ACD constructions such as a historically-grounded Nine-Spheres classification and unclustered countrywise distance, and geographic sample restrictions. Section 7 concludes.

2 Related Literature

The origins of comparative development extend into the prehistoric past through ancient population dispersals, long-run adaptations to regional ecological environments, and historical contingencies. The literature on comparative economic development has increasingly focused on the deep, historically determined factors that affect long-run prosperity and the bilateral interactions between societies (Nunn, 2009, 2020; Spolaore & Wacziarg, 2009, 2013). Economic development is often shaped by the characteristics of human populations rather than solely by the territories they inhabit, and long-term historical persistence operates primarily through the traits of population ancestors (Chanda et al., 2014; Putterman & Weil, 2010). The "Out of Africa" theory posits that anatomically modern humans dispersed through sequential migratory events where sub-samples of the ancestral population detached and settled at increasing distances from the East African origin (Ashraf & Galor, 2013; Becker et al., 2020). Not all migration before 1500 was of the "out of Africa" type and the sequential expansion of human populations led to a systematic reduction in intrapopulation genetic and phenotypic diversity, thereby shaping the foundational capacity of societies to adapt to environmental and social challenges, limiting opportunities for cultural innovation, and influencing the distribution of human diversity across populations (Ashraf & Galor, 2013; Galor et al., 2024).

Over time, within this broad prehistoric context, populations evolved their own distinct traits, social norms, and behavioral shortcuts shaped by the specific ecological environments they inhabited. Geoclimatic conditions influenced the emergence and persistence of cultural norms by affecting the incentives and constraints faced by populations over long periods. Regions characterized by highly productive agricultural environments tended to develop cultural traits associated with greater patience and future-oriented behavior, through both natural selection and cultural evolution (Galor & Ozak, 2016; Sunde et al., 2022). Similarly, paleoclimatic volatility across generations determined the cross-generational stability and persistence of these vertically transmitted traditions, with populations originating from stable climatic environments placing a significantly higher value on maintaining ancestral customs than those from highly variable ancestral environments (Giuliano & Nunn, 2021).

These processes of deep historical adaptation and ancient human migration established the pre-modern baseline of human relatedness. Our framework captures this pre-modern structure of human relatedness through the genetic and cultural clustering of populations observed around 1500 CE. Before the Columbian Exchange, most human populations lived in their indigenous ancestral homelands under conditions of relative geographic isolation (Chanda et al., 2014; Putterman & Weil, 2010). Under these conditions, neutral genetic drift and cultural drift (including linguistic divergence, religious differentiation, the propagation of oral narratives and folk motifs) proceeded in parallel as co-evolving dynamics (Ashraf & Galor, 2013; D’Amato & Russo, 2026; Spolaore & Wacziarg, 2009,

2013).

The post-1500 era introduced a demographic shock of unprecedented historical scale. Voluntary and involuntary population movements following the Columbian Exchange, encompassing the transatlantic slave trade, European colonial settlements and successive waves of international migration, fundamentally restructured global demographic distributions (Akren et al., 2026; Chanda et al., 2014; Putterman & Weil, 2010). Post-Columbian migrations weakened the historical correspondence between geographic locations and the ancestral traits of the populations residing there.

Existing research typically employs the World Migration Matrix of Putterman and Weil (2010) as an ancestry-adjustment tool, using historical population shares to construct ancestry-weighted country-level averages of inherited characteristics such as state history (Ang, 2013) or migration diversity (Rychlowska et al., 2015). Our framework departs from this tradition by treating these distributional shares as the primary object of analysis.

Bilateral cultural distance exhibits substantial persistence because cultural traits are transmitted vertically across generations through costly socialization processes and imitation-based learning mechanisms (Bisin & Verdier, 2000; Nunn, 2012, 2020; Spolaore & Wacziarg, 2013). As a result, historical similarities and differences between populations can continue to shape cultural outcomes long after the original conditions that produced them have disappeared. Spolaore and Wacziarg (2016a) demonstrate a positive correlation between genetic distance and cultural distance as measured by stated values from the World Values Survey (WVS).

The vertical transmission of ancestral preferences is documented across both explicit and implicit metrics of culture, and stated survey values capturing views on authority, politics, and social cooperation are highly persistent (Desmet et al., 2017; Spolaore & Wacziarg, 2016a). This persistence is also evident in individuals' behaviorally manifested preferences within digital contexts. Obradovich et al. (2022) demonstrate a significant correlation between behaviorally revealed preferences and traditional distance measures, including linguistic, geographic, religious, and ancestral distance, using a massive, bottom-up dataset of Facebook interests spanning two billion users. Historical measures of ancestral relatedness predict contemporary differences in economic preferences, including risk aversion, trust, and social preferences, with considerable reliability (Becker et al., 2020, 2026). These preference differences persist even when migrant populations share an identical host-country institutional environment, so they are rooted in vertical transmission rather than immediate institutional conditions (Becker et al., 2020; Bertocchi et al., 2026; Fouka & Tabellini, 2025; Giavazzi et al., 2019). D'Amato and Russo (2026) construct a folklore-based index of cultural similarity from Yuri Berezkin's database of oral narratives and myths. They demonstrate that while genetic and linguistic distances primarily capture demic transmission, folklore similarity additionally captures historical episodes of horizontal cultural diffusion. These findings suggest that enduring cultural similarity

need not coincide with close genetic relatedness, highlighting the importance of broader measures of historical connectedness among populations. Complementary evidence is provided by Michalopoulos and Xue (2021), who demonstrate that oral traditions contain persistent information about ancestral cultural traits. Using a global database of folklore motifs, they show that historically transmitted narratives predict contemporary trust, risk-taking, entrepreneurship, and gender norms, even among second-generation immigrants, highlighting the long-run persistence of cultural content transmitted across generations.

The persistence of ancestral cultural traits has implications beyond cultural outcomes themselves and extends to the accumulation of human capital. Deeply rooted, vertically transmitted cultural traits, particularly patience, cooperativeness, and long-term orientation, directly shape parental investments in education, children’s cognitive skill acquisition, and aggregate literacy (Galor & Ozak, 2016; Sunde et al., 2022). Consequently, countries with ancestrally distant populations exhibit diverging trajectories in human capital formation. Giavazzi et al. (2019) empirically document this long-run transmission across immigrant generations, demonstrating that although labor force participation and work attitudes converge relatively quickly, deep-seated family values and educational aspirations exhibit high persistence and continue to reflect the ancestral country of origin.

These two intermediate channels function interdependently. Vertically transmitted cultural traits govern investment in education and skills in systematic ways (Sunde et al., 2022), but educational attainment and economic modernization also change cultural values actively. Across generations, formal education accelerates secularization, erodes rigid kinship and ethnic cleavages, and builds civic capital, thereby driving cultural convergence (Alesina & Giuliano, 2015; Desmet et al., 2017). These macro-dynamics rest on micro-foundations and ultimately operate through household-level processes of socialization, partner selection, and intergenerational transmission. As individuals form families and social networks across ancestral backgrounds, cultural traits are reinforced, modified, or combined, affecting both future cultural outcomes and investments in human capital (Bisin & Verdier, 2000, 2001, 2011, 2021; Ghatak & Karaivanov, 2014).

Bilateral cultural and cognitive distances can act as transaction costs and information frictions, influencing international trade, capital allocation, and migration flows. First, cultural distance acts as a resistance factor in gravity-type relationships governing trade, migration, and financial flows. Guiso et al. (2009) find that bilateral trust falls significantly in response to culturally driven biases and somatic differences, with attendant reductions in trade, portfolio investment, and foreign direct investment. "Embedded culture" is conceptualized as a powerful source of comparative advantage, systematically shaping stock market participation, default behavior, and credit relations in international financial markets (Guiso et al., 2024). Pellegrino et al. (2025), employing a structural general equilibrium framework, show that cultural and linguistic distances function as persistent portfolio wedges, which explain global capital misallocation and help resolve the 'Lucas

puzzle’ of why capital does not flow from rich to poor nations. Fensore et al. (2022) find that greater ancestral distance between country pairs reduces both the probability of trade and, among pairs that do trade, the volume of goods exchanged. Coimbra Vieira et al. (2024) similarly find that greater cultural similarity, measured through social media interests in food and drink, strongly predicts both international migrant flows and stocks, substantially improving the fit of traditional gravity models.

Second, contemporary cultural distance and differences in human capital are closely associated with cross-country differences in income, human development, and institutional quality. This pathway extends the theoretical framework of Spolaore and Wacziarg (2009, 2013, 2018), who posit that cultural distance from a technological leader impedes the horizontal diffusion of development. Cultural proximity reduces the barriers and costs associated with adopting innovations across populations, while greater cultural divergence raises the costs of technology transfer from the frontier. Empirical studies find that greater genetic and cultural distance from the global technological frontier is associated with lower income levels and slower income convergence (Bove & Gokmen, 2018; D’Amato & Russo, 2026; Spolaore & Wacziarg, 2009).

Disparities in human capital contribute to cross-country variation in physical capital accumulation, total factor productivity, and institutional development. Individual traits such as patience and cognitive skills have been linked to the emergence and quality of formal institutions through their effects on long-term investment, technological progress, cooperation, and state capacity (Acemoglu et al., 2001; Ang, 2013; Comin et al., 2010; Sunde et al., 2022). While the deep-roots literature has primarily examined income differences, considerably less attention has been paid to bilateral differences in broader measures of human development, including health and educational outcomes summarized by the Human Development Index (HDI) (Sunde et al., 2022).

Finally, Contemporary cultural distance is also associated with geopolitical distance and foreign policy alignment. Political alignment in international forums, as measured by dynamic ideal points in United Nations General Assembly voting, has been linked to persistent cultural values and ancestral backgrounds (Bailey et al., 2017; Spolaore & Wacziarg, 2016b). This relatedness gravity paradigm is conceptually enriched by the relatedness-conflict paradox documented by Spolaore and Wacziarg (2016b) who demonstrate that ancestrally closer country pairs can actually be more prone to militarized disputes because their highly overlapping preferences over rival, excludable strategic resources increase the probability of friction.

Taken together, this literature suggests that historical relatedness between populations can influence a wide range of contemporary bilateral outcomes, including cultural proximity, exchange relationships, developmental divergences, and political alignment. This perspective motivates the use of ACD as an alternative and complementary dyadic measure of historically rooted relatedness between contemporary countries.

3 Construction of Ancestral Composition Distance (ACD)

To quantify how similarly contemporary nation-states assemble their populations from pre-1500 historical origins, we construct a novel bilateral metric *Ancestral Composition Distance (ACD)*, which quantifies the similarity of countries' ancestral population compositions.

Existing studies typically measure historical relatedness using genetic, linguistic, religious, or geographic distance. While informative, these measures largely reflect relationships among populations before the extensive population movements that followed the Columbian Exchange. Contemporary countries differ substantially in the extent to which their populations descend from multiple ancestral origins. Hence, similarities in ancestral composition may contain additional information about contemporary cultural proximity. *ACD* is designed to capture this dimension of historical relatedness.

Rather than focusing on the average characteristics inherited from ancestral populations, *ACD* captures the degree of overlap in countries' ancestral origins and thereby provides a novel measure of historical relatedness between contemporary societies. Rather than asking whether a single country draws on a diverse set of origins, *ACD* reflects whether any two countries are demographically assembled from similar ancestral backgrounds. Because it is determined by historical migration patterns that predate contemporary economic and political outcomes, *ACD* provides a predetermined measure of bilateral relatedness (that is not mechanically affected by contemporary bilateral interactions), allowing us to examine how similarities in ancestral composition are associated with present-day cross-country interactions.

We construct *ACD* in three steps. First, we extract historical ancestry profiles from the *World Migration Matrix (WMM)* of Putterman and Weil (2010), using the revision by Chanda et al. (2014).² Next, we partition pre-1500 source populations into discrete ancestral spheres and aggregate each country's WMM ancestry shares within these clusters. Finally, we evaluate pairwise vector distances across the resulting population profiles using alternative "distance" measures. Our baseline measures differ in whether they focus only on direct overlap in ancestry profiles or also allow for similarity between constructed ancestral clusters.

3.1 Ancestral Shares from the *World Migration Matrix (WMM)*

The World Migration Matrix (WMM) traces contemporary national populations back to their ancestral homelands as of 1500 CE, mapping post-Columbian demographic movements across global territories. WMM has primarily been used to ancestry-adjust historical

²The original WMM reconstructs the ancestral composition of populations observed around 2000. We use the Chanda et al. (2014) revision, which adjusts the matrix to represent populations in 1960.

country characteristics. We instead treat the ancestry shares themselves as the object of interest and use them to study bilateral historical relatedness. Crucially, we utilize the 1960 population benchmark variant constructed which adjusts the original year-2000 Putterman–Weil matrix back to 1960 demographic shares. This choice is methodologically vital: Because all contemporary socio-economic, cultural, trade, and institutional outcomes in our study are evaluated in outcomes measured from the 2000s onward., fixing ancestry composition as of 1960 ensures that ACD is predetermined. Hence, it prevents modern, post-1960 migration flows—which respond dynamically to bilateral economic integration, wage gaps, and cultural affinity—from contaminating our measure.

Rows of WMM correspond to destination countries in 1960 and columns to ancestral origin countries in 1500. For destination d and origin o , the entry $a_{do} \in [0, 1]$ gives the share of country d ’s 1960 population whose ancestors lived in origin o in 1500. These entries therefore do not measure bilateral migration flows. Instead, they trace the population of each country back to the historical locations of its ancestors.

For each destination d , the ancestry shares sum to one across origins, $\sum_O a_{do} = 1$. The row vector $\mathbf{a}_d = (a_{d1}, \dots, a_{dO})$ can therefore be interpreted as the country’s ancestral-composition profile. A country may draw ancestry from many different origins, while descendants of the same origin may appear across many destinations. These profiles provide the starting point for the ACD measures developed below, which differ in how distance between two ancestry compositions is calculated. As a simple example, an entry of 44.4 percent for Italy in Argentina’s WMM profile means that 44.4 percent of Argentina’s population in 1960 traced its ancestry to people living in Italy in 1500.

After harmonizing historical-origin and contemporary-country identifiers, we obtain a balanced universe of 163 countries.³ As Table 2 shows, the sample spans all seven contemporary World Bank regions. Measured using 1960 populations—the reference year for the WMM destination populations—the countries in the sample account for approximately 99.7% of the world population. The countries omitted from the ancestry matrix therefore have a very limited effect on global population coverage. The complete country list is reported in Appendix A, Table A1.

³The original WMM lists 172 destinations. Seven of them—the Channel Islands, Iceland, Niue, São Tomé and Príncipe, Tonga, Saint Vincent and the Grenadines, and Samoa—lack corresponding origin columns and therefore cannot appear on both sides of the ancestry matrix. We further exclude the composite Israel/Palestine (*ISR/PSE*) and Serbia and Montenegro (*SCG*) entries because they do not correspond to the contemporary country units used elsewhere in the analysis, yielding the final sample of 163 countries.

Table 1: Illustrative WMM Ancestral-Origin Profiles

Ancestral origin	Share (%)	Ancestral origin	Share (%)	Ancestral origin	Share (%)
<i>Panel A: Nigeria</i>					
Nigeria	88.5	Mali	7.7	Senegal	3.8
<i>Panel B: Argentina</i>					
Italy	44.4	Spain	27.3	Argentina	5.6
France	3.4	Germany	3.4	Poland	3.4
Russia	3.4	United Kingdom	3.3	Lebanon	2.4
Angola	0.6	Ghana	0.5	Israel/Palestine	0.5
Syria	0.5	Congo Republic	0.2	Gabon	0.2
Nigeria	0.2	Benin	0.1	DR Congo	0.1
Turkey	0.1	other origins*	0.4		
<i>Panel C: Singapore</i>					
China	83.7	Indonesia	11.6	Japan	1.0
Sri Lanka	1.0	United Kingdom	1.0	Bangladesh	0.9
Pakistan	0.9				

Notes: Entries are the positive elements of each country’s WMM row, expressed as percentages and ordered from largest to smallest; zero entries are omitted. Shares are rounded to one decimal place, while totals are computed from the unrounded shares. Each profile describes the ancestry composition of the country’s 1960 population by ancestral origin in 1500.

* *other origins* combines 17 origin countries with positive shares in Argentina’s underlying WMM data that individually account for less than 0.1 percent of its ancestry profile; together they account for 0.4 percent after rounding.

Table 2: Sample Composition by World Bank Regions

World Bank region	Countries	Share of sample (%)
East Asia and Pacific	21	12.9
Europe and Central Asia	45	27.6
Latin America and Caribbean	25	15.3
Middle East and North Africa	18	11.0
North America	2	1.2
South Asia	7	4.3
Sub-Saharan Africa	45	27.6
Total	163	100.0

Notes: Regions follow the seven-category World Bank classification used in the analysis; Taiwan is classified as East Asia and Pacific. Country shares may not sum to exactly 100 because of rounding. Population coverage is calculated from 1960 values in the World Bank’s population series. Share of sample (%) column is counting each country as an equal unit, not going by share of populations.

The empirical unit is an unordered country pair ($\{d, o\}$ and $\{o, d\}$ represent the same dyad). The balanced country universe therefore produces

$$\binom{163}{2} = 13,203$$

unique off-diagonal dyads (self-pairs are excluded). Both ancestral-distance measures are available for every dyad; individual regression samples become smaller only when an outcome or an additional control has more limited coverage.

3.2 Cultural Clusters in 1500

In the literature, genetic distance is not interpreted as having a direct biological effect on economic outcomes; rather, it serves as a proxy for broader patterns of historical separation and cultural divergence. Instead, genetic relatedness just before the modern era can be considered as a clean proxy for the complete range of vertically transmitted cultural habits, customs and implicit beliefs that populations inherited from their ancestors (Becker et al., 2020; Spolaore & Wacziarg, 2009). Following Spolaore and Wacziarg (2009), we interpret F_{ST} as a proxy for the historical separation of populations, rather than as a direct measure of cultural distance itself.

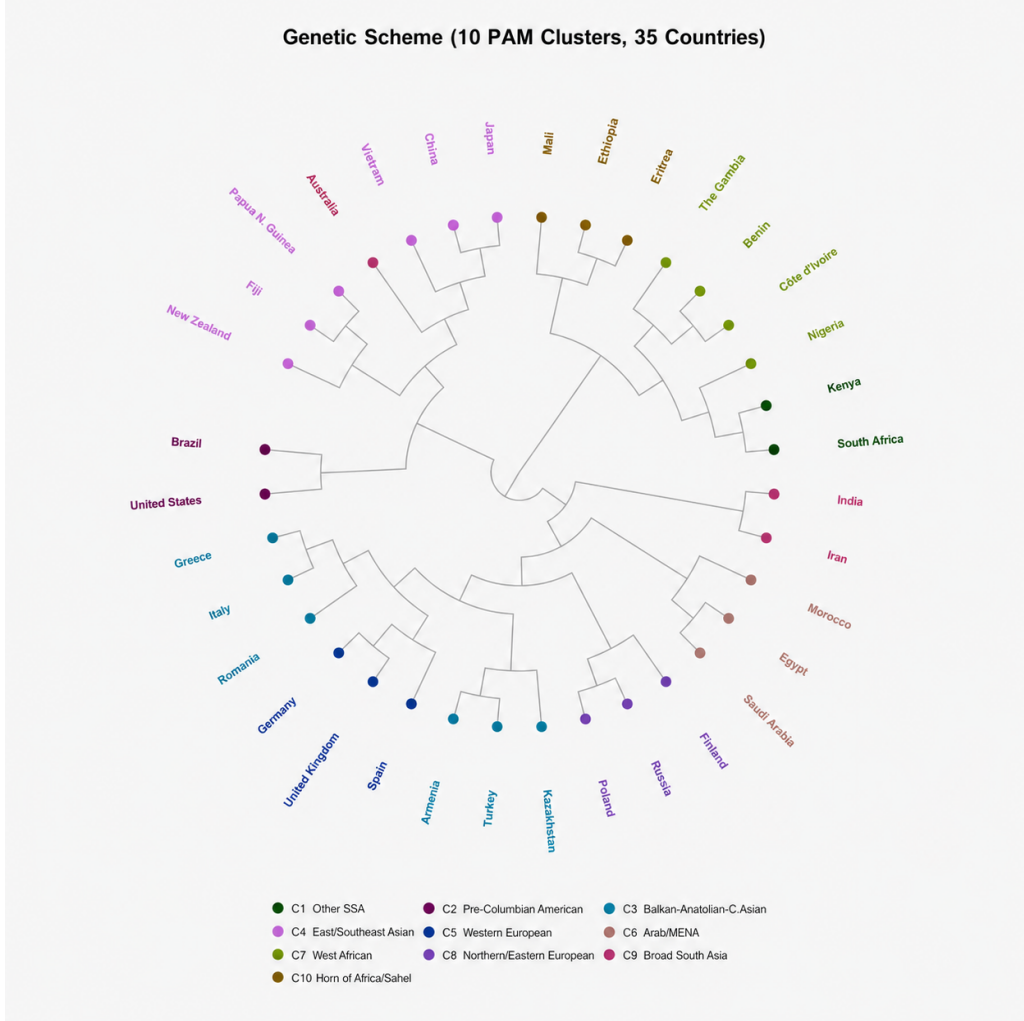
Our baseline grouping uses the year-1500 pairwise genetic distances of Spolaore and Wacziarg (2018), measured by F_{ST} . At this stage, the genetic data are used only to group ancestral origin countries; all country-level ancestry shares continue to come from the WMM. Here, the core idea is to compare countries by the similarity of the ancestral populations from which their modern populations were assembled.

We cluster the F_{ST} distance matrix using partitioning around medoids (PAM). We select 10 clusters as our baseline and refer to this partition as PAM10.⁴

Figure 2 illustrates the resulting structure. Cluster names are added only for ease of interpretation and play no role in the clustering procedure.

⁴Although the four-cluster solution has a marginally higher silhouette score (0.662 versus 0.648), it largely reproduces broad continental groupings and leaves too little variation in ancestral composition. Among the more granular alternatives, $k = 10$ performs best while avoiding very small clusters. Appendix B, Table A5 reports the full silhouette ranking and Table A6 the country assignments. PAM is estimated on the dense F_{ST} sample. Germany, which has incomplete coverage, is assigned using its available mean F_{ST} distances. Six countries without observed F_{ST} links are assigned ex post to their closest regional and historical counterpart clusters. All assignments are documented in Table A6.

Figure 2: PAM10 Ancestral Clusters in the 1500 Genetic-Distance Tree



Notes: Hierarchical structure of the 1500 pairwise F_{ST} distances of Spolaore and Wacziarg (2018), with leaves coloured by PAM10 cluster. The tree is drawn on the full distance matrix; for legibility the figure shows 35 of the 163 origin countries, chosen to cover all ten clusters. The complete assignment is in Appendix B, Table A6. Cluster labels are descriptive and are assigned after clustering.

With the PAM10 clusters in hand, we aggregate each country's WMM ancestry shares from origin countries to clusters. For country i , $c_{i,k}$ is the share of its 1960 population whose ancestors came from origin countries assigned to cluster k . The resulting ten shares form the country's ancestry profile, $\mathbf{c}_i$, and sum to one by construction. Table 3 illustrates these profiles for five countries with different ancestral compositions.

Table 3: *PAM10 Ancestry Profiles* for 5 Illustrative Countries

	Other SSA	Pre-Col. Amer.	Balkan-Anat.	East/SE Asian	W. European	Arab/MENA	West African	N./E. European	Broad South Asian	Horn/Sahel
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
United Kingdom	0.5	—	—	—	97.7	—	1.3	—	0.5	—
Nigeria	3.8	—	—	—	—	—	88.5	—	—	7.7
United States	5.4	6.4	7.3	2.9	63.2	0.4	4.2	9.2	1.0	—
Mexico	3.9	62.6	—	—	30.1	0.8	2.7	—	—	—
Haiti	47.0	—	—	—	2.6	—	50.4	—	—	—

Notes: Entries show the share (in %) of the contemporary population descending from each ancestral cluster. Numbers are obtained by summing the Chanda et al. (2014) origin-country shares within each cluster C_i . Rows sum to 100 up to rounding.

3.3 Calculating the Ancestral Composition Distance

To evaluate the dissimilarity between two country profile vectors $\mathbf{c}_i$ and $\mathbf{c}_j$, we construct two complementary bilateral distance metrics that capture distinct conceptual dimensions of relatedness. Both baseline measures start from the same PAM10 ancestry profile and differ only in how two country profiles are compared. **PAM10 Cosine Distance** measures direct overlap in ancestry shares, whereas **WMM10 HW Distance** additionally takes into account the similarity among PAM10 ancestry profiles.

PAM10 Cosine Distance: We first compare country ancestry profiles directly:

$$d_{ij}^C = 1 - \frac{\mathbf{c}_i' \mathbf{c}_j}{\|\mathbf{c}_i\|_2 \|\mathbf{c}_j\|_2}. \quad (1)$$

The measure is zero when the two normalized ancestry profiles are identical and approaches one as their ancestry is concentrated in different clusters. Genetic Distance in 1500 enters solely to define the initial sphere boundaries (PAM10 cultural clusters); once established, Cosine Distance treats all distinct ancestral clusters as mutually orthogonal.

WMM10 Hellinger–Wasserstein (HW) Distance: PAM10 Cosine Distance treats all PAM10 clusters as equally distinct once the partition is formed. This provides a transparent benchmark, however it ignores the possibility that different ancestral clusters may have similar destination footprints in the WMM. Two distinct ancestral spheres may

share similar historical migration footprints across destinations. To account for cross-sphere similarity, we construct an optimal "transport" metric using a Hellinger ground cost.

We first apply the PAM10 partition to destination countries and compute the average share of source cluster k across destination cluster ℓ , yielding an aggregated 10×10 migration matrix $\mathbf{P}$ with normalized rows $\mathbf{p}_k$. The Hellinger similarity between source clusters k and ℓ is calculated as the inner product of their square-root destination vectors. Let $\mathcal{C}_\ell$ denote the set of countries assigned to destination cluster ℓ . For source cluster k , define:

$$M_{k\ell} = \frac{1}{|\mathcal{C}_\ell|} \sum_{i \in \mathcal{C}_\ell} c_{i,k}. \quad (2)$$

Thus, $M_{k\ell}$ is the average share of ancestry from source cluster k among countries in destination cluster ℓ .⁵ We normalize each source-cluster row to obtain:

$$P_{k\ell} = \frac{M_{k\ell}}{\sum_{m=1}^K M_{km}}, \quad \sum_{\ell=1}^K P_{k\ell} = 1. \quad (3)$$

The row p_k of P therefore gives the normalized, equal-country-weighted WMM destination profile of source cluster k . Table 4 reports this 10-cluster representation of the WMM.

