

Ashraf, Quamrul; Galor, Oded

**Working Paper**

## The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development

Working Paper, No. 2010-7

**Provided in Cooperation with:**

Department of Economics, Brown University

*Suggested Citation:* Ashraf, Quamrul; Galor, Oded (2010) : The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development, Working Paper, No. 2010-7, Brown University, Department of Economics, Providence, RI

This Version is available at:

<https://hdl.handle.net/10419/62634>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The “Out of Africa” Hypothesis, Human Genetic Diversity, and Comparative Economic Development\*

Quamrul Ashraf  
Williams College

Oded Galor  
Brown University

August, 2010

## Abstract

This research argues that deep-rooted factors, determined tens of thousands of years ago, had a significant effect on the course of economic development from the dawn of human civilization to the contemporary era. It advances and empirically establishes the hypothesis that, in the course of the exodus of *Homo sapiens* out of Africa, variation in migratory distance from the cradle of humankind to various settlements across the globe affected genetic diversity and has had a long-lasting effect on the pattern of comparative economic development that is not captured by geographical, institutional, and cultural factors. In particular, the level of genetic diversity within a society is found to have a hump-shaped effect on development outcomes in both the pre-colonial and the modern era, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. While the intermediate level of genetic diversity prevalent among Asian and European populations has been conducive for development, the high degree of diversity among African populations and the low degree of diversity among Native American populations have been a detrimental force in the development of these regions. Further, the optimal level of diversity has increased in the process of industrialization, as the beneficial forces associated with greater diversity have intensified in an environment characterized by more rapid technological progress.

*Keywords:* The “Out of Africa” hypothesis, Human genetic diversity, Comparative development, Income per capita, Population density, Neolithic Revolution, Land productivity

*JEL Classification Numbers:* N10, N30, N50, O10, O50, Z10

---

\*We are grateful to Alberto Alesina, Kenneth Arrow, Alberto Bisin, Dror Brenner, John Campbell, Steve Davis, David Genesove, Douglas Gollin, Sergiu Hart, Saul Lach, Ross Levine, Ola Olsson, Mark Rosenzweig, Antonio Spilimbergo, Enrico Spolaore, Alan Templeton, Romain Wacziarg, David Weil, and seminar participants at Bar-Ilan, Ben-Gurion, Brown, Boston College, Chicago GSB, Copenhagen, Doshisha, Haifa, Harvard, Hebrew U, Hitotsubashi, IMF, Keio, Kyoto, Osaka, Tel Aviv, Tokyo, Tufts, UCLA Anderson, Williams, the World Bank, as well as conference participants at the CEPR “UGT Summer Workshop”, the NBER meeting on “Macroeconomics Across Time and Space”, the KEA “International Employment Forum”, the SED Annual Meeting, the NBER Summer Institute, and the NBER Political Economy meeting. We also thank attendees of the Klein Lecture, Osaka, the Kuznets Lecture, Yale, and the Nordic Doctoral Program for helpful comments and suggestions. We are especially indebted to Yona Rubinstein for numerous insightful discussions, and to Sohini Ramachandran for sharing her data with us. Financial support from the Watson Institute at Brown University is gratefully acknowledged. Galor’s research is supported by NSF grant SES-0921573.

# 1 Introduction

Existing theories of comparative development highlight a variety of proximate and ultimate factors underlying some of the vast inequities in living standards across the globe. The importance of geographical, cultural and institutional factors, human capital formation, ethnic, linguistic, and religious fractionalization, colonialism and globalization has been at the center of a debate regarding the origins of the differential timing of transitions from stagnation to growth and the remarkable transformation of the world income distribution in the last two centuries. While theoretical and empirical research has typically focused on the effects of such factors in giving rise to and sustaining the Great Divergence in income per capita in the pre-industrial era, attention has recently been drawn towards some deep-rooted factors that have been argued to affect the course of comparative economic development.

This paper argues that deep-rooted factors, determined tens of thousands of years ago, have had a significant effect on the course of economic development from the dawn of human civilization to the contemporary era. It advances and empirically establishes the hypothesis that, in the course of the exodus of *Homo sapiens* out of Africa, variation in migratory distance from the cradle of humankind in East Africa to various settlements across the globe affected genetic diversity and has had a long-lasting hump-shaped effect on the pattern of comparative economic development that is not captured by geographical, institutional, and cultural factors. Further, the optimal level of diversity appears to have increased in the process of industrialization, as the beneficial forces associated with greater diversity have intensified in an environment characterized by more rapid technological progress.

Consistent with the predictions of the theory, the empirical analysis finds that the level of genetic diversity within a society has a hump-shaped effect on development outcomes in the pre-colonial as well as in the modern era, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. While the intermediate level of genetic diversity prevalent among the Asian and European populations has been conducive for development, the high degree of diversity among African populations and the low degree of diversity among Native American populations have been a detrimental force in the development of these regions. In addition, the empirical findings suggest that, indeed, the optimal level of diversity has increased in the course of industrialization. While the optimal level of diversity in the year 1500 CE corresponded to that prevalent in the Far East, the optimal level in the year 2000 CE corresponds to the level present in the U.S. This paper thus highlights one of the deepest channels in comparative development, pertaining not to factors associated with the dawn of complex agricultural societies as in Diamond's (1997) influential hypothesis, but to conditions innately related to the very dawn of mankind itself.

The hypothesis rests upon two fundamental building blocks. First, migratory distance from the cradle of humankind in East Africa had an adverse effect on the degree of genetic diversity within ancient indigenous settlements across the globe. Following the prevailing hypothesis, commonly known as the *serial-founder effect*, it is postulated that, in the course of human expansion over planet Earth, as subgroups of the populations of parental colonies left to establish new settlements further away, they carried with them only a subset of the overall genetic diversity of their parental colonies. Indeed, as depicted in Figure 1, migratory distance from East Africa has an adverse effect on genetic diversity in the 53 ethnic groups across the globe that constitute the Human Genome Diversity Cell Line Panel.

Second, there exists an optimal level of diversity for each stage of development reflecting the interplay between conflicting effects of diversity on the development process. The adverse effect

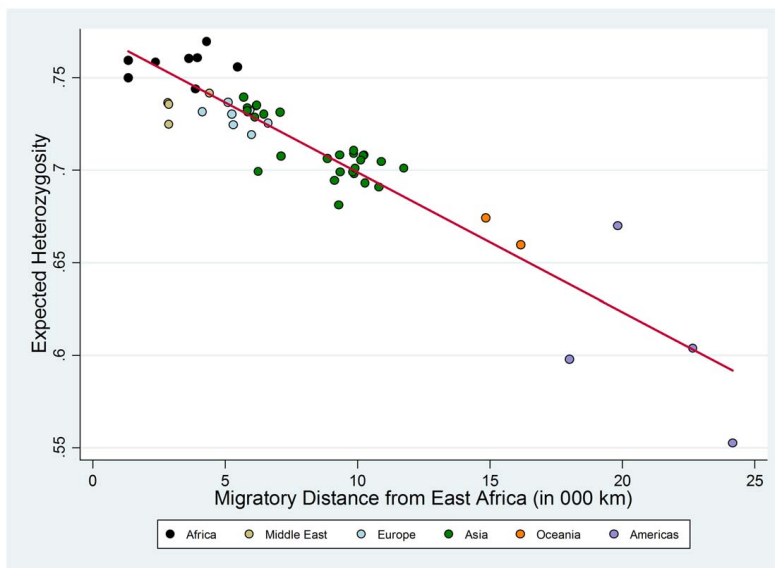


FIGURE 1: Expected Heterozygosity and Migratory Distance in the HGDP-CEPH Sample

pertains to the detrimental impact of diversity on the efficiency of the aggregate production process of an economy. Heterogeneity increases the likelihood of mis-coordination and distrust, reducing cooperation and disrupting the socioeconomic order. Greater population diversity is therefore associated with the social cost of a lower total factor productivity, which inhibits the ability of society to operate efficiently with respect to its production possibility frontier.

The beneficial effect of diversity, on the other hand, concerns the positive role of diversity in the expansion of society's production possibility frontier. A wider spectrum of traits is more likely to be complementary to the development and successful implementation of advanced technological paradigms. Greater heterogeneity therefore fosters the ability of a society to incorporate more sophisticated and efficient modes of production, expanding the economy's production possibility frontier and conferring the benefits of increased total factor productivity.<sup>1</sup>

Higher diversity in a society's population can therefore have conflicting effects on the level of its total factor productivity. Aggregate productivity is enhanced on the one hand by an increased capacity for technological advancement, while simultaneously diminished on the other by reduced cooperation and efficiency. However, if the beneficial effects of population diversity dominate at lower levels of diversity and the detrimental effects dominate at higher levels (i.e., if there are diminishing marginal returns to both diversity and homogeneity), the theory would predict an inverted-U relationship between genetic diversity and development outcomes over the course of the development process. Furthermore, the theory would also predict that the optimal level of diversity increases with the process of economic development, as the beneficial forces associated with greater population diversity become intensified in an environment characterized by more rapid technological progress.

In estimating the impact on economic development of migratory distance from East Africa via its effect on genetic diversity, this research overcomes limitations and potential concerns that are presented by the existing data on genetic diversity across the globe (i.e., measurement error,

<sup>1</sup>Indeed, this observation is broadly consistent with theoretical and empirical evidence on the creativity-promoting effects of diversity in the workforce (see, e.g., Alesina and La Ferrara, 2005).

data limitations, and potential endogeneity). Population geneticists typically measure the extent of diversity in genetic material across individuals within a given population (such as an ethnic group) using an index called *expected heterozygosity*. Like most other measures of diversity, this index may be interpreted simply as the probability that two individuals, selected at random from the relevant population, are genetically different from one another. Specifically, the expected heterozygosity measure for a given population is constructed by geneticists using sample data on *allelic frequencies*, i.e., the frequency with which a “gene variant” or allele (e.g., the brown vs. blue variant for the eye color gene) occurs in the population sample.<sup>2</sup> Given allelic frequencies for a particular gene or DNA locus, it is possible to compute a gene-specific heterozygosity statistic (i.e., the probability that two randomly selected individuals differ with respect to the gene in question), which when averaged over multiple genes or DNA loci yields the overall expected heterozygosity for the relevant population.<sup>3</sup>

The most reliable and consistent data for genetic diversity among indigenous populations across the globe consists, however, of only 53 ethnic groups from the Human Genome Diversity Cell Line Panel. According to anthropologists, these groups are not only historically native to their current geographical location but have also been isolated from genetic flows from other ethnic groups. Empirical evidence provided by population geneticists (e.g., Ramachandran et al., 2005) for these 53 ethnic groups suggest that, indeed, migratory distance from East Africa has an adverse linear effect on genetic diversity as depicted in Figure 1. Migratory distance from East Africa for each of the 53 ethnic groups was computed using the great circle (or geodesic) distances from Addis Ababa (Ethiopia) to the contemporary geographic coordinates of these ethnic groups, subject to five obligatory intermediate waypoints (i.e., Cairo (Egypt), Istanbul (Turkey), Phnom Penh (Cambodia), Anadyr (Russia) and Prince Rupert (Canada)), that capture paleontological and genetic evidence on prehistorical human migration patterns.

Nonetheless, while the existing data on genetic diversity pertain only to ethnic groups, data for examining comparative development are typically available at the country level. Moreover, many national populations today are composed of multiple ethnicities, some of which may not be indigenous to their current geographical locations. This raises two complex tasks. First, one needs to construct of a measure of genetic diversity for national populations, based on genetic diversity data at the ethnic group level, accounting for diversity not only *within* each component group but for diversity due to differences *between* ethnic groups as well. Second, it is necessary to account for the potential inducement for members of distinct ethnic groups to relocate to relatively more lucrative geographical locations.

To tackle these difficulties, this study adopts two distinct strategies. The first restricts attention to development outcomes in the pre-colonial era when, arguably, regional populations were indigenous to their current geographical location and largely homogenous in terms of their ethnic compositions, with the presence of multiple indigenous ethnicities in a given region having a negligible effect on the diversity of the regional population. The second, more complex strategy involves the construction of an index of genetic diversity for contemporary national populations that accounts for the expected heterozygosity within each sub-national group as well as the additional component of diversity at the country level that arises from the genetic distances between its pre-

---

<sup>2</sup>In molecular genetics, an allele is defined as any one of a number of viable DNA codings (formally, a sequence of nucleotides) that occupy a given locus (or position) in a chromosome. Chromosomes themselves are “packages” for carrying strands of DNA molecules in cells and comprise multiple loci that typically correspond to some of the observed discrete “units of heredity” (or genes) in living organisms. For further elaboration on basic concepts and definitions in genetics, the interested reader is referred to Griffiths et al. (2000).

<sup>3</sup>See Weir (1996) for the statistical theory underlying measures of genetic diversity.

colonial ancestral populations. The examination of comparative development under this second strategy would have to additionally account for the potential inducement for members of distinct ethnic groups to relocate to relatively more lucrative geographical locations.

The examination of comparative development in the pre-colonial era, when societies were in their agricultural stage of development, requires the interpretation of outcomes from a Malthusian equilibrium point of view. Improvements in the technological environment during the Malthusian epoch generated only temporary gains in income per capita, eventually leading to a larger, but not richer, population (Ashraf and Galor, 2010). Thus the relevant variable gauging comparative economic development during this era is population density as opposed to income per capita. In light of this argument, this study employs cross-country historical data on population density as the outcome variable of interest in the historical analysis and examines the hypothesized effect of human genetic diversity within societies on their population densities in the years 1 CE, 1000 CE and 1500 CE.<sup>4</sup>

Using data on genetic diversity observed at the ethnic group level, the historical analysis reveals, consistently with the proposed hypothesis, a highly significant hump-shaped effect of genetic diversity on log population density in the year 1500 CE. In particular, accounting for the influence of the timing of the Neolithic Revolution, the natural productivity of land for agriculture, as well as other geographical characteristics that may affect population density in the pre-industrial era, the estimated linear and quadratic coefficients associated with genetic diversity imply that a 1 percentage point increase in diversity for the least diverse society in the regression sample would raise its population density by 58.03%, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 23.36%. Despite the statistical significance and robustness of these effects, however, the analysis is subsequently expanded upon to lend further credence to these findings by alleviating concerns regarding sample size limitations and potential endogeneity bias.

The issue of data limitations encountered by the analysis stems from the fact that diversity data at the ethnic group level currently spans only a modest subset of the sample of countries for which historical population estimates are available. The potential endogeneity issue, on the other hand, arises from the possibility that genetic diversity within populations could partly reflect historical processes such as interregional migrations that were, in turn, determined by historical patterns of comparative development. Furthermore, the direction of the potential endogeneity bias is *a priori* ambiguous. For example, while historically better developed regions may have been attractive destinations to potential migrants, serving to increase genetic diversity in relatively wealthier societies, the more advanced technologies in these societies may also have conferred the necessary military prowess to prevent or minimize foreign invasions, thereby reducing the likelihood of greater genetic diversity in their populations.<sup>5</sup>

In surmounting the aforementioned data limitations and potential endogeneity issues, this research appeals to the “out of Africa” theory regarding the origins of *Homo sapiens*. According

---

<sup>4</sup>Admittedly, historical data on population density likely suffers from mismeasurement as well. However, while measurement error in explanatory variables leads to attenuation bias in OLS estimators, mismeasurement of the dependent variable in an OLS regression has the less serious consequence of yielding larger standard errors, a result that works against rejecting the “null hypothesis”. This statistical symptom, however, further strengthens the “alternative hypothesis” if the relevant coefficient estimates are statistically significant despite the mismeasurement of the dependent variable.

<sup>5</sup>The history of world civilization is abound with examples of both phenomena. The “Barbarian Invasions” of the Western Roman Empire in the Early Middle Ages is a classic example of historical population diffusion occurring along a prosperity gradient, whereas the The Great Wall of China, built and expanded over centuries to minimize invasions by nomadic tribes, serves (literally) as a landmark instance of the latter phenomenon.

to this well-established hypothesis, the human species, having evolved to its modern form in East Africa some 150,000 years ago, thereafter embarked on populating the entire globe in a stepwise migration process beginning about 70,000–90,000 BP.<sup>6</sup> Using archeological data combined with mitochondrial and Y-chromosomal DNA analysis to identify the most recent common ancestors of contemporary human populations, geneticists are able to not only offer evidence supporting the origin of humans in East Africa but also trace the prehistorical migration routes of the subsequent human expansion into the rest of the world.<sup>7</sup> In addition, population geneticists studying human genetic diversity have argued that the contemporary distribution of diversity across populations should reflect a serial-founder effect originating in East Africa. Accordingly, since the populating of the world occurred in a series of stages where subgroups left initial colonies to create new colonies further away, carrying with them only a portion of the overall genetic diversity of their parental colonies, contemporary genetic diversity in human populations should be expected to decrease with increasing distance along prehistorical migratory paths from East Africa.<sup>8</sup> Indeed, several studies in population genetics (e.g., Prugnolle et al., 2005; Ramachandran et al., 2005; Wang et al., 2007) have found strong empirical evidence in support of this prediction.

The present study exploits the explanatory power of migratory distance from East Africa for genetic diversity within ethnic groups in order to overcome the data limitations and potential endogeneity issues encountered by the initial analysis discussed above. In particular, the strong ability of prehistorical migratory distance from East Africa in explaining *observed* genetic diversity permits the analysis to generate *predicted* values of genetic diversity using migratory distance for countries for which diversity data are currently unavailable. This enables a subsequent analysis to estimate the effects of genetic diversity, as predicted by migratory distance from East Africa, in a much larger sample of countries. Moreover, given the obvious exogeneity of migratory distance from East Africa with respect to development outcomes in the period 1–1500 CE, the use of migratory distance to project genetic diversity alleviates concerns regarding potential endogeneity between *observed* genetic diversity and economic development.

The main results from the historical analysis, employing predicted genetic diversity in the extended sample of countries, indicate that, controlling for the influence of land productivity, the timing of the Neolithic Revolution, and continental fixed effects, a 1 percentage point increase in diversity for the most homogenous society in the sample would raise its population density in 1500 CE by 36.36%, whereas a 1 percentage point decrease in diversity for the most diverse

---

<sup>6</sup>An alternative to this “recent African origin” (RAO) model is the “multiregional evolution accompanied by gene flow” hypothesis, according to which early modern hominids evolved independently in different regions of the world and thereafter exchanged genetic material with each other through migrations, ultimately giving rise to a relatively uniform dispersion of modern *Homo sapiens* throughout the globe. However, in light of surmounting genetic and paleontological evidence against it, the multiregional hypothesis has by now almost completely lost ground to the RAO model of modern human origins (Stringer and Andrews, 1988).

<sup>7</sup>For studies accessible to a general audience, the reader is referred to Cavalli-Sforza et al. (1994), Cavalli-Sforza and Cavalli-Sforza (1995), Olson (2002), Wells (2002) and Oppenheimer (2003).

<sup>8</sup>In addition, population geneticists argue that the reduced genetic diversity associated with the founder effect is due not only to the subset sampling of alleles from parental colonies but also to a stronger force of “genetic drift” that operates on the new colonies over time. Genetic drift arises from the fundamental tendency of the frequency of any allele in an inbreeding population to vary randomly across generations as a result of random statistical sampling errors alone (i.e., the chance production of a few more or less progeny carrying the relevant allele). Thus, given the inherent “memoryless” (Markovian) property of allelic frequencies across generations as well as the absence of mutation and natural selection, the process ultimately leads to either a 0% or a 100% representation of the allele in the population (Griffiths et al., 2000). Moreover, since random sampling errors are more prevalent in circumstances where the law of large numbers is less applicable, genetic drift is more pronounced in smaller populations, thereby allowing this phenomenon to play a significant role in the founder effect.

society would raise its population density by 28.62%. Further, a 1 percentage point change in genetic diversity in either direction at the predicted optimum diversity level of 0.6832 would lower population density by 1.45%. Consistent with the predictions of the proposed hypothesis, the non-monotonic effect of genetic diversity on development outcomes is uncovered for earlier historical periods as well. Moreover, genetic diversity explains between 15% and 42% of the cross-country variation in log population density, depending on the historical period examined and the control variables included in the regression specification. Indeed, the impact of genetic diversity is robust to various regression specifications such as the inclusion of controls for the spatial influence of regional technological frontiers via trade and the diffusion of technologies, and controls for microgeographic factors gauging terrain quality and proximity to waterways.

Moving to the contemporary period, the analysis, as discussed earlier, constructs an index of genetic diversity at the country level that not only incorporates the expected heterozygosities of the pre-Columbian ancestral populations of contemporary sub-national groups, as predicted by the migratory distances of the ancestral populations from East Africa, but also incorporates the pairwise genetic distances between these ancestral populations, as predicted by their pairwise migratory distances. Indeed, the serial-founder effect studied by population geneticists not only predicts that expected heterozygosity declines with increasing distance along migratory paths from East Africa, but also that the genetic distance between any two populations will be larger the greater the migratory distance between them.

The baseline results from the contemporary analysis indicate that the genetic diversity of contemporary national populations has an economically and statistically significant hump-shaped effect on income per capita. Moreover, in line with the prediction that the beneficial impact of diversity has increased in the process of industrialization, the optimal level of diversity with respect to the modern world income distribution is higher than that obtained with respect to population density in the pre-colonial Malthusian era. The hump-shaped impact of diversity on income per capita is robust to continental fixed effects, and to controls for ethnic fractionalization and various measures of institutional quality, including an index gauging the extent of democracy, constraints on the power of chief executives, legal origins, and major religion shares, as well as to controls for years of schooling, disease environments, and other geographical factors that have received attention in the literature on cross-country comparative development.

The direct effect of genetic diversity on contemporary income per capita, once institutional, cultural, and geographical factors are accounted for, indicates that: (i) increasing the diversity of the most homogenous country in the sample (Bolivia) by 1 percentage point would raise its income per capita in the year 2000 CE by 38.63%, (ii) decreasing the diversity of the most diverse country in the sample (Ethiopia) by 1 percentage point would raise its income per capita by 20.52%, (iii) a 1 percentage point change in genetic diversity (in either direction) at the optimum level of 0.7208 (that most closely resembles the U.S. diversity level of 0.7206) would lower income per capita by 1.91%, (iv) increasing Bolivia's diversity to the optimum level prevalent in the U.S. would increase Bolivia's per capita income by a factor of 4.73, closing the income gap between the U.S. and Bolivia from 12:1 to 2.5:1, and (v) decreasing Ethiopia's diversity to the optimum level of the U.S. would increase Ethiopia's per capita income by a factor of 1.73 and, thus, close the income gap between the U.S. and Ethiopia from 47:1 to 27:1.

Reassuringly, the highly significant and stable hump-shaped effect of genetic diversity on income per capita in the year 2000 CE is not an artifact of post-colonial migrations towards prosperous countries and the concomitant increase in ethnic diversity in these economies. The hump-shaped effect of genetic diversity remains highly significant and the optimal diversity estimate



remains virtually intact if the regression sample is restricted to (a) non-OECD economies (i.e., economies that were less attractive to migrants), (b) non Neo-European countries (i.e., excluding the U.S., Canada, New Zealand and Australia), (c) non-Latin American countries, (d) non Sub-Saharan African countries, and perhaps most importantly (e) to countries whose indigenous population is larger than 97% of the entire population (i.e., under conditions that virtually eliminate the role of migration in the creation of diversity).

The remainder of the paper is organized as follows: Section 2 briefly reviews some related literature. Section 3 covers the historical analysis, discussing the empirical strategy as well as the relevant data and data sources before presenting the empirical findings. Section 4 does the same for the contemporary analysis, and, finally, Section 5 concludes.

## 2 Related Literature

The existing literature on comparative development has emphasized a variety of factors underlying some of the vast differences in living standards across the globe. The influence of geography, for instance, has been stressed from a historical perspective by Jones (1981), Diamond (1997) and Pomeranz (2000), and is highlighted empirically by Gallup et al. (1999) and Olsson and Hibbs (2005) amongst others. Institutions, on the other hand, are given historical precedence by North and Thomas (1973), Mokyr (1990), and Greif (1993), and are emphasized empirically by Hall and Jones (1999), La Porta et al. (1999), Glaeser and Shleifer (2002), Rodrik et al. (2004), and Acemoglu et al. (2005). In related strands of the literature on institutions, Engerman and Sokoloff (2000) and Acemoglu et al. (2005) have stressed the role of colonialism, while the effects of ethno-linguistic fractionalization are examined by Easterly and Levine (1997), Alesina et al. (2003), Montalvo and Reynal-Querol (2005) and others. Meanwhile, the historical impact of sociocultural factors has been highlighted by Weber (1905, 1922) and Landes (1998), with empirical support coming from Barro and McCleary (2003), Tabellini (2008) as well as Guiso et al. (2009). Finally, the importance of human capital formation has been underlined in the unified growth theories of Galor and Weil (2000), Galor and Moav (2002), Hansen and Prescott (2002), Lucas (2002), Lagerlöf (2003, 2006), Doepke (2004), Galor and Mountford (2006, 2008), Galor (2005), Galor et al. (2009) and Dalgaard and Strulik (2010), and has been demonstrated empirically by Glaeser et al. (2004).

This research is singular in its attempt to empirically establish the role of deep-rooted factors, determined tens of thousands of years ago on contemporary development. It is the first to argue that variation in migratory distance from the cradle of humankind to various settlements across the globe had a persistent effect on the process of development and on the contemporary variation in income per capita across the globe. The paper is also unique in its attempt to establish the role of genetic diversity genetic *within* a society as a significant determinant of its development path and, thus, its comparative economic performance across space and time.

