Massachusetts Institute of Technology
Department of Economics
Working Paper Series

The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development

Quamrul Ashraf
Oded Galor

Working Paper 11-08
May 5, 2011
Rev. July 3, 2012
Rev. Aug 12, 2012

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development

Quamrul Ashraf Oded Galor*

Forthcoming in the *American Economic Review*

Abstract

This research advances and empirically establishes the hypothesis that, in the course of the prehistoric exodus of *Homo sapiens* out of Africa, variation in migratory distance to various settlements across the globe affected genetic diversity and has had a long-lasting hump-shaped effect on comparative economic development, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. While intermediate levels of genetic diversity prevalent among Asian and European populations have been conducive for development, the high diversity of African populations and the low diversity of Native American populations have been detrimental for the development of these regions. (*JEL* N10, N30, N50, O10, O50, Z10)

*Ashraf: Department of Economics, Williams College, 24 Hopkins Hall Dr., Williamstown, MA 01267 (email: Quamrul.H.Ashraf@williams.edu); Galor: Department of Economics, Brown University, 64 Waterman St., Providence, RI 02912 (email: Oded_Galor@brown.edu). The authors are grateful to five anonymous referees, Alberto Alesina, Kenneth Arrow, Alberto Bisin, Dror Brenner, John Campbell, Steve Davis, David Genesove, Douglas Gollin, Sergiu Hart, Saul Lach, Ross Levine, Nathan Nunn, Ola Olsson, Mark Rosenzweig, Antonio Spilimbergo, Enrico Spolaore, Alan Templeton, Romain Wacziarg, and David Weil; seminar participants at Aix-Marseille, Bar-Ilan, Barcelona, Ben-Gurion, Brown, Boston College, Chicago GSB, Copenhagen, Doshisha, Groningen, Haifa, Harvard, Hebrew U, Hitotsubashi, the IMF, Keio, Kyoto, Luxembourg, MIT, Osaka, Porto, Sciences Po, Tel Aviv, Tokyo, Tufts, UCLA Anderson, UPF, Williams, the World Bank, and Yale; and conference participants of the CEPR EHRTN Summer Workshop on "From Stagnation to Growth: Unified Growth Theory" in Florence, the 2nd Annual Conference on "Macroeconomics Across Time and Space" at the Philadelphia Fed, the Korean Economic Association's "International Employment Forum" in Seoul, the SED Annual Meeting, the NBER Summer Institute, the NBER Political Economy Group Meeting, the 4th Migration and Development Conference at Harvard, the 9th IZA Annual Migration Meeting, the MOVE Workshop on "Social Economics" in Barcelona, the 8th BETA Workshop in Historical Economics in Strasbourg, and the International Conference on "Intergenerational Transmission of Entrepreneurship, Occupation, and Cultural Traits in the Process of Long-Run Economic Growth" in Naples for helpful comments and suggestions. The authors also thank attendees of the Klein Lecture, the Kuznets Lecture, and the Nordic Doctoral Program, and they are especially indebted to Yona Rubinstein for numerous insightful discussions and to Sohini Ramachandran for sharing her data. Desislava Byanova and Daniel Doran provided excellent research assistance. Financial support from the Watson Institute at Brown University is gratefully acknowledged. Galor's research is supported by NSF grant SES-0921573.

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 have typically focused on the effects of such factors in giving rise to and sustaining the divergence in income per capita in the postindustrial era, attention has recently been drawn towards some deep-rooted factors that have been argued to affect the course of comparative economic development.

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

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 precolonial as well as in the modern era, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. While the intermediate level of genetic diversity prevalent among the Asian and European populations has been conducive for development, the high degree of diversity among African populations and the low degree of diversity among Native American populations have been a detrimental force in the development of these regions. This research thus highlights one of the deepest channels in comparative development, pertaining not to factors associated with the dawn of complex agricultural societies as in Diamond's (1997) influential hypothesis, but to conditions innately related to the very dawn of mankind itself.

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

Second, there exists an optimal level of diversity for economic development, reflecting the interplay between the 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 of an economy. Heterogeneity increases the likelihood of miscoordination and distrust, reducing cooper-

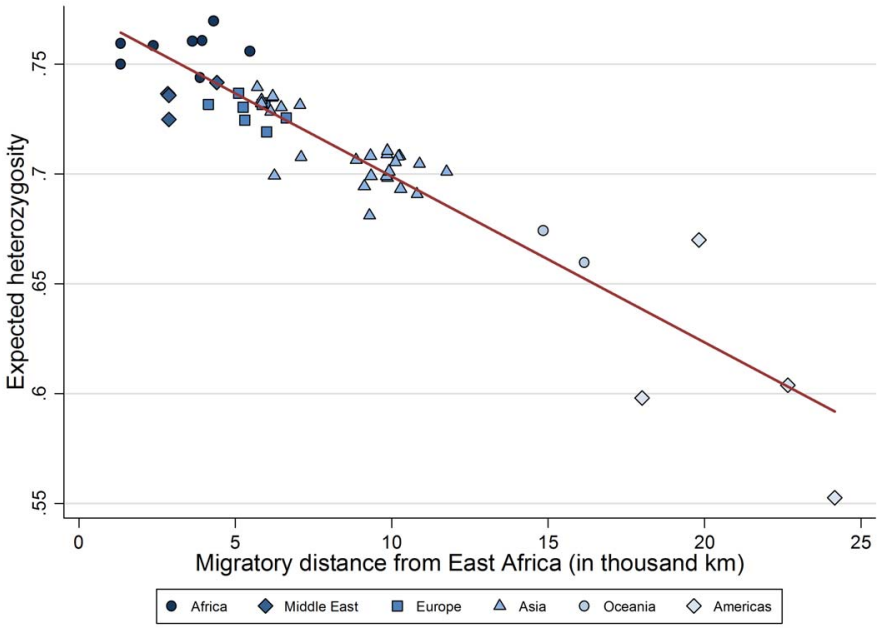


FIGURE 1: Expected Heterozygosity and Migratory Distance from East Africa

Notes: This figure depicts the negative impact of migratory distance from East Africa on expected heterozygosity (genetic diversity) across the 53 ethnic groups that constitute the HGDP-CEPH Human Genome Diversity Cell Line Panel.

