Ancestry and Development: New Evidence

Romain Wacziarg

UCLA and NBER

Enrico Spolaore

Tufts University and NBER

Ancestry and Development: New Evidence[*]

Enrico Spolaore

Tufts University and NBER

Romain Wacziarg

UCLA and NBER

March 2018

Abstract

We revisit the relationship between ancestral distance and barriers to the diffusion of development by replicating previous results with a new genomic dataset on human microsatellite variation. We again find a statistically and economically significant effect of ancestral distance from the technological frontier on income per capita. The historical pattern of the effect is hump shaped, peaking between 1870 and 1913, and declining steeply afterwards. This suggests that ancestral distance acts as a temporary barrier to the diffusion of development. We also confirm that ancestral distance from the frontier is a barrier to the spread of specific technologies and institutions in modern times.

1 Introduction

Does ancestry affect economic development? In this paper, we study the relationship between ancestral distance and the diffusion of development using new information on human microsatellite variation (Pemberton et al., 2013). This dataset leads to measures of relatedness between societies that differ in several respects from measures based on classic genetic markers (Cavalli-Sforza et al., 1994). These measures of relatedness can be used to analyze the determinants of development, shedding new light on the effects of long-term historical barriers on the spread of modern technologies and institutions.

This paper contributes to a recent literature that has explored how the characteristics of a society's ancestral population exert an influence on its current level of development (e.g., Spolaore and Wacziarg, 2009; Putterman and Weil, 2010; Ashraf and Galor, 2013; for an overview, see Spolaore and Wacziarg, 2013). In our previous work, we formulated the hypothesis that ancestry matters because more closely related populations face lower barriers to interacting and learning from each other. Thus, technological and institutional innovations are more likely to spread first across societies that share a more recent common history, and only later to more ancestrally distant societies (Spolaore and Wacziarg, 2009, 2012, 2013, 2014).

In this paper, we present further evidence in support of the hypothesis that ancestral distance acts as a significant but temporary barrier to the spread of economic development. Using data on human microsatellite variation from Pemberton et al. (2013), we construct measures of relatedness between societies, and use them to study the determinants of development. We find that ancestral barriers affect the diffusion of modern productivity and development across time and space. We document a statistically and economically significant effect of ancestral distance from the technological frontier on income per capita. The effect is robust to controlling for geographic factors; climatic differences and continental fixed effects; measures of language, religion, and common history; and the effect of genetic diversity within countries (as emphasized in the seminal contribution of Ashraf and Galor, 2013). The historical pattern of the impact of relative ancestral distance on income per capita is hump shaped, peaking between 1870 and 1913, and declining steeply afterwards. This hump shape is consistent with a gradual diffusion of development from the world technology frontier, where ancestral distance acts as a temporary barrier to the spread of modern technologies, behaviors and institutions, rather than as a permanent obstacle to economic development.[1] When

---

[1] Evidence on ancestral and cultural distance acting as temporary barriers to the spread of new social norms and

we compare our new estimates based on Pemberton et al. (2013) to the analogous regressions based on classic measures of genetic distance from Cavalli-Sforza at al. (1994), used in our previous work, we find that the results come out strengthened in terms of standardized magnitudes and levels of significance.

This paper is related to a recent empirical literature in economics that employs new information on human microsatellite variation to shed light on economic and political outcomes. However, existing contributions focus on the effects of genetic diversity *within* countries and ethnic groups. In contrast, ancestral barriers *between* populations are at the center of our analysis. Recent economic research using human microsatellite data includes several contributions: Ashraf, Galor and Klemp (2015) show a hump-shaped effect of genetic diversity on economic prosperity across ethnic groups. Depetris-Chauvin and Özak (2016) employ an ethnic-level dataset combining geo-coded ethnographic and microsatellite-based genetic data to study the positive effect of population diversity on specialization and the emergence of trade-related institutions. Finally Arbatli, Ashraf and Galor (2015) and Galor and Klemp (2016), respectively, estimate the effects of genetic diversity on internal conflict and the emergence of autocratic institutions across different ethnic groups.

In Section 2, we present the new data based on human satellite variation. Section 3 analyzes the relationships between the new measures of ancestral distance, economic development, technology and institutions. Section 4 concludes.

2   Ancestral Distance Based on Human Microsatellite Variation

Cavalli-Sforza et al. (1994), in a landmark study, provided measures of genetic distance between human populations using classic genetic markers, such as blood-group systems and variants of immunoglobulins. When studying genetic relatedness at the world level, Cavalli-Sforza and coauthors considered 42 representative populations, aggregating subpopulations characterized by a high level of genetic similarity, and reported bilateral genetic distances between these populations, computed from 120 genetic *loci*.

Advances in DNA sequencing and genotyping have allowed large-scale studies of human polymorphisms (genetic variants) directly at the molecular level. In particular, geneticists have been able to infer relatedness between human populations by studying microsatellite variation. Mi-

behavior regarding fertility is provided in Spolaore and Wacziarg (2016a).

2

crosatellites are tracts of DNA in which specific motifs, typically ranging in length from two to five base pairs, are repeated. Microsatellites have high mutation rates and high diversity, and have therefore been used by geneticists to infer how different populations are phylogenetically related to each other. Measures of genetic distance based on microsatellite variation, like those based on classic genetic markers, tend to capture mostly neutral change that is not subject to natural selection. Consequently, it is important to note that these measures do *not* capture overall differences in genetic endowments, but only the extent that different populations are related to each other - that is, the time since when two populations were the same population. This is a crucial point when interpreting the effects of such measures of ancestral distance on observable outcomes, as we will discuss in Section 3.

Early microsatellite studies of global human variation (e.g., Bowcock et al., 1994) were limited to a small number of indigenous populations. More recent research, including work based on the Human Genome Diversity Project (HGDP-CEPH), has gradually extended the data.[2] Pemberton et al. (2013) combine eight datasets covering 645 common microsatellite *loci* into a single dataset covering 267 worldwide populations, thus providing more comprehensive coverage of world populations than Cavalli-Sforza et al. (1994). The dataset from Pemberton et al. (2013) differs from Cavalli-Sforza et al. (1994) not only with respect to the genetic information on which it is based (microsatellites vs. classic genetic markers), but also in the number and specificity of populations that are covered. An important advantage of the new dataset is that it provides more detailed information on populations outside Europe - especially within Asia and Africa.

Pemberton et al. (2013), like Cavalli-Sforza et al. (1994), provide F_{ST} genetic distance data at the population level, not at the country level. Therefore, as we did in Spolaore and Wacziarg (2009), we match populations to countries, using ethnic composition data by country from Alesina et al. (2003). This source lists $1,120$ country-ethnic group categories.[3] Each ethnic group was matched to a genetic group from Pemberton et al. (2013). With this match in hand, we constructed two measures of F_{ST} genetic distance between countries. The first is the distance between the plurality groups of each country in a pair, defined as the groups with the largest shares of each country's population.[4] The second is a measure of weighted genetic distance. Denote $i = 1, ..., I$

---

[2]A description of the HGDP-CEPH is provided in Cann et al. (2002).

[3]For a more detailed explanation of our approach, see for instance Spolaore and Wacziarg (2016b).

[4]To assign the plurality match, we first cumulated the shares of groups matched to the same Pemberton et al.

the populations of country 1, $j = 1, ..., J$ those of country 2, s_{1i} the share of population i in country 1 (similarly for country 2) and d_{ij} the genetic distance between populations i and j. Then the weighted F_{ST} genetic distance between countries 1 and 2 is defined as:

$$F_{ST}^{W} = \sum_{i=1}^{I} \sum_{j=1}^{J} (s_{1i} \times s_{2j} \times d_{ij}) \tag{1}$$

The interpretation of F_{ST}^{W} is straightforward: it represents the expected genetic distance between two randomly selected individuals, one from each country.[5] In addition, we employ the data from Pemberton et al. (2013) to construct genetic distances matched to populations as they were in 1500 AD (F_{ST}^{1500}), before the movements that followed modern explorations and conquests. For this variable, in particular, New World countries are matched to their corresponding aboriginal populations. The resulting data series can be compared to its analog obtained using data from Cavalli-Sforza et al. (1994).

As already mentioned, an advantage of using the genetic-distance data from Pemberton et al. (2013) versus Cavalli-Sforza et al. (1994) is that it allows a finer match of populations to ethnic groups in Asian and African countries. For example, most ethnic groups in Afghanistan are matched to one large population from Cavalli-Sforza et al. ("Iranian"), while Pemberton et al.'s data allow an exact match to specific Afghan groups, such as "Balochi" and "Hazara." Because of such finer partitions, in the new dataset we are able to match about twice as many populations to ethnic groups from Alesina et al. (2003) as we did when using the data from Cavalli-Sforza et al. (1994).

Table 1 present summary statistics for all six measures of genetic distance - that is, plurality F_{ST}, weighted F_{ST}^{W} and per-modern F_{ST}^{1500}, each from Cavalli-Sforza et al. (1994) and Pemberton et al. (2013). Panel A describes the mean and variation of these six measures, while Panel B shows their pairwise correlations. Distances based on the Pemberton et al. (2013) data are highly but not perfectly correlated with the corresponding measures from Cavalli-Sforza et al. (1994), consistent with the fact that they capture conceptually analogous relations, but are based on different biological information, sampling, and population coverage. The correlation between the

(2013) genetic populations, and then picked in each country the group with the largest cumulative share, as we did using the Cavalli-Sforza data in our previous work.

