

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

Quamrul Ashraf
Williams College

Oded Galor
Brown University

April, 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 direct long-lasting effect on the pattern of comparative economic development that could not be captured by contemporary 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 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 (i.e., genetic diversity and genetic distance among and between its ancestral populations) has a similar non-monotonic effect on the contemporary levels of income per capita. 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. 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: “Out of Africa” hypothesis, Human genetic diversity, Comparative development, Population density, Neolithic Revolution, Land productivity, Malthusian stagnation

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

*We are grateful to Kenneth Arrow, Dror Brenner, 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/workshop participants at Bar-Ilan, Ben-Gurion, Brown, Boston College, Copenhagen, Doshisha, Haifa, the Hebrew U, Hitotsubashi, the International Monetary Fund, Keio, Kyoto, Osaka, Tel Aviv, Tokyo, Tufts, UCLA Anderson, Williams, the World Bank, as well as conference participants in the CEPR “UGT Summer Workshop”, the NBER meeting on “Macroeconomics Across Time and Space”, the KEA “International Employment Forum”, the SED 2008 Annual Meeting, and the NBER 2008 Summer Institute. We also thank attendees of the Klein Lecture, Osaka, 2008, Kuznets Lecture, Yale, 2009, 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 contemporaneous effects of such factors in giving rise to and sustaining the Great Divergence in income per capita since the Industrial Revolution, attention has recently been drawn towards some “deep-rooted” factors that have been argued to affect the course of comparative economic development from the dawn of human civilization to the modern era.

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 direct long-lasting hump-shaped effect on the pattern of comparative economic development that could not be captured by contemporary geographical, institutional, and cultural factors. 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.

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 era, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. Moreover, the level of genetic diversity within a country today (i.e., genetic diversity and genetic distance among and between its ancestral populations) has a similar non-monotonic effect on contemporary levels of income per capita. 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. Further, the empirical findings suggest that, indeed, the optimal level of diversity has increased in the course of industrialization.

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. Indeed, as depicted in Figure 1, migratory distance from East Africa has an adverse linear 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 pertains to the detrimental impact of diversity on the efficiency of the aggregate production process. Heterogeneity increases the likelihood of mis-coordination and distrust, reducing cooperation and

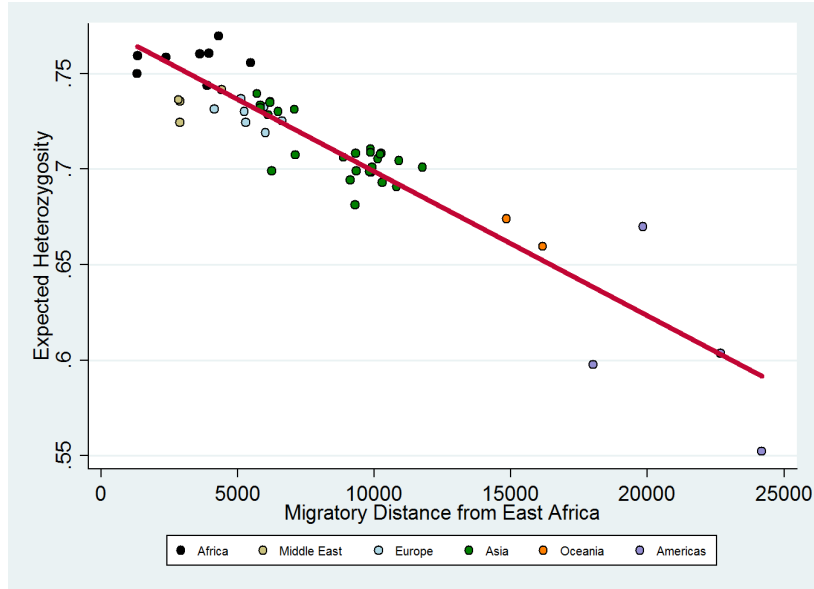


FIGURE 1: Expected Heterozygosity and Migratory Distance in the HGDP-CEPH Sample

disrupting the socioeconomic order. Greater population diversity is associated therefore 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 concerns the positive role of diversity in the expansion of society’s production possibility frontier. A wider spectrum of traits enhances the accumulation of universally-applicable human capital and 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 society’s production possibility frontier and conferring the benefits of increased total factor productivity.¹

Higher diversity in a society’s population can therefore have conflicting effects on the level of its total factor productivity. 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

¹Indeed, this observation is broadly consistent with theoretical and empirical evidence on the creativity-promoting effects of diversity in the workforce. Hong and Page (2001) provide a theoretical formalization of this idea, proving that a group of “cognitively diverse” problem solvers can find optimal solutions to difficult problems, and that a more diverse group of people with limited abilities can outperform a homogeneous group of high ability problem solvers. In the context of team theory, Prat (2002) has established that heterogeneity raises team productivity when jobs within teams are complementary to one another. Similarly, Lazear (1999a, 1999b) has also discussed how diversity in a productive unit can raise overall productivity once the costs associated with language barriers are taken into account. The benefits of heterogeneity are further highlighted by Ottaviano and Peri (2005, 2006) who have empirically demonstrated the productivity enhancing effects of cultural diversity in American cities. While not explicitly related to a genetic hypothesis, these studies offer supporting evidence on the beneficial effects of diversity, insofar as such diversity is broadly manifested in cultural or cognitive terms.

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 have 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, 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.² 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 yields the overall expected heterozygosity for the relevant population.³

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, that according to anthropologists, have been isolated from genetic flows from other ethnic groups.⁴ Empirical evidence for these 53 ethnic groups suggest that, indeed, migratory distance from East Africa has an adverse linear effect on genetic diversity (Ramachandran et al., 2005), as depicted in Figure 1.⁵ 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.

Nevertheless, while genetic diversity data pertains only to ethnic groups, data on development outcomes are typically available at the country level where most national populations today are composed of multiple ethnicities, some of which may not be indigenous to the current geographical location of these ethnic groups. 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 diversity due to differences *between* ethnic groups as well. Second, one needs to account for the potential induce-

²In 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).

³See Weir (1996) for the statistical theory underlying measures of genetic diversity. The expected heterozygosity index is discussed in greater technical detail in Section 3.1.1.

⁴Ramachandran et al. (2005) compute expected heterozygosity (i.e., genetic diversity) for these 53 ethnic groups from allelic frequencies associated with 783 chromosomal loci.

⁵For 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).

ment for members of distinct ethnic groups to relocate to potentially more lucrative geographical locations.

To tackle these difficulties, this study adopts two distinct strategies. The first strategy 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, while the presence of multiple indigenous ethnicities had a negligible effect on the genetic diversity of the regional population. The second, more complex strategy involves the construction of an index of genetic diversity for contemporary national populations, accounting for the expected heterozygosity within each sub-national group, and the additional component of diversity at the country level that arises from the genetic distances between its pre-colonial ancestral populations. Development outcomes under this second strategy would have to additionally account for the potential inducement for members of distinct ethnic groups to relocate to potentially 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.⁶ 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.⁷

Using data on genetic diversity observed at the ethnic group level, the initial 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

⁶To further elaborate, the Malthusian theory, proposed initially by Malthus (1798) and formalized more recently by Kremer (1993) and Lucas (2002), suggests that the global stagnation of income per capita in the pre-Industrial era of development reflected the counterbalancing effect of population growth on the expansion of resources, in an environment characterized by diminishing returns to labor. In particular, resource surpluses led to population growth as a natural result of the “passion between the sexes.” However, in the event of population expansions beyond resource capacity, population reduction occurred via the “preventative check” (i.e., the intentional reduction of fertility) as well as the “positive check” (i.e., the natural forces of disease, famine and warfare). Accordingly, periods of economic stagnation were characterized by stable population size and income per capita, while episodes of prosperity gave rise to only temporary income gains, triggering an increase in population size that eventually retracted income per capita back to its long-run (steady-state) equilibrium level. The theory therefore proposes that variation in population density across regions during the agricultural stage of development primarily reflected cross-regional variation in technology and land productivity.

⁷Admittedly, 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.

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.⁸

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 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.⁹ 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.¹⁰ 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.¹¹ Indeed, several studies

⁸The 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.

⁹An 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).

¹⁰For 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).

¹¹In 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

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.¹²

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 is 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 the concerns regarding potential endogeneity bias. In so doing, the paper highlights one of the deepest channels in comparative development, pertaining not to factors associated with the dawn of complex agricultural societies as in Jared Diamond’s (1997) influential hypothesis, but to conditions innately related to the very dawn of mankind itself.

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 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 declining expected heterozygosity with increasing distance along migratory paths from East Africa, but also predicts 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 imply that, in line with the prediction that the beneficial impact of diversity has increased in the process of development, 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. Indeed, the linear

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.

¹²The evidence uncovered by some of these studies is further elaborated upon in Section 3.1.1.

and quadratic regression coefficients associated with these economically significant effects of contemporary diversity are both statistically significant at the 1% level. Moreover, the hump-shaped impact of contemporary 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, and 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 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 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 optimum level of 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 11.7:1 to 2.5:1, and (v) decreasing the diversity of Ethiopia to the optimum level of the U.S. would increase Ethiopia's per capita income by a factor of 1.73, closing the income gap between the U.S. and Ethiopia from 47.4:1 to 27.4:1.

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

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 influence of geography has been stressed from a historical perspective by Jones (1981), Diamond (1997) and Pomeranz (2000), and is highlighted empirically by Gallup et al. (1999), Gylfason (2001), and Ols-son and Hibbs (2005). Institutions, on the other hand, are given historical precedence by North and Thomas (1973), North (1981), Greif (1993), and Mokyr (1990), and are emphasized empirically by Hall and Jones (1999), La Porta et al. (1999), Acemoglu et al. (2002), Glaeser and Shleifer (2002), Easterly and Levine (2003) and Rodrik et al. (2004). In related strands of the literature on institutions, Engerman and Sokoloff (2000), Acemoglu et al. (2001) and Bertocchi and Canova

(2002) 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. 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), Guiso et al. (2003, 2006) as well as Tabellini (2007). Finally, the importance of human capital formation has been underlined in the unified growth theories of Galor and Weil (2000), Fernández-Villaverde (2001), Galor and Moav (2002), Lucas (2002), Lagerlöf (2003, 2006), Doepke (2004), Galor and Mountford (2006, 2008), Galor (2005) and in the recent study of Galor et al. (2009), 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. In particular, 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 contemporary variations in income per capita across the globe. It highlights the deepest channels of geographical determinism in comparative development, pertaining not to factors associated with the dawn of complex agricultural societies as in Jared Diamond’s (1997) influential hypothesis that emphasizes the timing of the Neolithic Revolution, but to conditions innately related to the very dawn of mankind itself.

Second, the paper is 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.

The study however employs data and empirical results from research in population genetics, placing it 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.¹³ Spolaore and Wacziarg (2009) argue that genetic distance observed between populations captures their divergence in biological and cultural characteristics (transmitted vertically across the generations of 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.¹⁴ 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%.¹⁵

¹³See 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.

¹⁴The 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.

¹⁵It should be noted that Giuliano et al. (2006) have recently objected to 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

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, as already discussed, 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.

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, ultimately leading 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, enabling 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

geographic distance, is really a proxy for transportation costs associated with geographical (as opposed to biological or sociocultural) barriers. Nevertheless, both Spolaore and Wacziarg (2009) as well as Guiso et al. (2009) demonstrate that their results remain robust to controls for this alternative transportation cost hypothesis.

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) unicausal hypothesis, the composition of human populations with respect to their overall 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.¹⁶ Interestingly, however, unlike the conjecture of Diamond (1997), the timing of the Neolithic Revolution has no effect on contemporary income per capita.¹⁷

3 The Historical Analysis

3.1 Data and Empirical Strategy

This section discusses in detail the data and empirical strategy employed to examine the historical impact of the proposed genetic diversity channel while controlling for the impact of the Neolithic Revolution. Section 3.1.1, in particular, elaborates on the expected heterozygosity index of genetic diversity used by population geneticists and also discusses the empirical evidence, along with the underlying genetic data, regarding the negative relationship between human genetic diversity and prehistorical migratory distance from East Africa. Details pertaining to the data and methodology used to control for the proximate (i.e., agricultural transition timing) and ultimate (i.e., geographic and biogeographic) factors in Diamond’s (1997) hypothesis are explored in Section 3.1.2. Section 3.1.3 discusses controls for additional geographic factors gauging the natural productivity of land for agriculture, which, given the focus on Malthusian era comparative development, are necessary in examining the hypothesized effects of genetic diversity on total factor productivity in pre-colonial societies. Information regarding the outcome variables of interest, specifically population density in 1 CE, 1000 CE and 1500 CE, is provided in Section 3.1.4. Finally, Section 3.1.5 summarizes the overall empirical strategy adopted, and discusses some descriptive statistics of the regression samples employed by the analysis.