Nevertheless, the employment of data and empirical results from research in population genetics places this paper in the neighborhood of some recent insightful papers in the economic literature (e.g., Guiso et al., 2009; Spolaore and Wacziarg, 2009) that have appealed to data on genetic distance between human populations to instrument or proxy for the effect of sociocultural differences between societies on technological diffusion and trade.<sup>9</sup> Spolaore and Wacziarg (2009) argue that genetic distance observed between populations captures their divergence in biological

---

<sup>9</sup>See also Desmet et al. (2006) who demonstrate a strong correlation between genetic and cultural distances among European populations to argue that genetic distance can be employed as an appropriate proxy to study the effect of cultural distance on the formation of new political borders in Europe.

and cultural characteristics (transmitted intergenerationally within a population over time), acting as a barrier to the horizontal diffusion of technological innovations across populations. The authors show that  $F_{st}$  genetic distance, which reflects the time elapsed since two populations shared a common ancestor, bears a statistically significant positive relationship with both historical and contemporary pairwise income differences. In particular, they find that a standard deviation in genetic distance accounts for 20–30% of a standard deviation in income differences, a result that remains robust after controlling for various geographical, linguistic and religious differences.<sup>10</sup> Guiso et al. (2009), on the other hand, employ data on genetic distance between European populations as an instrument for measures of trust to estimate its effect on the volume of bilateral trade and foreign direct investment, finding that a one standard deviation increase in genetic distance reduces the level of trust by about 27%.<sup>11</sup>

The employment of the genetic distance metric *between* populations in the earlier studies permitted the analysis of the effect of cultural (and biological) differences, proxied by genetic distances, on the degree of spillovers across societies. In addition, Spolaore and Wacziarg’s (2009) finding that income differences between societies are a function of their relative genetic distance from the world technological frontier implicitly invokes the notion of a hierarchy of traits, whereby the most complementary traits for economic development are those that are predominant in the population at the frontier. In contrast, the genetic diversity metric *within* populations exploited in this paper facilitates the analysis of the effect of the variation in traits across individuals within a society on its development process, regardless of society’s proximity to the global technological frontier. Hence, unlike previous studies where interdependence across societies through trade or technological diffusion is a necessary condition for the effect of human genetics on the process of economic development, the current research advances the novel hypothesis that genetic diversity within a society plays a significant role in its development path, independently of its position in the world economy. Moreover, the genetic channel proposed in this study is entirely orthogonal to conceptual frameworks that posit a hierarchy of genetic traits in terms of their conduciveness to the process of development.

Furthermore, unlike earlier studies where genetic distance between populations diminishes the rate of technological diffusion and reduces productivity, the hypothesis advanced and tested in this paper suggests that genetic diversity within a population confers both social costs, in the form of lower social capital arising from differences amongst individual members, and social benefits in the form of diversity-driven knowledge accumulation. Thus, the overall effect of genetic diversity on developmental outcomes would be hump-shaped, rather than monotonically negative. The results of the empirical analysis conducted in this study suggest that the previously unexamined beneficial effect of genetic differences is indeed a significant factor in the overall influence of the genetic channel on comparative development.

---

<sup>10</sup>The coefficient estimates obtained from regressing genetic distance on income differences in Spolaore and Wacziarg’s (2009) study remain almost unaffected in both magnitude and significance when subjected to controls for cultural distance, proxied for with a set of variables including common colonial history, linguistic distance as well as religious distance. While this could be regarded as evidence for a biological interpretation of their results, the authors argue that the “barriers” arising from differences in vertically transmitted characteristics are not primarily linguistic or religious in nature.

<sup>11</sup>It should be noted that Giuliano et al. (2006) have raised concerns regarding the use of genetic distance as either a proxy or an instrument for cultural differences in these studies, arguing that genetic distance, being strongly correlated with geographic distance, is really a proxy for transportation costs associated with geographical (as opposed to biological or sociocultural) barriers. Nevertheless, Spolaore and Wacziarg (2009) and Guiso et al. (2009) demonstrate that their results remain robust to controls for this alternative transportation cost hypothesis.

The examination of the effects of genetic diversity along with the influence of the timing of agricultural transitions also places this paper in an emerging strand of the literature that has focused on empirically testing Diamond’s (1997) assertion regarding the long-standing impact of the Neolithic Revolution. Diamond (1997) has stressed the role of biogeographical factors in determining the timing of the Neolithic Revolution, which conferred a developmental head-start to societies that experienced an earlier transition from primitive hunting and gathering techniques to the more technologically advanced agricultural mode of production. According to this hypothesis, the luck of being dealt a favorable hand thousands of years ago with respect to biogeographic endowments, particularly exogenous factors contributing to the emergence of agriculture and facilitating the subsequent diffusion of agricultural techniques, is the single most important driving force behind the divergent development paths of societies throughout history that ultimately led to the contemporary global differences in standards of living. Specifically, an earlier transition to agriculture due to favorable environmental conditions gave some societies an early advantage by conferring the benefits of a production technology that generated resource surpluses and enabled the rise of a non-food-producing class whose members were crucial for the development of written language and science, and for the formation of cities, technology-based military powers and nation states. The early technological dominance of these societies subsequently persisted throughout history, being further sustained by the subjugation of less-developed societies through exploitative geopolitical and historical processes such as colonization.

While the long-standing influence of the Neolithic Revolution on comparative development remains a compelling argument, this research demonstrates that, contrary to Diamond’s (1997) uncausal hypothesis, the composition of human populations with respect to their genetic diversity has been an significant and persistent factor that affected the course of economic development from the dawn of human civilization to the present. In estimating the economic impact of human genetic diversity while controlling for the channel emphasized by Diamond (1997), the current research additionally establishes the historical significance of the timing of agricultural transitions for pre-colonial population density, which, as already argued, is the relevant variable capturing comparative economic development during the Malthusian epoch of stagnation in income per capita.<sup>12</sup> Interestingly, however, unlike the conjecture of Diamond (1997), the timing of the Neolithic Revolution has no effect on contemporary income per capita.<sup>13</sup>

---

<sup>12</sup>Note that, although the genetic diversity channel raised in this study is conceptually independent of the timing of the agricultural transition, an additional genetic channel that interacts with the time elapsed since the Neolithic Revolution has been examined by Galor and Moav (2002, 2007). These studies argue that the Neolithic transition triggered an evolutionary process resulting in the natural selection of certain genetic traits (such as preference for higher quality children and greater longevity) that are complementary to economic development, thereby implying a *ceteris paribus* positive relationship between the timing of the agricultural transition and the representation of such traits in the population. Indeed, the empirical evidence recently uncovered by Galor and Moav (2007) is consistent with this theoretical prediction. Thus, while the significant reduced-form effect of the Neolithic Revolution observed in this study may be associated with the Diamond hypothesis, it could also be partly capturing the influence of this additional genetic channel. See also Lagerlöf (2007) for a complementary evolutionary theory regarding the dynamics of human body mass in the process of economic development.

<sup>13</sup>Olsson and Hibbs (2005) and Putterman (2008) have suggested that there is empirical support for the Diamond hypothesis in that the timing of the Neolithic revolution affected the contemporary variation in income per capita across the globe. However, as established in Table 12, these results are non-robust. Once the genetic diversity channel is included in the analysis, the (direct or indirect) effect of the Neolithic Revolution on contemporary outcomes becomes statistically insignificant.

## 3 The Historical Analysis

### 3.1 Data and Empirical Strategy

This section discusses the data and empirical strategy employed to examine the impact of genetic diversity on comparative development in the time period 1-1500 CE.

#### 3.1.1 Dependent Variable: Historical Population Density

As argued previously, the relevant variable reflecting comparative development across countries in the pre-colonial Malthusian era is population density. The empirical examination of the proposed genetic hypothesis therefore aims to employ cross-country variation in observed genetic diversity and in that predicted by migratory distance from East Africa to explain cross-country variation in historical population density. Data on historical population density are obtained from McEvedy and Jones (1978) who provide figures at the country level, i.e., for regions defined by contemporary national borders, over the period 400 BCE–1975 CE.<sup>14</sup> However, given the greater unreliability (and less availability in terms of observations) of population data for earlier historical periods, the baseline regression specification adopts population density in 1500 CE as the preferred outcome variable to examine. The analysis additionally examines population density in 1000 CE and 1 CE to demonstrate the robustness of the genetic channel for earlier time periods.

#### 3.1.2 Independent Variable: Genetic Diversity

The most reliable and consistent data for genetic diversity among indigenous populations across the globe consists of 53 ethnic groups from the Human Genome Diversity Cell Line Panel, compiled by the Human Genome Diversity Project-Centre d’Etudes du Polymorphisme Humain (HGDP-CEPH).<sup>15</sup> According to anthropologists, these 53 ethnic groups are not only historically native to their current geographical location but have also been isolated from genetic flows from other ethnic groups. Population geneticists typically measure the extent of diversity in genetic material across individuals within a given population (such as an ethnic group) using an index called *expected heterozygosity*. Like most other measures of diversity, this index may be interpreted simply as the probability that two individuals, selected at random from the relevant population, are genetically different from one another. Specifically, the expected heterozygosity measure for a given population is constructed by geneticists using sample data on *allelic frequencies*, i.e., the frequency with which a “gene variant” or allele occurs in the population sample. Given allelic frequencies for a particular gene or DNA locus, it is possible to compute a gene-specific heterozygosity statistic (i.e., the probability that two randomly selected individuals differ with respect to a given gene), which when averaged over multiple genes or DNA loci yields the overall expected heterozygosity for the relevant population.

Consider a single gene or locus  $l$  with  $k$  observed variants or alleles in the population and let  $p_i$  denote the frequency of the  $i$ -th allele. Then, the expected heterozygosity of the population with respect to locus  $l$ ,  $H_{\text{exp}}^l$ , is:

$$H_{\text{exp}}^l = 1 - \sum_{i=1}^k p_i^2. \quad (1)$$

---

<sup>14</sup>The reader is referred to Appendix B for additional details.

<sup>15</sup>For a more detailed description of the HGDP-CEPH Human Genome Diversity Cell Line Panel data set, the interested reader is referred to Cann et al. (2002). A broad overview of the Human Genome Diversity Project is given by Cavalli-Sforza (2005). The 53 ethnic groups are listed in Appendix A.

Given allelic frequencies for each of  $m$  different genes or loci, the average across these loci then yields an aggregate expected heterozygosity measure of overall genetic diversity,  $H_{\text{exp}}$ , as:

$$H_{\text{exp}} = 1 - \frac{1}{m} \sum_{l=1}^m \sum_{i=1}^{k_l} p_i^2, \quad (2)$$

where  $k_l$  is the number of observed variants in locus  $l$ .

Empirical evidence uncovered by Ramachandran et al. (2005) for the 53 ethnic groups from the Human Genome Diversity Cell Line Panel suggests that migratory distance from East Africa has an adverse linear effect on genetic diversity.<sup>16</sup> They interpret this finding as providing support for a serial-founder effect originating in East Africa, reflecting a process where the populating of the world occurred in a series of discrete steps involving subgroups leaving initial settlements to establish new settlements further away and carrying with them only a subset of the overall genetic diversity of their parental colonies.

In estimating the migratory distance from East Africa for each of the 53 ethnic groups in their data set, Ramachandran et al. (2005) calculate great circle (or geodesic) distances using Addis Ababa (Ethiopia) as the point of common origin and the contemporary geographic coordinates of the sampled groups as the destinations. Moreover, these distance estimates incorporate five obligatory intermediate waypoints, used to more accurately capture paleontological and genetic evidence on prehistorical human migration patterns that are consistent with the widely-held hypothesis that, in the course of their exodus from Africa, humans did not cross large bodies of water. The intermediate waypoints, depicted on the world map in Figure 2 along with the spatial distribution of the ethnic groups from the HGDP-CEPH sample, are: Cairo (Egypt), Istanbul (Turkey), Phnom Penh (Cambodia), Anadyr (Russia) and Prince Rupert (Canada). For instance, as illustrated in Figure 2, the migration path from Addis Ababa to the Papuan ethnic group in modern-day New Guinea makes use of Cairo and Phnom Penh whereas that to the Karitiana population in Brazil incorporates Cairo, Anadyr and Prince Rupert as intermediate waypoints.<sup>17</sup> The migratory distance between endpoints (i.e., Addis Ababa and the location of a group) is therefore the sum of the great circle distances between these endpoints and the waypoint(s) in the path connecting them, and the distance(s) between waypoints if two or more such points are required.

The empirical analysis of Ramachandran et al. (2005) establishes migratory distance from East Africa as a strong negative predictor of genetic diversity at the ethnic group level. Based on the R-squared of their regression, migratory distance alone explains almost 86% of the cross-group variation in within-group diversity.<sup>18</sup> In addition, the estimated OLS coefficient is highly

<sup>16</sup>Ramachandran et al. (2005) compute expected heterozygosity (i.e., genetic diversity) for these 53 ethnic groups from allelic frequencies associated with 783 chromosomal loci.

<sup>17</sup>Based on mitochondrial DNA analysis, some recent studies (e.g., Oppenheimer, 2003; Macaulay et al., 2005) have proposed a southern exit route out of Africa whereby the initial exodus into Asia occurred not via the Levant but across the mouth of the Red Sea (between modern-day Djibouti and Yemen), thereafter taking a “beachcombing” path along the southern coast of the Arabian Peninsula to India and onward into Southeast Asia. Moreover, a subsequent northern offshoot from the Persian Gulf region ultimately lead to the settlement of the Near East and Europe. This scenario therefore suggests the use of Sana’a (Yemen) and Bandar Abbas (Iran) as intermediate waypoints instead of Cairo. Adopting this alternative route for computing migratory distances, however, does not significantly alter the main results presented in Section 3.2.

<sup>18</sup>These results are similar to those uncovered in an independent study by Prugnolle et al. (2005) that employs a subset of the HGDP-CEPH sample encompassing 51 ethnic groups whose expected heterozygosities are calculated from allelic frequencies for 377 loci. Despite their somewhat smaller sample at both the ethnic group and DNA analysis levels, Prugnolle et al. (2005) find that migratory distance from East Africa explains 85% of the variation in genetic diversity. On the other hand, using an expanded data set comprised of the 53 HGDP-CEPH ethnic groups

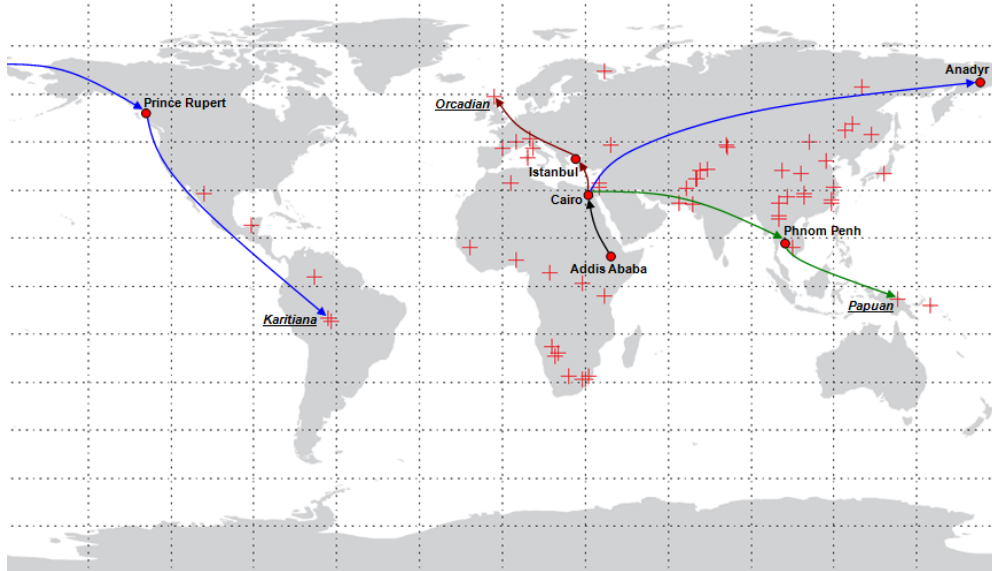


FIGURE 2: The 53 HGDP-CEPH Ethnic Groups and Migratory Paths from East Africa

statistically significant, possessing a  $t$ -statistic =  $-9.770$  ( $P$ -value  $< 10^{-4}$ ), and suggests that *predicted* expected heterozygosity falls by 0.0755 percentage points for every 10,000 km increase in migratory distance from Addis Ababa. This is the relationship depicted earlier on the scatter plot in Figure 1.

The present study exploits the explanatory power of migratory distance from East Africa for the cross-sectional variation in ethnic group expected heterozygosity in order to advance the empirical analysis of the effect of diversity on development in two dimensions. First, given the potential endogeneity between *observed* genetic diversity and economic development as discussed earlier, the use of genetic diversity values *predicted* by migratory distance from East Africa alleviates concerns regarding endogeneity bias. Specifically, the identifying assumption being employed here is that distances along prehistorical human migration routes from Africa have no direct effect on economic development during the Common Era. Second, the strong capacity of migratory distance in predicting genetic diversity implies that the empirical analysis of the genetic hypothesis proposed in this study need not be restricted to the 53 HGDP-CEPH ethnic groups that span only 21 countries, especially since data on the outcome variable of interest (i.e., population density in the year 1500 CE) are available for a much larger set of countries.

To further elaborate, the current analysis tests the proposed genetic hypothesis both using observed genetic diversity in a limited sample of 21 countries, spanned by the 53 ethnic groups in the HGDP-CEPH data set, and using genetic diversity predicted by migratory distance from East Africa in an extended sample of 145 countries. In the 21-country sample, genetic diversity and migratory distance are aggregated up to the country level by averaging across the set of ethnic

---

and an additional 24 Native American populations, Wang et al. (2007) find that migratory distance explains a more modest 74% of the variation in genetic diversity, based on allelic frequencies for 678 loci. The authors attribute their somewhat weaker results to the fact that the additional Native American ethnic groups in their augmented sample were historically subjected to a high degree of gene flow from foreign populations (i.e., European colonizers), which obscured the genetic legacy of a serial-founder effect in these groups.

groups located within a given country.<sup>19</sup> For the extended sample, however, the distance calculation methodology of Ramachandran et al. (2005) is adopted to first construct migratory distance from East Africa for each country, using Addis Ababa as the origin and the country’s modern capital city as the destination along with the aforementioned waypoints for restricting the migration route to landmasses as much as possible.<sup>20</sup> This constructed distance variable is then applied to obtain a predicted value of genetic diversity for each country based on the coefficient on migratory distance in Ramachandran et al.’s (2005) regression across the 53 HGDP-CEPH ethnic groups. Hence, it is this predicted genetic diversity at the country level that is employed as the explanatory variable of interest in the extended sample of countries.<sup>21</sup>

### 3.1.3 Control Variables: Neolithic Transition Timing and Land Productivity

Diamond’s (1997) hypothesis has identified the timing of the Neolithic Revolution as a proximate determinant of economic development, designating initial geographic and biogeographic conditions that governed the emergence and adoption of agricultural practices in prehistorical hunter-gatherer societies as the ultimate determinants in this channel. Some of these geographic and biogeographic factors, highlighted in the empirical analysis of Olsson and Hibbs (2005), include the size of the continent or landmass, the orientation of the major continental axis, type of climate, and the number of prehistorical plant and animal species amenable for domestication.<sup>22</sup>

---

<sup>19</sup>A population-weighted averaging method is infeasible in this case due to the current unavailability of population figures for the HGDP-CEPH ethnic groups.

<sup>20</sup>Clearly, there is some amount of measurement error that is introduced by following this methodology since actual migration paths are only approximated due to the use of five major intercontinental waypoints. For instance, using this general method to calculate the migratory distance to Iceland, which was settled in the 9th century CE by a Norwegian population, fails to capture Oslo as an additional case-specific waypoint. The overall sparsity of historical evidence, however, regarding the actual source of initial settlements in many regions makes a more refined analysis infeasible. Nonetheless, it is credibly postulated that the absence of case-specific waypoints from the analysis does not introduce significant mismeasurement at the global scale. The same argument applies in defense of using modern capital cities as destination points for the migratory paths, although historical evidence suggests that, at least for many cases in the “Old World”, modern capitals were also some of the major centers of urbanization throughout the Common Era (see, e.g., Bairoch, 1988; Chandler, 1987; and, McEvedy and Jones, 1978).

<sup>21</sup>As argued by Pagan (1984) and Murphy and Topel (1985), the OLS estimator for this two-step estimation method yields consistent estimates of the coefficients in the second stage regression, but inconsistent estimates of their standard errors as it fails to account for the presence of a generated regressor. This inadvertently causes naive statistical inferences to be biased in favor of rejecting the null hypothesis. To surmount this issue, the current study employs a two-step bootstrapping algorithm to compute the standard errors in all regressions that use the extended sample containing predicted genetic diversity at the country level. The bootstrap estimates of the standard errors are constructed in the following manner. A random sample with replacement is drawn from the HGDP-CEPH sample of 53 ethnic groups. The first stage regression is estimated on this random sample and the corresponding OLS coefficient on migratory distance is used to compute predicted genetic diversity in the extended sample of countries. The second stage regression is then estimated on a random sample with replacement drawn from the extended cross-country sample and the OLS coefficients are stored. This process of two-step bootstrap sampling and least squares estimation is repeated 1,000 times. The standard deviations in the sample of 1,000 observations of coefficient estimates from the second stage regression are thus the bootstrap standard errors of the point estimates of these coefficients.

<sup>22</sup>See also Weisdorf (2005). While the influence of the number of domesticable species on the likelihood of the emergence of agriculture is evident, the role of the geographic factors requires some elaboration. A larger size of the continent or landmass implied greater biodiversity and, hence, a greater likelihood that at least some species suitable for domestication would exist. In addition, a more pronounced East-West (relative to North-South) orientation of the major continental axis meant an easier diffusion of agricultural practices within the landmass, particularly among regions sharing similar latitudes and, hence, similar environments suitable for agriculture. This orientation factor is argued by Diamond (1997) to have played a pivotal role in comparative economic development by favoring the early rise of complex agricultural civilizations on the Eurasian landmass. Finally, certain climates are known to be more beneficial for agriculture than others. For instance, moderate zones encompassing the Mediterranean and marine

The current analysis controls for the ultimate and proximate determinants of development in the Diamond channel using cross-country data on the aforementioned geographic and biogeographic variables as well as on the timing of the Neolithic Revolution.<sup>23</sup> However, given the empirical link between the ultimate and proximate factors in Diamond’s hypothesis, the baseline specification focuses on the timing of the Neolithic transition to agriculture as the relevant control variable for this channel. The results from an extended specification that incorporates initial geographic and biogeographic factors as controls are presented to demonstrate robustness.

The focus of the historical analysis on economic development in the pre-colonial Malthusian era also necessitates controls for the natural productivity of land for agriculture. Given that in a Malthusian environment resource surpluses are primarily channeled into population growth with per capita incomes largely remaining at or near subsistence, regions characterized by natural factors generating higher agricultural crop yields should, *ceteris paribus*, also exhibit higher population densities (Ashraf and Galor, 2010).<sup>24</sup> If diversity in a society influences its development through total factor productivity (comprised of both social capital and technological know-how), then controlling for the natural productivity of land would constitute a more accurate test of the effect of diversity on the Malthusian development outcome – i.e., population density.

In controlling for the agricultural productivity of land, this study employs measurements of three geographic variables at the country level including the percentage of arable land, absolute latitude, and an index gauging the overall suitability of land for agriculture based on soil quality and temperature.<sup>25</sup>

### 3.1.4 The Baseline Regression Specifications

In light of the proposed genetic diversity hypothesis as well as the roles of the Neolithic transition timing and land productivity channels in agricultural development, the following specification is adopted to examine the influence of observed genetic diversity on economic development in the limited sample of 21 countries:

$$\ln P_{it} = \beta_{0t} + \beta_{1t}G_i + \beta_{2t}G_i^2 + \beta_{3t} \ln T_i + \beta'_{4t} \ln X_i + \beta'_{5t} \ln \Delta_i + \varepsilon_{it}, \quad (3)$$

where  $P_{it}$  is the population density of country  $i$  in a given year  $t$ ,  $G_i$  is the average genetic diversity of the subset of HGDP-CEPH ethnic groups that are located in country  $i$ ,  $T_i$  is the time in years

---

west coast subcategories in the Köppen-Geiger climate classification system are particularly amenable for growing annual, heavy grasses whereas humid subtropical, continental and wet tropical climates are less favorable in this regard, with agriculture being almost entirely infeasible in dry and Polar climates. Indeed, the hypothesized influence of these exogenous factors on the Neolithic Revolution has been established empirically by Olsson and Hibbs (2005) and Putterman (2008).

<sup>23</sup>The data source for the aforementioned geographic and biogeographic controls is Olsson and Hibbs (2005) whereas that for the timing of the Neolithic Revolution is Putterman (2008). See Appendix B for the definitions and sources of all primary and control variables employed by the analysis.

<sup>24</sup>It is important to note, in addition, that the type of land productivity being considered here is largely independent of initial geographic and biogeographic endowments in the Diamond channel and, thus, somewhat orthogonal to the timing of agricultural transitions as well. This holds due to the independence of natural factors conducive to domesticated species from those that were beneficial for the wild ancestors of eventual domesticates. As argued by Diamond (2002), while agriculture originated in regions of the world to which the most valuable domesticable wild plant and animal species were native, other regions proved more fertile and climatically favorable once the diffusion of agricultural practices brought the domesticated varieties to them.

<sup>25</sup>The data for these variables are obtained from the World Bank’s *World Development Indicators*, the CIA’s *World Factbook*, and Michalopoulos (2008) respectively. See Appendix B for additional details.



elapsed since country  $i$ 's transition to agriculture,  $X_i$  is a vector of land productivity controls,  $\Delta_i$  is a vector of continental dummies, and  $\varepsilon_{it}$  is a country-year specific disturbance term.<sup>26</sup>

Moreover, considering the remarkably strong predictive power of migratory distance from East Africa for genetic diversity, the baseline regression specification employed to test the proposed genetic channel in the extended cross-country sample is given by:

$$\ln P_{it} = \beta_{0t} + \beta_{1t}\hat{G}_i + \beta_{2t}\hat{G}_i^2 + \beta_{3t} \ln T_i + \beta_{4t}' \ln X_i + \beta_{5t}' \ln \Delta_i + \varepsilon_{it}, \quad (4)$$

where  $\hat{G}_i$  is the genetic diversity predicted by migratory distance from East Africa for country  $i$  using the methodology discussed in Section 3.1.2. Indeed, it is this regression specification that is estimated to obtain the main empirical findings.<sup>27</sup>

Before proceeding, it is important to note that the regression specifications in (3) and (4) above constitute reduced-form empirical analyses of the genetic diversity channel in Malthusian economic development. Specifically, according to the proposed hypothesis, genetic diversity has a non-monotonic impact on society's level of development through two opposing effects on the level of its total factor productivity: a detrimental effect on social capital and a beneficial effect on the knowledge frontier. However, given the absence of measurements for the proximate determinants of development in the genetic diversity channel, a more discriminatory test of the hypothesis is infeasible. Nonetheless, the results to follow are entirely consistent with the theoretical prediction that, in the presence of diminishing marginal effects of genetic diversity on total factor productivity in a Malthusian economy, the overall reduced-form effect of genetic diversity on cross-country population density should be hump-shaped – i.e., that  $\beta_{1t} > 0$  and  $\beta_{2t} < 0$ . Moreover, as will become evident, the unconditional hump-shaped relationship between genetic diversity and development outcomes does not differ significantly between the adopted quadratic and alternative non-parametric specifications.

