Ancestry and Development: New Evidence

Enrico Spolaore and Romain Wacziarg

PWP-CCPR-2017-004

April 6, 2017

California Center for Population Research
On-Line Working Paper Series

Ancestry and Development: New Evidence*

Enrico Spolaore
Tufts University and NBER

Romain Wacziarg
UCLA and NBER

June 2016

Abstract

We revisit the relationship between ancestral distance and barriers to the diffusion of development using a new genomic dataset on human microsatellite variation. With this new data we confirm past findings of 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. These findings are consistent with the hypothesis that ancestral distance acts as a temporary barrier to the diffusion of innovations and development.

---

*Spolaore: Department of Economics, Tufts University, Medford, MA 02155-6722, enrico.spolaore@tufts.edu. Wacziarg: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles CA 90095, wacziarg@ucla.edu.
We thank Omer Ali for outstanding research assistance, Trevor Pemberton for making the new genetic distance data available and Stefan Legge for useful comments.

1 Introduction

Does ancestry affect economic development? A recent literature in economics has found that the characteristics of a society's ancestral population exert a strong 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 own work, we argued 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).

This paper revisits the relationship between ancestral distance and the diffusion of development using new information on human microsatellite variation (Pemberton et al., 2013). This new 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). In turn, these measures of relatedness can be used to reassess and extend previous results on the determinants of development, shedding more light on the effects of long-term historical barriers on the spread of modern development.

The new results presented here confirm and strengthen our previous conclusions. We find 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 populations (as emphasized in 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 and institutions, rather than as a permanent obstacle to economic development.[1]

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

[1] Evidence on ancestral and cultural distance acting as temporary barriers to the spread of new social norms and behavior regarding fertility is provided in Spolaore and Wacziarg (2016a).

2 New Data on 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. Microsatellites 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 which is not subject to natural selection. Consequently, it is important to notice 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 data from Pemberton et al. (2013) differ 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

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

that are covered. In particular, an important advantage of the new data set 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 $i = 1, ..., I$ 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

[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. (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.

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 data set 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 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 correlation 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 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

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 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%. In column (2) we add several controls for geographical features (absolute latitude, landlocked dummy, island dummy) as well as for geographical barriers with the US (geodesic distance from the US and absolute differences in latitude and longitude to the US). 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 columns (3), we restrict the sample to countries outside of Sub-Saharan Africa to address a possible 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 bit in the sample that excludes Sub-Saharan Africa.[8] Finally in column (4) we add a control for 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-A6. Additionally, the new genetic distance data used here is available on the authors' websites.

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

[8] Table A1 in the Appendix conducts a more systematic analysis of regional effects. We find that the results are robust to the inclusion of a broad range of regional dummies, including dummies for Sub-Saharan Africa and Europe (entered either individually or jointly), and a full set of 6 continental dummies (Oceania being the excluded category). The results are also robust to the exclusion of European countries and the exclusion of both Sub-Saharan Africa and Europe.

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 (0.632), 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 beta equal to 61.9%.

In Table 3, building on a recent 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 from having a heterogeneous population, whereby heterogeneity is beneficial for development at lower levels but detrimental above a critical threshold. In Table 3 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 beta for genetic distance equal to 60.4%. In columns (2) and (3) we add geographical controls. Table 3 column (2) includes the same geographical controls used in Table 2, while in column (3) we add a dummy for Sub-Saharan Africa and the percentage of land in the tropics. The effects of the ancestral variables (genetic distance and genetic diversity) remain statistically significant, and the standardized beta on genetic distance equals 51.3% in column (2) and 34.7% in column (3). Finally, in column (4) we control for measures of cultural distance to the US, namely linguistic and religious distance.[9] We expect such measures to reduce the effect of

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

genetic distance, as language and religion form part of the intergenerationally generated 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]

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 the interpretation of the effect in terms of temporary barriers to the horizontal diffusion of modern economic development across historically and culturally distant societies.

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:

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

and Mecham, Fearon and Laitin (2006).