We compare these profiles after applying the Hellinger transformation. For source clusters k and ℓ , define:

$$S_{k\ell}^H = \sum_{m=1}^K \sqrt{P_{km}P_{\ell m}}. \quad (4)$$

$S_{k\ell}^H$ is the inner product of two unit-length square-root profiles and it equals one when the two destination profiles are identical and decreases as their overlap falls. Relative to an untransformed comparison, the square-root transformation reduces the dominance of large destination shares and increases the relative contribution of smaller shared destinations.

We convert this similarity into the ground cost $D_{k\ell}^H = 1 - S_{k\ell}^H$. Our WMM10 HW Distance then compares the country ancestry profiles $\mathbf{c}_i$ and $\mathbf{c}_j$ by finding the least-cost way of matching their ancestry shares:

$$d_{ij}^{WH} = \min_{\pi \geq 0} \sum_{k=1}^K \sum_{\ell=1}^K D_{k\ell}^H \pi_{k\ell}, \quad \pi \mathbf{1} = \mathbf{c}_i, \quad \pi' \mathbf{1} = \mathbf{c}_j. \quad (5)$$

Here, $\pi_{k\ell}$ denotes the amount of ancestry mass in cluster k of country i matched to cluster ℓ of country j . Identical ancestry profiles have distance zero. Different profiles can nevertheless remain relatively close when their unmatched ancestry shares belong to clusters with similar WMM destination profiles.⁶

⁵Destination countries receive equal weight. Population weighting would allow a small number of large countries to dominate the destination profile of a cluster.

⁶Since $D_{k\ell}^H = 1 - S_{k\ell}^H$, the ground cost corresponds to squared Hellinger distance under the usual normalization. Strictly, d_{ij}^{WH} is therefore an optimal "transport" cost induced by this Hellinger-based ground cost. We use "Wasserstein" to describe the "transport" construction rather than to assert that the

To clarify and summarize the distinction between the two measures: Our two baseline measures start from the same PAM10 partition and the same country-level ancestry profiles. But they differ in how the similarity among those profiles is evaluated. *PAM10 Cosine Distance* captures direct overlap in ancestral composition. Two countries are closer when similar shares of their populations trace to the same ancestral spheres. Whereas, *WMM10 HW Distance* additionally accounts for similarity among those spheres. Ancestry shares that do not overlap directly can therefore be matched across spheres, with the cost of that match determined by the similarity of their WMM destination profiles.

Table 4: WMM Ancestry Matrix Aggregated to PAM10 Clusters

Source cluster	Normalized destination profile of each source cluster (%)									
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1 Other SSA	79.95	8.87	0.12	—	0.05	1.74	6.33	—	—	2.95
C2 Pre-Columbian American	—	99.99	—	—	0.01	—	—	—	—	—
C3 Balkan–Anatolian–C. Asian	—	3.80	91.84	0.04	0.56	1.16	—	0.25	1.90	0.46
C4 East/Southeast Asian	0.02	0.56	0.27	97.97	0.22	0.16	—	—	0.54	0.26
C5 Western Europe	2.00	25.84	0.38	2.44	59.72	0.39	0.21	0.22	6.31	2.49
C6 Arab/MENA	0.07	0.24	0.10	—	0.17	98.50	0.27	—	0.40	0.25
C7 West African	6.22	8.29	—	—	0.11	0.05	84.83	—	—	0.49
C8 Northern/Eastern European	—	0.90	2.62	0.05	0.88	0.07	—	95.05	0.44	—
C9 Broad South Asian	0.12	3.72	0.04	2.24	0.10	2.34	—	—	76.13	15.31
C10 Horn of Africa/Sahel	0.23	—	—	—	—	0.05	7.57	—	—	92.15

Notes: Entries are percentages, with each row summing to 100 apart from rounding. Shading is proportional to the entry, row maxima are shown in bold, — denotes an exact zero, and (<0.01) denotes a positive value below the reported precision.

For specifications examining development and technology gaps relative to the world technological frontier (the United States), we transform bilateral ACD into a relative distance metric and construct a USA-relative version of each ACD measure, that will be called *ACD - relative distance to USA*. Denoting either WMM10 HW Distance or PAM10 Cosine Distance generically by $d(\cdot, \cdot)$, we define the relative distance between countries i and j as $\text{relACD}_{ij} = |d(i, \text{USA}) - d(j, \text{USA})|$, with $d(\text{USA}, \text{USA}) = 0$. Unlike bilateral ACD, this measure captures how differently the two countries are positioned relative to a common benchmark rather than their distance from one another directly. We use this construction in the development-gap specifications reported in Table 9, where the United States serves as a common reference point for the technological frontier.

resulting cost satisfies every metric axiom.

4 Outcomes and Control Variables

To assess whether *Ancestral Composition Distance (ACD)* operates as a predetermined bilateral friction across contemporary societies, we assemble a comprehensive dyadic dataset covering 163 countries for the 2000–2019 period. Our outcome variables are organized into four core domains:

First, we examine intermediate transmission channels *contemporary cultural distance* - the outcome most directly related to the persistence of ancestral differences embodied in ACD - alongside with *Human Capital Gap*. Second, we compare *developmental outcomes*, measured by bilateral gaps in income and human development (HDI). Third, we study *international exchange*, using bilateral migration and trade intensity. And fourth, we look at *political and governance alignment* (UN voting distance and differences in institutional quality). Taken together, these outcomes allow us to assess whether ancestral composition is associated not only with contemporary cultural differences, but also with broader economic and political dimensions of bilateral distance. We also introduce a set of well-known related bilateral measures, such as contemporary genetic distance, that may play a complementary role to ACD, together with the geographic, historical, and institutional controls commonly used in the literature.

We additionally incorporate complementary measures of population relatedness and a comprehensive set of dyadic control variables.

4.1 Transmission Channels: Cultural Distance & Human Capital Gap

We examine cultural distance and human capital disparities as potential proximate channels associated with *ACD* and contemporary bilateral outcomes. This approach allows us to assess whether ACD retains explanatory power beyond conventional measures of bilateral relatedness.

We first examine whether ACD is systematically associated with contemporary bilateral cultural differences. We focus on two complementary dependent variables, which we refer to throughout as *FB Cultural Distance* and *WVS Cultural Distance*.

- **FB Cultural Distance:** We use the bilateral cultural-distance measure developed by Obradovich et al. (2022). The measure represents each country by a high-dimensional vector containing the shares of Facebook users associated with nearly 60,000 interests across a broad range of topics. Bilateral cultural distance is calculated as the cosine distance between the corresponding country vectors, equivalent to one minus their cosine similarity. Higher values therefore indicate greater divergence in the behaviourally revealed interests and preferences of Facebook users in the two countries.

- **WVS Cultural Distance:** We use the updated cultural distance measure of Pellegrino et al. (2025), which extends the survey-based construction of Spolaore and Wacziarg (2016a) and captures bilateral differences in contemporary values and beliefs as the expected disagreement between respondents drawn from two countries.⁷

Countries assembled from more dissimilar ancestral origins may also exhibit greater larger differences in human capital, in addition to higher cultural distance. Together, they capture behavioral and attitudinal dimensions of cross-country distance.

- **Human-Capital Gap:** For each country pair i, j , let T_{ij} denote the years between 2000 and 2019 in which the Penn World Table 11.0 human-capital index (hc) is observed for both countries (Feenstra et al., 2015). Our baseline measure is the unweighted average absolute difference in the level of the index across common observed years. We do not interpolate missing observations, carry values forward or backward, or code missing country-year values as zero.⁸

4.2 Developmental Outcomes: Income & HDI

To test whether ACD impedes the cross-border diffusion of productivity and human development, we examine income and human-development gaps. These outcomes capture complementary dimensions of development: material living standards, broader achievements in health and education, and the stock of skills embodied in the population. Two variables use the 2000–2019 period.

- **Income Gap:** for each year, we take the absolute difference in GDP per capita between the two countries from the CEPII Gravity database (Conte et al., 2022) and then its natural logarithm. The outcome is the mean of these annual logged gaps over the years in which both countries are observed. Taking logs before averaging prevents the measure from being dominated by the few years in which a pair’s gap is largest. We also report estimates by five-year interval and over the longer available window as robustness checks.
- **HDI Gap:** the mean absolute bilateral gap in the UNDP Human Development Index (United Nations Development Programme, 2024). The index combines health, education, and living standards and therefore provides a broader measure of development than income alone.

⁷Further details on the construction of the measure and the underlying data are provided in Pellegrino et al. (2025) and on the Geopolitical Distance database website. The authors’ flexible procedure accommodates questions that are not observed in all countries, reducing data loss and extending coverage to 116 countries. Measures are available for 1984, 1993, 1998, 2002, 2009, 2014, and 2021. We estimate the relationship separately for each benchmark year, and the results remain stable across survey waves.

⁸The income-gap overlap exercise uses the analogous log measure, $|\log hc_{it} - \log hc_{jt}|$, averaged over the same pair-specific set of common observed years. Further robustness checks apply the common-year construction to harmonized learning outcomes, learning-adjusted years of schooling, and Barro–Lee mean years of schooling among the population aged 25 and older.

As a robustness check, we also create the "relative to technology frontier" version of these outcomes, using the United States as a common reference point.

4.3 International Exchange: Migration and Trade

To evaluate ACD as a bilateral trade and demographic resistance term, we construct size-normalized intensity metrics for trade flows and migrant stocks. For a bilateral variable of F between countries i and j , intensity of that variable is calculated as the sum of the values in both directions divided by the combined outward total of the variable of the two countries:

$$Intensity_{ij}^F = \frac{F_{ij} + F_{ji}}{\sum_{k \neq i} F_{ik} + \sum_{k \neq j} F_{jk}}. \quad (6)$$

The numerator captures the total bilateral relationship between countries i and j , while the denominator represents their combined external activity. The resulting measure expresses the bilateral relationship as a share of the two countries' overall migration or trade activity, limiting the mechanical influence of country size. It is symmetric by construction and bounded between zero and one.

- **Migration Intensity:** applies equation (6) to bilateral migrant stocks. The baseline averages the United Nations international migrant-stock vintages observed in 2000, 2005, 2010, 2015, and 2020, with no interpolation between vintages (United Nations, Department of Economic and Social Affairs, Population Division, 2024).⁹
- **Trade Intensity:** is constructed from annual bilateral export flows in the CEPII Gravity database (Conte et al., 2022). We apply equation (6) in each year over 2000–2019 and use the period average of the importer-reported Comtrade series as the baseline.¹⁰

4.4 Political Alignment and Differences in Governance (Quality of Institutions)

Finally, we examine whether ACD is associated with bilateral differences in foreign-policy positions and governance institutions.

- **UN-Voting Distance:** the mean annual absolute difference between the two countries' United Nations General Assembly ideal-point estimates. Higher values indicate greater divergence in foreign-policy positions. The ideal-point estimates are

⁹Robustness checks use the full UN DESA 1990–2024 window, the decennial Global Bilateral Migration Database stocks for 1960–2000 (Ozden et al., 2011), and annual bilateral migration-flow estimates for 1990–2023 from Gaskin and Abel (2026).

¹⁰The CEPII database provides four alternative series combining exporter- and importer-reported flows from Comtrade and IMF sources. We report all four series, together with five-year estimates over the longer 1960–2019 window, as robustness checks.

developed by Bailey et al. (2017) from the roll-call data compiled by Voeten et al. (2009).¹¹

- **WGI Distance:** the cosine distance between the two countries’ six-dimensional Worldwide Governance Indicators profiles—voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption—averaged across years (Kaufmann et al., 2011).¹²

4.5 Measures of Bilateral Distance with a Complementary Role to ACD

To verify that ACD captures substantive relational information beyond existing proxies of cultural relatedness, we incorporate three established distance measures: Contemporary Genetic Distance, Linguistic Distance, and Religious Distance. Contemporary Genetic Distance reweights pairwise F_{ST} distances using the ancestral composition of present-day populations and is therefore the closest conceptual alternative to ACD. Linguistic Distance and Religious Distance are the updated weighted measures of Pellegrino et al. (2025). Together, these variables capture three prominent dimensions of bilateral similarity: ancestry-weighted population relatedness, linguistic proximity, and religious composition.

Table 5 provides a quick comparison. WMM10 HW Distance and PAM10 Cosine Distance are highly correlated, at 0.921, as expected given that both are constructed from closely related ancestry profiles. Their remaining variation reflects the different treatment of similarity across clusters: PAM10 Cosine relies on direct compositional overlap, whereas WMM10 HW allows similarity across clusters through the transport structure.¹³ Appendix A, Table A2 reports the full correlation matrix and coverage.

Contemporary Genetic Distance is positively related to both ACD measures, with correlations of 0.466 for WMM10 and 0.426 for PAM10. The relationship is substantial but far from one-to-one, reflecting the conceptual difference between the measures. Contemporary Genetic Distance summarizes the expected genetic relatedness of the ancestral populations represented in two countries, whereas ACD compares the composition of those ancestral origins across the two populations. Linguistic and Religious Distance provide further comparisons with dimensions of cultural similarity that may themselves reflect long-run population history.

These measures are therefore best viewed as related rather than competing proxies for the same object. The empirical question is whether ACD retains explanatory content

¹¹The baseline variable averages 2000–2019. The underlying source covers 1946–2025; robustness checks use the 1960–2024 average, decade and five-year windows, and year-specific estimates.

¹²The baseline variable uses the full 2000–2019 period. The six components separately and their Euclidean aggregation are retained as robustness checks.

¹³Correlations reported elsewhere in the paper may differ slightly because they are calculated on the dyads available for the variables used in each exercise rather than on the common sample in this table.

once contemporary genetic, linguistic, and religious distance are taken into account. We examine this directly in the comparison specifications below.

Table 5: Correlations: ACD Measures, Genetic Distance, Geodesic Distance

	(1)	(2)	(3)	(4)	(5)
(1) WMM10 HW Distance	1.000				
(2) PAM10 Cosine Distance	0.921***	1.000			
(3) Genetic Distance (1500)	0.332***	0.315***	1.000		
(4) Genetic Distance (Contemporary)	0.466***	0.426***	0.694***	1.000	
(5) Geodesic Distance	0.396***	0.372***	0.745***	0.579***	1.000

Notes: Pearson correlations across 12,162 complete dyads. Genetic Distance (1500) is the pre-1500 F_{ST} measure; Genetic Distance (Contemporary) reweights the same F_{ST} data by the ancestral composition of present-day populations. Geodesic Distance is the log of capital-to-capital distance. Significance: *** $p < 0.01$.

Table 6 reports descriptive statistics for all bilateral variables introduced up to this point: the two principal ACD measures, the established comparison distances, geographic distance, and the cultural, economic, demographic, political, and institutional outcomes. The full pairwise correlation matrix and its sample coverage are reported in Appendix A, Table A2.

Table 6: Descriptive Statistics

Variable	Observations	Mean	Std. dev.	Minimum	Maximum
PAM10 Cosine Distance	13,203	.8430	.3209	.0000	1.0000
WMM10 HW Distance	13,203	.6965	.2679	.0000	.9994
FB Cultural Distance	12,403	.1928	.0558	.0336	.5242
WVS Cultural Distance	5,671	.1301	1.0259	-3.9470	3.7348
Human-Capital Gap	9,453	.7919	.5540	.0042	2.5086
Income Gap	13,041	1.7092	1.6520	-3.7517	4.7019
HDI Distance	12,561	.1880	.1331	.0021	.6127
Trade Intensity	13,203	.0020	.0085	.0000	.3161
Migration Intensity	13,203	.0045	.0270	.0000	.8381
UN-Voting Distance	12,720	.8707	.6542	.0264	4.0873
WGI Distance	13,203	.0208	.0175	.0002	.1680
Contemporary Genetic Distance	12,162	.0350	.0175	.0000	.0872
Linguistic Distance	13,203	.8878	.1485	.0151	1.0000
Religious Distance	13,203	.6174	.2168	.0117	.9985

Notes: The table reports dyad-level descriptive statistics for our main variables. Geodesic Distance is the logarithm of capital-to-capital distance. Income Gap, Human-Capital Gap, and Trade Intensity are averaged over 2000–2019; Migration Intensity uses the observed UN DESA stock vintages in 2000, 2005, 2010, 2015, and 2020. The UN-Voting, HDI, and WGI variables use the windows stated in the text. No common-sample restriction is imposed across rows.

4.6 Other Bilateral Controls

All specifications condition on a comprehensive set of dyadic controls. Following the dyadic control framework of Spolaore and Wacziarg (2009, 2018), we include two groups of time-invariant bilateral controls that may independently shape cultural, political, and economic proximity. The variables are drawn primarily from CEPII GeoDist (Conte et al., 2022) and the updated dyadic dataset of Pellegrino et al. (2025).

Geographic and Topological Controls: The first group accounts for different dimensions of physical proximity and geographic connectedness. Geodesic distance captures the direct separation between two countries and the associated costs of travel, communication, and exchange, while contiguity identifies pairs that share a land border and are therefore likely to experience more frequent interaction than distance alone would suggest. Absolute differences in latitude and longitude capture broader north–south and east–west separation, including differences in regional location and environmental conditions that may persist even among pairs with similar bilateral distances. We also control for whether the two countries share access to the same sea or ocean, which captures an additional dimension of maritime connectivity. Landlocked and island indicators account for forms of geographic isolation that may systematically affect international contact, transport costs, and historical population movements. Together, these controls absorb several dimensions of geographic proximity that could otherwise be related to both ancestral composition and contemporary bilateral outcomes.

Legacy of Colonial Ties: The second group captures persistent bilateral connections created by common history, institutions, and language. A direct colonial relationship accounts for the institutional, demographic, and cultural links established through colonial rule, while common colonizer captures similarities arising from shared exposure to the same imperial power even when no direct colonial relationship existed between the two countries. Common legal origin controls for inherited similarities in legal and administrative systems that may influence both institutional proximity and economic interaction. Common official language captures lower communication costs as well as a durable historical and cultural connection that can facilitate the transmission of information, norms, and practices. These variables therefore account for some of the main historical and institutional ties through which country pairs may remain similar independently of their ancestral composition. These variables account for the main historical and institutional ties through which country pairs may remain similar independently of their ancestral composition. Because shared colonial experience also generated part of the ancestral similarity we measure, conditioning on them removes some genuine ACD variation as well, and the resulting estimates should be read as conservative.

5 Empirical Analyses and Results

This section evaluates the empirical performance of Ancestral Composition Distance (ACD) across our four outcome domains. The empirical analysis proceeds in five steps. We first estimate the conditional association between ACD and two different measures of contemporary cultural distance. We then examine ACD’s association with developmental and human-capital gaps, bilateral trade and migration, political alignment and differences in institutional quality. We next assess the explanatory contribution of ACD relative to conventional bilateral-resistance terms. Next, we investigate whether ACD contains information beyond genetic, linguistic, and religious distances. Finally, we decompose current ACD into a pre-1500 benchmark and post-1500 changes in population composition. Section 6 evaluates sensitivity to other ancestral characteristics, alternative constructions of ACD, and geographic sample restrictions. Overall, our results suggest that ACD is a) a robust and economically meaningful predictor of bilateral cultural distance and developmental differences as well as b) an informative bilateral resistance term.

5.1 Econometric Specification

The unit of observation is an unordered country pair. For each dyad $\{i, j\}$, we estimate

$$y_{ij} = \beta ACD_{ij} + \mathbf{x}_{ij}'\boldsymbol{\gamma} + \delta_i + \delta_j + \varepsilon_{ij}, \quad (7)$$

where y_{ij} is the bilateral outcome, ACD_{ij} is one of the two ancestry-distance measures, $\mathbf{x}_{ij}$ contains the dyadic controls, and δ_i and δ_j denote country fixed effects.

We estimate dyadic regressions on 163 countries and 13,203 pairs, controlling throughout for geodesic distance, absolute differences in latitude and longitude, contiguity, shared sea, landlocked and island status, common official language, common legal origin, colonial relationship, common colonizer, and including fixed effects for both members of every pair. So that, all country-level determinants are absorbed and only bilateral variation identifies the coefficient. Standard errors are two-way clustered by country, and continuous variables are standardized, so each coefficient reads as a standard deviation of the outcome per standard deviation of the regressor.

Country fixed effects absorb additive country-specific characteristics that may affect a country’s distance from all partners, so identification comes from within-country differences across bilateral relationships. The coefficient β therefore asks whether, conditional on country characteristics and observed dyadic controls, the partners that are closer in ancestral composition are also closer in the outcome of interest. Remaining pair-specific heterogeneity is addressed through the geographic, historical, and institutional controls described above. Controls cannot rule out what we do not observe: something common to two countries may have shaped both the migrations that formed ancestral profiles and

the outcomes we study. The estimates are therefore conditional associations, not causal effects.

For the OLS specifications, the outcome and all continuous regressors are standardized within the estimation sample, while binary variables remain indicators. Coefficients can therefore be interpreted as the standard-deviation change in the outcome associated with a one-standard-deviation change in the regressor. Standard errors are clustered two-way by the two countries in the dyad, allowing for dependence across observations that share a country (Aronow et al., 2015; Cameron et al., 2011).¹⁴

Trade and migration are estimated instead by Poisson Pseudo-Maximum Likelihood (PPML) in levels, which accommodates zero flows and heteroskedasticity while retaining the same fixed effects, controls, and clustering structure (Correia et al., 2020; Santos Silva & Tenreyro, 2006). Their coefficients are therefore semi-elasticities and are not directly comparable in magnitude with the standardized OLS estimates.

5.2 ACD and Transmission Channels: Contemporary Cultural Distance & Human Capital Gap

Table 7 reports the baseline cultural-distance estimates. Consistent with our conceptual framework, ACD enters positively in every specification. For PAM10 Cosine Distance, adding the full set of baseline bilateral control block reduces the coefficient from 0.324 to 0.196 for FB Cultural Distance and from 0.286 to 0.221 for WVS Cultural Distance. The corresponding WMM10 HW Distance estimates decline from 0.380 to 0.247 for FB and from 0.296 to 0.217 for WVS.

This pattern is consistent with a broad literature on the persistence and intergenerational transmission of culture. Preferences, beliefs, and norms can remain partly persistent across generations while still adapting to the local environment (Bisin & Verdier, 2001; Giuliano & Nunn, 2021). ACD therefore need not imply that culture is fixed or mechanically determined by ancestry; it captures whether contemporary populations were formed from similar or different ancestral backgrounds, whose cultural traits may have been transmitted with some persistence over time. The fact that the relationship is similar for FB, a behavioral measure of revealed interests, and WVS, a survey measure of stated values, further suggests that the result is not specific to a single way of measuring culture.

Geography, common language, and shared historical ties therefore explain only a small part of the relationship between cultural distance and ACD. The estimates suggest that

¹⁴Because the dyads are unordered, we verify that the estimates do not depend on which country is listed first. We re-estimate the models using symmetric country fixed effects, assigning each country a single effect whenever it appears in the pair, and repeat the estimation after reversing the ordering of the countries within dyads. We also compute dyadic-robust standard errors that allow arbitrary dependence among all observations sharing either country. The estimated coefficients remain essentially unchanged across these checks. Although the dyadic-robust standard errors are somewhat larger, the signs, magnitudes, and statistical significance of the main results remain the same.

ancestral composition contains information about contemporary cultural differences beyond that captured by standard bilateral controls.¹⁵

Table 7: ACD and Contemporary Cultural Distance

	FB Cultural Distance		WVS Cultural Distance	
	(1)	(2)	(3)	(4)
<i>Panel A: PAM10 Cosine Distance</i>				
ACD	0.324*** (0.024)	0.196*** (0.025)	0.286*** (0.038)	0.221*** (0.043)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	12,399	12,399	5,667	5,667
R^2	0.593	0.652	0.543	0.581
Within R^2	0.192	0.307	0.132	0.204
<i>Panel B: WMM10 HW Distance</i>				
ACD	0.380*** (0.035)	0.247*** (0.036)	0.296*** (0.039)	0.217*** (0.044)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	12,399	12,399	5,667	5,667
R^2	0.616	0.659	0.542	0.576
Within R^2	0.236	0.322	0.129	0.194

Notes: The two panels estimate PAM10 Cosine Distance and WMM10 HW Distance separately. Parsimonious columns include ACD and country1 and country2 fixed effects. Controlled columns add Geodesic Distance, absolute latitude and longitude differences, and indicators for a shared border, common official language, common legal origin, colonial relationship, common colonizer, shared sea or ocean, either country being landlocked, and either country being an island. Each parsimonious–controlled comparison uses the same outcome-specific sample. Continuous outcomes and regressors are standardized within that sample. Standard errors are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We treat culture and human capital as two proximate channels that may connect ancestral composition to later outcomes. Table 8 therefore turns to differences in human capital, measured from PWT over 2000–2019. The controlled ACD coefficients are 0.178 for PAM10 and 0.267 for WMM10. Thus, countries with more dissimilar ancestral compositions tend to differ more in productive capabilities as well as in culture. Human capital is more closely tied to development than the cultural outcomes above, but these estimates remain conditional associations and do not identify a causal channel.

¹⁵The result is stable across WVS vintages (Appendix D, Figure A9), geographic restrictions (Table 18), climate controls (Appendix J, Table A25).

Table 8: ACD and Human-Capital Gap

	<i>Panel A: PAM10 Cosine Distance</i>		<i>Panel B: WMM10 HW Distance</i>	
	(1)	(2)	(3)	(4)
ACD	0.235*** (0.036)	0.178*** (0.033)	0.312*** (0.050)	0.267*** (0.045)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	9,312	9,312	9,312	9,312
R^2	0.309	0.385	0.337	0.402
Within R^2	0.067	0.170	0.104	0.192

Notes: The dependent variable is the Human-Capital Gap. Columns (1)–(2) use PAM10 Cosine Distance, while columns (3)–(4) use WMM10 HW Distance. Columns (1) and (3) include ACD and country1 and country2 fixed effects; columns (2) and (4) additionally include the baseline dyadic controls. Human-Capital Gap is constructed from the Penn World Table human-capital index by averaging annual absolute bilateral gaps over years observed for both countries between 2000 and 2019. All models are standardized OLS. Standard errors are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 ACD and Developmental Gaps

Table 9 considers income and HDI gaps. Each block reports absolute ACD without and with controls, followed by the controlled *ACD - relative distance to USA* measure. The two measures never enter jointly. It is seen that ACD is positively associated with comparative developmental disparities. In the controlled specifications, the PAM10 and WMM10 coefficients are 0.104 and 0.107 for income, 0.224 and 0.287 for HDI. The outcomes suggests that the relationship is not confined to income alone, but is also present for wider differences in living standards and productive capabilities.