### 3.2 Empirical Findings

This section presents the results from empirically investigating the relationship between genetic diversity and log population density in the pre-colonial Malthusian era of development. To this end, the analysis exploits cross-country variations in observed genetic diversity, migratory distance from East Africa and historical population density, as well as in variables used to control for the timing of the Neolithic transition and the natural productivity of land for agriculture. Consistent

---

<sup>26</sup>The fact that economic development has been historically clustered in certain regions of the world raises concerns that these disturbances could be non-spherical in nature, thereby confounding statistical inferences based on the OLS estimator. In particular, the disturbance terms may exhibit spatial autocorrelation, i.e.,  $Cov[\varepsilon_i, \varepsilon_j] > 0$ , within a certain threshold of distance from each observation. Keeping this possibility in mind, the limited sample analyses presented in the text are repeated in Appendix D, where the standard errors of the point estimates are corrected for spatial autocorrelation across disturbance terms, following the methodology of Conley (1999).

<sup>27</sup>Tables E.1–E.2 in Appendix E present the descriptive statistics of the limited 21-country sample employed in estimating equation (3) while Tables E.3–E.4 present those of the extended 145-country sample used to estimate equation (4). As reported therein, the finite-sample moments of the explanatory variables in the limited and extended cross-country samples are remarkably similar. Specifically, the range of values for predicted genetic diversity in the extended sample falls within the range of values for observed diversity in the limited sample. This is particularly reassuring because it demonstrates that the methodology used to generate the predicted genetic diversity variable did not project values beyond what is actually observed, indicating that the HGDP-CEPH collection of ethnic groups is indeed a representative sample for the worldwide variation in within-country genetic diversity. Moreover, the fact that the finite-sample moments of log population density in 1500 CE are not significantly different between the limited and extended cross-country samples foreshadows the encouraging similarity of the regression results that are obtained under observed and predicted values of genetic diversity.

with the theoretical predictions of the proposed diversity channel, the results demonstrate that genetic diversity has a highly statistically significant and robust hump-shaped relationship with historical log population density. Results for observed diversity in the limited 21-country sample are examined in Section 3.2.1. The remaining sections concern genetic diversity, predicted by migratory distance from East Africa, in the extended sample of 145 countries.

Section 3.2.2, in particular, discusses the baseline results associated with examining the effect of predicted diversity on log population density in 1500 CE. The analysis is subsequently expanded upon in Sections 3.2.3-3.2.7 to establish the robustness of the genetic diversity channel with respect to (i) explaining comparative development in earlier historical periods, specifically log population density in 1000 CE and 1 CE, (ii) alternative concepts of distance including the aerial distance from East Africa as well as migratory distances from several “placebo” points of origin across the globe, (iii) the technology diffusion hypothesis that postulates a beneficial effect on development arising from spatial proximity to regional technological frontiers, (iv) controls for microgeographic factors including the degree of variation in terrain and access to waterways, and finally, (v) controls for the exogenous geographic and biogeographic factors favoring an earlier onset of agriculture in the Diamond channel.

### 3.2.1 Results from the Limited Sample

The initial investigation of the proposed genetic diversity hypothesis using the limited sample of countries is of fundamental importance for the subsequent empirical analyses, performed using the extended sample, in three critical dimensions. First, since the limited sample contains observed values of genetic diversity whereas the extended sample comprises values predicted by migratory distance from East Africa, similarity in the results obtained from the two samples would lend credence to the main empirical findings associated with predicted genetic diversity in the extended sample of countries. Second, the fact that migratory distance from East Africa and observed genetic diversity are not perfectly correlated with each other makes it possible to test, using the limited sample of countries, the assertion that migratory distance affects economic development through genetic diversity only and is, therefore, appropriate for generating predicted genetic diversity in the extended sample of countries. Finally, having verified the above assertion, the limited sample permits an instrumental variables regression analysis of the proposed hypothesis with migratory distance employed as an instrument for genetic diversity. This then constitutes a more direct and accurate test of the genetic diversity channel given possible concerns regarding the endogeneity between genetic diversity and economic development. As will become evident, the results obtained from the limited sample indeed deliver on all three aforementioned fronts.

**Explaining Comparative Development in 1500 CE.** Table 1 presents the limited sample results from regressions explaining log population density in 1500 CE.<sup>28</sup> In particular, a number of specifications comprising different subsets of the explanatory variables in equation (3) are estimated to examine the independent and combined effects of the genetic diversity, transition timing, and land productivity channels.

Consistent with the predictions of the proposed diversity hypothesis, Column 1 reveals the unconditional cross-country hump-shaped relationship between genetic diversity and log population density in 1500 CE. Specifically, the estimated linear and quadratic coefficients, both statistically

---

<sup>28</sup>Corresponding to Tables 1 and 2 in the text, Tables D.1 and D.2 in Appendix D present results with standard errors and 2SLS point estimates corrected for spatial autocorrelation across observations.

TABLE 1: Observed Diversity and Economic Development in 1500 CE

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable is Log Population Density in 1500 CE				
Observed Diversity	413.504*** (97.320)			225.440*** (73.781)	203.814* (97.637)
Observed Diversity Sqr.	-302.647*** (73.344)			-161.158** (56.155)	-145.717* (80.414)
Log Transition Timing		2.396*** (0.272)		1.214*** (0.373)	1.135 (0.658)
Log % of Arable Land			0.730** (0.281)	0.516*** (0.165)	0.545* (0.262)
Log Absolute Latitude			0.145 (0.178)	-0.162 (0.130)	-0.129 (0.174)
Log Land Suitability			0.734* (0.381)	0.571* (0.294)	0.587 (0.328)
Optimal Diversity	0.683*** (0.008)			0.699*** (0.015)	0.699*** (0.055)
Continent Dummies	No	No	No	No	Yes
Observations	21	21	21	21	21
R-squared	0.42	0.54	0.57	0.89	0.90

*Note:* Heteroskedasticity robust standard errors are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

significant at the 1% level, imply that a 1 percentage point increase in genetic diversity for the most homogenous society in the regression sample would raise its population density in 1500 CE by 113.99%, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 63.71%. In addition, the coefficients also indicate that a 1 percentage point change in diversity in either direction at the predicted optimum of 0.6831 would lower population density by 2.98%.<sup>29</sup> Furthermore, based on the R-squared coefficient of the regression, the genetic diversity channel appears to explain 42% of the variation in log population density in 1500 CE across the limited sample of countries. The quadratic relationship implied by the OLS coefficients reported in Column 1 is depicted together with a non-parametric local polynomial regression line on the scatter plot in Figure 3. Reassuringly, as illustrated therein, the estimated quadratic falls within the 95% confidence interval band of the non-parametric relationship.<sup>30</sup>

The unconditional effects of the Neolithic transition timing and land productivity channels are reported in Columns 2 and 3 respectively. In line with the Diamond hypothesis, a 1% increase in the number of years elapsed since the transition to agriculture increases population density in 1500 CE by 2.40%, an effect that is also significant at the 1% level. Similarly, consistent with the predictions of the land productivity channel, population density in 1500 CE possesses statistically significant positive elasticities with respect to both the percentage of arable land as well as the index gauging the suitability of land for agriculture. Moreover, the agricultural transition timing and land productivity channels independently explain 54% and 57% of the limited cross-country

<sup>29</sup>The magnitude of these effects can be derived directly from the estimated linear and quadratic coefficients associated with genetic diversity. Specifically, letting  $\hat{\beta}_1$  and  $\hat{\beta}_2$  denote the estimated coefficients on genetic diversity and genetic diversity square, equation (3) can be used to show that the proportional effect on population density of a  $\Delta G$  change in diversity at the specified level  $\bar{G}$  is given by:  $\Delta P/P = \exp\{\Delta G(\hat{\beta}_1 + 2\hat{\beta}_2\bar{G} + \bar{G}\Delta G)\} - 1$ .

<sup>30</sup>For consistency with Figure 1, which depicts the negative effect of increasing migratory distance from East Africa on genetic diversity, the horizontal axes in Figures 3–7 and 9–10 represent genetic homogeneity (i.e., 1 minus genetic diversity) so as to reflect increasing as opposed to decreasing migratory distance from East Africa.

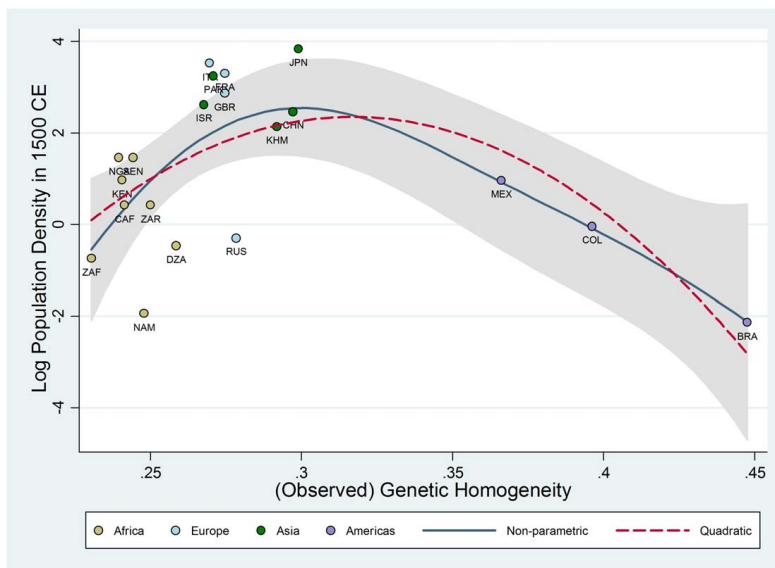


FIGURE 3: Observed Genetic Diversity and Population Density in 1500 CE – The Unconditional Relationship

sample variation in log population density in 1500 CE.

Column 4 presents the results obtained from exploiting the combined explanatory power of all three channels for log population density in the year 1500 CE. Not surprisingly, given the small sample size as well as the pairwise correlations between covariates reported in Table E.2 in Appendix E, the estimated conditional effects are sizeably reduced in magnitude in comparison to their unconditional estimates presented in earlier columns. Nonetheless, the OLS coefficients associated with all channels retain their expected signs and continue to remain highly statistically significant. To interpret the conditional effects of the genetic diversity channel, the estimated linear and quadratic coefficients associated with genetic diversity imply that, accounting for the influence of the transition timing and land productivity channels, a 1 percentage point increase in genetic diversity for the most homogenous society in the regression sample would raise its population density in 1500 CE by 58.03%, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 23.36%. Further, a 1 percentage point change in diversity in either direction at the predicted optimum of 0.6994 would lower population density by 1.60%. Additionally, by exploiting the combined explanatory power of all three channels, the estimated model explains an impressive 89% of the limited sample cross-country variation in log population density.

Finally, the results from estimating the regression model in equation (3) are reported in Column 5, which indicates that the results from previous columns were not simply reflecting the possible influence of some unobserved continent-specific attributes. In spite of the sample size limitations and the smaller variability of covariates within continents in comparison to that across continents, genetic diversity continues to exert significant influence in a manner consistent with theoretical predictions. Reassuringly, the estimated average within-continent effects of the diversity channel are very similar to the cross-continent effects reported in Column 4 and the implied optimal level of diversity remains intact, lending credence to the assertion that these effects are indeed due to genetic diversity as opposed to unobserved continental characteristics.

To summarize, the limited sample results presented in Table 1 demonstrate that genetic diversity has a statistically significant hump-shaped relationship with log population density in the year 1500 CE. The analysis, however, also reveals significant effects associated with the Neolithic transition timing and land productivity channels. Indeed, the non-monotonic effect of diversity on log population density prevails under controls for these other explanatory channels, and remains remarkably stable in magnitude regardless of whether the cross-country variations exploited by the analysis are within or across continents. While, given the obvious limitations of the sample employed, these results may initially appear to be more illustrative rather than conclusive, they are in fact reassuringly similar to those obtained in the extended sample of countries, as will become evident in Section 3.2.2 below. This similarity provides further assurance regarding the validity of the inferences made with the main empirical findings that are associated with predicted as opposed to observed values of genetic diversity.

**Establishing the Exogeneity of Migratory Distance.** As already mentioned, the fact that the limited cross-country sample comprises observed genetic diversity, which is strongly but not perfectly correlated with migratory distance from East Africa, permits a formal examination of whether migratory distance influences population density solely via the serial-founder effect on genetic diversity. This is a particularly important test since, if migratory distance from East Africa actually affects economic development either directly or via some other unobserved channels, then the main empirical analysis conducted using predicted values of diversity would be attributing this latent influence to the genetic diversity channel.<sup>31</sup>

To implement the aforementioned test, the current analysis examines a specification that includes migratory distance from East Africa rather than genetic diversity to explain the cross-country variation in log population density in 1500 CE. The associated results are then compared with those obtained from estimating an alternative specification including both migratory distance and genetic diversity as covariates. Unless migratory distance and genetic diversity are ultimate and proximate determinants within the same channel, then genetic diversity, when included in the regression, should not capture most of the explanatory power otherwise attributed to migratory distance. However, while Column 1 of Table 2 reveals a highly statistically significant unconditional hump-shaped effect of migratory distance from East Africa on log population density, this effect not only becomes insignificant but also drops considerably in magnitude once genetic diversity is accounted for in Column 2. Further, although the linear and quadratic coefficients associated with the effect of genetic diversity, conditional on migratory distance from East Africa, are admittedly somewhat weaker in magnitude when compared to their unconditional estimates in Table 1, they continue to remain statistically significant at conventional levels of significance.

The results of the “horse race” regression in Column 2 are perhaps even more striking given the prior that genetic diversity, as opposed to migratory distance, is likely to be afflicted by larger measurement errors. Nevertheless, since migratory distance is measured as the sum of aerial distances between intercontinental waypoints, it may also be viewed as a noisy proxy of the distance along actual migration routes taken by prehistorical humans during their exodus out of Africa. In order to test whether genetic diversity survives a “horse race” with a less noisy measure of migratory distance from East Africa, Columns 3–4 repeat the preceding analysis using migratory distance

---

<sup>31</sup>Figures C.3(a)–C.3(c) in Appendix C illustrate that, unlike the significant impact of migratory distance from East Africa on genetic diversity, migratory distance has no systematic relationship with a number of observed physiological characteristics of populations, including average skin reflectance, average height, and average weight, conditional on geographical factors such as the intensity of ultraviolet exposure, absolute latitude, the percentage of arable land, the shares of land in tropical and temperate zones, elevation, access to waterways, and continental fixed effects.

TABLE 2: Migratory Distance from East Africa and Economic Development in 1500 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	2SLS	2SLS
Dependent Variable is Log Population Density in 1500 CE						
Observed Diversity		255.219** (100.586)		361.421** (121.429)	233.761*** (86.884)	181.938** (71.933)
Observed Diversity Sqr.		-209.808** (73.814)		-268.514*** (87.342)	-167.566** (65.729)	-130.767** (59.268)
Migratory Distance	0.505*** (0.148)	0.070 (0.184)				
Migratory Distance Sqr.	-0.023*** (0.006)	-0.014 (0.009)				
Mobility Index			0.353** (0.127)	0.051 (0.154)		
Mobility Index Sqr.			-0.012*** (0.004)	-0.003 (0.006)		
Log Transition Timing					1.183*** (0.338)	1.166** (0.475)
Log % of Arable Land					0.531*** (0.170)	0.545** (0.219)
Log Absolute Latitude					-0.169 (0.106)	-0.118 (0.128)
Log Land Suitability					0.558** (0.256)	0.595** (0.256)
Optimal Diversity					0.698*** (0.015)	0.696*** (0.045)
Continent Dummies	No	No	No	No	No	Yes
Observations	21	21	18	18	21	21
R-squared	0.34	0.46	0.30	0.43	–	–
P-value for:						
Joint Sig. of Diversity and its Sqr.		0.023		0.027		
Joint Sig. of Distance and its Sqr.		0.235				
Joint Sig. of Mobility and its Sqr.				0.905		
Overidentifying Restrictions					0.889	0.861
Exogeneity of Distance and its Sqr.					0.952	0.804

*Note:* Heteroskedasticity robust standard errors are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

based on the index of human mobility employed previously by Ashraf et al. (2010). This index captures the average distance from Addis Ababa to the HGDP ethnic groups located within a given country, along “optimal” land-restricted routes that minimize the time cost of movement on the surface of the Earth. The index thus accounts for natural impediments to human mobility, including various meteorological and topographical conditions, and incorporates information on the time cost of travelling under such conditions. Reassuringly, as revealed in Columns 3–4, while distance from East Africa based on the mobility index possesses a significant hump-shaped correlation with log population density, this unconditional relationship virtually disappears once genetic diversity is accounted for by the analysis, lending further support to the claim that distance along prehistorical human migration routes from East Africa confers an effect on development outcomes through genetic diversity alone.<sup>32</sup>

<sup>32</sup>The difference in the number of observations between Columns 1–2 (21 obs.) and Columns 3–4 (18 obs.) arises due to the fact that the mobility index cannot be calculated for countries that can only be accessed from Addis Ababa

The analysis now turns to address concerns regarding the fact that diversity and economic development may be endogenously determined. In particular, Column 5 presents the results from estimating the preferred regression specification, with genetic diversity and its square instrumented by migratory distance and its square as well as the squares of the exogenous transition timing and land productivity variables. The results from a similar analysis that also accounts for continental fixed effects are reported in Column 6. Interestingly, in comparison to their OLS counterparts in Table 1, the estimated 2SLS coefficients associated with the diversity channel remain rather stable in magnitude and increase in statistical significance, particularly for the regression incorporating continental dummies. Moreover, the implied estimates for the optimal level of diversity remain virtually unchanged.

Finally, the 2SLS regressions in Columns 5 and 6 provide additional reassurance regarding the exogeneity of migratory distance with respect to population density. Specifically, since the estimated two-stage models are overidentified (i.e., the number of instruments exceed the number of endogenous regressors), the Sargan-Hansen test for overidentifying restrictions may be employed to examine the joint validity of the instruments. In addition, a difference-in-Sargan test may be used to investigate the orthogonality of a subset of these instruments. Encouragingly, the high P-values associated with these tests not only indicate that the set of instruments employed are plausibly exogenous, but also resonate the earlier finding that migratory distance does not impart independent influence on economic development other than via the serial-founder effect on genetic diversity. Overall, the results uncovered here provide support for the inferences made with predicted genetic diversity in the main empirical analysis to follow.

### 3.2.2 The Baseline Results

This section establishes the hump-shaped impact of genetic diversity, predicted by migratory distance from East Africa, on log population density in 1500 CE, using the extended sample of 145 countries. To reveal the independent and combined effects of the genetic diversity, transition timing, and land productivity channels, Table 3 presents the results from estimating a number of specifications spanning relevant subsets of the explanatory variables in equation (4).

The unconditional hump-shaped relationship between genetic diversity and log population density in 1500 CE is reported in Column 1. In particular, the estimated linear and quadratic coefficients, both statistically significant at the 1% level, imply that a 1 percentage point increase in genetic diversity for the least diverse society in the regression sample would raise its population density by 58.75%, whereas a 1 percentage point decrease in genetic diversity for the most diverse society would raise its population density by 24.56%.<sup>33</sup> Further, population density in 1500 CE is unconditionally predicted by the regression to be maximized at an expected heterozygosity value of about 0.7074, which roughly corresponds to that predicted for southern China by migratory distance from East Africa. Indeed, a 1 percentage point change in genetic diversity in either direction at the predicted optimum lowers population density by 1.76%. Moreover, based on the R-squared

---

by crossing at least one body of water. Restricting the sample used in Columns 1–2 to that in Columns 3–4 does not qualitatively alter the findings. In addition, the unavailability of the mobility index measure for several countries (due to the aforementioned strict land-accessibility constraint) makes this measure less suitable, in comparison to the baseline migratory distance measure of Ramachandran et al. (2005), to predict genetic diversity in the extended cross-country sample.

<sup>33</sup>Following the earlier discussion regarding the expected heterozygosity index, these effects are therefore associated with a 0.01 change in the probability that two randomly selected individuals from a given population are genetically different from one another. See Footnote 29 for details on how these effects may be computed based on the estimated linear and quadratic coefficients associated with genetic diversity.

TABLE 3: Predicted Diversity and Economic Development in 1500 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable is Log Population Density in 1500 CE					
Predicted Diversity	250.986*** (66.314)		213.537*** (61.739)	203.017*** (60.085)	195.416*** (55.916)	199.727** (80.281)
Predicted Diversity Sqr.	-177.399*** (48.847)		-152.107*** (45.414)	-141.980*** (44.157)	-137.977*** (40.773)	-146.167*** (56.251)
Log Transition Timing		1.287*** (0.170)	1.047*** (0.188)		1.160*** (0.143)	1.235*** (0.243)
Log % of Arable Land				0.523*** (0.117)	0.401*** (0.096)	0.393*** (0.103)
Log Absolute Latitude				-0.167* (0.093)	-0.342*** (0.096)	-0.417*** (0.124)
Log Land Suitability				0.189 (0.124)	0.305*** (0.094)	0.257*** (0.096)
Optimal Diversity	0.707*** (0.021)		0.702*** (0.025)	0.715*** (0.110)	0.708*** (0.051)	0.683*** (0.110)
Continent Dummies	No	No	No	No	No	Yes
Observations	145	145	145	145	145	145
R-squared	0.22	0.26	0.38	0.50	0.67	0.69

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.  
\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

of the regression, the cross-country variation in genetic diversity alone explains 22% of the cross-country variation in population density. The quadratic relationship implied by the OLS coefficients reported in Column 1 is depicted together with a non-parametric local polynomial regression line on the scatter plot in Figure 4. As before, the estimated quadratic falls within the 95% confidence interval band of the non-parametric relationship and, moreover, approximates the non-parametric regression line rather well.

Column 2 reports the unconditional effect of the timing of the agricultural transition on population density in 1500 CE. In line with the Diamond hypothesis, a 1% increase in the number of years elapsed since the Neolithic transition to agriculture is associated with a 1.28% increase in population density, an effect that is also statistically significant at the 1% level. Furthermore, 26% of the cross-country variation in population density is explained by the cross-country variation in the timing of the agricultural transition alone. Perhaps unsurprisingly, as foreshadowed by the sample correlations in Table E.4 in Appendix E, the unconditional effects of both the genetic diversity and agricultural transition timing channels are somewhat weakened in magnitude once they are simultaneously taken into account in Column 3, which reduces the omitted variable bias afflicting the coefficient estimates reported in earlier columns. The coefficients on both channels, however, retain their expected signs and continue to remain statistically significant at the 1% level with the combined cross-country variation in genetic diversity and transition timing explaining 38% of the cross-country variation in population density.

The results of examining the combined explanatory power of the genetic diversity and land productivity channels are reported in Column 4.<sup>34</sup> Once again, given the sample correlations, the linear and quadratic coefficients associated with genetic diversity are naturally somewhat weaker when compared to their unconditional estimates of Column 1. More importantly, the coefficients

<sup>34</sup>The cross-country variation in genetic diversity and in variables capturing the productivity of land for agriculture together explain 50% of the cross-country variation in population density.



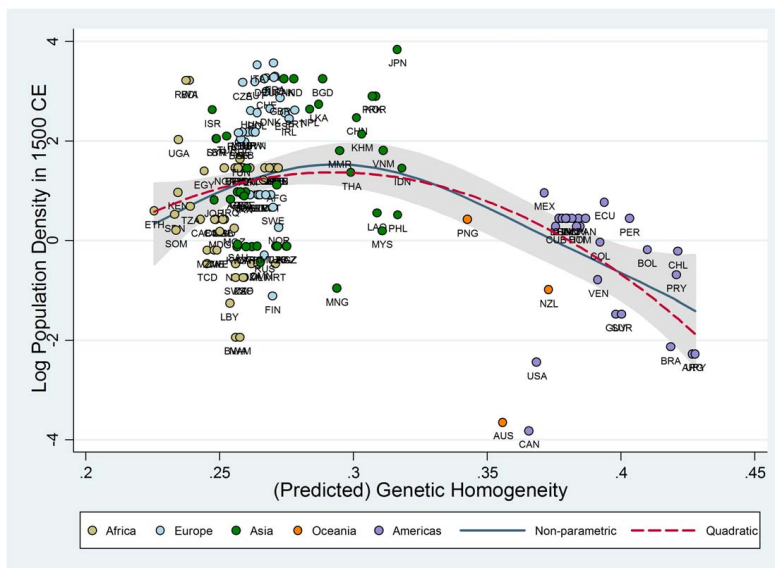


FIGURE 4: Predicted Genetic Diversity and Population Density in 1500 CE – The Unconditional Relationship

remain highly statistically significant and also rather stable in magnitude relative to those estimated while controlling for the timing of the Neolithic transition. In addition, the overall significance of the land productivity channel is also confirmed, particularly by the estimated coefficients on the log percentage of arable land and log absolute latitude variables, which indeed appear to possess their expected signs.<sup>35</sup> Nonetheless, these estimates continue to reflect some amount of omitted variable bias resulting from the exclusion of the transition timing channel. For instance, the fact that log agricultural transition timing has a sample correlation of 0.28 with genetic diversity and one of 0.32 with log absolute latitude implies that the estimated effects of these variables on log population density in Column 4 may be partially capturing the latent influence of the excluded Neolithic transition timing channel.