[10] The standardized β falls from 34.8% in column (3) to 23.6% in column (4), while the sample falls from 148 to 140 countries. The change in the sample is responsible for a 5.4 percentage point decline in the standardized β while the addition of linguistic and religious distance is responsible for a 5.8 percentage point decline - about 17% of the total effect.

where the diffusion framework predicts $\beta_1 = 0$ and $\beta_2 > 0$.[11] The baseline results are presented in Table 4. 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 largest 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. This is exactly as the model 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 5. 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 4, 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 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 and linguistic similarity. We expect, as before, the inclusion of these variables to 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. In sum, both the baseline results and the main robustness tests in Spolaore and Wacziarg (2009, 2012, 2013, 2014) carry over unchanged when using the new dataset of genetic distance.

[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).

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. We explore this hypothesis using the diffusion of the Industrial Revolution from England, starting in the first half of the 19^{th} century. 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. Table 6 provides strong evidence supportive of just such a pattern. The frontier is now defined as the United Kingdom, and we use data from Maddison 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 technological and institutional frontier.

4 Conclusion

In this paper, we used novel measures of ancestral distance between human societies to shed light on the diffusion of economic development.

First, we found that countries at a higher ancestral distance from the technological frontier (the United States) had a lower income per capita in 2005. The effect was robust to controlling for geographical barriers, climatic differences, a dummy for Sub-Saharan Africa, measures of linguistic and religious distance, and the effect of genetic diversity within populations (a variable emphasized in Ashraf and Galor, 2013).

Second, the effect of relative ancestral distance from the technological frontier had a statistically and economically significant effect on income differences, and dominated the effect of absolute ancestral distance in a horserace between the two variables. This is consistent with the hypothesis 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. – which facilitate 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 found that the historical pattern of the impact of relative ancestral distance from the frontier on income per capita is humped shaped, peaking between 1870 and 1913, and declining steeply afterwards. These results show that the effects of long-term divergence in inherited traits – captured by ancestral distance – are important but not fixed and immutable. The effects depend on dynamic factors, such as the location of the frontier and the gradual spread of innovations, and thus they change (and decline) over time.

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 study of such policies is an important topic for future research.

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth*, 8(2), 155-194.

Ashraf, Quamrul and Oded Galor. 2013. "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103(1), 1-46.

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

Cameron, Colin, Jonah Gelbach, and Douglas Miller. 2011. "Robust Inference with Multi-Way Clustering." *Journal of Business and Economic Statistics*, 29 (2), pp. 238-249.

Cann, Howard M., Claudia de Toma, Lucien Cazes, Marie-Fernande Legrand, Valerie Morel, Laurence Piouffre, Julia Bodmer et al. 2002. "A Human Genome Diversity Cell Line Panel." *Science* 296 (5566): 261-62

Cavalli-Sforza, Luigi Luca, Paolo Menozzi and Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Fearon, James. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8, pp. 195-222.

Mecham, Quinn, James Fearon and David Laitin. 2006. "Religious Classification and Data on Shares of Major World Religions." Unpublished, Stanford University.

Richerson, Peter J., and Robert Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Pemberton, Trevor J., Michael DeGiorgio, and Noah A. Rosenberg. 2013. "Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation." *G3-Genes/Genomes/Genetics,* 3: 903-919.

Putterman, Louis and David N. Weil. 2010. "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics* 125 (4): 1627-1682.

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

Spolaore, Enrico and Romain Wacziarg. 2012. "Long-Term Barriers to the International Diffusion of Innovations." Chapter 1 in Jeffrey Frankel and Christopher Pissarides (eds), *NBER International Seminar On Macroeconomics 2011*, Cambridge (MA): NBER.

Spolaore, Enrico and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51 (2): 1-45.

Spolaore, Enrico and Romain Wacziarg. 2014. "Long-Term Barriers to Economic Development" in Philippe Aghion and Steven Durlauf (eds.), *Handbook of Economic Growth*, vol. 2A, Chapter 3, pp. 121-176. Amsterdam: North Holland.

Spolaore, Enrico and Romain Wacziarg. 2016a. "Fertility and Modernity," UCLA and Tufts University. First draft: August 2014.

Spolaore, Enrico and Romain Wacziarg. 2016b. "Ancestry, Language and Culture." Chapter 7 in Victor Ginsburgh and Shlomo Weber (eds.), *The Palgrave Handbook of Economics and Language*, London: Palgrave Macmillan.