These results extend the empirical relevance of ACD beyond cultural and human capital differences. Countries with more dissimilar ancestral compositions also tend to differ more in their levels of development, while accounting for standard bilateral controls and country fixed effects. One possible mechanism is the diffusion of development. As emphasized by Spolaore and Wacziarg (2009) and Spolaore and Wacziarg (2013), historically rooted differences across populations may act as barriers to the diffusion of productivity-enhancing technologies and practices. Greater cultural distance may make communication, learning, adaptation, or the transfer of organizational practices more costly, so that otherwise available innovations spread less easily across societies. This interpretation is also consistent with evidence that differences in technology adoption are closely related to cross-country income differences (Comin & Hobijn, 2010). More broadly, persistent differences in norms, beliefs, and institutions may shape how societies

adopt and use new technologies or accumulate productive capabilities (Alesina & Giuliano, 2015; Guiso et al., 2006). In this sense, ACD may capture historically rooted frictions that matter for how development spreads across countries, rather than a direct effect of ancestry on economic performance.

Table 9: ACD and Differences in Income and Human Development

	Income Gap			HDI Gap		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: PAM10 Cosine Distance</i>						
ACD	0.129*** (0.023)	0.104*** (0.024)		0.275*** (0.037)	0.224*** (0.036)	
ACD - relative distance to USA			0.187*** (0.030)			0.258*** (0.040)
Additional controls	No	Yes	Yes	No	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,876	12,876	12,876	12,400	12,400	12,400
R^2	0.605	0.618	0.627	0.331	0.400	0.397
Within R^2	0.037	0.069	0.091	0.094	0.186	0.182
<i>Panel B: WMM10 HW Distance</i>						
ACD	0.137*** (0.026)	0.107*** (0.028)		0.335*** (0.049)	0.287*** (0.046)	
USA-relative distance			0.127*** (0.027)			0.171*** (0.036)
Additional controls	No	Yes	Yes	No	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,876	12,876	12,876	12,400	12,400	12,400
R^2	0.605	0.617	0.619	0.354	0.411	0.381
Within R^2	0.038	0.067	0.072	0.125	0.202	0.161

Notes: Within each outcome block, columns (1) and (4) include absolute ACD and country1 and country2 fixed effects. Columns (2) and (5) add the baseline dyadic controls; columns (3) and (6) retain those controls but replace absolute ACD with USA-relative ACD. Absolute and relative ACD never enter the same model. All three specifications use the same outcome-specific sample, and every dyad containing the United States is excluded. For country c , $ACDF_c = ACD_{\{c, USA\}}$ and $ACDF_{USA} = 0$, so $relACD_{ij} = |ACDF_i - ACDF_j|$. Income Gap and HDI Gap are 2000–2019 averages. All models are standardized OLS. Standard errors are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We then express these outcomes relative to the United States, using the US as a common reference point to the technological frontier. The estimated coefficients remain positive throughout. For income, they are larger than in the corresponding bilateral

specifications, while the comparison is less uniform for HDI. This pattern is consistent with the possibility that historically rooted differences are relevant for the diffusion of technologies and practices from the frontier.¹⁶

Complementary Transmission Channels: Culture and Human-Capital

Table 10 explores how much of the relationship between ACD and bilateral income differences is accounted by cultural and human-capital differences. To make the coefficients comparable across specifications, all columns use the same sample and retain the baseline bilateral controls and country fixed effects.

For PAM10, the baseline coefficient is 0.135. It falls to 0.028 when FB cultural distance is added and to 0.068 when we instead control for human-capital distance. Including both together reduces the coefficient further to 0.022. The pattern is similar, and somewhat stronger, for WMM10. Its baseline coefficient of 0.145 falls to -0.002 after controlling for cultural distance, to 0.042 after controlling for human-capital distance, and to -0.022 when both are included. Overall and within R^2 also increase substantially as these additional differences are taken into account.

Surely, we do not interpret this exercise as identifying a mediation mechanism. Culture, human capital, and income are jointly related, and these cross-sectional regressions cannot establish which of these relationships is causal. The more limited point is that the association between ACD and income differences overlaps considerably with both cultural and human-capital distance. But still, this attenuation helps clarify what lies behind the income gradient associated with ACD. A substantial part of the relationship appears to be shared with cross-country differences in culture and human capital. This is consistent with the broader interpretation of ACD as summarizing demographic histories that are also related to the transmission of cultural traits and to differences in productive capabilities.¹⁷

¹⁶Appendix E, Tables A13–A14, replicate the main specifications of Spolaore and Wacziarg (2009) and compare ACD with their genetic-distance measures. In the bilateral specifications, ACD is positively related to income differences when entered on its own and generally remains significant when weighted genetic distance is included, although the estimates become less precise. The pattern is clearer in the US-relative specifications, where ACD remains positive and precisely estimated both on its own and alongside relative genetic distance. Taken together, these results suggest that ACD contains information about historical distance that overlaps with, but is not fully captured by, conventional measures of inherited barriers.

¹⁷The underlying association remains positive in every five-year window (Appendix D, Figure A12).

Table 10: Income Gap with FB Cultural Distance and Human-Capital Distance

	Income Gap			
	(1)	(2)	(3)	(4)
<i>Panel A: PAM10 Cosine Distance</i>				
ACD	0.135*** (0.029)	0.028 (0.022)	0.068*** (0.021)	0.022 (0.018)
FB Cultural Distance		0.430*** (0.053)		0.240*** (0.052)
Human-capital distance			0.411*** (0.038)	0.325*** (0.046)
Controls	Yes	Yes	Yes	Yes
Observations	9,041	9,041	9,041	9,041
R^2	0.601	0.663	0.693	0.708
Within R^2	0.098	0.238	0.306	0.341
<i>Panel B: WMM10 HW Distance</i>				
ACD	0.145*** (0.034)	-0.002 (0.026)	0.042* (0.022)	-0.022 (0.019)
FB Cultural Distance		0.442*** (0.057)		0.259*** (0.055)
Human-capital distance			0.414*** (0.038)	0.327*** (0.045)
Controls	Yes	Yes	Yes	Yes
Observations	9,041	9,041	9,041	9,041
R^2	0.600	0.662	0.691	0.708
Within R^2	0.096	0.237	0.302	0.341

Notes: Every column uses the exact common sample, the full baseline dyadic controls, and country1 and country2 fixed effects. Column (1) includes ACD alone; column (2) adds FB Cultural Distance; column (3) adds Human-capital distance; and column (4) includes both. Human-capital distance is the PWT log measure used in the approved human-capital exercise. All continuous variables are standardized within the common sample. Standard errors are two-way clustered on country1 and country2. The coefficient movements describe overlap and are not mediation estimates. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.4 ACD and International Exchange: Trade and Migration

Estimating bilateral exchange via PPML demonstrates that ACD acts as a persistent demographic resistance term. Table 11 turns to bilateral trade and migration. Because both outcomes are estimated in levels using PPML, the coefficients are semi-elasticities and are not directly comparable with the standardized OLS coefficients.

The controlled specifications also account for policies that directly affect bilateral flows. For trade, we include regional and free trade agreements and joint GATT/WTO membership from the CEPII Gravity database (Conte et al., 2022). Since trade is averaged over 2000–2019, these variables measure the share of years in which each arrangement was in place. For migration, we similarly control for bilateral labour agreements using Shrestha et al. (2026), again measured as the share of years covered over 2000–2019.¹⁸ These controls help separate the association with ACD from some of the more immediate policy determinants of trade and migration.

ACD is negatively associated with both outcomes. In the controlled specifications, the PAM10 coefficient is -0.114 for trade and -0.130 for migration. The corresponding WMM10 coefficients are -0.140 and -0.184 . The estimates become smaller after adding the standard gravity controls and outcome-specific policy variables, but remain negative and precisely estimated.

These results are consistent with viewing ACD as an additional source of bilateral resistance. The cultural-distance results above provide one possible interpretation: populations with more similar ancestral backgrounds may also share greater familiarity, norms, information, or networks, reducing some of the costs involved in trading or moving across borders. The regressions do not allow us to distinguish among these channels, and some of the remaining association may reflect historical links that are not fully captured by the standard gravity controls. Another important point is that ACD retains information about bilateral frictions even after the standard geographic, historical, and policy determinants of cross-border flows are taken into account.¹⁹

¹⁸The original agreement data are directed and annual. We symmetrize the pair-year indicator so that an agreement recorded in either direction counts for the dyad, and then average it over 2000–2019. Dyads outside the source country panel are coded as having no recorded agreement and are flagged in the replication files.

¹⁹The relationship is not specific to a particular dataset or period. It remains visible across alternative trade sources and reporting sides, five-year windows, and alternative migration datasets; see Figures A10–A11.

Table 11: ACD, Trade, and Migration

	Trade Intensity		Migration Intensity	
	(1)	(2)	(3)	(4)
<i>Panel A: PAM10 Cosine Distance</i>				
ACD	-0.609*** (0.043)	-0.114*** (0.026)	-0.751*** (0.061)	-0.130** (0.062)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	13,037	13,037	13,155	13,155
Pseudo- R^2	0.181	0.245	0.213	0.348
<i>Panel B: WMM10 HW Distance</i>				
ACD	-0.687*** (0.046)	-0.140*** (0.032)	-0.885*** (0.065)	-0.184*** (0.070)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	13,037	13,037	13,155	13,155
Pseudo- R^2	0.186	0.245	0.224	0.348

Notes: Trade Intensity is averaged over 2000–2019, whereas Migration Intensity averages the actual UN DESA stock vintages in 2000, 2005, 2010, 2015, and 2020; no intermediate values are imputed. Both outcomes are estimated by PPML in levels. Continuous regressors are standardized, so the ACD coefficients are semi-elasticities for a one-standard-deviation increase in ACD. Parsimonious columns include ACD and country1 and country2 fixed effects. Controlled trade models add the baseline controls and RTA, FTA, and joint GATT/WTO shares. Controlled migration models add the baseline controls and the bilateral-labour-agreement share averaged over 2000–2019. Each parsimonious–controlled comparison uses the same effective sample. Standard errors are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5 ACD and Political Alignment & Institutional Differences

ACD is positively associated with foreign-policy divergence and governance-quality disparities. Table 12 examines whether the relationship with ACD also appears in political alignment and institutional differences. In the controlled specifications, the PAM10 coefficient is 0.148 for UN-voting distance and 0.053 for WGI distance. The corresponding WMM10 estimates are 0.167 and 0.045. All four coefficients are positive and precisely estimated, although the relationship is noticeably weaker for institutional distance.

These results suggest that countries with more dissimilar ancestral compositions also tend to be further apart in their international political positions and, to a lesser extent, in the quality and structure of their institutions. We do not interpret either outcome as a channel through which ACD affects the other outcomes considered above. Rather,

they provide additional evidence that the historical differences summarized by ACD are associated with a broader set of contemporary cross-country differences than culture alone.

20

Table 12: ACD and Political and Institutional Distance

	UN-Voting Distance		WGI Distance	
	(1)	(2)	(3)	(4)
<i>Panel A: PAM10 Cosine Distance</i>				
ACD	0.203*** (0.027)	0.148*** (0.030)	0.064*** (0.013)	0.053*** (0.016)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	12,718	12,718	13,199	13,199
R^2	0.437	0.506	0.699	0.701
Within R^2	0.063	0.178	0.012	0.022
<i>Panel B: WMM10 HW Distance</i>				
ACD	0.232*** (0.032)	0.167*** (0.034)	0.060*** (0.013)	0.045*** (0.015)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	12,718	12,718	13,199	13,199
R^2	0.444	0.507	0.698	0.701
Within R^2	0.074	0.179	0.010	0.019

Notes: UN-Voting Distance and WGI Distance are both averaged over 2000–2019; WGI Distance uses the six-indicator cosine measure. Parsimonious columns include ACD and country1 and country2 fixed effects. Controlled columns add the baseline dyadic controls listed in Table 7. Each parsimonious–controlled comparison uses the same sample. All continuous outcomes and regressors are standardized within that sample. Standard errors are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.6 Comparing ACD with Alternative Bilateral-Resistance Terms

The results so far show that ACD is systematically associated with a broad range of bilateral outcomes. We next ask a related but distinct question: how much additional information do PAM10 and WMM10 contribute relative to the conventional bilateral-resistance terms already used in our empirical specifications? The relevant benchmark is therefore not a separately selected set of predictors, but the same geographic, historical, institutional,

²⁰The UN-voting relationship is also stable over time. WMM10 remains positive in every five-year window, while the PAM10 estimates are precisely estimated from 1960 onward. Appendix D, Figure A8 reports these estimates.

and policy-related controls against which ACD has been evaluated throughout the paper. This comparison allows us to assess both whether ACD captures a distinct dimension of cross-country similarity beyond these conventional measures and how its explanatory contribution compares with that of the individual controls.²¹

We compare the contribution of ACD with these bilateral-resistance terms using two complementary measures of model fit. First, we measure the deterioration in fit when each bilateral variable is removed from the full specification. For the OLS outcomes, we report the corresponding partial R^2 ; for trade and migration, we use the proportional increase in PPML deviance. This exercise captures the explanatory contribution of each variable conditional on all other bilateral controls. Second, we compute Shapley values, which average a variable’s marginal contribution to model fit across all possible orderings of the regressors. The two approaches therefore answer slightly different questions: partial fit isolates the information that remains uniquely associated with a variable once the other controls are present, whereas Shapley values also assign credit for explanatory content shared among correlated bilateral measures.

Table 13 reports the individual-variable results. PAM10 and WMM10 each rank first among the 12 bilateral terms for both measures of cultural distance and for the income gap. PAM10 ranks fourth by partial fit and fifth by its Shapley contribution for trade, and eighth and fifth, respectively, for migration. WMM10 ranks fourth under both criteria for trade. For migration, it ranks seventh by partial fit but fourth by its Shapley contribution, suggesting that more of its explanatory content is shared with geography and historical links.

²¹To recall, the geographic and topological controls include geodesic distance, absolute differences in latitude and longitude, shared-border and shared-sea indicators, and indicators for whether either country is landlocked or an island. Historical and institutional controls include common official language, common legal origin, a direct colonial relationship, and a common colonizer. Trade specifications additionally control for the shares of years between 2000 and 2019 in which a dyad is covered by a regional trade agreement, a free trade agreement, or joint GATT/WTO membership, while migration specifications include the share of years covered by a bilateral labour agreement.

Table 13: Individual Ranking of PAM10 and WMM10 among Bilateral-Resistance Terms

Outcome	Rank by partial fit	Partial fit	Rank by Shapley share	Shapley share
<i>Panel A: PAM10 Cosine Distance</i>				
FB Cultural Distance	1st (of 12)	0.068	1st (of 12)	31.000%
WVS Cultural Distance	1st (of 12)	0.068	1st (of 12)	39.400%
Income Gap	1st (of 12)	0.020	1st (of 12)	36.700%
Trade Intensity	4th (of 15)	0.015	5th (of 15)	9.700%
Migration Intensity	8th (of 13)	0.011	5th (of 13)	9.400%
<i>Panel B: WMM10 HW Distance</i>				
FB Cultural Distance	1st (of 12)	0.088	1st (of 12)	37.800%
WVS Cultural Distance	1st (of 12)	0.057	1st (of 12)	37.500%
Income Gap	1st (of 12)	0.016	1st (of 12)	35.300%
Trade Intensity	4th (of 15)	0.019	4th (of 15)	10.900%
Migration Intensity	7th (of 13)	0.014	4th (of 13)	10.700%

Notes: Each panel separately compares the indicated ACD measure with the individual bilateral terms in each complete model; PAM10 and WMM10 never enter jointly. A position reported as “1st (of 12),” for example, means that the indicated ACD measure has the largest contribution among the 12 bilateral terms. Partial fit is leave-one-term-out partial R^2 for the OLS outcomes and proportional deviance reduction for PPML trade and migration. Shapley shares allocate within fit across individual terms. Cultural and income Shapley values are exact; trade and migration use deterministic balanced permutations. OLS and PPML fit magnitudes are not directly comparable.

The comparison therefore differs somewhat across outcomes. For culture and income, PAM10 and WMM10 stand out as the strongest individual bilateral variables and account for a sizeable share of explanatory fit even when they are compared with broad groups of conventional controls. For trade and migration, geography, historical links, and policy variables jointly explain much more, although WMM10 still contributes additional within fit. This is broadly what one would expect in gravity-type applications, where physical distance, borders, historical ties, and policy arrangements are already powerful predictors of bilateral flows.²²

²²We also repeat the exercise after grouping the conventional controls into broader geographic, historical, and, where relevant, policy blocks. Removing a block from the full model provides its partial contributions and Shapley Values. The block-level results tell much the same story, PAM10’s Shapley shares are 37.100 percent for FB Cultural Distance, 44.600 percent for WVS Cultural Distance, 39.800 percent for income, 14.600 percent for trade, and 16.100 percent for migration. The corresponding WMM10 shares are 44.300, 43.200, 38.700, 16.300, and 18.200 percent.

5.7 ACD complements Genetic, Linguistic, and Religious Distance Measures

This comparison asks how much ACD overlaps with other commonly used measures of historical and cultural distance, in particular genetic, linguistic, and religious distance. This is a demanding comparison. Contemporary weighted genetic distance already incorporates the ancestry composition of present-day populations, while linguistic and religious differences are themselves partly shaped by long-run migration and historical contact. The question is therefore not whether these measures are independent, but whether ACD retains explanatory content once these closely related dimensions are taken into account.²³

Across the twelve specifications, all ancestry coefficients remain positive and precisely estimated, although their magnitudes decline once the alternative distance measures are included. The competing measures generally enter with the expected signs as well. The results therefore point to substantial overlap, but not complete redundancy. Genetic distance captures the relatedness of ancestral populations, whereas ACD captures how those ancestral origins are combined across present-day populations.

This distinction matters empirically. Countries with similar average genetic relatedness can still differ considerably in the composition of their ancestral origins, just as countries with relatively distant ancestral populations may share similar mixtures. The comparison regressions show that ACD continues to explain variation in cultural distance after accounting for genetic relatedness, linguistic distance, and religious distance. At the same time, the attenuation of the coefficients makes clear that these measures capture partly overlapping features of long-run history.

²³The FB distance estimations appear in Appendix D, Table A11.

Table 14: WVS Cultural Distance: Comparison with Genetic, Linguistic, and Religious Distance

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: PAM10 Cosine Distance</i>						
ACD	0.221*** (0.043)	0.183*** (0.045)	0.179*** (0.035)	0.162*** (0.038)	0.144*** (0.039)	0.131*** (0.039)
Genetic Distance (Contemporary)		0.136** (0.062)			0.132** (0.058)	0.118* (0.067)
Linguistic Distance			0.229*** (0.055)		0.221*** (0.057)	
Religious Distance				0.347*** (0.050)		0.341*** (0.053)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,667	5,204	5,667	5,667	5,204	5,204
R^2	0.581	0.588	0.597	0.622	0.602	0.627
Within R^2	0.204	0.202	0.234	0.281	0.229	0.278
<i>Panel B: WMM10 HW Distance</i>						
ACD	0.217*** (0.044)	0.165*** (0.046)	0.173*** (0.035)	0.147*** (0.037)	0.124*** (0.038)	0.098*** (0.036)
Genetic Distance (Contemporary)		0.143** (0.064)			0.140** (0.060)	0.138** (0.067)
Linguistic Distance			0.239*** (0.057)		0.234*** (0.059)	
Religious Distance				0.351*** (0.052)		0.348*** (0.054)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,667	5,204	5,667	5,667	5,204	5,204
R^2	0.576	0.582	0.593	0.617	0.598	0.623
Within R^2	0.194	0.190	0.227	0.272	0.222	0.269

Notes: Comparison regressions; the outcome is WVS Cultural Distance, using the 2021 benchmark. The two panels estimate PAM10 Cosine Distance and WMM10 HW Distance separately; the measures never enter jointly. Columns (1)–(6) report, in order, the baseline ancestry specification; Genetic Distance (Contemporary) alone; Linguistic Distance alone; Religious Distance alone; Genetic Distance (Contemporary) plus Linguistic Distance; and Genetic Distance (Contemporary) plus Religious Distance. Religious Distance is measured in 2020. All columns include the baseline dyadic controls and country1 and country2 fixed effects. Continuous outcomes and distance regressors are standardized to mean 0 and standard deviation 1. Standard errors are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.8 Decomposing ACD: Cultural Clusters in 1500 vs. Post-1500 Demographic Change

As an additional descriptive exercise, we explore whether the relationship between ACD and contemporary cultural distance can be separated into a pre-1500 component and subsequent changes in population composition. To do so, we construct a simple pre-1500 benchmark (and for WMM HW Distance). Let c_i denote country i 's PAM10 cluster and

define:

$$D_{ij}^{1500} \equiv \begin{cases} 0, & \text{if } c_i = c_j, \\ 1, & \text{if } c_i \neq c_j. \end{cases} \quad (8)$$

The difference between this pre-1500 benchmark and current ACD is:

$$\Delta D_{ij} \equiv D_{ij}^{PAM10} - D_{ij}^{1500}. \quad (9)$$

For the purposes of this exercise, we separate this difference into two nonnegative components:

$$D_{ij}^{\text{apart}} \equiv |\Delta D_{ij}| \mathbf{1}\{\Delta D_{ij} > 0\}, \quad (10)$$

$$D_{ij}^{\text{closer}} \equiv |\Delta D_{ij}| \mathbf{1}\{\Delta D_{ij} < 0\}. \quad (11)$$

The first component captures differences in the direction of greater ancestral distance, while the second captures differences in the direction of greater ancestral similarity. By construction, the following identity holds for every dyad:

$$D_{ij}^{PAM10} = D_{ij}^{1500} + D_{ij}^{\text{apart}} - D_{ij}^{\text{closer}}. \quad (12)$$

This decomposition provides a simple way of asking whether the associations documented above appear to reflect primarily differences already present in the pre-1500 benchmark, or whether subsequent changes in population composition may also contain relevant information. The pre-1500 benchmark and the two change components enter jointly, so the main distances are not included separately.

The exercise should be interpreted as descriptive rather than as an attempt to identify the effect of post-1500 migration. If the relationship with contemporary cultural distance were concentrated entirely in the pre-1500 benchmark, one would expect relatively little systematic association with the subsequent change components once that benchmark is included. Associations with the latter would instead be consistent with post-1500 population re-composition containing additional information about present-day cultural differences. As elsewhere in the paper, we do not suggest and do not treat historical migration as exogenous.

Table 15 examines how these components relate to contemporary cultural distance. The pre-1500 benchmark is positively associated with cultural distance, while the moving-closer component is negative and precisely estimated across specifications. In standardized terms, a one-standard-deviation larger movement in the direction of a more similar ancestral composition is associated with approximately 0.190–0.290 standard deviations less cultural distance today.

This exercise provides suggestive evidence that the baseline association is not exhausted

by the coarse pre-1500 benchmark. At the same time, the decomposition has important limitations. Migration destinations were not randomly chosen, cultural similarity may itself have influenced historical migration patterns, and the binary benchmark necessarily provides a simplified representation of ancestral differences at 1500. We therefore view the results mainly as an informative descriptive exercise. Post-1500 demographic convergence appears to contain information about contemporary cultural similarity even after accounting for initial cluster membership. The resulting patterns are consistent with five centuries of global population reshuffling remaining reflected in contemporary cultural proximity.

Table 15: Decomposing PAM10 and WMM10 ACD and Contemporary Cultural Distance

	FB Cultural Distance		WVS Cultural Distance	
	(1)	(2)	(3)	(4)
<i>Panel A: PAM10 Cosine Distance</i>				
ACD at 1500	1.025*** (0.081)	0.589*** (0.080)	0.828*** (0.117)	0.557*** (0.129)
Moving apart	0.027 (0.021)	0.012 (0.020)	0.055** (0.021)	0.037* (0.019)
Moving closer	-0.153*** (0.023)	-0.130*** (0.022)	-0.170*** (0.041)	-0.178*** (0.035)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	12,399	12,399	5,667	5,667
<i>Panel B: WMM10 HW Distance</i>				
ACD at 1500	1.334*** (0.125)	0.884*** (0.127)	1.076*** (0.148)	0.781*** (0.158)
Moving apart	0.009 (0.025)	0.002 (0.022)	0.063** (0.028)	0.048* (0.025)
Moving closer	-0.292*** (0.039)	-0.224*** (0.037)	-0.205*** (0.058)	-0.186*** (0.051)
Additional controls	No	Yes	No	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	12,399	12,399	5,667	5,667

Notes: Each panel enters the three decomposition components jointly; current PAM10 Cosine Distance and current WMM10 HW Distance do not enter separately. PAM10 and WMM10 are estimated in separate panels and never enter jointly. Moving apart and moving closer are nonnegative absolute-change magnitudes and are never positive simultaneously. The two change components and the outcomes are standardized; Distance at 1500 remains a dummy variable. All models include country1 and country2 fixed effects. Controlled specifications add the full baseline dyadic controls. Standard errors, in parentheses, are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Robustness Checks

To verify that the empirical relationships documented in Section 5 capture a robust dimension of bilateral relatedness rather than omitted historical variables, methodological artifacts, or specific regional subsamples, we conduct extensive robustness checks across four principal domains. First, we check for omitted (deep-roots) variable bias. ACD may be capturing other persistent differences inherited from ancestral populations. In particular, we try to distinguish ACD from environmental/agricultural exposure (Agro-Climatic Distance) and pre-industrial institutional organization (Ethnographic Distance). Second, we test for data or algorithmic dependency. The results may depend on how ancestral origins are grouped and how similarity between population profiles is measured. We demonstrate that ACD does not depend on genetic data, Partitioning Around Medoids (PAM) clustering, or specific optimal-transport distance formulas by validating it against the Historical-Cultural Nine-Sphere Classification and unclustered Countrywise Cosine Distance. Third, we check for sample heterogeneity and Euro-Centric / New-World biases because the baseline relationship may be driven by dyads involving a particular region of the world. Finally, in section 6.4 we consider several additional robustness checks for endogeneity and mechanisms alternative to ACD.