Column 5 presents the results from exploiting the explanatory power of all three identified channels for log population density in 1500 CE. In line with the theoretical predictions of each hypothesis, the coefficient estimates possess their expected signs and are all statistically significant at the 1% level. Moreover, in comparison to their estimates in Columns 3 and 4, the linear and quadratic coefficients associated with the diversity channel remain largely stable. In particular, the estimated coefficients of interest imply that, controlling for the influence of land productivity and the timing of the Neolithic Revolution, a 1 percentage point increase in genetic diversity for the least diverse society in the sample would raise its population density in 1500 CE by 43.55%, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 18.38%. Further, population density in 1500 CE is predicted to be maximized at an expected heterozygosity value of 0.7081, where a 1 percentage point change in diversity in either direction

<sup>35</sup>To interpret the coefficients associated with the land productivity channel, a 1% increase in the fraction of arable land and in absolute latitude corresponds, respectively, to a 0.52% increase and a 0.17% decrease in population density. While this latter effect may seem unintuitive, given the positive relationship between absolute latitude and contemporary income per capita, it accurately reflects the fact that agricultural productivity in the past has typically been higher at latitudinal bands closer to the equator. In addition, this finding is also consistent with the “reversal of fortune” hypothesis documented by Acemoglu et al. (2005).

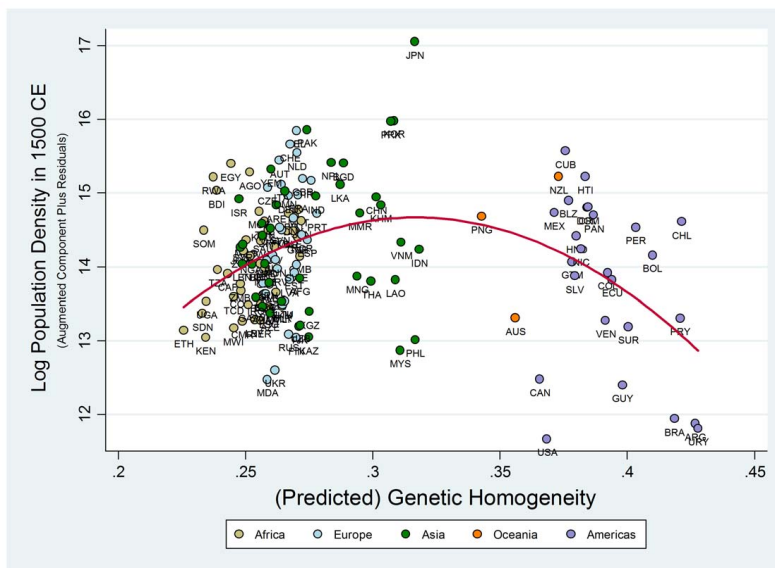


FIGURE 5: Predicted Genetic Diversity and Population Density in 1500 CE – Conditional on Transition Timing, Land Productivity, and Continental Fixed Effects

would lower population density by 1.37%. Overall, based on the R-squared of the regression, the cross-country variations in genetic diversity, agricultural transition timing, and land productivity together explain 67% of the cross-country variation in population density in 1500 CE.

Finally, Column 6 reports the results from estimating the baseline regression model, specified in equation (4), which allows the analysis to capture unobserved continent-specific attributes that could potentially have an influence on population density.<sup>36</sup> Despite the more modest cross-country variation in genetic diversity within continents as opposed to that across continents, the coefficients associated with diversity remain rather stable, increasing slightly in magnitude with the inclusion of continental dummies, although the statistical significance of the linear coefficient drops to the 5% level. Specifically, the coefficients associated with the diversity channel indicate that, controlling for the influence of land productivity, the timing of the Neolithic Revolution, and continental fixed effects, a 1 percentage point increase in diversity for the most homogenous society in the sample would raise its population density in 1500 CE by 36.36%, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 28.62%. In addition, a 1 percentage point change in genetic diversity in either direction at the predicted optimum diversity level of 0.6832, which roughly corresponds to that predicted for Japan by migratory distance from East Africa, would lower population density by 1.45%. Reassuringly, the optimal level of predicted diversity in the extended sample is quite similar to that obtained for observed diversity in the limited 21-country sample.

To place the worldwide effect of the diversity channel into perspective, the coefficients reported in Column 6 imply that increasing the expected heterozygosity of the most homogenous *native* South American populations by 11.1 percentage points to the predicted optimum would have raised their population density in 1500 CE by a factor of 6.07. On the other hand, decreasing the expected heterozygosity of the most heterogenous East African populations by 9.1 percentage points

<sup>36</sup>The excluded continent in all extended sample empirical specifications in this study that incorporate continental dummy variables is Oceania.

to the optimum would have raised their population density by a factor of 3.36. The non-monotonic effect of genetic diversity on log population density in 1500 CE, conditional on the timing of the Neolithic Revolution, land productivity, and continental fixed effects, is depicted on the scatter plot in Figure 5.<sup>37</sup>

To summarize the results reported in Table 3, genetic diversity as predicted by migratory distance from East Africa is found to have a highly statistically significant non-monotonic effect on population density in 1500 CE. This finding is entirely consistent with the theoretical prediction of the proposed genetic diversity channel that comprises both an adverse effect of diversity on Malthusian economic development, via diminished social capital, and a favorable effect arising from increased technological creativity. The analysis also confirms the significant beneficial effects of an earlier Neolithic transition to agriculture as well as geographical factors conducive to higher agricultural yields. Nevertheless, controlling for these additional explanatory channels hardly affects the hump-shaped relationship between genetic diversity and population density, a finding that remains robust to the inclusion of continental dummies as well.

### 3.2.3 Results for Earlier Historical Periods

This section examines the effects of genetic diversity on economic development in earlier historical periods of the Common Era and, in particular, establishes a hump-shaped relationship between genetic diversity, predicted by migratory distance from East Africa, and log population density in the years 1000 CE and 1 CE. In so doing, the analysis demonstrates the persistence of the diversity channel over a long expanse of time and indicates that the hump-shaped manner in which genetic diversity influences development, along with the optimal level of diversity, did not fundamentally change during the agricultural stage of development.

The results from replicating the analysis of the previous section to explain log population density in 1000 CE and 1 CE are presented in Tables 4 and 5 respectively. As before, the individual and combined explanatory powers of the genetic diversity, transition timing, and land productivity channels are examined empirically. The relevant samples, determined by the availability of data on the dependent variable of interest as well as all identified explanatory channels, are comprised of 140 countries for the 1000 CE regressions and 126 countries for the analysis in 1 CE. Despite the more constrained sample sizes, however, the empirical findings once again reveal a highly statistically significant hump-shaped relationship between genetic diversity, predicted by migratory distance from East Africa, and log population density in these earlier historical periods. Additionally, the magnitude and significance of the coefficients associated with the diversity channel in these earlier periods remain rather stable, albeit less so in comparison to the analysis for 1500 CE, when the regression specification is augmented with controls for the transition timing and land productivity channels as well as dummy variables capturing continental fixed effects.

In a pattern similar to that observed in Table 3, the unconditional effects of genetic diversity in Tables 4 and 5 decrease slightly in magnitude when subjected to controls for either the Neolithic transition timing or the land productivity channels, both of which appear to confer their expected

---

<sup>37</sup>Figures 5–7 are “augmented component plus residual” plots and not the typical “added variable” plots of residuals against residuals. In particular, the vertical axes in these figures represent the component of log population density that is explained by genetic homogeneity and its square plus the residuals from the corresponding regression. The horizontal axes, on the other hand, simply represent genetic homogeneity rather than the residuals obtained from regressing homogeneity on the covariates. This methodology permits the illustration of the overall non-monotonic effect of the genetic channel in one scatter plot per regression. Plots depicting the partial regression lines associated with the first-order and second-order effects of genetic homogeneity on log population density in 1500 CE are presented in Figures C.1(a)–C.1(b) in Appendix C.

TABLE 4: Predicted Diversity and Economic Development in 1000 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable is Log Population Density in 1000 CE					
Predicted Diversity	219.722*** (68.108)		158.631** (63.604)	179.523*** (65.981)	154.913** (61.467)	201.239** (97.612)
Predicted Diversity Sqr.	-155.442*** (50.379)		-113.110** (46.858)	-126.147*** (48.643)	-109.806** (44.967)	-145.894** (68.252)
Log Transition Timing		1.393*** (0.170)	1.228*** (0.180)		1.374*** (0.151)	1.603*** (0.259)
Log % of Arable Land				0.546*** (0.140)	0.371*** (0.106)	0.370*** (0.114)
Log Absolute Latitude				-0.151 (0.103)	-0.380*** (0.110)	-0.373*** (0.137)
Log Land Suitability				0.043 (0.135)	0.211** (0.104)	0.190* (0.106)
Optimal Diversity	0.707*** (0.039)		0.701*** (0.127)	0.712*** (0.146)	0.705** (0.108)	0.690** (0.293)
Continent Dummies	No	No	No	No	No	Yes
Observations	140	140	140	140	140	140
R-squared	0.15	0.32	0.38	0.36	0.61	0.62

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.  
\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

effects on population density in earlier historical periods. However, as argued previously, these estimates certainly reflect some amount of omitted variable bias resulting from the exclusion of one or more of the identified explanatory channels in Malthusian economic development. On the other hand, unlike the pattern in Table 3, the coefficients of the diversity channel also weaken moderately in statistical significance, dropping to the 5% level when controlling for transition timing in the 1000 CE analysis and to the 10% level under controls for the land productivity channel in the 1 CE analysis. Nonetheless, this reduction in significance is not entirely surprising when one accounts for the greater imprecision with which population density is recorded for these earlier periods, given that mismeasurement in the dependent variable of an OLS regression typically causes the resulting coefficient estimates to possess larger standard errors.

Column 5 in Tables 4 and 5 reveals the results from exploiting the combined explanatory power of the genetic diversity, transition timing, and land productivity channels for log population density in 1000 CE and 1 CE. Interestingly, in each case, the linear and quadratic coefficients associated with diversity remain rather stable when compared to the corresponding estimates obtained under a partial set of controls in earlier columns. In comparison to the corresponding results for population density in 1500 CE from Table 3, the coefficients of the diversity channel uncovered here are statistically significant at the 5% as opposed to the 1% level, a by-product of relatively larger standard errors that again may be partly attributed to the higher measurement error afflicting population density estimates reported for earlier historical periods.

Finally, the last column in each table augments the analysis with controls for continental fixed effects, demonstrating that the coefficients of the genetic diversity channel in each historical period maintain significance in spite of the lower degree of cross-country variation in diversity within each continent as compared to that observed worldwide. Moreover, the magnitudes of the diversity coefficients remain rather stable, particularly in the 1000 CE analysis, and increase somewhat for population density in 1 CE despite the smaller sample size and, hence, even lower within-continent variation in diversity exploited by the latter regression. Further, the estimated optimal levels of

TABLE 5: Predicted Diversity and Economic Development in 1 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable is Log Population Density in 1 CE					
Predicted Diversity	227.826*** (72.281)		183.142*** (57.772)	129.180* (66.952)	134.767** (59.772)	231.689** (113.162)
Predicted Diversity Sqr.	-160.351*** (53.169)		-132.373*** (42.177)	-88.040* (49.519)	-96.253** (43.718)	-166.859** (79.175)
Log Transition Timing		1.793*** (0.217)	1.636*** (0.207)		1.662*** (0.209)	2.127*** (0.430)
Log % of Arable Land				0.377** (0.158)	0.314** (0.125)	0.348*** (0.134)
Log Absolute Latitude				0.190 (0.125)	-0.121 (0.119)	-0.115 (0.135)
Log Land Suitability				0.160 (0.173)	0.238* (0.124)	0.210* (0.125)
Optimal Diversity	0.710*** (0.052)		0.692*** (0.027)	0.734** (0.347)	0.700*** (0.188)	0.694*** (0.194)
Continent Dummies	No	No	No	No	No	Yes
Observations	126	126	126	126	126	126
R-squared	0.16	0.42	0.46	0.32	0.59	0.61

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.  
\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

diversity in the two periods are relatively stable in comparison to that obtained under the baseline regression for the year 1500 CE. The coefficients associated with diversity from the 1000 CE analysis suggest that, accounting for both land productivity and the timing of the Neolithic transition, a 1 percentage point increase in genetic diversity for the least diverse society in the sample would raise its population density by 38.42%, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 26.15%. On the other hand, for the 1 CE analysis, a similar increase in genetic diversity for the least diverse society would raise its population density by 47.28%, whereas a similar decrease in diversity for the most diverse society would raise its population density by 28.45%.<sup>38</sup> The hump-shaped relationships, based on these coefficients, between genetic diversity and log population density in the years 1000 CE and 1 CE are depicted on the scatter plots in Figures 6 and 7.

In sum, the results presented in Tables 4 and 5 suggest that, consistent with the predictions of the proposed diversity channel, genetic diversity has indeed been a significant determinant of Malthusian economic development in earlier historical periods as well. The overall non-monotonic effect of diversity on population density in the years 1000 CE and 1 CE is robust, in terms of both magnitude and statistical significance, to controls for the timing of the agricultural transition, the natural productivity of land for agriculture and other unobserved continent-specific geographical and socioeconomic characteristics. More fundamentally, the analysis demonstrates the persistence of the diversity channel, along with the optimal level of diversity, over a long expanse of time during the agricultural stage of development.

<sup>38</sup>These effects are calculated directly via the methodology outlined in Footnote 29 earlier, along with the sample minimum and maximum genetic diversity values of 0.5733 and 0.7743, respectively, in both the 1000 CE and 1 CE regression samples.

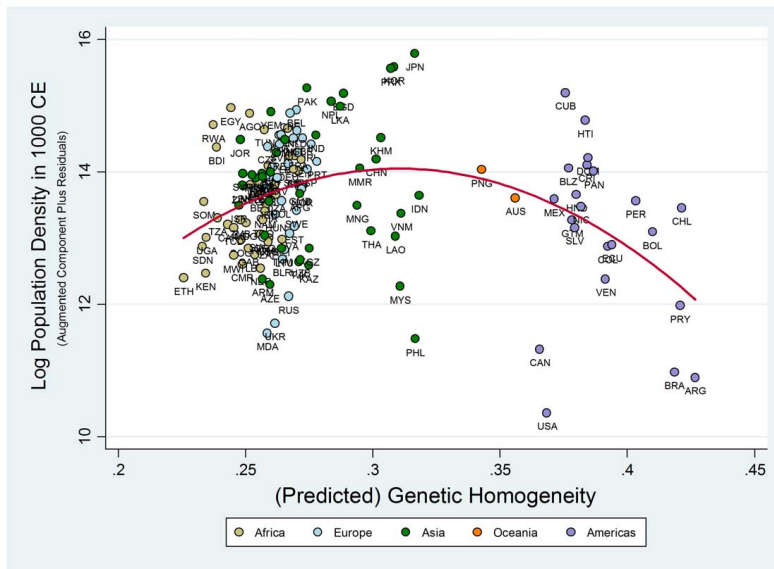


FIGURE 6: Predicted Genetic Diversity and Population Density in 1000 CE – Conditional on Transition Timing, Land Productivity, and Continental Fixed Effects

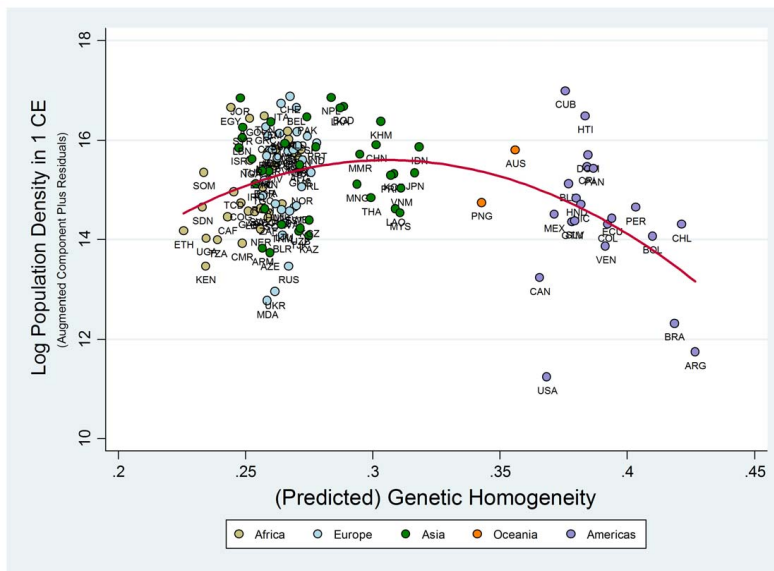


FIGURE 7: Predicted Genetic Diversity and Population Density in 1 CE – Conditional on Transition Timing, Land Productivity, and Continental Fixed Effects

### 3.2.4 Robustness to Aerial Distance and Migratory Distances from “Placebo” Points of Origin Across the Globe

The results from the limited sample analysis discussed earlier demonstrate that the cross-country variation in migratory distance from East Africa has a significant non-monotonic influence on comparative development in 1500 CE and that this impact runs exclusively via the serial-founder effect on genetic diversity. This finding, however, does not preclude the possibility that alternative measures of distance, potentially correlated with migratory distance from East Africa, may also explain the historical cross-country variation in economic development in a similar non-monotonic fashion. Indeed, if this is the case, then the role previously ascribed to the “out of Africa” migration of *Homo sapiens* as a deep determinant of comparative development becomes suspect, undermining the credibility of the proposed genetic diversity channel. Nonetheless, alternative distances, as will become evident, do not impart any significant influence, similar to that associated with migratory distance from East Africa, on log population density in 1500 CE.

The current analysis compares regression results obtained using migratory distance from Addis Ababa in the baseline specification with those obtained under several alternative concepts of distance. The alternative concepts of distance considered by the analysis include the aerial or “as the crow flies” distance from Addis Ababa as well as migratory distances from “placebo” points of origin in other continents across the globe, namely London, Tokyo, and Mexico City, computed using the same waypoints employed in constructing migratory distance from Addis Ababa.<sup>39</sup> As revealed in Table E.4 in Appendix E, with the exception of migratory distance from Tokyo, these other distances are rather strongly correlated with migratory distance from Addis Ababa. Despite some of these high correlations, however, the results presented in Table 6 indicate that migratory distance from Addis Ababa is the only concept of distance that confers a significant non-monotonic effect on log population density.

Specifically, consistent with the proposed diversity hypothesis, Column 1 reveals a highly statistically significant hump-shaped relationship between migratory distance from Addis Ababa and log population density in 1500 CE, conditional on controls for the Neolithic transition timing and land productivity channels. In contrast, the linear and quadratic effects of aerial distance from Addis Ababa, reported in Column 2, are not statistically different from zero at conventional levels of significance. Similarly, as shown in Columns 3–5, the migratory distances from “placebo” points of origin do not impart any statistically discernible effect, linear or otherwise, on log population density in the year 1500 CE.

These results strengthen the assertion that conditions innately related to the prehistorical migration of humans out of Africa have had a lasting impact on comparative development. Given the high correlations between migratory distance from Addis Ababa and some of these alternative distance concepts, the fact that these other distances fail to reveal any significant effects makes the argument in favor of the “out of Africa” hypothesis even stronger. Together with earlier findings establishing migratory distance from Addis Ababa and genetic diversity as ultimate and proximate determinants in the same channel, the findings from these “placebo” tests of distance lend further credence to the proposed diversity hypothesis.

---

<sup>39</sup>The choice of these alternative points of origin do not reflect any systematic selection process, other than the criterion that they belong to different continents in order to demonstrate, at a global scale, the neutrality of migratory distance from locations outside of East Africa. Indeed, other points of origin in Europe, Asia and the Americas yield qualitatively similar results.

TABLE 6: Robustness to Alternative Distances

	(1)	(2)	(3)	(4)	(5)
Distance from:	Addis Ababa	Addis Ababa	London	Tokyo	Mexico City
	Dependent Variable is Log Population Density in 1500 CE				
Migratory Distance	0.138** (0.061)		-0.040 (0.063)	0.052 (0.145)	-0.063 (0.099)
Migratory Distance Sqr.	-0.008*** (0.002)		-0.002 (0.002)	-0.006 (0.007)	0.005 (0.004)
Aerial Distance		-0.008 (0.106)			
Aerial Distance Sqr.		-0.005 (0.006)			
Log Transition Timing	1.160*** (0.144)	1.158*** (0.138)	1.003*** (0.164)	1.047*** (0.225)	1.619*** (0.277)
Log % of Arable Land	0.401*** (0.091)	0.488*** (0.102)	0.357*** (0.092)	0.532*** (0.089)	0.493*** (0.094)
Log Absolute Latitude	-0.342*** (0.091)	-0.263*** (0.097)	-0.358*** (0.112)	-0.334*** (0.099)	-0.239*** (0.083)
Log Land Suitability	0.305*** (0.091)	0.254** (0.102)	0.344*** (0.092)	0.178** (0.080)	0.261*** (0.092)
Observations	145	145	145	145	145
R-squared	0.67	0.59	0.67	0.59	0.63

*Note:* Heteroskedasticity robust standard errors are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

### 3.2.5 Robustness to the Technology Diffusion Hypothesis

The technology diffusion hypothesis, as mentioned earlier, suggests that spatial proximity to global and regional technological frontiers confers a beneficial effect on the development of less advanced societies by facilitating the diffusion of new technologies from more advanced societies through trade as well as sociocultural and geopolitical influences. In particular, the technology diffusion channel implies that, *ceteris paribus*, the greater the geographic distance from the global and regional technological “leaders” in a given period, the lower the level of economic development amongst the “followers” in that period. Indeed, several studies in international trade and economic geography have uncovered strong empirical support for this hypothesis in explaining comparative economic development in the contemporary era.<sup>40</sup> This section examines the robustness of the effects of genetic diversity on economic development during the pre-colonial era to controls for this additional hypothesis.

The purpose of the current investigation is to ensure that the preceding analyses were not ascribing to genetic diversity the predictive power that should otherwise have been attributed to the technology diffusion channel. To be specific, one may identify some of the waypoints employed to construct the prehistorical migratory routes from East Africa (such as Cairo and Istanbul) as origins of spatial technology diffusion during the pre-colonial era. This, coupled with the fact that genetic diversity decreases with increasing migratory distance from East Africa, raises the concern that what has so far been interpreted as evidence consistent with the beneficial effect of higher diversity may, in reality, simply be capturing the latent effect of the omitted technology diffusion channel in preceding regression specifications. As will become evident, however, while technology

<sup>40</sup>The interested reader is referred to Keller (2004) for a more comprehensive review of studies examining the technology diffusion hypothesis.



TABLE 7: The Regional Frontiers Identified for each Historical Period

City and Modern Location	Continent	Sociopolitical Entity	Relevant Period
Cairo, Egypt	Africa	Mamluk Sultanate	1500 CE
Fez, Morocco	Africa	Marinid Kingdom of Fez	1500 CE
London, UK	Europe	Tudor Dynasty	1500 CE
Paris, France	Europe	Valois-Orléans Dynasty	1500 CE
Constantinople, Turkey	Asia	Ottoman Empire	1500 CE
Peking, China	Asia	Ming Dynasty	1500 CE
Tenochtitlan, Mexico	Americas	Aztec Civilization	1500 CE
Cuzco, Peru	Americas	Inca Civilization	1500 CE
Cairo, Egypt	Africa	Fatimid Caliphate	1000 CE
Kairwan, Tunisia	Africa	Berber Zirite Dynasty	1000 CE
Constantinople, Turkey	Europe	Byzantine Empire	1000 CE
Cordoba, Spain	Europe	Caliphate of Cordoba	1000 CE
Baghdad, Iraq	Asia	Abbasid Caliphate	1000 CE
Kaifeng, China	Asia	Song Dynasty	1000 CE
Tollan, Mexico	Americas	Classic Maya Civilization	1000 CE
Huari, Peru	Americas	Huari Culture	1000 CE
Alexandria, Egypt	Africa	Roman Empire	1 CE
Carthage, Tunisia	Africa	Roman Empire	1 CE
Athens, Greece	Europe	Roman Empire	1 CE
Rome, Italy	Europe	Roman Empire	1 CE
Luoyang, China	Asia	Han Dynasty	1 CE
Seleucia, Iraq	Asia	Seleucid Dynasty	1 CE
Teotihuacán, Mexico	Americas	Pre-classic Maya Civilization	1 CE
Cahuachi, Peru	Americas	Nazca Culture	1 CE

diffusion is indeed found to have been a significant determinant of comparative development in the pre-colonial era, the baseline findings for genetic diversity remain robust to controls for this additional influential hypothesis.

To account for the technology diffusion channel, the current analysis constructs, for each historical period examined, a control variable measuring the great circle distance from the closest regional technological frontier in that period. Following the well-accepted notion that the process of pre-industrial urban development was typically more pronounced in societies that enjoyed higher agricultural surpluses, the analysis adopts historical city population size as an appropriate metric to identify the period-specific sets of regional technological frontiers. Specifically, based on historical urban population data from Chandler (1987) and Modelski (2003), the procedure commences with assembling, for each period, a set of regional frontiers comprising the two largest cities, belonging to different civilizations or disparate sociopolitical entities, from each of Africa, Europe, Asia and the Americas.<sup>41</sup> The effectiveness of this procedure in yielding an outcome that is consistent with what one might expect from a general familiarity with world history is evident in the regional frontiers obtained for each period as shown in Table 7.<sup>42</sup> In constructing the variable measuring distance to

<sup>41</sup>The exclusion of Oceania from the list of continents employed is not a methodological restriction but a natural result arising from the fact that evidence of urbanization does not appear in the historical record of this continent until after European colonization. Moreover, the consideration of the Americas as a single unit is consistent with the historical evidence that this landmass only harbored two distinct major civilizational sequences – one in Mesoamerica and the other in the Andean region of South America. Indeed, the imposition of the criteria that the selected cities in each continent (or landmass) should belong to different sociopolitical units is meant to capture the notion that technology diffusion historically occurred due to civilizational influence, broadly defined, as opposed to the influence of only major urban centers that were developed by these relatively advanced societies.

<sup>42</sup>Note that for the year 1 CE there are four cities appearing within the territories of the Roman Empire, which at first glance seems to violate the criterion that the regional frontiers selected should belong to different sociopolitical entities. This is simply a by-product of the dominance of the Roman Empire in the Mediterranean basin during

the closest regional frontier for a given historical period, the analysis then selects, for each country in the corresponding regression sample, the smallest of the great circle distances from the regional frontiers to the country's capital city.

To anticipate the robustness of the baseline results for predicted diversity to controls for the technology diffusion hypothesis, it may be noted that migratory distance from East Africa possesses a correlation coefficient of only 0.02 with the great circle distance from the closest regional frontier in the 1500 CE sample. Furthermore, for the 1000 CE and 1 CE regression samples, migratory distance is again weakly correlated with distance from the closest regional technological frontier in each period, with the respective correlation coefficients being -0.04 and 0.03.<sup>43</sup> These encouragingly low sample correlations are indicative of the fact that the earlier regression specifications estimated by the analysis were indeed not simply attributing to genetic diversity the effects possibly arising from the technology diffusion channel.

