

SIMON FRASER UNIVERSITY

Department of Economics

Working Papers

13-11

“Colonization and Genetics
of Comparative
Development”

Ideen Riahi

Oct, 2013



Colonization and Genetics of Comparative Development*

Ideen Riahi †

Job Market Paper

October 2013

Abstract

This paper examines the hypothesis that genealogical closeness to populations that are at the frontier of technological innovations facilitates the diffusion of development (e.g., Spolaore and Wacziarg, 2009). Proponents of this hypothesis argue that it is supported by a strong negative cross-country correlation between genetic distance to the United States and income per capita. This paper argues that there is no causal relationship between these two variables. This is because there are important variables, such as European settler mortality at the time of colonization that simultaneously affect countries' genetic distance to the technological frontier and their current economic performance. Controlling for confounding factors removes the statistical association between genetic distance and economic performance. IV estimates also confirm the lack of significant statistical relationship between genetic distance and economic performance. Further, fixed effect estimates show that those countries that became genetically closer to the world technological frontier over the past 500 years did not become richer because of it.

JEL classification: O10, O11, O50, P51, Z10.

Keywords: Genetic Distance, Comparative Development, Colonization, Institutions.

*I am thankful to Fernando Aragon, Greg Dow, Alexander Karaivanov, Chris Muris, Nikolaus Robalino, James Robinson, Arthur Robson, and Simon Woodcock, and the participants of Research in Economic Development Meeting (RECODE)-University of Ottawa, 2013, for comments and conversations.

†PhD candidate, Department of Economics, Simon Fraser University, Vancouver, Canada, email: iar2@sfu.ca.

1 Introduction

The broad message emerging after more than a decade of empirical research on the fundamental causes of economic development is that the history of populations matters a great deal for their current standards of living.¹ There is, however, considerable disagreement among scholars regarding which parts of history are relevant for contemporary comparative development and why. While some maintain that history matters because societal structures and institutions are highly persistent (e.g. Acemoglu et al., 2001; Acemoglu and Robinson, 2012; Diamond, 1997), others believe that the relevance of history is mostly due to the fact that social norms, beliefs, customs, etc. - that might be conducive or unfavourable for development - are transferred, biologically and/or culturally, across generations (e.g. Ashraf and Galor, 2013; Demset et al., 2007; Spolaore and Wacziarg, 2009).

Among the latter group, Spolaore and Wacziarg (2009) - hereafter SW - provide what appears to be strong empirical support for the effect of culture and/or biology on contemporary economic development. They document a negative cross-country correlation between *genetic distance to the United States* and contemporary income per capita, depicted in Figure 1-1, and show that this relationship remains robust to controlling for various measures of geographical distance, climate differences, transportation costs, and measures of historical, linguistic, and religious distance.² The robustness of their results prompts them to suggest that countries' genetic distance to the United States, as proxy of their cultural and biological divergence from the world technological frontier, is a barrier to the diffusion of development.³

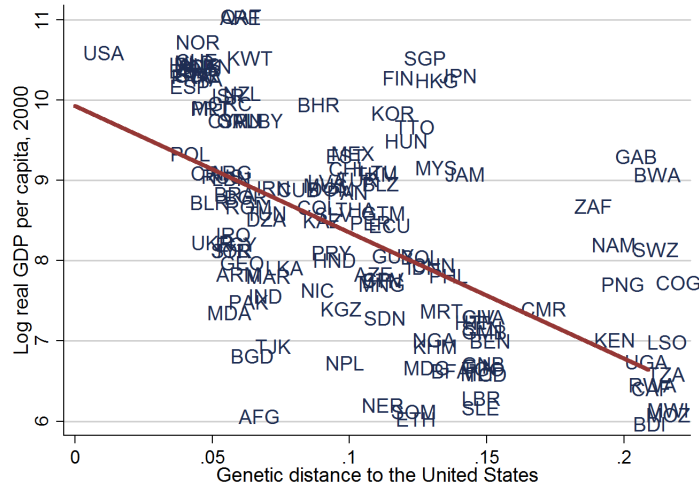
This paper re-examines the relationship between genetic distance to the technological frontier and economic performance. I show that genetic distance to the United States is highly correlated with other historical determinants of comparative development that have been suggested in the literature, e.g., the initial biogeographical conditions (Diamond, 1997; Olsson and Hibbs, 2005), and the hostility of the disease environment of former colonies to Europeans (Acemoglu et al., 2001). I show that, after controlling for confounding factors, genetic distance to the United States has no robust statistically significant power to explain contemporary comparative development.

¹See for example, Acemoglu, Johnson, and Robinson (2001, 2002), Acemoglu and Robinson (2012), Ashraf and Galor (2013), Coming, Easterly, and Gong (2010), Bockstette, Chanda, and Putterman (2002), Diamond (1997), Putterman and Weil (2012), Spolaore and Wacziarg (2013).

²Note that the spectrum of traits which are the base of calculation of genetic distance refer only to "random drifts" in the distribution of genes, and not to the divergence in genetic characteristics that occurs due to natural selection. As such, genetic distance measures the elapsed time since two currently separated populations shared a common ancestor. Larger values of genetic distance refers to a longer period of separation between populations, or, equivalently, less genealogical relatedness between them. See section 4 for more technical definition of genetic distance to the United States.

³SW's key hypothesis is that this "long-term (and *mainly random*) [cultural and biological] divergence has created barriers to the diffusion of technological and institutional innovations across societies in more recent times" (p.471, the emphasize on "mainly random" is mine).

Figure 1-1. Income per Capita and Genetic Distance to the United States



Notes: This figure is a reproduction of Spolaore and Wacziarg’s Figure III (2009, p.489). The regression represented by the fitted line yields a coefficient of -15.723 (standard error=1.807), $N=136$, $R^2=0.35$. The measure of genetic distance used in this figure is the current match of weighted F_{ST} genetic distance to the United States, which is the same measure used in Spolaore and Wacziarg (2009). This measure refers to expected genetic distance between two randomly selected individuals, one from the United States and one from another country, and is measured on the scale of 0 to 0.2088, where a larger value means less genealogical relatedness. See Section 4 for more detailed discussion of F_{ST} genetic distance.

Figure 1-1 implies that there is a negative correlation between countries’ current economic performance and the elapsed time since they shared a common ancestor with the United States population - a genetically diverse but mostly Western European population.⁴ SW’s interpretation of this correlation has two components: first, they suggest that genetic distance is a proxy for the cultural and the biological differences between populations.⁵ I accept this part of their interpretation and refer to “genetic distance” and “biocultural divergence” interchangeably throughout the paper. Second, they suggest that since it is easier for societies that are genealogically more closely related “to learn from each other and to adopt each other’s innovations” (2009, p.470), genetic distance to the United States is an obstacle to the diffusion of ideas and innovations from the world technological frontier. The focus of this paper is the second component of SW’s interpretation.

SW’s identifying assumption is that genetic distance captures mostly *random* biocultural divergences between populations. The empirical evidence presented in this paper suggests that this assumption is not

⁴According to Putterman and Weil’s (2010) *World Migration Matrix*, 1500-2000, the ancestors of the current population of the United States are 75 percent European and 68 percent Western European. The top three ethnicities of the United States are British, German, and Irish.

⁵There is a consensus among proponents of cultural explanations of the world income inequality that genetic distance between countries is a proxy of a mixture of cultural and biological differences between them; see, for example, Ashraf and Galor (2013), Desmet et al. (2007), Guiso et al. (2004, 2009), and SW (2009, 2012, 2013), and Section 2 for a review of this literature.

valid at least for genetic distance to the United States (or Western Europe). To see this, note that the two fundamental determinants of genetic distance to the United States are the early migrations of Homo sapiens out of Africa (which determines populations' genetic distance to Western Europeans before the start of colonization), and the great migration of Europeans to the New World after 1500 (which affect contemporary genetic distance to the United States as a direct result of European settlement). However, neither the early migration of Homo sapiens nor the great migration of Europeans is independent of the geographic distribution of pre-and post-colonial determinants of economic development.

With regard to the pre-colonial era (before 1500), the empirical evidence suggests that the prehistoric ecological prerequisites of good economic performance were absent in those societies that were genetically farther from Western Europeans. For example, Africans had a high genetic distance to Western Europeans before the start of European colonization, because only a fraction of Homo sapiens who evolved in East Africa around 200,000 years ago migrated out of Africa.⁶ At same time, according to Diamond (1997), Africa initial biogeographical conditions were not suitable for transition to sedentary life and development of agriculture.⁷ Since both genetic distance and economic performance are highly persistent, and since the majority of the current population of the United States is of Western European origin, the negative prehistoric correlation between genetic distance and economic performance is partly reflected in the contemporary correlation of Figure 1-1 (observe a cluster of sub-Saharan countries in the lower right quadrant).

With regard to the colonial era (after 1500), former colonies that experienced more extensive settlement by Europeans are genetically closer to Western Europe and the United States today. Acemoglu, Johnson, and Robinson (2001, 2002) - henceforth AJR - argue that European settlement is among the most important determinants of the *institutional* path of former colonies. That is, former colonies that experienced relatively larger shares of European settlement were also relatively immune to the adverse effects of colonization. AJR argue that this is because European settlement brought (European) "inclusive" institutions. This implies that the colonization strategies of Europeans have a simultaneous and opposing effect on the genetic distance of former colonies to their colonizers and their contemporary economic performance. For example, sub-Saharan Africa has a relatively higher genetic distance to Western Europe and the United states compared to other regions, partly because of the lack of European settlement in sub-Saharan Africa during the colonial era. While, according to AJR, lack of settlement in sub-Saharan Africa resulted in colonial "extractive"

⁶The earliest fossils of Homo sapiens have been found at Omo Kibish in Ethiopia and originated from 200,000 years ago. The "out of Africa" hypothesis of the origins of modern humans is widely accepted among scholars. See Ashraf and Galor (2013) for more detailed discussion of this hypothesis and Section 3 for an argument regarding the effect of the migratory path of Homo sapiens on genetic distance to Western Europeans.

⁷Ashraf and Galor (2011a) and Olsson and Hibbs (2005) provide empirical evidence confirming Diamond's theory, and suggesting that the timing of transition to agriculture has a persisting effect on historical and contemporary economic performance of nations.

institutions, with a persisting negative effect on the quality of contemporary institutions and economic performance of this region.

Thus, while SW's barrier hypothesis implies that the technological backwardness of sub-Saharan Africa is partly explained by the biocultural divergence of its population from Western Europeans and/or the United States population, this paper suggests that sub-Sahara's biocultural divergence from the West could be an obstacle to the diffusion of development, only if one is willing to ignore: (a) the importance of the ecological disadvantages of this region (e.g., Diamond, 1997; Ashraf and Galor, 2011a; Olsson and Hibbs, 2005), and (b) the significance of the colonial history of this sub-Saharan Africa (e.g., AJR; Glaeser et al., 2004; Easterly and Levine, 2012).

Empirically, I show that, consistent with the aforementioned argument, after controlling for the simultaneous effect of time-invariant pre-and post-colonial factors on genetic distance and economic performance, genetic distance to the world technological frontier does not explain the diverging development paths of nations. To be more specific, in a sample that includes both former colonies and non-colonies, controlling for continents fixed effects and proxies for the influence of Western Europe during colonization (e.g. either measures of contemporary institutions, Percent European descent, or absolute latitude) removes the statistical association between genetic distance to the United States and income per capita. This result remains robust when I control for the possible endogeneity of institutions in an IV strategy.

Next, I focus on a sample that includes only former European colonies and document a positive correlation between contemporary genetic distance to the United States and log of European settler mortality at the time of colonization. I show that under various econometric specifications, controlling for log settler mortality completely removes the statistical association between genetic distance to the United States and income per capita. AJR argue that European settler mortality is among the key determinants of colonization strategies. Hence this result provides direct and strong support for this paper's main argument regarding the simultaneous effect of colonization on colonies' genetic distance to the technological frontier and their economic performance. Moreover, I show that once the effect of other factors that European brought to their colonies (e.g., institutions or human capital) is accounted for, the biocultural divergence of former colonies from the United States does not explain their divergent development paths.⁸

To address the effect of colonization on countries' genetic distance to the technological frontier, SW use

⁸The results remain robust when I estimate the effect of genetic distance on economic performance using country-pairs in a bilateral analysis similar to SW. In addition, the base line results remain robust to controlling for geographic factors (log ancestry adjusted timing of Neolithic Revolution, log arable land area, absolute latitude, and a landlock dummy), continent, legal origin, and OPEC fixed effects, and the fraction of the population affiliated with major religions. To further check the robustness of the results in the sample of former colonies I show that potential measurement errors in the data on settler mortality (suggested by Albouy, 2012) do not affect the results, and, that, one obtains similar results using other proxies of colonization strategy e.g., log of population density in 1500.

genetic distance to the United Kingdom in 1500 as an instrument for contemporary genetic distance to the United States. The empirical evidence that I present suggest that this instrumental variable strategy is not valid in the sample of former colonies, and in fact, it might exacerbate the bias of their OLS estimates. This is because, on average, European colonists experienced higher mortality rates when they encountered populations that had been separated from them for a longer time (e.g., in sub-Saharan Africa). This suggests that SW's instrument is correlated with countries' current income per capita through channels that are not accounted for in their second stage of instrumental variable regressions. Including proxies of colonization strategies in the IV regressions restores SW's exclusion restriction, but after doing so I find no evidence that genetic distance to the technological frontier acts as an obstacle to the diffusion of development.

To further examine the validity of SW's barrier interpretation of genetic distance and the possibility of a causal interpretation of the correlation depicted in Figure 1-1, I also explore the relationship between the within country variation in genetic distance and economic performance over the past 500 years.⁹ The results suggest that, after controlling for common (colonial) shocks to genetic distance and economic performance, the reduction in the biocultural divergence of those former colonies that got genetically closer to their colonizer over the past 500 years does not explain their economic success. This result provides further evidence against SW's interpretation of genetic distance as a barrier to the diffusion of development.

The paper proceeds as follows. Section 2 reviews the related literature. In Section 3, I present my argument and descriptive evidence regarding the endogeneity of genetic distance to the technological frontiers. Section 4 describes the data. The empirical analysis is presented in Section 5. Section 6 concludes.

2 Related Literature

This study contributes to the empirical literature on the fundamental and the deep-rooted determinants of economic development.¹⁰ The most closely related lines of research are the empirical literature that use genetic distance to examine the effect of cultural and biological characteristics of populations on economic outcomes, and the literature that studies the effect of historical experience of countries during colonization

⁹Note that SW suggest that the result of their study can help policy-makers to “reduce barriers to the flows of ideas and innovations across populations” (Callaway, 2012, p.154). Therefore, their interpretation of their results has an inherent causal theme.

¹⁰In the contemporary literature, the works of Diamond (1997), Gallup et al. (1998), and Sachs (2001) refers to the effect of geography. The studies by Ahsraf and Galor (2013), Guiso et al. (2006, 2009), SW (2009, 2012, 2013) emphasize on cultural and/or biological characteristics of populations. AJR (2001, 2002, 2004, 2012), Engerman and Sokoloff (1997), Hall and Jones (1999), La Porta et al. (1999), Knack and Keefer (1995), North (1971), North and Thomas (1973), and Rodrik (1999) highlight the effect of institutions. Michalopoulos and Papaioannou (2013) explain the role of pre-colonial institutions. Reviewing all of these studies is out of the scope of this paper. Readers who are interested in these topics are referred to Acemoglu and Robinson (2012), who provide an excellent non-technical review of the literature, and a recent paper by SW (2013) on the historical and deep-rooted factors of contemporary comparative development.

on institutions.