[5] The weighted measure is not to be interpreted as F_{ST} genetic distance between the *whole* population of a country (say, all Australians) and the *whole* population of another country (say, all Americans), as if each country were formed by one randomly mating population.

4

two measures of plurality F_{ST} is 0.785, while the correlation between the two weighted F_{ST}^W is 0.829 and the correlation between the two pre-modern distances F_{ST}^{1500} is 0.757. The correlations between pairs of distances within each dataset are similar. For example, in the Pemberton-based dataset the correlation between plurality F_{ST} and weighted F_{ST}^W is 0.917, while the correlation between weighted F_{ST}^W and pre-modern F_{ST}^{1500} is 0.632, while the corresponding correlations in the Cavalli-Sforza-based dataset are respectively 0.938 and 0.732. In the rest of this paper, we use the new Pemberton-based measures to study the relation between ancestry and development.

3   Ancestral Distance and the Dynamics of Income Differences

In our previous work (starting with Spolaore and Wacziarg, 2009) we studied the diffusion of economic development using measures of ancestral distance between countries based on data from Cavalli-Sforza et al. (1994), testing the hypothesis that ancestral distance from the technological frontier acts as a barrier to the spread of innovations, institutions and development. The underlying idea was that populations at a greater distance from each other had more time to diverge in terms of intergenerationally transmitted traits, such as cultural norms, values, beliefs, habits, language, religion, etc. Empirical evidence on this close association between ancestry, language and culture is provided in Spolaore and Wacziarg (2016b). Such a long-term divergence in cultural traits is hypothesized to be an obstacle to communication, social interaction and learning across different societies, therefore hindering the diffusion of economic development to societies which are historically and culturally farther from the world technological frontier. In this section, we revisit the analysis and test these hypotheses using the new genetic distance measures constructed from the dataset in Pemberton et al. (2013).[6]

3.1   Income Levels

Baseline results.   We first test whether countries that are at a higher ancestral distance from the frontier have lower incomes per capita in 2005.[7] We consider the United States as the technological frontier, and measure ancestral distance from the US using our new weighted F_{ST}^W from the

---

[6] All the empirical results discussed in this section can be readily compared to their exact analogs using the Cavalli-Sforza data, to be found in the Appendix to this paper, Tables A2-A10. Additionally, the new genetic distance data used here are available on the authors' websites.

[7] Income per capita data is from the Penn World Tables, version 6.3.

Pemberton et al. (2013) dataset. The specification is:

$$\log y_i^{2005} = \alpha_0 + \alpha_1 G_{i,USA}^D + \alpha_2' X_i + \varepsilon_i \tag{2}$$

where $G_{i,USA}^D$ is defined as F_{ST}^W between country i and the US and X_i is a vector of control variables. The results are presented in Table 2. In all columns ancestral distance from the US has the expected negative sign and is statistically significant. In column (1), where ancestral distance is entered alone in the sample of 174 countries for which we have data, the standardized β on ancestral distance from the US is 54.5%. Since genetic distance is correlated with geographic distance, it is important to control for geographic variables: in column (2) we add several controls for geographical features, such as absolute latitude (i.e., distance to the equator), a landlocked country dummy, an island country dummy. We also add controls for geographical barriers with the US (geodesic distance from the US and absolute differences in latitude and longitude to the US, etc.). Ancestral distance from the US continues to have a high and significant effect on income per capita, with a standardized β equal to 44.5%. In column (3) we add a control for the percentage of country i's land area that is located in the tropics. The standardized β on ancestral distance to the US declines slightly, but its effect remains statistically and economically significant.

Ancestral distance today could be related to income not because it hindered the diffusion of development but because frontier populations settled in regions prone to generating high incomes. In order to control for the possible endogeneity of ancestral distance with respect to income differences, in column (5) we instrument for contemporary ancestral distance from the US using ancestral distance from the English in 1500 AD. We use pre-modern genetic distance to the English as an instrument because it is highly correlated with current genetic distance to the US, but was determined before the large movements of people due to post-Columbian exploration and conquests. In addition, this IV approach can address measurement error due to imperfect matching between populations and ethnic groups in modern times, to the extent that errors in measurement across F_{ST}^W and F_{ST}^{1500} are independent. Indeed, when using IV, the effect of ancestral distance is slightly higher than in the OLS regressions, with a standardized β equal to 61.9% (column 4).

Continental Effects. Historical separation times between populations tend to be larger across continents than within continents, because in pre-modern times (and especially before the discovery of the New World), it was harder for populations to travel from one continental mass to another. To address the possibility that genetic distance may be capturing continental effects, Table 3 conducts a

systematic investigation of the robustness of our findings to continental effects. To do so we pursue two strategies. First, in columns (1) and (2) we add sets of regional dummies to our baseline specification (that of column 2 of Table 2). Column (1) includes a full set of continental effects, where the continents are defined as the 6 major populated continents (Oceania is the excluded category). The effect of genetic distance falls from 44.5% to 34.8% and remains highly significant statistically. Column (2) conducts a more demanding test by including 11 dummies representing macro-regions (as defined by the World Bank). These are finer groupings of countries compared to continents (for instance Africa is composed of two macro-regions, Sub-Saharan Africa and North Africa). Interestingly, with this broader set of regional dummies the standardized magnitude of the effect of genetic distance to the frontier is similar to that in column 1.

Our second strategy to address continental effects is to exclude certain continents from the sample. There is a particular concern that Sub-Saharan Africa might drive the result (being a region that is both poor and genetically distant from the frontier). We find on the contrary that the standardized magnitude of ancestral distance to the US *rises* a little in the sample that excludes Sub-Saharan Africa (column 3 of Table 3). There is also a concern that European countries might drive the result, since European countries are both rich and genetically close to the frontier. Indeed the effect of genetic distance falls to some degree compared to the baseline, to a standardized magnitude of 36.1% (column 4). Finally we removed both Sub-Saharan African countries and European countries, finding again that genetic distance to the US has a negative, statistically significant, and sizable effect on log per capita income (column 5). In sum, our basic conclusion is preserved under both approaches to allow for continental effects.[8]

Genetic Diversity, Genetic Distance and Cultural Distance. In Table 4, building on the contribution by Ashraf and Galor (2013), we add controls for the effect of genetic diversity *within* each country. Ashraf and Galor (2013) construct measures of genetic diversity within modern countries using microsatellite-based genetic information about 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel. They find that genetic diversity has a non-monotonic hump-shaped effect on development, increasing at lower levels and decreasing at higher levels. They interpret their finding as resulting from a trade-off between the costs and benefits

---

[8]Even if continental effects significantly altered our conclusions, the issue of how to interpret them would remain. Continental effects may capture geographic features or barriers that prevent the diffusion of development from the frontier, but they may also capture cultural and historical barriers of the type we emphasize.

from having a heterogeneous population. In Table 4 column (1) we enter our new measure of genetic distance alongside genetic diversity and its square (from Ashraf and Galor, 2013). We find statistically significant effects for all the estimated coefficients, with the standardized β for genetic distance equal to 60.4%. In column (2) we add the same geographical controls used in Table 2. The effects of the ancestral variables (genetic distance and genetic diversity) remain statistically significant, and the standardized β on genetic distance remains high at 56.9%. Finally, in column (3) we add measures of cultural distance to the US, namely linguistic and religious distance.[9] We expect such measures to reduce the effect of genetic distance, as language and religion form part of the intergenerationally transmitted traits that could account for human barriers from the US. This is indeed what we find, as the magnitude of the effect of genetic distance falls when including linguistic and religious distance to the US.[10]

Interpretation. It is important to remember that our measures of ancestral distance are based on parts of the DNA that tend to vary through random mutation and drift, not as the result of natural selection. Hence, the relation between ancestral distance and income should not be interpreted as the effect of specific differences in genetic endowments between populations. Instead, the effect of ancestral distance from the technological frontier can be interpreted as the outcome of barriers across societies that are more distantly related. Such barriers result from divergence in intergenerationally transmitted traits that hinder interaction and communication. As pointed out in the scientific literature on human evolution, a large part of the variance in intergenerationally-transmitted traits among humans stems from cultural transmission (e.g., see Richerson and Boyd, 2005, Spolaore and Wacziarg, 2013). In the rest of this section, we provide further evidence consistent with an interpretation in terms of temporary barriers to the horizontal diffusion of modern economic development across historically and culturally distant societies.

---

[9]See Spolaore and Wacziarg (2009, 2016b) for details on these measures. The source data is from Fearon (2003) and Mecham, Fearon and Laitin (2006).

[10]The standardized β falls from 56.9% in column (2) to 46.9% in column (4), while the sample is reduced from 148 to 140 countries. The change in the sample is responsible for a 4.6 percentage point decline in the standardized β while the addition of linguistic and religious distance is responsible for a 5.5 percentage point decline - about 10% of the total effect.