6.1 Controlling for the Effects of Agro-Climatic Distance and Ethnographic Distance

Galor et al. (2026) argues that adaptation to different ecological niches is an important factor behind cross-societal variation in cultural traits. It is possible that ACD is partly picking up the differences in ecological environments as a potential deep-rooted source of cultural variation. To account for this, we add to our baseline model a measure of agro-climatic distance (*AgroClimDist*) as the cosine distance between country profiles constructed from 27 agricultural, land-use, crop suitability, and climatic characteristics. The sources and construction of *AgroClimDist* is documented in Appendix F.

Second, we probe the sensitivity of our results to accounting for bilateral differences in inherited social organization rather than physical environment. We compute a measure *EthnoDist* that compares ancestry-weighted profiles of pre-industrial settlement, kinship, subsistence, and gendered agricultural organization. The sources and construction of *EthnoDist* is documented in Appendix G.

ACD, *AgroClimDist* and *EthnoDist* capture related but distinct aspects of ancestral history. *AgroClimDist* describes the environments in which ancestral populations lived, *EthnoDist* describes features of their social organization, and ACD describes how those populations are represented in present-day countries. Using the same sample throughout ensures that changes in the ACD coefficient are driven by the additional controls rather than

by changes in country coverage. Table 16 shows that both *AgroClimDist* and *EthnoDist* are positively associated with contemporary cultural distance. At the same time, the ACD coefficient remains positive and statistically significant at the one-percent level for both cultural outcomes and under both ancestry measures. Its magnitude falls when the additional historical controls are introduced, indicating that the measures share some explanatory content, but the association with ACD remains sizeable.

The results suggest that the baseline relationship is not simply a reflection of ancestral ecology or pre-industrial social organization. Differences in these characteristics clearly matter for contemporary cultural distance, but they do not fully account for the information contained in the present-day composition of ancestral origins.

Table 16: ACD with AgroClimDist and EthnoDist controls

	PAM10 Cosine Distance				WMM10 HW Distance			
	Baseline	+AgroClimDist	+EthnoDist	+AgroClimDist +EthnoDist	Baseline	+AgroClimDist	+EthnoDist	+AgroClimDist +EthnoDist
<i>Panel A: FB Cultural Distance</i>								
ACD	0.204*** (0.027)	0.180*** (0.024)	0.186*** (0.024)	0.169*** (0.022)	0.254*** (0.038)	0.224*** (0.035)	0.235*** (0.036)	0.212*** (0.034)
AgroClimDist		0.188*** (0.029)		0.159*** (0.028)		0.175*** (0.028)		0.147*** (0.028)
EthnoDist			0.160*** (0.027)	0.134*** (0.026)			0.156*** (0.027)	0.133*** (0.026)
Observations	11,323	11,323	11,323	11,323	11,323	11,323	11,323	11,323
Within R^2	0.328	0.356	0.355	0.374	0.341	0.365	0.367	0.383
<i>Panel B: WVS Cultural Distance</i>								
ACD	0.228*** (0.047)	0.202*** (0.041)	0.194*** (0.039)	0.180*** (0.036)	0.218*** (0.048)	0.185*** (0.042)	0.184*** (0.040)	0.165*** (0.037)
AgroClimDist		0.319*** (0.058)		0.226*** (0.048)		0.319*** (0.059)		0.225*** (0.049)
EthnoDist			0.362*** (0.052)	0.316*** (0.047)			0.366*** (0.053)	0.320*** (0.048)
Observations	5,149	5,149	5,149	5,149	5,149	5,149	5,149	5,149
Within R^2	0.228	0.297	0.356	0.388	0.215	0.284	0.346	0.378

Notes: The left block uses PAM10 Cosine Distance as the ACD measure and the right block uses WMM10 HW Distance; the two never enter jointly. Agro-Climatic Distance (AgroClimDist) and Ethnographic Distance (EthnoDist) use their baseline constructions. Within each panel the estimation sample is held fixed across all columns at the dyads for which the outcome, ACD, baseline controls, AgroClimDist, and EthnoDist are all observed. Outcomes and continuous distance regressors are standardized. All specifications include the baseline dyadic controls and country1 and country2 fixed effects; standard errors (in parentheses) are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.2 Alternative Ways to Construct ACD

To establish that our empirical findings do not depend on genetic distance inputs, the PAM clustering algorithm, or specific optimal-transport distance formulas, we need to

examine whether the results depend on how ACD is constructed. Table 17 compares two alternative measures of ancestral distance with our baseline measures of ACD across all core outcomes. The first one, Countrywise Cosine Distance, works directly with the original WMM ancestry shares across 163 source countries, without grouping origins into broader clusters. Two countries are therefore considered similar only when their populations trace comparable shares of ancestry to the same source countries. This construction uses neither the PAM10 clustering nor genetic-distance information.

The second alternative measure, Historical-Cultural ACD, aggregates the WMM ancestry shares into nine broad historical-cultural spheres. We construct this classification specifically to approximate major affinities among populations around 1500, using historical evidence on the institutional, religious, linguistic, and broader cultural characteristics of pre-Columbian societies. The resulting nine-sphere partition is an independently constructed historical classification rather than a grouping derived from the PAM10 algorithm or from genetic-distance data. Each contemporary country is represented by the shares of its population tracing ancestry to these nine spheres, and distance is calculated from the similarity of these ancestry profiles. In constructing the classification, we follow the methodological spirit of Putterman and Weil (2010)’s World Migration Matrix, rather than relying on a single quantitative criterion. We combine information from historical sources with our judgment to assign populations to the historical-cultural sphere that best represents their position around 1500. The sphere definitions and assignments themselves are developed specifically for this paper. Detailed definitions, historical rationale, sources, and origin-country assignments are reported in Appendix C, Table A9.²⁴

The results are consistent across the four constructions. Greater ACD is associated with greater cultural distance and larger income and human-capital gaps, and with lower trade and migration. Every coefficient has the predicted sign and is statistically significant at least at the five-percent level. The important result is therefore the stability of the sign and statistical relationship across quite different ways of constructing ancestry distance. Neither clustering itself, the use of genetic information to define the groups, nor a particular distance metric appears to drive the baseline findings.²⁵

²⁴The PAM10 partition nevertheless aligns closely, though not perfectly, with our independently constructed historical-cultural classification. Among the 1,531 unordered country dyads whose native populations belong to the same PAM10 cluster, 1,164 (76.0%) are also assigned to the same historical-cultural sphere.

²⁵Beyond these constructions, we considered several alternative clustering and aggregation methods, as well as classifications based on non-genetic information. The main results are broadly robust across these alternative approaches. (Appendix C for details)

Table 17: Alternative Clustering and Distance Metrics across Cultural, Demographic, and Economic Outcomes

ACD construction	FB Cultural Distance	WVS Cultural Distance	Income Gap	Human-Capital Gap	Trade Intensity	Migration Intensity
Countrywise Cosine Distance	0.242*** (0.022)	0.267*** (0.044)	0.121*** (0.025)	0.188*** (0.036)	-0.101*** (0.029)	-0.133** (0.062)
Historical-Cultural ACD	0.360*** (0.033)	0.384*** (0.051)	0.249*** (0.032)	0.350*** (0.050)	-0.100** (0.039)	-0.153** (0.076)
PAM10 Cosine Distance	0.196*** (0.025)	0.221*** (0.043)	0.110*** (0.025)	0.184*** (0.034)	-0.114*** (0.026)	-0.130** (0.062)
WMM10 HW Distance	0.247*** (0.036)	0.217*** (0.044)	0.109*** (0.028)	0.268*** (0.045)	-0.140*** (0.032)	-0.184*** (0.070)

Notes: Each cell is the coefficient on the indicated ACD measure from a separate baseline regression. FB Cultural Distance, WVS Cultural Distance, Income Gap, and Human-Capital Gap are standardized outcomes estimated by OLS. Income Gap and Human-Capital Gap use 2000–2019 data; the latter is the PWT measure over common observed years. Trade Intensity uses the 2000–2019 average, whereas Migration Intensity averages the actual UN DESA stock vintages in 2000, 2005, 2010, 2015, and 2020, without interpolation. Both flow outcomes are estimated by PPML in levels, so their coefficients are semi-elasticities for a one-standard-deviation increase in ACD and are not comparable in size with the OLS columns. Trade models include period-matched RTA, FTA, and joint GATT/WTO shares; migration models include the bilateral-labour-agreement share averaged over 2000–2019. All models include the baseline dyadic controls and country1 and country2 fixed effects; standard errors (in parentheses) are two-way clustered on country1 and country2. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.3 Robustness to Regional Restrictions of the Sample and Differences in European Ancestry

A substantial share of the cross-country variation in ancestral composition is associated with post-Columbian settlement in the Americas and European settler colonies, raising the concern that the baseline results may be disproportionately driven by dyads that include a New World country. Hence, we evaluate whether our results are driven by New World dyads or European ancestral gradients. Table 18 and Appendix H, Tables A22–A23, address this concern. We report estimates for the full sample, the Old World, samples excluding either Sub-Saharan Africa or the Americas, and the Old World excluding Sub-Saharan Africa. Across all sub-samples, both PAM10 Cosine Distance and WMM10 HW Distance remain positive and significantly associated with cultural distance.

One may alternatively wonder how much of the baseline estimates for ACD are driven by European countries or differences in ancestry from Western Europe. To address this question, we first exclude from the sample any dyad that has at least one country grouped in the Western Europe cluster (C5). Then, we alternatively exclude all observations involving a geographically European country. The main results remain broadly robust to both exclusions (Appendix H, Table A22). Finally, we construct “European Ancestry Share” for each country using WMM. Then we include *Absolute European Share Distance*

(European-share gap) as an additional dyadic control variable. Table A23 presents the results. ACD coefficients remain significant for all outcomes. Not surprisingly, estimated ACD coefficients are somewhat reduced once we control for gaps in European ancestry in cultural distance and income gap regressions. The main takeaway is that ACD is not merely picking up the influence of its specific European ancestry component.

Table 18: Robustness to geographic subsample: excluding the Americas & Oceania, and Sub-Saharan Africa

	FB Cultural Distance	WVS Cultural Distance
<i>PAM10 Cosine Distance</i>		
Full Sample (baseline)	0.196*** (0.025)	0.221*** (0.043)
Old World only (excl. Americas + Oceania)	0.125*** (0.022)	0.132*** (0.040)
Excl. Sub-Saharan Africa	0.105*** (0.022)	0.209*** (0.044)
Old World excl. Sub-Saharan Africa	0.053** (0.023)	0.124*** (0.043)
Excl. Americas only (keeps Oceania)	0.162*** (0.027)	0.214*** (0.054)
<i>WMM10 HW Distance</i>		
Full sample (baseline)	0.247*** (0.036)	0.217*** (0.044)
Old World only (excl. Americas + Oceania)	0.198*** (0.036)	0.124*** (0.040)
Excl. Sub-Saharan Africa	0.097*** (0.024)	0.192*** (0.045)
Old World excl. Sub-Saharan Africa	0.049* (0.026)	0.100** (0.042)
Excl. Americas only (keeps Oceania)	0.231*** (0.038)	0.204*** (0.053)

Notes: Separate regressions on the indicated samples using the baseline specification in Section 5.1. Old World excludes the Pre-Columbian American and Oceanian spheres; the remaining labels are literal exclusions. Sphere definitions follow Appendix C. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.4 Additional Robustness Checks

We conduct several additional exercises (permutation-based validation, sensitivity checks, and Instrumental-Variable analysis) that extend the robustness analysis along four dimen-

sions. First, permutation tests show that the results based on the observed clusters are considerably stronger than those generated by random cluster assignments (Appendix B.2). Second, we vary the clustering method, distance calculation method, time period, data source. The results remain stable (Appendix C, Appendix D). The main results are also similar when ACD is reconstructed using the 2000 population matrix rather than the baseline matrix (Appendix D, Table A12). Third, we consider alternative empirical specifications. The Instrumental-Variable (IV) analysis reports the corresponding first-stage estimates, 2SLS coefficients, and weak-instrument robust inference (Appendix I, Table A24). Lastly, we also replace latitude and longitude differences with bilateral differences in temperature and precipitation, and alternatively include both geographic and climatic controls. The estimated associations between ACD and cultural distance remain positive and precisely estimated across these specifications (Appendix J, Table A25). These analyses indicate that the baseline findings are not specific to arbitrary cluster assignments, New World or European dyads, a particular family of outcomes, a single time window or population snapshot, or one particular treatment of geographic and climatic differences.

7 Conclusion

Post-1500 migration did more than redistribute people across territories. It changed the ancestral mixture of national populations and, with it, the structure of cultural, cognitive, and economic similarity between modern societies. We introduce *Ancestral Composition Distance (ACD)* to bring this vital dimension of demographic history directly into dyadic empirical analysis. Using the 1960 version of the World Migration Matrix (Chanda et al., 2014; Putterman & Weil, 2010), we describe each country by the shares of its population descending from a common set of pre-1500 ancestral spheres. The distance between two such profiles records whether their populations were assembled from similar origins in similar proportions.

The cultural results provide the most direct test of this idea. Country pairs with more dissimilar ancestry profiles are further apart in both the Facebook-based measure of revealed interests and the WVS measure of values and beliefs, alongside with human capital gaps. The coefficients remain sizeable after including country fixed effects and controls for geography, language, legal origin, and colonial history. Adding ancestral ecology and pre-industrial social organization reduces the estimates but does not remove them. Nor do the results depend on one particular partition of the world. Random groupings fail to reproduce the findings. Alternative clustering methods and a historical-cultural classification constructed without genetic information yield similar patterns. The agreement between the results obtained using behavioral and survey measures is important here, since the two are built from very different kinds of evidence.

ACD also tracks differences outside culture, though the strength of the relationship is not uniform. More distant pairs differ more in income, human development, foreign-policy positions, and, more modestly, governance; they also trade less and exchange fewer migrants. The fit rankings help delimit the claim. Both ACD variants are the strongest individual bilateral terms for the cultural outcomes and the income gap, but they play a smaller role in trade and migration, where geographic, historical, and policy variables already account for much of the variation. Hence, ACD is not a substitute for the standard gravity controls. For bilateral flows, it adds a historical-demographic term to models that remain dominated by physical distance, established ties, and formal agreements.

Conceptually, ACD is distinct from weighted genetic distance, even though the two draw on related historical information. Weighted genetic distance averages the relatedness of individuals drawn from two contemporary populations, whereas ACD compares the shares in which broader ancestral origins are represented in those populations. When genetic, linguistic, or religious distance enters the same regression, the ACD coefficient becomes smaller but remains precisely estimated. ACD carries information beyond average inter-group distance. The income regressions reveal a different kind of overlap: once cultural and human-capital differences are included, little of the original ACD coefficient remains. Those findings suggest that the income association attributed to ancestral relatedness overlaps substantially with differences in cultural norms and human capital. But, although that exercise is informative about where the income association lies, it is not a mediation test and does not separate the direction of causality among culture, human capital, and income.

By decomposing ACD into pre-1500 cluster separation (D^{1500}) and post-Columbian demographic movements (D^{closer}), we demonstrate that post-1500 population reshuffling remains clearly legible in modern cultural relationships and places post-Columbian migration more directly in the analysis. Pairs whose native populations began in different ancestral spheres remain more culturally distant today, but the initial separation does not account for the whole relationship. Conditional on that benchmark, pairs whose ancestral compositions became more similar after 1500 are also culturally closer today, whereas the evidence for movement in the opposite direction is much weaker. The decomposition is an accounting exercise, not an identification of the effects of a migration shock. The benchmark is coarse, migration destinations were selected, and unobserved bilateral history may have influenced both population movements and subsequent outcomes. The weaker estimates in the most restrictive Old World samples provide a related and expected qualification that much of the ACD’s useful variation comes from the regions most heavily reshaped after 1500.

For empirical research, ACD provides a predetermined, demography-based measure that belongs alongside the conventional bilateral-resistance terms rather than in place of them. The ancestry shares are fixed at 1960, before the outcomes studied here, so later

trade, migration, or cultural convergence cannot feed back into ACD. This does not make historical migration exogenous, but it avoids the immediate simultaneity problem that arises when contemporary cultural distance is used to explain contemporary interaction. ACD is a measure of bilateral similarity that can be used when culture, human capital, information, or inherited networks are likely to matter but are hard to observe directly. Post-Columbian migration left a demographic map that is still reflected in cultural proximity, development gaps, and cross-border interaction and ACD has the potential to bring that map into dyadic empirical work.

References

- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5), 1369–1401.
- Akren, M. B., Ertan, A. S., & Putterman, L. (2026). Land, power and population density in the global population shifts of 1500–1960 [Available at SSRN]. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.6943319>
- Alesina, A., & Giuliano, P. (2015). Culture and Institutions. *Journal of Economic Literature*, 53(4), 898–944.
- Ang, J. B. (2013). Are Modern Financial Systems Shaped by State Antiquity? *Journal of Banking & Finance*, 37(11), 4038–4058. <https://doi.org/10.1016/j.jbankfin.2013.07.023>
- Aronow, P. M., Samii, C., & Assenova, V. A. (2015). Cluster-Robust Variance Estimation for Dyadic Data. *Political Analysis*, 23(4), 564–577.
- Ashraf, Q., & Galor, O. (2013). The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *American Economic Review*, 103(1), 1–46.
- Bailey, M. A., Strezhnev, A., & Voeten, E. (2017). Estimating Dynamic State Preferences from United Nations Voting Data. *Journal of Conflict Resolution*, 61(2), 430–456.
- Becker, A., Borner, C., Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2026). *The Global Variation in Risk and Time Preferences* [CESifo Working Paper No. 12589].
- Becker, A., Enke, B., & Falk, A. (2020). Ancient Origins of the Global Variation in Economic Preferences. *AEA Papers and Proceedings*, 110, 319–323.
- Bertocchi, G., Dimico, A., & Tedeschi, G. L. (2026). *Ancestral Inequality and Preferences for Redistribution* [CEPR Discussion Paper No. DP21490].
- 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. <https://doi.org/10.1162/0033555300554953>
- Bisin, A., & Verdier, T. (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, 97(2), 298–319.
- Bisin, A., & Verdier, T. (2011). The Economics of Cultural Transmission and Socialization. In J. Benhabib, A. Bisin, & M. O. Jackson (Eds.), *Handbook of Social Economics* (pp. 339–416, Vol. 1). North-Holland. <https://doi.org/10.1016/B978-0-444-53187-2.00009-7>
- Bisin, A., & Verdier, T. (2021). Phase Diagrams in Historical Economics: Culture and Institutions. In A. Bisin & G. Federico (Eds.), *The Handbook of Historical Economics* (pp. 491–522). Academic Press. <https://doi.org/10.1016/B978-0-12-815874-6.00024-1>

- Bockstette, V., Chanda, A., & Putterman, L. (2002). States and Markets: The Advantage of an Early Start. *Journal of Economic Growth*, 7(4), 347–369.
- Bove, V., & Gokmen, G. (2018). Genetic Distance, Trade, and the Diffusion of Development. *Journal of Applied Econometrics*, 33(4), 617–623.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference with Multiway Clustering. *Journal of Business & Economic Statistics*, 29(2), 238–249.
- Chanda, A., Cook, C. J., & Putterman, L. (2014). Persistence of Fortune: Accounting for Population Movements, There Was No Post-Columbian Reversal. *American Economic Journal: Macroeconomics*, 6(3), 1–28.
- Coimbra Vieira, C., Lohmann, S., & Zagheni, E. (2024). The Value of Cultural Similarity for Predicting Migration: Evidence from Food and Drink Interests in Digital Trace Data. *Population and Development Review*, 50(1), 149–176. <https://doi.org/10.1111/padr.12607>
- Comin, D., Easterly, W., & Gong, E. (2010). Was the Wealth of Nations Determined in 1000 BC? *American Economic Journal: Macroeconomics*, 2(3), 65–97.
- Comin, D., & Hobijn, B. (2010). An Exploration of Technology Diffusion. *American Economic Review*, 100(5), 2031–2059. <https://doi.org/10.1257/aer.100.5.2031>
- Conte, M., Cotterlaz, P., & Mayer, T. (2022). *The CEPII Gravity Database* [CEPII Working Paper No. 2022-05].
- Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson Estimation with High-Dimensional Fixed Effects. *The Stata Journal*, 20(1), 95–115.
- D’Amato, M., & Russo, F. F. (2026). Cultural Doorways in the Barriers to Development. *Journal of Economic Growth*, 31(1), 125–178.
- Desmet, K., Ortuno-Ortin, I., & Wacziarg, R. (2017). Culture, Ethnicity, and Diversity. *American Economic Review*, 107(9), 2479–2513.
- Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review*, 105(10), 3150–3182. <https://doi.org/10.1257/aer.20130954>
- Fensore, I., Legge, S., & Schmid, L. (2022). Ancestry and International Trade. *Journal of Comparative Economics*, 50(1), 33–51.
- Fouka, V., & Tabellini, M. (2025). *Culture and Contemporary Political Preferences* [NBER Working Paper No. 33786].
- Fulford, S. L., Petkov, I., & Schiantarelli, F. (2020). Does It Matter Where You Came From? Ancestry Composition and Economic Performance of US Counties, 1850–2010. *Journal of Economic Growth*, 25(3), 341–380.
- Galor, O., Klemp, M., & Wainstock, D. C. (2024). *Roots of Cultural Diversity* [IZA Discussion Paper No. 17481].
- Galor, O., & Ozak, O. (2016). The Agricultural Origins of Time Preference. *American Economic Review*, 106(10), 3064–3103.

- Galor, O., Klemp, M., & Wainstock, D. C. (2026). *Cultural evolution* (NBER Working Paper No. 35656). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w35656>
- Gaskin, T., & Abel, G. J. (2026). Deep Learning Four Decades of Human Migration. *Nature*, 655, 148–157. <https://doi.org/10.1038/s41586-026-10611-7>
- Ghatak, M., & Karaivanov, A. (2014). Contractual Structure in Agriculture with Endogenous Matching. *Journal of Development Economics*, 110, 239–249.
- Giavazzi, F., Petkov, I., & Schiantarelli, F. (2019). Culture: Persistence and Evolution. *Journal of Economic Growth*, 24(2), 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.
- Guiso, L., Sapienza, P., & Zingales, L. (2009). Cultural Biases in Economic Exchange? *Quarterly Journal of Economics*, 124(3), 1095–1131.
- Guiso, L., Sapienza, P., & Zingales, L. (2024). *Embedded Culture as a Source of Comparative Advantage* [NBER Working Paper No. 33268]. <https://doi.org/10.3386/w33268>
- Guiso, L., Sapienza, P., & Zingales, L. (2006). Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives*, 20(2), 23–48. <https://doi.org/10.1257/jep.20.2.23>
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law*, 3(2), 220–246.
- Michalopoulos, S., & Xue, M. M. (2021). Folklore. *The Quarterly Journal of Economics*, 136(4), 1993–2046. <https://doi.org/10.1093/qje/qjab003>
- Nunn, N. (2009). The Importance of History for Economic Development. *Annual Review of Economics*, 1(1), 65–92.
- Nunn, N. (2012). Culture and the Historical Process. *Economic History of Developing Regions*, 27(S1), S108–S126. <https://doi.org/10.1080/20780389.2012.664864>
- Nunn, N. (2020). The Historical Roots of Economic Development. *Science*, 367(6485), eaaz9986.
- Nunn, N. (2021). History as Evolution. In A. Bisin & G. Federico (Eds.), *The Handbook of Historical Economics* (pp. 41–91). Academic Press.
- Obradovich, N., Ozak, O., Martin, I., Ortuno-Ortin, I., Awad, E., Cebrian, M., Cuevas, R., Desmet, K., Rahwan, I., & Cuevas, A. (2022). Expanding the Measurement of Culture with a Sample of Two Billion Humans. *Journal of the Royal Society Interface*, 19(190), 20220085.
- Ozden, C., Parsons, C. R., Schiff, M., & Walmsley, T. L. (2011). Where on Earth Is Everybody? The Evolution of Global Bilateral Migration 1960–2000. *World Bank Economic Review*, 25(1), 12–56.