ation and disrupting the socioeconomic order. Greater population diversity is therefore associated with lower total factor productivity, which inhibits the ability of society to operate efficiently with respect to its production possibility frontier. The beneficial effect of diversity, on the other hand, concerns the positive role of diversity in the expansion of society's production possibility frontier. A wider spectrum of traits is more likely to be complementary to the development and successful implementation of advanced technological paradigms.[1] Greater heterogeneity therefore fosters the ability of a society to incorporate more sophisticated and efficient modes of production, expanding the economy's production possibility frontier and conferring the benefits of increased total factor productivity.

Higher diversity in a society's population can therefore have conflicting effects on the level of its total factor productivity. Aggregate productivity is enhanced on the one hand by an increased capacity for technological advancement while diminished on the other by reduced cooperation and efficiency.[2] 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

[1]The following two mechanisms further illustrate this argument. First, in an economy where the labor force is characterized by genetic heterogeneity in a wide array of traits, to the extent that some of these traits lead to specialization in task-oriented activities, higher diversity will increase productivity for society as a whole, given complementarities across different tasks. Second, in an environment in which only individuals with sufficiently high levels of cognitive abilities can contribute to technological innovation, greater variance in the distribution of these traits across the population will lead to higher productivity.

[2]This hypothesis is consistent with evidence on the costs and benefits associated with intrapopulation heterogeneity, primarily in the context of ethnic diversity, as reviewed by Alesina and La Ferrara (2005).

returns to both diversity and homogeneity), the theory would predict an inverted-U relationship between genetic diversity and development outcomes throughout the development process.

The hypothesized channels through which genetic diversity affects aggregate productivity follow naturally from separate well-established mechanisms in the field of evolutionary biology and from experimental evidence from scientific studies on organisms that display a relatively high degree of social behavior in nature (e.g., living in task-directed hierarchical societies and engaging in cooperative rearing of offspring).[3] The benefits of genetic diversity, for instance, are highlighted in the Darwinian theory of evolution by natural selection, according to which diversity, by permitting the forces of natural selection to operate over a wider spectrum of traits, increases the adaptability and, hence, the survivability of a population to changing environmental conditions.[4] On the other hand, to the extent that genetic diversity is associated with a lower average degree of relatedness amongst individuals in a population, kin selection theory, which emphasizes that cooperation amongst genetically related individuals can indeed be collectively beneficial as it ultimately facilitates the propagation of shared genes to the next generation, is suggestive of the hypothesized mechanism through which diversity confers costs on aggregate productivity.

In estimating the impact on economic development of migratory distance from East Africa via its effect on genetic diversity, this research overcomes issues that are presented by the existing data on genetic diversity across the globe (i.e., measurement error and data limitations) as well as concerns about 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 with respect to a given spectrum of traits. 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. Given allelic frequencies for a particular gene or DNA locus, it is possible to compute a gene-specific heterozygosity statistic (i.e., the probability that two randomly selected individuals differ with respect to the gene in question), which when averaged over multiple genes or DNA loci yields the overall expected heterozygosity for the relevant population.

The most reliable and consistent data for genetic diversity among indigenous populations across the globe consists of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel. According to anthropologists, these groups are not only historically native to their current geographical locations but have also been isolated from genetic flows from other ethnic groups. Empirical evidence provided by population geneticists (e.g., Ramachandran et al., 2005) for these 53 ethnic groups suggest that, indeed, migratory distance from East Africa has an adverse linear effect on genetic diversity as depicted in Figure 1. Migratory distance from East Africa for each

[3] Section H of the appendix provides a detailed discussion of the evidence from evolutionary biology on the costs and benefits of genetic diversity.

[4] Moreover, according to a related hypothesis, genetically diverse honeybee colonies may operate more efficiently and productively, as a result of performing specialized tasks better as a collective, and thereby gain a fitness advantage over colonies with uniform gene pools (Robinson and Page, 1989).

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 prehistoric human migration patterns.

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

To tackle these difficulties, this study adopts two distinct strategies. The first restricts attention to development outcomes in the precolonial era when, arguably, regional populations were indigenous to their current geographical locations. Specifically, in light of the serial founder effect, the presence of multiple indigenous ethnicities in a given region would have had a negligible impact on the diversity of the regional population during this period. The second, more complex strategy involves the construction of an index of genetic diversity for contemporary national populations that accounts for the expected heterozygosity within each subnational group as well as the additional component of diversity at the country level that arises from the genetic distances between its precolonial ancestral populations. The examination of comparative development under this second strategy would have to additionally account for the potential inducement for members of distinct ethnic groups to relocate to relatively more lucrative geographical locations.