Column 1 of Table 8 reports the results from estimating the baseline specification for log population density in 1500 CE, while controlling for technology diffusion as originating from the regional frontiers identified for this period. In comparison to the baseline estimates revealed in Column 6 of Table 3, the regression coefficients associated with the genetic diversity channel remain relatively stable, decreasing only moderately in magnitude and statistical significance. Some similar robustness characteristics may be noted for the transition timing and land productivity channels as well. Importantly, however, the estimate for the optimal level of diversity remains virtually unchanged and highly statistically significant. Interestingly, the results also establish the technology diffusion channel as a significant determinant of comparative development in the pre-colonial Malthusian era. In particular, a 1% increase in distance from the closest regional frontier is associated with a decrease in population density by 0.19%, an effect that is statistically significant at the 1% level.

Columns 2–3 establish the robustness of the diversity channel in 1000 CE and 1 CE to controls for technology diffusion arising from the technological frontiers identified for these earlier historical periods. Specifically, comparing Column 2 with the relevant baseline (i.e., Column 6 in Table 4), the linear and quadratic coefficients of genetic diversity for the 1000 CE regressions remain largely stable under controls for technology diffusion, decreasing slightly in magnitude but maintaining statistical significance. A similar stability pattern also emerges for the coefficients capturing the influence of the diversity channel in the 1 CE regressions. Indeed, the estimates for optimal diversity in these earlier periods remain rather stable relative to their respective baselines in Tables 4 and 5. Finally, in line with the predictions of the technology diffusion hypothesis, a statistically significant negative effect of distance from the closest regional frontier on economic development is observed for these earlier historical periods as well.

The results uncovered herein demonstrate the persistence of the significant non-monotonic effect of diversity on comparative development over the period 1–1500 CE, despite controls for the clearly influential role of technology diffusion from technological frontiers that were relevant during this period of world history. Indeed, these findings lend further credence to the proposed genetic diversity channel by demonstrating that the empirical analyses thus far have not been ascribing to

---

that period. In fact, historical evidence suggests that the cities of Athens, Carthage and Alexandria had long been serving as centers of regional diffusion prior to their annexation to the Roman Empire. Moreover, the appearance of Constantinople under Europe in 1000 CE and Asia in 1500 CE is an innocuous classification issue arising from the fact that the city historically fluctuated between the dominions of European and Asian civilizations.

<sup>43</sup>These correlations differ slightly from those presented in Table E.4 in Appendix E, where the correlations are presented for the entire 145-country sample used in the regressions for 1500 CE.

TABLE 8: Robustness to the Technology Diffusion Hypothesis

	(1)	(2)	(3)
	Dependent Variable is Log Population Density in:		
	1500 CE	1000 CE	1 CE
Predicted Diversity	156.736** (75.572)	183.771** (88.577)	215.858** (105.286)
Predicted Diversity Sqr.	-114.626** (52.904)	-134.609** (61.718)	-157.724** (73.681)
Log Transition Timing	0.909*** (0.254)	1.253*** (0.339)	1.676*** (0.434)
Log % of Arable Land	0.363*** (0.104)	0.323*** (0.121)	0.342*** (0.131)
Log Absolute Latitude	-0.492*** (0.134)	-0.454*** (0.149)	-0.212 (0.142)
Log Land Suitability	0.275*** (0.090)	0.239** (0.105)	0.191 (0.120)
Log Distance to Frontier in 1500 CE	-0.187*** (0.070)		
Log Distance to Frontier in 1000 CE		-0.230* (0.121)	
Log Distance to Frontier in 1 CE			-0.297*** (0.102)
Optimal Diversity	0.684*** (0.169)	0.683*** (0.218)	0.684** (0.266)
Continent Dummies	Yes	Yes	Yes
Observations	145	140	126
R-squared	0.72	0.64	0.66

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

genetic diversity the explanatory power that should otherwise have been attributed to the impact of spatial technology diffusion.

### 3.2.6 Robustness to Microgeographic Factors

This section addresses concerns regarding the possibility that the baseline results for predicted genetic diversity could in fact be reflecting the latent impact of microgeographic factors, such as the degree of variation in terrain and proximity to waterways, if these variables happen to be correlated with migratory distance from East Africa. There are several conceivable channels through which such factors could affect a society's aggregate productivity and thus its population density in the Malthusian stage of development. For instance, the degree of terrain variation within a region can directly affect its agricultural productivity by influencing the arability of land. Moreover, terrain ruggedness may also have led to the spatial concentration of economic activity, which has been linked with increasing returns to scale and higher aggregate productivity through agglomeration by the new economic geography literature.<sup>44</sup> On the other hand, by geographically isolating population subgroups, a rugged landscape could also have nurtured their ethnic differentiation over time (Michalopoulos, 2008), and may thus confer an adverse effect on society's aggregate

<sup>44</sup>The classic reference on economies of agglomeration is Krugman (1991). A detailed survey of the new economic geography literature is conducted by Fujita et al. (1999).

productivity via the increased likelihood of ethnic conflict. Similarly, while proximity to waterways can directly affect crop yields by making beneficial practices such as irrigation possible, it may also have augmented productivity indirectly by lowering transportation costs and, thereby, fostering urban development, trade and technology diffusion.<sup>45</sup>

To ensure that the significant effects of genetic diversity in the baseline regressions are not simply reflecting the latent influence of microgeographic factors, the current analysis examines variants of the baseline specification augmented with controls for terrain quality and proximity to waterways. In particular, the terrain controls are derived from the G-ECON data set compiled by Nordhaus (2006) and include mean elevation and a measure of surface roughness, aggregated up to the country level from grid-level data at a granularity of 1° latitude x 1° longitude. In light of the possibility that the impact of terrain undulation could be non-monotonic, the specifications examined also control for the squared term of the roughness index. The control variables gauging access to waterways, obtained from the Gallup et al. (1999) data set, include the expected distance from any point within a country to the nearest coast or sea-navigable river as well as the percentage of a country's land area located within 100 km of a coast or sea-navigable river.<sup>46</sup> Foreshadowing the robustness of the baseline results, mean elevation, roughness and roughness square possess only moderate correlation coefficients of -0.11, 0.16 and 0.09, respectively, with migratory distance from East Africa. Moreover, migratory distance is also only moderately correlated with the measures of proximity to waterways, possessing sample correlations of -0.20 and 0.19 with the distance and land area variables described above.

The results from estimating augmented regression specifications for log population density in 1500 CE, incorporating controls for either terrain quality or access to waterways, are shown in Columns 1 and 2 of Table 9. In each case, the coefficients associated with the diversity channel remain statistically significant and relatively stable, experiencing only a moderate decrease in magnitude, when compared to the baseline results from Table 3. Interestingly, the control variables for terrain quality in Column 1 and those gauging access to waterways in Column 2 appear to confer statistically significant effects on population density in 1500 CE, and mostly in directions consistent with priors. The results suggest that terrain roughness does indeed have a non-monotonic impact on aggregate productivity, with the beneficial effects dominating at relatively lower levels of terrain roughness and the detrimental effects dominating at higher levels. Further, regions with greater access to waterways are found to support higher population densities.

The final column of Table 9 examines the influence of the genetic diversity channel when subjected to controls for both terrain quality and access to waterways. As anticipated by the robustness of the results from preceding columns, genetic diversity continues to exert a significant non-monotonic effect on population density in 1500 CE, without exhibiting any drastic reductions in the magnitude of its impact. Moreover, the estimate for the optimal level of diversity remains fully intact in comparison to the baseline estimate from Column 6 in Table 3. The results of this section therefore suggest that the significant non-monotonic impact of genetic diversity on population density in 1500 CE is indeed not a spurious relationship arising from the omission of microgeographic factors in the baseline regression specification.

---

<sup>45</sup>Indeed, a significant positive relationship between proximity to waterways and contemporary population density has been demonstrated by Gallup et al. (1999).

<sup>46</sup>For completeness, specifications controlling for the squared terms of the other microgeographic factors were also examined. The results from these additional regressions, however, did not reveal any significant non-linear effects and are therefore not reported.

TABLE 9: Robustness to Microgeographic Factors

	(1)	(2)	(3)
	Dependent Variable is Log Population Density in 1500 CE		
Predicted Diversity	160.346** (78.958)	157.073** (79.071)	157.059** (69.876)
Predicted Diversity Sqr.	-118.716** (55.345)	-112.780** (55.694)	-114.994** (48.981)
Log Transition Timing	1.131*** (0.225)	1.211*** (0.201)	1.215*** (0.197)
Log % of Arable Land	0.397*** (0.099)	0.348*** (0.099)	0.374*** (0.087)
Log Absolute Latitude	-0.358*** (0.124)	-0.354*** (0.132)	-0.352*** (0.122)
Log Land Suitability	0.188* (0.101)	0.248*** (0.082)	0.160** (0.081)
Mean Elevation	-0.404 (0.251)		0.502* (0.273)
Terrain Roughness	5.938*** (1.870)		4.076** (1.840)
Terrain Roughness Sqr.	-7.332** (2.922)		-7.627*** (2.906)
Mean Distance to Nearest Waterway		-0.437** (0.178)	-0.390** (0.181)
% of Land within 100 km of Waterway		0.731** (0.310)	1.175*** (0.294)
Optimal Diversity	0.675*** (0.224)	0.696*** (0.188)	0.683*** (0.083)
Continent Dummies	Yes	Yes	Yes
Observations	145	145	145
R-squared	0.72	0.75	0.78

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

### 3.2.7 Robustness to Exogenous Factors in the Diamond Hypothesis

This section demonstrates the robustness of the effects of genetic diversity to additional controls for the Neolithic transition timing channel. In particular, the analysis is intended to alleviate concerns that the significant effects of genetic diversity presented in Section 3.2.2, although estimated while controlling for the timing of the Neolithic Revolution, may still capture some latent influence of this other explanatory channel if spurious correlations exist between migratory distance from East Africa and exogenous factors governing the timing of the Neolithic transition. The results from estimating some extended specifications, constructed by augmenting equation (4) with controls for the ultimate determinants in the Diamond hypothesis, for log population density in 1500 CE are presented in Table 10.

Following the discussion in Section 3.1.3 on the geographic and biogeographic determinants in the transition timing channel, the additional control variables employed by the current analysis include: (i) climate, measured as a discrete index with higher integer values assigned to countries in Köppen-Geiger climatic zones that are increasingly favorable to agriculture; (ii) orientation of continental axis, measured as the ratio of the longitudinal distance to the latitudinal distance of the continent or landmass to which a country belongs; (iii) size of continent, measured as the total land

TABLE 10: Robustness to Ultimate Determinants in the Diamond Hypothesis

	(1)	(2)	(3)	(4)	(5)
Dependent Variable is Log Population Density in 1500 CE					
Predicted Diversity	216.847*** (62.764)	252.076*** (71.098)	174.414*** (62.505)	212.123*** (70.247)	274.916*** (73.197)
Predicted Diversity Sqr.	-154.750*** (45.680)	-180.650*** (52.120)	-125.137*** (45.568)	-151.579*** (51.463)	-197.120*** (53.186)
Log Transition Timing	1.300*** (0.153)				1.160*** (0.298)
Log % of Arable Land	0.437*** (0.116)	0.431*** (0.119)	0.441*** (0.111)	0.411*** (0.116)	0.365*** (0.112)
Log Absolute Latitude	-0.212** (0.102)	-0.426*** (0.131)	-0.496*** (0.154)	-0.487*** (0.163)	-0.332** (0.145)
Log Land Suitability	0.288** (0.135)	0.184 (0.143)	0.297** (0.146)	0.242* (0.146)	0.280** (0.122)
Climate		0.622*** (0.137)		0.419 (0.268)	0.374* (0.225)
Orientation of Axis		0.281 (0.332)		0.040 (0.294)	-0.169 (0.255)
Size of Continent		-0.007 (0.015)		-0.005 (0.013)	-0.006 (0.012)
Domesticable Plants			0.015 (0.019)	-0.005 (0.023)	0.003 (0.021)
Domesticable Animals			0.154** (0.063)	0.121 (0.074)	-0.013 (0.073)
Optimal Diversity	0.701*** (0.021)	0.698*** (0.019)	0.697*** (0.051)	0.700*** (0.078)	0.697*** (0.020)
Observations	96	96	96	96	96
R-squared	0.74	0.70	0.70	0.72	0.78

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

area of the country's continent; (iv) the number of domesticable wild plant species known to have existed in prehistory in the region to which a country belongs; and (v) the number of domesticable wild animal species known to have been native to the region in prehistory, as reported in the data set of Olsson and Hibbs (2005).

To demonstrate the robustness of the baseline effects of genetic diversity across the various extended specifications examined in this section, Column 1 first presents the results from estimating the baseline specification for log population density in 1500 CE using the restricted 96-country sample of Olsson and Hibbs (2005). Reassuringly, the highly significant coefficients associated with diversity, as well as the other explanatory channels, remain rather stable in magnitude relative to their estimates obtained with the unrestricted sample in Column 5 of Table 3, implying that any sampling bias that may have been introduced inadvertently by the use of the restricted sample in the current analysis is indeed negligible.<sup>47</sup>

Columns 2–4 reveal the results from estimating variants of the baseline specification where the Diamond channel is controlled for not by its proximate determinant but by one or more of its

<sup>47</sup>Note that the specifications estimated in the current analysis do not incorporate continental dummies since a sizeable portion of possible continent-specific effects are captured by some of the (bio)geographic variables in the Diamond channel that are measured at either continental or macro-regional levels. Augmenting the specifications with continental dummies, however, does not significantly alter the results for genetic diversity.

ultimate determinants – i.e., either the set of geographical factors or the set of biogeographical factors or both. The results indicate that the coefficients associated with diversity continue to remain highly significant and relatively stable in magnitude in comparison to their baseline estimates of Column 1. Interestingly, when controlling for only the geographical determinants of the Diamond channel in Column 2, climate alone is significant amongst the additional factors and likewise, when only the biogeographical determinants are controlled for in Column 3, the number of domesticable animal species, rather than plants, appears to be important. In addition, none of the ultimate factors in the Diamond channel appear to possess statistical significance when both geographic and biogeographic determinants are controlled for in Column 4, a result that possibly reflects the high correlations amongst these control variables. Regardless of these tangential issues, however, genetic diversity, as already mentioned, continues to exert significant influence in a manner consistent with theoretical predictions.

The final column in Table 10 establishes the robustness of the effects of genetic diversity on Malthusian development in 1500 CE to controls for both the proximate and ultimate determinants in the Diamond channel. Perhaps unsurprisingly, the Neolithic transition timing variable, being the proximate factor in this channel, captures most of the explanatory power of the ultimate exogenous determinants of comparative development in the Diamond hypothesis. More importantly, the linear and quadratic coefficients of the diversity channel maintain relative stability, increasing slightly in magnitude when compared to their baseline estimates, but remaining highly statistically significant in their expected directions. Overall, the results in Table 10 suggest that the baseline estimates of the impact of genetic diversity presented in Section 3.2.2 earlier are indeed not simply reflecting some latent effects of the influential agricultural transition timing channel.

## 4 The Contemporary Analysis

### 4.1 Data and Empirical Strategy

This section discusses the data and empirical strategy employed to examine the impact of genetic diversity on contemporary comparative development.

#### 4.1.1 The Index of Contemporary National Population Diversity

The construction of the index of genetic diversity for contemporary national populations is partly based on their ethnic compositions resulting from population flows amongst countries in the post-Columbian era. Specifically, given the genetic diversity of the ancestral populations of the source countries, data on post-Columbian population flows can be used to construct a weighted average expected heterozygosity measure for the national population of each country in the contemporary period.<sup>48</sup> This measure alone, however, would not capture the full extent of genetic diversity in contemporary national populations as it would fail to account for the diversity arising from differences between sub-national ethnic groups.

To additionally incorporate the between-group component of diversity in contemporary national populations, the index makes use of the concept of  $F_{st}$  genetic distance from field of population genetics. Specifically, for any sub-population pair, the  $F_{st}$  genetic distance between the two sub-populations captures the proportion of their combined genetic diversity that is unexplained

---

<sup>48</sup>The data on ethnic compositions are obtained from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2009) who compile, for each country in their data set, the share of the country’s population in 2000 CE that is descended from the population of every other country in 1500 CE.

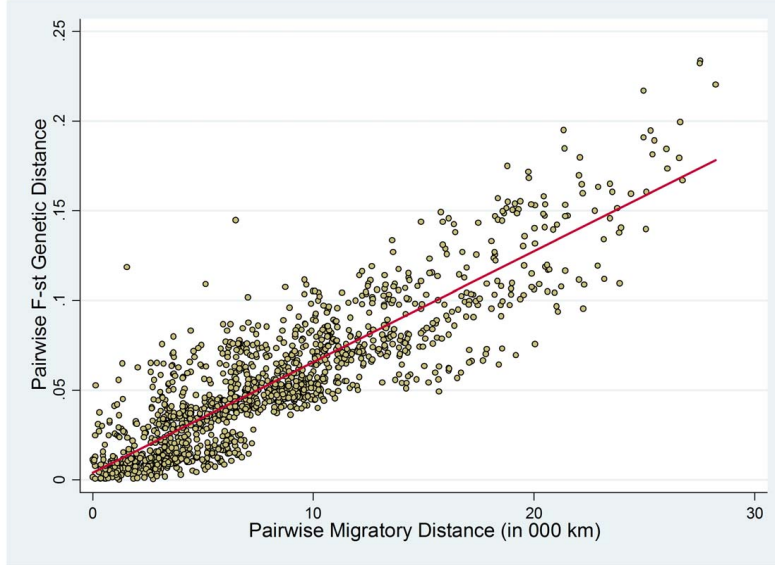


FIGURE 8: Pairwise  $F_{st}$  Genetic and Migratory Distances in the HGDP-CEPH Sample

by the weighted average of their respective genetic diversities. Consider, for instance, a population comprised of two ethnic groups or sub-populations,  $A$  and  $B$ . The  $F_{st}$  genetic distance between  $A$  and  $B$  would then be defined as:

$$F_{st}^{AB} = 1 - \frac{\theta_A H_{\text{exp}}^A + \theta_B H_{\text{exp}}^B}{H_{\text{exp}}^{AB}}, \quad (5)$$

where  $\theta_A$  and  $\theta_B$  are the shares of groups  $A$  and  $B$ , respectively, in the combined population;  $H_{\text{exp}}^A$  and  $H_{\text{exp}}^B$  are their respective expected heterozygosities; and  $H_{\text{exp}}^{AB}$  is the expected heterozygosity of the combined population. Thus, given (i) genetic distance,  $F_{st}^{AB}$ , (ii) the expected heterozygosities of the component sub-populations,  $H_{\text{exp}}^A$  and  $H_{\text{exp}}^B$ , and (iii) their respective shares in the overall population,  $\theta_A$  and  $\theta_B$ , the overall diversity of the combined population is:

$$H_{\text{exp}}^{AB} = \frac{\theta_A H_{\text{exp}}^A + \theta_B H_{\text{exp}}^B}{(1 - F_{st}^{AB})}. \quad (6)$$

In principle, the methodology described above could be applied recursively to arrive at a measure of overall diversity for any contemporary national population comprised of an arbitrary number of ethnic groups, provided sufficient data on the expected heterozygosities of all ethnicities worldwide as well as the genetic distances amongst them are available. In reality, however, the fact that the HGDP-CEPH sample provides such data for only 53 ethnic groups (or pairs thereof) implies that a straightforward application of this methodology would necessarily restrict the calculation of the index of contemporary diversity to a small set of countries. Moreover, unlike the earlier historical analysis, exploiting the predictive power of migratory distance from East Africa for genetic diversity would, by itself, be insufficient since, while this would overcome the problem of data limitations with respect to expected heterozygosities at the ethnic group level, it does not address the problem associated with limited data on genetic distances.

To surmount this issue, the current analysis appeals to a second prediction of the serial-founder effect regarding the genetic differentiation of populations through isolation by geographical



distance. Accordingly, in the process of the initial step-wise diffusion of the human species from Africa into the rest of the world, offshoot colonies residing at greater geographical distances from parental ones would also be more genetically differentiated from them. This would arise due to the larger number intervening migration steps, and a concomitantly larger number of genetic diversity sub-sampling events, that are associated with offshoots residing at locations farther away from parental colonies. Indeed, this second prediction of the serial-founder effect is borne out in the data as well. Based on data from Ramachandran et al. (2005), Figure 8 shows the strong positive correlation between pairwise migratory distances and pairwise genetic distances across all pairs of ethnic groups in the HGDP-CEPH sample. Specifically, according to the regression, variation in migratory distance explains 78% of the variation in  $F_{st}$  genetic distance across the 1378 ethnic group pairs. Moreover, the estimated OLS coefficient is highly statistically significant, possessing a t-statistic = 53.62, and suggests that *predicted*  $F_{st}$  genetic distance falls by 0.0617 percentage points for every 10,000 km increase in pairwise migratory distance. The construction of the index of diversity for contemporary national populations thus employs  $F_{st}$  genetic distance values *predicted* by pairwise migratory distances.

In particular, using the hypothetical example of a contemporary population comprised of two groups whose ancestors originate from countries  $A$  and  $B$ , the overall diversity of the combined population would be calculated as:

$$\hat{H}_{\text{exp}}^{AB} = \frac{\theta_A \hat{H}_{\text{exp}}^A(d_A) + \theta_B \hat{H}_{\text{exp}}^B(d_B)}{\left[1 - \hat{F}_{st}^{AB}(d_{AB})\right]}, \quad (7)$$

where, for  $i \in \{A, B\}$ ,  $\hat{H}_{\text{exp}}^i(d_i)$  denotes the expected heterozygosity predicted by the migratory distance,  $d_i$ , of country  $i$  from East Africa (i.e., the predicted genetic diversity of country  $i$  in the historical analysis); and  $\theta_i$  is the contribution of country  $i$ , as a result of post-Columbian migrations, to the combined population being considered. Moreover,  $\hat{F}_{st}^{AB}(d_{AB})$  is the genetic distance predicted by the migratory distance *between* countries  $A$  and  $B$ , obtained by applying the coefficients associated with the regression line depicted in Figure 8. In practice, since contemporary national populations are typically composed of more than two ethnic groups, the procedure outlined in equation (7) is applied recursively in order to incorporate a larger number of component ethnic groups in modern populations.

Reassuringly, the ancestry-adjusted measure of genetic diversity dominates the unadjusted measure in predicting economic development in the contemporary period. In line with the diversity hypothesis, Column 1 in Table 11 reveals a significant unconditional hump-shaped relationship between the adjusted measure of diversity and income per capita in the year 2000 CE. This relationship is depicted together with a non-parametric local polynomial regression line on the scatter plot in Figure 9. As in the historical analysis, the estimated quadratic fit falls within the 95% confidence interval band of the non-parametric relationship.

Column 2 establishes that the unconditional quadratic relationship from Column 1 remains qualitatively intact when conditioned for the impact of continent fixed effects. As revealed in Columns 3 and 4, however, while the unadjusted measure also possesses a significant unconditional hump-shaped relationship with income per capita across countries, the relationship disappears once the regression is augmented to account for continental dummies. Moreover, examining jointly the explanatory powers of the ancestry-adjusted and unadjusted measures of genetic diversity for income per capita, Columns 5 and 6 demonstrate the superior relative performance of the adjusted measure, regardless of whether continent fixed effects are accounted for by the analysis, lending

TABLE 11: Adjusted versus Unadjusted Diversity

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable is Log Income Per Capita in 2000 CE						
Predicted Diversity (Ancestry Adjusted)	556.439*** (129.697)	254.906*** (88.619)			533.983*** (164.216)	387.314** (188.300)
Predicted Diversity Sqr. (Ancestry Adjusted)	-397.224*** (90.784)	-176.907*** (62.730)			-377.365*** (117.645)	-273.925** (136.442)
Predicted Diversity (Unadjusted)			140.903*** (51.614)	10.152 (52.732)	1.670 (69.101)	-64.226 (81.419)
Predicted Diversity Sqr. (Unadjusted)			-107.686*** (38.133)	-7.418 (38.000)	-4.057 (52.990)	51.016 (64.295)
Continent Dummies	No	Yes	No	Yes	No	Yes
Observations	143	143	143	143	143	143
R-squared	0.13	0.47	0.08	0.45	0.14	0.48
P-value for:						
Joint Sig. of Adjusted Diversity and its Sqr.					0.009	0.038
Joint Sig. of Unadjusted Diversity and its Sqr.					0.399	0.741

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.  
 \*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

further credence *ex post* to the methodology employed in constructing the index of contemporary population diversity.<sup>49</sup>

#### 4.1.2 The Empirical Model

Maintaining symmetry with the earlier historical analysis, a regression specification similar to that employed for the historical regressions is adopted initially to examine the contemporary impact of genetic diversity along with the transition timing and land productivity channels. The current specification, however, is further augmented with controls for institutional, cultural, and additional geographical factors that have received attention in the literature. This permits the examination of the direct impact of the diversity channel, as opposed to its overall impact that additionally captures indirect effects potentially correlated with these other determinants.

Formally, the following specification is adopted as a baseline to examine the direct influence of contemporary population diversity on the modern world income distribution:

$$\ln y_i = \gamma_0 + \gamma_1 \hat{G}_i + \gamma_2 \hat{G}_i^2 + \gamma_3 \ln T_i + \gamma_4' \ln X_i + \gamma_5' \ln \Lambda_i + \gamma_6 \ln \Gamma_i + \eta_i, \quad (8)$$

where  $y_i$  is the income per capita of country  $i$  in the year 2000 CE;  $\hat{G}_i$  is the index of contemporary population diversity for country  $i$ , as discussed above;  $T_i$  and  $X_i$  are the Neolithic transition timing and land productivity controls for country  $i$ ;  $\Lambda_i$  is a vector of institutional and cultural controls

<sup>49</sup>Table D.3 in Appendix D establishes that migratory distance from Addis Ababa, adjusted to reflect the weighted average of migratory distances of the pre-Columbian ancestral populations of a country today, is the only distance concept that confers a significant hump-shaped effect on income per capita in 2000 CE. As shown in the table, the other distance concepts, including (i) the unadjusted measure of migratory distance from Addis Ababa (used in the historical analysis), (ii) the aerial distance from Addis Ababa, and (iii) the ancestry-adjusted aerial distance from Addis Ababa, do not confer any systematic non-monotonic effect on income per capita in 2000 CE, given that the ancestry-adjusted migratory distance measure is accounted for by the regression.