¹⁶Note 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.

¹⁷Olsson 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 15, these results are in fact non-robust. Even the inclusion of continental fixed effects make the (direct or indirect) effect of the Neolithic Revolution on contemporary outcomes insignificantly different than zero.

3.1.1 Expected Heterozygosity and Migratory Distance from East Africa

The expected heterozygosity index, as mentioned previously, is a measure of genetic diversity across individuals within a given population (such as an ethnic or regional group), reflecting the probability that two randomly-selected individuals from the population are different with respect to their genetic makeup. In particular, individuals in any population can differ in terms of the allele that they possess for a given gene (or locus in a chromosome), where an allele is any one of a number of viable DNA codings for the gene in question, with overall genetic differentiation arising from such variation across numerous genes or chromosomal loci. The expected heterozygosity of a population is therefore calculated using sample data on allelic frequencies (i.e., the probability that an individual, selected at random from the relevant population, will carry a particular gene variant or allele). Consider the case of 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 is:

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

which, given allelic frequencies for each of m different genes or loci, can be averaged across all the loci to yield an aggregate expected heterozygosity measure of overall genetic diversity as:

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

where the locus indexed by l is assumed to have k_l observed variants.

Using a worldwide sample comprising 1027 individuals spanning 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), Ramachandran et al. (2005) compute the expected heterozygosity of each group in the data set from allelic frequencies associated with 783 chromosomal loci.¹⁸ They then regress expected heterozygosity on distance along migration routes of the prehistorical human expansion out of East Africa to establish a highly statistically significant inverse linear relationship between these variables. The authors 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 belief that, until recently, humans did not generally cross large bodies of water while migrating. The intermediate waypoints, depicted on the world map in Figure 2 along with the spatial distribution of the ethnic

¹⁸For 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.

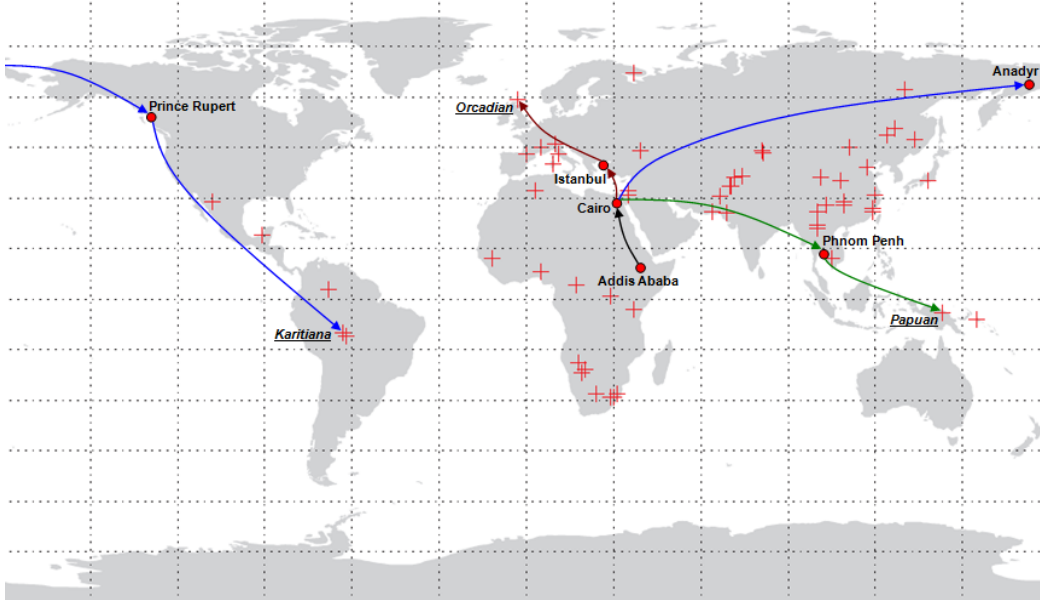


FIGURE 2: The 53 HGDP-CEPH Ethnic Groups and Migratory Paths from East Africa

groups from the HGDP-CEPH sample, are as follows: 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.¹⁹ 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, plus the distance(s) between waypoints if two or more such points are required.

As mentioned earlier, the regression analysis conducted by Ramachandran et al. (2005) at the ethnic group level establishes migratory distance from East Africa as a strong negative predictor of genetic diversity. Specifically, based on the R -squared of their regression, migratory distance alone explains almost 86% of the cross-group variation in within-group diversity.²⁰ In

¹⁹Based 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.

²⁰These 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 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

addition, the estimated OLS coefficient is highly 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 of *observed* genetic diversity and Malthusian era economic development as discussed earlier, the use of genetic diversity values *predicted* by migratory distance from East Africa alleviates concerns regarding endogeneity bias. This is possible since there is no clear reason to believe that distance determined by prehistorical human migration routes from Africa can have a 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 1500 CE) is available for a much larger set of countries or, to be precise, regions defined by modern national borders.

To further elaborate, the current research tests the proposed genetic hypothesis both using actual 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 groups located within a given country.²¹ 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.²² 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 by the current analysis.²³

obscured the genetic legacy of a serial-founder effect in these groups.

²¹A population-weighted averaging method is infeasible in this case due to the current unavailability of population figures for the HGDP-CEPH ethnic groups.

²²Clearly, 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).

²³As 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

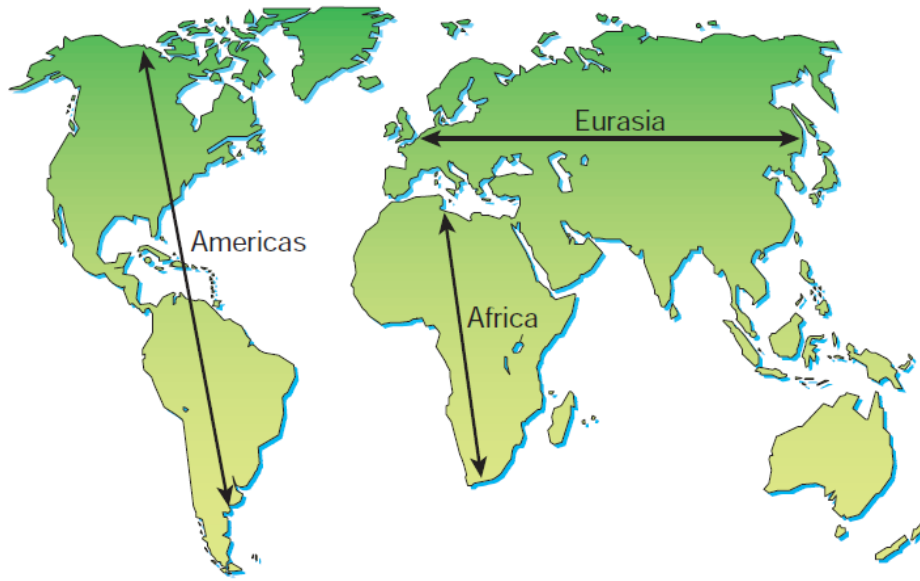


FIGURE 3: The Orientation of the Major Axes of the Continents

3.1.2 The Timing of the Transition to Agriculture

Diamond's (1997) hypothesis, as discussed earlier, identifies the timing of the Neolithic transition to agriculture as a proximate determinant of institutional and economic development, thereby designating initial geographic and biogeographic conditions, governing 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.

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. Briefly, 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, compared across continents on the world map in Figure 3, is argued by Diamond (1997) to have played a pivotal role

sample containing predicted genetic diversity at the country level.

Briefly, 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 1000 times. The standard deviations in the sample of 1000 observations of coefficient estimates from the second stage regression are thus the bootstrap standard errors of the point estimates of these coefficients.

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 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).

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 from Olsson and Hibbs (2005) as well as on the timing of agricultural transitions from Putterman (2008). However, given the well-established empirical link between the ultimate and proximate factors in Diamond’s hypothesis, the baseline specification focuses only on the timing of the transition to agriculture as the relevant control variable for this channel. The results of an extended specification that incorporates initial geographic and biogeographic factors as controls are presented to demonstrate robustness.

3.1.3 The Productivity of Land for Agriculture

The focus of the current research on economic development in the pre-colonial Malthusian era necessitates controls for the natural productivity of land with respect to 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. If the diversity in a society’s population 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 empirical test of the effect of diversity on the Malthusian development outcome - i.e., population density.

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) and illustrated in Figure 4, 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.

In controlling for the agricultural productivity of land, this study employs measurements of three geographic variables at the country level including the arable percentage of land, absolute latitude, and an index gauging the overall suitability of land for agriculture based on soil quality and temperature. The data for these variables are obtained from the World Bank’s World Development Indicators, the CIA’s World Factbook and Michalopoulos (2007) respectively.²⁴

²⁴Specifically, Michalopoulos (2007) provides a country-specific average value of the suitability index using more refined data compiled at a 0.5° latitude x 0.5° longitude grid level by Ramankutty et al. (2002).

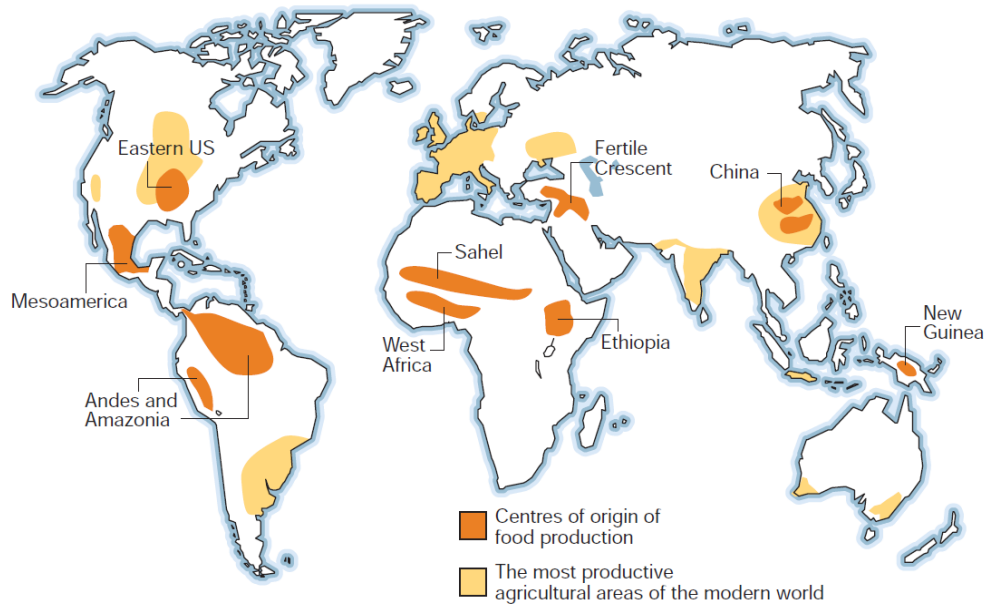


FIGURE 4: The Independence of Land Productivity and Agricultural Origins

3.1.4 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 both in actual 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 is 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.²⁵ 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, for which data is available for 184 countries, as the preferred outcome variable to examine. Nonetheless, to demonstrate the robustness of the genetic channel for earlier time periods, the empirical analysis additionally examines population density in 1000 CE and 1 CE despite their somewhat smaller sample sizes of 177 and 155 observations respectively.

3.1.5 Regression Specifications and Descriptive Statistics

In light of the proposed genetic diversity hypothesis as well as the Neolithic transition timing and land productivity channels in agricultural development, the following specification is adopted to

²⁵In particular, the cross-sectional unit of observation in McEvedy and Jones (1978) is a region delineated by its international borders in 1975. Historical population estimates are provided for regions corresponding to either individual countries in 1975 or, in some instances, 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 USSR) in 1975.

examine the influence of actual genetic diversity in the limited sample of 21 countries:

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

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 elapsed since country i 's transition to agriculture, X_i is a vector of land productivity controls and ε_{it} is a country-year specific disturbance term.²⁶

On the other hand, 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:

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

where \hat{G}_i is the genetic diversity predicted by migratory distance from East Africa for country i using the methodology discussed in Section 3.1.1. Indeed, it is this regression specification that is estimated to obtain the main empirical findings.

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 influence on society's level of development through two opposing marginal effects on 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, as will become evident shortly, the results from estimating the regression models in (3) and (4) 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 $\beta_{2t} < 0$.