The examination of comparative development in the precolonial era, when societies were in their agricultural stage of development, requires the interpretation of outcomes from a Malthusian equilibrium point of view. Improvements in the technological environment during the Malthusian epoch generated only temporary gains in income per capita, eventually leading to a larger but not richer population (Ashraf and Galor, 2011). 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 year 1500 CE.

Using data on genetic diversity observed at the ethnic group level, the historical analysis reveals, consistently with the proposed hypothesis, a highly significant hump-shaped relationship between genetic diversity and 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 preindustrial 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 be associated with a 58 percent rise in its population density, whereas a 1 percentage point decrease in diversity for the most diverse society would be associated with a 23 percent rise in it its population density. Despite the statistical significance and robustness of these relationships, 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.[5]

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.[6] 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 prehistoric migration routes of the subsequent human expansion into the rest of the world. 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 prehistoric migratory paths from East Africa.[7] Indeed, several studies in population genetics (e.g.,

[5] The history of world civilization is abound with examples of both phenomena. The so-called 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.

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

[7] In addition, population geneticists argue that the reduced genetic diversity associated with the founder effect is due not only to the subset sampling of alleles from parental colonies but also to a stronger force of genetic drift that operates on the new colonies over time. Genetic drift arises from the fundamental tendency of the frequency of any allele in an inbreeding population to vary randomly across generations as a result of random statistical sampling errors

Prugnolle, Manica and Balloux, 2005; Ramachandran et al., 2005; Wang et al., 2007) have found strong empirical evidence in support of this prediction.

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

The main results from the historical analysis, employing predicted genetic diversity in the extended sample of countries, indicate that, controlling for the influence of land productivity, the timing of the Neolithic Revolution, and continent 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 percent, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 29 percent. Further, a 1 percentage point change in diversity in either direction at the predicted optimum of 0.683 would lower population density by 1.5 percent.[8]

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 precolonial ancestral populations of contemporary subnational groups, as predicted by the migratory distances of the ancestral populations from East Africa, but also incorporates the pairwise genetic distances between these ancestral populations, as predicted by their pairwise migratory distances. Indeed, the serial founder effect studied by population geneticists not only predicts that expected heterozygosity declines with increasing distance along migratory paths from East Africa but also that the genetic distance between any two populations will be larger the greater the migratory distance between them.

The baseline results from the contemporary analysis indicate that the genetic diversity of contemporary national populations has an economically and statistically significant hump-shaped effect

alone (i.e., the random production of a few more or less progeny carrying the relevant allele). Thus, given the inherent memoryless (Markovian) property of allelic frequencies across generations, the process ultimately leads, in the absence of mutation and natural selection, to either a 0 percent or a 100 percent representation of the allele in the population (Griffiths et al., 2000). Moreover, since random sampling errors are more prevalent in circumstances where the law of large numbers is less applicable, genetic drift is more pronounced in smaller populations, thereby allowing this phenomenon to play a significant role in the founder effect.

[8] Moreover, the partial R^2 associated with diversity suggests that residual genetic diversity explains roughly 7 percent of the cross-country variation in residual log population density in 1500 CE, conditional on land productivity, the timing of the Neolithic Revolution, and continent fixed effects. Consistent with the predictions of the proposed hypothesis, the robustness analysis in Section A of the appendix demonstrates that the nonmonotonic effect of genetic diversity on development outcomes is prevalent in earlier historical periods as well. Further, the impact of genetic diversity on economic development in the preindustrial era is robust to 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.

on income per capita. This hump-shaped impact is robust to controls for continent fixed effects, ethnic fractionalization, various measures of institutional quality (i.e., social infrastructure, an index gauging the extent of democracy, and constraints on the power of chief executives), legal origins, major religion shares, the share of the population of European descent, years of schooling, disease environments, and other geographical factors that have received attention in the empirical literature on cross-country comparative development.

The direct effect of genetic diversity on contemporary income per capita, once institutional, cultural, and geographical factors are accounted for, indicates that: (i) increasing the diversity of the most homogenous country in the sample (Bolivia) by 1 percentage point would raise its income per capita in the year 2000 CE by 41 percent, (ii) decreasing the diversity of the most diverse country in the sample (Ethiopia) by 1 percentage point would raise its income per capita by 21 percent, (iii) a 1 percentage point change in genetic diversity (in either direction) at the optimum level of 0.721 (that most closely resembles the diversity level of the U.S.) would lower income per capita by 1.9 percent, (iv) increasing Bolivia's diversity to the optimum level prevalent in the U.S. would increase Bolivia's per capita income by a factor of 5.4, closing the income gap between the U.S. and Bolivia from a ratio of 12:1 to 2.2:1, and (v) decreasing Ethiopia's diversity to the optimum level of the U.S. would increase Ethiopia's per capita income by a factor of 1.7 and thus close the income gap between the U.S. and Ethiopia from a ratio of 47:1 to 27:1. Moreover, the partial R^2 associated with diversity suggests that residual genetic diversity explains about 16 percent of the cross-country variation in residual log income per capita in 2000 CE, conditional on the institutional, cultural, and geographical covariates in the baseline regression model.