3.2 Income Differences

To more precisely assess the role of ancestral distance as a barrier to development, we turn to a bilateral approach where a measure of economic distance - the absolute difference in the log of per capita income between two countries i and j - is regressed on measures of geographic and genetic distance between them. Define *absolute* genetic distance, G_{ij}^D as equal to F_{ST}^W between countries i and j, and *relative* genetic distance, $G_{ij}^R = \left| G_{i,USA}^D - G_{j,USA}^D \right|$. The simple models of diffusion in Spolaore and Wacziarg (2009, 2014) predict that economic distance should be positively related to G_{ij}^D, but that G_{ij}^R should be a stronger predictor of economic distance and trump the effect of G_{ij}^D when both measures are entered together. The specification is now:

$$\left| \log y_i^{2005} - \log y_j^{2005} \right| = \beta_0 + \beta_1 G_{ij}^D + \beta_2 G_{ij}^R + \beta_3' X_{ij} + \nu_{ij} \tag{3}$$

where the diffusion framework predicts $\beta_1 = 0$ and $\beta_2 > 0$.[11] The baseline results are presented in Table 5. In columns (1) and (2) we find indeed that both absolute and relative genetic distance positively predict income differences when these variables are entered separately, and that the magnitude of the effect of relative genetic distance is the larger of the two. In column (3), when entering both measures together, we see that the coefficient on G_{ij}^R remains positive and significant, while the coefficient on G_{ij}^D becomes statistically indistinguishable from zero, as the diffusion models predicts. Finally, in column (4) we instrument for G_{ij}^R using relative distance to the US using the 1500 match. The coefficient barely changes from the baseline.

Several extensions and robustness tests are presented in Table 6. In the first column, we include a broad set of continental dummies. For each continent, we define a dummy for both countries in a pair belonging to that continent, and another dummy for whether one and only one country in a pair belongs to that continent. The effect of relative genetic distance is reduced but not eliminated. In column (2) we remove every pair involving at least one country from the New World (Americas, Oceania) from the sample. The idea is to further reduce the possible endogeneity of genetic distance to the frontier induced by post-Columbian population movements. The standardized effect of G_{ij}^R (33.7%) is actually larger than in the corresponding full sample baseline of Table 5, column (1) (23.5%). Column (3), in another attempt to control for continental effects, removes all pairs involving at least one country from Sub-Saharan Africa from the sample. The effect of G_{ij}^R, while

---

[11] To account for the effects of spatial correlation induced by the presence of $\log y^{2005}$ for countries i and j in multiple pairs of countries, we two-way cluster standard errors at the level of i and j (Cameron, Gelbach and Miller, 2011).

smaller, remains positive and significant. Column (4) controls for climatic similarity, defined as the average absolute difference in the shares of each country's area in each of twelve climatic zones. The effect of G_{ij}^R remains positive, large, and significant. Finally, in column (5) we add measures of common history, religious distance and linguistic distance. As before, we expect that including these variables may reduce the effect of genetic distance relative to the frontier. This is only barely the case, as the standardized β on G_{ij}^R is 33.6%, while it is 34.8% in the same sample without the common history variables. Linguistic and religious distance themselves bear the expected positive signs. In sum, both the baseline results and the main robustness tests in Spolaore and Wacziarg (2009, 2012, 2013, 2014) are qualitatively unchanged when using the new dataset of genetic distance.[12]

3.3   Historical Pattern

An additional prediction of our diffusion hypothesis is that the effect of genetic distance relative to the frontier should be hump shaped: in the early phases of the diffusion process, only the frontier has adopted modern methods of production. Subsequently, societies that are ancestrally close start to industrialize, so relative genetic distance has a larger effect on economic differences. Later, economic modernity reaches more distant populations, and the effect of genetic distance fades away as populations at farther and farther distances from the frontier adopt modern methods of production. We explore this hypothesis using the diffusion of the Industrial Revolution from England, starting in the first half of the 19th century. Table 7 provides strong evidence supportive of just such a pattern. The frontier is now defined as the United Kingdom, and we use data from Maddison (2003) on income per capita in 1820 and 1913. We find that the standardized magnitude of G_{ij}^R estimated in a balanced sample of 820 country pairs (from 41 countries) starts at a modest 12.6% in 1820, peaks at 28.2% in 1870, and declines gradually thereafter to reach 12.7% in 2005 (Figure 1). This hump-shaped effect of G_{ij}^R is strongly supportive of the hypothesis that ancestral distance constitutes a temporary barrier to the diffusion of development from the world's

---

[12]In addition, when comparing the results of Table 6 to those of Table A6 in the Appendix, where the Cavalli-Sforza et al (1994) data on genetic distance is used instead of the data from Permberton et al. (2013), we see that the standardized magnitude of the effect of relative F_{ST} genetic distance to the frontier is generally higher using the new data.

technological and institutional frontier.[13]

3.4 Ancestral Distance, Institutions and Technologies

To further establish how ancestral distance affects the diffusion of development, our past research sought to show that ancestral barriers also hinder the flow of variables that are more proximate determinants of income levels, such as institutions and specific technologies. Here we revisit the relationship between ancestral distance, institutions and technologies using our new genetic distance dataset. We adopt a specification where bilateral differences in measures of institutional quality or technological sophistication are regressed on relative distance to the institutional and technological frontier, as well as a set of geographic distance controls. This specification is analogous to that in equation (3) but with different dependent variables.

3.4.1 Institutions and the Diffusion of the Third Wave of Democratization

To assess whether ancestral distance affects institutional differences we closely follow the approach and specification in Spolaore and Wacziarg (2016c), but we use the Pemberton et al. (2013) data instead of the Cavalli-Sforza et al. (1994) data. We rely on two sets of commonly used indicators of institutional quality, all measured in 1990 to maximizes the size of the sample. The first is the POLITY2 score, a composite democracy score from the Polity IV project (Marshall, Gurr and Jaggers, 2016). The variable can take 21 possible values, with -10 denoting autocracy and 10 democracy, and every shade inbetween. The composite score is based on underlying variables that capture the breadth of political participation in competitive elections and constraints on the exercise of executive power. The second set is a series of three subjective indicators of institutional quality from the International Country Risk Guide (ICRG), which have often been used in the literature on institutions and economic outcomes (ICRG, 2013). We consider the risk of repudiation of contracts, the risk of expropriation, and an index of the rule of law. For all these measures, we compute the absolute average difference across pairs of countries, a measure of their institutional distance, and use the resulting differences as dependent variables.

Table 8 presents estimation results from a specification that includes our basic set of geographic

---

[13]We examined the robustness of these findings to changing the frontier from the UK to the USA, to answer the concern that the decline of the effect after 1870 may be due to the decline of the UK as a technological leader. The results are presented in Appendix Table A1. The magnitude of the effect of relative genetic distance to the frontier at various dates is very similar with the US as the frontier, and the hump-shaped pattern through time is preserved.

distance controls. Across all four dependent variables, we find positive and statistically significant effects of genetic distance relative to the United States on institutional differences. The standardized magnitude of the effect varies from 16.9% (Polity2 score) to 26.1% (rule of law indicator). These results are analogous to those found for income levels. Moreover, the magnitude of the effect is everywhere larger than in the corresponding regressions (with the same sample of country pairs) using the Cavalli-Sforza data (compare Table 8 to Appendix Table A8).

To further assess the role of ancestral distance as a barrier to the diffusion of institutions, we conduct an exercise across time analogous to that presented in Section 3.3 for income per capita in the aftermath of the Industrial Revolution. Here, the specific instance of institutional diffusion that we examine is that of the third wave of democratization (Huntington, 1993). From its inception in Portugal in 1974, this wave of democracy spread gradually to more and more countries during the ensuing two decades. We study the period before and during the third wave by regressing differences in the Polity2 score of democracy on relative genetic distance to the institutional frontier (the US), at different dates starting in 1960 and ending in 2005, using a common sample of 4,278 country pairs based on 93 underlying countries. The results appear in Table 9. We find that genetic distance relative to the frontier positively predicts differences in the democracy score at all dates. However, the effect is smaller at the beginning and at the end of the period, and is maximized in 1990 when the third wave reached its peak. In other words, we find a hump-shaped pattern similar to that uncovered for income, consistent with a model where ancestral distance acts as a barrier to the diffusion of institutional change.

3.4.2   The Diffusion of Technologies

The final exercise that we carry out consists of looking at the effect of ancestral distance on the diffusion of specific kinds of technologies. To do so we follow closely the approach and specification in Spolaore and Wacziarg (2012, 2014). We use measures of technological adoption from Comin, Easterly and Gong (2010) for the period 1970-2000. These authors used underlying technology usage data for nine technologies from Comin et al. (2008).[14] From usage data of these technologies, the authors construct an index of usage lag relative to the frontier (the United States), and group the resulting technology usage intensity metrics into four major categories: agricultural, transporta-

---

[14]Comin et al. (2008) specify that "these technologies are electricity (in 1990), Internet (in 1996), PCs (in 2002), cell phones (in 2002), telephones (in 1970), cargo and passenger aviation (in 1990), trucks (in 1990), cars (in 1990), and tractors (in 1970) all in per capita terms." (p. 75-76).

tion, communication, and industrial technologies. A fifth category is obtained by aggregating all technologies into a single index of overall technological usage intensity. We further calculated the absolute difference in technological usage intensity between pairs of countries for each of the five major indicators. These are our dependent variables. We seek to relate ancestral distance relative to the frontier to bilateral differences in technological usage intensities.[15]

The results are presented in Table 10. Across all measures, we find evidence that higher genetic distance relative to the United States is positively correlated with absolute differences in technology usage intensity. The magnitudes, as revealed by the standardized β coefficients across the various technology categories, varies between 18.5% and 35.5% (the latter for the difference in usage intensity for industrial technologies). These estimates are larger in magnitude and more significant statistically than results using the Cavalli-Sforza et al. (1994) data (compare Table 10 to Appendix Table A10). In sum, the results reported above for per capita income and institutions carry over to indices capturing more specific technological differences.