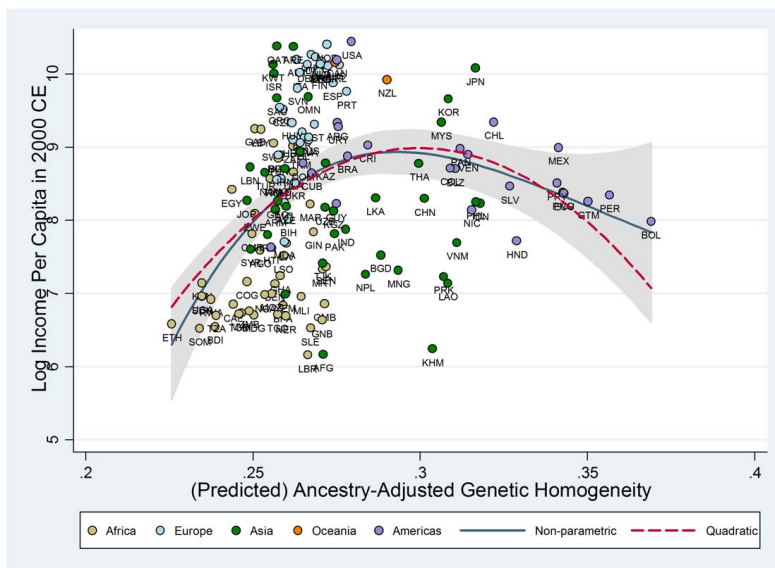


FIGURE 9: Adjusted Genetic Diversity and Income Per Capita in 2000 CE – The Unconditional Relationship

for country  $i$ ;  $\Gamma_i$  is a vector of additional geographical controls for country  $i$ ; and, finally,  $\eta_i$  is a country specific disturbance term.<sup>50</sup>

## 4.2 Empirical Findings

The empirical findings indicate that the highly significant hump-shaped effect of genetic diversity on macroeconomic outcomes in the pre-industrial period is present in the contemporary period as well. Furthermore, the persistent hump-shaped impact of genetic diversity on the pattern of comparative economic development is a direct effect that is not captured by contemporary geographical, institutional, and cultural factors. Moreover, in line with the theory, the findings suggest that optimal level of diversity has increased in the process of industrialization, as the beneficial forces associated with greater diversity have intensified in an environment characterized by more rapid technological progress.

Using a sample of 143 countries for which data are available for the entire set of control variables used in the baseline regression for the year 1500 CE, Column 1 of Table 12 reveals a significant hump-shaped effect of genetic diversity on income per capita in 2000 CE, accounting for the set of baseline controls employed in the historical analysis – i.e., the logs of the timing of the Neolithic transition, the percentage of arable land, absolute latitude and the suitability of land for agriculture, as well as continental fixed effects.<sup>51</sup> Further, as predicted by the theory, the optimal

<sup>50</sup>The data on income per capita are from the *Penn World Table*, version 6.2. The institutional and cultural controls include the social infrastructure index of Hall and Jones (1999), an index of institutionalized democracy from the Polity IV data set, legal origin dummies and the shares of the population affiliated with major world religions from the data set of La Porta et al. (1999), as well as the ethnic fractionalization index of Alesina et al. (2003). The additional geographical controls include the share of the population at risk of contracting falciparum malaria from Gallup and Sachs (2001), as well as the share of the population living in Köppen-Geiger tropical zones and distance from the nearest coast or sea-navigable river, both from the data set of Gallup et al. (1999). See Appendix B for further details.

<sup>51</sup>Tables E.5–E.6 in Appendix E present the relevant descriptive statistics for this 143-country sample.

TABLE 12: Diversity and Economic Development in 2000 CE and 1500 CE

	(1)	(2)	(3)	(4)
	Dependent Variable is:			
	Log Income Per Capita in 2000 CE			Log Population Density in 1500 CE
Predicted Diversity (Ancestry Adjusted)	204.610** (86.385)	237.238*** (85.031)	244.960*** (83.379)	
Predicted Diversity Sqr. (Ancestry Adjusted)	-143.437** (61.088)	-166.507*** (60.474)	-171.364*** (59.386)	
Predicted Diversity (Unadjusted)				198.587** (79.225)
Predicted Diversity Sqr. (Unadjusted)				-145.320*** (55.438)
Log Transition Timing (Ancestry Adjusted)		0.061 (0.262)	0.002 (0.305)	
Log Transition Timing (Unadjusted)	-0.151 (0.186)			1.238*** (0.241)
Log % of Arable Land	-0.110 (0.100)	-0.119 (0.107)	-0.137 (0.111)	0.378*** (0.108)
Log Absolute Latitude	0.164 (0.125)	0.172 (0.119)	0.192 (0.143)	-0.423*** (0.122)
Log Land Suitability	-0.193** (0.095)	-0.177* (0.102)	-0.189* (0.102)	0.264*** (0.095)
Log Population Density in 1500 CE			0.047 (0.097)	
Optimal Diversity	0.713*** (0.100)	0.712*** (0.036)	0.715*** (0.118)	0.683*** (0.095)
Continent Dummies	Yes	Yes	Yes	Yes
Observations	143	143	143	143
R-squared	0.57	0.57	0.57	0.68

*Note:* Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

level of diversity appears to increase in an environment characterized by more rapid technological progress. While the estimate for the optimal level in 1500 CE is 0.6833 (Column 4), the estimated optimum in 2000 CE, under the same specification, is 0.7132.

Interestingly, contrary to Diamond's (1997) conjecture, the timing of the Neolithic transition has no explanatory power for comparative development in the contemporary era. Moreover, as shown in Column 2, while the hump-shaped effect of diversity on income per capita in 2000 CE remains virtually intact, the impact of the Neolithic transition, adjusted to capture the average time elapsed since the pre-Columbian ancestral populations of each country today experienced the transition to agriculture (i.e., traits that are embodied in the country's population today, rather than the country's geographical attributes), remains statistically insignificant.

In particular, as established in Column 2, the estimated linear and quadratic coefficients on genetic diversity are both statistically significant at the 1% level. They imply that increasing the diversity of the most homogenous country in the sample (Bolivia) by 1 percentage point would raise its income per capita in 2000 CE by 29.01%, whereas decreasing the diversity of the most diverse country in the sample (Ethiopia) by 1 percentage point would raise its income per capita by 20.86%. Further, a 1 percentage point change in diversity (in either direction) at the optimum

level of 0.7124 would lower income per capita by 1.65%.<sup>52</sup>

Importantly, the hump-shaped effect of genetic diversity on income per capita in 2000 CE does not reflect an inertia originating from its effect on technology and, thus, population density in 1500 CE. As established in Column 4, the results are essentially unchanged if the regression accounts for the potentially confounding effect of population density in 1500 CE. Namely, the effect of genetic diversity on income per capita in 2000 CE does not operate through its impact on population density in the year 1500 CE.

The findings uncovered by the analysis thus far suggest that genetic diversity has a highly significant hump-shaped effect on income per capita in the year 2000 CE. Moreover, as established by the analysis to follow, this overall effect comprises a direct impact that does not operate through institutional, cultural and other geographical factors.

Using a sample of 109 countries for which data are available for the institutional and cultural controls that are employed in the examination, Column 1 of Table 13 demonstrates that genetic diversity has a hump-shaped effect on income per capita in the year 2000 CE, accounting for the set of baseline controls employed in the historical analysis – i.e., the logs of the weighted timing of the Neolithic transition, the percentage of arable land, and absolute latitude, as well as continental fixed effects.<sup>53</sup> The estimated linear and quadratic coefficients associated with the diversity channel are both statistically significant at the 1% level and the estimate for the optimal level of diversity is 0.7134. The regression in Column 2 examines the robustness of the results to the inclusion of a measure of institutional quality, as captured by the social infrastructure index of Hall and Jones (1999). The estimated hump-shaped effect of genetic diversity remains highly statistically significant and rather stable, while the optimal level of diversity increases to 0.7247. As indicated by the regression in Column 3, the inclusion of the Polity IV democracy index as an alternative measure of institutional quality does not affect the results and, since it is insignificant, it is dropped from subsequent regressions.

The regression in Column 4 is designed to examine whether the effect of genetic diversity operates via ethnic fractionalization. It demonstrates that the effect of genetic diversity is virtually unaffected by the potentially confounding impact of ethnic fractionalization.<sup>54</sup> While, as established earlier in the literature, ethnic fractionalization does indeed confer a significant adverse effect on income per capita in the year 2000 CE, the hump-shaped impact of genetic diversity remains highly statistically significant. Moreover, the estimate for the optimal level of diversity, 0.7243, is effectively unchanged in comparison to earlier columns.

Column 5 demonstrates the robustness of the hump-shaped effect of genetic diversity to the inclusion of additional cultural and institutional controls (i.e., legal origins and the fraction of the population affiliated with major religions). The coefficients associated with genetic diversity remain highly significant statistically and rather stable in magnitude, while the estimated optimal level of diversity, 0.7215, remains virtually intact.

---

<sup>52</sup>Following the earlier discussion regarding the expected heterozygosity index, these effects are therefore associated with a 0.01 change in the probability that two randomly selected individuals from a given population are genetically different from one another. See Footnote 29 for details on how these effects may be computed based on the estimated linear and quadratic coefficients associated with genetic diversity.

<sup>53</sup>The agricultural suitability index was not found to enter significantly in any of the specifications examined in Table 13 and is therefore dropped from the analysis. Tables E.7–E.8 in Appendix E present the relevant descriptive statistics for the 109-country sample employed in Tables 13–14.

<sup>54</sup>Results (not shown) from estimating a similar specification that included ethnic fractionalization square as an additional explanatory variable did not reveal any discernible non-monotonic impact of ethnic fractionalization on income per capita in 2000 CE. Importantly, the regression coefficients associated with genetic diversity, as well as the estimate for the optimal level of diversity, were unaffected.

TABLE 13: Diversity and Other Determinants of Economic Development in 2000 CE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	Full Sample	Full Sample	Full Sample	Full Sample	Full Sample	Schooling Sample	Schooling Sample
Dependent Variable is Log Income Per Capita in 2000 CE								
Predicted Diversity (Ancestry Adjusted)	315.282*** (85.309)	225.858*** (67.857)	204.102*** (69.201)	219.453*** (67.261)	245.377*** (69.200)	277.342*** (69.452)	271.249*** (73.818)	215.675*** (67.984)
Predicted Diversity Sqr. (Ancestry Adjusted)	-220.980*** (60.304)	-155.826*** (48.011)	-140.850*** (48.752)	-151.489*** (47.512)	-170.036*** (48.737)	-192.386*** (49.056)	-188.899*** (52.806)	-150.871*** (48.295)
Log Transition Timing (Ancestry Adjusted)	-0.273 (0.269)	-0.092 (0.200)	-0.062 (0.207)	0.032 (0.189)	0.352 (0.242)	0.396* (0.233)	0.112 (0.234)	-0.046 (0.208)
Log % of Arable Land	-0.218*** (0.061)	-0.159*** (0.049)	-0.163*** (0.048)	-0.171*** (0.049)	-0.211*** (0.047)	-0.183*** (0.051)	-0.084 (0.056)	-0.084 (0.056)
Log Absolute Latitude	0.123 (0.122)	0.083 (0.100)	0.080 (0.099)	0.038 (0.103)	0.119 (0.097)	0.009 (0.108)	0.007 (0.099)	-0.006 (0.087)
Social Infrastructure		2.359*** (0.269)	2.069*** (0.378)	2.389*** (0.257)	2.072*** (0.375)	1.826*** (0.417)	1.470*** (0.426)	0.880** (0.418)
Democracy			0.036 (0.031)					
Ethnic Fractionalization				-0.684** (0.277)	-0.505 (0.319)	-0.333 (0.280)	0.051 (0.280)	-0.122 (0.265)
% of Population at Risk of Contracting Malaria						-0.502 (0.351)	-0.734* (0.389)	-0.723** (0.353)
% of Population Living in Tropical Zones						-0.319 (0.204)	-0.411* (0.214)	-0.185 (0.199)
Mean Distance to Nearest Waterway						-0.368** (0.186)	-0.179 (0.216)	-0.062 (0.188)
Years of Schooling								0.134*** (0.042)
Optimal Diversity	0.713*** (0.014)	0.725*** (0.032)	0.725*** (0.045)	0.724*** (0.049)	0.722*** (0.014)	0.721*** (0.068)	0.718*** (0.052)	0.715*** (0.073)
OPEC Dummy	No	No	No	No	No	Yes	Yes	Yes
Legal Origin Dummies	No	No	No	No	Yes	Yes	Yes	Yes
Major Religion Shares	No	No	No	No	Yes	Yes	Yes	Yes
Observations	109	109	109	109	109	109	94	94
R-squared	0.74	0.84	0.85	0.85	0.87	0.90	0.91	0.93

Notes: Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses. All regressions include sub-Saharan Africa and continent dummies.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

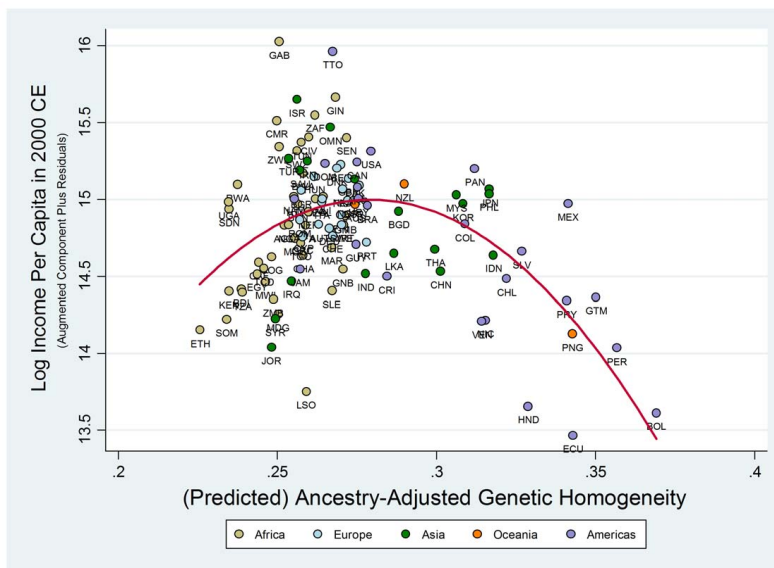


FIGURE 10: Adjusted Genetic Diversity and Income Per Capita in 2000 CE – Conditional on Transition Timing, Land Productivity, Institutional and Geographical Determinants, and Continental Fixed Effects

Finally, Column 6 demonstrates the robustness of the results to the inclusion of controls for the health environment (i.e., percentage of the population at risk of contracting malaria, and percentage of the population in tropical zones), additional geographical controls gauging access to waterways, and an OPEC dummy. The results in this column therefore reveal the direct effect of genetic diversity, once institutional, cultural, and geographical factors are accounted for. The direct hump-shaped impact of genetic diversity on log income per capita in 2000 CE, as established in Column 6, is depicted on the scatter plot in Figure 10.<sup>55</sup>

The coefficients associated with the diversity channel in Column 6 imply that: (i) increasing the diversity of the most homogenous country in the sample (Bolivia) by 1 percentage point would raise its income per capita in the year 2000 CE by 38.63%, (ii) decreasing the diversity of the most diverse country in the sample (Ethiopia) by 1 percentage point would raise its income per capita by 20.52%, (iii) a 1 percentage point change in genetic diversity (in either direction) at the optimum level of 0.7208 (that most closely resembles the U.S. diversity level of 0.7206) would lower income per capita by 1.91%, (iv) increasing the diversity of Bolivia to the level prevalent in the U.S. would increase Bolivia’s per capita income by a factor of 4.73, closing the income gap between the two countries from 12:1 to 2.5:1, and (v) decreasing the diversity of Ethiopia to the level prevalent in the U.S. would increase Ethiopia’s per capita income by a factor of 1.73 and thus close the income gap between the two countries from 47:1 to 27:1.

Moreover, as reported in Column 8, even if one accounts for the contribution of human capital formation over the time period 1960–2000, the hump-shaped effect of genetic diversity on income per capita in 2000 CE remains highly statistically significant. Further, the estimate for the optimal level of diversity drops only moderately from 0.7180 (as presented in Column 7, that

<sup>55</sup>Similar to Figures 5–7 in the historical analysis, Figure 10 is an “augmented component plus residual” plot. See Footnote 37 for an explanation of this type of plot. Plots depicting the partial regression lines associated with the first-order and second-order effects of genetic homogeneity on log income per capita in 2000 CE are presented in Figures C.2(a)–C.2(b) in Appendix C.

TABLE 14: Addressing Endogenous Post-Columbian Migrations

	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample	Non OECD	w/o Neo Europes	w/o Latin America	w/o Sub Saharan	>0.97 Indigenous
Dependent Variable is Log Income Per Capita in 2000 CE						
Predicted Diversity (Ancestry Adjusted)	277.342*** (69.452)	271.979*** (84.232)	261.367*** (69.946)	412.222*** (148.781)	264.805** (107.492)	304.735** (111.588)
Predicted Diversity Sqr. (Ancestry Adjusted)	-192.386*** (49.056)	-188.974*** (59.200)	-181.811*** (49.261)	-287.067*** (102.021)	-183.863** (77.684)	-213.389** (77.255)
Log Transition Timing (Ancestry Adjusted)	0.396* (0.233)	0.390 (0.281)	0.355 (0.231)	0.518* (0.298)	0.068 (0.442)	0.448* (0.254)
Log % of Arable Land	-0.183*** (0.051)	-0.236*** (0.060)	-0.201*** (0.055)	-0.189*** (0.050)	-0.211** (0.097)	-0.104 (0.061)
Log Absolute Latitude	0.009 (0.108)	-0.021 (0.119)	-0.025 (0.111)	-0.139 (0.126)	0.218 (0.242)	-0.074 (0.130)
Social Infrastructure	1.826*** (0.417)	1.313** (0.579)	1.416*** (0.507)	2.044*** (0.545)	1.585*** (0.486)	1.311* (0.716)
Ethnic Fractionalization	-0.333 (0.280)	-0.437 (0.375)	-0.390 (0.300)	-0.752** (0.348)	0.104 (0.408)	-0.044 (0.412)
% of Population at Risk of Contracting Malaria	-0.502 (0.351)	-0.605 (0.381)	-0.591 (0.370)	-0.308 (0.486)	-0.425 (0.581)	-0.153 (0.434)
% of Population Living in Tropical Zones	-0.319 (0.204)	-0.196 (0.239)	-0.302 (0.219)	-0.520** (0.252)	-0.528 (0.341)	-0.339 (0.312)
Mean Distance to Nearest Waterway	-0.368** (0.186)	-0.387* (0.222)	-0.452** (0.210)	-0.494*** (0.186)	-0.743 (0.469)	-0.367* (0.201)
Optimal Diversity	0.721*** (0.068)	0.720*** (0.085)	0.719*** (0.015)	0.718*** (0.023)	0.720*** (0.180)	0.714*** (0.012)
Observations	109	83	105	87	71	37
R-squared	0.90	0.82	0.89	0.93	0.86	0.98

Notes: Bootstrapped standard errors, accounting for the use of generated regressors, are reported in parentheses. All regressions include controls for major religion shares as well as OPEC, legal origin, sub-Saharan Africa and continent dummies.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.



accounts for the smaller sample of 94 countries for which data on education and all other variables are available) to 0.7148.

Reassuringly, the highly significant and stable hump-shaped effect of genetic diversity on income per capita in 2000 CE is not an artifact of post-colonial migration towards prosperous countries and the concomitant increase in ethnic diversity in these economies. Importantly, for the sample of countries whose national population is largely indigenous to their current geographical location, the hump-shaped effect of genetic diversity on contemporary income per capita is highly significant and virtually identical to the one observed in the entire sample. Thus, since genetic diversity in these populations is the level of diversity *predicted* by migratory distance from East Africa, rather than the actual one, the potential concern about endogeneity between genetic diversity and income per capita in the modern world is alleviated.

In particular, as established in Table 14, the hump-shaped effect of genetic diversity remains highly significant and the optimal diversity estimate remains virtually intact if the sample is restricted to (a) non-OECD economies (i.e., economies that were less attractive to migrants) in Column 2, (b) non Neo-European countries (i.e., excluding USA, Canada, New Zealand and Australia) in Column 3, (c) non-Latin American countries in Column 4, (d) non Sub-Saharan African countries in Column 5, and (e) countries whose indigenous population is larger than 97% of the entire population (i.e., under conditions that virtually eliminate the role of migration in the creation of diversity over the last 500 years) in Column 6.<sup>56</sup>

## 5 Concluding Remarks

This paper argues that deep-rooted factors, determined tens of thousands of years ago, had a significant effect on the course of economic development from the dawn of human civilization to the contemporary era. It advances and empirically establishes the hypothesis that, in the course of the exodus of *Homo sapiens* out of Africa, variation in migratory distance from the cradle of humankind to various settlements across the globe affected genetic diversity, and has had a long-lasting effect on the pattern of comparative economic development that is not captured by geographical, institutional, and cultural factors.

The level of genetic diversity within a society is found to have a hump-shaped effect on development outcomes in the pre-colonial era, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. Moreover, the level of genetic diversity in each country today (as determined by the genetic diversities and genetic distances amongst its ancestral populations), has a similar non-monotonic effect on income per capita in the modern world. While the intermediate level of genetic diversity prevalent among Asian and European populations has been conducive for development, the high degree of diversity among African populations and the low degree of diversity among Native American populations have been a detrimental force in the development of these regions. Further, the optimal level of diversity appears to have increased in the process of industrialization, as the beneficial forces associated with greater diversity have intensified in an environment characterized by more rapid technological progress.

The direct effect of genetic diversity on contemporary income per capita, once institutional, cultural, and geographical factors are accounted for, indicates that: (i) increasing the diversity of the most homogenous country in the sample (Bolivia) by 1 percentage point would raise its income

---

<sup>56</sup>This result reflects the well-known fact from the field of population genetics that the overwhelming majority of genetic diversity in human populations stems from the diversity *within* groups, as opposed to the diversity *between* groups (see, e.g., Lewontin, 1972; Barbujani et al., 1997).

per capita in the year 2000 CE by 38.63%, (ii) decreasing the diversity of the most diverse country in the sample (Ethiopia) by 1 percentage point would raise its income per capita by 20.52%, (iii) a 1 percentage point change in genetic diversity (in either direction) at the optimum level of 0.7208 (that most closely resembles the U.S. diversity level of 0.7206) would lower income per capita by 1.91%, (iv) increasing Bolivia's diversity to the optimum level prevalent in the U.S. would increase Bolivia's per capita income by a factor of 4.73, closing the income gap between the U.S. and Bolivia from 12:1 to 2.5:1, and (v) decreasing Ethiopia's diversity to the optimum level of the U.S. would increase Ethiopia's per capita income by a factor of 1.73 and, thus, close the income gap between the U.S. and Ethiopia from 47:1 to 27:1.

## A The HGDP-CEPH Sample of 53 Ethnic Groups

Ethnic Group	Migratory Distance (in km)	Country	Region
Bantu (Kenya)	1,338.94	Kenya	Africa
Bantu (Southeast)	4,306.19	South Africa	Africa
Bantu (Southwest)	3,946.44	Namibia	Africa
Biaka Pygmy	2,384.86	Central African Republic	Africa
Mandenka	5,469.91	Senegal	Africa
Mbuti Pygmy	1,335.50	Zaire	Africa
San	3,872.42	Namibia	Africa
Yoruba	3,629.65	Nigeria	Africa
Bedouin	2,844.95	Israel	Middle East
Druze	2,887.25	Israel	Middle East
Mozabite	4,418.17	Algeria	Middle East
Palestinian	2,887.25	Israel	Middle East
Adygei	4,155.03	Russia	Europe
Basque	6,012.26	France	Europe
French	5,857.48	France	Europe
Italian	5,249.04	Italy	Europe
Orcadian	6,636.69	United Kingdom	Europe
Russian	5,956.40	Russia	Europe
Sardinian	5,305.81	Italy	Europe
Tuscan	5,118.37	Italy	Europe
Balochi	5,842.06	Pakistan	Asia
Brahui	5,842.06	Pakistan	Asia
Burusho	6,475.60	Pakistan	Asia
Cambodian	10,260.55	Cambodia	Asia
Dai	9,343.96	China	Asia
Daur	10,213.13	China	Asia
Han	10,123.19	China	Asia
Han (North China)	9,854.75	China	Asia
Hazara	6,132.57	Pakistan	Asia
Hezhen	10,896.21	China	Asia
Japanese	11,762.11	Japan	Asia
Kalash	6,253.62	Pakistan	Asia
Lahu	9,299.63	China	Asia
Makrani	5,705.00	Pakistan	Asia
Miao	9,875.32	China	Asia
Mongola	9,869.85	China	Asia
Naxi	9,131.37	China	Asia
Oroqen	10,290.53	China	Asia
Pathan	6,178.76	Pakistan	Asia
She	10,817.81	China	Asia
Sindhi	6,201.70	Pakistan	Asia
Tu	8,868.14	China	Asia
Tujia	9,832.50	China	Asia
Uygur	7,071.97	China	Asia
Xibo	7,110.29	China	Asia
Yakut	9,919.11	Russia (Siberia)	Asia
Yi	9,328.79	China	Asia
Melanesian	16,168.51	Papua New Guinea	Oceania
Papuan	14,843.12	Papua New Guinea	Oceania
Colombian	22,662.78	Colombia	Americas
Karitiana	24,177.34	Brazil	Americas
Maya	19,825.71	Mexico	Americas
Pima	18,015.79	Mexico	Americas

## B Variable Definitions and Sources

### *Outcome Variables:*

**Population Density in 1 CE, 1000 CE, and 1500 CE.** Population density (in persons per square km) for given year is calculated as population in that year, as reported by McEvedy and Jones (1978), divided by total land area, as reported by the World Bank’s *World Development Indicators*. The cross-sectional unit of observation in McEvedy and Jones’ (1978) data set is a region delineated by its international borders in 1975. Historical population estimates are provided for regions corresponding to either individual countries or, in some cases, to sets comprised of 2–3 neighboring countries (e.g., India, Pakistan and Bangladesh). In the latter case, a set-specific population density figure is calculated based on total land area and the figure is then assigned to each of the component countries in the set. The same methodology is also employed to obtain population density for countries that exist today but were part of a larger political unit (e.g., the former Yugoslavia) in 1975. The data reported by the authors are based on a wide variety of country and region-specific historical sources, the enumeration of which would be impractical for this appendix. The interested reader is therefore referred to McEvedy and Jones (1978) for more details on the original data sources cited therein.