Tables 1(a)-1(b) present some descriptive statistics of the limited 21-country sample employed in estimating equation (3). Similarly, the descriptive statistics of the extended 145-country sample used to estimate the regression model in (4) appear in Tables 2(a)-2(b). Both samples reflect countries for which data is available on all the explanatory variables in the regression specifications along with population density figures in 1500 CE. A number of inferences may be drawn from these statistics that are worth further elaboration.

Comparing the figures in Table 1(a) with those in Table 2(a), the finite-sample moments of the explanatory variables in the limited and extended cross-country samples appear to be remarkably similar indeed. For instance, while actual genetic diversity in the limited sample possesses a mean and a standard deviation of 0.713 and 0.056 respectively, the corresponding moments for predicted diversity in the extended sample of countries are 0.711 and 0.053. Moreover, the range of values for predicted genetic diversity in the extended sample falls within the observed range of values for actual diversity in the limited sample. This is particularly reassuring because it demonstrates

²⁶The 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).

TABLE 1: Descriptive Statistics of the Limited Cross-Country Sample

(a) Means and Standard Deviations

	Obs.	Mean	Std. Dev.	Min.	Max.
Actual Genetic Diversity	21	0.713	0.056	0.552	0.770
Migratory Distance	21	8.238	6.735	1.335	24.177
Log Agricultural Transition Timing	21	8.342	0.539	7.131	9.259
Log Arable Percentage of Land	21	2.141	1.168	-0.799	3.512
Log Absolute Latitude	21	2.739	1.178	0.000	4.094
Log Suitability for Agriculture	21	-1.391	0.895	-3.219	-0.288
Log Population Density in 1500 CE	21	1.169	1.756	-2.135	3.842

(b) Pairwise Correlations

	1	2	3	4	5	6	7
1 Actual Genetic Diversity	1.000						
2 Migratory Distance	-0.968	1.000					
3 Log Agricultural Transition Timing	-0.117	0.024	1.000				
4 Log Arable Percentage of Land	0.173	-0.183	0.521	1.000			
5 Log Absolute Latitude	0.055	-0.012	0.392	0.453	1.000		
6 Log Suitability for Agriculture	-0.218	0.282	0.299	0.376	0.049	1.000	
7 Log Population Density in 1500 CE	0.244	-0.226	0.735	0.670	0.336	0.561	1.000

that the methodology used to generate the predicted genetic diversity variable did not project values beyond what is observed for actual genetic diversity, indicating that the HGDP-CEPH collection of ethnic groups is indeed a representative sample for the worldwide variation in within-country genetic diversity. Finally, 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 regression samples, foreshadows the encouraging similarity of the regression results that are obtained under actual and predicted values of genetic diversity.

Turning to the correlation matrices in Tables 1(b) and 2(b), the overall positive correlations of 0.24 and 0.39 between log population density in 1500 CE and genetic diversity in the limited and extended regression samples, respectively, are consistent with diversity being predominantly beneficial for Malthusian economic development. Indeed, while the regression analysis in the next section reveals an additional statistically significant adverse effect of diversity on development, the overwhelmingly beneficial effect of diversity is confirmed therein as well. Further, in line with the predictions of the agricultural transition timing and land productivity channels, log population density in 1500 CE is positively correlated with the timing of the Neolithic Revolution as well as with the measures chosen to reflect the productivity of land for agriculture in both the limited and extended regression samples. With the exception of the correlation of 0.65 between the arable percentage of total land area and the agricultural suitability index in the extended sample, these geographic measures are in general only weakly correlated with each other, implying that they possibly capture different dimensions of overall land productivity. In addition, the rather weak correlations (ranging from -0.13 to 0.32 in the extended sample) between these variables and the timing of the agricultural transition is consistent with the orthogonality of land productivity with respect to the exogenous determinants in the Diamond hypothesis discussed earlier. Last but not least, the fact that genetic diversity in both samples is only weakly correlated with the controls for

TABLE 2: Descriptive Statistics of the Extended Cross-Country Sample

(a) Means and Standard Deviations

	Obs.	Mean	Std. Dev.	Min.	Max.
Predicted Genetic Diversity	145	0.711	0.053	0.572	0.774
Log Agricultural Transition Timing	145	8.343	0.595	5.991	9.259
Log Arable Percentage of Land	145	2.232	1.203	-2.120	4.129
Log Absolute Latitude	145	3.003	0.924	0.000	4.159
Log Suitability for Agriculture	145	-1.409	1.313	-5.857	-0.041
Log Population Density in 1500 CE	145	0.881	1.500	-3.817	3.842

(b) Pairwise Correlations

	1	2	3	4	5	6
1 Predicted Genetic Diversity	1.000					
2 Log Agricultural Transition Timing	0.275	1.000				
3 Log Arable Percentage of Land	0.132	0.158	1.000			
4 Log Absolute Latitude	0.106	0.322	0.272	1.000		
5 Log Suitability for Agriculture	-0.251	-0.133	0.649	-0.044	1.000	
6 Log Population Density in 1500 CE	0.391	0.511	0.582	0.101	0.364	1.000

the transition timing and land productivity channels is a reassuring indicator that the significant effects of diversity, uncovered by the empirical analysis to follow, are not simply reflecting the influence of these other explanatory channels.

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 actual 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 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 actual 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 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 actual 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 of 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 3 presents the limited sample results from regressions explaining log population density in 1500 CE.²⁷ 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 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 would lower population density by 2.98%.²⁸ 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 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

²⁷Corresponding to Tables 3 and 4 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.

²⁸The 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 ΔG change in diversity at the specified level \bar{G} is given by: $\Delta P/P = \exp\{\Delta G(\hat{\beta}_1 + 2\hat{\beta}_2\bar{G} + \bar{G}\Delta G)\} - 1$.

TABLE 3: Actual Diversity and Economic Development in 1500 CE

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable is Log Population Density in 1500 CE				
Genetic Diversity	413.505*** (97.320)			225.441*** (73.781)	203.815* (97.637)
Genetic Diversity Sqr.	-302.647*** (73.344)			-161.159** (56.155)	-145.717* (80.414)
Log Transition Timing		2.396*** (0.272)		1.214*** (0.373)	1.135 (0.658)
Log Arable % of 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 Agri. 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

Notes: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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 arable percentage of 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 sample variation in log population density in 1500 CE.

Column 4 presents the results from estimating the regression model in (3), exploiting the combined explanatory power of all three identified channels, for log population density in the year 1500 CE. Not surprisingly, given the small sample size and the pairwise correlations between covariates reported in Table 1(b), the estimated conditional effects are sizeably reduced in magnitude in comparison to their unconditional estimates presented in earlier columns. Nonetheless, the OLS coefficients on 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 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%. 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 equation (3) but with continental dummies included as additional explanatory variables are reported in Column 6. The purpose of this regression is to ensure that the earlier results 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. In addition, the

TABLE 4: 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						
Genetic Diversity		255.220** (100.586)		196.041* (92.029)	233.758*** (86.883)	181.932** (71.934)
Genetic Diversity Sqr.		-209.808** (73.814)		-128.524 (79.082)	-167.564** (65.729)	-130.762** (59.269)
Mig. Distance	0.505*** (0.148)	0.070 (0.184)	0.293* (0.147)	0.097 (0.192)		
Mig. Distance Sqr.	-0.023*** (0.006)	-0.014 (0.009)	-0.015** (0.006)	0.001 (0.006)		
Log Transition Timing			1.532*** (0.352)	1.583* (0.742)	1.183*** (0.338)	1.166** (0.475)
Log Arable % of Land			0.415** (0.180)	0.471*** (0.149)	0.531*** (0.170)	0.545** (0.219)
Log Absolute Latitude			-0.308 (0.184)	-0.183 (0.184)	-0.169 (0.106)	-0.118 (0.128)
Log Agri. Suitability			0.585* (0.302)	0.524 (0.308)	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	21	21	21	21
R-squared	0.34	0.46	0.88	0.90	0.89	0.90
P-value for:						
Joint Sig. of Diversity and its Sqr.		0.023		0.007		
Joint Sig. of Distance and its Sqr.		0.235		0.769		
Overidentifying Restrictions					0.889	0.861
Exogeneity of Distance and its Sqr.					0.952	0.804

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

estimated within-continent effects of the diversity channel are very similar to the cross-continent effects reported in Column 5, 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 3 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 actual values of genetic diversity.

Establishing the Exogeneity of Migratory Distance As already mentioned, the fact that the limited sample comprises actual genetic diversity, which is strongly but not perfectly correlated with

migratory distance from East Africa, permits a formal empirical examination of whether migratory distance influences population density only via the serial-founder effect on genetic diversity. This is a particularly important test since, if migratory distance actually affects economic development either directly or through some other unobserved channels, then the main empirical analysis conducted using predicted values of diversity would be naively attributing this latent influence to the genetic diversity channel.

To implement the aforementioned test, the current analysis examines specifications that include migratory distance 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 alternative specifications including both migratory distance and genetic diversity as covariates. If migratory distance and genetic diversity are ultimate and proximate determinants within the same channel, then genetic diversity, when included in the regression, should capture most of the explanatory power otherwise attributed to migratory distance. The results reported in Columns 1-4 of Table 4 indicate that this is indeed the case. Thus, while Column 1 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 3, they continue to remain statistically significant at conventional levels in line with expectations.

Reassuringly, a pattern similar to that observed in the first two columns also emerges in Columns 3 and 4 where the analysis is augmented with controls for the Neolithic transition timing and land productivity channels. In addition, to overcome potentially confounding results given the presence of multicollinearity, the analysis conducts F -tests on the joint significance of the linear and quadratic coefficients associated with diversity and migratory distance when both determinants are included in the same specification. Indeed, the corresponding P -values indicate that it is diversity and not distance that exerts a significant non-monotonic effect on population density, verifying the assertion that migratory distance from East Africa has no independent influence on development besides its effect through the genetic diversity channel.

Having established that migratory distance affects population density only via the genetic diversity channel, 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 3, 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.

Finally, the 2SLS regressions in Columns 5 and 6 provide another opportunity to verify the exogeneity of migratory distance with respect to population density. Given that 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 indeed valid, but also verify the earlier finding that migratory distance does not impart any 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

The results from regressions explaining log population density in 1500 CE, using the extended sample of countries, are summarized in Table 5. To examine the independent and combined effects of the genetic diversity, transition timing and land productivity channels, a number of specifications spanning relevant subsets of the explanatory variables in equation (4) are estimated.

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%.²⁹ 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 of the regression, the cross-country variation in genetic diversity alone explains 22% of the cross-country variation in population density.

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 2(b), 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 4.³⁰ 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

²⁹Following 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 28 for details on how these effects may be computed based on the estimated linear and quadratic coefficients associated with genetic diversity.

³⁰The 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.

TABLE 5: Predicted Diversity and Economic Development in 1500 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable is Log Population Density in 1500 CE					
Pred. Diversity	250.986*** (68.257)		213.537*** (63.502)	203.017*** (61.049)	195.416*** (56.086)	199.727** (80.513)
Pred. Diversity Sqr.	-177.399*** (50.221)		-152.107*** (46.648)	-141.980*** (44.827)	-137.977*** (40.839)	-146.167*** (56.257)
Log Transition Timing		1.287*** (0.177)	1.047*** (0.191)		1.160*** (0.147)	1.235*** (0.243)
Log Arable % of Land				0.523*** (0.115)	0.401*** (0.094)	0.393*** (0.101)
Log Absolute Latitude				-0.167* (0.094)	-0.342*** (0.091)	-0.417*** (0.123)
Log Agri. Suitability				0.189 (0.122)	0.305*** (0.095)	0.257*** (0.098)
Optimal Diversity	0.707*** (0.116)		0.702*** (0.035)	0.715*** (0.108)	0.708*** (0.073)	0.683*** (0.252)
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

Notes: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

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 arable percentage of land and log absolute latitude variables, which indeed appear to possess their expected signs.³¹ 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 0.27 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 estimating equation (4) for log population density in 1500 CE and, thus, encompasses the explanatory power of all three identified channels. 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 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, which roughly corresponds

³¹To interpret the estimated coefficients associated with the land productivity channel, a 1% increase in the arable fraction of 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. (2002).