- Pellegrino, B., Spolaore, E., & Wacziarg, R. (2025). Barriers to Global Capital Allocation. *Quarterly Journal of Economics*, 140(4), 3067–3131. <https://doi.org/10.1093/qje/qjaf031>
- Pemberton, T. J., DeGiorgio, M., & Rosenberg, N. A. (2013). Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation. *G3: Genes, Genomes, Genetics*, 3(5), 891–907.
- 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.
- Rychlowska, M., Miyamoto, Y., Matsumoto, D., Hess, U., Gilboa-Schechtman, E., Kambale, S., Muluk, H., Masuda, T., & Niedenthal, P. M. (2015). Heterogeneity of Long-History Migration Explains Cultural Differences in Reports of Emotional Expressivity and the Functions of Smiles. *Proceedings of the National Academy of Sciences*, 112(19), E2429–E2436. <https://doi.org/10.1073/pnas.1413661112>
- Santos Silva, J. M. C., & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4), 641–658.
- Shrestha, M., Kaila, H., Cha’ngom, N., & Adhikari, S. (2026). Do Bilateral Labor Agreements Increase Migration? Global Evidence from 1960 to 2020. *Journal of Development Economics*, 179, 103673. <https://doi.org/10.1016/j.jdeveco.2025.103673>
- Spolaore, E., & Wacziarg, R. (2009). The Diffusion of Development. *Quarterly Journal of Economics*, 124(2), 469–529.
- Spolaore, E., & Wacziarg, R. (2013). How Deep Are the Roots of Economic Development? *Journal of Economic Literature*, 51(2), 325–369.
- Spolaore, E., & Wacziarg, R. (2016a). Ancestry, Language and Culture. In V. Ginsburgh & S. Weber (Eds.), *The Palgrave Handbook of Economics and Language* (pp. 174–211). Palgrave Macmillan.
- Spolaore, E., & Wacziarg, R. (2016b). War and Relatedness. *Review of Economics and Statistics*, 98(5), 925–939.
- Spolaore, E., & Wacziarg, R. (2018). Ancestry and Development: New Evidence. *Journal of Applied Econometrics*, 33(5), 748–762.
- Sunde, U., Dohmen, T., Enke, B., Falk, A., Huffman, D., & Meyerheim, G. (2022). Patience and Comparative Development. *Review of Economic Studies*, 89(5), 2806–2840.
- United Nations, Department of Economic and Social Affairs, Population Division. (2024). International Migrant Stock 2024. *United Nations*. Retrieved July 27, 2026, from <https://www.un.org/development/desa/pd/content/international-migrant-stock>
- United Nations Development Programme. (2024). Human Development Index [Human Development Reports, <https://hdr.undp.org/data-center>].
- Voeten, E., Strezhnev, A., & Bailey, M. A. (2009). *United Nations General Assembly Voting Data*. Harvard Dataverse. <https://doi.org/10.7910/DVN/LEJUQZ>

APPENDIX

Outline

- A. Country Coverage and Summary Statistics:** Lists the 163-country universe, reports coverage and missingness, and presents lower-triangular correlations for the principal variables together with descriptive statistics for the baseline controls.
- B. Genetic-Distance Clustering and Baseline ACD:** Documents cluster selection, PAM10 membership and medoids, the direct PAM10 Cosine Distance and WMM10 HW Distance constructions, sample construction, and permutation evidence.
- C. Alternative ACD Constructions:** Varies every choice behind the measure—the input distance used to form the clusters, the clustering algorithm, the number of clusters, whether origins are grouped at all, and the rule used to compare ancestry profiles—and rebuilds the measure from a historical-cultural classification that uses no genetic information.
- D. Robustness Checks to Alternative Measurements, Time Periods, Sample:** Holds the measure fixed and varies its surroundings—outcome vintages and reporting sources, the period over which outcomes are dated, and the WMM population vintage. All bilateral outcomes are symmetric, so no directional specifications are reported.
- E. Replication of Spolaore and Wacziarg (2009) Using ACD:** Re-estimates Spolaore and Wacziarg (2009) using the direct bilateral and frontier-relative Income Gap designs using the final WMM10 HW Distance, separately and alongside Contemporary Genetic Distance.
- F. Controlling for Agro-Climatic Distance:** Documents the 27-component construction and its role as an additional control.
- G. Controlling for Ethnographic Distance:** Documents the retained components, exclusions, correlations, and additional-control estimates.
- H. Additional Checks for the Potential "European Driven" Effect on Results:** Excludes the Western Europe cluster and geographically European countries. Also, adds European Share Distance as a new control.
- I. IV Estimations Using Genetic Distance (1500) as Instrument:** Instruments ACD with pre-1500 genetic distance, reports first-stage and weak-instrument diagnostics, and compares the 2SLS estimates across the two cultural outcomes and the income gap.
- J. Controlling For Climatic Differences:** Replaces absolute latitude and longitude differences with absolute temperature and precipitation differences while holding the remaining specification fixed.

A Country Coverage and Summary Statistics

Table A1 lists all 163 countries in the estimation sample, with ISO 3166-1 alpha-3 codes. Country names follow the master dataset.

Table A2 presents Pearson correlations of all variables using all available dyads.

Table A3 presents descriptive statistics of the baseline control variables.

Table A4 presents correlations between baseline controls and the main variables

Table A1: Country Sample

Country	ISO	Country	ISO	Country	ISO
Afghanistan	AFG	Georgia	GEO	Nigeria	NGA
Albania	ALB	Germany	DEU	Norway	NOR
Algeria	DZA	Ghana	GHA	Oman	OMN
Angola	AGO	Greece	GRC	Pakistan	PAK
Argentina	ARG	Guatemala	GTM	Panama	PAN
Armenia	ARM	Guinea	GIN	Papua New Guinea	PNG
Australia	AUS	Guinea-Bissau	GNB	Paraguay	PRY
Austria	AUT	Guyana	GUY	Peru	PER
Azerbaijan	AZE	Haiti	HTI	Philippines	PHL
Bahrain	BHR	Honduras	HND	Poland	POL
Bangladesh	BGD	Hong Kong SAR, China	HKG	Portugal	PRT
Belarus	BLR	Hungary	HUN	Puerto Rico	PRI
Belgium	BEL	India	IND	Qatar	QAT
Belize	BLZ	Indonesia	IDN	Romania	ROU
Benin	BEN	Iran, Islamic Rep.	IRN	Russian Federation	RUS
Bhutan	BTN	Iraq	IRQ	Rwanda	RWA
Bolivia	BOL	Ireland	IRL	Saudi Arabia	SAU
Bosnia and Herzegovina	BIH	Italy	ITA	Senegal	SEN
Botswana	BWA	Jamaica	JAM	Sierra Leone	SLE
Brazil	BRA	Japan	JPN	Singapore	SGP
Bulgaria	BGR	Jordan	JOR	Slovak Republic	SVK
Burkina Faso	BFA	Kazakhstan	KAZ	Slovenia	SVN
Burundi	BDI	Kenya	KEN	Somalia	SOM
Cambodia	KHM	Korea, Dem. Rep.	PRK	South Africa	ZAF
Cameroon	CMR	Korea, Rep.	KOR	Spain	ESP
Canada	CAN	Kuwait	KWT	Sri Lanka	LKA
Cape Verde	CPV	Kyrgyz Republic	KGZ	Sudan	SDN
Central African Republic	CAF	Lao PDR	LAO	Swaziland	SWZ
Chad	TCD	Latvia	LVA	Sweden	SWE
Chile	CHL	Lebanon	LBN	Switzerland	CHE
China	CHN	Lesotho	LSO	Syrian Arab Republic	SYR

Table A1 continued

Country	ISO	Country	ISO	Country	ISO
Colombia	COL	Liberia	LBR	Taiwan	TWN
Comoros	COM	Libya	LBY	Tajikistan	TJK
Congo, Rep.	COG	Lithuania	LTU	Tanzania	TZA
Costa Rica	CRI	Luxembourg	LUX	Thailand	THA
Cote d'Ivoire	CIV	Macedonia, FYR	MKD	Timor-Leste	TLS
Croatia	HRV	Madagascar	MDG	Togo	TGO
Cuba	CUB	Malawi	MWI	Trinidad and Tobago	TTO
Cyprus	CYP	Malaysia	MYS	Tunisia	TUN
Czech Republic	CZE	Mali	MLI	Turkey	TUR
Democratic Republic of Congo	COD	Malta	MLT	Turkmenistan	TKM
Denmark	DNK	Mauritania	MRT	Uganda	UGA
Dominican Republic	DOM	Mauritius	MUS	Ukraine	UKR
Ecuador	ECU	Mexico	MEX	United Arab Emirates	ARE
Egypt, Arab Rep.	EGY	Moldova	MDA	United Kingdom	GBR
El Salvador	SLV	Mongolia	MNG	United States	USA
Equatorial Guinea	GNQ	Morocco	MAR	Uruguay	URY
Eritrea	ERI	Mozambique	MOZ	Uzbekistan	UZB
Estonia	EST	Myanmar	MMR	Venezuela, RB	VEN
Ethiopia	ETH	Namibia	NAM	Vietnam	VNM
Fiji	FJI	Nepal	NPL	Yemen, Rep.	YEM
Finland	FIN	Netherlands	NLD	Zambia	ZMB
France	FRA	New Zealand	NZL	Zimbabwe	ZWE
Gabon	GAB	Nicaragua	NIC		
Gambia, The	GMB	Niger	NER		

Table A2: Correlation Matrix for All Main Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) WMM10 HW Distance	1.000																
(2) PAM10 Cosine Distance	0.922***	1.000															
(3) FB Cultural Distance	0.380***	0.333***	1.000														
(4) WVS Cultural Distance	0.175***	0.220***	0.189***	1.000													
(5) UN-Voting Distance	0.170***	0.171***	0.164***	0.385***	1.000												
(6) Trade Intensity	-0.241***	-0.245***	-0.215***	-0.127***	-0.041***	1.000											
(7) Income Gap	0.062***	0.078***	0.110***	0.364***	0.430***	0.023***	1.000										
(8) Human-Capital Gap	0.234***	0.218***	0.281***	0.444***	0.431***	-0.136***	0.321***	1.000									
(9) Migration Intensity	-0.202***	-0.204***	-0.187***	-0.076***	-0.008	0.474***	-0.037***	-0.066***	1.000								
(10) Genetic Distance (1500)	0.332***	0.315***	0.290***	0.225***	0.003	-0.234***	-0.027***	0.184***	-0.165***	1.000							
(11) Genetic Distance (Contemporary)	0.466***	0.426***	0.327***	0.232***	-0.022**	-0.228***	-0.100***	0.272***	-0.165***	0.694***	1.000						
(12) Religious Distance	0.284***	0.246***	0.204***	0.287***	0.121***	-0.047***	0.062***	0.101***	-0.086***	0.109***	0.173***	1.000					
(13) Linguistic Distance	0.479***	0.458***	0.362***	0.184***	0.060***	-0.231***	-0.063***	0.223***	-0.221***	0.277***	0.348***	0.413***	1.000				
(14) HDI Distance	0.248***	0.253***	0.315***	0.459***	0.345***	-0.153***	0.482***	0.789***	-0.063***	0.251***	0.368***	0.077***	0.233***	1.000			
(15) WGI Distance	0.106***	0.090***	0.317***	0.180***	0.056***	-0.120***	0.016*	0.112***	-0.048***	0.018**	0.068***	0.155***	0.188***	0.120***	1.000		
(16) Agro-Climatic Distance (AgroClimDist)	0.287***	0.246***	0.145***	0.314***	0.372***	-0.160***	0.342***	0.354***	-0.119***	0.130***	0.141***	0.138***	0.206***	0.309***	0.078***	1.000	
(17) Ethnographic Distance (EthnoDist)	0.153***	0.160***	0.136***	0.286***	0.189***	-0.156***	0.117***	0.293***	-0.097***	0.169***	0.206***	0.084***	0.222***	0.356***	0.083***	0.271***	1.000

Notes: Income, Human-Capital, HDI, WGI, UN-Voting, and Trade measures use 2000–2019 data; Migration Intensity uses the observed 2000–2020 five-year UN DESA vintages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Descriptive Statistics of the Baseline Control Variables

Variable	Observations	Mean	Std. dev.	Minimum	Maximum
Geodesic Distance	13,203	7393.5661	4244.8978	130.5370	19803.4220
Latitude Difference	13,203	28.1822	20.6645	.0000	104.4167
Longitude Difference	13,203	62.6923	47.2584	.0000	180.0000
Shared Border	13,203	.0208	.1428	.0000	1.0000
Common Official Language	13,203	.1331	.3397	.0000	1.0000
Common Legal Origin	13,203	.4304	.4952	.0000	1.0000
Colonial Relationship	13,201	.0105	.1017	.0000	1.0000
Common Colonizer	13,203	.0796	.2707	.0000	1.0000
Shared Sea/Ocean	13,203	.1574	.3642	.0000	1.0000
Either Landlocked	13,203	.4035	.4906	.0000	1.0000
Either Island	13,203	.2944	.4558	.0000	1.0000

Notes: Baseline dyadic controls in their original scales; binary controls are indicators.

Table A4: Correlations between Baseline Controls and the Main Variables

	WMM10 HW Distance	PAM10 Cosine Distance	FB Cultural Distance	WVS Cultural Distance	UN-Voting Distance	Trade Intensity	Income Gap	Human-Capital Gap	Migration Intensity	Genetic Distance (1500)	Genetic Distance (Contemporary)	Religious Distance	Linguistic Distance	HDI Distance	WGI Distance	AgroClimDist	EthnoDist
Geodesic Distance	.30***	.28***	.23***	.23***	.03***	-.20***	.03***	.14***	-.15***	.71***	.50***	.23***	.26***	.04***	.04***	.12***	.06***
Latitude difference	.17***	.15***	.09***	.24***	.26***	-.15***	.18***	.23***	-.11***	.30***	.24***	.00	.09***	.24***	-.06***	.45***	.26***
Longitude difference	.25***	.23***	.17***	.18***	-.01	-.14***	.01	-.07***	-.12***	.58***	.35***	.28***	.21***	-.08***	.07***	.04***	-.05***
Shared border	-.24***	-.24***	-.19***	-.15***	-.10***	.40***	-.10***	-.12***	.44***	-.20***	-.18***	-.13***	-.22***	-.12***	-.04***	-.19***	-.13***
Common official language	-.28***	-.25***	-.23***	-.09***	-.12***	.11***	-.06***	-.07***	.15***	-.11***	-.06***	-.16***	-.37***	-.03***	-.05***	-.13***	-.06***
Common legal origin	-.04***	-.04***	-.03***	-.17***	-.15***	.03***	-.17***	-.11***	.04***	.02*	-.00	-.15***	-.12***	-.08***	.05***	-.14***	-.04***
Colonial relationship	-.07***	-.05***	-.06***	.03**	.10***	.07***	.08***	.03***	.13***	-.03***	-.04***	-.04***	-.10***	.03***	-.00	.01	-.01
Common colonizer	-.08***	-.09***	-.09***	-.17***	-.19***	.04***	-.15***	-.13***	.06***	-.08***	-.05***	.00	-.02*	-.06***	-.01	-.16***	-.03***
Shared sea/ocean	-.31***	-.29***	-.14***	-.05***	-.02**	.23***	.05***	-.05***	.17***	-.08***	-.10***	-.08***	-.21***	-.04***	-.02***	-.12***	-.09***
Either landlocked	-.01	.01	-.01	-.01	-.08***	-.09***	-.12***	.12***	-.05***	-.06***	.01	-.03***	.12***	.10***	.02**	.07***	.10***
Either island	-.03***	-.04***	.03***	-.01	-.02**	-.02***	.11***	-.05***	-.03***	.09***	.04***	.06***	.03***	-.03***	-.01	.03***	.04***

Notes: Income, Human-Capital, HDI, WGI, UN-Voting, and Trade measures use 2000–2019 data; Migration Intensity uses the observed 2000–2020 five-year UN DESA vintages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Genetic-Distance Clustering and Baseline ACD

B.1 PAM clustering

Partitioning around medoids (PAM) is well suited to our setting because the starting point is a pairwise distance matrix. In contrast to k -means, PAM does not construct artificial centroids; instead, each cluster is represented by an observed country, the medoid, chosen to minimize its average distance to the other members of the cluster.

We choose the number of clusters using average silhouette width, which measures how well countries fit within their assigned cluster relative to neighboring clusters. Table A5 reports the resulting silhouette values and the ranking of the alternative partitions.

Table A5: Silhouette Ranking for PAM, $k = 2$ –10

Rank	k	Silhouette	Smallest Group	Group Sizes
1	4	0.662	25	63, 42, 26, 25
2	10	0.648	4	26, 26, 21, 21, 14, 14, 12, 10, 8, 4
3	5	0.642	12	63, 30, 26, 25, 12
4	3	0.610	26	88, 42, 26
5	9	0.593	4	26, 26, 23, 22, 15, 14, 14, 12, 4
6	8	0.580	4	37, 26, 26, 22, 15, 14, 12, 4
7	7	0.548	12	37, 30, 26, 22, 15, 14, 12
8	6	0.531	12	39, 30, 26, 25, 24, 12
9	2	0.471	42	114, 42

Notes: Average silhouette width from PAM applied to the 1500 pairwise F_{ST} distances of Spolaore and Wacziarg (2018) for the 156-country fitting sample.

The highest silhouette score is obtained at $k = 4$. This solution, however, is very coarse, separating the sample largely into the Americas, Sub-Saharan Africa, East Asia, and a broad Eurasian group. Such a partition leaves relatively little variation in the ancestry profiles that ACD is intended to capture. Among the more granular solutions, $k = 10$ performs best. Its average silhouette width is 0.648, second only to $k = 4$ and slightly above $k = 5$ at 0.642, while its smallest cluster still contains four countries. We therefore use $k = 10$ as the baseline, balancing within-cluster fit against the need to retain meaningful variation across ancestral origins.

Table A6 reports the resulting assignments. The cluster names are added only to make the groups easier to read and have no role in the estimation.

Table A6: Baseline Ancestral Clusters: PAM, $k = 10$

Cluster	Medoid	N	Countries
C1: Other SSA	Angola	27	Angola, Botswana, Burundi, Cape Verde, Central African Republic, Chad, Comoros, Congo, Rep., Democratic Republic of Congo, Equatorial Guinea, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mozambique, Namibia, Rwanda, Senegal, Sierra Leone, South Africa, Sudan, Swaziland, Tanzania, Uganda, Zambia, Zimbabwe
C2: Pre-Columbian Americas	Argentina	27	Argentina, Belize, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Trinidad and Tobago, United States, Uruguay, Venezuela, RB
C3: Balkan–Anatolian–Central Asia	Albania	22	Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Georgia, Greece, Italy, Kazakhstan, Kyrgyz Republic, Libya, Macedonia, FYR, Malta, Moldova, Romania, Slovenia, Somalia, Turkey, Turkmenistan, Uzbekistan
C4: East/Southeast Asia	China	22	Bhutan, Cambodia, China, Fiji, Hong Kong SAR, China, Indonesia, Japan, Korea, Dem. Rep., Korea, Rep., Lao PDR, Malaysia, Mongolia, Myanmar, Nepal, New Zealand, Papua New Guinea, Philippines, Singapore, Taiwan, Thailand, Timor-Leste, Vietnam
C5: Western Europe	Austria	15	Austria, Belgium, Czech Republic, Denmark, France, Germany, Ireland, Luxembourg, Netherlands, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom
C6: Arab/MENA	Algeria	15	Algeria, Bahrain, Egypt, Arab Rep., Iraq, Jordan, Kuwait, Lebanon, Morocco, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates, Yemen, Rep.
C7: West Africa	Benin	13	Benin, Burkina Faso, Cameroon, Cote d'Ivoire, Gabon, Gambia, The, Guinea, Guinea-Bissau, Liberia, Mauritania, Niger, Nigeria, Togo
C8: Northern/Eastern Europe	Belarus	10	Belarus, Estonia, Finland, Hungary, Latvia, Lithuania, Norway, Poland, Russian Federation, Ukraine

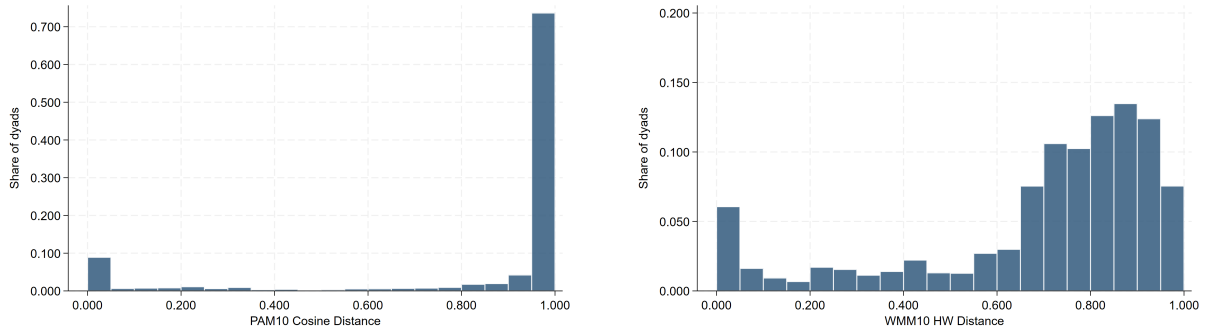
Table A6 continued

Cluster	Medoid	N	Countries
C9: Broad South Asia	Bangladesh	8	Afghanistan, Australia, Bangladesh, India, Iran, Islamic Rep., Pakistan, Sri Lanka, Tajikistan
C10 Horn of Africa/Sahel	Eritrea	4	Eritrea, Ethiopia, Mali, Mauritius

Notes: PAM10 is fitted on 156 countries using the 1500 pairwise F_{ST} distances of Spolaore and Wacziarg (2018). Bosnia and Herzegovina, Puerto Rico, Taiwan, Tanzania, Timor-Leste, Togo, and Yemen are assigned by the documented nearest-neighbour procedure. Counts include these additions, so the table covers all 163 WMM origins. Cluster labels are only descriptive.

Figures A1a and A1b show the full-sample distributions of the two principal ACD measures.

Figure A1: Histograms for Two Main ACD Measures



(a) Distribution of WMM10 HW Distance

(b) Distribution of PAM10 Cosine Distance

Notes: Both histograms use all 13,203 unordered country dyads. The vertical axis reports the share of dyads in each bin; distances are shown on their raw zero-to-one scales.

Table A7 summarizes the same bilateral 1500 F_{ST} input by PAM10 cluster pair. The pair-weighted mean is 0.009 within clusters and 0.047 between clusters, a 5.28-fold difference, as expected from a partition fitted to these distances.

Table A7: Average Genetic Distance (1500) within and between PAM10 clusters

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	.009									
C2	.086	.011								
C3	.046	.062	.006							
C4	.063	.059	.045	.018						
C5	.049	.059	.008	.042	.000					
C6	.041	.060	.011	.043	.014	.001				
C7	.023	.093	.043	.069	.047	.038	.004			
C8	.051	.056	.013	.041	.009	.018	.051	.000		
C9	.047	.051	.020	.029	.019	.021	.049	.018	.012	
C10	.023	.072	.031	.051	.034	.026	.028	.036	.033	.000

Notes: Mean 1500 F_{ST} over unique unordered country dyads. Diagonal cells are within-cluster means; lower-triangular cells are between-cluster means. The matrix uses all 163 assignments and 13,203 dyads in Table A6. Cluster-pair sizes follow from the cluster counts in that table.

B.2 Permutation Tests

We randomize the assignment of the 163 ancestral origins to ten clusters and rebuild both principal measures from each draw. For PAM10 Cosine Distance, this means reaggregating the WMM ancestry shares into the randomized clusters and comparing the resulting country profiles directly. For WMM10 HW Distance, the randomized grouping changes the country ancestry vectors and the WMM source-to-destination kernel; the optimal-transport problem is then solved again for every dyad. We use 999 size-preserving assignments and 999 free-size assignments. The reported finite-sample probability is as below, so the smallest attainable value is $1/1000 = 0.001$.

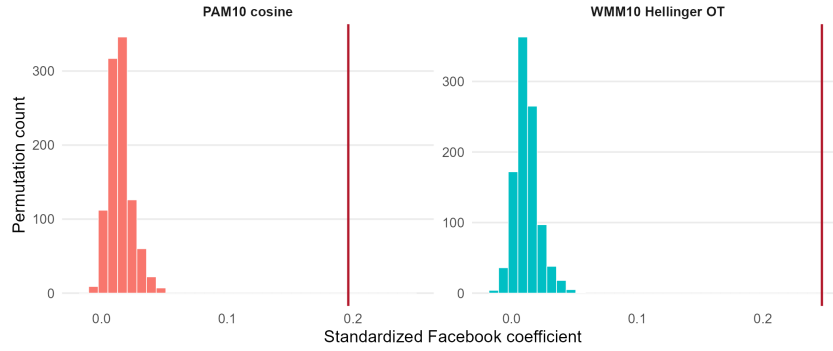
$$\hat{p} = \frac{1 + \sum_{b=1}^B \mathbf{1}\{|T_b| \geq |T_{\text{actual}}|\}}{B + 1},$$

Table A8: Permutation Results - Outcome: FB Cultural Distance

Design	Measure	Actual	Mean	P95	Maximum	p
Size-preserving	PAM10 Cosine Distance	0.196	0.015	0.033	0.048	0.001
Size-preserving	WMM10 HW Distance	0.247	0.012	0.030	0.054	0.001
Free-size	PAM10 Cosine Distance	0.196	0.016	0.034	0.057	0.001
Free-size	WMM10 HW Distance	0.247	0.012	0.029	0.049	0.001

Notes: Randomized standardized coefficients come from the baseline cultural-distance regression. Size-preserving draws retain PAM10 group sizes; free-size draws require non-empty groups. Both measures are rebuilt in every draw, and no randomized absolute coefficient reaches its actual value.

Figure A2: Size-Preserving Permutation Distributions



Notes: Histograms show 999 exact randomized coefficients for the two principal measures. Red vertical lines mark the actual coefficients.

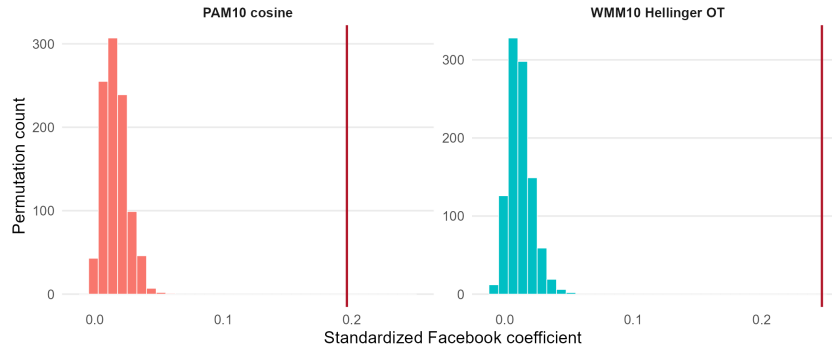


Figure A3: Free-Size Permutation Distributions

Notes: Histograms show 999 exact randomized coefficients when cluster sizes may vary. Red vertical lines mark the actual coefficients.

C Alternative ACD Constructions

Constructing ACD requires choices about the distance used to form ancestral clusters, the number of clusters, and the metric used to compare ancestry profiles. This section varies each of these choices and also reconstructs the measure using a classification using no genetic data.

C.1 Alternative PAM10 Partitions and Clustering Procedures

We vary the clustering input, algorithm, number of groups, and whether origins are clustered at all. We then hold the partition fixed and vary the distance used to compare ancestry profiles.

Alternative Clustering Inputs

The baseline clusters are formed from Genetic Distance (1500). Figure A4 instead constructs PAM-based ACD using five alternative input matrices: Genetic Distance

(1500), Geodesic Distance, Linguistic Distance, Religious Distance, and baseline EthnoDist. We consider the silhouette-ranked resolutions $k = 3, 4, 5, 9, 10$. Across all inputs and resolutions, the ancestry coefficient remains positive for both FB and WVS Cultural Distance.