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 postcolonial migrations towards prosperous countries and the concomitant increase in ethnic diversity in these economies. The hump-shaped effect of genetic diversity remains highly significant and the optimal diversity estimate remains virtually intact if the regression sample is restricted to (i) non-OECD economies (i.e., economies that were less attractive to migrants), (ii) non-Neo-European countries (i.e., excluding the U.S., Canada, Australia, and New Zealand), (iii) non-Latin American countries, (iv) non-Sub-Saharan African countries, and, perhaps most importantly, (v) countries whose indigenous population is larger than 97 percent of the entire population (i.e., under conditions that virtually eliminate the role of migration in contributing to diversity). Moreover, consistently with the overall hump-shaped effect of diversity on the contemporary standard of living, the analysis indicates that genetic diversity is negatively associated with the extent of cooperative behavior, as measured by the prevalence of interpersonal trust, and positively associated with innovative activity, as measured by the intensity of scientific knowledge creation.

The remainder of the paper is organized as follows. Section 1 briefly reviews some related literature. Section 2 presents a basic model that predicts a hump-shaped effect of diversity on economic development. Sections 3 and 4 cover the historical analysis, discussing the empirical strategy as well as the relevant data and data sources before presenting the empirical findings. Sections 5 and 6 do the same for the contemporary analysis, and, finally, Section 7 concludes.

1 Related Literature

The existing literature on comparative development has emphasized a variety of factors underlying some of the vast differences in living standards across the globe. The influence of geography has been stressed from a historical perspective by Jones (1981), Diamond (1997), and Pomeranz (2000), and it has been highlighted empirically by Gallup, Sachs and Mellinger (1999) and Olsson and Hibbs (2005). Institutions, on the other hand, are given historical precedence by North and Thomas (1973), Mokyr (1990), and Greif (1993), and they have been emphasized empirically by Hall and Jones (1999), La Porta et al. (1999), Rodrik, Subramanian and Trebbi (2004), and Acemoglu, Johnson and Robinson (2005). In related strands of the literature on institutions, Engerman and Sokoloff (2000) and Acemoglu, Johnson and Robinson (2005) have stressed the role of colonialism, while the effects of ethnolinguistic fractionalization are examined by Easterly and Levine (1997) and Alesina et al. (2003). Moreover, the historical impact of sociocultural factors has been highlighted by Weber (1905) and Landes (1998), and their importance is supported empirically by Barro and McCleary (2003), Tabellini (2008), and Guiso, Sapienza and Zingales (2009). Finally, the importance of human capital formation has been underlined in unified growth theory (e.g., Galor, 2011) and has been demonstrated empirically by Glaeser et al. (2004).

This research is the first to argue that the variation in prehistoric migratory distance from the cradle of humankind to various settlements across the globe has had a persistent effect on the process of development and on the contemporary variation in income per capita across the globe. The paper is also unique in its attempt to establish the role of genetic (rather than ethnic) diversity within a society as a significant determinant of its development path and thus its comparative economic performance across space and time.

The employment of data and empirical results from the field of population genetics places this research in proximity to a recent insightful paper by Spolaore and Wacziarg (2009) who have appealed to data on genetic distance *between* human populations to proxy for the effect of sociocultural differences between societies on the diffusion of economic development.[9] Specifically, the authors argue that genetic distance between populations, which captures their divergence in biological and cultural characteristics over time, has been a barrier to the horizontal diffusion of technological innovations across populations. They show that F_{st} genetic distance, a measure that reflects the time elapsed since two populations shared a common ancestor, confers a statistically significant positive effect on both historical and contemporary pairwise income differences. In contrast, the genetic diversity metric *within* populations exploited by this paper facilitates the analysis of the effect of the variation in traits across individuals within a society on its development process.

Unlike Spolaore and Wacziarg (2009) where genetic distance between populations diminishes the rate of technological diffusion and reduces productivity, the hypothesis advanced and tested by the current analysis suggests that genetic diversity within a population confers both social costs, in the form of miscoordination and distrust arising from genetic differences across members of society, and social benefits in the form of diversity-driven knowledge accumulation. Hence, the overall effect

[9] See also Desmet et al. (2011) 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.

of genetic diversity on developmental outcomes would be hump shaped, rather than monotonically negative. Indeed, the results of the empirical analysis conducted in this study suggest that the previously unexamined beneficial effect of genetic differences is 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.[10] Diamond (1997) has stressed the role of biogeographic factors in determining the timing of the Neolithic Revolution, which conferred a developmental head start to societies that experienced an earlier transition from primitive hunting and gathering techniques to the more technologically advanced agricultural mode of production. According to this hypothesis, the luck of being dealt a favorable hand thousands of years ago with respect to biogeographic endowments, particularly exogenous factors contributing to the emergence of agriculture and facilitating the subsequent diffusion of agricultural techniques, is the single most important driving force behind the divergent development paths of societies throughout history that ultimately led to the contemporary global differences in standards of living. Specifically, an earlier transition to agriculture due to favorable environmental conditions gave some societies an early advantage by conferring the benefits of a production technology that generated resource surpluses and enabled the rise of a non-food-producing class whose members were crucial for the development of written language and science and for the formation of cities, technology-based military powers, and nation states. The early technological dominance of these societies subsequently persisted throughout history, being further sustained by the subjugation of less-developed societies through exploitative geopolitical and historical processes such as colonization.

While the long-standing influence of the Neolithic Revolution on comparative development in the precolonial era remains a compelling argument, this research demonstrates that, contrary to Diamond's (1997) unicausal hypothesis, the composition of human populations with respect to their genetic diversity has been a significant and persistent factor that affected the course of economic development from the dawn of human civilization to the present. Moreover, 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 precolonial population density, which, as already argued, is the relevant variable capturing economic development during the Malthusian epoch of stagnation in income per capita.[11]

[10] See, for example, Olsson and Hibbs (2005) and Putterman (2008).