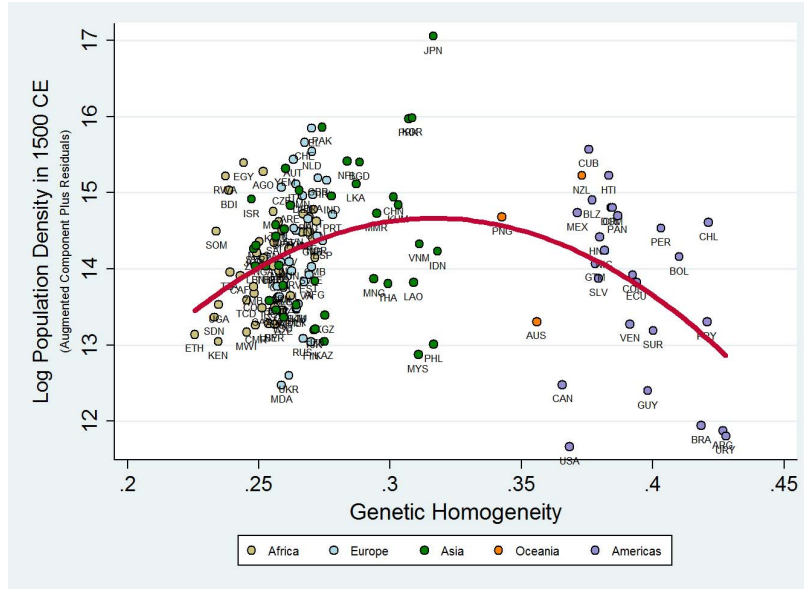


FIGURE 5: Predicted Diversity and Population Density in 1500 CE - Conditional on Transition Timing, Land Productivity and Continental Fixed Effects

to that predicted for the Indian subcontinent by migratory distance from East Africa. Indeed, a 1 percentage point change in diversity in either direction at the optimum would lower population density by 1.37%. To place the worldwide effect of the diversity channel into perspective, increasing the expected heterozygosity of the most homogenous *native* South American populations by 13.6 percentage points to the predicted optimum would have raised their population density by a factor of 12.84. On the other hand, decreasing the expected heterozygosity of the most heterogenous East African populations by 6.6 percentage points to the optimum would have raised their population density by a factor of 1.83. Overall, 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 specification augmented with continental dummies as additional explanatory variables, which allows the analysis to capture unobserved continent-specific attributes that could potentially have an influence on population density.³² Nonetheless, 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. The Neolithic transition timing and land productivity channels broadly retain their expected effects in magnitude and significance as well. Given the robustness of these results, it is not surprising that the estimated coefficients on the continental dummies do not reveal any significant continent-specific effects on population density in 1500 CE. Indeed, the inclusion of continental dummies does not appear to significantly increase the explanatory power of the analysis beyond that of the baseline regression specification examined in Column 5. The non-monotonic effect of genetic diversity on log population

³²The excluded continent in all extended sample empirical specifications in this study that incorporate continental dummy variables is Oceania.

density in 1500 CE, conditional on agricultural transition timing, land productivity and continental fixed effects, is depicted on the scatter plot in Figure 5.³³

To summarize the results reported in Table 5, 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 Robustness 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 suggests that the results of the baseline investigation are indeed more plausibly associated with the proposed genetic diversity channel as opposed to being generated some other unobserved factor that not only explained the world population distribution in 1500 CE but also happened to be correlated with migratory distance from East Africa. More broadly, the analysis demonstrates a strong persistence of the diversity channel across time and thereby implies, consistently with the theory, that the manner in which genetic diversity influences development did not fundamentally change as a result of technological progress over the period 1–1500 CE.³⁴ This finding is particularly reassuring given that the conceptual framework of the proposed genetic diversity channel abstracts from interactions that may exist, at least in theory, between the manner in which diversity effects economic outcomes and the overall level of technological advancement.³⁵ Finally, by examining historical periods when major population

³³For consistency with Figure 1, which depicts the negative effect of increasing migratory distance from East Africa on genetic diversity, the horizontal axes in Figures 5-7 represent genetic homogeneity (i.e., 1 minus genetic diversity) so as to reflect increasing as opposed to decreasing migratory distance from East Africa. It should also be noted that 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 are collected in Appendix C.

³⁴According to McEvedy and Jones (1978), the global population grew from 170 million in 1 CE to 425 million in 1500 CE, representing a two and a half fold increase over this period. This, of course, reflects a similar growth in global total factor productivity over the same 1500-year period given the Malthusian feedback between technology and population in the agricultural stage of development.

³⁵For instance, one could argue that the marginal detrimental effects of diversity on total factor productivity are exacerbated at higher levels of technological sophistication where the necessity for a well-functioning socioeconomic system with properly coordinated activities is even greater. If there is no significant change in the beneficial effects of diversity, the argument would then imply that at sufficiently high levels of global technological advancement the detrimental effects of diversity would outweigh its beneficial effects at all observed levels of diversity, thereby yielding a monotonically decreasing relationship between diversity and population density. While such an argument may be valid in principle, it does not appear to hold at least for the growth in global technology that occurred during the

TABLE 6: Predicted Diversity and Economic Development in 1000 CE

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable is Log Population Density in 1000 CE						
Pred. Diversity	219.722*** (66.804)		158.631** (62.786)	179.523*** (66.506)	154.913** (62.390)	201.239** (95.577)
Pred. Diversity Sqr.	-155.442*** (49.298)		-113.110** (46.160)	-126.147** (49.006)	-109.806** (45.702)	-145.894** (66.791)
Log Transition Timing		1.393*** (0.170)	1.228*** (0.176)		1.374*** (0.151)	1.603*** (0.271)
Log Arable % of Land				0.546*** (0.145)	0.371*** (0.104)	0.370*** (0.115)
Log Absolute Latitude				-0.151 (0.106)	-0.380*** (0.104)	-0.373*** (0.135)
Log Agri. Suitability				0.043 (0.143)	0.211** (0.103)	0.190* (0.107)
Optimal Diversity	0.707*** (0.055)		0.701*** (0.133)	0.712*** (0.162)	0.705** (0.290)	0.690** (0.323)
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

Notes: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

migrations and, hence, genetic mixtures arguably occurred with even less frequency than in 1500 CE, the current analysis performs, in principle, a somewhat cleaner test of the proposed diversity hypothesis.

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 6 and 7 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 5, the unconditional effects of genetic diversity in Tables 6 and 7 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 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 5, 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

expanse of time examined in this study.

TABLE 7: Predicted Diversity and Economic Development in 1 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable is Log Population Density in 1 CE					
Pred. Diversity	227.826*** (70.543)		183.142*** (59.578)	129.180* (67.847)	134.767** (63.446)	231.689** (115.826)
Pred. Diversity Sqr.	-160.351*** (52.009)		-132.373*** (43.559)	-88.040* (50.172)	-96.253** (46.491)	-166.859** (81.126)
Log Transition Timing		1.793*** (0.216)	1.636*** (0.209)		1.662*** (0.210)	2.127*** (0.443)
Log Arable % of Land				0.377** (0.152)	0.314*** (0.116)	0.348*** (0.128)
Log Absolute Latitude				0.190 (0.122)	-0.121 (0.116)	-0.115 (0.133)
Log Agri. Suitability				0.160 (0.169)	0.238* (0.123)	0.210* (0.122)
Optimal Diversity	0.710*** (0.038)		0.692*** (0.018)	0.734* (0.397)	0.700** (0.313)	0.694*** (0.098)
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

Notes: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

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 6 and 7 reveals the results from estimating the preferred specification for log population density in 1000 CE and 1 CE, exploiting the combined explanatory power of all three identified channels. 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. The estimated coefficients 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 32.21%, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 15.07%. 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 26.43%, whereas a similar decrease in diversity for the most diverse society would raise its population density by 14.26%.³⁶ In comparison to the corresponding results for population density in 1500 CE from Table 5, 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

³⁶These effects are calculated directly via the methodology outlined in Footnote 28 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.

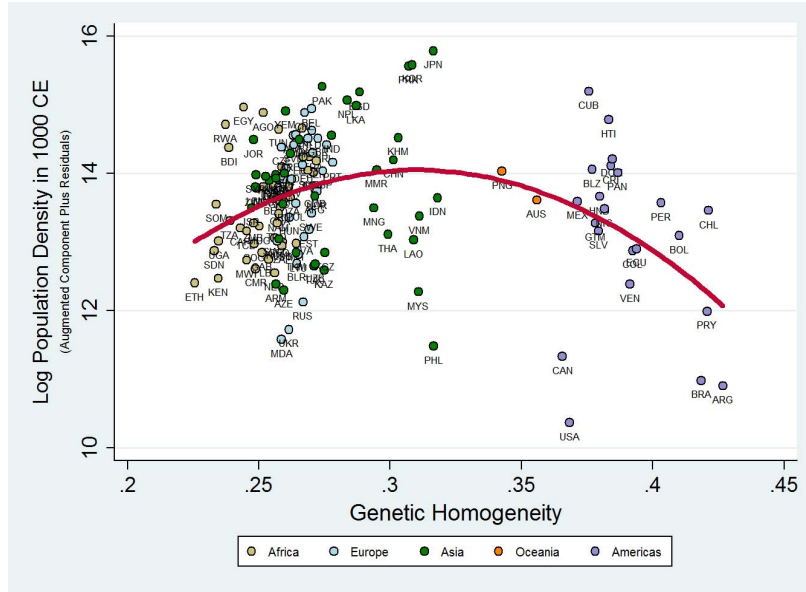


FIGURE 6: Predicted Diversity and Population Density in 1000 CE - Conditional on Transition Timing, Land Productivity and Continental Fixed Effects

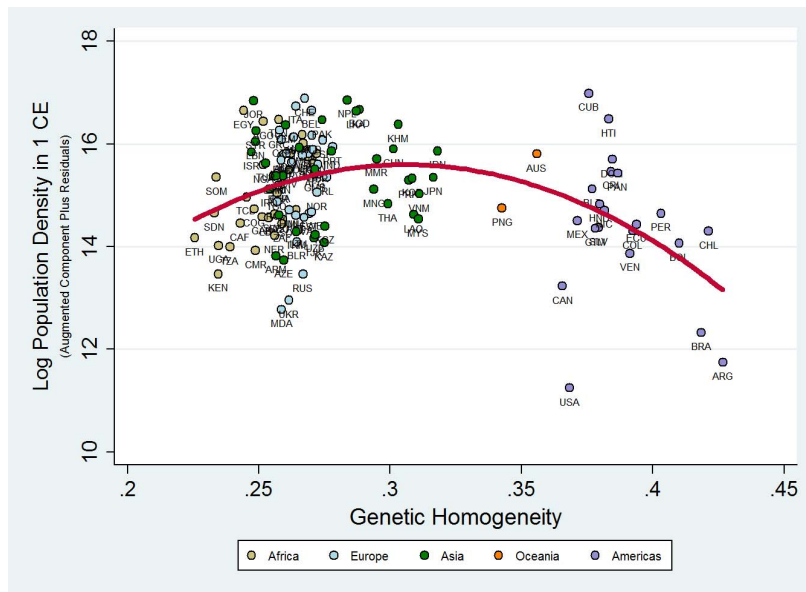


FIGURE 7: Predicted Diversity and Population Density in 1 CE - Conditional on Transition Timing, Land Productivity and Continental Fixed Effects

TABLE 8: Pairwise Correlations Between Alternative Distances

	1	2	3	4	5
1 Migratory Distance from Addis Ababa	1.000				
2 Aerial Distance from Addis Ababa	0.934	1.000			
3 Migratory Distance from London	0.899	0.801	1.000		
4 Migratory Distance from Tokyo	0.266	0.172	0.418	1.000	
5 Migratory Distance from Mexico City	-0.822	-0.759	-0.675	-0.025	1.000

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. The hump-shaped relationships, as implied by 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 6 and 7 suggest that, consistent with the predictions of the proposed genetic 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 robustness of the diversity channel to possible interactions with the level of technological sophistication that increased in the course of 1500 years of global agricultural development.

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.³⁷ As

³⁷The 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

TABLE 9: 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					
Mig. Distance	0.138** (0.061)		-0.040 (0.063)	0.052 (0.145)	-0.063 (0.099)
Mig. 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 Arable % of 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 Agri. 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

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

shown in Table 8, 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 9 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.

distance from locations outside of East Africa. Indeed, other points of origin in Europe, Asia and the Americas yield qualitatively similar results.

TABLE 10: 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

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 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, including Keller (2001, 2002) and Eaton and Kortum (2002), have uncovered strong empirical support for this hypothesis in explaining comparative development in the contemporary era.³⁸ 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 shortly, however, while the

³⁸The literature on international technology spillovers in the contemporary era is vast and, as such, the list of works cited above is meant to be illustrative rather than exhaustive. The interested reader is referred to Keller (2004) for a more comprehensive review of studies examining the technology diffusion hypothesis.

diffusion channel is indeed found to have been a significant determinant of comparative development in the pre-colonial era, the baseline results for genetic diversity remain highly 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, reported for that period and belonging to different civilizations or disparate sociopolitical entities, from each of Africa, Europe, Asia and the Americas.³⁹ 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 10.⁴⁰ In constructing the variable measuring distance to the closest regional frontier for a given historical period, the analysis then selects, for each country in the corresponding regression sample, the minimum value from the set of great circle distances between the country's capital city and the regional frontiers identified as being relevant for that period.