Each panel reports standardized ACD coefficients from the baseline specification under PAM clustering based on genetic, geodesic, linguistic, religious, or ethnographic input distances. Panel (a) uses PAM10 Cosine Distance. Panel (b) uses WMM10 HW Distance where the input distance determines only the partition, while the similarity kernel and transport costs are constructed from WMM source-to-destination profiles.

Ward Hierarchical Clustering

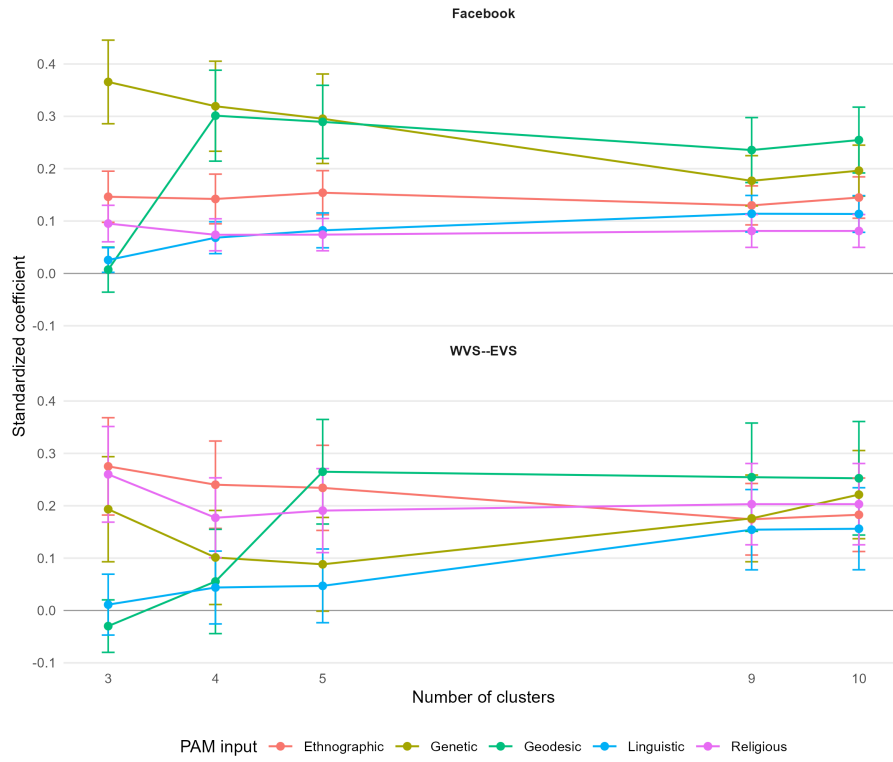
The algorithm matters as well as the input. Figure A5 replaces PAM with Ward's minimum-variance hierarchical clustering, applied to the same genetic-distance matrix and the same procedures. Both principal measures are rebuilt under each algorithm, so the figure isolates the effect of the clustering procedure alone.

Alternative WMM10 HW Distance calculations

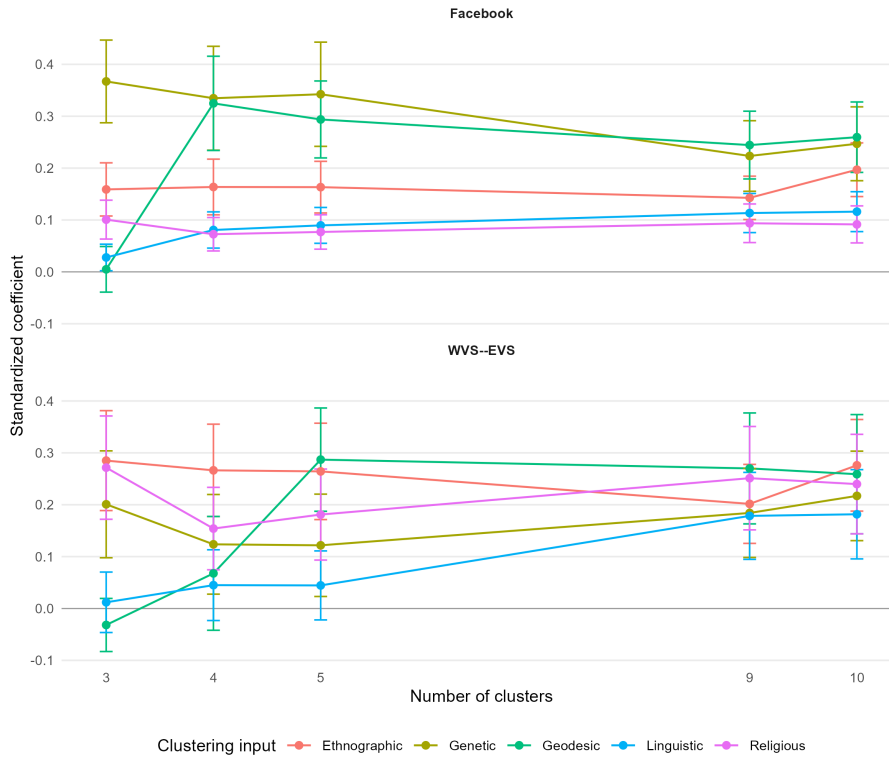
The last choice is how to compare two ancestry profiles once the partition is settled. Figure A6 holds the genetic PAM10 partition fixed and varies both the WMM-based kernel (cosine, Hellinger, and diffusion-2) and the comparison rule (soft cosine or optimal transport), for the two cultural outcomes. The baseline pairs the Hellinger kernel with optimal transport; the figure shows where that combination sits among the alternatives.

Figure A7 closes the subsection by showing pairwise Pearson correlations across available country dyads.

Figure A4: ACD Coefficients under Five Alternative Clustering Inputs, $k = 3, 4, 5, 9, 10$



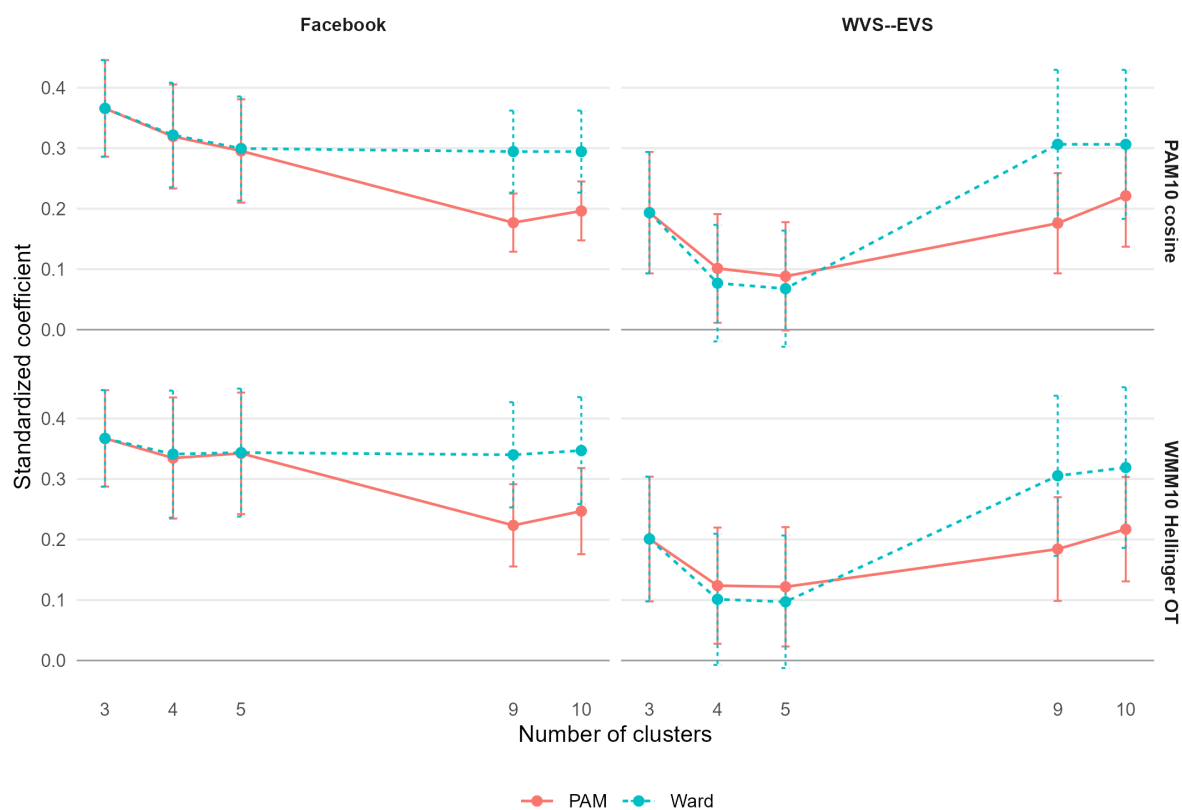
(a) PAM10 Cosine Distance



(b) WMM10 HW Distance

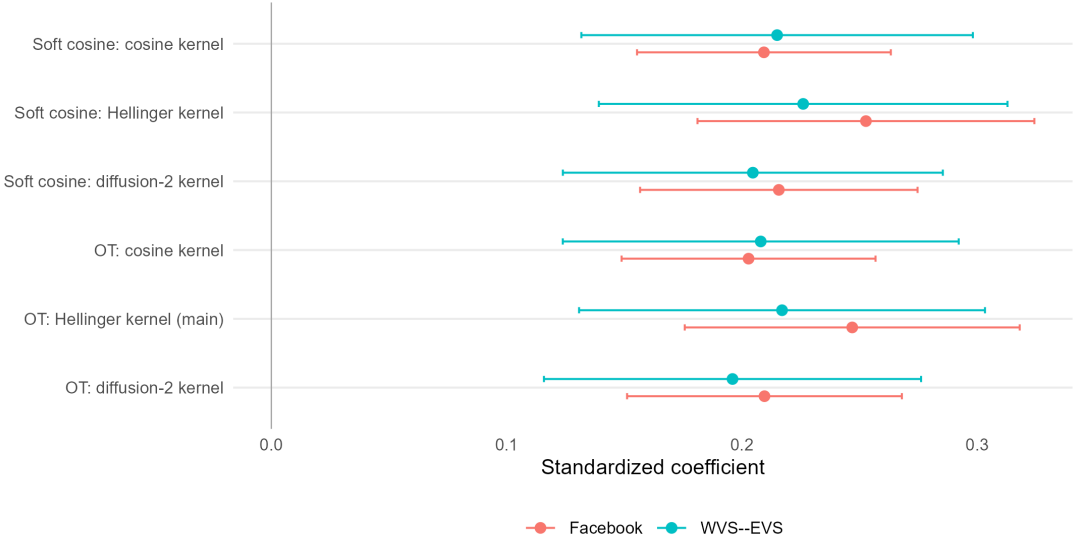
Notes: Vertical bars denote 95 percent confidence intervals based on two-way clustered standard errors.

Figure A5: PAM versus Ward Clustering



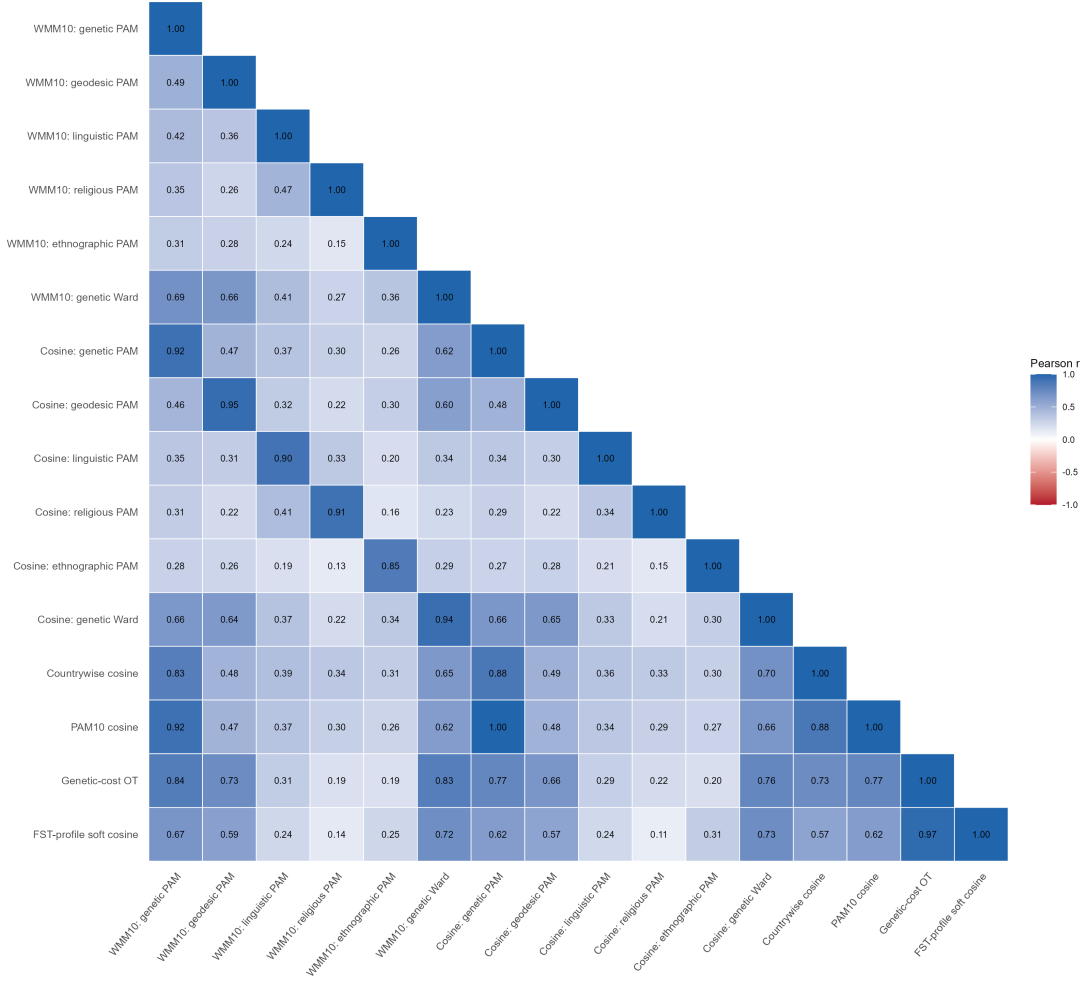
Note: Each panel compares standardized coefficients on ACD under Ward hierarchical clustering against genetic PAM for $k = 3, 4, 5, 9, 10$. The rows distinguish PAM10 Cosine Distance and WMM10 HW Distance; the columns show the two cultural outcomes. Vertical bars denote 95 percent confidence intervals.

Figure A6: Alternative WMM10 HW Distance Calculations and Cultural Distance



Notes: Points are standardized coefficients from separate baseline regressions. Bars represent 95 percent confidence intervals. The baseline WMM10 HW Distance is marked as the main specification.

Figure A7: Correlations between Main and Alternative ACD Measures



C.2 Historical-Cultural ACD

This construction answers an expected question a reader can put to the measure, which is whether the results are an artifact of the F_{ST} information used to form the baseline clusters. Here no genetic input enters at any stage. Each country's historical ancestry vector is aggregated to the nine spheres of Table A9. Ordinary cosine distance treats those spheres as mutually orthogonal. For soft cosine, we instead use

$$d_{ij}^S = 1 - \frac{x'_i S x_j}{\sqrt{(x'_i S x_i)(x'_j S x_j)}},$$

where x_i and x_j are the two nine-sphere ancestry profiles.

Historical-Cultural Classification

The nine historical-cultural spheres and their country assignments are defined below.

1. **Western Christian:** The Western Christian category is intended to capture the

historical space of Latin Christendom. Its coherence rested not only on shared adherence to the Catholic Church, but also on common ecclesiastical institutions, the authority of the papacy, and the spread of Latin legal and liturgical forms across much of medieval Europe. In this sense, the category refers to a long-run civilizational zone shaped by a shared religious-institutional framework rather than to a narrowly confessional label. Although anchored around 1500, the classification also reflects the longer historical consolidation of this Latin-Christian sphere and its enduring influence on later European political and cultural development (Bartlett, 1993; Le Goff, 2005; Moore, 2000; Rubin & Simons, 2009).

2. **Eastern Christian:** This category encompasses the Byzantine-Orthodox world as well as the older Christian traditions of the eastern Mediterranean and the Caucasus. Its coherence lies not in strict doctrinal uniformity, but in the historical continuity of Eastern Christian institutions, liturgical traditions, and ecclesiastical networks that developed outside Latin Christendom. In this sense, the category captures a broad Christian macro-zone shaped by Byzantine legacies and parallel eastern traditions rather than a narrowly defined confessional bloc (Obolensky, 1971; Parry, 2007; Shepard, 2008; Taddesse, 1972).
3. **Islamic / Arab:** This category encompasses the Arabic-speaking and Ottoman-influenced Islamic world of the eastern Mediterranean, North Africa, Arabia, and adjacent Indian Ocean spaces. Its coherence lies not in strict political uniformity, but in a shared Islamic religious tradition, the centrality of Arabic in religious and intellectual life, and long-run historical connections shaped by imperial structures, commercial exchange, and regional mobility. In this sense, the category captures a broad Islamicate macro-zone extending across the Arab heartlands and their connected maritime peripheries rather than a narrowly bounded political bloc. Although anchored around 1500, it also reflects the longer historical durability of these interconnected religious, imperial, and commercial formations (Faroghi, 2004; Hodgson, 1974; Hourani, 1991; İnalcık & Quataert, 1994; Streusand, 2011).¹
4. **Turkic / Persianate** This category encompasses Iran, Central Asia, Afghanistan, and adjacent regions historically shaped by the Persianate world. Its coherence lies in the long-run role of Persian as a language of literary production, elite culture, and political communication, as well as in shared courtly, intellectual, and imperial traditions that connected these societies across time and space. In this sense, the category captures a transregional Islamicate sphere that is analytically distinct

¹Comoros is retained in the Sub-Saharan African category for classificatory purposes despite its longer-run connections to the western Indian Ocean Muslim sphere. This choice reflects the practical difficulty of assigning a single small archipelago to the Arab world on the basis of religion alone, given that its demographic and cultural formation is also deeply shaped by Bantu African and Austronesian influences. For the historical context, see (Beaujard, 2007; Walker, 2019).

from the Ottoman/Arab world, even though both formed part of the wider Islamic ecumene (Amanat & Ashraf, 2019; Canfield, 1991; Green, 2019; Hodgson, 1974; Peacock & Tor, 2015).

5. **East Asian / Sinographic** This category encompasses China, Korea, Japan, and related cases historically shaped by the Sinographic world. Its coherence lies in the long-run influence of Literary Sinitic, Confucian political thought, and shared bureaucratic and textual institutions that connected these societies despite clear linguistic and political differences. Rather than denoting a uniform civilization, the category captures a historically recognizable regional formation in which Chinese models of governance, learning, and elite culture provided a common frame of reference across multiple polities (Denecke & Nguyen, 2017; Elman, 2013; Huang & Kang, 2022; Kang, 2010; Woodside, 2006).
6. **South–Southeast Asian** This category encompasses South and Southeast Asia as a historically connected macro-region. Its coherence lies not in strict civilizational uniformity, but in long-run interactions shaped by Sanskritic influence, Buddhist and commercial networks, and broader monsoon-world connections linking the two regions across both mainland and maritime settings. In this sense, the category captures a broad zone of cultural and historical interconnection rather than a single homogeneous tradition. It is therefore best understood as a parsimonious regional grouping anchored around 1500, while also reflecting longer-term patterns of exchange and mutual influence that extended beyond that benchmark (Lieberman, 2003, 2009; Manguin et al., 2011; Pollock, 2006; Reid, 1988, 1993; Subrahmanyam, 1997).²
7. **Pre-Columbian American** This category encompasses the territories of the broader Americas as shaped by indigenous civilizational formations prior to and at the time of European conquest. Its coherence lies not in cultural uniformity, but in the fact that these regions developed outside sustained Old World contact and were structured by distinct pre-Columbian political, social, and ecological systems. In this sense, the category captures the deep precolonial historical substrate upon which later colonial and national orders were built, rather than the present-day demographic composition of the countries concerned (Clendinnen, 1991; D’Altroy, 2015; Denevan, 1992; Hemming, 1970; Mann, 2005; Salomon & Schwartz, 1999; Trigger & Washburn, 1996).
8. **Sub-Saharan African** This category encompasses polities south of the Sahara whose long-run historical development was shaped primarily by African ecologies,

²Mauritius is included in the South–Southeast Asian category as a post-1500 adjustment, reflecting its later demographic and cultural formation through Indian Ocean and South Asian connections rather than its status at the exact benchmark date of 1500 (Allen, 1999; Vaughan, 2005).

migration patterns, agrarian and pastoral systems, regional exchange, and local processes of political formation. Its coherence lies not in strict cultural uniformity, but in a shared macro-regional historical framework that distinguished much of sub-Saharan Africa from the Mediterranean and the Islamicate polities to its north, even where extensive interaction existed. In this sense, the category is best understood as a broad historical grouping with important frontier zones³⁴ rather than as a homogeneous civilizational bloc (Ehret, 2016; Levtzion & Pouwels, 2000; Thornton, 1998; Vansina, 1990).

9. **Oceanian** This category encompasses Australia, Papua New Guinea, New Zealand, and the wider Pacific cases grouped here within a broad Oceanian macro-region. Its coherence lies not in cultural uniformity, but in the long-run historical significance of Pacific migration, seafaring mobility, and interconnected island and coastal worlds shaped by successive waves of settlement. In this sense, the category captures a pragmatic but historically grounded regional formation linking Aboriginal Australian, Papuan/Melanesian, and wider Oceanian traditions rather than a single homogeneous civilization (Bedford, 2022; Denham et al., 2003; Flood, 2019; Kirch, 2000).

Table A9: Civilizational Classification and Country Assignments

Civilizational Category	Countries
Western Christian ($N = 26$)	Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom.
Eastern Christian ($N = 13$)	Albania, Armenia, Belarus, Bosnia and Herzegovina, Bulgaria, Cyprus, Georgia, Greece, Moldova, North Macedonia, Romania, Russia, Ukraine.
Islamic / Arab ($N = 17$)	Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates, Yemen.
Turkic / Persianate ($N = 11$)	Afghanistan, Azerbaijan, Iran, Kazakhstan, Kyrgyzstan, Mongolia, Pakistan, Tajikistan, Turkey, Turkmenistan, Uzbekistan.
East Asian ($N = 6$)	China, Hong Kong, Japan, North Korea, South Korea, Taiwan.

³This applies especially to the East African coast. Kenya, Tanzania, and Mozambique are retained in the Sub-Saharan African category at the country level because their Islamic and Swahili coastal connections did not define the entirety of these territories in the benchmark period, even if coastal regions were deeply integrated into the western Indian Ocean world (Beaujard, 2007; Levtzion & Pouwels, 2000).

⁴Madagascar is retained in the Sub-Saharan African category for classificatory purposes, but it should be understood as a hybrid Indian Ocean case whose longer-run historical formation reflects both African and extra-African, especially Austronesian, influences (Beaujard, 2007).

Table A9 continued

Civilizational Category	Countries
South–Southeast Asian ($N = 16$)	Bangladesh, Bhutan, Cambodia, India, Indonesia, Laos, Malaysia, Mauritius, Myanmar, Nepal, Philippines, Singapore, Sri Lanka, Thailand, Timor-Leste, Vietnam.
Pre-Columbian American ($N = 27$)	Argentina, Belize, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Trinidad & Tobago, United States, Uruguay, Venezuela.
Sub-Saharan African ($N = 43$)	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, DR Congo, Equatorial Guinea, Eritrea, Eswatini, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, Sudan, Tanzania, Togo, Uganda, Zambia, Zimbabwe.
Oceanian ($N = 4$)	Australia, Fiji, New Zealand, Papua New Guinea.

As a direct country-pair check, we compare whether each of the 13,203 unordered dyads belongs to the same PAM10 cluster and to the same historical-cultural sphere. Among the 1,531 dyads whose countries share a PAM10 cluster, 1,164 (76.0 percent) also share a historical-cultural sphere. Thus, PAM10 co-membership is associated with a roughly 10.8-fold increase in the probability of historical-cultural co-membership. In 9,999 country-level permutations that preserve the observed PAM10 cluster sizes, the expected within-PAM10 historical-cultural co-membership rate is 15.1 percent; none of the simulated partitions reaches the observed 76.0 percent, yielding a finite-sample permutation p -value of 0.0001.

Table A10: Historical-Cultural Measures and the Two Principal ACD Measures

	(1)	(2)	(3)	(4)
(1) Historical-cultural cosine	1.000			
(2) Historical-Cultural Soft-Cosine Distance	0.982	1.000		
(3) PAM10 Cosine Distance	0.625	0.625	1.000	
(4) WMM10 HW Distance	0.660	0.671	0.922	1.000

Notes: Pairwise Pearson correlations on common dyads. Historical measures use the nine ex ante spheres; the other measures use genetic PAM10.

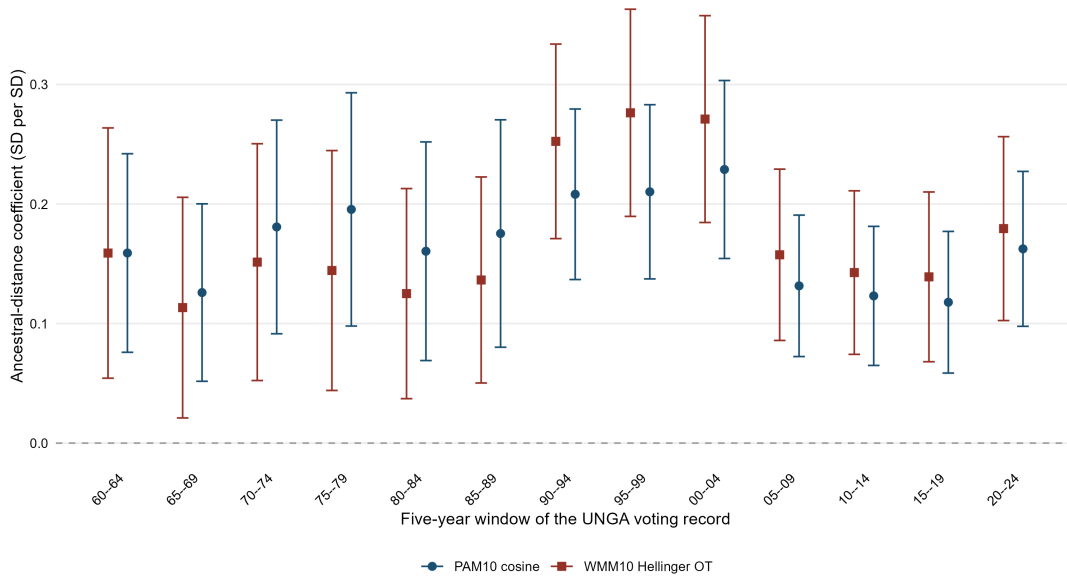
D Robustness Checks to Alternative Measurements, Time Periods, Sample

The previous section varied the construction of ACD. This one holds the measure fixed and varies everything around it—how the outcomes are measured and dated, which countries enter the sample, and which specification is estimated. The organizing question is whether the main results depend on choices made about the data rather than on the ancestry variable itself.⁵

D.1 Time-Specific Estimates: UN Voting and Cultural Distance

The UN-voting baseline averages 2000–2019, whereas the main cultural benchmark uses the 2021 WVS–EVS vintage. Figure A8 decomposes the UN General Assembly voting record into thirteen non-overlapping five-year windows, and the same picture holds, since the WMM10 HW Distance coefficient is positive in every window, while the PAM10 Cosine Distance coefficient is small before 1990 and statistically distinguishable from zero in every window thereafter.