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)	(2)	(3)	(4)	(5)
	Univariate	**Dist. & geo. controls**	**Without Sub-Saharan African countries**	**Add tropics control**	**IV using 1500 gen. dist.**
F_{ST} genetic distance to the USA, weighted, Pemberton et al.	**-43.594 (9.12)*** **	**-35.610 (5.67)*** **	**-33.081 (4.51)*** **	**-37.720 (5.16)*** **	**-53.372 (3.67)*** **
Absolute latitude		0.025 (3.75)***	0.013 (1.46)	0.030 (3.10)***	0.021 (1.56)
Landlocked dummy		-0.549 (3.14)***	-0.596 (2.73)***	-0.477 (2.67)***	-0.395 (2.11)**
Island dummy		0.750 (3.70)***	0.787 (3.80)***	0.486 (1.78)*	0.519 (1.84)*
Geodesic distance to the USA		0.812 (1.29)	-0.137 (0.18)	1.317 (1.84)*	1.098 (1.31)
Absolute difference in latitude to the USA		-0.167 (0.22)	0.473 (0.52)	-0.211 (0.27)	0.305 (0.30)
Absolute difference in longitude to the USA		-0.967 (2.11)**	-0.250 (0.44)	-1.213 (2.37)**	-0.994 (1.58)
Dummy for common sea/ocean with the USA		-0.161 (0.91)	-0.374 (1.97)*	0.011 (0.05)	0.082 (0.39)
Dummy for contiguity to the USA		0.575 (1.82)*	0.845 (2.20)**	0.551 (1.60)	0.693 (1.84)*
% land area in the tropics				-0.008 (0.02)	-0.018 (0.05)
Constant	10.171 (67.41)***	9.646 (25.34)***	9.988 (24.31)***	9.336 (17.56)***	9.825 (14.37)***
# of observations	171	171	126	150	150
Adjusted R^2	0.29	0.48	0.32	0.52	0.51
Standardized β on genetic distance (%)	54.492	44.512	48.465	43.772	61.935

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

Table 3: Genetic Distance, Genetic Diversity, Language and Religion
(Dependent variable: log income per capita 2005)

	(1)	(2)	(3)	(4)
	Gen. div. controls	**Gen. div. & dist. & geo. controls**	**Gen. div. & all dist. & geo. controls**	**Add linguistic and religious distance**
F_{ST} genetic distance to the USA, weighted, Pemberton et al.	**-48.289**	**-40.949**	**-48.962**	**-39.482**
	(10.90)* **	**(5.88)*** **	**(5.29)*** **	**(4.35)*** **
Predicted genetic diversity, Ashraf-Galor (2013)	156.853	163.694	199.496	176.082
	(3.69)***	(2.83)***	(3.16)***	(2.86)***
Predicted genetic diversity squared, Ashraf-Galor (2013)	-120.518	-122.318	-146.501	-129.630
	(3.83)***	(2.87)***	(3.16)***	(2.88)***
% land area in the tropics			-0.305	-0.472
			(0.94)	(1.51)
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)
Constant	-40.004	-44.035	-56.764	-48.522
	(2.82)***	(2.29)**	(2.71)***	(2.35)**
# of observations	169	169	148	140
Adjusted R^2	0.42	0.51	0.54	0.60
Standardized β on genetic distance (%)	60.438	51.252	56.908	46.874

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

Table 4: Income difference regressions
(Dependent variable: absolute difference in log per capita income, 2005)

	(1)	(2)	(3)	(4)
	Relative GD	**Simple GD**	**Horserace between simple and relative GD**	**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)
Constant	1.142 (13.181)***	1.135 (12.209)***	1.127 (12.384)***	1.159 (13.414)***
R^2	0.07	0.05	0.07	0.04
Standardized Beta on genetic distance (%)	23.470	15.914	22.255	20.450

t-statistics based on two-way clustered standard errors, in parentheses; * p<0.1; ** p<0.05; *** p<0.01.
All regressions are based on 14,365 country pair observations from 170 countries.