**Income Per Capita in 2000 CE.** Real GDP per capita, in constant 2000 international dollars, as reported by the *Penn World Table*, version 6.2.

### *Genetic Diversity Variables:*

**Observed Genetic Diversity in the Limited Historical Sample.** The average expected heterozygosity across ethnic groups from the Human Genome Diversity Cell Line Panel that are located within a given country. The expected heterozygosities of the ethnic groups are from Ramachandran et al. (2005).

**Predicted Genetic Diversity in the Extended Historical Sample.** The expected heterozygosity (genetic diversity) of a given country as predicted by (the extended sample definition of) migratory distance from Addis Ababa (Ethiopia). This measure is calculated by applying the regression coefficients obtained from regressing expected heterozygosity on migratory distance at the ethnic group level, using the worldwide sample of 53 ethnic groups from the Human Genome Diversity Cell Line Panel. The expected heterozygosities and geographical coordinates of the ethnic groups are from Ramachandran et al. (2005).

**Predicted Genetic Diversity (Ancestry Adjusted).** The expected heterozygosity (genetic diversity) of a country’s population, predicted by migratory distances from Addis Ababa (Ethiopia) to the year 1500 CE locations of the ancestral populations of the country’s component ethnic groups in 2000 CE, as well as by pairwise migratory distances between these ancestral populations. The source countries of the year 1500 CE ancestral populations are identified from the *World Migration Matrix, 1500–2000*, discussed in Putterman and Weil (2009), and the modern capital cities of these countries are used to compute the aforementioned migratory distances. The measure of genetic diversity is then calculated by applying (i) the regression coefficients obtained from regressing expected heterozygosity on migratory distance from Addis Ababa at the ethnic group level, using the worldwide sample of 53 ethnic groups from the Human Genome Diversity Cell Line Panel, (ii) the regression coefficients obtained from regressing pairwise F<sub>st</sub> genetic distances on pairwise migratory distances between these ethnic groups, and (iii) the ancestry weights representing the fractions of the year 2000 CE population (of the country for which the measure is being computed) that can trace their ancestral origins to different source countries in the year 1500 CE. The construction of this measure is discussed in detail in Section 4.1.1. The expected heterozygosities, geographical coordinates, and

pairwise F-st genetic distances of the 53 ethnic groups are from Ramachandran et al. (2005). The ancestry weights are from the *World Migration Matrix, 1500–2000*.

*Distance Variables:*

**Migratory Distance from Addis Ababa in the Limited Historical Sample.** The average migratory distance across ethnic groups from the Human Genome Diversity Cell Line Panel that are located within a given country. The migratory distance of an ethnic group is the great circle distance from Addis Ababa (Ethiopia) to the location of the group, along a land-restricted path forced through one or more of five intercontinental waypoints, including Cairo (Egypt), Istanbul (Turkey), Phnom Penh (Cambodia), Anadyr (Russia) and Prince Rupert (Canada). Distances are calculated using the Haversine formula and are measured in units of 1,000 km. The geographical coordinates of the ethnic groups and the intercontinental waypoints are from Ramachandran et al. (2005).

**Migratory Distance from Addis Ababa in the Extended Historical Sample.** The great circle distance from Addis Ababa (Ethiopia) to the country’s modern capital city, along a land-restricted path forced through one or more of five aforementioned intercontinental waypoints. Distances are calculated using the Haversine formula and are measured in units of 1,000 km. The geographical coordinates of the intercontinental waypoints are from Ramachandran et al. (2005), while those of the modern capital cities are from the CIA’s *World Factbook*.

**Migratory Distance from Addis Ababa (Ancestry Adjusted).** The cross-country weighted average of (the extended sample definition of) migratory distance from Addis Ababa (Ethiopia), 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 (2009).

**Migratory Distance from a “Placebo” Point of Origin.** The great circle distance from a “placebo” location (i.e., other than Addis Ababa, Ethiopia) to the country’s modern capital city, along a land-restricted path forced through one or more of five aforementioned intercontinental waypoints. Distances are calculated using the Haversine formula and are measured in units of 1,000 km. The geographical coordinates of the intercontinental waypoints are from Ramachandran et al. (2005), while those of the modern capital cities are from the CIA’s *World Factbook*. The “placebo” locations for which results are presented in the text include London (UK), Tokyo (Japan), and Mexico City (Mexico).

**Aerial Distance from Addis Ababa.** The great circle distance “as the crow flies” from Addis Ababa (Ethiopia) to the country’s modern capital city. Distances are calculated using the Haversine formula and are measured in units of 1,000 km. The geographical coordinates of capital cities are from the CIA’s *World Factbook*.

**Aerial Distance from Addis Ababa (Ancestry Adjusted).** The cross-country weighted average of aerial distance from Addis Ababa, 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 from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2009).

**Distance to Frontier in 1 CE, 1000 CE, and 1500 CE.** The great circle distance from a country’s capital city to the closest regional technological frontier for a given year. The year-specific set of regional frontiers

comprises the two most populous cities, reported for that year and belonging to different civilizations or sociopolitical entities, from each of Africa, Europe, Asia, and the Americas. Distances are calculated using the Haversine formula and are measured in km. The historical urban population data, used to identify the frontiers, are obtained from Chandler (1987) and Modelski (2003), and the geographical coordinates of ancient urban centers are obtained using Wikipedia.

**Human Mobility Index.** The average migratory distance across ethnic groups from the Human Genome Diversity Cell Line Panel that are located within a given country. The migratory distance of an ethnic group is the distance from Addis Ababa (Ethiopia) to the location of the group, along an “optimal” land-restricted path that minimizes the time cost of travelling on the surface of the Earth. The optimality of a path is determined by incorporating information on natural impediments to human spatial mobility such as the meteorological and topographical conditions prevalent along the path, as well as information on the time cost of travelling under such conditions. Distances are measured in units of 1,000 km. The geographical coordinates of the ethnic groups are from Ramachandran et al. (2005). The methodology underlying the construction of this index is discussed in greater detail by Ashraf et al. (2010).

*Transition Timing Variables:*

**Neolithic Transition Timing.** The number of thousand years elapsed, until the year 2000, since the majority of the population residing within a country’s modern national borders began practicing sedentary agriculture as the primary mode of subsistence. This measure, reported by Putterman (2008), is compiled using a wide variety of both regional and country-specific archaeological studies as well as more general encyclopedic works on the transition from hunting and gathering to agriculture during the Neolithic. The reader is referred to the author’s web site for a detailed description of the primary and secondary data sources employed by the author in the construction of this variable.

**Neolithic Transition Timing (Ancestry Adjusted).** 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 (2009).

*Geographical Variables:*

**Percentage of Arable Land.** The fraction of a country’s total land area that is arable, as reported by the World Bank’s *World Development Indicators*.

**Absolute Latitude.** The absolute value of the latitude of a country’s approximate geodesic centroid, as reported by the CIA’s *World Factbook*.

**Land Suitability for Agriculture.** An index of the suitability of land for agriculture, based on geospatial soil pH and temperature data, reported at a half-degree resolution by Ramankutty et al. (2002) and aggregated to the country level by Michalopoulos (2008).

**Mean Elevation.** The mean elevation of a country in km above sea level, calculated using geospatial elevation data reported by the G-ECON project (Nordhaus, 2006) at a 1-degree resolution. The measure is thus the average elevation across the grid cells within a country. The interested reader is referred to the G-ECON project web site for additional details.

**Terrain Roughness.** The degree of terrain roughness of a country, calculated using geospatial surface undulation data reported by the G-ECON project (Nordhaus, 2006) at a 1-degree resolution. The measure is thus the average degree of terrain roughness across the grid cells within a country. The reader is referred to the G-ECON project web site for additional details.

**Mean Distance to Nearest Waterway.** The distance, in thousands of km, from a GIS grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country. This variable was originally constructed by Gallup et al. (1999) and is part of Harvard University's CID Research Datasets on *General Measures of Geography*.

**Percentage of Land within 100 km of Waterway.** The percentage of a country's total land area that is located within 100 km of an ice-free coastline or sea-navigable river. This variable was originally constructed by Gallup et al. (1999) and is part of Harvard University's CID Research Datasets on *General Measures of Geography*.

**Percentage of Population Living in Tropical Zones.** The percentage of a country's population in 1995 that resided in areas classified as tropical by the Köppen-Geiger climate classification system. This variable was originally constructed by Gallup et al. (1999) and is part of Harvard University's CID Research Datasets on *General Measures of Geography*.

**Percentage of Population at Risk of Contracting Malaria.** The percentage of a country's population in 1994 residing in regions of high malaria risk, multiplied by the proportion of national cases involving the fatal species of the malaria pathogen, *P. falciparum* (as opposed to other largely non-fatal species). This variable was originally constructed by Gallup and Sachs (2001) and is part of Columbia University's Earth Institute data set on malaria.

**Climate.** An index of climatic suitability for agriculture, based on the Köppen-Geiger climate classification system. This variable is obtained from the data set of Olsson and Hibbs (2005).

**Orientation of Continental Axis.** The orientation of a continent (or landmass) along a North-South or East-West axis. This measure, reported in the data set of Olsson and Hibbs (2005), is calculated as the ratio of the largest longitudinal (East-West) distance to the largest latitudinal (North-South) distance of the continent (or landmass).

**Size of Continent.** The total land area of a continent (or landmass) as reported in the data set of Olsson and Hibbs (2005).

**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. This variable is obtained from the data set of 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. This variable is obtained from the data set of Olsson and Hibbs (2005).

*Institutional, Cultural, and Human Capital Variables:*

**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.

**Democracy.** The 1960–2000 mean of an index that quantifies the extent of institutionalized democracy, as reported in the Polity IV data set. The Polity IV democracy index for a given year is an 11-point categorical variable (from 0 to 10) that is additively derived from Polity IV codings on the (i) competitiveness of political participation, (ii) openness of executive recruitment, (iii) competitiveness of executive recruitment, and (iv) constraints on the chief executive.

**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.

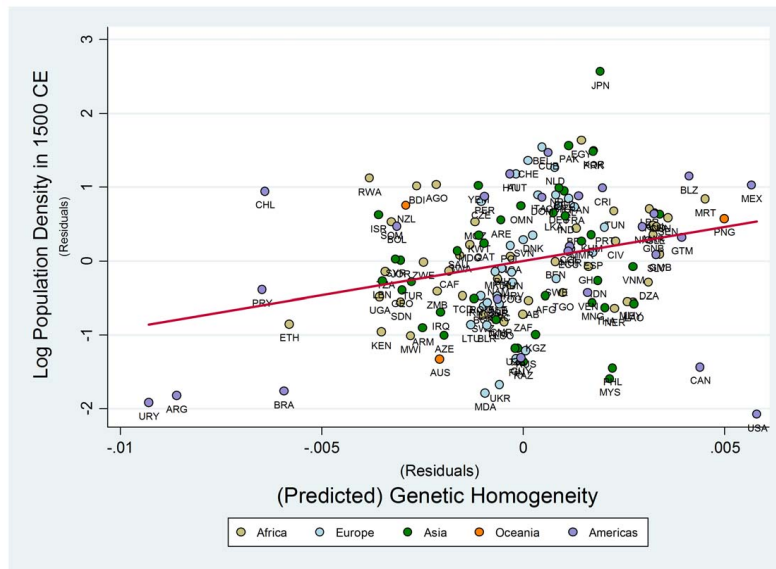
**Major Religion Shares.** A set of variables, from La Porta et al. (1999), that identifies the percentage of a country’s population belonging to the three most widely spread religions of the world. The religions identified are: (i) Roman Catholic, (ii) Protestant, and (iii) Muslim.

**Ethnic Fractionalization.** A fractionalization index, constructed by Alesina et al. (2003), that captures the probability that two individuals, selected at random from a country’s population, will belong to different ethnic groups.

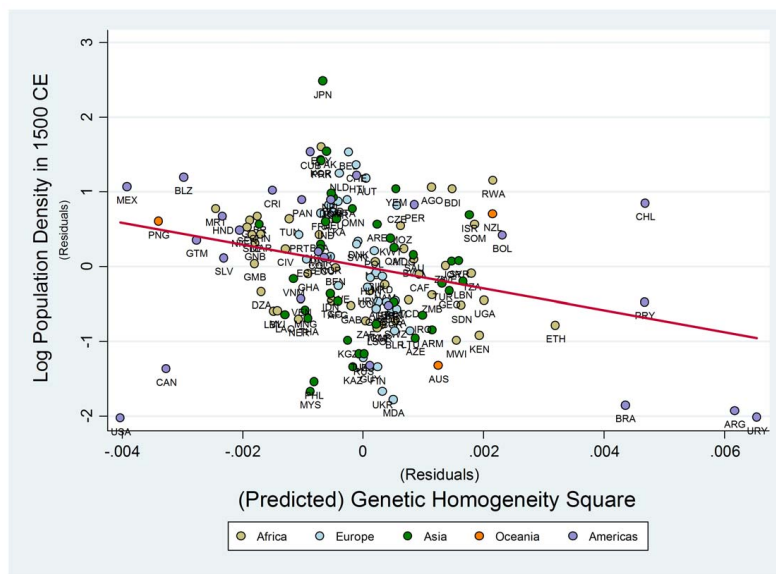
**Average Years of Schooling.** The mean, over the 1960–2000 time period, of the 5-yearly figure, reported by Barro and Lee (2001), on average years of schooling amongst the population aged 25 and over.



## C Supplementary Figures

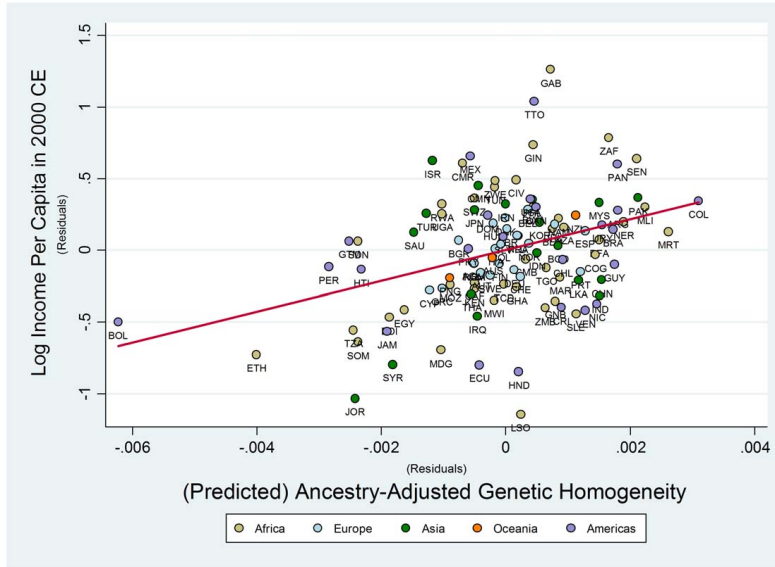


(a) The First-Order Effect

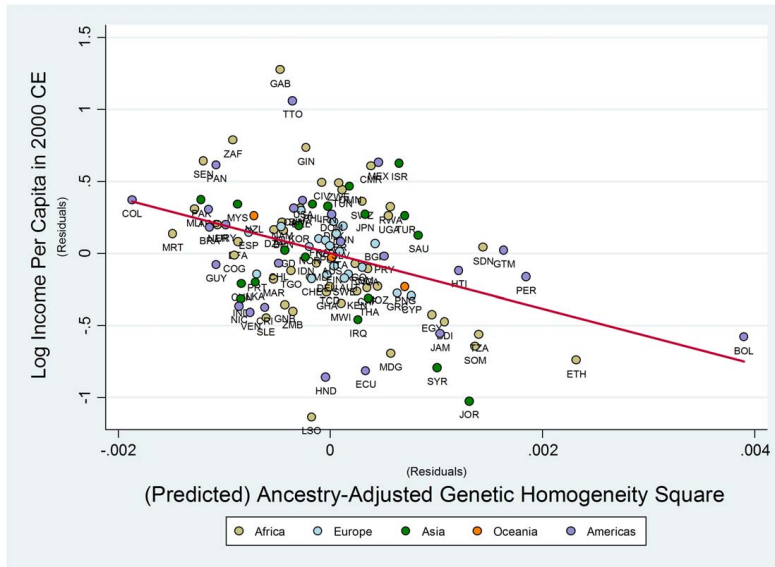


(b) The Second-Order Effect

FIGURE C.1: The First- and Second-Order Partial Effects of Predicted Diversity on Population Density in 1500 CE – Conditional on Transition Timing, Land Productivity, and Continental Fixed Effects

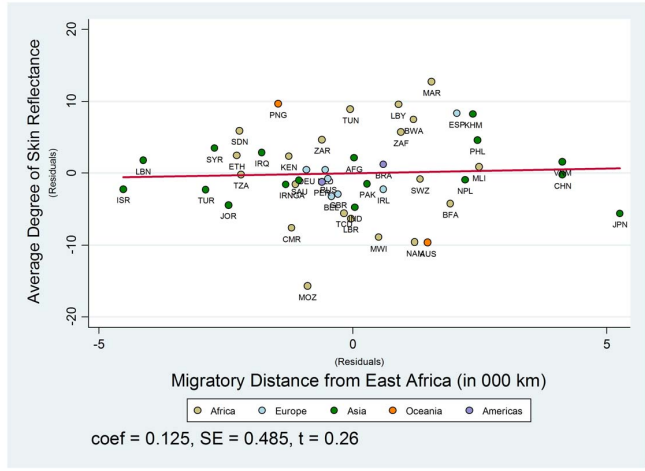


(a) The First-Order Effect

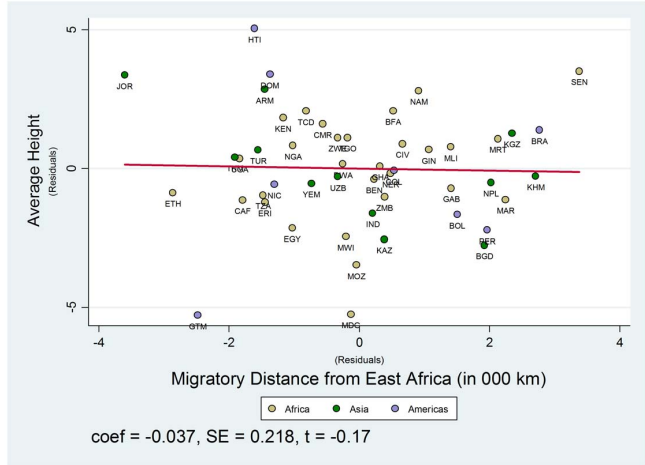


(b) The Second-Order Effect

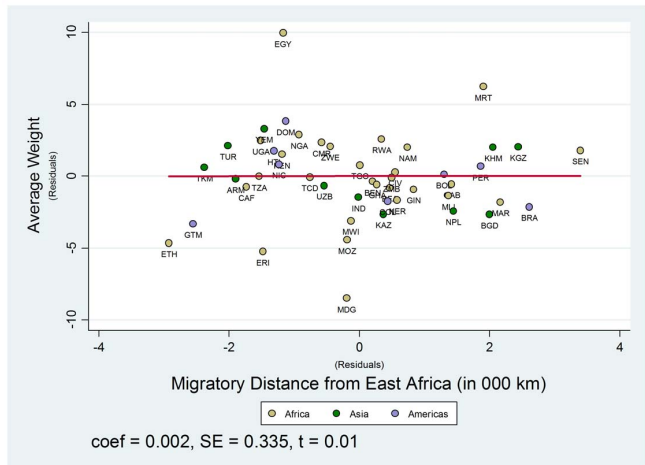
FIGURE C.2: The First- and Second-Order Partial Effects of Adjusted Diversity on Income Per Capita in 2000 CE – Conditional on Transition Timing, Land Productivity, Institutional and Geographical Determinants, and Continental Fixed Effects



(a) Migratory Distance and Average Skin Reflectance



(b) Migratory Distance and Average Height



(c) Migratory Distance and Average Weight

FIGURE C.3: Migratory Distance and Some Average Physiological Characteristics of Populations

## D Supplementary Results

TABLE D.1: Results of Table 1 with Correction for Spatial Dependence in Errors

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable is Log Population Density in 1500 CE				
Observed Diversity	413.504*** [85.389]			225.440*** [55.428]	203.814*** [65.681]
Observed Diversity Sqr.	-302.647*** [64.267]			-161.158*** [42.211]	-145.717*** [53.562]
Log Transition Timing		2.396*** [0.249]		1.214*** [0.271]	1.135*** [0.367]
Log % of Arable Land			0.730*** [0.263]	0.516*** [0.132]	0.545*** [0.178]
Log Absolute Latitude			0.145 [0.180]	-0.162* [0.084]	-0.129 [0.101]
Log Land Suitability			0.734* [0.376]	0.571** [0.240]	0.587** [0.233]
Continent Dummies	No	No	No	No	Yes
Observations	21	21	21	21	21
R-squared	0.42	0.54	0.57	0.89	0.90

*Notes:* Standard errors corrected for spatial autocorrelation, following Conley (1999), are reported in brackets. To perform this correction, the spatial distribution of observations was specified on the Euclidean plane using aerial distances between all pairs in the sample, and the autocorrelation was modelled as declining linearly away from each location upto a threshold of 5,000 km. This threshold excludes spatial interactions between the Old World and the New World, which is appropriate given the historical period being considered.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

TABLE D.2: Results of Table 2 with Correction for Spatial Dependence in Errors

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	Spatial GMM	Spatial GMM
Dependent Variable is Log Population Density in 1500 CE						
Observed Diversity		255.219*** [77.933]		361.420*** [108.692]	242.190*** [76.047]	203.164*** [67.420]
Observed Diversity Sqr.		-209.808*** [58.315]		-268.514*** [77.740]	-173.830*** [57.614]	-147.461*** [53.983]
Migratory Distance	0.505*** [0.110]	0.070 [0.138]				
Migratory Distance Sqr.	-0.023*** [0.004]	-0.014* [0.008]				
Mobility Index			0.353*** [0.108]	0.051 [0.125]		
Mobility Index Sqr.			-0.012*** [0.003]	-0.003 [0.005]		
Log Transition Timing					1.128*** [0.234]	1.027*** [0.366]
Log % of Arable Land					0.513*** [0.179]	0.570*** [0.219]
Log Absolute Latitude					-0.127 [0.097]	-0.130* [0.077]
Log Land Suitability					0.578*** [0.188]	0.591*** [0.207]
Continent Dummies	No	No	No	No	No	Yes
Observations	21	21	18	18	21	21
R-squared	0.34	0.46	0.30	0.43	–	–
Overid. Restrictions Test (P-value)					0.775	0.547

*Notes:* Standard errors corrected for spatial autocorrelation, following Conley (1999), are reported in brackets. To perform this correction, the spatial distribution of observations was specified on the Euclidean plane using aerial distances between all pairs in the sample, and the autocorrelation was modelled as declining linearly away from each location upto a threshold of 5,000 km. This threshold effectively excludes spatial interactions between the Old World and the New World, which is appropriate given the historical period being considered. Columns 5–6 present the results from estimating the corresponding 2SLS specifications in Table 2 using Conley’s (1999) spatial GMM estimation procedure.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

TABLE D.3: Adjusted Migratory Distance vs. Alternative Distances

	(1)	(2)	(3)	(4)
Dependent Variable is				
Log Income Per Capita in 2000 CE				
Migratory Distance (Ancestry Adjusted)	0.601*** (0.073)	0.499*** (0.124)	0.532*** (0.160)	0.564** (0.235)
Migratory Distance Sqr. (Ancestry Adjusted)	-0.030*** (0.004)	-0.026*** (0.006)	-0.027*** (0.008)	-0.029*** (0.010)
Migratory Distance (Unadjusted)		0.078 (0.084)		
Migratory Distance Sqr. (Unadjusted)		-0.002 (0.003)		
Aerial Distance (Unadjusted)			0.064 (0.201)	
Aerial Distance Sqr. (Unadjusted)			-0.002 (0.011)	
Aerial Distance (Ancestry Adjusted)				0.043 (0.330)
Aerial Distance Sqr. (Ancestry Adjusted)				-0.001 (0.018)
Observations	109	109	109	109
R-squared	0.28	0.29	0.29	0.29

*Note:* Heteroskedasticity robust standard errors are reported in parentheses.

\*\*\* Significant at 1%, \*\* Significant at 5%, \* Significant at 10%.