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 only -0.04 and 0.03. 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 11 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 5 of Table 5, the coefficients on genetic diversity continue to remain reassuringly stable in both magnitude and statistical significance. The same robustness characteristics may be noted for the transition timing and land productivity channels as well. Interestingly, the results also

³⁹The 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.

⁴⁰Note 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 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.

TABLE 11: Robustness to the Technology Diffusion Hypothesis

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Population Density in 1500 CE		Log Population Density in 1000 CE		Log Population Density in 1 CE	
Pred. Diversity	199.020*** (55.055)	156.736** (77.979)	182.903*** (61.415)	183.771** (91.195)	184.976*** (61.643)	215.858** (106.499)
Pred. Diversity Sqr.	-140.115*** (40.097)	-114.626** (54.672)	-129.824*** (44.815)	-134.609** (63.650)	-132.731*** (45.103)	-157.724** (74.815)
Log Transition Timing	0.986*** (0.155)	0.909*** (0.263)	1.165*** (0.167)	1.253*** (0.311)	1.328*** (0.208)	1.676*** (0.437)
Log Arable % of Land	0.367*** (0.092)	0.363*** (0.099)	0.327*** (0.103)	0.323*** (0.115)	0.304*** (0.114)	0.342*** (0.121)
Log Absolute Latitude	-0.402*** (0.092)	-0.492*** (0.129)	-0.430*** (0.109)	-0.454*** (0.146)	-0.232** (0.117)	-0.212 (0.138)
Log Agri. Suitability	0.317*** (0.086)	0.275*** (0.089)	0.265*** (0.097)	0.239** (0.103)	0.213* (0.114)	0.191* (0.114)
Log Dist. to Regional Frontier in 1500 CE	-0.190*** (0.064)	-0.187*** (0.070)				
Log Dist. to Regional Frontier in 1000 CE			-0.226** (0.095)	-0.230** (0.108)		
Log Dist. to Regional Frontier in 1 CE					-0.323*** (0.087)	-0.297*** (0.099)
Optimal Diversity	0.710*** (0.100)	0.684*** (0.172)	0.704*** (0.040)	0.683*** (0.172)	0.697*** (0.069)	0.684** (0.313)
Continent Dummies	No	Yes	No	Yes	No	Yes
Observations	145	145	140	140	126	126
R-squared	0.70	0.72	0.63	0.64	0.65	0.66

Notes: Bootstrap standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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.

Column 2 of Table 11 presents the results from repeating the regression exercise of the first column but with additional controls for continental fixed effects. Here again the robustness of the proposed genetic diversity channel is established. Specifically, in comparison to the regression results presented in Column 6 of Table 5, the estimated linear and quadratic coefficients of genetic diversity remain rather stable in magnitude and statistical significance, although less so relative to their high stability exhibited in the absence of continental dummies. This is symptomatic of the fact that the within-continent correlations between migratory distance from East Africa and distance from the closest regional frontier actually differ from the corresponding cross-continent or sample-wide correlation.⁴¹ Thus, once technology diffusion is accounted for, the reduction in omitted variable bias on the average within-continent influence of genetic diversity also differs from that which occurs for its cross-continent influence. Despite these issues, however, the results demonstrate that, having controlled for the technology diffusion channel, genetic diversity continues to remain as significant a determinant of economic development within continents as it is across continents in the year 1500 CE.

⁴¹For instance, while the sample-wide correlation between migratory distance and distance from the closest regional frontier in 1500 CE is 0.02, the corresponding correlations for the continent-specific sub-samples are: -0.09 for Africa (44 obs.), -0.02 for Europe (33 obs.), 0.42 for Asia (40 obs.), and 0.13 for the Americas (25 obs.).

Finally, Columns 3-6 establish the robustness of the genetic diversity channel in 1000 CE and 1 CE to controls for technology diffusion as originating from the regional technological frontiers identified for these earlier historical periods. Comparing Columns 3 and 4 with their respective baselines (i.e., Columns 5 and 6 in Table 6), the linear and quadratic coefficients of genetic diversity for the 1000 CE regressions remain largely stable under controls for technology diffusion, increasing moderately in both magnitude and statistical significance for the specification without continental dummies and decreasing slightly only in magnitude when continental fixed effects are accounted for by the analysis. A similar stability pattern also emerges for the coefficients capturing the influence of the genetic diversity channel in the 1 CE regressions. Moreover, 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, regardless of whether the exploited variation in the relevant distance variable is within continents or across continents.

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 not only lend further credence to the proposed genetic diversity channel itself, but also to the notion that the manner in which genetic diversity influences development did not fundamentally change as a result of possible interactions with the level of technological sophistication, which undoubtedly increased over the 1500-year period examined by the current analysis.

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.⁴² On the other hand, by isolating population subgroups geographically, a rugged landscape could also have nurtured their ethnic differentiation over time and may therefore confer an adverse effect on society’s aggregate productivity via the increased likelihood of ethnic conflict. Similarly, while proximity to waterways can directly affect agricultural crop yields by making beneficial practices such as irrigation possible, it may also have augmented aggregate productivity indirectly by lowering transportation costs and, thereby, fostering urban development, trade and technology diffusion.⁴³

To ensure that the significant effects of genetic diversity revealed by the baseline exercise 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

⁴²The 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).

⁴³Indeed, a significant positive relationship between proximity to waterways and contemporary population density has been demonstrated by Gallup et al. (1999).

TABLE 12: Robustness to Microgeographic Factors

	(1)	(2)	(3)	(4)	(5)	(6)
	Terrain Quality Controls		Waterway Access Controls		Combined Controls	
	Dependent Variable is Log Population Density in 1500 CE					
Pred. Diversity	159.924*** (56.001)	160.346** (77.314)	153.198*** (53.394)	157.073** (78.815)	150.016*** (49.359)	157.059** (68.611)
Pred. Diversity Sqr.	-110.390*** (41.077)	-118.716** (54.328)	-105.325*** (39.105)	-112.780** (55.478)	-102.757*** (36.232)	-114.994** (48.256)
Log Transition Timing	1.060*** (0.148)	1.131*** (0.225)	1.090*** (0.121)	1.211*** (0.201)	1.041*** (0.126)	1.215*** (0.197)
Log Arable % of Land	0.384*** (0.094)	0.397*** (0.099)	0.346*** (0.094)	0.348*** (0.099)	0.354*** (0.084)	0.374*** (0.087)
Log Absolute Latitude	-0.307*** (0.097)	-0.358*** (0.124)	-0.372*** (0.106)	-0.354*** (0.132)	-0.370*** (0.097)	-0.352*** (0.122)
Log Agri. Suitability	0.273*** (0.092)	0.188* (0.101)	0.262*** (0.081)	0.248*** (0.082)	0.219*** (0.077)	0.160** (0.081)
Mean Elevation	-0.475** (0.234)	-0.404 (0.251)			0.513* (0.271)	0.502* (0.273)
Roughness	5.145*** (1.773)	5.938*** (1.870)			3.086* (1.740)	4.076** (1.840)
Roughness Sqr.	-7.051** (3.113)	-7.332** (2.922)			-7.048** (2.960)	-7.627*** (2.906)
Mean Dist. to Nearest Waterway			-0.485*** (0.177)	-0.437** (0.178)	-0.474** (0.184)	-0.390** (0.181)
% Land within 100 km of Waterway			0.697** (0.279)	0.731** (0.310)	1.108*** (0.293)	1.175*** (0.294)
Optimal Diversity	0.724*** (0.201)	0.675*** (0.233)	0.727*** (0.190)	0.696*** (0.187)	0.730*** (0.229)	0.683*** (0.089)
Continent Dummies	No	Yes	No	Yes	No	Yes
Observations	145	145	145	145	145	145
R-squared	0.69	0.72	0.74	0.75	0.76	0.78

Notes: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

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.⁴⁴ 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 -0.19 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

⁴⁴For 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.

Columns 1 and 3 of Table 12. In each case, the coefficients associated with the diversity channel remain highly statistically significant and relatively stable, experiencing only a moderate decrease in magnitude, when compared to the baseline results from Table 5. Moreover, a similar stability pattern for the influence of genetic diversity emerges once continental fixed effects are also taken into account in Columns 2 and 4.

Interestingly, the control variables for terrain quality in Columns 1-2 and those gauging access to waterways in Columns 3-4 appear to confer statistically significant effects on population density, 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.⁴⁵ Further, regions possessing high mean elevations are on average not conducive for sustaining large populations whereas those with greater access to coasts and sea-navigable rivers are found to support higher population densities.

The final two columns of Table 12 examine 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. Indeed, this holds regardless of whether the regression specification incorporates continental dummies or not, assuring that the influence of genetic diversity remains robust both within and across continents. The results for the microgeographic factors, on the other hand, indicate that the linear effect of surface roughness on aggregate productivity loses some of its explanatory power to the measures gauging access to waterways while the effect of elevation switches direction in comparison to the estimates from Columns 1-2. This suggests that some of the effects of terrain quality revealed earlier were largely reflecting the latent influence of proximity to waterways due to the fact that these two dimensions of microgeography are obviously not orthogonal to one another. Nonetheless, it is apparent 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.

3.2.7 Robustness to Exogenous Factors in the Diamond Hypothesis

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 13. The purpose of this analysis is to demonstrate more rigorously 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 transition to agriculture.

Following the discussion in Section 3.1.2 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

⁴⁵Specifically, the roughness index in the regression sample ranges from a minimum value of 0.0127 to a maximum value of 0.6022. According to the coefficient estimates presented in Column 1, the optimal level of terrain roughness for population density is 0.3648, which implies that the effect of roughness on productivity is indeed non-monotonic over the sample and not just non-linear.

TABLE 13: Robustness to Ultimate Determinants in the Diamond Hypothesis

	(1)	(2)	(3)	(4)	(5)
Dependent Variable is Log Population Density in 1500 CE					
Pred. Diversity	216.847*** (62.059)	252.076*** (70.812)	174.414*** (62.754)	212.123*** (72.132)	274.916*** (72.117)
Pred. Diversity Sqr.	-154.750*** (45.185)	-180.650*** (51.890)	-125.137*** (45.720)	-151.579*** (52.794)	-197.120*** (52.402)
Log Transition Timing	1.300*** (0.156)				1.160*** (0.307)
Log Arable % of Land	0.437*** (0.108)	0.431*** (0.115)	0.441*** (0.106)	0.411*** (0.111)	0.365*** (0.106)
Log Absolute Latitude	-0.212** (0.106)	-0.426*** (0.133)	-0.496*** (0.154)	-0.487*** (0.162)	-0.332** (0.146)
Log Agri. Suitability	0.288** (0.130)	0.184 (0.137)	0.297** (0.139)	0.242* (0.141)	0.280** (0.119)
Climate		0.622*** (0.142)		0.419 (0.266)	0.374* (0.222)
Orientation of Axis		0.281 (0.333)		0.040 (0.296)	-0.169 (0.265)
Size of Continent		-0.007 (0.015)		-0.005 (0.013)	-0.006 (0.012)
Domesticable Plants			0.015 (0.018)	-0.005 (0.022)	0.003 (0.020)
Domesticable Animals			0.154** (0.062)	0.121 (0.074)	-0.013 (0.074)
Optimal Diversity	0.701*** (0.123)	0.698*** (0.016)	0.697*** (0.159)	0.700*** (0.045)	0.697*** (0.041)
Observations	96	96	96	96	96
R-squared	0.74	0.70	0.70	0.72	0.78

Notes: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

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 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. However, since data on these variables are obtained from the more limited sample employed by the study of Olsson and Hibbs (2005), the current analysis is necessarily restricted to a subset of the baseline sample, with the sub-sample being comprised of 96 as opposed to 145 countries.

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 regression specification for log population density in 1500 CE using the restricted sample of 96 countries. 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 5, 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.⁴⁶

⁴⁶Note 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

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 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. However, this somewhat unintuitive latter result is suspect given the strong correlation of 0.88 between the biogeographic variables. In addition, the high correlations of 0.82 and 0.78 between climate and the numbers of domesticable wild plants and animals, respectively, may also explain why 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. 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 13 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 13 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 the proposed genetic diversity channel on contemporary comparative development. Section 4.1.1, in particular, explores how the construction of the index of genetic diversity for contemporary national populations is accomplished based on data on their ethnic compositions arising from migrations amongst countries over the last 500 years. The analysis also discusses how the concept of genetic diversity between contemporary sub-national ethnic groups is incorporated into this index by applying the measure of F_{st} genetic distance, predicted by the pairwise migratory distances between the ancestral populations of these groups. To this effect, the discussion presents corresponding evidence from population genetics regarding the positive relationship between pairwise F_{st} genetic distances and pairwise migratory distances across all ethnic group pairs in the HGDP-CEPH sample of ethnicities. Section 4.1.2, on the other hand, outlines the overall empirical strategy adopted by the subsequent regression analysis.

with continental dummies, however, does not significantly alter the results for genetic diversity.