Table 5 - Income difference regressions, robustness and extensions
(Dependent variable: absolute difference in log per capita income, 2005)

	(1)	(2)	(3)	(4)	(5)
	Continent dummies	**Excl. New World**	**Excl. Sub-Saharan Africa**	**Climatic Difference**	**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)*
Constant	1.661 (7.036)***	1.046 (10.462)***	0.966 (13.957)***	0.711 (6.127)***	0.940 (10.661)***
R^2	0.15	0.12	0.04	0.14	0.16
Observations (countries)	14,365 (170)	8,256 (129)	7,750 (125)	11,026 (149)	10,296 (144)
Standardized Beta (%)	16.768	33.695	15.081	31.473	33.621

t-statistics based on two-way clustered standard errors, in parentheses; * p<0.1; ** p<0.05; *** p<0.01.
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 6 - Regressions using Historical Data
(Dependent variable: absolute difference in log per capita income, various dates as in row 1)

	(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 (%)	12.312	27.837	22.781	29.999	23.967
Standardized β on genetic distance (%) for a common sample (a)	12.590	28.175	25.540	20.514	12.668

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 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.

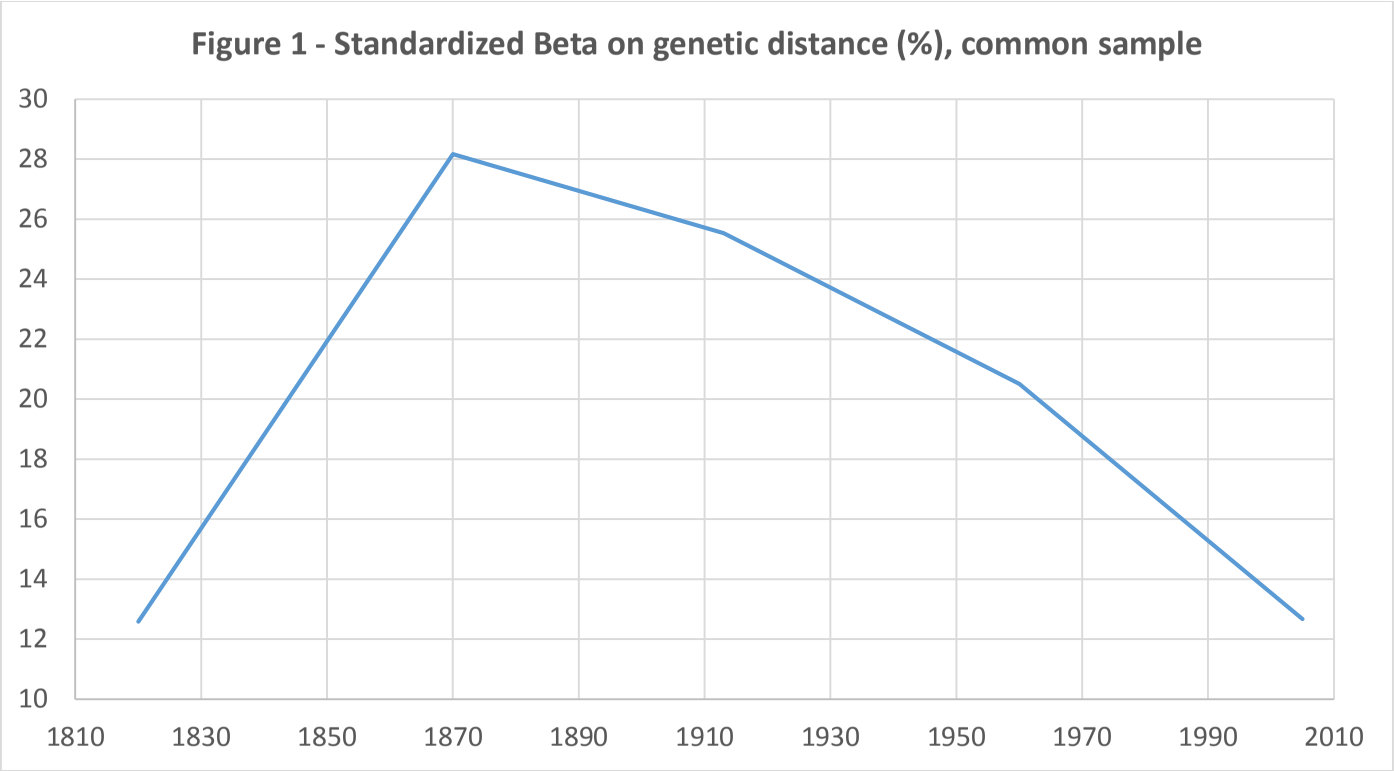
Figure 1 - Standardized Beta on genetic distance (%), common sample
30
28
26
24
22
20
18
16
14
12
10
1810
1830
1850
1870
1890
1910
1930
1950
1970
1990
2010