## E Descriptive Statistics

TABLE E.1: Summary Statistics for the 21-Country Historical Sample

	Obs.	Mean	S.D.	Min.	Max.
Log Population Density in 1500 CE	21	1.169	1.756	-2.135	3.842
Observed Genetic Diversity	21	0.713	0.056	0.552	0.770
Migratory Distance from Addis Ababa	21	8.238	6.735	1.335	24.177
Human Mobility Index	18	10.965	8.124	2.405	31.360
Log Neolithic Transition Timing	21	8.342	0.539	7.131	9.259
Log % of Arable Land	21	2.141	1.168	-0.799	3.512
Log Absolute Latitude	21	2.739	1.178	0.000	4.094
Log Land Suitability for Agriculture	21	-1.391	0.895	-3.219	-0.288

TABLE E.2: Pairwise Correlations for the 21-Country Historical Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Log Population Density in 1500 CE	1.000						
(2) Observed Genetic Diversity	0.244	1.000					
(3) Migratory Distance from Addis Ababa	-0.226	-0.968	1.000				
(4) Human Mobility Index	-0.273	-0.955	0.987	1.000			
(5) Log Neolithic Transition Timing	0.735	-0.117	0.024	0.011	1.000		
(6) Log % of Arable Land	0.670	0.172	-0.183	-0.032	0.521	1.000	
(7) Log Absolute Latitude	0.336	0.055	-0.012	0.044	0.392	0.453	1.000
(8) Log Land Suitability for Agriculture	0.561	-0.218	0.282	0.245	0.299	0.376	0.049

TABLE E.3: Summary Statistics for the 145-Country Historical Sample

	Obs.	Mean	S.D.	Min.	Max.
Log Population Density in 1500 CE	145	0.881	1.500	-3.817	3.842
Log Population Density in 1000 CE	140	0.463	1.445	-4.510	2.989
Log Population Density in 1 CE	126	-0.070	1.535	-4.510	3.170
Predicted Genetic Diversity	145	0.711	0.053	0.572	0.774
Log Neolithic Transition Timing	145	8.343	0.595	5.991	9.259
Log % of Arable Land	145	2.232	1.203	-2.120	4.129
Log Absolute Latitude	145	3.003	0.924	0.000	4.159
Log Land Suitability for Agriculture	145	-1.409	1.313	-5.857	-0.041
Log Distance to Frontier in 1500 CE	145	7.309	1.587	0.000	9.288
Log Distance to Frontier in 1000 CE	145	7.406	1.215	0.000	9.258
Log Distance to Frontier in 1 CE	145	7.389	1.307	0.000	9.261
Mean Elevation	145	0.555	0.481	0.024	2.674
Terrain Roughness	145	0.178	0.135	0.013	0.602
Mean Distance to Nearest Waterway	145	0.350	0.456	0.014	2.386
% of Land within 100 km of Waterway	145	0.437	0.368	0.000	1.000
Climate	96	1.531	1.046	0.000	3.000
Orientation of Continental Axis	96	1.521	0.685	0.500	3.000
Size of Continent	96	30.608	13.605	0.065	44.614
Domesticable Plants	96	13.260	13.416	2.000	33.000
Domesticable Animals	96	3.771	4.136	0.000	9.000
Migratory Distance from Addis Ababa	145	8.399	6.970	0.000	26.771
Aerial Distance from Addis Ababa	145	6.003	3.558	0.000	14.420
Migratory Distance from London	145	8.884	7.104	0.000	26.860
Migratory Distance from Tokyo	145	11.076	3.785	0.000	19.310
Migratory Distance from Mexico City	145	15.681	6.185	0.000	25.020



TABLE E.4: Pairwise Correlations for the 145-Country Historical Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Log Population Density in 1500 CE	1.000											
(2) Log Population Density in 1000 CE	0.963	1.000										
(3) Log Population Density in 1 CE	0.876	0.936	1.000									
(4) Predicted Genetic Diversity	0.391	0.312	0.341	1.000								
(5) Log Neolithic Transition Timing	0.511	0.567	0.645	0.275	1.000							
(6) Log % of Arable Land	0.582	0.501	0.455	0.132	0.157	1.000						
(7) Log Absolute Latitude	0.101	0.090	0.284	0.106	0.322	0.272	1.000					
(8) Log Land Suitability for Agriculture	0.364	0.302	0.248	-0.251	-0.133	0.649	-0.044	1.000				
(9) Log Distance to Frontier in 1500 CE	-0.360	-0.367	-0.401	-0.021	-0.396	-0.188	-0.318	-0.012	1.000			
(10) Log Distance to Frontier in 1000 CE	-0.243	-0.328	-0.389	-0.084	-0.522	-0.101	-0.307	0.175	0.606	1.000		
(11) Log Distance to Frontier in 1 CE	-0.326	-0.390	-0.503	-0.082	-0.521	-0.177	-0.341	-0.002	0.457	0.703	1.000	
(12) Mean Elevation	-0.028	-0.046	-0.047	0.106	0.069	-0.051	-0.026	0.018	0.018	0.033	0.028	1.000
(13) Terrain Roughness	0.197	0.199	0.218	-0.161	0.215	0.126	0.068	0.287	-0.060	-0.110	-0.229	0.626
(14) Mean Distance to Nearest Waterway	-0.305	-0.331	-0.366	0.195	-0.017	-0.178	-0.014	-0.230	0.155	0.118	0.170	0.429
(15) % of Land within 100 km of Waterway	0.383	0.362	0.391	-0.192	0.110	0.294	0.244	0.290	-0.210	-0.084	-0.219	-0.526
(16) Climate	0.527	0.567	0.633	0.080	0.621	0.438	0.563	0.101	-0.414	-0.456	-0.460	-0.178
(17) Orientation of Continental Axis	0.479	0.500	0.573	0.159	0.644	0.302	0.453	0.066	-0.202	-0.241	-0.336	-0.019
(18) Size of Continent	0.308	0.339	0.408	0.465	0.454	0.244	0.327	-0.159	-0.125	-0.218	-0.267	0.148
(19) Domesticable Plants	0.510	0.538	0.663	0.346	0.636	0.371	0.642	-0.112	-0.439	-0.507	-0.502	-0.197
(20) Domesticable Animals	0.580	0.615	0.699	0.249	0.768	0.366	0.640	-0.055	-0.427	-0.467	-0.461	-0.124
(21) Migratory Distance from Addis Ababa	-0.391	-0.312	-0.341	-1.000	-0.275	-0.132	-0.106	0.251	0.021	0.084	0.082	-0.106
(22) Aerial Distance from Addis Ababa	-0.287	-0.238	-0.283	-0.934	-0.331	-0.044	-0.017	0.334	-0.053	0.079	0.073	-0.153
(23) Migratory Distance from London	-0.537	-0.473	-0.518	-0.899	-0.497	-0.271	-0.385	0.176	0.215	0.224	0.256	0.001
(24) Migratory Distance from Tokyo	-0.420	-0.403	-0.353	-0.266	-0.559	-0.187	-0.316	0.056	0.164	0.227	0.224	-0.122
(25) Migratory Distance from Mexico City	0.198	0.128	0.167	0.822	-0.034	0.009	-0.006	-0.210	0.169	0.189	0.180	0.002

TABLE E.4: Pairwise Correlations for the 145-Country Historical Sample (Continued)

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(13) Terrain Roughness	1.000											
(14) Mean Distance to Nearest Waterway	-0.002	1.000										
(15) % of Land within 100 km of Waterway	0.039	-0.665	1.000									
(16) Climate	0.147	-0.497	0.431	1.000								
(17) Orientation of Continental Axis	0.260	-0.236	0.285	0.482	1.000							
(18) Size of Continent	-0.079	0.104	-0.170	0.327	0.668	1.000						
(19) Domesticable Plants	0.074	-0.333	0.334	0.817	0.613	0.487	1.000					
(20) Domesticable Animals	0.132	-0.322	0.349	0.778	0.743	0.516	0.878	1.000				
(21) Migratory Distance from Addis Ababa	0.161	-0.195	0.192	-0.080	-0.159	-0.465	-0.346	-0.249	1.000			
(22) Aerial Distance from Addis Ababa	0.170	-0.220	0.292	-0.074	-0.058	-0.429	-0.281	-0.191	0.934	1.000		
(23) Migratory Distance from London	0.088	-0.081	-0.042	-0.407	-0.455	-0.569	-0.683	-0.612	0.899	0.800	1.000	
(24) Migratory Distance from Tokyo	-0.264	-0.128	-0.118	-0.220	-0.672	-0.310	-0.309	-0.626	0.266	0.172	0.418	1.000
(25) Migratory Distance from Mexico City	-0.284	0.094	-0.194	-0.014	-0.035	0.268	0.148	0.075	-0.822	-0.759	-0.675	-0.025

TABLE E.5: Summary Statistics for the 143-Country Contemporary Sample

	Obs.	Mean	S.D.	Min.	Max.
Log Income Per Capita in 2000 CE	143	8.416	1.165	6.158	10.445
Log Population Density in 1500 CE	143	0.891	1.496	-3.817	3.842
Predicted Genetic Diversity (Unadjusted)	143	0.712	0.052	0.572	0.774
Predicted Genetic Diversity (Ancestry Adjusted)	143	0.727	0.027	0.631	0.774
Log Neolithic Transition Timing (Unadjusted)	143	8.343	0.599	5.991	9.259
Log Neolithic Transition Timing (Ancestry Adjusted)	143	8.498	0.455	7.213	9.250
Log % of Arable Land	143	2.251	1.180	-2.120	4.129
Log Absolute Latitude	143	3.014	0.921	0.000	4.159
Log Land Suitability for Agriculture	143	-1.416	1.321	-5.857	-0.041

TABLE E.6: Pairwise Correlations for the 143-Country Contemporary Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Log Income Per Capita in 2000 CE	1.000							
(2) Log Population Density in 1500 CE	-0.011	1.000						
(3) Predicted Genetic Diversity (Unadjusted)	-0.212	0.378	1.000					
(4) Predicted Genetic Diversity (Ancestry Adjusted)	-0.157	0.058	0.759	1.000				
(5) Log Neolithic Transition Timing (Unadjusted)	0.191	0.512	0.276	0.006	1.000			
(6) Log Neolithic Transition Timing (Ancestry Adjusted)	0.440	0.241	-0.165	-0.161	0.756	1.000		
(7) Log % of Arable Land	-0.008	0.571	0.097	0.078	0.156	0.153	1.000	
(8) Log Absolute Latitude	0.502	0.083	0.082	0.035	0.322	0.411	0.248	1.000
(9) Log Land Suitability for Agriculture	-0.105	0.369	-0.251	-0.240	-0.133	-0.078	0.670	-0.041

TABLE E.7: Summary Statistics for the 109-Country Contemporary Sample

	Obs.	Mean	S.D.	Min.	Max.
Log Income Per Capita in 2000 CE	109	8.455	1.189	6.524	10.445
Predicted Genetic Diversity (Ancestry Adjusted)	109	0.727	0.029	0.631	0.774
Log Neolithic Transition Timing (Ancestry Adjusted)	109	8.422	0.480	7.213	9.250
Log % of Arable Land	109	2.248	1.145	-2.120	4.129
Log Absolute Latitude	109	2.841	0.960	0.000	4.159
Social Infrastructure	109	0.453	0.243	0.113	1.000
Democracy	109	4.132	3.701	0.000	10.000
Ethnic Fractionalization	109	0.460	0.271	0.002	0.930
% of Population at Risk of Contracting Malaria	109	0.369	0.438	0.000	1.000
% of Population Living in Tropical Zones	109	0.361	0.419	0.000	1.000
Mean Distance to Nearest Waterway	109	0.298	0.323	0.008	1.467
Average Years of Schooling	94	4.527	2.776	0.409	10.862
Migratory Distance from Addis Ababa (Unadjusted)	109	9.081	7.715	0.000	26.771
Migratory Distance from Addis Ababa (Ancestry Adjusted)	109	6.322	3.897	0.000	18.989
Aerial Distance from Addis Ababa (Unadjusted)	109	6.332	3.861	0.000	14.420
Aerial Distance from Addis Ababa (Ancestry Adjusted)	109	5.161	2.425	0.000	12.180

TABLE E.8: Pairwise Correlations for the 109-Country Contemporary Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Log Income Per Capita in 2000 CE	1.000									
(2) Predicted Genetic Diversity (Ancestry Adjusted)	-0.261	1.000								
(3) Log Neolithic Transition Timing (Ancestry Adjusted)	0.535	-0.241	1.000							
(4) Log % of Arable Land	0.050	0.087	0.214	1.000						
(5) Log Absolute Latitude	0.563	-0.014	0.367	0.242	1.000					
(6) Social Infrastructure	0.808	-0.219	0.363	0.119	0.476	1.000				
(7) Democracy	0.717	-0.232	0.321	0.122	0.406	0.724	1.000			
(8) Ethnic Fractionalization	-0.644	0.195	-0.406	-0.255	-0.590	-0.468	-0.433	1.000		
(9) % of Population at Risk of Contracting Malaria	-0.784	0.406	-0.628	-0.204	-0.599	-0.593	-0.552	0.672	1.000	
(10) % of Population Living in Tropical Zones	-0.418	-0.167	-0.256	-0.001	-0.723	-0.380	-0.207	0.396	0.414	1.000
(11) Mean Distance to Nearest Waterway	-0.444	0.331	-0.280	-0.189	-0.238	-0.309	-0.349	0.417	0.443	-0.079
(12) Average Years of Schooling	0.866	-0.138	0.426	0.138	0.556	0.758	0.697	-0.478	-0.665	-0.436
(13) Migratory Distance from Addis Ababa (Unadjusted)	0.284	-0.734	0.253	-0.099	-0.015	0.099	0.284	-0.168	-0.424	0.267
(14) Migratory Distance from Addis Ababa (Ancestry Adjusted)	0.261	-1.000	0.241	-0.087	0.014	0.219	0.232	-0.195	-0.406	0.167
(15) Aerial Distance from Addis Ababa (Unadjusted)	0.335	-0.804	0.203	-0.047	0.064	0.195	0.362	-0.229	-0.443	0.239
(16) Aerial Distance from Addis Ababa (Ancestry Adjusted)	0.328	-0.943	0.173	-0.021	0.131	0.297	0.319	-0.261	-0.406	0.108

TABLE E.8: Pairwise Correlations for the 109-Country Contemporary Sample (Continued)

	(11)	(12)	(13)	(14)	(15)
(11) Mean Distance to Nearest Waterway	1.000				
(12) Average Years of Schooling	-0.360	1.000			
(13) Migratory Distance from Addis Ababa (Unadjusted)	-0.279	0.191	1.000		
(14) Migratory Distance from Addis Ababa (Ancestry Adjusted)	-0.331	0.138	0.734	1.000	
(15) Aerial Distance from Addis Ababa (Unadjusted)	-0.338	0.261	0.939	0.804	1.000
(16) Aerial Distance from Addis Ababa (Ancestry Adjusted)	-0.365	0.233	0.665	0.943	0.819

## References

- Acemoglu, Daron, Simon Johnson, and James A. Robinson**, “Institutions as a Fundamental Cause of Long-Run Growth,” in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth, Vol IA*, Amsterdam, The Netherlands: Elsevier North-Holland, 2005.
- Alesina, Alberto and Eliana La Ferrara**, “Ethnic Diversity and Economic Performance,” *Journal of Economic Literature*, September 2005, *43* (3), 762–800.
- , **Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg**, “Fractionalization,” *Journal of Economic Growth*, June 2003, *8* (2), 155–194.
- Ashraf, Quamrul and Oded Galor**, “Dynamics and Stagnation in the Malthusian Epoch,” 2010. Working Paper, Department of Economics, Williams College.
- , —, and **Ömer Özak**, “Isolation and Development,” *Journal of the European Economic Association*, April-May 2010, *8* (2-3), 401–412.
- Bairoch, Paul**, *Cities and Economic Development: From the Dawn of History to the Present*, Trans. by Christopher Braider. Chicago, IL: The University of Chicago Press, 1988.
- Barbujani, Guido, Arianna Magagni, Eric Minch, and L. Luca Cavalli-Sforza**, “An Apportionment of Human DNA Diversity,” *Proceedings of the National Academy of Sciences*, April 1997, *94* (9), 4516–4519.
- Barro, Robert J. and Jong-Wha Lee**, “International Data on Educational Attainment: Updates and Implications,” *Oxford Economic Papers*, July 2001, *53* (3), 541–563.
- and **Rachel M. McCleary**, “Religion and Economic Growth across Countries,” *American Sociological Review*, October 2003, *68* (5), 760–781.
- Cann, Howard M., Claudia de Toma, Lucien Cazes, Marie-Fernande Legrand, Valerie Morel, Laurence Piouffre et al.**, “A Human Genome Diversity Cell Line Panel,” *Science*, April 2002, *296* (5566), 261–262.
- Cavalli-Sforza, L. Luca**, “The Human Genome Diversity Project: Past, Present and Future,” *Nature Reviews Genetics*, April 2005, *6* (4), 333–340.
- and **Francesco Cavalli-Sforza**, *The Great Human Diasporas: The History of Diversity and Evolution*, Trans. by Sarah Thorne. New York, NY: Addison-Wesley, 1995.
- , **Paolo Menozzi, and Alberto Piazza**, *The History and Geography of Human Genes*, Princeton, NJ: Princeton University Press, 1994.
- Chandler, Tertius**, *Four Thousand Years of Urban Growth: An Historical Census*, Lewiston, NY: The Edwin Mellen Press, 1987.
- Conley, Timothy G.**, “GMM Estimation with Cross Sectional Dependence,” *Journal of Econometrics*, September 1999, *92* (1), 1–45.
- Dalgaard, Carl-Johan and Holger Strulik**, “The Physiological Foundations of the Wealth of Nations,” 2010. Discussion Paper No. 10-05, University of Copenhagen.

- Desmet, Klaus, Michel Le Breton, Ignacio Ortuño-Ortín, and Shlomo Weber**, “Nation Formation and Genetic Diversity,” 2006. CEPR Discussion Paper No. 5918.
- Diamond, Jared**, *Guns, Germs and Steel: The Fates of Human Societies*, New York, NY: W. W. Norton & Co., 1997.
- , “Evolution, Consequences and Future of Plant and Animal Domestication,” *Nature*, August 2002, *418* (6898), 700–707.
- Doepke, Matthias**, “Accounting for Fertility Decline During the Transition to Growth,” *Journal of Economic Growth*, September 2004, *9* (3), 347–383.
- Easterly, William and Ross Levine**, “Africa’s Growth Tragedy: Policies and Ethnic Divisions,” *Quarterly Journal of Economics*, November 1997, *112* (4), 1203–1250.
- Engerman, Stanley L. and Kenneth L. Sokoloff**, “History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World,” *Journal of Economic Perspectives*, Summer 2000, *14* (3), 217–232.
- Fujita, Masahisa, Paul Krugman, and Anthony J. Venables**, *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge, MA: The MIT Press, 1999.
- Gallup, John L. and Jeffrey D. Sachs**, “The Economic Burden of Malaria,” *The American Journal of Tropical Medicine and Hygiene*, January 2001, *64* (1-2), 85–96.
- , —, and **Andrew D. Mellinger**, “Geography and Economic Development,” *International Regional Science Review*, August 1999, *22* (2), 179–232.
- Galor, Oded**, “From Stagnation to Growth: Unified Growth Theory,” in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth, Vol IA*, Amsterdam, The Netherlands: Elsevier North-Holland, 2005.
- and **Andrew Mountford**, “Trade and the Great Divergence: The Family Connection,” *American Economic Review*, May 2006, *96* (2), 299–303.
- and —, “Trading Population for Productivity: Theory and Evidence,” *Review of Economic Studies*, October 2008, *75* (4), 1143–1179.
- and **David N. Weil**, “Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond,” *American Economic Review*, September 2000, *90* (4), 806–828.
- and **Omer Moav**, “Natural Selection and the Origin of Economic Growth,” *Quarterly Journal of Economics*, November 2002, *117* (4), 1133–1191.
- and —, “The Neolithic Revolution and Contemporary Variations in Life Expectancy,” 2007. Working Paper, Department of Economics, Brown University.
- , —, and **Dietrich Vollrath**, “Inequality in Land Ownership, the Emergence of Human Capital Promoting Institutions, and the Great Divergence,” *Review of Economic Studies*, January 2009, *76* (1), 143–179.

- Giuliano, Paola, Antonio Spilimbergo, and Giovanni Tonon**, “Genetic, Cultural and Geographical Distances,” 2006. IZA Discussion Paper No. 2229.
- Glaeser, Edward L. and Andrei Shleifer**, “Legal Origins,” *Quarterly Journal of Economics*, November 2002, *117* (4), 1193–1229.
- , **Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer**, “Do Institutions Cause Growth?,” *Journal of Economic Growth*, September 2004, *9* (3), 271–303.
- Greif, Avner**, “Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders’ Coalition,” *American Economic Review*, June 1993, *83* (3), 525–548.
- Griffiths, Anthony J. F., Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart**, *An Introduction to Genetic Analysis*, New York, NY: W. H. Freeman & Co., 2000.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, “Cultural Biases in Economic Exchange?,” *Quarterly Journal of Economics*, August 2009, *124* (3), 1095–1131.
- Hall, Robert E. and Charles I. Jones**, “Why Do Some Countries Produce So Much More Output Per Worker Than Others?,” *Quarterly Journal of Economics*, February 1999, *114* (1), 83–116.
- Hansen, Gary D. and Edward C. Prescott**, “Malthus to Solow,” *American Economic Review*, September 2002, *92* (4), 1205–1217.
- Jones, Eric L.**, *The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia*, Cambridge, UK: Cambridge University Press, 1981.
- Keller, Wolfgang**, “International Technology Diffusion,” *Journal of Economic Literature*, September 2004, *42* (3), 752–782.
- Krugman, Paul**, “Increasing Returns and Economic Geography,” *Journal of Political Economy*, June 1991, *99* (3), 483–499.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny**, “The Quality of Government,” *Journal of Law, Economics, and Organization*, March 1999, *15* (1), 222–279.
- Lagerlöf, Nils-Petter**, “From Malthus to Modern Growth: Can Epidemics Explain the Three Regimes?,” *International Economic Review*, May 2003, *44* (2), 755–777.
- , “The Galor-Weil Model Revisited: A Quantitative Exercise,” *Review of Economic Dynamics*, January 2006, *9* (1), 116–142.
- , “Long-Run Trends in Human Body Mass,” *Macroeconomic Dynamics*, June 2007, *11* (3), 367–387.
- Landes, David S.**, *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*, New York, NY: W. W. Norton & Co., 1998.

- Lewontin, Richard C.**, “The Apportionment of Human Diversity,” in Theodosius G. Dobzhansky, Max K. Hecht, and William C. Steere, eds., *Evolutionary Biology, Vol. 6*, New York, NY: Appleton-Century-Crofts, 1972.
- Lucas, Jr., Robert E.**, “The Industrial Revolution: Past and Future,” in Robert E. Lucas, Jr., ed., *Lectures on Economic Growth*, Cambridge, MA: Harvard University Press, 2002.
- Macauley, Vincent, Catherine Hill, Alessandro Achilli, Chiara Rengo, Douglas Clarke, William Meehan et al.**, “Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes,” *Science*, May 2005, *308* (5724), 1034–1036.
- McEvedy, Colin and Richard Jones**, *Atlas of World Population History*, New York, NY: Penguin Books Ltd., 1978.
- Michalopoulos, Stelios**, “The Origins of Ethnolinguistic Diversity: Theory and Evidence,” 2008. Working Paper, Department of Economics, Tufts University.
- Modelski, George**, *World Cities: -3000 to 2000*, Washington, DC: FAROS 2000, 2003.
- Mokyr, Joel**, *The Lever of Riches: Technological Creativity and Economic Progress*, New York, NY: Oxford University Press, 1990.
- Montalvo, Jose G. and Marta Reynal-Querol**, “Ethnic Diversity and Economic Development,” *Journal of Development Economics*, April 2005, *76* (2), 293–323.
- Murphy, Kevin M. and Robert H. Topel**, “Estimation and Inference in Two-Step Econometric Models,” *Journal of Business and Economic Statistics*, October 1985, *3* (4), 370–379.
- Nordhaus, William D.**, “Geography and Macroeconomics: New Data and New Findings,” *Proceedings of the National Academy of Sciences*, March 2006, *103* (10), 3510–3517.
- North, Douglass C. and Robert P. Thomas**, *The Rise of the Western World: A New Economic History*, Cambridge, UK: Cambridge University Press, 1973.
- Olson, Steve**, *Mapping Human History: Discovering the Past Through Our Genes*, New York, NY: Houghton Mifflin, 2002.
- Olsson, Ola and Douglas A. Hibbs Jr.**, “Biogeography and Long-Run Economic Development,” *European Economic Review*, May 2005, *49* (4), 909–938.
- Oppenheimer, Stephen**, *The Real Eve: Modern Man’s Journey Out of Africa*, New York, NY: Carroll & Graf, 2003.
- Pagan, Adrian**, “Econometric Issues in the Analysis of Regressions with Generated Regressors,” *International Economic Review*, February 1984, *25* (1), 221–247.
- Pomeranz, Kenneth**, *The Great Divergence: Europe, China and the Making of the Modern World Economy*, Princeton, NJ: Princeton University Press, 2000.
- Prugnolle, Franck, Andrea Manica, and François Balloux**, “Geography Predicts Neutral Genetic Diversity of Human Populations,” *Current Biology*, March 2005, *15* (5), R159–R160.



- Putterman, Louis**, “Agriculture, Diffusion, and Development: Ripple Effects of the Neolithic Revolution,” *Economica*, November 2008, 75 (300), 729–748.
- and **David N. Weil**, “Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality,” 2009. Working Paper, Department of Economics, Brown University.
- Ramachandran, Sohini, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza**, “Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa,” *Proceedings of the National Academy of Sciences*, November 2005, 102 (44), 15942–15947.
- Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney**, “The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change,” *Global Ecology and Biogeography*, September 2002, 11 (5), 377–392.
- Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi**, “Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development,” *Journal of Economic Growth*, June 2004, 9 (2), 131–165.
- Sachs, Jeffrey D. and Andrew Warner**, “Economic Reform and the Process of Global Integration,” *Brookings Papers on Economic Activity*, Spring 1995, 26 (1), 1–118.
- Spolaore, Enrico and Romain Wacziarg**, “The Diffusion of Development,” *Quarterly Journal of Economics*, May 2009, 124 (2), 469–529.
- Stringer, Chris B. and Peter Andrews**, “Genetic and Fossil Evidence for the Origin of Modern Humans,” *Science*, March 1988, 239 (4845), 1263–1268.
- Tabellini, Guido**, “Institutions and Culture,” *Journal of the European Economic Association*, April-May 2008, 6 (2-3), 255–294.
- Wang, Sijia, Cecil M. Lewis Jr., Mattias Jakobsson, Sohini Ramachandran, Nicolas Ray, Gabriel Bedoya et al.**, “Genetic Variation and Population Structure in Native Americans,” *PLoS Genetics*, November 2007, 3 (11), 2049–2067.
- Weber, Max**, *The Protestant Ethic and the Spirit of Capitalism*, Trans. by Talcott Parsons and Anthony Giddens, 1930. London, UK: Allen & Unwin, 1905.
- , *The Religion of China: Confucianism and Taoism*, Trans. and ed. by Hans H. Gerth, 1951. Glencoe, IL: Free Press, 1922.
- Weir, Bruce S.**, *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*, Sunderland, MA: Sinauer Associates, 1996.
- Weisdorf, Jacob L.**, “From Foraging to Farming: Explaining the Neolithic Revolution,” *Journal of Economic Surveys*, September 2005, 19 (4), 561–586.
- Wells, Spencer**, *The Journey of Man: A Genetic Odyssey*, Princeton, NJ: Princeton University Press, 2002.