[11] Note that, although the genetic diversity channel raised in this study is conceptually independent of the timing of the agricultural transition, an additional genetic channel that interacts with the time elapsed since the Neolithic Revolution has been examined by Galor and Moav (2002; 2007). These studies argue that the Neolithic transition triggered an evolutionary process resulting in the natural selection of certain genetic traits (such as preference for higher quality children and greater longevity) that are complementary to economic development, thereby implying a *ceteris paribus* positive relationship between the timing of the agricultural transition and the representation of such traits in the population. Indeed, the empirical evidence recently uncovered by Galor and Moav (2007) is consistent with this theoretical prediction. Thus, while the significant reduced-form effect of the Neolithic Revolution observed in this study may be associated with the Diamond hypothesis, it could also be partly capturing the influence of this additional genetic channel. See also Lagerlöf (2007), Dalgaard and Strulik (2010), and Galor and Michalopoulos (2011)

2 Diversity and Productivity: A Basic Model

Consider an economy where the level of productivity is affected by the degree of genetic diversity in society. Specifically, genetic diversity generates conflicting effects on productivity. A wider spectrum of traits is complementary to the adoption or implementation of new technologies. It enhances knowledge creation and fosters technological progress, thereby expanding the economy's production possibility frontier. However, a wider spectrum of traits also reduces the likelihood of cooperative or trustful behavior, generating inefficiencies in the operation of the economy relative to its production possibility frontier.

Suppose that the degree of genetic diversity, $\omega \in [0,1]$, has a positive but diminishing effect on the level of technology that is available for production. Specifically, the level of technology, A, and thus the economy's production possibility frontier, is determined by a vector of institutional, geographical, and human capital factors, $\mathbf{z}$, as well as by the degree of diversity, ω.[12] That is,

$$A = A(\mathbf{z}, \omega), \tag{1}$$

where $A(\mathbf{z},\omega) > 0$, $A_\omega(\mathbf{z},\omega) > 0$, and $A_{\omega\omega}(\mathbf{z},\omega) < 0$ for all $\omega \in [0,1]$, and the marginal effect of diversity on the level of technology satisfies the boundary conditions $\lim_{\omega \longrightarrow 0} A_\omega(\mathbf{z},\omega) = \infty$ and $\lim_{\omega \longrightarrow 1} A_\omega(\mathbf{z},\omega) = 0$.

Suppose further that the position of the economy relative to its production possibility frontier is adversely affected by the degree of genetic diversity. In particular, a fraction, $\alpha\omega$, of the economy's potential productivity, $A(\mathbf{z},\omega)$, is lost due to lack of cooperation and resultant inefficiencies in the production process.

Output per worker is therefore determined by the level of employment of factors of production, $\mathbf{x}$, the level of productivity, $A(\mathbf{z},\omega)$, and the degree of inefficiency in production, $\alpha \in (0,1)$.

$$y = (1 - \alpha\omega)A(\mathbf{z},\omega)f(\mathbf{x}) \equiv y(\mathbf{x},\mathbf{z},\omega), \tag{2}$$

where $\mathbf{x}$ is a vector of factor inputs per worker and $\alpha\omega$ is the extent of erosion in productivity due to inefficiencies in the production process.[13] Hence, as follows from (2), $y(\mathbf{x},\mathbf{z},\omega)$ is a strictly concave

for complementary evolutionary theories regarding the dynamics of human body size and entrepreneurial spirit in the process of economic development.

[12] Several mechanisms could generate this reduced-form relationship. Suppose that the labor force is characterized by heterogeneity in equally productive traits, each of which permit individuals to perform complementary specialized tasks. The quantity of trait i in the population is x_i, and it is distributed uniformly over the interval $[0,\omega]$. The level of productivity is therefore,

$$A(\mathbf{z},\omega) = \mathbf{z}\int_0^\omega x_i^\theta di; \qquad \theta \in (0,1).$$

Hence, an increase in the spectrum of traits, ω, (holding the aggregate supply of productive traits constant) will increase productivity at a diminishing rate. Alternatively, if there exists a hierarchy of traits and only traits above the cutoff $\xi \in (0,\omega)$ contribute to productivity, then an increase in the spectrum of traits, ω, could increase productivity at a diminishing rate.

[13] If degree of inefficiency is $\alpha(\omega)$, the results of the model would remain intact as long as the contribution of homogeneity for efficiency is diminishing (i.e., as long as $\alpha(\omega)$ is nondecreasing and weakly convex in ω).

hump-shaped function of ω. Specifically,

$$y_\omega(\mathbf{x},\mathbf{z},\omega) = [(1-\alpha\omega)A_\omega(\mathbf{z},\omega) - \alpha A(\mathbf{z},\omega)]f(\mathbf{x});$$

$$y_{\omega\omega}(\mathbf{x},\mathbf{z},\omega) = [(1-\alpha\omega)A_{\omega\omega}(\mathbf{z},\omega) - 2\alpha A_\omega(\mathbf{z},\omega)]f(\mathbf{x}) < 0; \tag{3}$$

$$\lim_{\omega\longrightarrow 0} y_\omega(\mathbf{x},\mathbf{z},\omega) > 0; \quad \text{and} \quad \lim_{\omega\longrightarrow 1} y_\omega(\mathbf{x},\mathbf{z},\omega) < 0.$$

Thus, there exists an intermediate level of diversity, $\omega^* \in (0,1)$, that maximizes the level of output per worker. In particular, ω^* satisfies

$$(1-\alpha\omega^*)A_\omega(\mathbf{z},\omega^*) = \alpha A(\mathbf{z},\omega^*). \tag{4}$$