Among the first group, Guiso et al. (2004, 2009) suggest that somatic and genetic differences between European countries have an indirect effect on their contemporary level of development. In their view, higher genetic distance between Europeans (among other factors) leads to lower level of trust between them, which then lead to lower investment and bilateral exchange. Demset et al. (2007) provide empirical evidence that genetic distance is a proxy of cultural heterogeneity, and show that cultural heterogeneity have a direct effect on the stability of European nations. Ashraf and Galor (2013) also provide support for cultural and biological explanations of the causes of underdevelopment by showing that genetic diversity, a measure which like genetic distance is based on expected heterozygosity, has a hump-shaped effect on historical and contemporary income per capita, i.e., very high and very low levels of genetic diversity are not conducive for development. They suggest that sub-Saharan Africa high genetic diversity decreases the level of trust and corporation and lead to socioeconomic disorder. On the other hand, Latin Americans low levels of genetic diversity decreases the production possibility frontier of these societies and is an obstacle to “successful implementation of superior technological paradigms” (2013, p.2).

In this paper, I accept the consensus among scholars that genetic distance (or diversity) is a proxy of biocultural heterogeneity. I argue, however, that these measures are endogenous with respect to historical determinants of contemporary comparative development and do not have a direct or indirect *causal* effect on economic performance.

Following the consensus among scholars, SW (2009) suggest that countries’ genetic distance to the United States is a proxy of their biocultural divergence from the the technological frontier, and maintain, as their key hypothesis, that “such long-term (and mainly random) divergence has created barriers to the diffusion of technological and institutional innovations across societies in more recent times” (p.471). Further, they argue that if this hypothesis is correct, then a measure of relative genetic distance from the United States should have more explanatory power for contemporary income per capita, than a measure of absolute genetic distance does. They define the absolute genetic distance between countries i and j as G_{ij}^D , which refers to the expected elapsed time since population of countries i and j shared a common ancestor; and the relative genetic distance from the United States as $G_{ij}^R = |G_{i,US}^D - G_{j,US}^D|$, where $D_{i,US}$ and $D_{j,US}$ denote genetic distance of country i and j to the United States, respectively. To test their hypothesis, SW calculate income differences between all 9316 pairs of countries (based on 137 countries) in their sample, and show that, in regressions that use the absolute value of income differences, $|Y_i - Y_j|$, as the dependent variable, the magnitude of the effect of G_{ij}^R is bigger than G_{ij}^D . They take this difference in the magnitude of the effect of

relative and absolute genetic distance on income differences as evidence in favour of the validity of their key hypothesis.

Note that this paper’s argument regarding the endogeneity of genetic distance to the technological frontier applies to both of SW’s measures of genetic distance. For example, my argument implies that both the relative and the absolute genetic distance between pairs of colonies, and also the absolute and the relative genetic distance between any pair involving a former colony and a non-colony are affected by the extent of European settlement during the colonial era. Therefore, without explicitly addressing the simultaneous effect of colonization on genetic distance and economic performance, the estimated effect of both the relative and the absolute genetic distance on income per capita is biased and inconsistent.¹¹

Two main objections to SW’s barrier interpretation of genetic distance to the United States have been advanced in the literature. Angeles (2012) suggests that SW’s results are not robust to controlling for the fraction of population descended from Europeans.¹² Campbell and Pyun (2011) show that the effect of genetic distance to the United States on income per capita is sensitive to controlling for distance from the equator, and suggest that SW are capturing the effect of geographic and climate similarity. This paper is related to these studies, but is more comprehensive in clearly distinguishing between the effect of time-invariant pre-and post-colonial factors on genetic distance and economic performance. To the best of my knowledge, the argument regarding the endogeneity of genetic distance to the technological frontier and the emphasize on the effect of colonization is unique to this paper.

In the literature on the effect of European colonialism on economic development, this paper is most closely related to AJR (2001, 2002), who maintain that colonizers’ mortality rate and the density of indigenous population at the time of colonization were among the most important determinants of colonization strategies.¹³ In places where Europeans faced high mortality rates and/or high population density, settlement was costly so they imposed “extractive” institutions to exploit natural and human resources of those regions. On the other hand, locations where Europeans faced low mortality rates and/or low population density, encouraged settlement by European colonists. Settlers generally established “inclusive” institutions, which mitigated the

¹¹Section 5 (footnote 26) presents this argument mathematically. Appendix A presents the results of pairwise income differences on the relative genetic distance in the sample of former colonies (2,123 pairs of countries based on underlying 68 former colonies for which data on settler mortality are available) and shows that the relative genetic distance is not a predictor of income differences once I control for European settler mortality and Diamond gap.

¹²However, SW (2013) control for Putterman and Weil’s (2012) share of descendent of Europeans and show that their result remains robust.

¹³The effect of colonial experience on institutions and the development path of countries has been emphasized by scholars such as Hayek (1960), La Porta et al. (1998, 1999), and North et al. (1998), among many others. These authors point to the effect of colonization on shaping the early institutions of former colonies. Acemoglu et al. (2008), Acemoglu and Robinson (2012), Engerman and Sokoloff (1997), Knack and Keefer (1995), North (1971), North and Thomas (1973), and Rodrik (1999), among others, further suggest that historical experience of countries during colonization has a persistent effect on their contemporary level of development.

adverse effect of colonization¹⁴ AJR argument, therefore, is that African underdevelopment has nothing to do with cultural and biological characteristics but rather is a direct result of the colonial institutions brought by Europeans.¹⁵

By testing the validity of biocultural explanations of the causes of underdevelopment against institutional alternatives, this paper provide strong empirical support for the latter, and contributes to our understanding of the fundamental and the deep-rooted determinants of comparative development.

3 Colonization, Genetic Distance, and Economic Performance

This section explains how time-invariant factors simultaneously affect countries' genetic distance and their economic performance in pre-and post-colonial eras. To this end, I explain the effect of two major human migrations, the early migrations of Homo sapiens out of Africa and the great migration of Europeans during colonization, on the historical and the contemporary genetic distance of populations. The goal is to highlight the implications of these two major human migrations for statistical inference, especially causal inference, based on genetic distance to the United States.

3.1 The Pre-Colonial Era

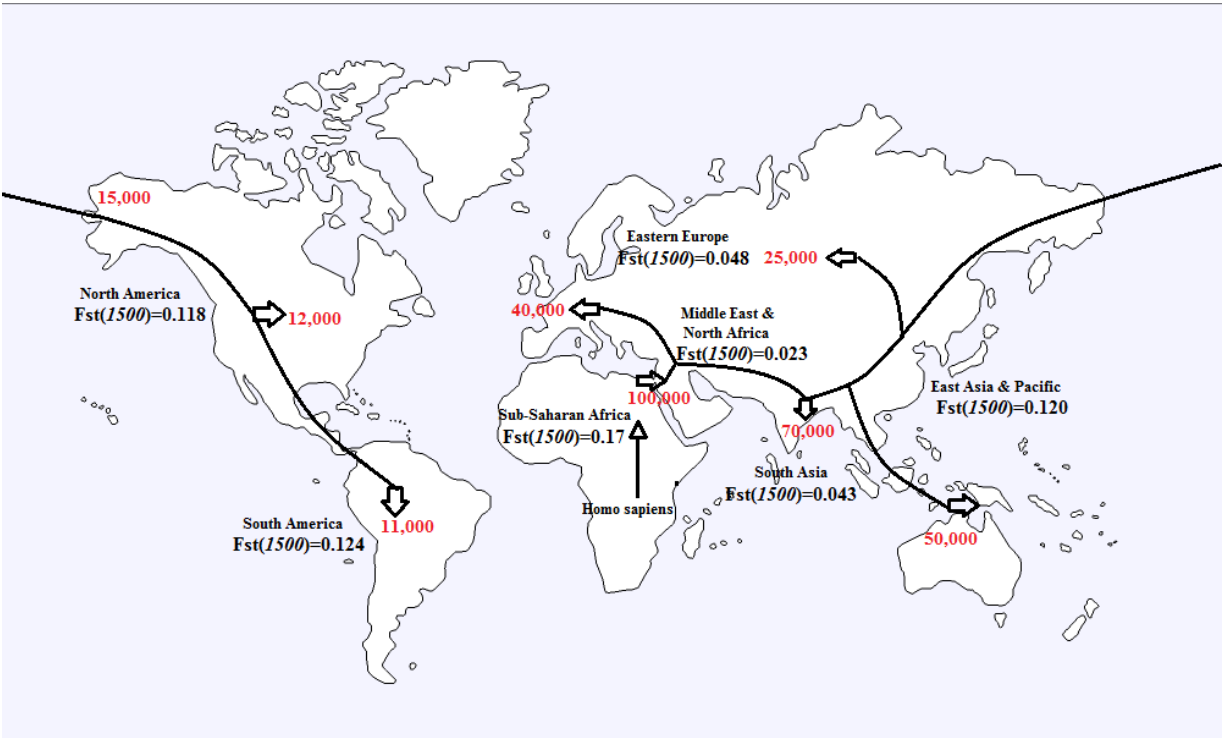
The fundamental determinant of contemporary genetic distance is the timing of separation of early modern humans, Homo sapiens, from each other in the prehistoric era. Map 1 depicts the consequence of the migratory path of Homo sapiens for the genetic distance between populations outside Western Europe and Western Europeans. Modern humans first evolved in East Africa and over a period of 200,000 years migrated to the north of this continent.¹⁶ Around 100,000 years ago a fraction of them reached Near East, and through there, colonized the rest of the planet by propagating first to South and East Asia and then, to Americas. On the map, the numbers beside arrows refer to the time elapsed since early modern humans has arrived in a specific geographical area, and " $F_{ST}(1500)$ " refers to the regional average of the genetic distance of countries to Western Europe in 1500 A.D. While early human migration had been generally eastward, at some point a fraction of the human population, who were residing in the Middle East, started migrating to

¹⁴The importance of European settlement during colonization for contemporary comparative development also confirmed by Glaeser et al. (2004) and Easterly and Levine (2012). Although Glaeser et al. (2004) suggest that European human capital has a stronger effect compared to institutions, and Easterly and Levine (2012) find evidence for the effect of both institutions and human capital.

¹⁵In fact, Acemoglu and Robinson (2012) suggest that culture is endogenous with respect to institutions and can not explain the huge differences in contemporary economic performance of nations.

¹⁶An alternative hypothesis regarding the origins of Homo sapiens is that they evolved as a result of the transfer of genetic materials between hominids that evolved independently across the planet. However, genetic and paleontological evidence do not confirm this alternative hypothesis (Ashraf and Galor, 2013).

Map 1. Migration of Homo sapiens and Genetic Distance to Western Europe in 1500 A.D.



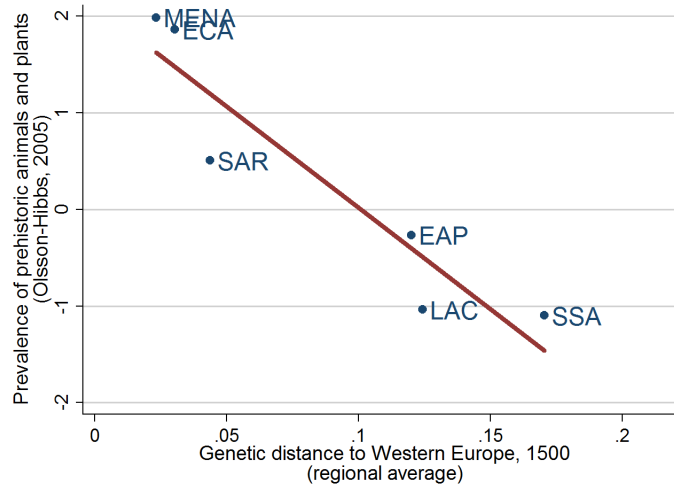
Notes: This map illustrates the “out of Africa” hypothesis regarding the origins of modern humans, and its effect on the genetic distance of populations to Western European in 1500 A.D. “ F_{ST} (1500)” refers to the regional average of genetic distance to British population in 1500. However, since genetic distance of Belgium, France, Iceland, Ireland, Luxembourg and Spain to British population in 1500 is 0 and genetic distance of Austria, Denmark, Germany, Netherlands, Norway, Sweden and Switzerland is very close to zero (0.0021), the genetic distance of countries to British population in 1500 could be interpreted as their genetic distance to Western Europe in 1500. Information on the migratory paths of Homo sapiens out of Africa is gathered from resources available on the World Wide Web, For example from Smithsonian Foundation (<http://www.smithsonianmag.com/history-archaeology/human-migration.html>), National Geographic (<https://genographic.nationalgeographic.com/human-journey>). Data on genetic distance is taken from Sploaore and Wacziarg (2009). Data definition and sources are presented in Appendix C.

the west and reached Western Europe around 40,000 years ago.

Two important points emerge out of Map 1: First, in 1500, roughly before the start of European colonization, sub-Saharan Africans had the highest genetic distance to Western Europeans. This is because only a subset of Homo sapiens who evolved in Sub-Saharan Africa migrated to the north of Africa, and only a fraction of this group migrated to the Middle East. As a result, human populations outside sub-Saharan African were genealogically much more closely related to each other than to sub-Saharan Africans.¹⁷ Second, Map 1 shows that outside Africa, there is a proportional relationship between migratory distance of

¹⁷The average of genetic distance of sub-Saharan Africans to Western Europeans in 1500 is 0.1704 (on the scale of 0 to 0.2288), while the average of the genetic distance of the rest of the world (excluding Western Europe itself) is 0.0835. To have a sense of the magnitude of this difference in terms of visible demographic characteristics of population, note that 0.0869 (=0.1704-0.0835) units difference in contemporary genetic distance of countries to the United Kingdom approximately corresponds to the difference between Netherlands and Guatemala (the current match of weighted F_{ST} genetic distance of Netherlands and Guatemala to the United Kingdom is 0.005 and 0.091 respectively).

Figure 3-1. The Initial Biogeographical Conditions and Genetic Distance to Western Europe in 1500.



Notes: Genetic distance to Western Europe in 1500 refers to F_{ST} genetic distance between plurality groups. Prevalence of prehistoric animals and plants refers to the first principal component of the regional average of the number of prehistoric animals and plants available for domestication. Both variables are taken from Olson and Hibbs (2005). ECA=Europe and Central Asia, MENA=Middle East and North Africa, EAP=East Asia and Pacific, SAR=South Asian Region, LAC=Latin America, SSA=sub-Saharan Africa, according to the World Bank classification of regions. Data definitions and sources are presented in Appendix C.