4   Conclusion

This paper uses novel measures of ancestral distance between human societies to shed light on the diffusion of economic development, institutions and technologies. We reach three main conclusions:

First, countries at a higher ancestral distance from the technological frontier (the United States) had a lower income per capita in 2005. The effect is robust to controlling for geographical barriers, climatic differences, measures of linguistic and religious distance, and the effect of genetic diversity within populations. These results confirm those first presented in Spolaore and Wacziarg (2009), and extend them by showing robustness to the inclusion of additional control variables such as regional fixed effects and genetic diversity.

Second, the effect of relative ancestral distance from the technological frontier has a statistically and economically significant effect on income differences, and dominates the effect of absolute ancestral distance in a horserace between the two variables. This is consistent with the hypoth-

---

[15]Differences in technological usage intensities are positively correlated with each other, but these correlations are of moderate size. Among the four sectoral indices, the correlations range from 0.43 to 0.63. The correlations between each of these sectoral indices and the overall index of technological usage intensity range from 0.70 to 0.79. Thus, there is merit in considering these different indicators separately. Spolaore and Wacziarg (2014), p. 154 contains further details.

esis that ancestral distance acts as a barrier to the diffusion of economic development from the technological frontier. Our interpretation is that societies more closely related to the innovators share more similar traits with them – such as cultural norms, habits, communication styles etc. – facilitating learning and imitation. Instead, societies that are more distant, on average, have diverged more in those cultural traits, and therefore face greater obstacles when interacting with the technological innovators.

Finally, we find that the historical pattern of the impact of relative ancestral distance from the frontier on income per capita is hump shaped, peaking between 1870 and 1913, and declining steeply afterwards. This shows that the effect of long-term divergence in inherited traits – captured by ancestral distance – is important but not fixed and immutable. The effect depends on dynamic factors, such as the location of the frontier and the gradual spread of innovations, and thus it declines over time as more and more societies adopt the institutions and technologies of the frontier innovator.

In sum, ancestry matters but it is not permanent destiny. A widespread concern when considering the effects of ancestry and long-term history on development is that not much can be done today to change those factors. However, if a substantial share of the variation in income per capita is due to temporary barriers to the diffusion of innovations, there is scope for policy action. Economic development could be fostered through policies that reduce obstacles to communication and interaction across different cultures and societies. The exploration of such policies is an important topic for future research.

References

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2), 155-194.

Arbatli, C., Ashraf, Q., & Galor, O. (2015). The nature of conflict. *NBER Working Paper #21079*.

Ashraf, Q., & Galor, O. (2013). The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic Development. *American Economic Review*, 103(1), 1-46.

Ashraf, Q., Galor, O., & Klemp, M. (2015). Heterogeneity and productivity. *Working Paper #2015-4*, Brown University, Department of Economics.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., et al. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature,* 368, 455-457.

Cameron, C., Gelbach, J., & Miller, D. (2011). Robust inference with multi-way clustering. *Journal of Business and Economic Statistics*, 29 (2), 238-249.

Cann, H. M., de Toma, C., Cazes, L., Legrand, M.-F., Morel, V., Piouffre, L., Bodmer, J., et al. (2002). A human genome diversity cell line panel. *Science,* 296(5566), 261-62.

Cavalli-Sforza, L. L., Menozzi P., & Piazza, A. (1994). *The history and geography of human genes.* Princeton, NJ: Princeton University Press.

Comin, D., Easterly, W., & Gong, E. (2010). Was the wealth of nations determined in 1000 B.C.? *American Economic Journal: Macroeconomics,* 2(3), 65-97.

Comin, D., Hobijn, B., & Rovito, E. (2008). World technology usage lags, *Journal of Economic Growth,* 13(4), 237-256.

Depetris-Chauvin, E., & Özak, Ö. (2016). Population diversity, division of labor and comparative development. *Working Paper* #1605, Southern Methodist University, Department of Economics.

Fearon, J. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2), 195-222.

Galor, O. & Klemp, M. (2016). Roots of autocracy. NBER Working Paper #23301.

Huntington, S. (1993). *The Third Wave: Democratization in the late 20th century.* Norman, OK: University of Oklahoma Press.

International Country Risk Guide (ICRG) (2013). International Country Risk Guide (ICRG) researchers dataset. hdl:1902.1/21446, Harvard Dataverse, V3.

Maddison, A. (2003). *The World economy: Historical statistics.* Paris, France: OECD Development Center.

Marshall, M., Gurr, T., & Jaggers, K. (2016). *POLITY IV PROJECT Political regime characteristics and transitions, 1800-2015 dataset users' manual.* Center for Systemic Peace.

Mecham, Q., Fearon, J., & Laitin, D. (2006). Religious classification and data on shares of major world religions. Unpublished, Stanford University.

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, 903-919.

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.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution.* Chicago, IL: University of Chicago Press.

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

Spolaore, E., & Wacziarg, R. (2012). Long-term barriers to the international diffusion of innovations. Chapter 1 in Frankel, J., & Pissarides, C. (eds.), *NBER International Seminar On Macroeconomics 2011*, Chicago, IL: University of Chicago Press.

Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of Economic Literature*, 51(2), 1-45.

Spolaore, E., & Wacziarg, R. (2014). Long-term barriers to economic development. Chapter 3 in Aghion, P., & Durlauf, S. (eds.), *Handbook of Economic Growth*, vol. 2A, 121-176. Amsterdam: North Holland.

Spolaore, E., & Wacziarg, R. (2016a). Fertility and modernity. *Working Paper* #799, Tufts University, Department of Economics.

Spolaore, E., & Wacziarg, R. (2016b). Ancestry, language and culture. Chapter 7 in Ginsburgh, V., & Weber, S. (eds.), *The Palgrave Handbook of Economics and Language*, London, UK: Palgrave Macmillan.

Spolaore, E., & Wacziarg, R. (2016c). The diffusion of institutions. Chapter 9 in Wilson, D. S., & Kirman, A. (eds), *Complexity and Evolution: Toward a New Synthesis for Economics.* Strüngmann Forum Reports, vol. 19, 147-166. Cambridge, MA: MIT Press.

Table 1 – Summary Statistics for the Genetic Distance measures, from both Pemberton et al. (2013) and Cavalli-Sforza et al. (1994)

Panel A – Mean and Variation

Variable	Mean	Std. Dev.	Min	Max
F_{ST} genetic distance, plurality match, Pemberton et al.	0.037	0.022	0.000	0.106
F_{ST} genetic distance, plurality match, Cavalli-Sforza et al.	0.117	0.081	0.000	0.338
F_{ST} genetic distance, weighted, Pemberton et al.	0.037	0.019	0.000	0.095
F_{ST} Genetic Distance, weighted, Cavalli-Sforza et al.	0.115	0.070	0.000	0.355
F_{ST} genetic distance, 1500 match, Pemberton et al.	0.045	0.025	0.000	0.106
F_{ST} Genetic Distance, 1500 match, Cavalli-Sforza et al.	0.125	0.079	0.000	0.356

(All statistics are computed from 15,051 country pair observations based on 174 countries)

Panel B – Correlations

	Plurality F_{ST}, Pemberton	Plurality F_{ST}, Cavalli-Sforza	Weighted F_{ST}, Pemberton	Weighted F_{ST}, Cavalli-Sforza	1500 F_{ST}, Pemberton
F_{ST} genetic distance, plurality match, Cavalli-Sforza et al.	0.785	1			
F_{ST} genetic distance, weighted, Pemberton et al.	0.917	0.786	1		
F_{ST} Genetic Distance, weighted, Cavalli-Sforza et al.	0.737	0.938	0.829	1	
F_{ST} genetic distance, 1500 match, Pemberton et al.	0.574	0.454	0.632	0.494	1
F_{ST} Genetic Distance, 1500 match, Cavalli-Sforza et al.	0.510	0.694	0.589	0.732	0.757

(All statistics are computed from 15,051 country pair observations based on 174 countries)

Table 2 – Income Level Regressions, controlling for geographic distance
(Dependent variable: log income per capita 2005)

	(1) Univariate	(2) Distance and geography controls	(3) Add tropics control	(4) IV using 1500 genetic distance
F_{ST} genetic distance to the USA, weighted, Pemberton et al.	-43.594 (9.12)***	-35.610 (5.67)***	-37.720 (5.16)***	-53.372 (3.67)***
Absolute latitude		0.025 (3.75)***	0.030 (3.10)***	0.021 (1.56)
Landlocked dummy		-0.549 (3.14)***	-0.477 (2.67)***	-0.395 (2.11)**
Island dummy		0.750 (3.70)***	0.486 (1.78)*	0.519 (1.84)*
Geodesic distance to the USA		0.812 (1.29)	1.317 (1.84)*	1.098 (1.31)
Absolute difference in latitude to the USA		-0.167 (0.22)	-0.211 (0.27)	0.305 (0.30)
Absolute difference in longitude to the USA		-0.967 (2.11)**	-1.213 (2.37)**	-0.994 (1.58)
Dummy for common sea/ocean with the USA		-0.161 (0.91)	0.011 (0.05)	0.082 (0.39)
Dummy for contiguity to the USA		0.575 (1.82)*	0.551 (1.60)	0.693 (1.84)*
% land area in the tropics			-0.008 (0.02)	-0.018 (0.05)
# of observations	171	171	150	150
Adjusted R²	0.29	0.48	0.52	0.51
Standardized β on genetic distance (%)	54.49	44.51	43.77	61.94

Robust t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$
All regressions include an intercept (estimates not reported).