3 The Historical Analysis: Data and Empirical Strategy

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

3.1 Dependent Variable: Historical Population Density

As argued previously, the relevant variable reflecting comparative development across countries in the precolonial Malthusian era is population density. The empirical examination of the proposed genetic hypothesis therefore aims to employ cross-country variation in observed genetic diversity and in that predicted by migratory distance from East Africa to explain cross-country variation in historical population density.[14] Data on historical population density are obtained from McEvedy and Jones (1978) who provide figures at the country level, i.e., for regions defined by contemporary national borders, over the period 400 BCE–1975 CE.[15] However, given the greater unreliability (and less availability in terms of observations) of population data for earlier historical periods, the baseline regression specification adopts population density in 1500 CE as the preferred outcome variable to examine. The analysis in Section A of the appendix additionally examines population density in 1000 CE and 1 CE to demonstrate the robustness of the genetic channel for earlier time periods.

3.2 Independent Variable: Genetic Diversity

The most reliable and consistent data for genetic diversity among indigenous populations across the globe consists of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel.[16] According to anthropologists, these 53 ethnic groups are not only historically native to

[14] Admittedly, historical data on population density is afflicted by measurement error. However, while measurement error in explanatory variables leads to attenuation bias in OLS estimators, mismeasurement of the dependent variable in an OLS regression, as a result of yielding larger standard errors for coefficient estimates, leads to rejecting the null when it is in fact true. As such, if OLS coefficients are precisely estimated, then confidence that the true coefficients are indeed different from zero rises even in the presence of measurement error in the dependent variable.

[15] The reader is referred to Section F of the appendix for additional details.

[16] For a more detailed description of the HGDP-CEPH Human Genome Diversity Cell Line Panel data set, the interested reader is referred to Cann et al. (2002). A broad overview of the HGDP is given by Cavalli-Sforza (2005). The 53 ethnic groups are listed in Section E of the appendix.

their current geographical locations but have also been isolated from genetic flows from other ethnic groups. Population geneticists typically measure the extent of diversity in genetic material across individuals within a given population (such as an ethnic group) using an index called expected heterozygosity. Like most other measures of diversity, this index may be interpreted simply as the probability that two individuals, selected at random from the relevant population, are genetically different from one another. Specifically, the expected heterozygosity measure for a given population is constructed by geneticists using sample data on allelic frequencies, i.e., the frequency with which a gene variant or allele occurs in the population sample. Given allelic frequencies for a particular gene or DNA locus, it is possible to compute a gene-specific heterozygosity statistic (i.e., the probability that two randomly selected individuals differ with respect to a given gene), which when averaged over multiple genes or DNA loci yields the overall expected heterozygosity for the relevant population.[17]

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

$$H^l_{\text{exp}} = 1 - \sum_{i=1}^{k} p_i^2. \tag{5}$$

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

$$H_{\text{exp}} = 1 - \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{k_l} p_i^2, \tag{6}$$

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

Empirical evidence uncovered by Ramachandran et al. (2005) for the 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel suggests that migratory distance from East Africa has an adverse linear effect on genetic diversity. They interpret this finding as providing support for a serial founder effect originating in East Africa, reflecting a process where the populating

[17] It should be noted that sources other than HGDP-CEPH exist for expected heterozygosity data. Specifically, the online Allele Frequency Database (ALFRED) represents one of the largest repositories of such data, pooled from across different data sets used by numerous studies in human population genetics. However, the data from ALFRED, while corresponding to a much larger sample of populations (ethnic groups) than the HGDP-CEPH sample, are problematic for a number of reasons. First, the expected heterozygosity data in ALFRED are not comparable across populations from the individual data sets in the collection because they are based on different DNA sampling methodologies (as dictated by the scientific goals of the different studies). Second, the vast majority of the individual data sets in ALFRED do not provide global coverage in terms of the different populations that are sampled and, even when they do, the sample size is considerably less than that of the HGDP-CEPH panel. Third, in comparison to the 783 loci employed by Ramachandran et al. (2005) to compute the expected heterozygosities for the 53 HGDP-CEPH populations, those reported for the non-HGDP populations in ALFRED are on average based on allelic frequencies for less than 20 DNA loci, which introduces a significant amount of potentially systematic noise in the heterozygosity estimates for these other populations. Fourth, unlike the microsatellite loci used by Ramachandran et al. (2005) for the HGDP-CEPH populations, the expected heterozygosities reported for many non-HGDP populations in ALFRED capture allelic variations across individuals in loci that reside in protein-coding regions of the human genome, thus reflecting diversity in phenotypic expressions that may have been subject to the environmental forces of natural selection. Finally, in contrast to the HGDP-CEPH populations, many of the non-HGDP populations in ALFRED represent ethnic groups that have experienced significant genetic admixture in their recent histories, particularly during the post-1500 era, and this introduces an endogeneity problem for the current analysis since genetic admixtures are, in part, the result of migrations occurring along spatial economic prosperity gradients.