4.1.1 The Index of Contemporary National Population Diversity

The construction of the index of genetic diversity for contemporary national populations, as mentioned previously, 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 (i.e., the expected heterozygosity index employed in the earlier historical analysis), 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 era.⁴⁷ 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 entities in the pair captures the proportion of the combined genetic diversity of the pair that is unexplained by the average of the genetic diversities of the two sub-populations. Consider, for instance, a population comprised of two ethnic groups or sub-populations, A and B . The 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 θ_A and θ_B represent the shares of groups A and B , respectively, in the combined population; H_{exp}^A and H_{exp}^B are the respective expected heterozygosities of groups A and B ; and H_{exp}^{AB} is the expected heterozygosity of the combined population comprised of A and B . Thus, given information on genetic distance, F_{st}^{AB} , the expected heterozygosities of the component sub-populations, H_{exp}^A and H_{exp}^B , and their respective shares in the overall population, θ_A and θ_B , applying the definition of genetic distance permits the calculation of the overall diversity of the combined population as follows:

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

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 is available. In practice, 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 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

⁴⁷The data on ethnic compositions is obtained from the World Migration Matrix, 1500-2000 of Putterman and Weil (2009) who compile, for each country in their dataset, the share of the country's population in the year 2000 CE that is descended from the population of every other country in the year 1500 CE.

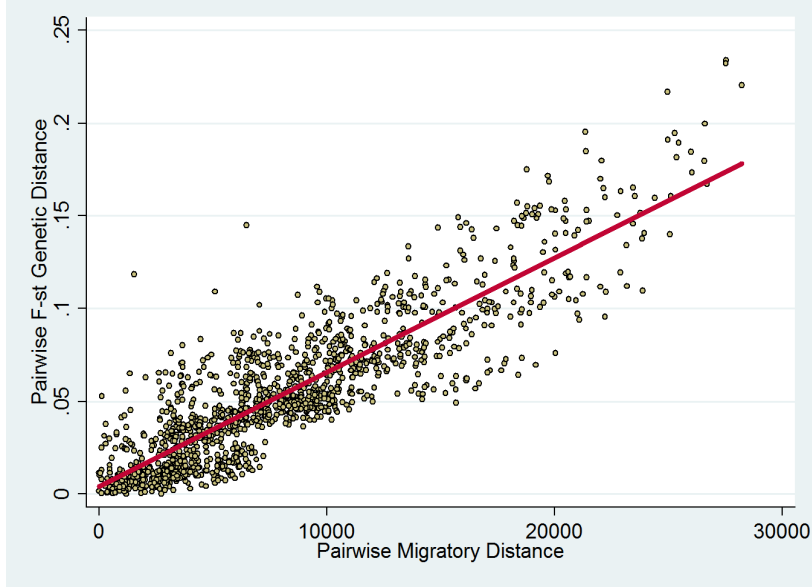


FIGURE 8: Pairwise F_{st} Genetic and Migratory Distances in the HGDP-CEPH Sample

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 ultimately residing at greater geographical distances from parental ones would also be more genetically differentiated from parental colonies. 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 prediction of the serial-founder effect is borne out in the population genetics data as well. Based on the analysis of Ramachandran et al. (2005), Figure 8 shows the strong positive correlation between pairwise migratory distances and pairwise genetic distances across all ethnic group pairs in the HGDP-CEPH sample. Specifically, according to the depicted 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.

To be precise, 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:

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

where, for $i \in \{A, B\}$, $\hat{H}_{\text{exp}}^i(d_i)$ is 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 θ_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 associ-

TABLE 14: Adjusted versus Unadjusted Diversity

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable is Log Income Per Capita in 2000 CE						
Adj. Genetic Diversity	556.439*** (134.034)	254.906*** (89.814)			533.983*** (167.245)	387.314** (188.496)
Adj. Genetic Diversity Sqr.	-397.224*** (93.793)	-176.907*** (63.553)			-377.365*** (119.747)	-273.925** (136.677)
Genetic Diversity			140.903*** (53.395)	10.152 (54.148)	1.670 (70.175)	-64.226 (82.555)
Genetic Diversity Sqr.			-107.686*** (39.408)	-7.418 (39.003)	-4.057 (53.740)	51.016 (65.051)
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

Notes: Bootstrap standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

ated with the regression depicted in Figure 8. In reality, since contemporary national populations are typically composed of more than two ethnic groups, the aforementioned procedure is applied recursively to incorporate a larger number of component ethnic groups in modern populations.

Table 14 presents some preliminary findings comparing the index of diversity for contemporary populations with that employed by the historical analysis. In line with the diversity hypothesis, Column 1 reveals a significant unconditional hump-shaped relationship between the contemporary measure of diversity and income per capita in the year 2000 CE. This relationship remains qualitatively intact when conditioned for the impact of continent fixed effects in Column 2. As shown in Columns 3 and 4, however, while the historical 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 contemporary and historical measures of diversity for income per capita, Columns 5 and 6 demonstrate the superior relative performance of the contemporary measure, regardless of whether continent fixed effects are accounted for by the comparative analyses, lending further credence *ex post* to the methodology employed in constructing the index of contemporary population diversity.

4.1.2 The Empirical Model

Maintaining symmetry with the earlier historical analysis, a specification similar to that employed for the historical regressions is adopted to examine the contemporary impact of genetic diversity along with the Neolithic 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 to examine the direct influence of contem-

porary 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 in detail above; T_i and X_i are the Neolithic transition timing and land productivity controls for country i ; Λ_i is a vector of institutional and cultural controls for country i ; Γ_i is a vector of additional geographical controls for country i ; and, finally, η_i is a country specific disturbance term.⁴⁸

4.2 Empirical Findings

The empirical finding suggests that the highly significant hump-shaped effect of genetic diversity on macroeconomic outcomes in the pre-industrial period is present in the post-industrial period as well. Furthermore, it has a direct long-lasting hump-shaped effect on the pattern of comparative economic development that could not be captured by contemporary geographical, institutional, and cultural factors. Moreover, it shows, as predicted by the theory, that the optimal level of diversity indeed increases in an environment characterized by more rapid technological progress.

Using a sample of 143 countries for which data is available for the entire set of control variables used in the baseline regression for the year 1500 CE, Column 1 of Table 15 shows a hump-shaped effect of genetic diversity on income per capita in the year 2000 CE, accounting for the set of controls employed in the baseline regression for the year 1500 CE - i.e., the logs of the timing of the Neolithic transition, the arable percentage of land, absolute latitude and the suitability of land for agriculture, as well as continental fixed effects. Further, as predicted by the theory, the optimal level of diversity indeed increases in an environment characterized by more rapid technological progress. While the optimal level of diversity in the year 1500 CE is 0.683 (Column 4), the optimal level of diversity in the year 2000 CE, under the same specification, is 0.713.

Interestingly, unlike the conjecture of Diamond (1997), the timing of the Neolithic Revolution has no effect on contemporary economic outcomes. Moreover, as established in Column 2, the hump-shaped effect of genetic diversity on income per capita in the year 2000 CE remains virtually intact, and the effect of the Neolithic Revolution, capturing the average time elapsed since the ancestors of the population of each country today experienced the transition to agriculture (i.e., traits that are embodied in the current population, rather than the country's geographical attributes), remains insignificant statistically.

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 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%. Furthermore,

⁴⁸The data on income per capita is 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 democracy from the Polity IV dataset, legal origin dummies and the shares of the population affiliated with major world religions from the dataset 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 dataset of Gallup et al. (1999).

TABLE 15: 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
Adj. Genetic Diversity	204.610** (88.466)	237.238*** (86.278)	244.960*** (85.454)	
Adj. Genetic Diversity Sqr.	-143.437** (62.545)	-166.507*** (61.363)	-171.364*** (60.843)	
Genetic Diversity				198.587** (79.110)
Genetic Diversity Sqr.				-145.320*** (55.472)
Log Adj. Transition Timing		0.061 (0.262)	0.002 (0.305)	
Log Transition Timing	-0.151 (0.186)			1.238*** (0.230)
Log Arable % of Land	-0.110 (0.100)	-0.119 (0.107)	-0.137 (0.111)	0.378*** (0.100)
Log Absolute Latitude	0.164 (0.125)	0.172 (0.119)	0.192 (0.143)	-0.423*** (0.124)
Log Agri. Suitability	-0.193** (0.095)	-0.177* (0.102)	-0.189* (0.102)	0.264*** (0.096)
Log Population Density in 1500 CE			0.047 (0.097)	
Optimal Diversity	0.713*** (0.094)	0.712*** (0.037)	0.715*** (0.133)	0.683*** (0.177)
Continent Dummies	Yes	Yes	Yes	Yes
Observations	143	143	143	143
R-squared	0.57	0.57	0.57	0.68

Notes: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

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%.⁴⁹

Importantly, the effect of genetic diversity on income per capita in the year 2000 CE does not reflect any historical inertia from its effect on technology and, thus, population density in the year 1500 CE. As established in Column 4, the qualitative results remain virtually intact if the regression accounts for the potentially confounding effect of population density in the year 1500 CE. Namely, the effect of genetic diversity on income per capita the year 2000 CE does not operate through its effect on population density in the year 1500 CE.

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

Using a sample of 109 countries for which data is available for the institutional and cultural controls that are employed in the examination, Column 1 of Table 16 demonstrates that genetic diversity has a hump-shaped effect on income per capital in the year 2000 CE, accounting for the set of controls employed in the baseline regression for the year 1500 CE - i.e., the logs of the weighted timing of the Neolithic transition, the arable percentage of land, absolute latitude, and the suitability of land for agriculture, as well as continental fixed effects. The estimated linear and quadratic coefficients are both statistically significant at the 1% level and the optimal level of diversity is 0.713. Columns 2 and 3 examine the robustness of the results to the inclusion of measures of institutional quality, as captured by the social infrastructure index of Hall and Jones (1999). The estimated effect of genetic diversity remains highly significant and rather stable. In particular the optimal level of diversity increases to 0.725. As indicated by the regression in Column 3, the inclusion of the Polity IV democracy index does not affect the results, and since it is insignificant it is dropped from subsequent regressions. Column 4 demonstrates the robustness of the 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 religious). The estimated effect of genetic diversity remains highly significant and rather stable. In particular the optimal level of diversity, 0.723, remains virtually intact.

The regression in Column 5 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 effect of ethnic fractionalization. While, as was established earlier in the literature, ethnic fractionalization has a significant adverse effect on income per capita in the year 2000 CE, the hump-shaped effect of genetic diversity remains highly significant and the optimal level of diversity, 0.722, is virtually unchanged. 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, and an OPEC dummy.

⁴⁹Following 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 28 for details on how these effects may be computed based on the estimated linear and quadratic coefficients associated with genetic diversity.

TABLE 16: 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								
Adj. Genetic Diversity	315.282*** (84.215)	225.858*** (67.669)	204.102*** (66.984)	252.826*** (66.150)	245.377*** (66.845)	277.342*** (70.232)	271.249*** (75.904)	215.675*** (63.954)
Adj. Genetic Diversity Sqr.	-220.980*** (59.562)	-155.826*** (47.962)	-140.850*** (47.393)	-174.955*** (46.808)	-170.036*** (47.195)	-192.386*** (49.675)	-188.899*** (54.142)	-150.871*** (45.554)
Log Adj. Transition Timing	-0.273 (0.269)	-0.092 (0.200)	-0.062 (0.203)	0.296 (0.246)	0.352 (0.242)	0.396* (0.233)	0.112 (0.234)	-0.046 (0.208)
Log Arable % of Land	-0.218*** (0.061)	-0.159*** (0.049)	-0.163*** (0.050)	-0.209*** (0.048)	-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 (0.101)	0.164* (0.091)	0.119 (0.097)	0.009 (0.108)	0.007 (0.099)	-0.006 (0.087)
Social Infrastructure		2.359*** (0.269)	2.069*** (0.377)	2.018*** (0.376)	2.072*** (0.375)	1.826*** (0.417)	1.470*** (0.426)	0.880** (0.418)
Democracy			0.036 (0.029)					
Ethnic Fractionalization					-0.505 (0.319)	-0.333 (0.280)	0.051 (0.280)	-0.122 (0.265)
% Population at Risk of Contracting Malaria						-0.502 (0.351)	-0.734* (0.389)	-0.723** (0.353)
% Population Living in Tropical Zones						-0.319 (0.204)	-0.411* (0.214)	-0.185 (0.199)
Mean Dist. to Nearest Waterway						-0.368** (0.186)	-0.179 (0.216)	-0.062 (0.188)
Avg. Years of Schooling								0.134*** (0.042)
Optimal Diversity	0.713*** (0.014)	0.725*** (0.032)	0.725*** (0.269)	0.723*** (0.013)	0.722*** (0.014)	0.721*** (0.060)	0.718*** (0.048)	0.715*** (0.073)
OPEC Dummy	No	No	No	No	No	Yes	Yes	Yes
Legal Origin Dummies	No	No	No	Yes	Yes	Yes	Yes	Yes
Major Religion Shares	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	109	109	109	109	109	109	94	94
R-squared	0.74	0.84	0.85	0.86	0.87	0.90	0.91	0.93

Notes: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; All regressions include sub-Saharan Africa and continent dummies.