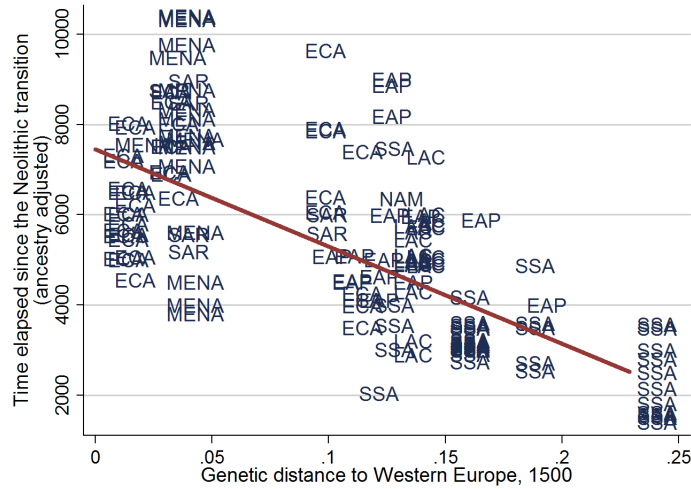
populations and their genetic distance to Western Europeans i.e., as we move eastward along the migratory path of *Homo sapiens*, genetic distance to Western Europeans increases.¹⁸

This pattern of early human migration is relevant for statistical inference based on genetic distance. Figure 3-1 plots the first principal component of the regional average of the number of prehistoric animals and plants available for domestication against the regional average of genetic distance to Western Europe in 1500.¹⁹ There is a strong negative correlation between the two variables (their covariance is -0.93). Similarly, Figure 3-2 shows that there is a negative correlation between the number of years since the Neolithic transition and genetic distance to Western Europe in 1500. Diamond (1997) argues that the ecological advantages of Eurasia provided suitable conditions for early transition to sedentary life and the development of agricultural technologies, and the East-West Orientation of Eurasia facilitated the diffusion of agricultural resources and technology across this continent. Diamond partly attributes the subsequent economic success of Europe to its initial biogeographical advantages. Olsson and Hibbs (2005) and Ashraf and Galor (2011a) provide confirmatory econometric evidence regarding the effect of initial biogeographical conditions on historical and

¹⁸This phenomenon is known as the serial founder effect in population genetics. It refers to the loss of genetic variations when populations migrate over a long distance, resulting in a positive proportional relationship between genetic distance and migratory distance outside Africa.

¹⁹Number of prehistoric domesticated animals refers to the geographical distribution of the worlds 14 domesticable herbivorous or omnivorous, terrestrial mammals weighing more than 45 kg. Number of prehistoric domesticated plants refer to the geographical distribution of the worlds 56 heaviest wild grasses.

Figure 3-2. The Timing of The Neolithic Revolution and Genetic Distance to Western Europe in 1500.



Notes: Genetic distance to Western Europe in 1500 refers to F_{ST} genetic distance between plurality groups. The vertical axis is the weighted average of Neolithic transition timing, where the weight associated with a given country represents the fraction of the year 2000 population that can trace its ancestral origins to the given country in 1500. ECA=Europe and Central Asia, MENA=Middle East and North Africa, EAP=East Asia and Pacific, SAR=South Asian Region, LAC=Latin America, NAM=North America, SSA=sub-Saharan Africa, according to the World Bank classification of regions. Data definitions and sources are presented in Appendix C.

contemporary levels of development.²⁰

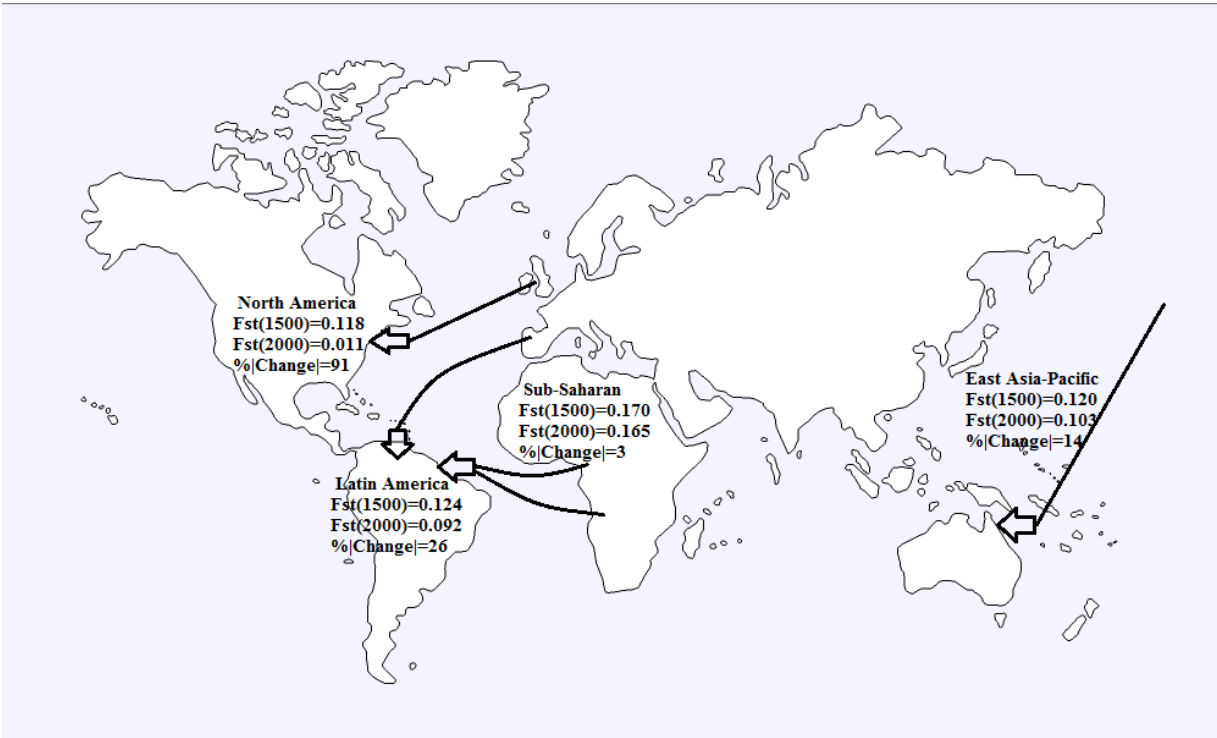
These ecological advantages were not present in other continents and this is reflected in Figures 3-1 and 3-2. These figures show that regions with populations that were genetically more distant from European also had a fewer domesticable animals and plants, and as a result, underwent the Neolithic transformation later. As a consequence, populations' genetic distance to Western Europeans before the start of European colonization was not independent of the distribution of the prehistoric determinants of economic performance. Since both genetic distance and economic performance are highly persistent, this endogeneity is transferred to the contemporary era.

3.2 The Colonial Era

Colonization and the great migration of Europeans to the New World exacerbated the problem of the endogeneity of genetic distance to the technological frontier as a determinant of contemporary economic development. Map 2 depicts the effect of European migration to the New World on genetic distance of former

²⁰Further indirect evidence in favour of Diamond's argument is provided by Coming, Easterly, and Gong (2006, 2010), who show that the level of technological advancement of countries even 3000 years ago has explanatory power for current level of development; and by Bockstette, Chanda, and Putterman (2002), who suggest that an index capturing experience with state level institutions from year 1 to 1950 is significantly correlated with contemporary measures of political stability and institutional quality, and the level and the rate of growth of income.

Map2. Migration of Europeans and Contemporary Genetic Distance to the United Kingdom.



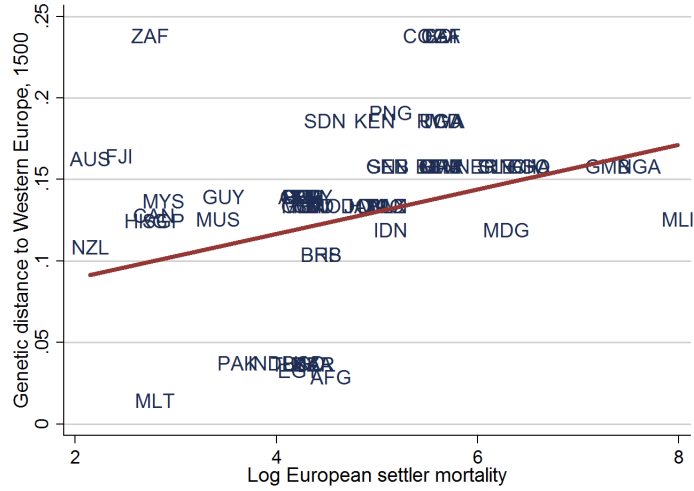
Notes: This map refers to the effect of European colonization on the regional average of the genetic distance of former colonies to the British population. “ F_{ST} (1500)” refers to the regional average of genetic distance to British population in 1500, and “ F_{ST} (2000)” refers to the regional average of the current match of weighted F_{ST} genetic distance to United Kingdom. % |Change| in genetic distance is calculated as the absolute value of $(F_{ST}(2000)-F_{ST}(1500))*100/F_{ST}(1500)$. Data on genetic distance is taken from Sploaore and Wacziarg (2009). Data definition and sources are presented in Appendix C.

colonies. On this map, “ $F_{ST}(2000)$ ” refers to the regional average of the contemporary genetic distance to the United Kingdom, and “% Change” refers to the percentage change in the regional average of genetic distance over the past 500 years.²¹ European settlement significantly reduced the genetic distance of former colonies outside sub-Saharan Africa to Western Europeans (in absolute value, 91 and 26 percent in North and South America receptively and 14 percent in Pacific region). While, Europeans generally didn’t settle in sub-Saharan Africa due to an unfavourable disease environment, leaving genetic distance to Europeans relativity unchanged. It is well known in the literature that European settlement brought inclusive institutions (AJR, 2001) and/or European human capital (Glaeser et al.; 2004). Therefore, colonization induced a negative cross-country correlation between genetic distance to the United States and current income per capita that might be unrelated to SW’s barrier effect of genetic distance.

Clarifying the underlying mechanism, Figure 3-3 plots the genetic distance of former colonies to Western

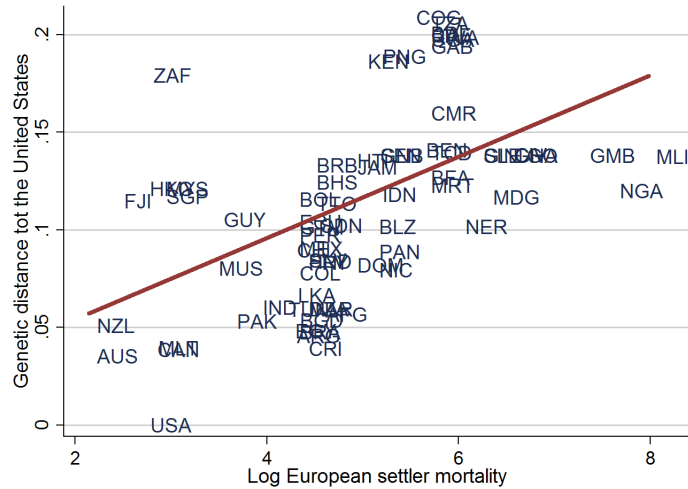
²¹In practice, one can use genetic distance to the United States and genetic distance to the United Kingdom interchangeably, since the cross-country correlation between the two measures is 0.99.

Figure 3-3. European Settler Mortality and Genetic Distance to Western Europe in 1500.



Notes: The sample includes former colonies. The regression represented by the fitted line yields a coefficient of 0.014 (standard error=0.062), N=70, $R^2=0.10$. Genetic distance to Western Europe in 1500 refers to F_{ST} genetic distance between plurality groups. The covariance between the two variables is 0.38. Data definition and sources are presented in Appendix C.

Figure 3-4. European Settler Mortality and Contemporary Genetic Distance to the United States.



Notes: The sample includes former colonies. The regression represented by the fitted line yields a coefficient of 0.020 (standard error=0.012), N=71, $R^2=0.28$. The measure of genetic distance is the current match of weighted F_{ST} genetic distance to the United States. The covariance between the two variables is 0.57. Data definition and sources are presented in Appendix C.

Europe in 1500 against log of European settler mortality. The positive correlation between the two variables suggests that, *on average*, Europeans experienced higher mortality rates when they encountered populations

that have been separated from them for a longer time.²² Figure 3-4 plots the contemporary genetic distance of former colonies to the United States against log of European settler mortality. Compared to Figure 3-3, the fitted line is steeper, highlighting the impact of European settlement on genetic distance of former colonies to the frontier of technological innovation.

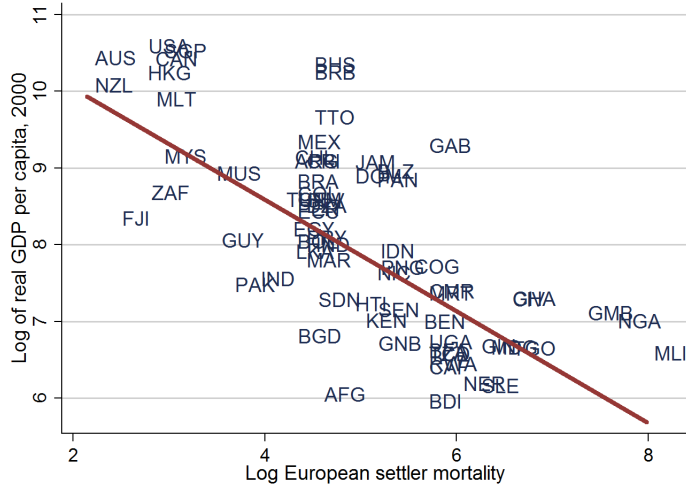
Figure 3-5 plots log of real GDP per capita against log of European settler mortality in the sample former colonies, and shows a strong negative correlation between the two variables. AJR famously argued that this correlation is a direct result of European settlement, because colonizers generally didn't settle in geographic regions with an unfavourable disease environment, and they followed an extractive strategy to exploit natural and human resources of those regions. According to Glaeser et al. (2004) this association reflects the fact that Europeans brought growth-inducing human capital to those regions in which they settled. Thus, if one does not properly account for the simultaneous effect of colonization on genetic distance and economic performance, the effect of all other factors that European brought (or didn't bring) to their former colonies might be attributed to these countries genetic distance to the United States or Western Europe (Figure A1 of Appendix B presents this argument using a diagram).

The correlations depicted in Figures 3-3/3-5 suggest that genetic distance to the United states does not capture the random biocultural differences between populations. To see this with an example, note that in the post-colonial epoch, being African corresponds to having a high genetic distance to Western Europe and to the United States, because of the hostility of the disease environment of sub-Saharan Africa to Europeans as well as the migratory path of Homo sapiens. At the same time, sub-Saharan countries have a low income per capita, because of the biogeographical disadvantages of this region (Diamond, 1997), and the European colonization strategy in Africa.

To summarize, the prehistoric ecological prerequisites of good economic performance were absent in those societies that were genetically farther away from Western Europeans. This results in a negative cross-country correlation between genetic distance to Western Europe and measures of economic performance before the start of colonization in 1500. Since both genetic distance and economic performance are highly persistent, this historical correlation is partly reflected in Figure 1-1. Further, European settlement (or lack of) in former colonies reinforced this negative correlation. A empirical strategy that does not properly control for the confounding effect of pre-and post-colonial factors suffers from bias due correlated missing variables and yields inconsistent estimates, which may lead to misleading interpretations.

²²Note that I am not implying that the correlation between the two variables is causal. In fact the existence or lack of casual relationship between the two variables is inconsequential for my argument.

Figure 3-5. Income per Capita and European Settler Mortality.



Notes: This figures illustrates AJR (2001) main argument regarding the colonial origins of comparative development. The regression represented by the fitted line yields a coefficient of -0.726 (standard error=0.080), N=71, R²=0.50. The sample includes former colonies Data definition and sources are presented in Appendix C.

4 Data and Descriptive Statistics

4.1 Genetic Distance

Data on genetic distance is from SW (2009) and are publicly available on the authors’ website.²³ The original source of the genetic distance data is Cavalli-Sforza, Menozzi, and Piazza (1994). Here I briefly review the main concepts and the definition of genetic distance; readers are referred to SW (2009, 2012, 2013) for a more detailed and technical discussion.