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 prehistoric human migration patterns that are consistent with the widely-held hypothesis that, in the course of their exodus from Africa, humans did not cross large bodies of water. The intermediate waypoints, depicted on the world map in Figure 2 along with the spatial distribution of the ethnic groups from the HGDP-CEPH sample, are: Cairo (Egypt), Istanbul (Turkey), Phnom Penh (Cambodia), Anadyr (Russia), and Prince Rupert (Canada). For instance, as illustrated in Figure 2, the migration path from Addis Ababa to the Papuan ethnic group in modern-day New Guinea makes use of Cairo and Phnom Penh whereas that to the Karitiana population in Brazil incorporates Cairo, Anadyr, and Prince Rupert as intermediate waypoints.[18] The migratory distance between endpoints (i.e., Addis Ababa and the location of a group) is therefore the sum of (i) the great circle distances between these endpoints and the waypoint(s) in the path connecting them and (ii) the distance(s) between waypoints if two or more such points are required.

The empirical analysis of Ramachandran et al. (2005) establishes migratory distance from East Africa as a strong negative predictor of genetic diversity at the ethnic group level. Based on the R^2 of their regression, migratory distance alone explains almost 86 percent of the cross-group variation in within-group diversity.[19] In addition, the estimated OLS coefficient is highly statistically significant, possessing a t-statistic of -9.770 (P-value $< 10^{-4}$), and suggests that expected heterozygosity falls by 0.076 percentage points for every 10,000 km increase in migratory distance from East Africa.[20] This is the relationship depicted earlier 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

[18] Based on mitochondrial DNA analysis, some recent studies (e.g., 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 qualitatively alter the main results.

[19] These results are similar to those uncovered in an independent study by Prugnolle, Manica and Balloux (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, Manica and Balloux (2005) find that migratory distance from East Africa explains 85 percent 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 percent of the variation in genetic diversity based on allelic frequencies for 678 loci. The authors attribute their somewhat weaker results to the fact that the additional Native American ethnic groups in their augmented sample were historically subjected to a high degree of gene flow from foreign populations (i.e., European colonizers), which obscured the genetic legacy of a serial founder effect in these groups.

[20] This effect corresponds to roughly one-third of the full (worldwide) range of expected heterozygosity values observed across the HGDP-CEPH sample of ethnic groups.

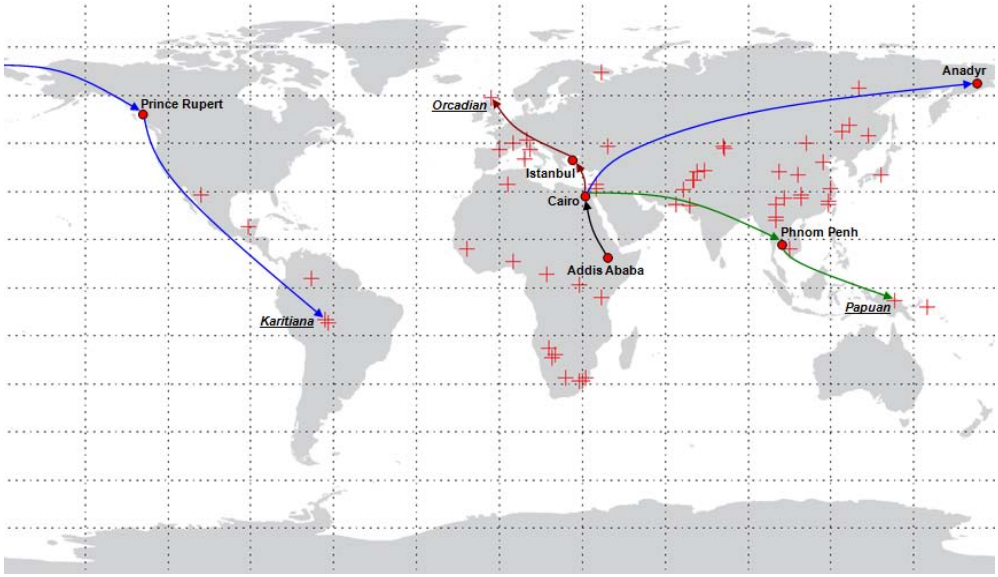


FIGURE 2: The 53 HGDP-CEPH Ethnic Groups and Migratory Paths from East Africa

Notes: This figure depicts on a world map (i) the locations (denoted by crosses) of the 53 HGDP-CEPH ethnic groups, (ii) the locations (denoted by circles) of the intermediate waypoints used to construct migratory paths from Addis Adaba to these ethnic groups, and (iii) some migratory paths (denoted by solid lines) based on these waypoints.

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

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

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

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 in the extended sample of countries.[23]

3.3 Control Variables: Neolithic Transition Timing and Land Productivity

Diamond's (1997) hypothesis has identified the timing of the Neolithic Revolution as a proximate determinant of economic development, designating initial geographic and biogeographic conditions that governed the emergence and adoption of agricultural practices in prehistoric 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 numbers of prehistoric plant and animal species amenable for domestication.

The current analysis controls for the ultimate and proximate determinants of development in the Diamond channel using cross-country data on the aforementioned geographic and biogeographic variables as well as on the timing of the Neolithic Revolution.[24] However, given the empirical link between the ultimate and proximate factors in Diamond's hypothesis, the baseline specification focuses on the timing of the Neolithic transition to agriculture as the relevant control variable for this channel.[25] The results from an extended specification that incorporates initial geographic

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

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

[24] The data source for the aforementioned geographic and biogeographic controls is Olsson and Hibbs (2005) whereas that for the timing of the Neolithic Revolution is Putterman (2008). See Section F of the appendix for the definitions and sources of all primary and control variables employed by the analysis.