Table 3 – Robustness of the Effect of Genetic Distance to Continental Effects

(Dependent variable: log income per capita 2005)

	(1) All Continent dummies	(2) All Macro-region dummies	(3) Removing SS African countries	(4) Removing European countries	(5) Removing European and SS African countries
F_{ST} genetic distance to the USA, weighted, Pemberton et al.	-27.808 (3.77)***	-28.072 (3.71)***	-33.081 (4.51)***	-31.433 (4.83)***	-26.435 (3.09)***
# of observations	171	171	126	135	90
Adjusted R^2	0.52	0.58	0.32	0.37	0.22
Standardized β on gen. dist. (%)	34.76	35.09	48.47	36.09	37.66

Robust t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

Continent dummies in column (1) are for Europe, Africa, Latin America & Caribbean, Asia, North America, and Oceania (omitted category). Macro-region dummies in column (2) are for Sub-Saharan Africa, North America, Western & Southern Europe, Eastern & Central Europe, Latin America (excl. Caribbean), Caribbean, North Africa, East & Southeast Asia, South Asia, Southwestern & Central Asia, Middle East, Oceania (omitted category).

All regressions include additional controls for absolute latitude, landlocked dummy, island dummy, geodesic distance to the USA, absolute difference in latitude to the USA, absolute difference in longitude to the USA, dummy for common sea/ocean with the USA, dummy for contiguity to the USA.

19

Table 4 – Genetic Distance, Genetic Diversity, Language and Religion
(Dependent variable: log income per capita 2005)

	(1) Gen. div. controls	(2) Gen. div. & dist. & geo. controls	(3) Add linguistic and religious distance
F_{ST} genetic distance to the USA, weighted, Pemberton et al.	**-48.289 (10.90)***	**-48.962 (5.29)***	**-39.482 (4.35)***
Predicted genetic diversity, Ashraf-Galor (2013)	156.853 (3.69)***	199.496 (3.16)***	176.082 (2.86)***
Predicted genetic diversity squared, Ashraf-Galor (2013)	-120.518 (3.83)***	-146.501 (3.16)***	-129.630 (2.88)***
Linguistic distance to the USA, Fearon measure, weighted			-0.073 (0.14)
Religious distance to the USA, Mecham-Fearon-Laitin, weighted			-0.835 (1.24)
# of observations	169	148	140
Adjusted R^2	0.42	0.54	0.60
Standardized β on genetic distance (%)	60.44	56.91	46.87

Robust t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

Columns (2) and (3) include controls for absolute latitude, landlocked dummy, island dummy, geodesic distance to the USA, absolute difference in latitude to the USA, absolute difference in longitude to the USA, dummy for common sea/ocean with the USA, dummy for contiguity to the USA, and percentage of land area in the tropics.

Table 5 – Income difference regressions
(Dependent variable: absolute difference in log per capita income, 2005)

	(1) Relative GD	(2) Simple GD	(3) Horserace between simple and relative GD	(4) 2SLS with 1500 GD
Relative F_{ST} genetic distance to the USA, weighted, Pemberton et al.	**18.525 (5.099)***		**17.565 (4.772)***	**16.348 (2.931)***
Simple F_{ST} genetic distance, weighted, Pemberton et al.		**8.502 (3.101)***	**1.272 (0.527)**	
Absolute difference in latitudes	-0.495 (2.167)**	0.117 (0.469)	-0.452 (2.020)**	-0.437 (1.703)*
Absolute difference in longitudes	-0.624 (2.806)***	-0.432 (2.025)**	-0.593 (2.805)***	-0.627 (2.807)***
Geodesic Distance (1000s of km)	0.056 (2.110)**	0.019 (0.704)	0.050 (1.964)**	0.056 (2.137)**
1 for contiguity	-0.522 (8.854)***	-0.539 (9.294)***	-0.516 (8.849)***	-0.532 (8.454)***
=1 if either country is an island	-0.039 (0.584)	-0.017 (0.255)	-0.039 (0.582)	-0.035 (0.541)
=1 if either country is landlocked	0.152 (1.788)*	0.138 (1.563)	0.151 (1.752)*	0.152 (1.736)*
=1 if pair shares at least one sea or ocean	0.006 (0.084)	-0.005 (0.072)	0.004 (0.057)	0.006 (0.086)
R^2	0.07	0.05	0.07	0.04
Standardized β on genetic distance (%)	23.47	15.91	22.26	20.45

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

All regressions are based on 14,365 country pair observations from 170 countries.

Table 6 – Income difference regressions, robustness and extensions
(Dependent variable: absolute difference in log per capita income, 2005)

	(1) Continent dummies	(2) Excl. New World	(3) Excl. Sub-Saharan Africa	(4) Climate Similarity	(5) Common history controls
Relative F$_{ST}$ genetic distance to the USA, weighted, Pemberton et al.	13.234 (3.733)***	29.025 (5.164)***	9.536 (2.807)***	26.803 (6.226)***	28.384 (6.464)***
Measure of climatic difference of land areas, by 12 KG zones				0.026 (4.153)***	
1 if countries were or are the same country					-0.404 (4.703)***
1 for pairs ever in colonial relationship					0.188 (2.110)**
1 for common colonizer post-1945					-0.028 (0.389)
1 for pairs currently in colonial relationship					-0.716 (4.155)***
Religious distance index, relative to USA, weighted					0.957 (4.152)***
Linguistic distance index, relative to USA, weighted					0.336 (1.685)*
R^2	0.15	0.12	0.04	0.14	0.16
Observations	14,365	8,256	7,750	11,026	10,296
(countries)	(170)	(129)	(125)	(149)	(144)
Standardized β on genetic distance (%)	16.77	33.70	15.08	31.47	33.62

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regression includes an intercept (estimates not reported).

All columns include controls for: absolute difference in latitudes, absolute difference in longitudes, geodesic distance, dummy for contiguity, dummy for either country being an island, dummy for either country being landlocked, dummy = 1 if pair shares at least one sea or ocean. Column 1 includes a full set of continental dummy variables: both in Asia Dummy, both in Africa Dummy, both in Europe Dummy, both in Latin America/Caribbean dummy, Both in Oceania Dummy, Dummy if one and only one country is in Asia, Dummy if one and only one country is in Africa, dummy if one and only one country is in Europe, dummy if one and only one country is in North America, dummy if one and only one country is in South America.

22

Table 7 – Regressions using Historical Income Data

(Dependent variable: absolute difference in log per capita income, various dates as in row 2)

	(1)	(2)	(3)	(4)	(5)
	Income 1820	Income 1870	Income 1913	Income 1960	Income 2005
Relative F_{ST} genetic distance to the UK, weighted, Pemberton et al.	3.026 (2.130)**	9.906 (4.085)***	9.135 (3.194)***	14.336 (5.117)***	15.626 (5.208)***
R^2	0.27	0.22	0.19	0.19	0.08
Observations	1,081	1,540	1,711	5,460	14,365
(countries)	(47)	(56)	(59)	(105)	(170)
Standardized β on genetic distance (%)	12.31	27.84	22.78	30.00	23.97
Standardized β on genetic distance (%) for a common sample [a]	12.59	28.18	25.54	20.51	12.67

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

[a]. the common sample is composed of 820 pairs (41 countries).

All regressions include an intercept (estimates not reported)

All columns include controls for: absolute difference in latitudes, absolute difference in longitudes, geodesic distance, dummy for contiguity, dummy for either country being an island, dummy for either country being landlocked, dummy = 1 if pair shares at least one sea or ocean.