The main measure of genetic distance used in this paper (and SW) is F_{ST} genetic distance to the United States. This measure is based on indices of expected heterozygosity, or the probability of the dissimilarity of alleles between two randomly selected individuals from the relevant populations.²⁴ By construction, F_{ST} distance is a measure of the degree of genealogical relatedness, i.e., the time elapsed since two currently separated populations were one. Higher values of F_{ST} distance refer to longer periods of separation between populations and smaller degrees of relatedness between them. Since data on genetic distance are available at the population level, and not at the country level, SW use the ethnic composition data of Alesina et al. (2003) to match the population level genetic distance to the country level. Further, to account for the fact that the United States is a genetically diverse society, they weight the country level F_{ST} genetic distance by

²³See <http://sites.tufts.edu/enricospolaore>.

²⁴An allele is one of the various forms a gene can take.

shares of ethnic groups residing in each country. More specifically, the weighted F_{ST} genetic distance of a country k from the United States (US) is calculated according to the formula below:

$$F_{ST}^W = \sum_{i=1}^I \sum_{j=1}^J (s_i^k \times s_j^{US} \times d_{ij})$$

where s_i^k and s_j^{US} refer to the share of population $i = 1, \dots, I$ and $j = 1, \dots, J$ in countries k and the United States, respectively, and d_{ij} is the F_{ST} genetic distance between groups i and j . SW refer to this index as “the current match of weighted F_{ST} genetic distance to the United States”. To save space, I refer to this measure simply as “genetic distance to the United States”. This index captures the expected genetic distance between two randomly selected individuals, one from country k and one from the United States, and is measured on the scale of 0 to 0.2088, where a greater value means less genealogical relatedness between populations.

Following SW, in Section 5.3 I use genetic distance to the United Kingdom in 1500 as an instrument for genetic distance to the United States. This index is not weighted and refers to genetic distance between the major ethnic group of a country and the British population, based on the global distribution of ethnicities in 1500. I refer to this index as “genetic distance to Western Europe in 1500”. This is because the genetic distance of Belgium, France, Iceland, Ireland, Luxembourg and Spain to the British population in 1500 is 0, and genetic distance of Austria, Denmark, Germany, Netherlands, Norway, Sweden and Switzerland to the British population is very close to zero (0.0021).

4.2 Other Variables and Summary Statistics

My analysis is based on two samples. The first is a world sample that includes all countries for which data on income per capita, genetic distance to the United States, and other control variables are available. The second is a sample of all former European colonies for which data on income per capita, genetic distance to the United States, and log of European settler mortality (or log of population density in 1500) are available. The world sample includes 136 countries and the sample of former colonies includes 68. Table 4-1 presents summary statistics of the main variables used in empirical analysis. The dependent variable in all regressions is log of real GDP per capita in 2000, in purchasing power parity and constant 2005 US dollars from the Penn World Tables. Data on institutions are from AJR (2001), Acemoglu et al. (2008), Glaser et al. (2004), Hall and Jones (1999), and the Worldwide Governance Indicator (WGI). Data definitions and sources are presented in Appendix C.

Table 4-1: Summary Statistics

	Mean	Std. Dev.	Min	Max	N
Sample: The World					
Log of real GDP per capita, 2000	8.442	1.365	5.965	11.041	136
Genetic distance to the United States	0.094	0.051	0	0.209	136
Genetic distance to Western Europe, 1500	0.097	0.068	0	0.229	134
Latitude (absolute value)	26.467	17.104	0.422	67.47	136
Percent of descendant of European	30.836	40.53	0	100	136
Average of government effectiveness (1996-1998)	-0.003	0.987	-1.921	2.364	136
Log Neolithic transition timing (ancestry adjusted)	8.507	0.436	7.244	9.25	136
Log arable land area	9.786	2.065	2.301	14.381	136
Sample: Former Colonies					
Log of real GDP per capita, 2000	8.004	1.236	5.965	10.588	68
Genetic distance to the United States	0.11	0.05	0	0.209	68
Genetic distance to the United Kingdom	0.105	0.063	0.008	0.226	67
Genetic distance to Germany	0.098	0.053	0.009	0.174	68
Genetic distance to Western Europe, 1500	0.127	0.053	0.005	0.229	67
Log European settler mortality	4.746	1.24	2.146	7.986	68
Latitude (absolute value)	16.208	12.29	0.422	61.063	68
Percent of descendant of European	19.191	29.011	0	99.5	68
Percent European settler, 1900	14.158	24.766	0	99	66
Log Neolithic transition timing (ancestry adjusted)	8.386	0.359	7.617	9.101	68
Log arable land area	9.936	2.085	2.301	14.381	68
Social infrastructure, (1986-1995)	0.393	0.211	0.118	0.973	66
Average years of schooling (1960-2000)	3.625	2.529	0.409	10.862	57
Log population density, 1500	0.582	1.503	-3.817	4.135	66

Notes: Data definitions and sources are presented in Appendix C.

5 Empirical Analysis

Section 3 suggests that time-invariant pre-and post-colonial factors might be responsible for driving the cross-country correlation between genetic distance to the United States and income per capita that we observe in Figure 1-1. Let Y and D denote current income per capita and genetic distance to the United States, respectively, and let M denote a historical determinant of economic performance which is correlated with both D and Y , for example log of European settler mortality at the time of colonization. Suppose that the true relationship between Y , and D , and M is

$$Y = \alpha D + \beta M + u \quad (5.1)$$

where u is a disturbance term.²⁵ The hypothesis I am interested in is that, apart from its correlation with the pre-and the post-colonial determinants of economic performance (for example, continents' initial biogeographical conditions and proxies of colonization strategies) genetic distance to the technological frontier does not have any effect on income per capita; that is, $\alpha=0$.

²⁵For the sake of exposition, other covariates of income per capita are excluded from this model.

As is well-known, if M is correlated with both D and Y , its inclusion in the regression is necessary for obtaining a consistent estimate of α . Remember that in the case of settler mortality, Figures 3-4 and 3-5 show that $\text{Cov}(D, M) > 0$ and $\text{Cov}(Y, M) < 0$. Let $D = \theta M + v$ denote the statistical association between genetic distance to the United States and log European settler mortality. Also assume that D and M are independent of u and that v is independent of M . Suppose that one ignores M and estimates the effect of genetic distance on economic performance in a regression like $Y = aD + \epsilon$. Standard argument implies that the probability limit of the OLS estimate from this regression, \hat{a} , is

$$\text{plim } \hat{a} = \alpha + \beta\theta \tag{5.2}$$

note that Figure 3-4 shows that θ is positive and Figure 3-5 suggests that β is negative, therefore according to equation (5.2), the OLS estimate of a has a downward bias.²⁶ The same argument holds for the other historical determinants of income per capita that were discussed in Section 3. For example, the exclusion of other factors related to the effect of colonization, such as institutions and human capital, also induces a negative bias in the OLS estimate of the effect of genetic distance on economic performance.²⁷ Also, Section 3.1 suggests that controlling for heterogeneity between continents is necessary for obtaining a consistent estimate of α . This is because both the fundamental determinants of populations' genetic distance to the United States (the timing of separation of *Homo sapiens* from each other) and their biogeographical endowments are determined by their geographical location.²⁸

The aforementioned bias is present in all of SW's estimates of the effect of genetic distance on economic performance. This is because SW do not control for the simultaneous effect of colonization on genetic distance to the United States and income per capita.²⁹ Further, they do not properly control for heterogeneity between continents. This is because in their bilateral regressions, SW include two sets of continent dummies, one set of six dummies (one for each continent) that equal one if the two countries in a pair were on the same continent, and a set of six dummies each equal to one if one country belonged to a given continent and the

²⁶Note that using country-pairs and regressing income differences on the differences in countries genetic distance to the United States, which is SW's measure of relative genetic distance, does not solve the problem of omitted variable bias. In this case, the regression equation is $\Delta Y = a\Delta D + \Delta\epsilon$, and an argument similar to the one used in deriving equation (5.2) suggests that $\text{Cov}(\Delta D, \Delta\epsilon) = \beta\theta \neq 0$.

²⁷In practice, the size of the bias could be very large, since the empirical literature on the effect of colonization on contemporary economic performance of nations suggest that the the effect of colonization on institutions or human capital, which is captured by β is sizeable and persistent (see for example, AJR 2001, 2002; Glaeser et al., 2004; Easterly and Levine, 2012).

²⁸In this case, M could be thought of as continents fixed effects.

²⁹SW control for variables such as absolute differences in latitude and longitude relative to the United States, absolute differences in tropical land area, differences in countries' elevation, measure of geographical distance to the United States, transportation cost, climate differences, measures of linguistic and religious distance. None of these variables are proper proxy for colonization strategies or the other factors that European brought to their former colonies, such as institutions and human capital.

other did not. Neither set of dummies properly controls for the differences between continents. The following example illustrates the reason: take three countries, The United States (North America), China (Asia), and Congo (sub-Saharan Africa), as an example. In an econometric model where everything expressed in levels, the dummy variable for sub-Saharan Africa will be $(0, 0, 1)$, for the United States, China, and Congo, respectively. While in a bilateral approach, which includes US-China, US-Congo, and China-Congo as pairs of observations, SW's first and second sets of continent dummies corresponding to sub-Saharan Africa will be $(0, 0, 0)$, and $(0, 1, 1)$, respectively. In other words, the first set of continent dummies imposes an assumption of homogeneity across all three countries, and the second set implies homogeneity between China and the United States relative to Congo.³⁰

With the assumption of the independence of u from D and M , a simple test of the hypothesis of this paper is to run an OLS regression of Y , on D and M :

$$Y = aD + bM + \nu \tag{5.3}$$

to obtain the estimates \hat{a} and \hat{b} , and test for $\hat{a}=0$. Next sections present the empirical results.

5.1 Genetic Distance and Economic Performance in the World Sample

This section examines the robustness of the statistical association between genetic distance to the United States and income per capita in the world sample. Because there are important geographic factors that are correlated with both genetic distance to Western Europe in 1500 and the pre-colonial economic performance (recall Section 3.1), I include continents fixed effects in the empirical specification. The most important dimension of heterogeneity between continents is the difference between sub-Saharan Africa and the rest of the world. This is for three reasons: First, due to the fact that *Homo sapiens* evolved in sub-Saharan Africa and only a subset of them migrated out of Africa, human populations outside sub-Saharan Africa are genealogically much more closely related to each other than to sub-Saharan Africans (see Map 1). Second, compared to other regions, sub-Saharan Africa had a lower biogeographical endowments (see Figure 3-1). And third, European settled in relatively higher proportions in colonies outside sub-Saharan Africa (see Figure 3-4).³¹ Therefore, in the baseline specification, I only include a fixed effect for sub-Saharan

³⁰Campbell and Pyun (2011) show that in an econometric specification that is similar to SW's bilateral strategy, relative genetic distance is not a significant determinants of income differences, once they properly control for a fixed effect for sub-Saharan Africa and absolute latitude.

³¹This argument is consistent with the consensus among scholars that it is the history of the ancestor populations, as opposed to the history of geographical location, which is relevant for contemporary economic performance of nations (see for example, Putterman and Weil, 2010; and SW, 2013). Since colonization made populations of former colonies outside sub-Saharan Africa genetically more European (as a result of European settlement), the fact that Americas was also relatively less endowed with

Africa. In other specifications that examine the robustness of the baseline results, I include fixed effects for other continents, the ancestry-adjusted timing of the Neolithic transition, arable land area, and a landlock dummy.³²

To capture the simultaneous effect of colonization on genetic distance and economic performance, I control for either the average of government effectiveness (1996-1998) as proxy of current institutions, percent of current population of countries with an ancestor who lived in Europe in 1500, or the absolute value of Latitude. The absolute value of latitude partially captures the extent to which countries were influenced by Western Europeans and is used by Hall and Jones (1999) as an instrument for social infrastructure. The use of percent of European descent and government effectiveness capture AJR's (2001, 2002) argument regarding the importance of European settlement in affecting the institutional path of former colonies. To examine the robustness of the results, in some specifications I also include legal origin fixed effects to capture La Portal et al.'s (1999) argument regarding the importance of the legal origin of countries for the quality of government, and the fraction of the population affiliated with major religions to account for the importance of religion as an institutional barrier (or facilitator) for development (e.g. Weber, 1930; Huntington, 1991; Fish, 2002).

Estimates of equation (5.3) are presented in Table 5-1. Column (1) is the simple regression of income per capita on genetic distance to the United States. The coefficient on genetic distance has the expected sign and is significant at 1 percent. To have a sense of the magnitude of the effect, note that the coefficient of -15.723 implies that, if we take a country like Burkina Faso (a sub-Saharan country in the low income group according to the World Bank's classification), and replace a fraction of Burkinabé with Europeans with identical characteristics, to reduce genetic distance from 0.12 to 0.04 (which is the level of Egypt), Burkinabé's income per capita should increase, permanently, from \$753 to \$1684. This means 123.5 percent growth in the level of income per capita, in a country that, according to the World Bank, has experienced an average annual growth rate of 1.9 percent in its income per capita from 1961-2012. If SW's interpretation of genetic distance is correct, this increase in income per capita of Burkina Faso is a result of a reduction in biocultural barriers to the diffusion of innovations from the world technological frontier.

Column (2) adds a dummy for sub-Saharan Africa and the average of government effectiveness (1996-1998) in Panel A, percent of European descent in Panel B, and the absolute value of latitude in Panel C. Estimates in all three panels confirm that the inclusion of a dummy for sub-Saharan Africa and proxies of

prehistoric domesticable plants and animals becomes less important for the current economic performance of this continent, because the majority of the populations of America today are of European descent.

³²Timing of the Neolithic transition captures Diamond (1997) hypothesis. Since after 1500 the world witnessed mass migration of Europeans to the New World, following the literature (e.g. Putterman and Weil, 2010; Ashraf and Galor, 2013), the ancestry adjusted timing of Neolithic Revolution is used. Arable land area captures the feasibility of agriculture, and the landlock dummy captures the potential effect of geographical isolation on the diffusion of development.

Table 5-1: OLS Regression of Income per Capita.

Dependent Variable: Log Real GDP per Capita, 2000 (Penn World Tables) / Sample: The World						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Genetic distance to the United States	-15.723*** (1.808)	-2.599 (1.716)	-1.943 (1.907)	-0.161 (2.092)	-0.308 (2.145)	-0.214 (2.009)
Government effectiveness (1996-1998)		0.872*** (0.062)	0.831*** (0.063)	0.814*** (0.072)	0.819*** (0.093)	0.824*** (0.101)
R ²	0.350	0.751	0.780	0.787	0.792	0.800
Panel B						
Genetic distance to the United States	-15.723*** (1.808)	-2.337 (2.750)	-0.552 (2.894)	1.473 (3.134)	3.521 (2.892)	2.584 (2.572)
Percent European descent		0.012*** (0.003)	0.014*** (0.002)	0.027*** (0.006)	0.030*** (0.005)	0.030*** (0.005)
R ²	0.350	0.513	0.617	0.638	0.715	0.730
Panel C						
Genetic distance to the United States	-15.723*** (1.808)	-3.804 (2.683)	-1.551 (3.026)	0.259 (3.335)	2.342 (3.108)	-0.442 (2.991)
Latitude (absolute value)		0.021*** (0.006)	0.033*** (0.005)	0.031*** (0.008)	0.036*** (0.007)	0.033*** (0.007)
R ²	0.350	0.472	0.601	0.627	0.701	0.715
Observations	136	136	136	136	136	136
sub-Saharan fixed effect	-	Yes	Yes	Yes	Yes	Yes
Geographical controls	-	-	Yes	Yes	Yes	Yes
Other continents fixed effects	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: Genetic distance to the United States is the current match of weighted F_{ST} genetic distance to this country. Column (1) report the result of cross-sectional OLS regression where genetic distance is entered alone. The estimated coefficient of genetic distance is comparable to the reported result in Spolaore and Wacziarg (2009, Table I, column 1, p.488); they report a coefficient of -12.906 (standard error=1.383), $R^2=0.39$, for 137 countries using income per capita in PPP in 1995 from the World Bank. Percent European descent is from Putterman and Weil's (2010) *World Migration Matrix* and refers to the share of current population of countries with ancestors who lived in Europe in 1500. Government effectiveness is from Worldwide Governance Indicator (WGI) and is measured on a scale of -2.5 to 2.5, where a higher score means more effective government, averaged over 1996-1998. Geographical controls are log of ancestry adjusted timing of Neolithic transition, log arable land area, and a dummy variable taking value of 1 if a country is landlock. Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

Western influence during colonization dramatically reduce the magnitude of the coefficient on genetic distance to the point of statistical insignificance. The remaining columns in Table 5-1 examine the robustness of the baseline results by controlling for other historical determinants of contemporary comparative development. These robustness checks produce fairly similar results. To avoid unnecessary repetition, I therefore discuss estimates in Panel B, where the share of European descent is used to capture the effect of colonization.