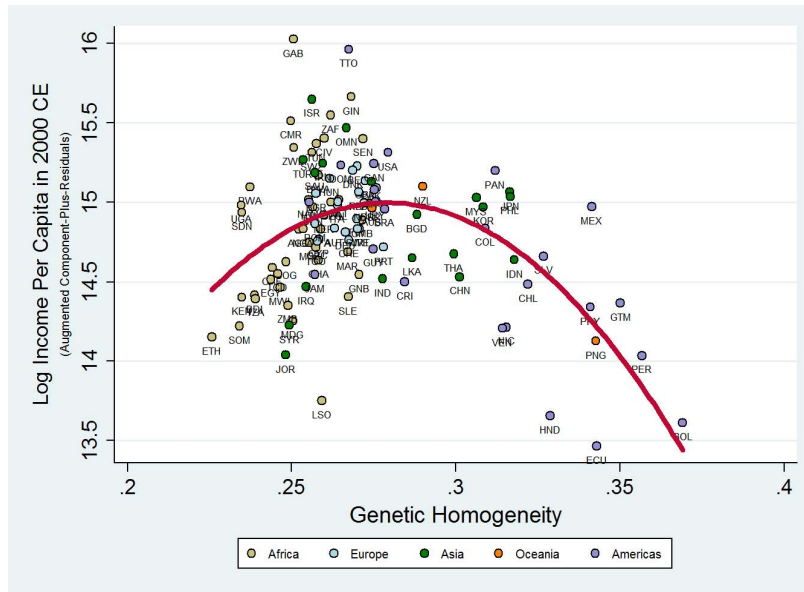


FIGURE 9: Adjusted Diversity and Income Per Capita in 2000 CE - Conditional on Transition Timing, Land Productivity, Institutional and Geographical Determinants, and Continental Fixed Effects

Hence, the regression in Column 6, as depicted in Figure 9, reveals the direct effect of genetic diversity, once institutional, cultural, and geographical factors are accounted for.⁵⁰ It shows that (i) increasing the diversity of the most homogenous country in the sample (Bolivia) by 1 percentage point would raise its income per capita 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 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 optimum level of 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 11.7:1 to 2.5:1, and (v) decreasing the diversity of Ethiopia to the optimum level of the U.S. would increase Ethiopia's per capita income by a factor of 1.73, closing the income gap between the U.S. and Ethiopia from 47.4:1 to 27.4:1.

Moreover, even if one accounts for the contribution of human capital formation in the time period 1960-2000, as reported in Column 8, the hump-shaped effect of genetic diversity remains highly significant and the optimal diversity slightly drops from 0.718 (as presented in Column 7, that accounts for the smaller sample of 94 countries for which data on education and all other variables are available) to 0.715.

⁵⁰The corresponding first- and second-order effects are depicted in Figures 2(a)-2(b) in the appendix.

TABLE 17: 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						
Adj. Genetic Diversity	277.342*** (70.232)	271.979*** (88.479)	261.367*** (70.533)	412.222*** (148.584)	264.805*** (111.365)	304.735*** (111.588)
Adj. Genetic Diversity Sqr.	-192.386*** (49.675)	-188.974*** (62.096)	-181.811*** (49.671)	-287.067*** (101.906)	-183.863*** (80.398)	-213.389*** (77.255)
Log Adj. Transition Timing	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 Arable % of 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)
% 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)
% 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 Dist. 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*** (0.083)	0.720*** (0.083)	0.719*** (0.016)	0.718*** (0.025)	0.720*** (0.171)	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: Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; All regressions include controls for major religion shares as well as OPEC, legal origin, sub-Saharan Africa and continent dummies.

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 migration towards prosperous countries and the increased ethnic diversity in these economies. As established in Table 17, the hump-shaped effect of genetic diversity remains highly significant and the optimal diversity remains virtually intact in the interval 0.714 and 0.721 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) to countries whose indigenous population is larger the 97% of the entire population, virtually eliminating the role of migration in the creation of diversity, in Column 6.

5 Concluding Remarks

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 direct long-lasting effect on the pattern of comparative economic development that could not be captured by contemporary 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 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 (i.e., genetic diversity and genetic distance among and between its ancestral populations), has a similar non-monotonic effect on contemporary levels of income per capita. 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. 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.

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 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 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 optimum level of 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 11.7:1 to 2.5:1, and (v) decreasing the diversity of Ethiopia to the optimum level of the U.S. would increase Ethiopia's per capita income by a factor of 1.73, closing the income gap between the U.S. and Ethiopia from 47.4:1 to 27.4:1.

APPENDIX A: THE HGDP-CEPH SAMPLE OF 53 ETHNIC GROUPS

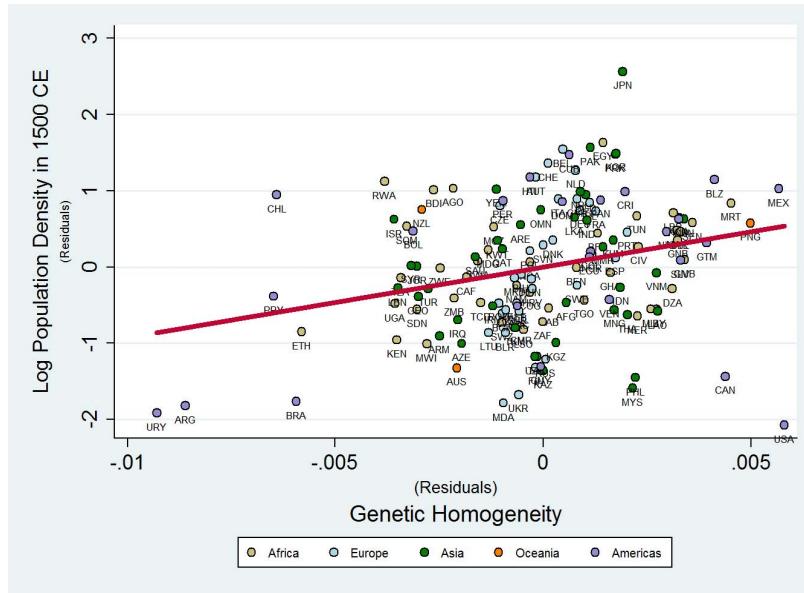
Ethnic Group	Migratory Distance (in km)	Country	Region
Bantu (Kenya)	1,338.94	Kenya	Africa
Bantu (Southeast)	4,306.19	South Africa	Africa
Bantu (Southwest)	3,946.44	Namibia	Africa
Biaka Pygmy	2,384.86	Central African Republic	Africa
Mandenka	5,469.91	Senegal	Africa
Mbuti Pygmy	1,335.50	Zaire	Africa
San	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

APPENDIX B: VARIABLE DEFINITIONS AND SOURCES

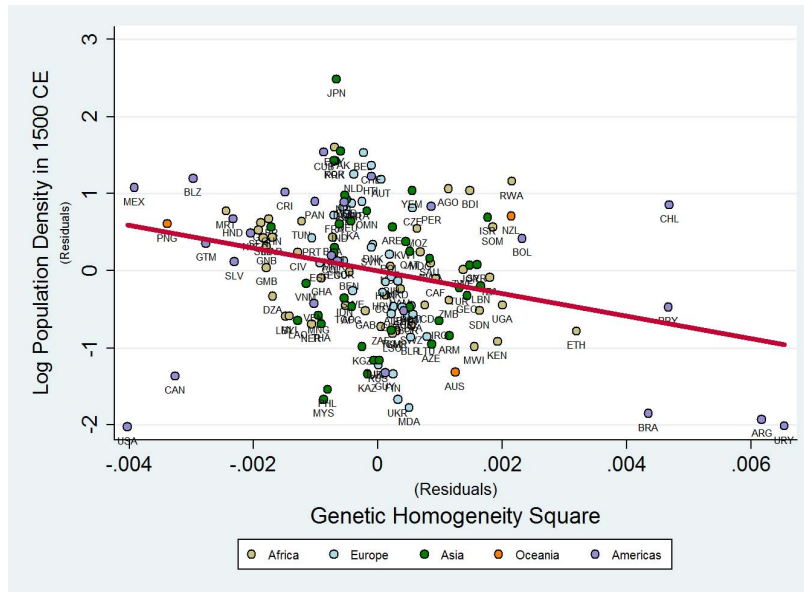
Variable	Definition
Population Density in 1 CE, 1000 CE and 1500 CE	Population density calculated as total population divided by total land area in 1 CE, 1000 CE and 1500 CE, respectively. Footnote 25 provides some additional details. <i>Source:</i> McEvedy and Jones (1978).
Migratory Distance from E. Africa in the Limited Country Sample	The average migratory distance of ethnic groups from the HGDP-CEPH sample that are located within a country. The migratory distance of a given 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 as described in Section 3.1.1. Distance is calculated using the Haversine formula and measured in units of 1000 km. <i>Source:</i> Ethnic groups, waypoints and their coordinates are from Ramachandran et al. (2005).
Actual Diversity in the Limited Country Sample	Expected heterozygosity (genetic diversity) averaged across ethnic groups from the HGDP-CEPH sample that are located within a country. <i>Source:</i> Expected heterozygosities of ethnic groups are from Ramachandran et al. (2005).
Migratory Distance from E. Africa in the Extended Country 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 intercontinental waypoints as described in Section 3.1.1. Distance is calculated using the Haversine formula. <i>Source:</i> Waypoints and their coordinates are from Ramachandran et al. (2005); modern capital cities and their coordinates are from the CIA World Factbook online.
Predicted Diversity in the Extended Country Sample	Expected heterozygosity (genetic diversity) as predicted by migratory distance from East Africa. Calculated by applying the beta coefficient obtained from regressing expected heterozygosity on migratory distance in the HGDP-CEPH sample of ethnic groups. <i>Source:</i> Authors' calculations, based on ethnic group expected heterozygosity data from Ramachandran et al. (2005).
Transition Timing	The total number of years elapsed since the transition to agriculture. <i>Source:</i> Putterman (2008).
Arable % of Land	The arable percentage of total land area. <i>Source:</i> World Bank, WDI online.
Absolute Latitude	The absolute value of the latitude of the country's centroid. <i>Source:</i> CIA World Factbook online.
Agricultural Suitability	An index of the suitability of land for agriculture based on soil pH levels and temperature. For more details, see Footnote 24. <i>Source:</i> Michalopoulos (2007).
Aerial Distance from E. Africa	The great circle distance from Addis Ababa, Ethiopia to the country's modern capital city. <i>Source:</i> Authors' calculations, based on coordinates from the CIA World Factbook online.
Migratory Distance from a "Placebo" Point of Origin	The great circle distance from a "placebo" location to the country's modern capital city, along a land-restricted path forced through one or more of five intercontinental waypoints as described in Section 3.1.1. Distance is calculated using the Haversine formula. <i>Source:</i> Waypoints and their coordinates are from Ramachandran et al. (2005); modern capital cities and their coordinates are from the CIA World Factbook online.