Table 8 – Determinants of Institutional Differences
(Dependent variables: measures of institutional quality for 1990, as in row 2)

	(1) Polity score 1990	(2) Repudiation of contracts 1990	(3) Risk of expropriation 1990	(4) Rule of law 1990
Relative genetic distance to the USA, weighted, Pemberton et al.	94.502***	41.522***	35.836***	26.699***
	(3.47)	(4.62)	(4.72)	(4.49)
Absolute difference in latitudes	-2.242	-0.545	-0.719*	1.917***
	(-1.24)	(-0.93)	(-1.65)	(4.48)
Absolute difference in longitudes	-2.709	-1.288**	-1.090**	0.514
	(-1.43)	(-2.02)	(-2.10)	(1.28)
Geodesic Distance (1000s of km)	0.441**	0.132*	0.094	-0.074
	(2.11)	(1.80)	(1.58)	(-1.56)
1 for contiguity	-2.891***	-0.563**	-0.446*	-0.146
	(-4.69)	(-2.41)	(-1.84)	(-1.49)
=1 if either country is an island	0.008	0.104	-0.045	0.156*
	(0.02)	(0.64)	(-0.41)	(1.69)
=1 if either country is landlocked	0.339	0.196	0.152	-0.074
	(1.06)	(0.92)	(0.67)	(-0.75)
=1 if pair shares at least one sea or ocean	-0.215	-0.022	-0.057	0.041
	(-0.43)	(-0.15)	(-0.37)	(0.48)
Observations	8,128	5,671	5,671	5,671
(countries)	(128)	(107)	(107)	(107)
R-squared	0.054	0.091	0.082	0.182
Standardized β on genetic distance (%)	16.88	25.19	24.64	26.06

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

Table 9 – Determinants of Differences in Democracy

(Dependent variable: absolute difference in Polity 2 democracy score at various dates listed in row 2)

	(1)	(2)	(3)	(4)	(5)	(6)
	1960	1970	1980	1990	2000	2005
Relative genetic distance to the USA, weighted, Pemberton et al.	**55.709***	**92.778***	**112.169***	**161.489***	**41.601**	**34.618***
	(1.94)	**(2.89)**	**(3.20)**	**(4.86)**	**(2.11)**	**(1.80)**
Absolute difference in latitudes	2.918*	5.887***	3.624**	-5.159***	-4.506***	-4.061***
	(1.88)	(3.07)	(2.08)	(-2.66)	(-3.30)	(-2.92)
Absolute difference in longitudes	1.895*	3.027**	2.179	-4.332**	-1.871	-1.036
	(1.80)	(2.09)	(1.43)	(-2.18)	(-1.08)	(-0.64)
Geodesic Distance (1000s of km)	-0.277**	-0.426**	-0.300	0.586***	0.271	0.205
	(-2.21)	(-2.41)	(-1.60)	(2.68)	(1.53)	(1.23)
1 for contiguity	-2.102***	-1.388*	-2.366***	-3.818***	-2.382***	-2.675***
	(-3.23)	(-1.81)	(-3.56)	(-5.63)	(-4.05)	(-4.37)
=1 if either country is an island	1.023*	1.742**	1.086*	-0.586**	-1.429***	-1.457***
	(1.67)	(2.11)	(1.70)	(-2.39)	(-5.79)	(-6.27)
=1 if either country is landlocked	0.115	-0.527	-0.580	0.570	-0.342	-0.044
	(0.28)	(-1.32)	(-1.61)	(1.33)	(-0.47)	(-0.07)
=1 if pair shares at least one sea or ocean	-0.666	-0.221	0.171	-0.114	-1.229**	-1.333**
	(-1.59)	(-0.37)	(0.38)	(-0.25)	(-2.29)	(-2.41)
R-squared	0.026	0.064	0.055	0.108	0.034	0.035
Standardized β on genetic distance (%)	9.66	15.63	18.19	28.29	8.07	6.64

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

All regressions run on a sample of 4,278 country pairs from 93 countries.

Table 10 – Technological distance and genetic distance in the contemporary period (1970–2000)
(Dependent variables: measures of technological usage from Comin et al. as described in row 2)

	(1) Agricultural Technology	(2) Communications Technology	(3) Transportation Technology	(4) Industrial Technology	(5) Overall Technology
Relative genetic distance to the USA, weighted, Pemberton et al.	**3.946*****	**4.261*****	**2.422*****	**6.407*****	**4.299*****
	(3.84)	**(4.84)**	**(3.54)**	**(6.29)**	**(5.45)**
Absolute difference in latitudes	0.565***	0.196***	0.243***	0.203***	0.266***
	(5.18)	(3.11)	(3.82)	(2.93)	(3.58)
Absolute difference in longitudes	0.343***	0.080	0.248***	0.112*	0.192**
	(2.99)	(1.44)	(3.17)	(1.85)	(2.39)
Geodesic Distance (1000s of km)	-0.044***	-0.015**	-0.030***	-0.017***	-0.027***
	(-3.46)	(-2.36)	(-3.61)	(-2.68)	(-3.15)
1 for contiguity	-0.032**	-0.063***	-0.055***	-0.080***	-0.062***
	(-2.41)	(-4.73)	(-4.11)	(-4.50)	(-4.88)
=1 if either country is an island	0.136*	0.068**	0.099**	0.076***	0.129**
	(1.74)	(2.45)	(2.06)	(3.44)	(2.57)
=1 if either country is landlocked	-0.005	0.018	-0.004	0.022	-0.013
	(-0.18)	(1.14)	(-0.38)	(1.03)	(-0.96)
=1 if pair shares at least one sea or ocean	0.020	-0.015	0.011	-0.006	0.002
	(0.76)	(-0.94)	(0.62)	(-0.29)	(0.08)
Observations	5,886	7,140	6,216	5,356	7,260
(countries)	(109)	(120)	(112)	(104)	(121)
R-squared	0.248	0.136	0.116	0.192	0.185
Standardized β on genetic distance (%)	20.46	26.38	18.46	35.47	28.74

t-statistics based on two-way clustered standard errors, in parentheses; *** p<0.01, ** p<0.05, * p<0.1;

All regressions include an intercept (estimates not reported).

Figure 1 - Standardized Beta on genetic distance (%), common sample
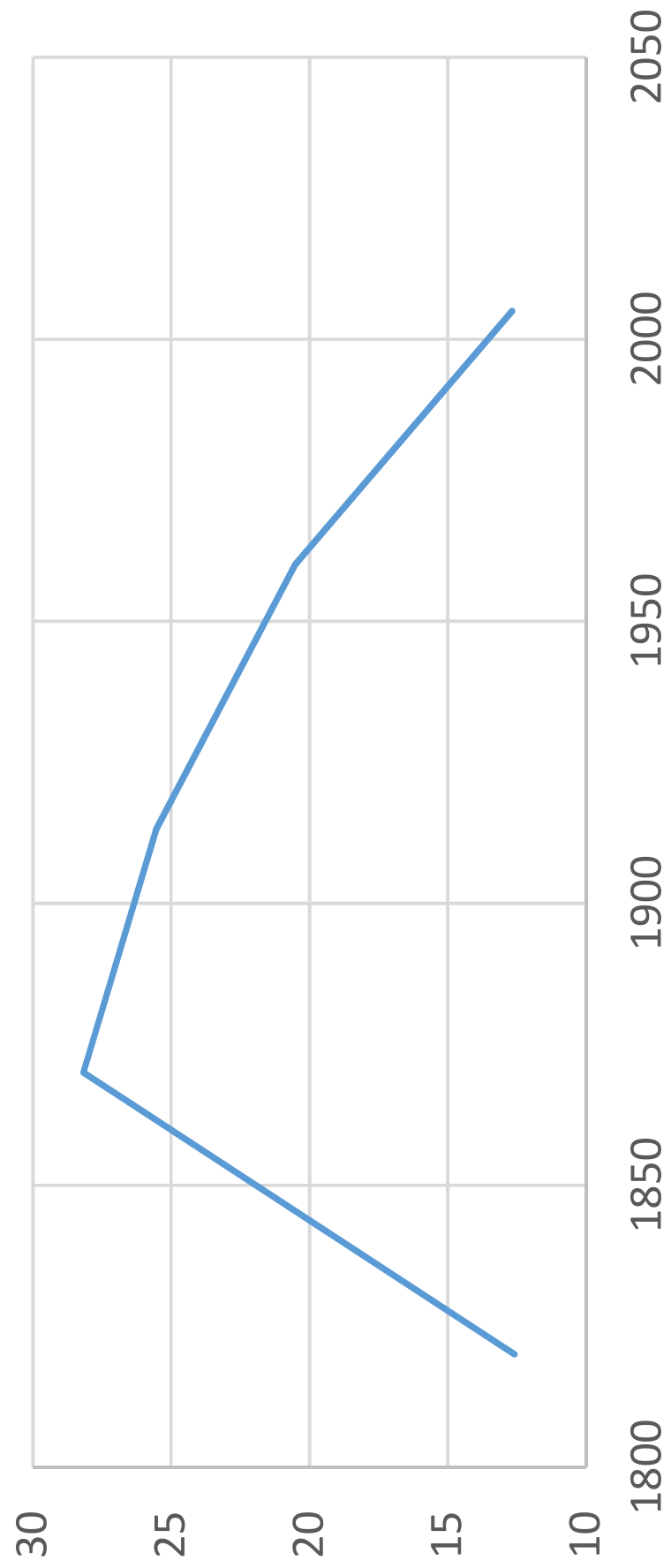
Online Appendix

Ancestry and Development: New Evidence

Enrico Spolaore and Romain Wacziarg

March 2018

<u>**NOTE**</u>: Tables A2-A10 in this Appendix are numbered to correspond to the numbering of tables in the main text.   So for instance Table A2 is to be compared to Table 2 in the main text of the paper.