Column (3) includes a set of geographical controls that the literature suggest might have an effect on the contemporary level of development (timing of the Neolithic transition, arable land area, and a landlock dummy). Adding these controls does not alter the statistical insignificance of genetic distance to the United States. Column (4) includes fixed effects for other continents and shows that the coefficient on genetic distance has the “wrong” (positive) sign and remains statistically insignificant. Adding more controls for institutional and cultural factors (columns 5 and 6) does not alter the results: the percent of European descent remains highly significant, and the coefficient on genetic distance to the United States remains positive and statistically insignificant.

A valid concern regarding the estimates in Panel A is the possible endogeneity of institutions. As is well known, more developed countries might be able to afford better institutions, and/or good institutions and high levels of income could be driven by historical factors that are not captured in the econometric model. To address this concern, I instrument for government effectiveness using the absolute value of latitude or percent of European descent in Table 5-2.³³ Hall and Jones (1999) use latitude to instrument for social infrastructure, and it is generally accepted among scholars that distance from the equator does not have a direct effect on contemporary economic outcomes.³⁴ AJR’s argument regarding the effect of European settlement on shaping the institutional path of former colonies motivates using percent of European descent as an instrument for government effectiveness.³⁵

The top panel of Table 5-2 reports the estimates using the latitude and the percent of European descent instruments in the top and the bottom panels, respectively. The estimated coefficients on genetic distance and government effectiveness in both Panels A-I and B-I are comparable to the OLS estimates reported in Table 5-1. In both panels, the coefficient on genetic distance to the United States is quantitatively small and

³³It is worth mentioning that the results presented in Tables 5-1 and 5-2 do not depend on the specific measure of institutions used in the regressions, i.e., one could obtain similar result using for example, constraint on executive, risk of expropriation, social infrastructure, indices of the quality of governance from WGI dataset.

³⁴There is a relative consensus among scholars that the effect of geographic and climatic factors, captured with variables such as latitude and longitude, on comparative development is not direct and work through the historical effect of initial geographic conditions on factors such as institutions, social capital, and culture. See Spolaore and Wacziarg (2013) for a comprehensive discussion of the effect of geography on comparative development.

³⁵AJR (2001) use percent of European settlers in 1900 as an instrument for risk of expropriation in their overidentification tests, and find that the effect of European settlement on income per capita is through institutions. I use Putterman and Weil’s (2011) share of European descent instead. The two variables are tightly correlated (the covariance between them is 0.89). This allows me to keep the sample size at 136 countries (percent of European settler in 1900 is available for 118 countries).

Table 4-3: IV Regression of Income per Capita.

Dependent Variable: Log Real GDP per Capita, 2000 (Penn World Tables) / Sample: The World						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A-I: Two-Stage Least Square / Instrument: Latitude (absolute value)						
Genetic distance to the United States	-2.795 (1.931)	0.049 (2.087)	1.856 (2.438)	4.205 (2.734)	2.167 (2.067)	3.213 (2.399)
Government effectiveness	0.836*** (0.137)	1.070*** (0.113)	1.259*** (0.242)	1.546*** (0.348)	1.391*** (0.246)	1.365*** (0.255)
Panel A-II: First Stage for Government Effectiveness, 1996-1998						
Genetic distance to the United States	-1.208 (2.025)	-1.495 (2.700)	-1.268 (2.825)	-1.717 (2.836)	0.897 (2.522)	-1.943 (2.244)
Latitude (absolute value)	0.025*** (0.006)	0.030*** (0.005)	0.025*** (0.008)	0.023*** (0.008)	0.029*** (0.007)	0.029*** (0.007)
R ² (first stage)	0.307	0.410	0.464	0.473	0.673	0.718
Panel B-I: Two-Stage Least Square / Instrument: Percent of Descendant of European						
Genetic distance to the United States	-2.127 (1.984)	0.305 (2.326)	1.633 (2.541)	3.386 (2.388)	1.723 (1.877)	3.002 (2.211)
Government effectiveness	0.959*** (0.159)	1.100*** (0.148)	1.210*** (0.210)	1.377*** (0.225)	1.247*** (0.160)	1.332*** (0.199)
Panel B-II: First Stage for Government Effectiveness, 1996-1998						
Genetic distance to the United States	-0.219 (1.935)	-0.778 (2.505)	-0.132 (2.846)	-0.567 (2.824)	2.228 (2.437)	-0.077 (2.314)
Percent of descendant of European	0.012*** (0.002)	0.013*** (0.002)	0.022*** (0.005)	0.022*** (0.005)	0.026*** (0.004)	0.023*** (0.004)
R ² (first stage)	0.366	0.427	0.483	0.493	0.704	0.726
Observations	136	136	136	136	136	136
sub-Saharan fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	-	Yes	Yes	Yes	Yes	Yes
Other continents fixed effects	-	-	Yes	Yes	Yes	Yes
OPEC fixed effect	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: Genetic distance to the United States is the current match of weighted F_{ST} genetic distance to this country. Percent European descent is from Putterman and Weil's (2002) *World Migration Matrix* and refers to the share of current population of countries with ancestors who lived in Europe in 1500. Government effectiveness is from Worldwide Governance Indicator (WGI) and is measured on a scale of -2.5 to 2.5, where a higher score means more effective government, averaged over 1996-1998. Geographical controls are log of ancestry adjusted timing of Neolithic transition, log arable land area, and a dummy variable taking value of 1 if a country is landlock. Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

statistically insignificant, and government effectiveness is significant at the 1-percent level. The estimates are robust to the inclusion of additional control variables, as we move from column (1) to (6).

The first stage regression results are of independent interest. Because they shed light on SW's hypothesis - that genetic distance to the United States is a barrier to the diffusion of technological and institutional innovations. The reported results in Panels A-II and B-II do not confirm their hypothesis. After accounting for the effect Western influence during the colonial era, genetic distance to the United States is not a statistically significant determinant of government effectiveness.

Overall, these results are strong evidence in favour of the importance of the simultaneous effect of colonization on genetic distance to the technological frontier and economic performance and casts serious doubt on SW's barrier interpretation of genetic distance.

5.2 Genetic Distance and Economic Performance in Former Colonies

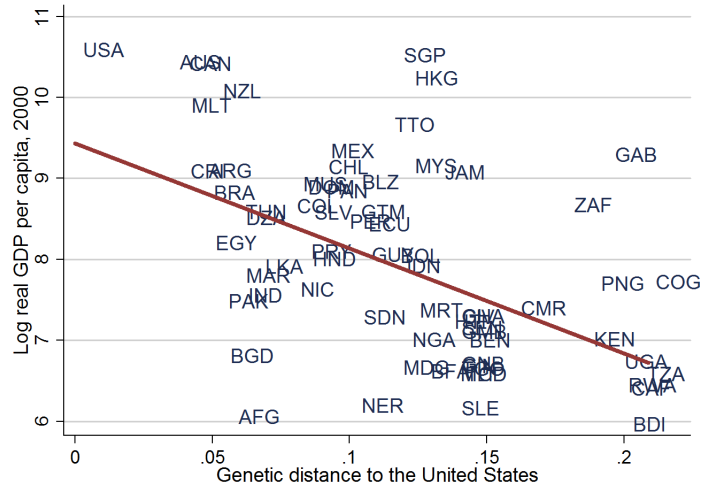
Figure 5-1 plots log of real GDP per capita in 2000 against genetic distance to the United States in the sample of 68 former colonies. The coefficient on genetic distance in the regression represented by the fitted line is -12.967 (standard error=2.615), so the relationship between genetic distance and economic performance in this sample is comparable to the world sample. As mentioned before, this is unlikely to be an unbiased estimate of the effect of genetic distance to the technological frontier on economic performance because of omitted variables.

Table 5-3 presents estimates that address this bias by controlling for European settler mortality as a proxy of colonization strategy. In Panel A genetic distance to the United States is used as the proxy of genetic distance to the world technological frontier. Panels B and C use genetic distance to the United Kingdom and Germany, instead, as a robustness check.³⁶ The estimated coefficients in column (1) suggest that the genetic distance of former colonies to the world technological frontiers is negatively correlated with their income per capita and is highly statistically significant. Once we control for the European settler mortality (column 2), the magnitude of the coefficients on genetic distance falls substantially, and the coefficients are no longer statistically significant. The coefficients on log settler mortality, on the other hand, are negative and significant at 1 percent.

The remaining columns in Table 5-1 show that the regression results presented in column (2) are robust to controlling for other historical correlates of income per capita. In column (3) log of ancestry-adjusted timing

³⁶According to Putterman and Weil's (2010) *World Migration Matrix*, almost 55 percent of the United States population have British or German ancestors. Further, both countries are in the front-end of technological and institutional innovations compared to former colonies.

Figure 5-1. Income per Capita and Genetic Distance to the United States in Former Colonies



Notes: The regression represented by the fitted line yields a coefficient of -12.967 (standard error=2.615), N=68, R²=0.27.

of the Neolithic transition, log percentage of arable land, and a landlock dummy are added to the model. Columns (4) and (5) include continents and legal origins fixed effects, and column (6) adds major religion shares. In all three panels, the inclusion of the set of control variables does not affect the baseline results of column (2), i.e., genetic distance remains statistically insignificant, sometimes with an estimated positive effect, and log of European settler mortality has the expected sign and is highly significant, statistically. Overall, the results presented in Table 5-1 provide strong empirical support in favour of the simultaneous effect of colonization on genetic distance and economic performance and the endogeneity of genetic distance.³⁷

An advantage of limiting the sample to include only former colonies is that since these countries were directly affected by the feasibility and the extent of European settlement during the colonial era, and since we have a relatively clear story regarding the effect of colonization on contemporary economic performance of these nations, a horse race between biocultural and institutional explanations of the causes of underdevelopment is possible in this sample. Besides, such a test provides an additional robustness check for the

³⁷A possible concern with the result presented in Table 4-4 could be that the potential measurement errors in settler mortality data are affecting the results. For example, Albouy (2012) suggest the AJR's bishop mortality rates in Latin America are not a reliable source of Spanish and Portuguese's mortality rates at the time of colonization, and that AJR do not properly distinguish between peacetime and campaign episodes. Albouy (2012) shows that correcting for these two potential flaws in AJR's settler mortality data would result in settler mortality to be a weak instrument for institutions. Although AJR (2012) provide a comprehensive answer to Albouy's comments, to remove any source of concern, in regression performed but not presented here I have examined the robustness of the results presented in Table 4-4 to limiting the sample to include only former colonies for which according to Albouy the data on settler mortality are the most reliable (this is 26 out of 68 countries). The obtained results is exactly the same as the results reported Table 4-4, i.e., genetic distance is significant at the 1-percent level when entered alone, and not statistically significant when settler mortality is included on the right-hand side. Besides, Appendix A suggests that one could obtain virtually similar results, using log of population density in 1500 as the proxy of colonization strategies.

Table 5-3: OLS Regression of Income per Capita, Controlling for European Settler Mortality

Dependent Variable: Log of Real GDP per Capita, 2000 / Sample: Former Colonies						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Genetic Distance to the United States						
Genetic distance to the United States	-12.967*** (2.615)	-3.825 (2.594)	-3.645 (3.255)	0.742 (3.843)	0.192 (4.171)	-1.772 (3.939)
Log European settler mortality		-0.661*** (0.090)	-0.563*** (0.118)	-0.455*** (0.136)	-0.444*** (0.128)	-0.360*** (0.130)
Observations	68	68	68	68	68	68
R ²	0.271	0.575	0.654	0.722	0.726	0.749
Panel B: Genetic Distance to the United Kingdom						
Genetic distance to the United Kingdom	-9.506*** (2.071)	-2.422 (2.122)	-2.200 (2.623)	1.871 (3.108)	1.497 (3.508)	0.139 (3.192)
Log European settler mortality		-0.654*** (0.093)	-0.532*** (0.118)	-0.403*** (0.128)	-0.400*** (0.124)	-0.336** (0.126)
Observations	67	67	67	67	67	67
R ²	0.246	0.551	0.646	0.719	0.721	0.738
Panel C: Genetic Distance to Germany						
Genetic distance to Germany	-12.138*** (2.377)	-3.232 (2.512)	-2.881 (2.688)	2.246 (3.366)	1.830 (3.480)	0.060 (3.400)
Log European settler mortality		-0.680*** (0.102)	-0.565*** (0.111)	-0.462*** (0.122)	-0.444*** (0.118)	-0.418*** (0.128)
Observations	74	74	74	74	74	74
R ²	0.274	0.554	0.636	0.712	0.717	0.723
Geographical controls	-	-	Yes	Yes	Yes	Yes
Continent fixed effects	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: In all three panels, the measure of genetic distance is the current match of weighted F_{ST} genetic distance. Geographical controls are absolute value of latitude, percentage of tropical land, and a dummy variable taking value of 1 if a country is land-lock. Continent fixed effect are dummies for Africa, Asia, and Other continents with America being the omitted group. Data definitions and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

argument of Section 3 and the regression results presented in Table 4-4.

Table 5-4 control for the social infrastructure index of Hall and Jones (1999), instrumented with log European settler mortality (AJR), and asks whether it is the genetic distance of former colonies to the technological frontier, or the colonial heritage that is embedded in their institutions that affect their income per capita. Panel A reports the result of the second stage of the IV regressions, and Panels B and C report the corresponding first stage and OLS, respectively.