Figure A8: ACD and UN-Voting Distance by Five-Year Intervals

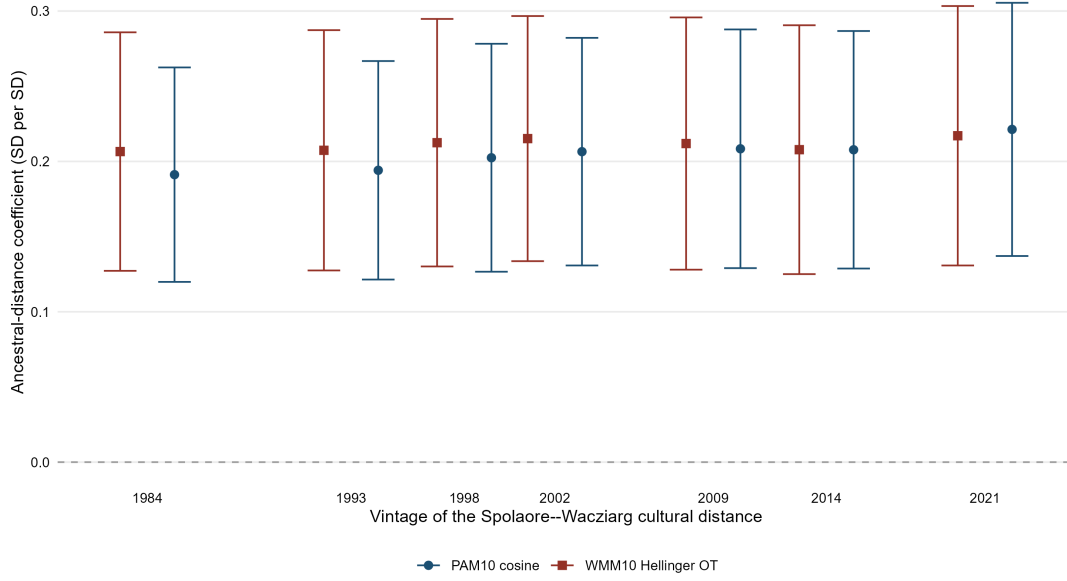


Notes: Thirteen five-year regressions; dyads require at least three common years. In 1975–1979 the clustered variance is zero, so the bars collapse to the point estimates.

Similarly, Figure A9 re-estimates the baseline across all seven. Both ACD coefficients remain positive and precisely estimated, and stable in size, so the cultural result is not tied to a particular vintage.

⁵Unless stated otherwise, figures and tables in this section standardize continuous outcomes and distance regressors, include the baseline dyadic controls and country1 and country2 fixed effects, and use standard errors two-way clustered by country. Figure bars are 95% confidence intervals.

Figure A9: ACD and Cultural Distance by Data Vintage

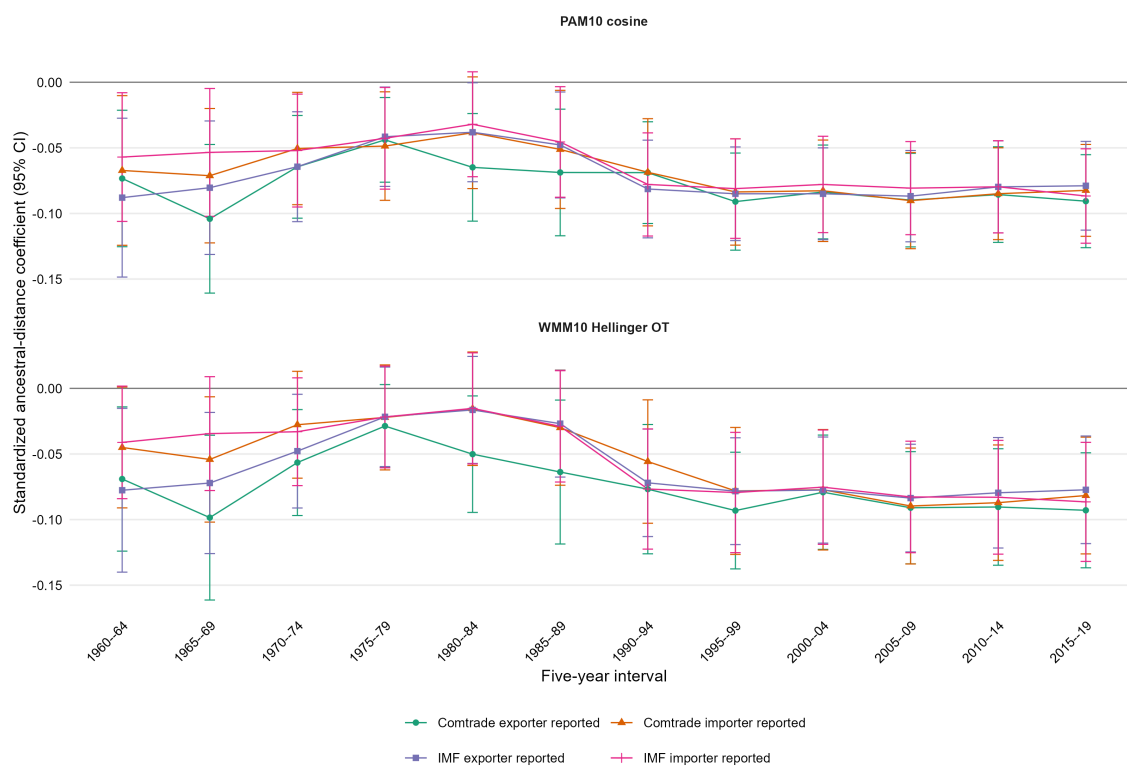


Notes: Seven WVS Cultural Distance vintages; points are the ancestry coefficients.

D.2 Period-Specific Estimates: Trade, Migration, Income

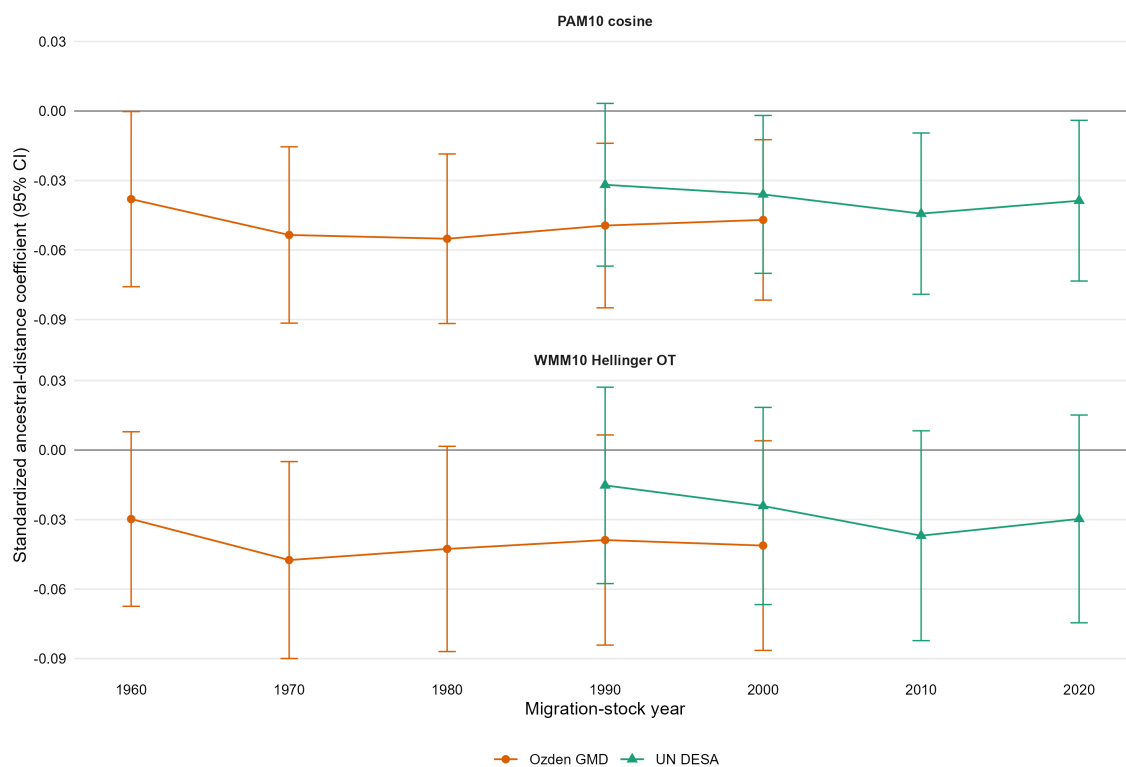
Figures A10–A12 examine the stability of the main relationships over time for WMM10 HW Distance and PAM10 Cosine Distance. Trade Intensity and the absolute Income Gap are reconstructed from annual CEPII data in twelve non-overlapping five-year intervals from 1960–1964 to 2015–2019. For trade, we use all four combinations of Comtrade and IMF data with exporter- and importer-reported flows. Migration is measured from decennial bilateral stocks in the Global Migration Database for 1960–2000 and UN DESA for 1990–2020. The results are highly stable.

Figure A10: ACD and Trade Intensity by Five-Year Intervals



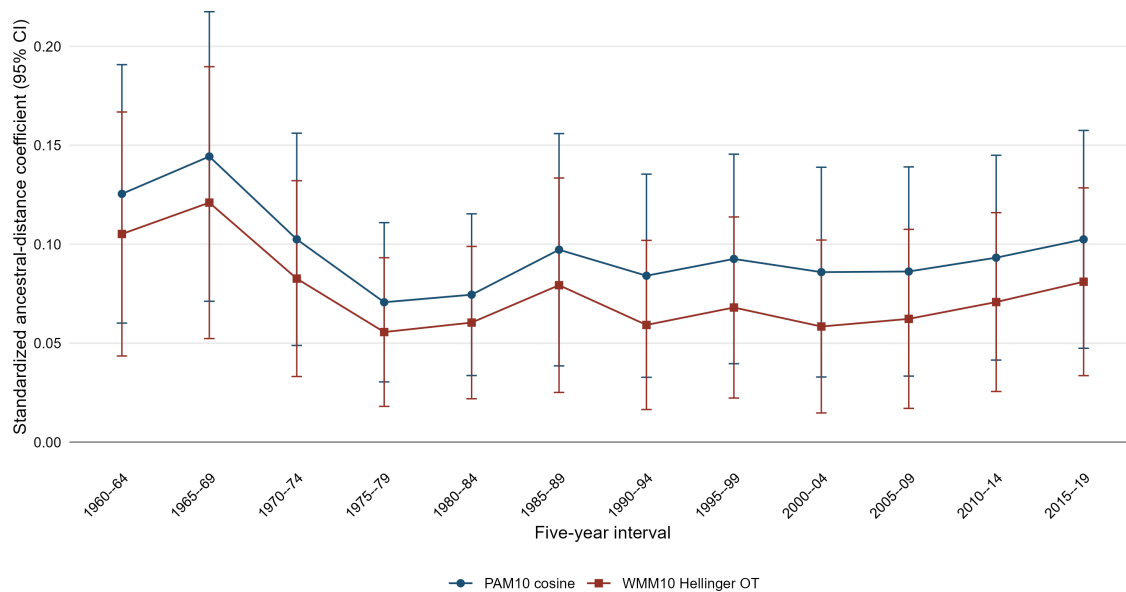
Notes: Five-year CEPII series; annual bilateral intensities are averaged within intervals. Points are the ancestry coefficients.

Figure A11: ACD and Migration Intensity (Özden GMD and UN DESA)



Notes: Özden GMD stocks: 1960–2000 by decade; UN DESA stocks: 1990–2020 by decade. No intermediate values are imputed. Each regression covers 13,199 dyads among 163 countries.

Figure A12: ACD and Income Gap by Five-Year Intervals



Notes: The outcome is Income Difference. Points are the ancestry coefficients.

D.3 Comparing with Alternative Contemporary Cultural Outcomes

Table A11: Contemporary Genetic Distance, Linguistic Distance, Religious Distance

Dep. Var. FB Cultural Distance	(1)	(2)	(3)	(4)
ACD (PAM10 Cosine)	0.196*** (0.025)	0.042** (0.019)	0.135*** (0.023)	0.179*** (0.026)
Contemporary Genetic Distance		0.394*** (0.049)		
Linguistic Distance			0.252*** (0.030)	
Religious Distance				0.120*** (0.025)
Observations	12399	11391	12399	12399
R ²	0.652	0.697	0.676	0.657
Within R ²	0.307	0.389	0.355	0.318

Notes: FB Cultural Distance is the outcome. ACD is measured using PAM10 Cosine Distance. Column (1) reports the baseline specification. Columns (2)–(4) add Contemporary Genetic Distance, Linguistic Distance, and Religious Distance separately. All models use the Section D common specification. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.4 ACD Construction with WMM 2000

The baseline in this study reconstructs ancestry for the 1960 population; the original Putterman–Weil matrix describes the population 2000. As Table A12 shows each measure correlates closely.

Table A12: Correlations between 1960- and 2000-Vintage Principal ACD Measures

	(1)	(2)	(3)	(4)
(1) WMM10 HW Distance (Baseline Vintage)	1.000			
(2) PAM10 Cosine Distance (Baseline Vintage)	0.9218	1.000		
(3) WMM10 HW Distance (Alternative Vintage)	0.9971	0.9144	1.000	
(4) PAM10 Cosine Distance (Alternative Vintage)	0.9224	0.9989	0.9172	1.000

Notes: Baseline uses the revised 1960 WMM; Alternative uses the original 2000 WMM.

E Replication of Spolaore and Wacziarg (2009) Using ACD

Spolaore and Wacziarg (2009) show that Genetic Distance predicts bilateral Income Gaps, consistent with their barriers to the diffusion of development argument. We replicate their 1995 income specification using WMM10 HW Distance and ask whether ACD captures

the same gradient and retains explanatory power alongside weighted F_{ST} .⁶ Following their design, we regress the absolute bilateral log-income gap on ACD while sequentially adding macro-geography, micro-geography, freight costs, and climate controls.

WMM10 HW Distance is positive and significant when entered alone, with a fully controlled coefficient of 0.307, but becomes insignificant once weighted F_{ST} is included jointly. Also, in the technological-frontier specification, based on $|d(i, \text{USA}) - d(j, \text{USA})|$, the results are stronger, Table A14 shows a fully controlled coefficient of 1.309, which remains positive and significant at 1.218 after adding relative weighted F_{ST} . Thus, ACD reproduces the original development-gradient result most clearly when distance is defined relative to the technological frontier; and adds additional information over Genetic Distance.

Table A13: Replication of the Direct Bilateral Income-Gap Specification

	(1)	(2)	(3)	(4)	(5)
<i>Panel A: WMM10 HW Distance</i>					
ACD	.505*** (.100)	.497*** (.095)	.468*** (.105)	.468*** (.105)	.307*** (.104)
Observations	10,585	10,585	10,585	10,585	9,453
R^2	.022	.043	.058	.058	.086
<i>Panel B: WMM10 HW Distance and Genetic Distance (Contemporary)</i>					
ACD	.241*** (.080)	.253*** (.081)	.225** (.088)	.224** (.088)	.104 (.095)
Genetic Distance (Contemporary)	2.115*** (.628)	2.401*** (.637)	2.265*** (.623)	2.270*** (.622)	1.949*** (.613)
Observations	10,585	10,585	10,585	10,585	9,453
R^2	.044	.065	.078	.078	.101
Macro-geography controls	No	Yes	Yes	Yes	Yes
Micro-geography controls	No	No	Yes	Yes	Yes
Freight-cost control	No	No	No	Yes	Yes
Climate control	No	No	No	No	Yes

Notes: Income Gap is measured in 1995; regressors use the replication scales. Columns add the control blocks shown; (1)–(4) use the common non-climate sample and (5) the climate sample. Standard errors in parentheses are two-way clustered by country. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

⁶Results are similar when PAM10 Cosine Distance is used.

Table A14: Replication of the Frontier-Relative Income-Gap Specification

	(1)	(2)	(3)	(4)	(5)
<i>Panel A: USA-relative ACD</i>					
ACD	.893*** (.226)	1.086*** (.233)	1.108*** (.234)	1.110*** (.234)	1.309*** (.239)
Observations	10,585	10,585	10,585	10,585	9,453
R^2	.039	.072	.093	.093	.150
<i>Panel B: USA-relative ancestry and genetic distance (Contemporary)</i>					
ACD	.838*** (.201)	.980*** (.209)	1.019*** (.211)	1.016*** (.211)	1.218*** (.216)
Genetic distance	5.883*** (.913)	5.671*** (.898)	5.510*** (.885)	5.527*** (.885)	5.504*** (.871)
Observations	10,585	10,585	10,585	10,585	9,453
R^2	.134	.143	.158	.158	.214
Macro-geography controls	No	Yes	Yes	Yes	Yes
Micro-geography controls	No	No	Yes	Yes	Yes
Freight-cost control	No	No	No	Yes	Yes
Climate control	No	No	No	No	Yes

Notes: Income Gap and estimation conventions follow Table A13. USA-relative distance is $|d(i, \text{USA}) - d(j, \text{USA})|$ for both WMM10 HW Distance and Genetic Distance (Contemporary). Columns add the same control blocks and use the same samples. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F Controlling for Agro-Climatic Distance

Data and Construction

AgroClimDist captures how different two countries are in the environmental conditions and agricultural endowments that shape production possibilities. The country-level database covers the 163-country project universe and combines five blocks: (i) irrigation potential and physical geography; (ii) ecological endowments, including elevation, ruggedness, agricultural suitability, tropical and desert shares, and coastal access; (iii) average yields for eleven major crops; (iv) historical land use; and (v) ERA5 temperature and precipitation conditions from the World Bank Climate Change Knowledge Portal.

The irrigation and topographic variables are drawn from the replication archive of Bentzen et al. (2017). That archive provides irrigation potential and the country-level mean and variance of elevation; its underlying irrigation measure originates in the FAO Global Agro-Ecological Zones database. Absolute latitude, terrain ruggedness, and mean distance to the coast are taken from the country-geography data assembled by Nunn and Puga (2012).

The eleven crop-yield variables are annual country means calculated from the FAOSTAT *Crops and Livestock Products* series of the Food and Agriculture Organization of the United Nations (Food and Agriculture Organization of the United Nations, 2025). We access the harmonized series through the *Land Use* data collection maintained by (Ritchie & Roser,

2019). Built-up land, grazing land, and cropland are historical land-use series from the HYDE 3.2 database documented by Klein Goldewijk et al. (2017). These data are also distributed through Our World in Data, together with the associated FAO series (Ritchie & Roser, 2019).

Finally, mean temperature, temperature variability, mean precipitation, precipitation variability, and the warmest, coldest, wettest, and driest climate indices are calculated from the ERA5 historical. We obtain the country-level ERA5 series from the World Bank’s *Climate Change Knowledge Portal*, which distributes the Copernicus Climate Change Service/ECMWF reanalysis at a common spatial resolution (World Bank, 2025). Table A15 maps each included component to the corresponding source.

Variables are first min–max scaled across countries to $[0, 1]$. If a_i denotes the resulting agro-climatic profile of country i , bilateral distance is

$$d_{ij}^A = 1 - \frac{a_i' a_j}{\|a_i\| \|a_j\|}.$$

Scaling prevents variables measured in large physical units from mechanically dominating the cosine. Missing crop yields are coded as zero under the maintained interpretation that the crop is not grown. The baseline excludes irrigation potential because that variable is unavailable for Cape Verde, Mauritius, and Timor-Leste; it therefore uses 27 variables and recovers substantially more dyads. Hong Kong and Taiwan lack the CCKP block, leaving 12,880 observed dyads in the baseline variant.

Table A15: Components and Sources of AgroClimDist

Component	Source	Component	Source
Irrigation Potential	FAO GAEZ/BKW	Absolute latitude	Nunn–Puga
Terrain Ruggedness	Nunn–Puga	Distance to coast	Nunn–Puga
Mean Elevation	NCEI/NOAA via BKW	Elevation variance	NCEI/NOAA via BKW
Wheat Yield	FAOSTAT/OWID	Rice yield	FAOSTAT/OWID
Banana Yield	FAOSTAT/OWID	Maize yield	FAOSTAT/OWID
Soybean Yield	FAOSTAT/OWID	Potato yield	FAOSTAT/OWID
Bean Yield	FAOSTAT/OWID	Pea yield	FAOSTAT/OWID
Cassava Yield	FAOSTAT/OWID	Cocoa-bean yield	FAOSTAT/OWID
Barley Yield	FAOSTAT/OWID	Built-up land share	HYDE/FAO via OWID
Grazing-Land Share	HYDE/FAO via OWID	Cropland share	HYDE/FAO via OWID
Mean Temperature	World Bank CCKP/ERA5	Temperature variability	World Bank CCKP/ERA5
Mean Precipitation	World Bank CCKP/ERA5	Precipitation variability	World Bank CCKP/ERA5
Warmest Condition	World Bank CCKP/ERA5	Coldest condition	World Bank CCKP/ERA5
Wettest Condition	World Bank CCKP/ERA5	Driest condition	World Bank CCKP/ERA5

Notes: BKW denotes the Bentzen–Kaarsen–Wingender replication data; OWID denotes Our World in Data.

Table A16: AgroClimDist variants

Variant	Variables	Excluded block	Dyads
AgroClimDist - baseline (no irrigation)	27	Irrigation potential	12,880
AgroClimDist - Full	28	None	12,403
AgroClimDist - No Land Use	25	Historical land use	12,403
AgroClimDist - No Crops / Land Use	14	Crop yields and land use	12,403

Notes: Counts refer to non-missing distance values in the 13,203-dyad master universe. The estimation samples below are slightly smaller when outcome or control variables are unavailable.

Correlations and Alternative Specifications

The four variants are strongly correlated, as shown in Table A17. Their correlations with both principal ACD measures are much smaller, indicating that AgroClimDist is not simply another representation of the ancestry measures.

Table A17: Correlations among WMM10 HW Distance, PAM10 Cosine Distance, and Agro-Climatic Distance (AgroClimDist) variants

	(1)	(2)	(3)	(4)	(5)	(6)
(1) WMM10 HW Distance	1.000					
(2) PAM10 Cosine Distance	0.922***	1.000				
(3) AgroClimDist - baseline	0.287***	0.246***	1.000			
(4) AgroClimDist - full	0.309***	0.263***	0.989***	1.000		
(5) AgroClimDist - no land use	0.300***	0.259***	0.978***	0.988***	1.000	
(6) AgroClimDist - no crops/land use	0.318***	0.248***	0.874***	0.896***	0.900***	1.000

Notes: Pairwise Pearson correlations using all available dyads. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A18 enters each standardized agro-climatic variant alongside WMM10 HW Distance and PAM10 Cosine Distance in separate regressions. AgroClimDist remains positive under both constructions, while each ACD measure retains distinct predictive content.

Table A18: Agro-Climatic Distance (AgroClimDist) across alternative constructions

	FB Cultural Distance		WVS Cultural Distance	
	WMM10 HW Distance	PAM10 Cosine Distance	WMM10 HW Distance	PAM10 Cosine Distance
Agro-Climatic Distance (AgroClimDist) baseline	0.164*** (0.028)	0.176*** (0.028)	0.305*** (0.059)	0.305*** (0.057)
WMM10 HW Distance / PAM10 Cosine Distance	0.224*** (0.035)	0.179*** (0.023)	0.188*** (0.040)	0.201*** (0.039)
Agro-Climatic Distance (AgroClimDist): full	0.162*** (0.027)	0.174*** (0.028)	0.309*** (0.059)	0.310*** (0.057)
WMM10 HW Distance / PAM10 Cosine Distance	0.220*** (0.034)	0.176*** (0.023)	0.185*** (0.039)	0.198*** (0.039)
Agro-Climatic Distance (AgroClimDist): no land use	0.161*** (0.028)	0.173*** (0.029)	0.304*** (0.057)	0.305*** (0.056)
WMM10 HW Distance / PAM10 Cosine Distance	0.220*** (0.034)	0.176*** (0.023)	0.184*** (0.039)	0.198*** (0.039)
Agro-Climatic Distance (AgroClimDist): no crops/land use	0.097*** (0.021)	0.113*** (0.022)	0.221*** (0.054)	0.229*** (0.054)
WMM10 HW Distance / PAM10 Cosine Distance	0.234*** (0.036)	0.189*** (0.025)	0.196*** (0.043)	0.211*** (0.043)

Notes: Each outcome-measure column is a separate regression that includes the indicated agro-climatic-distance variant, either WMM10 HW Distance or PAM10 Cosine Distance, the baseline controls, and country1 and country2 fixed effects. Outcomes and distance regressors are standardized. Standard errors are two-way clustered by country. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

G Controlling for Ethnographic Distance

Data and Construction

Ethnographic Distance (*EthnoDist*) captures bilateral differences in inherited forms of social organization using the *Ancestral Characteristics of Modern Populations* database of Giuliano and Nunn (2018). The measure is based on ten established variables from the Ethnographic Atlas tradition: ancestral latitude; family organization (v9); preferred cousin marriage (v25); agricultural intensity (v28); settlement complexity (v30); jurisdictional hierarchy (v33); gendered agricultural work (v54); absence of class stratification (v66); absence of slavery (v70); and consensual or electoral local succession (v72). These variables are widely used in the empirical literature and are not constructed specifically for this study.