Table A1 - Regressions using Historical Data, using the USA as the frontier instead of the UK

(Dependent variable: absolute difference in log per capita income, various dates as in row 2)

	(1) Income 1820	(2) Income 1870	(3) Income 1913	(4) Income 1960	(5) Income 2005
Relative genetic distance to the USA, weighted, Pemberton et al.	**3.512** **(2.01)****	**12.036** **(4.00)*****	**10.991** **(3.11)*****	**17.186** **(4.95)*****	**18.525** **(5.10)*****
Absolute difference in latitudes	1.111 (5.44)***	1.270 (4.91)***	1.199 (4.91)***	0.976 (4.15)***	-0.495 (2.17)**
Absolute difference in longitudes	0.633 (4.19)***	0.716 (2.30)**	0.854 (2.96)***	0.645 (2.61)***	-0.624 (2.81)***
Geodesic Distance (1000s of km)	-0.082 (3.89)***	-0.096 (2.59)***	-0.102 (3.21)***	-0.067 (2.39)**	0.056 (2.11)**
1 for contiguity	-0.203 (4.09)***	-0.213 (4.62)***	-0.235 (5.40)***	-0.164 (2.82)***	-0.522 (8.85)***
=1 if either country is an island	-0.106 (3.47)***	0.024 (0.25)	0.024 (0.27)	-0.038 (0.77)	-0.039 (0.58)
=1 if either country is landlocked	0.129 (3.77)***	0.173 (2.56)**	0.222 (3.07)***	0.279 (3.33)***	0.152 (1.79)*
=1 if pair shares at least one sea or ocean	0.010 (0.29)	0.092 (2.03)**	0.131 (2.91)***	0.107 (1.96)*	0.006 (0.08)
R^2	0.27	0.22	0.19	0.18	0.07
N	1,081	1,540	1,711	5,460	14,365
Standardized β on genetic distance (%)	11.492	27.125	21.976	28.955	23.470
Standardized β on genetic distance (%) for a common sample [a]	11.948	26.856	24.154	19.333	11.978

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

[a]: the common sample is composed of 820 pairs (41 countries).

All regressions include an intercept (estimates not reported)

All columns include controls for: absolute difference in latitudes, absolute difference in longitudes, geodesic distance, dummy for contiguity, dummy for either country being an island, dummy for either country being landlocked, dummy = 1 if pair shares at least one sea or ocean.

Table A2 – Income Level Regressions, controlling for geographic distance, using Cavalli-Sforza et al. data
(Dependent variable: log income per capita 2005)

	(1) Univariate	(2) Distance and geography controls	(3) Add tropics control	(4) IV using 1500 genetic distance
F_{ST} genetic distance to the USA, weighted, Cavalli-Sforza et al. (1994)	-12.646 (8.90)***	-6.330 (2.96)***	-6.079 (2.79)***	-7.503 (2.68)***
Absolute latitude		0.032 (4.42)***	0.041 (4.35)***	0.039 (3.91)***
Landlocked dummy		-0.567 (3.05)***	-0.501 (2.68)***	-0.460 (2.37)**
Island dummy		0.629 (3.11)***	0.434 (1.49)	0.440 (1.50)
Geodesic distance to the USA		0.294 (0.42)	1.297 (1.77)*	1.169 (1.52)
Absolute difference in latitude to the USA		0.161 (0.19)	-0.287 (0.35)	-0.013 (0.01)
Absolute difference in longitude to the USA		-0.485 (0.92)	-1.130 (2.10)**	-0.987 (1.67)*
Dummy for common sea/ocean with the USA		-0.190 (1.02)	-0.053 (0.28)	-0.028 (0.14)
Dummy for contiguity to the USA		0.528 (1.72)*	0.388 (1.09)	0.430 (1.23)
% land area in the tropics			0.069 (0.22)	0.081 (0.26)
# of observations	171	171	150	150
Adjusted R^2	0.30	0.44	0.48	0.48
Standardized β on genetic distance (%)	55.027	27.546	26.774	33.043

Robust t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.
All regressions include an intercept (estimates not reported).

Table A3 – Robustness of the Effect of Genetic Distance to Continental Effects, using the Cavalli-Sforza et al. data
(Dependent variable: log income per capita 2005)

	(1) All Continent dummies	(2) All Macro-region dummies	(3) Removing SS African countries	(3) Removing European countries	(5) Removing European and SS African countries
F_{ST} genetic distance to the USA, weighted, Cavalli-Sforza et al.	-4.565 (1.90)*	-4.815 (1.71)*	-8.121 (3.00)***	-5.203 (2.18)**	-5.042 (1.28)
# of observations	171	171	126	135	90
Adjusted R^2	0.49	0.56	0.28	0.32	0.18
Standardized β on gen. dist. (%)	19.864	20.951	30.585	22.315	18.252

Robust t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

Continent dummies in column (1) are for Europe, Africa, Latin America & Caribbean, Asia, North America, and Oceania (omitted category).

Macro-region dummies in column (2) are for Sub-Saharan Africa, North America, Western & Southern Europe, Eastern & Central Europe, Latin America (excl. Caribbean), Caribbean, North Africa, East & Southeast Asia, South Asia, Southwestern & Central Asia, Middle East, Oceania (omitted category).

All regressions include additional controls for absolute latitude, landlocked dummy, island dummy, geodesic distance to the USA, absolute difference in latitude to the USA, absolute difference in longitude to the USA, dummy for common sea/ocean with the USA, dummy for contiguity to the USA.

Table A4 – Genetic Distance, Genetic Diversity, Language and Religion, using Cavalli-Sforza et al. (1994) data
(Dependent variable: log income per capita 2005)

	(1) Gen. div. controls	(2) Gen. div. & dist. & geo. controls	(3) Add linguistic and religious distance
F_{ST} **genetic distance to the USA, weighted, Cavalli-Sforza et al. (1994)**	**-11.911** **(8.09)***	**-6.966** **(2.64)***	**-5.345** **(2.04)****
Predicted genetic diversity, Ashraf-Galor (2013)	115.816 (2.53)**	104.750 (1.68)*	108.449 (1.74)*
Predicted genetic diversity squared, Ashraf-Galor (2013)	-87.604 (2.58)**	-77.032 (1.70)*	-80.110 (1.77)*
Linguistic distance to the USA, Fearon measure, weighted			-0.269 (0.48)
Religious distance to the USA, Mecham-Fearon-Laitin, weighted			-1.697 (2.33)**
# of observations	169	148	140
Adjusted R^2	0.33	0.48	0.57
Standardized β on genetic distance (%)	51.998	30.827	24.088

Robust t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

Columns (2) and (3) include controls for absolute latitude, landlocked dummy, island dummy, geodesic distance to the USA, absolute difference in latitude to the USA, absolute difference in longitude to the USA, dummy for common sea/ocean with the USA, dummy for contiguity to the USA, and percentage of land area in the tropics.

Table A5 – Income difference regressions, using genetic distance from Cavalli-Sforza et al. (1994).
(Dependent variable: difference in log per capita income, 2005)

	(1) Relative GD	(2) Simple GD	(3) Horserace between simple and relative GD	(4) 2SLS with 1500 GD
Relative F_{ST} genetic distance to the USA, weighted, Cavalli-Sforza et al.	5.094 (4.82)***		4.902 (4.37)***	9.113 (5.21)***
Weighted Fst Genetic Distance, Cavalli-Sforza et al.		2.042 (3.23)***	0.235 (0.40)	
Absolute difference in latitudes	-0.248 (1.10)	0.112 (0.45)	-0.226 (1.02)	-0.444 (1.81)*
Absolute difference in longitudes	-0.353 (1.75)*	-0.414 (2.03)**	-0.337 (1.77)*	-0.120 (0.56)
Geodesic Distance (1000s of km)	0.031 (1.26)	0.024 (0.96)	0.028 (1.21)	0.007 (0.26)
1 for contiguity	-0.480 (8.19)***	-0.541 (9.52)***	-0.478 (8.23)***	-0.381 (6.16)***
=1 if either country is an island	0.022 (0.37)	-0.003 (0.06)	0.021 (0.35)	0.043 (0.71)
=1 if either country is landlocked	0.115 (1.35)	0.124 (1.40)	0.113 (1.32)	0.087 (1.06)
=1 if pair shares at least one sea or ocean	-0.011 (0.18)	0.014 (0.20)	-0.010 (0.15)	-0.026 (0.41)
R^2	0.08	0.05	0.08	0.08
Standardized β on genetic distance (%)	23.475	14.378	22.588	41.345

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

All regressions are based on 14,365 country pair observations from 170 countries.