Column (1)-Panel C shows that when genetic distance is the only independent variable, it's coefficient is significant at 1 percent. Column (2)-Panel A adds social infrastructure (instrumented with log settler

Table 5-4: IV Regression of Income per Capita / Instrumented Variable: Social Infrastructure

Dependent Variable: Log of Real GDP per Capita, 2000 (Penn World Tables) / Sample: Former Colonies						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Two-Stage Least Square / Instrument: Log European Settler Mortality						
Genetic distance to the United States		-3.002 (2.780)	-2.428 (3.284)	3.852 (2.932)	3.865 (3.119)	3.779 (2.993)
Social infrastructure		6.289*** (0.917)	5.823*** (1.263)	4.325*** (1.043)	4.327*** (1.067)	4.282*** (1.309)
Panel B: First Stage for Social Infrastructure, 1986-1995						
Genetic distance to the United States		-0.264 (0.477)	-0.068 (0.692)	-0.652 (0.789)	-0.722 (0.840)	-1.145 (1.079)
Log European settler mortality		-0.103*** (0.024)	-0.088*** (0.029)	-0.093*** (0.034)	-0.092*** (0.034)	-0.076** (0.036)
R ² (first stage)		0.429	0.444	0.478	0.480	0.514
Panel C: Ordinary Least Square						
Genetic distance to the United States	-13.799*** (2.489)	-7.540*** (2.112)	-1.621 (2.982)	4.097 (3.261)	3.907 (3.532)	2.621 (2.877)
Social infrastructure		3.645*** (0.396)	2.816*** (0.462)	2.691*** (0.383)	2.678*** (0.391)	2.463*** (0.417)
Observations	66	66	66	66	66	66
Geographical controls	-	-	Yes	Yes	Yes	Yes
Continent fixed effects	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: Panel A report the result of the second stage of the instrumental variable regression, where social infrastructure (1986-1996) is instrumented with log of European settler mortality. Panel B reports to the results of the corresponding first stage, and Panel C report the result of Ordinary Least Square regressions. Geographical controls are log of the ancestry adjusted timing of Neolithic Transition, log percentage of arable land, a landlock dummy, and the absolute value of latitude. Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

Table 5-5: IV Regression of Income per Capita / Instrumented Variable: Schooling

Dependent Variable: Log of Real GDP per Capita, 2000 (Penn World Tables) / Sample: Former Colonies						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Two-Stage Least Square / Instrument: Log European Settler Mortality						
Genetic distance to the United States		-0.626 (2.283)	-2.232 (2.530)	-1.333 (3.292)	-1.122 (3.408)	1.491 (4.123)
Years of schooling, 1960-2000		0.496*** (0.065)	0.444*** (0.060)	0.513*** (0.079)	0.532*** (0.082)	0.602*** (0.117)
Panel B: First Stage for Years of Schooling, 1960-2000						
Genetic distance to the United States		-7.302 (4.899)	-7.687 (7.587)	0.867 (9.059)	-0.085 (9.126)	-5.938 (9.369)
Log European settler mortality		-1.400*** (0.213)	-1.464*** (0.300)	-1.082*** (0.273)	-0.993*** (0.237)	-0.743*** (0.258)
R ² (first stage)		0.573	0.606	0.741	0.754	0.808
Panel C: Ordinary Least Square						
Genetic distance to the United States	-13.452*** (2.556)	-3.441* (1.932)	-2.469 (2.698)	0.186 (3.244)	0.198 (3.383)	0.998 (4.319)
Years of schooling, 1960-2000		0.387*** (0.044)	0.354*** (0.038)	0.377*** (0.048)	0.378*** (0.049)	0.376*** (0.055)
Observations	57	57	57	57	57	57
Geographical controls	-	-	Yes	Yes	Yes	Yes
Continent fixed effects	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: Panel A report the result of the second stage of the instrumental variable regression, where average years of schooling is instrumented with log of European settler mortality. Panel B reports to the results of the corresponding first stage, and Panel C report the result of Ordinary Least Square regressions. Geographical controls are log of the ancestry adjusted timing of Neolithic Transition, log percentage of arable land, a landlock dummy, and the absolute value of latitude. Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

mortality) and shows that genetic distance to the United States is no longer statistically significant. The comparison of OLS and IV results (column 2-Panels C and A) suggests that OLS regression underestimates the effect of institutions by almost 50 percent. Column (3) demonstrates that the result presented in column (2) is robust once I control for log ancestry adjusted timing of Neolithic transition, log arable land area, landlock dummy, and absolute latitude. Controlling for continent and legal origin fixed effects and major religion shares (columns 4-6) does not alter the results, i.e., social infrastructure remains significant at the 1 percent, and the coefficient on genetic distance to the United States is positive and insignificant. Panel B demonstrates that in none of the regressions, the biocultural divergence of former colonies from the United States has any effect on their institutions.

Glaeser et al. (2004) suggest that Europeans brought their growth-inducing human capital to former colonies in which they settled, and that human capital captures relatively “deeper” factors compared to institutions. To address their argument, Table 5-5 controls for the average years of schooling (1960-2000) in IV regressions. Column (1)-Panel C shows that the coefficient on genetic distance to the United States is quantitatively large and highly statistically significant in the OLS regression (and when genetic distance is the only independent variable). The IV and OLS estimates (Panels A and C) suggest that, with and without the set of control variables, the inclusion of the average years of schooling significantly reduces the magnitude of the effect of genetic distance and removes its statistical significance (except column 2-Panel C, in which genetic distance to the United States is significant at 10 percent in the OLS regressions). The first stage results in Panel B suggest that the genetic distance between former colonies and the world technological frontier is not a significant determinant of these countries human capital.

Overall, the estimates presented in Tables 5-4 and 5-5 suggest that, once the effect of other factors that European brought to their colonies is accounted for, genetic distance to the technological frontier does not have a significant explanatory power for contemporary comparative development.

5.3 SW’s Instrumental Variable Strategy

To control for the possible endogeneity of genetic distance to the United States, that might “arises from the pattern of colonization of the New World starting after 1500” (SW, 2009 p.502), SW use genetic distance to Western Europe in 1500 as an instrument for contemporary genetic distance to the United States. They argue that since genetic distance is highly persistent, and because the majority of the current population of the United States has Western European ancestors, genetic distance to Western Europe is a good instrument for genetic distance to the United States. The problem with SW’s instrumental variable strategy, however, is that the correlation between genetic distance to Western Europe in 1500 and log European settler mortality, depicted in Figure 3-2, leads to a mechanical violation of their exclusion restriction. This is because their instrument is correlated with contemporary economic performance through channels that is not accounted for in their first stage.

To clarify, Figure 5-2 depicts the reduced-form effect of colonization on genetic distance and economic performance. Since Figure 3-3 shows that genetic distance to the Western Europe in 1500, denoted by D_0 in the diagram, is positively correlated with log of European settler mortality (M), a line connects D_0 to M . Since European settler mortality has a positive effect on genetic distance to the United States (D), and a negative effect on current economic performance (Y), two arrows connect M to D and Y . SW’s exclusion

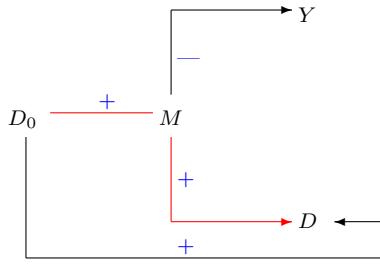


Figure 5-2. The Reduced Form Effect of Colonization on Genetic Distance and Economic Outcomes.
Notes: D_0 and D refer to genetic distance to Western Europe in 1500 and contemporary genetic distance to the United States, respectively. M refers to European settler mortality rate, and Y refers to contemporary economic performance e.g., income per capita. The arrows refer to the direction of causal effect, and the sign beside each arrow denotes the sign of this effect.

restriction is based on the path that connect D_0 to D , while Figure 5-2 shows that D_0 is associated with the contemporary economic performance, Y , through its association with European settler mortality at the time of colonization (M). Since SW do not control for the effect of colonization in their second stage of IV regressions, this violate their exclusion restrictions. In fact, their IV regressions might exacerbate the bias in OLS estimates. This is because in the first stage, the fitted values of genetic distance to the United States are inflated as a result of ignoring the path that connect D_0 to D through M .

To avoid the aforementioned problem, I include log of European settler mortality in the IV regression for the sample of former colonies. In the world sample, adding a fixed effect for sub-Saharan Africa and percent European descent restores SW's exclusion restriction.³⁸ Panels A and B of Table 5-6 report the results for the former colonies sample and the world sample, respectively.

The estimates without controlling for the simultaneous effect of colonization on genetic distance and economic performance is reported in column (1). The coefficients on genetic distance to the United States is significant at 5 percent and 10 percent in Panels A and B, respectively. These estimates are consistent with SW's finding and provide evidence in favour of their barrier effect of genetic distance to the technological frontier. Column (2) adds proxies of the effect of colonialism and shows that, neither in the world sample nor in the sample of former colonies genetic distance to the United States confers a statistically significant effect on income per capita. The remaining columns in Table 5-6 show that these results remain robust to controlling for other historical correlates of contemporary development. Therefore, the IV regressions confirm the argument of this paper regarding the endogeneity of genetic distance to the world technological frontier and show no sign of causal effect of genetic distance on economic performance.

³⁸Remember that Figures 3-1 and 3-2 show that sub-Saharan Africa had a high genetic distance to Western Europe and relatively inferior economic performance before the start of colonization in 1500 due to time-invariant geographic factors. This is captured by a fixed effect for sub-Saharan Africa. Percent of European descent is simply a proxy of the colonization strategy of Europeans.

Table 5-6: IV Regression of Income per Capita / Instrumented Variable: Genetic Distance to the U.S

Dependent Variable: Log of Real GDP per Capita, 2000 (Penn World Tables)						
Instrument: Genetic Distance to Western Europe, 1500						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Two-Stage Least Square / Sample: Former Colonies						
Genetic distance to the United States	-8.169** (3.578)	0.183 (3.896)	8.828 (6.322)	8.039 (6.932)	7.822 (7.248)	11.436 (10.639)
Log European settler mortality		-0.727*** (0.109)	-0.479*** (0.126)	-0.340*** (0.132)	-0.338*** (0.129)	-0.263** (0.130)
Observations	67	67	67	67	67	67
R ²	0.209	0.539	0.574	0.706	0.707	0.709
Panel B: Two-Stage Least Square / Sample: The World						
Genetic distance to the United States	-15.518*** (2.103)	-2.887 (3.416)	0.282 (3.791)	3.769 (3.403)	5.334* (3.057)	4.190 (3.165)
Percent European descent		0.011*** (0.003)	0.014*** (0.003)	0.027*** (0.006)	0.030*** (0.005)	0.030*** (0.006)
sub-Saharan fixed effect		-1.218*** (0.276)	-1.074*** (0.266)	-1.749*** (0.397)	-1.898*** (0.380)	-1.790*** (0.389)
Observations	134	134	134	134	134	134
R ²	0.333	0.503	0.620	0.644	0.717	0.727
Geographical controls	-	-	Yes	Yes	Yes	Yes
Continent fixed effects	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: Panel A report the result of the second stage of the instrumental variable regression for the sample of former colonies and the world sample, respectively, genetic distance to the United States is instrumented with instrumented with genetic distance to Western Europe in 1500. Geographical controls are log of the ancestry adjusted timing of Neolithic Transition, log percentage of arable land, a landlock dummy, and the absolute value of latitude. Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

5.4 Genetic Distance and Economic Performance In the Long-Run

The argument of Section 3 refers to the simultaneous effect of colonization on genetic distance and economic performance over the past 500 years. This subsection provides a more direct test of this argument by exploring the within country variation in genetic distance and economic performance over the the long-run. The idea is to ask whether it is more likely for a country to *become* relatively richer as it *becomes* genetically closer to the world technological frontier. If SW's key hypothesis is correct, one would expect to see that, for example, former colonies that got genetically closer to Western Europeans (and consequently to the United States) over the past 500 years, to be relatively richer today compared to their genetically-remote counterparts, as a result of the change in their populations' genetic characteristics.

Figure 5-3 plots the change in the log of real income per capita against the change in genetic distance in the sample of 43 former colonies that got genealogically *closer* to Angelo-American population over the past 500 years. As the slope of the fitted line suggests, to a first approximation, SW's barrier hypothesis is correct: in the regression represented by the fitted line change in genetic distance is significant at -22.250 (standard error=5.586).³⁹ However, the long-run association between these two variables refers not only to the potential effect of a reduction in the biocultural differences between former colonies and Western Europeans on development, but also to the effect of colonization on the development paths of former colonies through non-genetic channels (e.g., institutions). Therefore, the estimated slope of the fitted line could be inconsistent and biased. To see this more clearly, assume that the following statistical processes governing income per capita and genetic distance to the technological frontier are

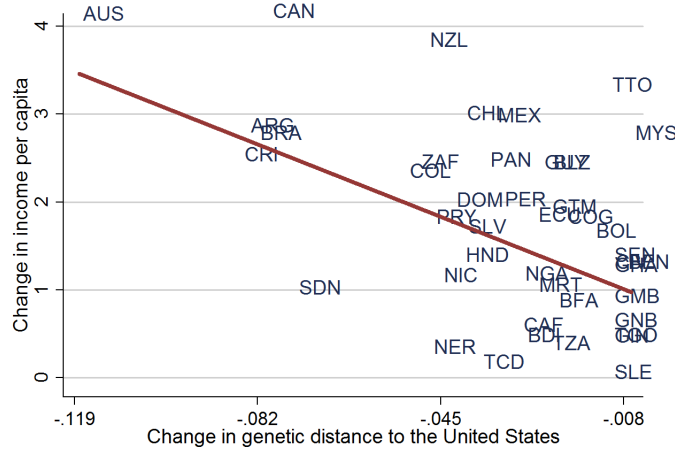
$$Y_{it} = \alpha D_{it} + \delta_i^Y + \epsilon_{it}^Y \quad \text{and} \quad D_{it} = \delta_i^D + \epsilon_{it}^D$$

where Y_{it} and D_{it} are the levels of income per capita and genetic distance at time t , respectively, and δ_i^Y and δ_i^D are fixed differences in the levels of income and genetic distance. ϵ_{it}^Y and ϵ_{it}^D are error terms. Note that I have observations on income and genetic distance at two points in time, $T - S = 1500$ and $T = 2000$.

After some manipulations, the equation capturing the changes in income and genetic distance over the past

³⁹A close inspection of the sample 65 countries that got genetically closer to Western Europeans over the past 500 years reveals that 54 out of 65 of these countries, or 86 percent of this sample, are former colonies. Non-colonies that got genetically close to Angelo-American populations are Estonia, Ethiopia, Kazakhstan, Kyrgyzstan, Latvia, Liberia, Lithuania, Suriname, and Thailand. The sample used in Figure 4-5 includes only a subset of these countries for which I have data on proxies of colonization strategies, and therefore includes only former colonies. Also, regressions performed but not presented here suggest that the change in genetic distance does not possess a significant explanatory power for the change in income in countries that got genetically away from Western Europeans over the past 500 years. For these reasons, in this subsection I focus on the sample of 43 former colonies that got genetically closer to their colonizers over the past 500 years.

Figure 5-3. Change in Income per Capita and Genetic Distance to the United States, 1500-2000



Notes: The regression represented by the fitted line yields a coefficient of -22.250 (standard error=5.586), N=43, R²=0.28.

500 years could be written as:

$$Y_{iT} - Y_{iT-S} = \alpha \Delta \epsilon^D + \Delta \epsilon^Y$$

where $\Delta \epsilon^Y = \epsilon_{iT}^Y - \epsilon_{iT-S}^Y$ and $\Delta \epsilon^D = \epsilon_{iT}^D - \epsilon_{iT-S}^D$. The slope of the fitted line in Figure 4-4 is the OLS estimate of α . Standard argument implies that the probability limit of this OLS estimates is

$$\text{plim } \hat{\alpha} = \alpha + \frac{\text{Cov}(\Delta \epsilon^Y, \Delta \epsilon^D)}{\text{Var}(\Delta \epsilon^D)} \quad (5.4)$$

therefore, according to equation (5.3), the OLS estimate of α is consistent if and only if $\text{Cov}(\Delta \epsilon^Y, \Delta \epsilon^D) = 0$. However, the empirical evidence presented so far refers to the existence of large and common shocks, caused by colonization, affecting former colonies' genetic distance to the United States and their income per capita in opposite directions. This means that $\text{Cov}(\Delta \epsilon^Y, \Delta \epsilon^D) < 0$. Therefore, there is a downward bias in the OLS estimate of α .⁴⁰

In practice, one has to control for proxies of the colonial shocks to genetic distance and economic performance to obtain consistent estimates. The empirical evidence presented before suggest that European settler mortality, population density in 1500, and the extent of European settlement in former colonies could

⁴⁰This model is a simpler version of a framework suggested by Acemoglu et al. (2008) to study the effect of shock or "critical juncture" to the changes in variables over the long-run. The argument regarding the inconsistency of the OLS estimate of α is valid if first, the colonial shocks to genetic distance and economic performance have happened during the period in which the data is observed (between 1500 and 2000), and, second, if these shocks have a persistent effect on both variables.