Giuliano and Nunn (2018) build the underlying country-level measures from George P. Murdock’s digitized *Ethnographic Atlas*, supplemented with the Eastern Europe, Siberia, and World Ethnographic Sample files. Ethnicity-level characteristics are mapped to present-day populations using information on their ancestral composition and aggregated into population-weighted country characteristics.

Each feature is min–max scaled to $[0, 1]$, and bilateral distance is one minus the cosine similarity between the two country profiles, using the same formula as for the agro-climatic measure. The baseline drops preferred cousin marriage (v25), the lowest-coverage characteristic, leaving nine dimensions and 12,246 observed distances in the 13,203-dyad master universe. Four cumulative alternatives progressively drop further characteristics,

trading dimensions for dyad coverage; they are listed in Table A19.

Table A19: EthnoDist variants

EthnoDist variant	Characteristics excluded	Observed dyads
EthnoDist baseline	v25	12,246
All characteristics	None	11,175
Drop v25, v72	Cousin marriage; succession	12,561
Drop v25, v72, v70	Above; slavery	12,561
Drop v25, v72, v70, v54	Above; gendered agricultural work	12,880

Correlations and Alternative Specifications

The ethnographic variants are highly correlated with one another but only weakly related to either principal ACD measure. Table A20 reports both the WMM10 HW Distance and PAM10 Cosine Distance correlations.

Table A20: Correlations among WMM10 HW Distance, PAM10 Cosine Distance, Ethnographic Distance (EthnoDist) variants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) WMM10 HW Distance	1.000						
(2) PAM10 Cosine Distance	0.922***	1.000					
(3) Ethnographic Distance (EthnoDist) baseline	0.153***	0.160***	1.000				
(4) EthnoDist - all ten	0.168***	0.166***	0.915***	1.000			
(5) EthnoDist - drop v25,72	0.153***	0.158***	0.957***	0.884***	1.000		
(6) EthnoDist - drop v25,72,70	0.102***	0.105***	0.877***	0.809***	0.916***	1.000	
(7) EthnoDist - drop v25,72,70,54	0.065***	0.068***	0.854***	0.780***	0.890***	0.973***	1.000

Notes: Pairwise Pearson correlations using all available dyads. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A21 enters each standardized ethnographic variant alongside WMM10 HW Distance and PAM10 Cosine Distance in separate regressions. The ethnographic and ancestry coefficients remain positive across the reported constructions.

Table A21: Ethnographic Distance (EthnoDist) across Alternative ACD Constructions

	FB Cultural Distance		WVS Cultural Distance	
	WMM10 HW Distance	PAM10 Cosine Distance	WMM10 HW Distance	PAM10 Cosine Distance
EthnoDist - baseline	0.149*** (0.026)	0.153*** (0.026)	0.352*** (0.053)	0.348*** (0.052)
WMM10 HW Distance / PAM10 Cosine Distance	0.233*** (0.036)	0.183*** (0.024)	0.183*** (0.039)	0.192*** (0.038)
EthnoDist - all ten	0.154*** (0.028)	0.160*** (0.028)	0.386*** (0.056)	0.384*** (0.055)
WMM10 HW Distance / PAM10 Cosine Distance	0.232*** (0.037)	0.178*** (0.024)	0.168*** (0.038)	0.178*** (0.038)
EthnoDist - drop v25,72	0.144*** (0.025)	0.149*** (0.025)	0.346*** (0.056)	0.342*** (0.055)
WMM10 HW Distance / PAM10 Cosine Distance	0.235*** (0.036)	0.184*** (0.024)	0.180*** (0.038)	0.189*** (0.037)
EthnoDist - drop v25,72,70	0.094*** (0.025)	0.097*** (0.026)	0.294*** (0.056)	0.291*** (0.055)
WMM10 HW Distance / PAM10 Cosine Distance	0.246*** (0.036)	0.194*** (0.025)	0.204*** (0.041)	0.211*** (0.040)
EthnoDist - drop v25,72,70,54	0.066** (0.026)	0.068** (0.026)	0.297*** (0.055)	0.295*** (0.054)
WMM10 HW Distance / PAM10 Cosine Distance	0.249*** (0.037)	0.196*** (0.025)	0.206*** (0.040)	0.211*** (0.039)

Notes: Each outcome-measure column is a separate regression that includes the indicated ethnographic-distance variant, either WMM10 HW Distance or PAM10 Cosine Distance, the baseline controls, and country1 and country2 fixed effects. Outcomes and distance regressors are standardized. Standard errors are two-way clustered by country. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

H Additional Checks for the Potential "European Driven" Effect on Results

A natural concern is that a measure rooted in deep population history may largely reflect distance from Europe. We address this concern with two increasingly restrictive sample exclusions. First, we remove all dyads involving any of the 15 countries in C5, our Western Europe cluster. Second, we remove all dyads involving a geographically European country.⁷

The pattern is fairly clear. Dropping the C5 countries leaves the main results largely unchanged, 14 of the 16 coefficients remain statistically significant at the 5 percent level, and all retain their expected signs. Removing all European countries is a considerably stronger test, but most of the main relationships survive. Under both ACD measures, ancestry distance remains significantly associated with FB and WVS cultural distance, income gaps, migration intensity, UN voting distance, and HDI distance.

There are two notable exceptions. The trade coefficient falls essentially to zero once European dyads are removed, suggesting that there is a concentrated trade in pairs involving Europe. The WGI coefficient also becomes statistically insignificant, although it

⁷We define the latter group using the World Bank's Europe and Central Asia classification, retaining only countries that are geographically European. This leaves 35 European countries in our sample. Russia is included, while Armenia, Azerbaijan, Cyprus, Georgia, Kazakhstan, the Kyrgyz Republic, Tajikistan, Turkmenistan, Türkiye, and Uzbekistan are excluded from the geographic definition of Europe.

is already small in the unrestricted sample. Overall, the results do not support the view that ACD is simply a European gradient. Its association with cultural and socioeconomic differences is also present among non-European countries, while the trade result is less general and should be interpreted accordingly.

On top of that exercise, we augment each controlled specification with the absolute difference in countries' European ancestry shares. For country i , the European ancestry share is the sum of its 1960 WMM ancestry shares from Chanda et al. (2014) originating from the same fixed list of 35 geographically European countries used in the European-gradient exercise (including Russia and excluding Central Asian countries):

$$ES_i = \sum_{o \in \mathcal{E}} w_{io}^{1960}, \quad ESgap_{ij} = |ES_i - ES_j|.$$

The baseline and augmented columns use identical samples, controls, fixed effects, standardization, and inference procedures.

Table A23 shows that adding the European-share gap reduces the ACD coefficient for both measures of cultural distance and for the income gap. Nevertheless, ACD remains positive and statistically significant in all six comparisons. The European-share gap is itself positive and precisely estimated for cultural distance and income, indicating that this broad ancestry gradient contains relevant information. It does not, however, fully account for the association with ACD.

Table A23: ACD with the European-Ancestry-Share Gap

	FB Cultural Distance		WVS Cultural Distance		Income Gap		Trade Intensity		Migration Intensity	
	Baseline ACD	ACD + ES gap	Baseline ACD	ACD + ES gap	Baseline ACD	ACD + ES gap	Baseline ACD	ACD + ES gap	Baseline ACD	ACD + ES gap
<i>Panel A: PAM10 Cosine Distance</i>										
ACD	0.196*** (0.025)	0.144*** (0.021)	0.221*** (0.043)	0.142*** (0.029)	0.104*** (0.024)	0.058*** (0.018)	-0.114*** (0.026)	-0.104*** (0.029)	-0.130** (0.062)	-0.139** (0.055)
European-share gap	—	0.213*** (0.032)	—	0.306*** (0.046)	—	0.195*** (0.034)	—	-0.050 (0.040)	—	0.042 (0.094)
Existing controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,399	12,399	5,667	5,667	12,876	12,876	13,037	13,037	13,155	13,155
<i>Panel B: WMM10 HW Distance</i>										
ACD	0.247*** (0.036)	0.201*** (0.033)	0.217*** (0.044)	0.146*** (0.031)	0.107*** (0.028)	0.066*** (0.022)	-0.141*** (0.032)	-0.131*** (0.035)	-0.184*** (0.070)	-0.199*** (0.066)
European-share gap	—	0.213*** (0.031)	—	0.312*** (0.047)	—	0.198*** (0.034)	—	-0.049 (0.040)	—	0.055 (0.097)
Existing controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,399	12,399	5,667	5,667	12,876	12,876	13,037	13,037	13,155	13,155

Notes: For each outcome, the *Baseline ACD* column excludes the European-share gap, whereas the *ACD + ES gap* column includes both jointly. All specifications retain the same outcome-specific controls and country1 and country2 fixed effects. FB Cultural Distance, WVS Cultural Distance, and Income Gap are standardized outcomes estimated by OLS. Trade and migration are estimated by PPML in levels, so coefficients are semi-elasticities. ACD, the European-share gap, and continuous controls are standardized within each outcome-specific sample. Standard errors are two-way clustered by country. PAM10 and WMM10 ACD enter separate regressions. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A22: Robustness to Excluding European Dyads

Outcome	Baseline	Excluding C5 dyads	Excluding European Dyads
<i>Panel A: WMM10 HW Distance</i>			
	0.247*** (0.036)	0.246*** (0.039)	0.229*** (0.037)
FB Cultural Distance	0.217*** (0.044)	0.159*** (0.034)	0.137*** (0.035)
WVS Cultural Distance	0.109*** (0.028)	0.091*** (0.027)	0.072*** (0.028)
Income Gap	-0.140*** (0.032)	-0.120*** (0.044)	0.010 (0.064)
Trade Intensity	-0.184*** (0.070)	-0.296*** (0.068)	-0.239*** (0.073)
Migration Intensity	0.167*** (0.034)	0.185*** (0.039)	0.050** (0.023)
UN Voting Distance	0.289*** (0.047)	0.312*** (0.048)	0.281*** (0.054)
HDI Distance	0.045*** (0.015)	0.016 (0.013)	0.023 (0.015)
WGI Distance			
<i>Panel B: PAM10 Cosine Distance</i>			
	0.196*** (0.025)	0.181*** (0.026)	0.182*** (0.025)
FB Cultural Distance	0.221*** (0.043)	0.119*** (0.027)	0.119*** (0.029)
WVS Cultural Distance	0.110*** (0.025)	0.061*** (0.019)	0.059*** (0.020)
Income Gap	-0.114*** (0.026)	-0.088** (0.038)	0.006 (0.052)
Trade Intensity	-0.130** (0.062)	-0.209*** (0.062)	-0.184*** (0.066)
Migration Intensity	0.148*** (0.030)	0.129*** (0.027)	0.050*** (0.019)
UN Voting Distance	0.230*** (0.036)	0.198*** (0.033)	0.197*** (0.037)
HDI Distance	0.053*** (0.016)	0.015 (0.011)	0.019 (0.013)
WGI Distance			
<i>Panel C: Observations</i>			
FB Cultural Distance	12,399	10,149	7,499
WVS Cultural Distance	5,667	4,182	2,552
Income Gap	13,037	10,727	7,997
Trade Intensity	13,037	10,727	7,997
Migration Intensity	13,155	10,832	8,089
UN Voting Distance	12,718	10,438	7,748
HDI Distance	12,558	10,293	7,623
WGI Distance	13,199	10,874	8,124

Notes: Each cell in Panels A and B reports the standardized ACD coefficient from a separate regression, with two-way country-clustered standard errors in parentheses. All specifications include the baseline dyadic controls and country-1 and country-2 fixed effects. Trade and migration are estimated by PPML in levels and additionally include the corresponding policy controls; all other outcomes are estimated by OLS. Income Gap, Trade Intensity, UN-Voting Distance, HDI Distance, and WGI Distance use the 2000–2019 baseline window. Migration Intensity averages the observed 2000, 2005, 2010, 2015, and 2020 UN DESA stock vintages without interpolation. Continuous variables are standardized within each estimation sample. C5 contains the 15 countries in the Western Europe cluster. “European Dyads” are dyads in which at least one country belongs to the geographically European subset of the World Bank’s Europe and Central Asia region, as defined in the text. Panel C reports the number of observations used for each outcome; sample sizes are identical across Panels A and B. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I Robustness Check: IV Estimations Using Genetic Distance (1500) as Instrument

This section provides an instrumental-variable robustness check using predetermined deep-history variation. We treat ACD as endogenous and use Genetic Distance (1500) as the excluded instrument. Within each outcome, OLS and 2SLS are estimated on the same sample and retain the full set of baseline dyadic controls, country fixed effects, and two-way clustered standard errors. All continuous variables are standardized within the corresponding estimation sample.

The first stage is strong for FB Cultural Distance and the income gap. For FB, the first-stage Wald statistics are 40.06 for WMM10 HW Distance and 35.54 for PAM10 Cosine Distance. The corresponding 2SLS coefficients are 0.764 and 0.849, and the Anderson–Rubin tests reject a zero coefficient at the one-percent level. For the income gap, the first-stage statistics are 41.56 and 36.04, with 2SLS estimates of 0.384 and 0.429; the Anderson–Rubin tests again yield $p < 0.001$.

The evidence is weaker for WVS Cultural Distance. Here the first-stage statistics fall to 13.38 for WMM10 and 8.38 for PAM10, while the 2SLS coefficients are 0.166 and 0.211. In both cases, the Anderson–Rubin p -value is 0.267. We therefore do not obtain comparable IV evidence for the WVS outcome. The discussion focuses on the two principal ACD measures used throughout the paper.

The reported first-stage statistic is the two-way-clustered Wald F produced by the estimator rather than a Kleibergen–Paap rk statistic. Because the models are exactly identified, no overidentification test is available. In the smaller WVS sample, the estimator also applies a positive-semidefinite correction to some clustered covariance matrices.

These results should be interpreted as a robustness exercise rather than as causal identification. Genetic Distance (1500) may affect cultural and development outcomes through channels other than ACD, and it also contributes to the information used to construct the PAM partition. The exclusion restriction is therefore demanding and cannot be validated by a strong first stage alone. The IV results nevertheless show that the positive associations with FB Cultural Distance and the income gap remain under this alternative source of historical variation, whereas the evidence for WVS Cultural Distance is considerably weaker.

Table A24: IV Estimation Results

	Instrument: Genetic Distance (1500)		
	FB Cultural Distance	WVS Cultural Distance	Income Gap
<i>Panel A: WMM10 HW Distance</i>			
OLS			
ACD	0.247*** (0.037)	0.213*** (0.044)	0.108*** (0.028)
2SLS			
ACD	0.764*** (0.103)	0.166 (0.142)	0.384*** (0.084)
First-stage Wald F	40.06	13.38	41.56
Anderson–Rubin p	< 0.001	0.267	< 0.001
Observations	11,391	5,204	12,003
<i>Panel B: PAM10 Cosine Distance</i>			
OLS			
ACD	0.194*** (0.025)	0.220*** (0.044)	0.106*** (0.025)
2SLS			
ACD	0.849*** (0.123)	0.211 (0.180)	0.429*** (0.096)
First-stage Wald F	35.54	8.38 ^w	36.04
Anderson–Rubin p	< 0.001	0.267	< 0.001
Observations	11,391	5,204	12,003

Notes: Each column treats ACD as endogenous and uses Genetic Distance (1500) as the only instrument. FB Cultural Distance is the Facebook-based cultural-distance outcome constructed from Facebook interest profiles; WVS Cultural Distance is the source-provided standardized 2021 WVS–EVS measure; Income Gap is the log absolute GDP-per-capita gap averaged over 2000–2019. Each panel reports OLS and 2SLS on the identical outcome-specific sample. All models include country1 and country2 fixed effects and the full baseline dyadic controls. Outcomes, ancestral distance, the instrument, and continuous controls are standardized within each complete-case sample; binary controls remain indicators. Standard errors and Wald tests are two-way clustered by country1 and country2. The F row is the cluster-robust first-stage Wald statistic; ^w marks $F < 10$. Anderson–Rubin p is weak-instrument robust. The models are exactly identified, so no Hansen J statistic is available. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

J Controlling For Climatic Differences

This check re-estimates the principal specifications after removing absolute latitude and longitude differences from the baseline controls and replacing them with absolute differences in average temperature and precipitation. Bilateral Geodesic Distance and all remaining geographic, historical, institutional, and policy controls are retained. Continuous variables are standardized within each estimation sample, both country fixed effects remain included, and standard errors continue to be clustered two-way by country. The exercise tests whether ACD is proxying for broad climatic differences rather than geographic position. Table A25 shows that both ACD coefficients retain their expected positive signs and remain significant at the one-percent level in every outcome and specification. The estimates are also shown with location and climate controls included jointly.

Table A25: ACD after substituting climate differences for latitude and longitude differences

	FB Cultural Distance	WVS Cultural Distance
<i>WMM10 HW Distance</i>		
Baseline location controls	0.247*** (0.036)	0.217*** (0.044)
Climate controls replace location	0.247*** (0.038)	0.190*** (0.042)
Location and climate controls jointly	0.230*** (0.035)	0.188*** (0.041)
<i>PAM10 Cosine Distance</i>		
Baseline location controls	0.196*** (0.025)	0.221*** (0.043)
Climate controls replace location	0.192*** (0.025)	0.199*** (0.041)
Location and climate controls jointly	0.183*** (0.024)	0.200*** (0.041)

Notes: Cells report standardized ACD coefficients with two-way country-clustered standard errors in parentheses. Each specification retains bilateral distance, border, language, legal-origin, colonial, sea/ocean, landlocked, and island controls plus country1 and country2 fixed effects. The baseline and alternative specifications are estimated on each specification's available sample; the climate variables cover 13,041 of the 13,203 master dyads. Stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix References

- Allen, R. B. (1999). *Slaves, Freedmen, and Indentured Laborers in Colonial Mauritius*. Cambridge University Press.
- Amanat, A., & Ashraf, A. (Eds.). (2019). *The Persianate World: Rethinking a Shared Sphere*. Brill.
- Bartlett, R. (1993). *The Making of Europe: Conquest, Colonization, and Cultural Change, 950–1350*. Princeton University Press.
- Beaujard, P. (2007). East Africa, the Comoros Islands and Madagascar before the Sixteenth Century: On a Neglected Part of the World System. *Azania: Journal of the British Institute in Eastern Africa*, 42(1), 15–35. <https://doi.org/10.1080/00672700709480448>
- Bedford, S. (2022). Austronesian Colonization of the Pacific Islands, 1200 BCE–1250 CE. In R. T. Jones & M. K. Matsuda (Eds.), *The Cambridge History of the Pacific Ocean* (pp. 434–456, Vol. 1). Cambridge University Press.
- Bentzen, J. S., Kaarsen, N., & Wingender, A. M. (2017). Irrigation and Autocracy. *Journal of the European Economic Association*, 15(1), 1–53.
- Canfield, R. L. (Ed.). (1991). *Turko-Persia in Historical Perspective*. Cambridge University Press.
- Chanda, A., Cook, C. J., & Putterman, L. (2014). Persistence of Fortune: Accounting for Population Movements, There Was No Post-Columbian Reversal. *American Economic Journal: Macroeconomics*, 6(3), 1–28.
- Clendinnen, I. (1991). *Aztecs: An Interpretation*. Cambridge University Press.
- D’Altroy, T. N. (2015). The Americas: The Age of Indigenous Empires. In J. H. Bentley, S. Subrahmanyam, & M. E. Wiesner-Hanks (Eds.), *The Cambridge World History* (pp. 245–273, Vol. 6). Cambridge University Press.
- Denecke, W., & Nguyen, N. (2017). Shared Literary Heritage in the East Asian Sinographic Sphere. In W. Denecke, W.-Y. Li, & X. Tian (Eds.), *The Oxford Handbook of Classical Chinese Literature* (pp. 510–532). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199356591.013.33>
- Denevan, W. M. (1992). The Pristine Myth: The Landscape of the Americas in 1492. *Annals of the Association of American Geographers*, 82(3), 369–385.
- Denham, T. P., Haberle, S. G., Lentfer, C., Fullagar, R., Field, J., Therin, M., Porch, N. A., & Winsborough, B. (2003). Origins of Agriculture at Kuk Swamp in the Highlands of New Guinea. *Science*, 301(5630), 189–193. <https://doi.org/10.1126/science.1085255>
- Ehret, C. (2016). *The Civilizations of Africa: A History to 1800* (2nd ed.). University of Virginia Press.
- Elman, B. A. (2013). *Civil Examinations and Meritocracy in Late Imperial China*. Harvard University Press.
- Faroghi, S. (2004). *The Ottoman Empire and the World Around It*. I. B. Tauris.

- Flood, J. (2019). *The Original Australians: The Story of the Aboriginal People* (2nd ed.). Allen & Unwin.
- Food and Agriculture Organization of the United Nations. (2025). *FAOSTAT: Crops and Livestock Products*. Retrieved July 28, 2026, from <https://www.fao.org/faostat/>
- Giuliano, P., & Nunn, N. (2018). Ancestral Characteristics of Modern Populations. *Economic History of Developing Regions*, 33(1), 1–17.
- Green, N. (Ed.). (2019). *The Persianate World: The Frontiers of a Eurasian Lingua Franca*. University of California Press.
- Hemming, J. (1970). *The Conquest of the Incas*. Harcourt.
- Hodgson, M. G. S. (1974). *The Venture of Islam: Conscience and History in a World Civilization*. University of Chicago Press.
- Hourani, A. (1991). *A History of the Arab Peoples*. Harvard University Press.
- Huang, C.-H., & Kang, D. C. (2022). *State Formation through Emulation: The East Asian Model*. Cambridge University Press.
- İnalçık, H., & Quataert, D. (Eds.). (1994). *An Economic and Social History of the Ottoman Empire, 1300–1914*. Cambridge University Press.
- Kang, D. C. (2010). *East Asia Before the West: Five Centuries of Trade and Tribute*. Columbia University Press.
- Kirch, P. V. (2000). *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact*. University of California Press.
- Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic Land Use Estimates for the Holocene: HYDE 3.2. *Earth System Science Data*, 9, 927–953. <https://doi.org/10.5194/essd-9-927-2017>
- Le Goff, J. (2005). *The Birth of Europe*. Blackwell.
- Levtzion, N., & Pouwels, R. L. (Eds.). (2000). *The History of Islam in Africa*. Ohio University Press.
- Lieberman, V. (2003). *Strange Parallels: Southeast Asia in Global Context, c. 800–1830* (Vol. 1). Cambridge University Press.
- Lieberman, V. (2009). *Strange Parallels: Southeast Asia in Global Context, c. 800–1830* (Vol. 2). Cambridge University Press.
- Manguin, P.-Y., Mani, A., & Wade, G. (Eds.). (2011). *Early Interactions between South and Southeast Asia: Reflections on Cross-Cultural Exchange*. ISEAS Publishing.
- Mann, C. C. (2005). *1491: New Revelations of the Americas Before Columbus*. Knopf.
- Moore, R. I. (2000). *The First European Revolution, c. 970–1215*. Blackwell.
- Nunn, N., & Puga, D. (2012). Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics*, 94(1), 20–36.
- Obolensky, D. (1971). *The Byzantine Commonwealth: Eastern Europe, 500–1453*. Weidenfeld & Nicolson.
- Parry, K. (Ed.). (2007). *The Blackwell Companion to Eastern Christianity*. Blackwell.

- Peacock, A. C. S., & Tor, D. G. (Eds.). (2015). *Medieval Central Asia and the Persianate World: Iranian Tradition and Islamic Civilisation*. I. B. Tauris.
- Pollock, S. (2006). *The Language of the Gods in the World of Men: Sanskrit, Culture, and Power in Premodern India*. University of California Press.
- Reid, A. (1988). *Southeast Asia in the Age of Commerce, 1450–1680* (Vol. 1). Yale University Press.
- Reid, A. (1993). *Southeast Asia in the Age of Commerce, 1450–1680* (Vol. 2). Yale University Press.
- Ritchie, H., & Roser, M. (2019). *Land Use*. Our World in Data. Retrieved July 28, 2026, from <https://ourworldindata.org/land-use>
- Rubin, M., & Simons, W. (Eds.). (2009). *The Cambridge History of Christianity, Volume 4: Christianity in Western Europe, c.1100–c.1500*. Cambridge University Press.
- Salomon, F., & Schwartz, S. B. (Eds.). (1999). *The Cambridge History of the Native Peoples of the Americas: South America* (Vol. 3). Cambridge University Press.
- Shepard, J. (Ed.). (2008). *The Cambridge History of the Byzantine Empire, c. 500–1492*. Cambridge University Press.
- Spolaore, E., & Wacziarg, R. (2009). The Diffusion of Development. *Quarterly Journal of Economics*, 124(2), 469–529.
- Spolaore, E., & Wacziarg, R. (2018). Ancestry and Development: New Evidence. *Journal of Applied Econometrics*, 33(5), 748–762.
- Streusand, D. E. (2011). *Islamic Gunpowder Empires: Ottomans, Safavids, and Mughals*. Westview Press.
- Subrahmanyam, S. (1997). Connected Histories: Notes Towards a Reconfiguration of Early Modern Eurasia. *Modern Asian Studies*, 31(3), 735–762.
- Taddesse, T. (1972). *Church and State in Ethiopia, 1270–1527*. Clarendon Press.
- Thornton, J. K. (1998). *Africa and Africans in the Making of the Atlantic World, 1400–1800* (2nd ed.). Cambridge University Press.
- Trigger, B. G., & Washburn, W. E. (Eds.). (1996). *The Cambridge History of the Native Peoples of the Americas: North America* (Vol. 1). Cambridge University Press.
- Vansina, J. (1990). *Paths in the Rainforests: Toward a History of Political Tradition in Equatorial Africa*. University of Wisconsin Press.
- Vaughan, M. (2005). *Creating the Creole Island: Slavery in Eighteenth-Century Mauritius*. Duke University Press.
- Walker, I. (2019). *Islands in a Cosmopolitan Sea: A History of the Comoros*. Hurst.
- Woodside, A. (2006). *Lost Modernities: China, Vietnam, Korea, and the Hazards of World History*. Harvard University Press.
- World Bank. (2025). *Climate Change Knowledge Portal: Observed Climate Data, ERA5 0.25-Degree*. <https://doi.org/10.57966/128g-6s70>