Variable	Definition
Distance to the Regional Frontier in 1 CE, 1000 CE and 1500 CE	The great circle distance to the closest regional frontier in 1 CE, 1000 CE and 1500 CE, respectively, from the country's capital city. Regional frontiers are identified with a selection criterion using urbanization estimates as described in Section 3.2.5. <i>Source</i> : Historical urbanization data are from Chandler (1987) and Modelski (2003); coordinates of ancient cities are obtained using Wikipedia online and the Google Earth program.
Mean Elevation	The mean elevation of a country calculated using gridded elevation data from the G-ECON project at a granularity of 1° latitude x 1° longitude by averaging across the grid cells assigned to the country. <i>Source</i> : Nordhaus (2006), available online at the website of the G-ECON project.
Terrain Roughness	The surface roughness of a country calculated using gridded roughness data from the G-ECON project at a granularity of 1° latitude x 1° longitude by averaging across the grid cells assigned to the country. The definition of roughness may be found in the G-ECON documentation. <i>Source</i> : Nordhaus (2006), available online at the website of the G-ECON project.
Mean Distance to Nearest Waterway	The expected distance from any GIS grid cell within a country to the nearest ice-free coastline or sea-navigable river. <i>Source</i> : Gallup et al. (1999), available online at the website of the Center for International Development.
% of Land within 100km of Waterway	The percentage of a country's land area located within 100km of the nearest ice-free coastline or sea-navigable river. <i>Source</i> : Gallup et al. (1999), available online at the website of the Center for International Development.
Climate	An index of climatic suitability for agriculture based on the Köppen-Geiger climate classification system. <i>Source</i> : Olsson and Hibbs (2005).
Orientation of Axis	Major axis orientation of the continent (or landmass) calculated as the ratio of the largest longitudinal (or East-West) distance to the largest latitudinal (or North-South) distance. <i>Source</i> : Olsson and Hibbs (2005).
Size of Continent	Size calculated as total land area of the continent (or landmass). <i>Source</i> : Olsson and Hibbs (2005).
Domesticable Plants and Animals	The number of species of plants and animals, respectively, prehistorically native to the continent (or landmass) and amenable for domestication. <i>Source</i> : Olsson and Hibbs (2005).

APPENDIX C: ADDITIONAL SCATTER PLOTS

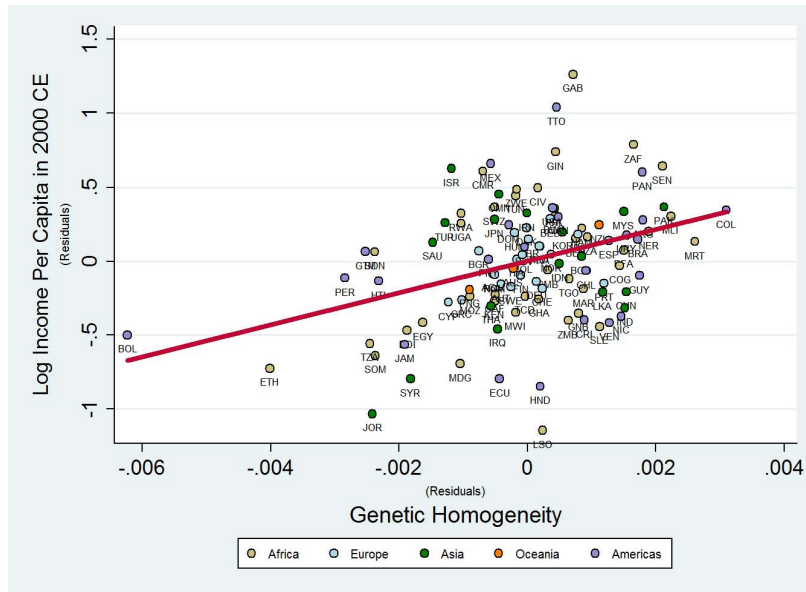


(a) The First-Order Effect

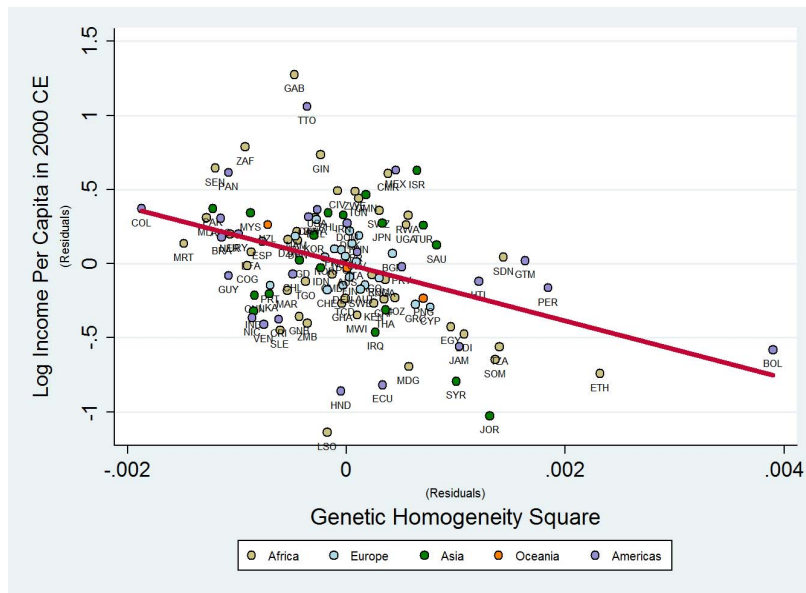


(b) The Second-Order Effect

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

APPENDIX D: LIMITED-SAMPLE RESULTS UNDER SPATIAL DEPENDENCE

TABLE D.1: Results of Table 3 with Correction for Spatial Dependence in Errors

	(1)	(2)	(3)	(4)	(5)
Dependent Variable is Log Population Density in 1500 CE					
Genetic Diversity	413.505*** [83.891]			225.441*** [57.654]	203.815*** [68.352]
Genetic Diversity Sqr.	-302.647*** [63.420]			-161.159*** [43.910]	-145.717*** [55.743]
Log Transition Timing		2.396*** [0.252]		1.214*** [0.268]	1.135*** [0.394]
Log Arable % of Land			0.730*** [0.270]	0.516*** [0.138]	0.545*** [0.189]
Log Absolute Latitude			0.145 [0.194]	-0.162* [0.090]	-0.129 [0.101]
Log Agri. Suitability			0.734** [0.357]	0.571*** [0.221]	0.587*** [0.219]
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: (1) Standard errors corrected for spatial autocorrelation, following Conley (1999), are reported in brackets; (2) The spatial distribution of observations is specified on the Euclidean plane using aerial distances between all pairs in the sample; (3) The spatial autocorrelation is modelled as linearly declining away from each location upto a threshold of 5000km, beyond which the correlation is assumed to be zero; (4) This threshold excludes spatial interactions between the Old World and the New World, which is appropriate for the historical period being considered; (5) *** p<0.01, ** p<0.05, * p<0.1.

TABLE D.2: Results of Table 4 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						
Genetic Diversity		255.220*** [82.086]		196.041** [76.684]	241.209*** [72.264]	194.499** [89.631]
Genetic Diversity Sqr.		-209.808*** [61.177]		-128.524* [67.586]	-173.105*** [54.782]	-139.686** [67.063]
Mig. Distance	0.505*** [0.108]	0.070 [0.139]	0.293*** [0.104]	0.097 [0.149]		
Mig. Distance Sqr.	-0.023*** [0.004]	-0.014* [0.007]	-0.015*** [0.004]	0.001 [0.005]		
Log Transition Timing			1.532*** [0.232]	1.583*** [0.594]	1.117*** [0.225]	1.017*** [0.283]
Log Arable % of Land			0.415*** [0.158]	0.471*** [0.108]	0.509*** [0.183]	0.542** [0.245]
Log Absolute Latitude			-0.308** [0.133]	-0.183 [0.115]	-0.125 [0.100]	-0.135* [0.073]
Log Agri. Suitability			0.585*** [0.223]	0.524** [0.218]	0.582*** [0.179]	0.586*** [0.216]
Continent Dummies	No	No	No	No	No	Yes
Observations	21	21	21	21	21	21
R-squared	0.34	0.46	0.88	0.90	0.89	0.90
Overid. Restrictions Test (P-value)					0.898	0.935

Notes: (1) Standard errors corrected for spatial autocorrelation, following Conley (1999), are reported in square brackets; (2) The spatial distribution of observations is specified on the Euclidean plane using aerial distances between all pairs in the sample; (3) The spatial autocorrelation is modelled as linearly declining away from each location upto a threshold of 5000km, beyond which the correlation coefficient is assumed to be zero; (4) This threshold effectively excludes spatial interactions between the Old World and the New World, which is appropriate for the historical period being considered; (5) Columns 5-6 present the results from estimating the corresponding 2SLS specifications in Table 4 using Conley's spatial GMM estimation procedure (Conley, 1999); (6) *** p<0.01, ** p<0.05, * p<0.1.

References

- Acemoglu, Daron, Simon Johnson, and James A. Robinson**, “The Colonial Origins of Comparative Development: An Empirical Investigation,” *American Economic Review*, December 2001, *91* (5), 1369–1401.
- , —, and —, “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution,” *Quarterly Journal of Economics*, November 2002, *117* (4), 1231–1294.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Roman 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.
- 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.
- Barro, Robert J. and Rachel M. McCleary**, “Religion and Economic Growth across Countries,” *American Sociological Review*, October 2003, *68* (5), 760–781.
- Bertocchi, Graziella and Fabio Canova**, “Did Colonization Matter for Growth? An Empirical Exploration into the Historical Causes of Africa’s Underdevelopment,” *European Economic Review*, December 2002, *46* (10), 1851–1871.
- 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.
- 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.
- and —, “Tropics, Germs and Crops: How Endowments Influence Economic Development,” *Journal of Monetary Economics*, January 2003, *50* (1), 3–39.
- Eaton, Jonathan and Samuel Kortum**, “Technology, Geography, and Trade,” *Econometrica*, September 2002, *70* (5), 1741–1779.
- 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.
- Fernández-Villaverde, Jesús**, “Was Malthus Right? Economic Growth and Population Dynamics,” 2001. Working Paper, Department of Economics, University of Pennsylvania.
- 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**, “People’s Opium? Religion and Economic Attitudes,” *Journal of Monetary Economics*, January 2003, *50* (1), 225–282.
- , —, and —, “Does Culture Affect Economic Outcomes?,” *Journal of Economic Perspectives*, Spring 2006, *20* (2), 23–48.
- , —, and —, “Cultural Biases in Economic Exchange?,” *Quarterly Journal of Economics*, August 2009, *124* (3), 1095–1131.
- Gylfason, Thorvaldur**, “Natural Resources, Education, and Economic Development,” *European Economic Review*, May 2001, *45* (4-6), 847–859.
- 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.
- Hong, Lu and Scott E. Page**, “Problem Solving by Heterogeneous Agents,” *Journal of Economic Theory*, March 2001, *97* (1), 123–163.
- 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**, “Knowledge Spillovers at the World’s Technology Frontier,” 2001. CEPR Discussion Paper No. 2815.
- , “Geographic Localization of International Technology Diffusion,” *American Economic Review*, March 2002, *92* (1), 120–142.
- , “International Technology Diffusion,” *Journal of Economic Literature*, September 2004, *42* (3), 752–782.
- Kremer, Michael**, “Population Growth and Technological Change: One Million B.C. to 1990,” *Quarterly Journal of Economics*, August 1993, *108* (3), 681–716.
- 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.
- Lazear, Edward P.**, “Culture and Language,” *Journal of Political Economy*, December 1999, *107* (6-2), S95–S126.
- , “Globalisation and the Market for Team-Mates,” *Economic Journal*, March 1999, *109* (454), C15–C40.
- 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.
- Malthus, Thomas R.**, *An Essay on the Principle of Population*, London, UK: J. Johnson, in St. Paul’s Church-Yard, 1798.
- McEvedy, Colin and Richard Jones**, *Atlas of World Population History*, New York, NY: Penguin Books Ltd., 1978.
- Michalopoulos, Stelios**, “Ethnolinguistic Diversity: Origins and Implications,” 2007. Mimeo, Department of Economics, Brown 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.**, *Structure and Change in Economic History*, New York, NY: W. W. Norton & Co., 1981.
- 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.
- Ottaviano, Gianmarco I. P. and Giovanni Peri**, “Cities and Cultures,” *Journal of Urban Economics*, September 2005, 58 (2), 304–337.
- and — , “The Economic Value of Cultural Diversity: Evidence from U.S. Cities,” *Journal of Economic Geography*, January 2006, 6 (1), 9–44.
- 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.
- Prat, Andrea**, “Should a Team be Homogenous?,” *European Economic Review*, July 2002, 46 (7), 1187–1207.
- 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.
- 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.
- 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**, “Culture and Institutions,” 2007. CEPR Discussion Paper No. 6589.
- 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.
- Wells, Spencer**, *The Journey of Man: A Genetic Odyssey*, Princeton, NJ: Princeton University Press, 2002.