Table 5-7: Income and Genetic Distance Over the Past 500 Years

Dependent Variable: Change in Real GDP per Capita, 1500-2000					
	(1)	(2)	(3)	(4)	(5)
Panel A: Change in genetic distance (the United States - Western Europe in 1500)					
Change in genetic distance, 1500-2000	-22.250*** (5.586)	-6.055 (4.391)	-2.931 (6.533)	0.974 (4.806)	3.935 (5.604)
Log European settler mortality		-0.612*** (0.104)			-0.356*** (0.110)
Log population density, 1500			-0.492*** (0.109)		-0.039 (0.117)
Percent European settler, 1900				0.034*** (0.004)	0.022*** (0.005)
Observations	43	43	43	43	43
R ²	0.279	0.633	0.460	0.668	0.754
Panel B: Change in genetic distance (the United Kingdom - Western Europe in 1500)					
Change in genetic distance, 1500-2000	-15.749*** (3.684)	-5.630* (3.118)	-4.692 (5.112)	0.629 (3.824)	3.110 (4.409)
Log European settler mortality		-0.590*** (0.108)			-0.431*** (0.119)
Log population density, 1500			-0.392*** (0.124)		-0.078 (0.104)
Percent European settler, 1900				0.032*** (0.005)	0.018*** (0.006)
Observations	46	46	46	46	46
R ²	0.293	0.635	0.404	0.562	0.717

Notes: Countries in the sample in both panels are former colonies that got genetically closer to Americans and Western Europeans over the past 500 years. In Panel A, change in genetic distance is calculated as difference between the current match of weighted F_{ST} genetic distance to the United States and genetic distance to Western Europe (between plurality groups) in 1500. In Panel B, change in genetic distance is calculated as difference between the current match of weighted F_{ST} genetic distance to the United Kingdom and genetic distance to Western Europe (between plurality groups) in 1500. Data on change in income per capita over the past 500 years is taken from Acemoglu, Johnson, Robinson, and Yared (2008). Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

be used for this purpose.⁴¹ Table 5-7 presents the results of OLS estimation of α , with and without proxies of colonization strategies. The sample includes former colonies that got genetically closer to the Anglo-American population over the past 500 years. Panel A reports estimates when the change in genetic distance is calculated as the difference between the current match of weighted F_{ST} genetic distance to the United States and the F_{ST} genetic distance (between plurality groups) to Western Europe in 1500. Panel B shows estimates where the current match of weighted F_{ST} genetic distance to the United Kingdom is used as the measure of contemporary genetic distance.

⁴¹AJR (2002) suggest that population density of former colonies is also among the key determinants of colonization strategies. This is because colonizers generally didn't settle in those region that were relatively rich in natural resources and as a result, were densely populated during the Malthusian era (e.g. Aztec and Inca empires), and followed an extractive strategy to exploit those regions (thereby creating a "reversal of fortune").

Column (1) of both panels is the simple regression of the change in income on the changes in genetic distance. The estimates suggest that the effect of the change in genetic distance is quantitatively very large and significant at the 1-percent level. Column (2) adds log of European settler mortality. This induces dramatic reductions in the coefficient on the change in genetic distance in Panels A and B, and make this variable statistically insignificant in Panel A, and significant at 10 percent in Panel B. Column (3) controls for log population density in 1500. This completely removes the statistical association between the change in genetic distance and the change in economic performance in both panels. Similarly, controlling for the extent of European settlement in 1900 (column 4) reduces the magnitude of the effect of the change in genetic distance to the point of statistical insignificance. Finally, column (5) includes all of the three proxies of the common shocks simultaneously, and shows that those former colonies that became genetically closer to the world technological frontier over the past 500 years did not become richer because of it.

6 Concluding Remarks

Recent empirical literature on the fundamental causes of economic development refers to the effect of deep-rooted genetic factors on the prosperity of nations. For example, Spolaore and Wacziarg (2009) document a strong negative cross-country association between genetic distance to the United States and income per capita, and suggest that countries' genetic distance to the world technological frontier is a barrier to the diffusion of development. In their view, this is because it is easier for populations that are genealogically close to each other to adopt each others innovations. This paper argues that there is no causal relationship between genetic distance to the technological frontier and economic performance. The reason, is that, the two fundamental determinants of contemporary genetic distance to the United States - the pattern of the separation of Homo sapiens from each other during their migration out of Africa, and the extent of European settlement in their former colonies during the colonial era - are not independent of the geographic distribution of the pre-and the post-colonial determinants of economic performance. More specifically, I show that: a) the prehistoric ecological prerequisites of good economic performance (Diamond, 1997) were absent in those societies that were genetically far from Western Europeans, and b) during the colonial era Europeans generally didn't settled in those former colonies that were genetically remote from them, and imposed extractive institutions to exploit the natural and the human resources of those regions (e.g. sub-Saharan Africa).

I shows that controlling for confounding factors completely removes the association between genetic distance to the technological frontier and economic performance. To be more specific, I demonstrate that, in

a sample that includes both former colonies and non-colonies, the association between genetic distance and economic performance is not robust to controlling for a fixed effect for sub-Saharan Africa and proxies of the influence of Western Europe during the colonial era (either measures of contemporary institutions, share of European descent, or absolute latitude). Further, in the sample of former colonies, controlling for log of European settler mortality at the time of colonization removes the statistical association between genetic distance and income per capita. The results remain robust once I control for institutions and/or human capital instrumented with log of European settler mortality or log of population density in 1500. Fixed effect estimates also suggest that those former colonies that became genetically closer to Western Europeans over the past 500 years did not become richer because of it. My results are in stark contrast to Spolaore and Wacziarg (2009, 2012, 2013) and provide further empirical support for Acemoglu et al. (2001, 2002) and Glaeser et al. (2004).

By providing strong empirical support for *institutional* explanations of the causes of economic development, this study contributes to a better understanding of the fundamental causes of development, especially the literature on the effect of European colonialism on comparative development, and sheds light on recent controversial claims regarding the effect of deep-rooted genetic factors on the prosperity of nations. For example, Ashraf and Galor (2013) use genetic distance between population to construct their measure of “predicted genetic diversity”, and show that sub-Saharan Africa high genetic diversity and Latin America low genetic diversity is not conducive for development. However, Ashraf and Galor (2013) do not control for the effect of colonization on the institutional paths of former colonies. This is because they do not address the possible endogeneity of institutions and/or human capital in Africa and Latin America, and, therefore, overestimate the effect of genetic diversity on contemporary economic development.

Appendices

A Additional Robustness Checks

Robustness Check with Log Population Density in 1500

A concern regarding the regression results presented in Table 5-4 and 5-6 is the possibility that the potential measurement errors in settler mortality data has any effect on the estimated coefficient of genetic distance. Albouy (2012) suggest the AJR’s bishop mortality rates in Latin America are not a reliable source of Spanish and Portuguese’s mortality rates at the time of colonization, and that AJR do not properly distinguish between peacetime and campaign episodes. Albouy (2012) shows that correcting for these problems result

in settler mortality to be a weak instrument for institutions. Although AJR (2012) dismissed Albouy's comments, to remove any source of concern regarding the IV results of Section 5, in this appendix I use log of population density in 1500 as an instrument for institutions and human capital. According to AJR (2002), the density of indigenous populations at the time of colonization is among the most important determinants of colonization strategies. This is because colonizers generally didn't settle in those region that were relatively rich in natural resources and were densely populated during the Malthusian era (e.g. Aztec and Inca empires), and followed an extractive strategy to exploit those regions, thereby creating a "reversal of fortune".

Tables A1 and A2 report the results for social infrastructure and schooling, respectively. The OLS estimates in Panel C of both tables suggests that, when genetic distance to the United States is entered alone, its coefficient is significant at -12.191 (standard error=2.326) and -12.766 (standard error=2.383), in Tables A1 and A2, respectively. With an exception of column (2)-Panel A of Table A1, the result of second stage of IV regressions suggest, however, that genetic distance to the United States loses its explanatory power once I control for institutions and human capital. These results suggest that the statistical insignificance of genetic distance to the United States, and the statistical significance and the magnitude of the effect of institutions and schooling do not depend on using log of settler mortality as instrument.

The estimates in Panel B of both tables suggest that, unless the set of control variables are entirely excluded from the models (like in column 2 of both tables), genetic distance is not a significant determinant of institutions and schooling. And finally, Panel C of both tables demonstrates that in most of the specifications, genetic distance to the United States is not a statistically significant determinant of economic performance of former colonies, even when one ignores the endogeneity of institutions and schooling.

Table A1: IV Regression of Income per Capita / Instrumented Variable: Social Infrastructure

Dependent Variable: Log of Real GDP per Capita, 2000 (Penn World Tables) / Sample: Former Colonies						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Two-Stage Least Square / Instrument: Log Population Density, 1500						
Genetic distance to the United States		-4.632** (2.168)	-0.513 (3.768)	2.105 (3.341)	2.708 (3.560)	1.508 (4.159)
Social infrastructure		6.300*** (1.291)	7.160*** (1.440)	6.136*** (1.415)	6.377*** (1.564)	6.249*** (1.595)
Panel B: First Stage for Social Infrastructure, 1986-1995						
Genetic distance to the United States		-1.054*** (0.343)	0.083 (0.587)	0.376 (0.646)	0.194 (0.674)	-0.234 (0.932)
Log population density, 1500		-0.050*** (0.015)	-0.051*** (0.015)	-0.060*** (0.019)	-0.056*** (0.019)	-0.053*** (0.019)
R ² (first stage)		0.268	0.359	0.468	0.487	0.521
Panel C: Ordinary Least Square						
Genetic distance to the United States	-12.191*** (2.326)	-7.529*** (1.834)	-0.465 (3.142)	3.008 (3.167)	2.855 (3.406)	-0.199 (3.029)
Social infrastructure		3.886*** (0.411)	3.354*** (0.418)	2.951*** (0.388)	2.912*** (0.423)	2.615*** (0.430)
Observations	73	73	73	73	73	73
Geographical controls	-	-	Yes	Yes	Yes	Yes
Continent fixed effects	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: Panel A report the result of the second stage of the instrumental variable regression, where social infrastructure is instrumented with log of European settler mortality. Panel B reports to the results of the corresponding first stage, and Panel C report the result of Ordinary Least Square regressions. Geographical controls are log of the ancestry adjusted timing of Neolithic Transition, log percentage of arable land, a landlock dummy, and the absolute value of latitude. Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

Table A2: IV Regression of Income per Capita / Instrumented Variable: Schooling

Dependent Variable: Log of Real GDP per Capita, 2000 (Penn World Tables) / Sample: Former Colonies						
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Two-Stage Least Square / Instrument: Log Population Density, 1500						
Genetic distance to the United States		-2.733 (2.503)	-5.441 (3.313)	-5.618 (4.041)	-5.599 (4.101)	0.861 (4.466)
Years of schooling, 1960-2000		0.463*** (0.093)	0.536*** (0.094)	0.633*** (0.131)	0.643*** (0.146)	0.651*** (0.163)
Panel B: First Stage for Years of Schooling, 1960-2000						
Genetic distance to the United States		-18.831*** (4.414)	1.850 (8.333)	12.701 (9.078)	11.575 (9.054)	-5.742 (9.703)
Log population density, 1500		-0.680*** (0.179)	-0.725*** (0.166)	-0.613*** (0.132)	-0.566*** (0.145)	-0.537*** (0.145)
R ² (first stage)		0.425	0.521	0.721	0.742	0.807
Panel C: Ordinary Least Square						
Genetic distance to the United States	-12.766*** (2.383)	-5.347** (2.236)	-6.226** (3.002)	-3.128 (3.162)	-3.167 (3.176)	-1.777 (4.364)
Years of schooling, 1960-2000		0.343*** (0.046)	0.323*** (0.047)	0.363*** (0.053)	0.344*** (0.053)	0.337*** (0.053)
Observations	63	63	63	63	63	63
Geographical controls	-	-	Yes	Yes	Yes	Yes
Continent fixed effects	-	-	-	Yes	Yes	Yes
Legal origin fixed effects	-	-	-	-	Yes	Yes
Major religion shares	-	-	-	-	-	Yes

Notes: Panel A report the result of the second stage of the instrumental variable regression, where average years of schooling is instrumented with log of population density in 1500. Panel B reports to the results of the corresponding first stage, and Panel C report the result of Ordinary Least Square regressions. Geographical controls are log of the ancestry adjusted timing of Neolithic Transition, log percentage of arable land, a landlock dummy, and the absolute value of latitude. Robust standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

Bilateral Regressions

As mentioned in Section 2, SW argue that *if* their key hypothesis is correct, i.e., if genetic distance to the United States is a barrier to the diffusion of development from the world technological frontier, then a measure of relative genetic distance from the United States should have more explanatory power for contemporary income per capita, compared to a measure of absolute genetic distance. They define the absolute genetic distance between countries i and j as G_{ij}^D , which refers to the expected elapsed time since population of countries i and j shared a common ancestor; and the relative genetic distance from the United States as $G_{ij}^R = |G_{i,US}^D - G_{j,US}^D|$, where $D_{i,US}$ and $D_{j,US}$ denote genetic distance of country i and j to the United States, respectively. SW follow a bilateral approach in their econometric analysis and calculate income differences between all 9316 pairs of countries (based on 137 countries) in their sample, and show that, in regressions that use the absolute value of income differences ($|Y_i - Y_j|$) as the dependent variable, the magnitude of the effect of G_{ij}^R is bigger than G_{ij}^D . They take this difference in the magnitude of the effect of relative and absolute genetic distance on income differences as evidence in favour of the validity of their key hypothesis.

In this section I estimate the effect of relative genetic distance on pairwise income differences in the sample of former colonies, controlling for the differences in European settler mortality. The dependent variable is the absolute value of income differences between 2278 pairs of countries (based on 68 countries in the sample of former colonies). All of the other independent variables of the model are calculated as pairwise differences. As noted in SW, a problem with bilateral specification is the spatial correlation resulting from the construction of the dependent variable. To address this problem, I follow SW and rely on two-way clustering of standard errors (see Cameron, Gelbach, and Miller, 2006 for more information).

Table A4 presents the results. Column (1) relative genetic distance is the only independent variable, and its coefficient is significant at 1 percent. Note that this coefficient is comparable to the reported result in SW (Table III-column (4), p.496), they report a coefficient of 6.357 (standard error=0.996) for a sample that includes both former colonies and non-colonies. Column (2) adds the absolute differences in log of European settler mortality. The coefficient on genetic distance remains significant at 5 percent, but its magnitude decreases by 43 percent. Column (3) includes the geographical controls and shows that the relative genetic distance is now statistically insignificant. Column (4) suggests that this result do not change once I control for the difference in major religion shares. Overall, the results of bilateral analysis are consistent with my argument in Section 3 and further confirm the empirical results of Section 5.