6

Table A6 – Income difference regressions, robustness and extensions, using genetic distance data from Cavalli-Sforza et al. (1994)
(Dependent variable: absolute difference in log per capita income, 2005)

	(1) Continent dummies	(2) Excl. New World	(3) Excl. Sub-Saharan Africa	(4) Climatic Difference	(5) Common history controls
Relative F_{ST} genetic distance to the USA, weighted, Cavalli-Sforza et al.	3.035 (2.73)***	4.226 (3.07)***	4.320 (4.07)***	5.440 (4.81)***	5.458 (4.61)***
Measure of climatic difference of land areas, by 12 KG zones				0.032 (5.00)***	
1 if countries were or are the same country					-0.361 (3.93)***
1 for pairs ever in colonial relationship					0.211 (1.94)*
1 for common colonizer post 1945					-0.101 (1.35)
1 for pairs currently in colonial relationship					-0.938 (4.91)***
Religious distance index, relative to USA, weighted					0.886 (3.60)***
Linguistic distance index, relative to USA, weighted					0.204 (1.08)
R^2	0.14	0.08	0.05	0.11	0.12
Observations (countries)	14,365 (170)	8,256 (129)	7,750 (125)	11,026 (149)	10,296 (144)
Standardized β on genetic distance (%)	13.987	19.897	17.082	25.781	26.007

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

All columns include controls for: absolute difference in latitudes, absolute difference in longitudes, geodesic distance, dummy for contiguity, dummy for either country being an island, dummy for either country being landlocked, dummy = 1 if pair shares at least one sea or ocean.

Column 1 includes a full set of continental dummy variables: both in Asia Dummy, both in Africa Dummy, both in Europe Dummy, both in Latin America/Caribbean dummy, Both in Oceania Dummy, Dummy if one and only one country is in Asia, Dummy if one and only one country is in Africa, dummy if one and only one country is in Europe, dummy if one and only one country is in North America, dummy if one and only one country is in South America.

Table A7 – Regressions using Historical Data, using genetic distance from Cavalli-Sforza et al. (1994)
(Dependent variable: absolute difference in log per capita income, various dates as in row 2)

	(1)	(2)	(3)	(4)	(5)
	Income 1820	Income 1870	Income 1913	Income 1960	Income 2005
Relative F_{ST} genetic distance to the UK, weighted	**0.622** **(1.76)***	**1.663** **(2.07)****	**1.705** **(2.02)****	**2.569** **(3.76)******	**3.973** **(4.70)******
R^2	0.26	0.17	0.16	0.16	0.07
Observations	1,081	1,540	1,711	5,460	14,365
(countries)	(47)	(56)	(59)	(105)	(170)
Standardized β on genetic distance (%)	7.978	14.521	13.079	21.975	22.840
Standardized β on genetic distance (%) for a common sample [a]	9.150	14.781	13.591	7.493	3.935

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

[a]: the common sample is composed of 820 pairs (41 countries).

All regressions include an intercept (estimates not reported).

All columns include controls for: absolute difference in latitudes, absolute difference in longitudes, geodesic distance, dummy for contiguity, dummy for either country being an island, dummy for either country being landlocked, dummy = 1 if pair shares at least one sea or ocean.

Table A8 – Determinants of Institutional Differences, using Cavalli-Sforza et al. data
(Dependent variables: measures of institutional quality for 1990, as in row 2)

	(1) Polity score 1990	(2) Repudiation of contracts 1990	(3) Risk of expropriation 1990	(4) Rule of law 1990
Relative F_{ST} genetic distance to the USA, weighted, Cavalli-Sforza et al.	**16.943***	**9.100***	**6.504***	**4.271***
	(2.70)	**(3.83)**	**(2.70)**	**(2.80)**
Absolute difference in latitudes	-1.243	-0.108	-0.300	2.247***
	(-0.67)	(-0.19)	(-0.73)	(5.09)
Absolute difference in longitudes	-2.224	-1.034*	-0.953**	0.581
	(-1.26)	(-1.73)	(-2.03)	(1.56)
Geodesic Distance (1000s of km)	0.389**	0.102	0.076	-0.084*
	(1.97)	(1.48)	(1.42)	(-1.88)
1 for contiguity	-2.941***	-0.555**	-0.482**	-0.192*
	(-4.90)	(-2.56)	(-2.08)	(-1.88)
=1 if either country is an island	0.060	0.127	-0.035	0.160*
	(0.15)	(0.80)	(-0.33)	(1.69)
=1 if either country is landlocked	0.057	0.110	0.092	-0.113
	(0.19)	(0.53)	(0.39)	(-0.97)
=1 if pair shares at least one sea or ocean	-0.333	-0.084	-0.106	0.007
	(-0.69)	(-0.62)	(-0.72)	(0.08)
Observations	8,128	5,671	5,671	5,671
R-squared	0.042	0.076	0.056	0.147
Standardized β on genetic distance (%)	12.18	21.59	17.49	16.31

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

Table A9 – Genetic Distance and Differences in Democracy, using the Cavalli-Sforza et al. data
(Dependent variable: absolute difference in Polity 2 democracy score at various dates listed in row 2)

	(1) 1960	(2) 1970	(3) 1980	(4) 1990	(5) 2000	(6) 2005
Relative F_{ST} genetic distance to the USA, weighted, Cavalli-Sforza et al.	**6.143**	**15.004****	**22.748*****	**37.383*****	**15.255****	**13.368****
	(0.96)	**(2.08)**	**(2.74)**	**(4.26)**	**(2.25)**	**(2.12)**
Absolute difference in latitudes	3.576**	6.911***	4.793**	-3.545*	-4.174***	-3.795***
	(2.24)	(3.46)	(2.55)	(-1.82)	(-3.12)	(-2.80)
Absolute difference in longitudes	1.925*	3.264**	2.645*	-3.482*	-1.433	-0.645
	(1.79)	(2.22)	(1.68)	(-1.81)	(-0.86)	(-0.41)
Geodesic Distance (1000s of km)	-0.289**	-0.462**	-0.361*	0.482**	0.225	0.164
	(-2.24)	(-2.54)	(-1.85)	(2.24)	(1.31)	(1.01)
1 for contiguity	-2.354***	-1.675**	-2.585***	-4.004***	-2.274***	-2.567***
	(-3.48)	(-2.08)	(-3.66)	(-6.10)	(-3.90)	(-4.23)
=1 if either country is an island	0.990	1.706**	1.061	-0.602***	-1.410***	-1.439***
	(1.60)	(2.04)	(1.62)	(-2.73)	(-5.65)	(-6.06)
=1 if either country is landlocked	0.119	-0.526	-0.584	0.559	-0.352	-0.053
	(0.29)	(-1.28)	(-1.54)	(1.31)	(-0.48)	(-0.09)
=1 if pair shares at least one sea or ocean	-0.739*	-0.352	0.005	-0.363	-1.304**	-1.397***
	(-1.78)	(-0.60)	(0.01)	(-0.80)	(-2.49)	(-2.60)
R-squared	0.019	0.049	0.041	0.085	0.038	0.039
Standardized β on genetic distance (%)	3.739	8.878	12.952	22.996	10.389	9.002

t-statistics based on two-way clustered standard errors, in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

All regressions include an intercept (estimates not reported).

All regressions run on a sample of 4,278 country pairs from 93 countries.

Table A10 – Genetic distance and technological distance (1970–2000), using Cavalli-Sforza et al. data
(Dependent variables: measures of technological usage from Comin et al. as described in row 2)

	(1) Agricultural Technology	(2) Communications Technology	(3) Transportation Technology	(4) Industrial Technology	(5) Overall Technology
Relative F_{ST} genetic distance to the USA, weighted, Cavalli-Sforza et al.	**0.392**	**0.465****	**0.573*****	**1.063*****	**0.706*****
	(1.35)	**(2.12)**	**(2.82)**	**(3.55)**	**(3.05)**
Absolute difference in latitudes	0.613***	0.249***	0.270***	0.269***	0.315***
	(5.57)	(3.81)	(4.39)	(3.75)	(4.14)
Absolute difference in longitudes	0.328***	0.066	0.267***	0.118**	0.196**
	(2.73)	(1.17)	(3.25)	(2.02)	(2.27)
Geodesic Distance (1000s of km)	-0.043***	-0.013**	-0.032***	-0.019***	-0.028***
	(-3.26)	(-2.10)	(-3.64)	(-2.98)	(-2.97)
1 for contiguity	-0.047***	-0.077***	-0.054***	-0.093***	-0.071***
	(-3.43)	(-5.94)	(-3.97)	(-4.99)	(-5.69)
=1 if either country is an island	0.131*	0.061**	0.098**	0.072***	0.124**
	(1.65)	(2.21)	(2.05)	(3.00)	(2.48)
=1 if either country is landlocked	-0.010	0.017	-0.009	0.015	-0.017
	(-0.34)	(0.98)	(-0.74)	(0.63)	(-1.19)
=1 if pair shares at least one sea or ocean	0.018	-0.016	0.008	-0.011	-0.001
	(0.68)	(-0.97)	(0.50)	(-0.60)	(-0.07)
Observations	5,886	7,140	6,216	5,356	7,260
R-squared	0.220	0.093	0.113	0.138	0.149
Standardized β on genetic distance (%)	8.360	11.73	17.63	24.26	18.88

t-statistics based on two-way clustered standard errors, in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
All regressions include an intercept (estimates not reported).

Figure A1 - Standardized Beta on genetic distance (%), common sample, Cavalli-Sforza et al. (1994) data