Table A4: Bilateral Regression of Income per Capita.

Dependent Variable: Log Real GDP per Capita, 2000 (Penn World Tables) / Sample: Former Colonies				
	(1)	(2)	(3)	(4)
Relative genetic distance	6.458*** (1.945)	3.689** (1.563)	1.512 (1.484)	1.392 (1.456)
Log European settler mortality		0.437*** (0.066)	0.358*** (0.062)	0.362*** (0.061)
Log Neolithic transition timing (ancestry adjusted)			0.445* (0.233)	0.440* (0.231)
Log arable land area			0.077*** (0.027)	0.078*** (0.027)
Latitude (absolute value)			0.010** (0.005)	0.010** (0.005)
Percent Protestant				0.001 (0.003)
Percent Catholic				0.002 (0.001)
Percent Muslim				0.001 (0.002)
Observations	2278	2278	2278	2278
R-squared	0.069	0.270	0.317	0.323

Notes: Relative genetic distance is the absolute value of the difference in the genetic distance of countries to the United States. Two-way clustered standard errors are reported in parentheses. Data definition and sources are presented in Appendix C.

*** Significant at 1 percent.

** Significant at 5 percent.

* Significant at 10 percent.

B The Effect of Colonization on Genetic Distance and Economic Performance

Figure A1 illustrates the argument of Section 3.2, diagrammatically. Take two hypothetical countries, i and j , with country i having a lower genetic distance to Western Europe in 1500 ($D_{i0} < D_{j0}$). The panel at the left-hand side of the dashed line marked by 1500 refers to the historical path of country i , and the panel at the right-hand side of this line refers to the historical path of country j . In both panels, the direction of arrows refers to the direction of causal effect between variables. Remember that Figure 3-3 suggests that, *on average*, Europeans faced a higher mortality rates in former colonies with higher genetic distance to Europeans in 1500. Therefore, the lines connecting genetic distance in 1500, D_{k0} , to European settler mortality, M_k for $k=\{i, j\}$, refer to this positive correlation (therefore on average $M_i < M_j$). The arrows connecting settler mortality (M) to institutions and contemporary economic performance (Y) reconstruct AJR's argument regarding the colonial origins of comparative development: Europeans settled with more (less) intensity in those geographic regions in which they had lower (higher) rate of mortality, installed inclusive (extractive) institutions that led to good (bad) contemporary economic performance.

Map 2 and Figure 3-4 refer to the effect of colonization on the contemporary genetic distance of former colonies to the United States. Take country i for example. Low rate of European mortality in this country

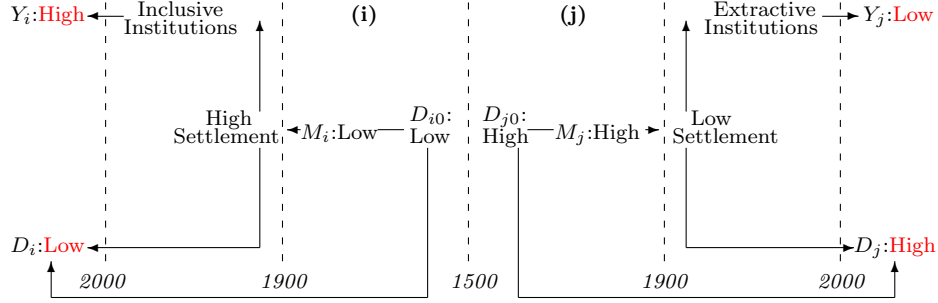


Figure A1. Simultaneous Effect of Colonization on Genetic Distance and Economic Performance.
Notes: D_{i0} and D_i refer to genetic distance of population of country i to Western Europe and the United States in 1500 and 2000, respectively. M_i refers to European settler mortality rate in country i , and Y_i refers to contemporary economic performance e.g., income per capita of country i .

(M_i :Low) resulted in relatively high settlement by Europeans during colonization period. This would result in country i to be currently genetically closer to Western Europe and the United States (D_i :Low). The exact opposite argument holds for country j , resulting in D_j to be bigger than D_i . Finally, since genetic distance is a measure persistence of in nature (the correlation between contemporary genetic distance to the United States and genetic distance to Western Europe in 1500 is 0.8) two arrows directly connect D_{i0} and D_{j0} to D_i and D_j .

Therefore, if one estimates the effect of genetic distance on contemporary income per capita in an equation like $Y_k = \alpha + \beta D_k + x'_k \gamma + \epsilon_k$ for $k=\{i, j\}$, where α and ϵ_k are the intercept and the error term respectively, and x_k is the vector of all other covariates of income (for example the set of control variables used in SW), the estimated coefficient, β , will be negative and possibly large. This is because the pairs (Y_i, Y_j) :(High,Low) and (D_i, D_j) :(Low,High) move in opposite directions. Further, this correlation should remain robust to controlling for various historical, geographical, linguistic, and religious controls as long as x_k does not include variables that are proper proxies of the effect of colonization on genetic distance and economic performance.

Figure A1 shows that, without addressing the simultaneous effect of colonization on genetic distance and economic performance, the effect all other factors that are related to colonization and affecting economic development (e.g. institutions or human capital) are embedded genetic distance to the United States. In other words, $\text{Cov}(D_k, \epsilon_k) \neq 0$ for $k=\{i, j\}$. Therefore, a causal interpretation of the effect of genetic distance on economic performance is not valid in this case.

C Data Definitions and Sources

Variable	Definition and Source
Genetic distance to the United States	The F_{ST} genetic distance, as reported by Spolaore and Wacziarg (2009), between the contemporary national populations of a given country and the United States, calculated as the average pairwise genetic distance across all ethnic group pairs, where each pair comprises two distinct ethnic groups, one from each country, and is weighted by the product of the proportional representations of the two groups in their respective national populations. <i>Source:</i> Spolaore and Wacziarg (2009).
Genetic distance to Western Europe (1500)	The F_{ST} genetic distance, as reported by Spolaore and Wacziarg (2009), between the year 1500 CE populations of a given country and the United Kingdom, calculated as the genetic distance between the two ethnic groups comprising the largest shares of each country's population in the year 1500 CE. However, since genetic distance of Belgium, France, Iceland, Ireland, Luxembourg and Spain to British population in 1500 is 0 and genetic distance of Austria, Denmark, Germany, Netherlands, Norway, Sweden and Switzerland is 0.0021 (which is very small in 0-0.2288 scale), the genetic distance of countries to British population in 1500 could be interpreted as their genetic distance to Western Europe in 1500. <i>Source:</i> Spolaore and Wacziarg (2009).
Government Effectiveness	Captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. <i>Source:</i> WGI, http://info.worldbank.org/governance/wgi/resources.htm .
Social Infrastructure	An index, calculated by Hall and Jones (1999), that quantifies the wedge between private and social returns to productive activities. To elaborate, this measure is computed as the average of two separate indices. The first is a government anti-diversion policy (GADP) index, based on data from the International Country Risk Guide, that represents the average across five categories, each measured as the mean over the 1986-1995 time period: (i) law and order, (ii) bureaucratic quality, (iii) corruption, (iv) risk of expropriation, and (v) government repudiation of contracts. The second is an index of openness, based on Sachs and Warner (1995), that represents the fraction of years in the time period 1950-1994 that the economy was open to trade with other countries, where the criteria for being open in a given year includes: (i) non-tariff barriers cover less than 40% of trade, (ii) average tariff rates are less than 40%, (iii) any black market premium was less than 20% during the 1970s and 80s, (iv) the country is not socialist, and (v) the government does not monopolize over major exports. <i>Source:</i> Hall and Jones (1999).
Log European Settler Mortality	Log of mortality rate faced by European settlers at the time of colonization. <i>Source:</i> Acemoglu, Johnson, and Robinson (2001).
Log Population Density	Log of total population divided by total arable land in 1500 AD. <i>Source:</i> McEvedy and Jones (1978) as cited in Acemoglu, Johnson, and Robinson (2002).
Percent European Settler	Percent of European resident of countries in 1900. <i>Source:</i> Acemoglu, Johnson, and Robinson (2001).
Percent European descent	The fraction of the year 2000 CE population (of the country for which the measure is being computed) that can trace its ancestral origins to the European continent due to migrations occurring as early as the year 1500 CE. <i>Source:</i> Putterman and Weil's (2010) World Migration Matrix.
Years of Schooling	The mean, over the 1960-2000 time period, of the 5-yearly figure, reported by Barro and Lee (2001), on average years of schooling amongst the population aged 25 and over. <i>Source:</i> Barro and Lee (2000) as cited in Glaeser et al. (2004)
GDP per Capita	Log of real GDP per capita in 2000, in Purchasing Power Parity (PPP) and in constant 2005 US\$. <i>Source:</i> Penn World Tables, https://pwt.sas.upenn.edu .
Change in GDP per Capita	Represent change in log of real GDP per capita from 1500-2000. Data on income in 1500 is from Maddison (2003) <i>Source:</i> Acemoglu, Johnson, Robinson, and Yared (2008).
Timing of Neolithic Transition	The cross-country weighted average of Neolithic transition timing, where the weight associated with a given country in the calculation represents the fraction of the year 2000 CE population (of the country for which the measure is being computed) that can trace its ancestral origins to the given country in the year 1500 CE. The ancestry weights are obtained from the World Migration Matrix, 1500-2000 of Putterman and Weil (2010). <i>Source:</i> Ashraf and Galor (2013).
Arable land area	Log of The fraction of a country's total land area that is arable, as reported by the World Bank's World Development Indicators. <i>Source:</i> Ashraf and Galor (2013).

C Data Definitions and Sources (Continued)

Variable	Definition and Source
Domesticable plants	The number of annual and perennial wild grass species, with a mean kernel weight exceeding 10 mg, that were prehistorically native to the region to which a country belongs. <i>Source:</i> Olsson and Hibbs (2005).
Domesticable animals	The number of domesticable large mammalian species, weighing in excess of 45 kg, that were prehistorically native to the region to which a country belongs. <i>Source:</i> Olsson and Hibbs (2005).
Legal Origins	A set of dummy variables, reported by La Porta et al. (1999), that identifies the legal origin of the Company Law or Commercial Code of a country. The five legal origin possibilities are: (i) English Common Law, (ii) French Commercial Code, (iii) German Commercial Code, (iv) Scandinavian Commercial Code, and (v) Socialist or Communist Laws. <i>Source:</i> Ashraf and Galor (2013).
Major Religion Shares	A set of variables, from La Porta et al. (1999), that identifies the percentage of a country's population belonging to the three most widely spread religions of the world. The religions identified are: (i) Roman Catholic, (ii) Protestant, and (iii) Muslim. <i>Source:</i> Ashraf and Galor (2013).

References

- Acemoglu, Daron, James Robinson.** 2012. *Why Nations Fail*. New York, NY: Crown Business.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369-1401.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics*, 117(4): 1231-194.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2005. "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American Economic Review*, 95(3): 546-79.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2005. "Institutions as a Fundamental Cause of Long-Run Growth." In *Handbook of Economic Growth*. Vol 1A, edited by Philippe Aghion and Steven N. Durlauf, 385-472. Amsterdam: North-Holland.
- Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared (2008).** "Income and Democracy." *American Economic Review*, 98: 808-42.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg.** 2003. "Fractionalization." *Journal of Economic Growth*, 8, 55-194.
- Angeles, Luis.** 2012. "Is There a Role for Genetics in Economic Development?" accessible at: <http://www.gla.ac.uk>.
- Ashraf, Quamrul, and Oded Galor.** 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review* 101 (5): 2003-41.
- 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.
- Bockstette, Valerie, Areendam Chanda, and Louis Putterman.** 2002. "States and Markets: The Advantage of an Early Start." *Journal of Economic Growth* 7 (4): 347-69.
- Callaway, Ewen.** 2012. "Economics and Genetic Meet in Uneasy Union." *Nature*, 490(11): 154-55.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller.** 2008. "Robust Inference with Multi-Way Clustering." NBER Technical Working Paper No. T0327.
- Campbell, Douglas L. and Ju Hyun Pyun.** 2011. "The Diffusion of Development: Along Genetic or Geographic Lines?". MPRA Paper No. 35178.
- Comin, Diego, William Easterly, and Erick Gong.** 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics* 2 (3): 659-7.
- Desmet, Klaus, Michel Le Breton, Ignacio Ortuno Ortin, and Shlomo Weber.** 2007. "Stability of Nations

and Genetic Diversity.” working paper, Universidad Carlos III.

Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. New York, NY: Norton.

Easterly, William, and Ross Levine. 2012. “The European Origins of Comparative Development.” NBER Working Paper No. 18162.

Engerman, Stanley L. and Sokoloff, Kenneth L. 1997. “Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies.” in Stephen Haber, ed., *How Latin America fell behind*. Stanford, CA: Stanford University Press, pp. 260-304.

Fish, Steven M. 2002. “Islam and Authoritarianism.” *World Politics*, 55(1): 4-37.

Gallup, John L., Andrew D. Mellinger, and Jeffrey D. Sachs. 1998. “Geography and Economic Development.” NBER Working Paper No. 6849.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. “Do Institutions Cause Growth? *Journal of Economic Growth*” *Journal of Economic Growth* 9 (3): 271-303.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. “Cultural Biases in Economic Exchange.” NBER Working Paper No. 11005.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. “Does Culture Affect Economic Outcomes?.” *Journal of Economic Perspectives*, 20, 2348.

Hall, Robert E., and Charles I. Jones. 1999. “Why Do Some Countries Produce So Much More Output Per Worker Than Others?” *Quarterly Journal of Economics* 114 (1): 831-16.

Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman OK: University of Oklahoma Press.

Knack, Stephen and Keefer, Philip. 1995. “Institutions and Economic Performance: Cross-Country Tests Using Alternative Measures.” *Economics and Politics*, N7(3), pp. 207-27.

La Porta, Rafael; Lopez-de-Silanes, Florencio; Shleifer, Andrei and Vishny, Robert W. 1998. “Law and Finance.” *Journal of Political Economy*, 106(6), pp. 1113-55.

La Porta, Rafael; Lopez-de-Silanes, Florencio; Shleifer, Andrei and Vishny, Robert W. 1999. “The Quality of Government.” *Journal of Law, Economics, and Organization*, 15(1), pp. 222-79.

North, Douglass C. 1981. *Structure and Change in Economic History*. New York: W.W. Norton & Co.

North, Douglass C., and Robert P. Thomas. 1973. *The Rise of the Western World: A New Economic History*. Cambridge: Cambridge University Press.

Olsson, Ola, and Douglas A. Hibbs. 2005. “Biogeography and Long-Run Economic Development.” *European Economic Review* 49 (4): 909-38.

Papaioannou, Elias, and Gregorios Siourounis. 2013. “Pre-Colonial Ethnic Institutions and Contemporary African Development.” *Econometrica*, 81(1): 113-52.

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): 1627-82.

Rodrik, Dani. 1999. “Where Did All the Growth Go?” *Journal of Economic Growth*, 4(4), pp. 385-412.

Sachs, Jeffrey. 2001 “Tropical Underdevelopment.” NBER Working Paper No. 8119.

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.” NBER Working Paper No. 17271.

Spolaore, Enrico, and Romain Wacziarg. 2013. “How Deep Are the Roots of Economic Development?” *Journal of Economic Literature* 51(2): 325-369.

von Hayek, Frederich A. 1960. *The constitution of liberty*. Chicago: University of Chicago Press.

Weber, Max. 1930. *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons and Anthony Giddens. London: Allen and Unwin.