[25] The Neolithic transition timing variable, employed throughout the current analysis, reflects the number of years elapsed, as of the year 2000 CE, since the onset of sedentary agriculture. Tables D.15 and D.16 in Section D of the appendix demonstrate that all the results of the historical analysis are qualitatively robust to the use of an alternative

and biogeographic factors as controls are presented in Section A of the appendix to demonstrate robustness.

The focus of the historical analysis on economic development in the precolonial Malthusian era also necessitates controls for the natural productivity of land for agriculture. Given that in a Malthusian environment resource surpluses are primarily channeled into population growth with per capita incomes largely remaining at or near subsistence, regions characterized by natural factors generating higher agricultural crop yields should, *ceteris paribus*, also exhibit higher population densities (Ashraf and Galor, 2011).[26] If diversity in a society influences its development through total factor productivity (comprised of both social capital and technological know-how), then controlling for the natural productivity of land would constitute a more accurate test of the effect of diversity on the Malthusian development outcome, i.e., population density.

In controlling for the agricultural productivity of land, this study employs measurements of three geographical variables at the country level: (i) the percentage of arable land, (ii) absolute latitude, and (iii) an index gauging the overall suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH.[27]

3.4 The Baseline Regression Specifications

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

$$\ln P_{it} = \beta_{0t} + \beta_{1t} G_i + \beta_{2t} G_i^2 + \beta_{3t} \ln T_i + \boldsymbol{\beta}_{4t}' \ln \mathbf{X}_i + \boldsymbol{\beta}_{5t}' \ln \boldsymbol{\Delta}_i + \varepsilon_{it}, \tag{7}$$

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, $\mathbf{X}_i$ is a vector of land productivity controls, $\boldsymbol{\Delta}_i$ is a vector of continent fixed effects, and ε_{it} is a country-year-specific disturbance term.[28]

definition of the Neolithic transition timing variable where this variable reflects the number of years elapsed, as of the year 1500 CE, since the onset of sedentary agriculture.

[26] 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 is thus somewhat orthogonal to the timing of agricultural transitions as well. This holds due to the independence of natural factors conducive to domesticated species from those that were beneficial for the wild ancestors of eventual domesticates. As argued by Diamond (2002), while agriculture originated in regions of the world to which the most valuable domesticable wild plant and animal species were native, other regions proved more fertile and climatically favorable once the diffusion of agricultural practices brought the domesticated varieties to them.

[27] The data for these variables are obtained from the World Bank's *World Development Indicators*, the CIA's *World Factbook*, and Michalopoulos (2011) respectively. The country-level aggregate data on the land suitability index from Michalopoulos (2011) are, in turn, based on more disaggregated geospatial data on this index from the ecological study of Ramankutty et al. (2002). See Section F of the appendix for additional details.

[28] The fact that economic development has been historically clustered in certain regions of the world raises concerns that these disturbances could be nonspherical in nature, thereby confounding statistical inferences based on the OLS estimator. In particular, the disturbance terms may exhibit spatial autocorrelation, i.e., $\text{cov}[\varepsilon_i, \varepsilon_j] > 0$, within a certain

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

$$\ln P_{it} = \beta_{0t} + \beta_{1t}\hat{G}_i + \beta_{2t}\hat{G}_i^2 + \beta_{3t}\ln T_i + \boldsymbol{\beta}'_{4t}\ln \mathbf{X}_i + \boldsymbol{\beta}'_{5t}\ln \boldsymbol{\Delta}_i + \varepsilon_{it}, \tag{8}$$

where $\hat{G}_i$ is the genetic diversity predicted by migratory distance from East Africa for country i using the methodology discussed in Section 3.2. Indeed, it is this regression specification that is estimated to obtain the main empirical findings.[29]

Before proceeding, it is important to note that the regression specifications in (7) and (8) 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 nonmonotonic impact on society's level of development through two opposing effects on the level of its total factor productivity: a detrimental effect on social capital and a beneficial effect on the knowledge frontier. However, given the absence of measurements for the proximate determinants of development in the genetic diversity channel, a more discriminatory test of the hypothesis is infeasible. Nonetheless, the results to follow are entirely consistent with the theoretical prediction that, in the presence of diminishing marginal effects of genetic diversity on total factor productivity in a Malthusian economy, the overall reduced-form effect of genetic diversity on cross-country population density should be hump shaped, i.e., that $\beta_{1t} > 0$ and $\beta_{2t} < 0$. Moreover, as will become evident, the unconditional hump-shaped relationship between genetic diversity and development outcomes does not differ significantly between the adopted quadratic and alternative nonparametric specifications.

4 The Historical Analysis: Empirical Findings

This section presents the results from empirically investigating the relationship between genetic diversity and log population density in the precolonial Malthusian era. Results for observed diversity in the limited 21-country sample are examined in Section 4.1. Section 4.2 discusses the baseline results associated with examining the effect of predicted diversity on log population density in 1500 CE in the extended sample of 145 countries. The robustness of the diversity channel with respect to 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, are presented in Section 4.3.

threshold of distance from each observation. Keeping this possibility in mind, the limited-sample analyses presented in the text are repeated in Tables D2–D3 in Section D of the appendix, where the standard errors of the point estimates are corrected for spatial autocorrelation across disturbance terms, following the methodology of Conley (1999).

[29] Tables G1–G2 in Section G of the appendix present the descriptive statistics of the limited 21-country sample employed in estimating equation (7), while Tables G3–G4 present those of the extended 145-country sample used to estimate equation (8). As reported therein, the finite-sample moments of the explanatory variables in the limited and extended cross-country samples are remarkably similar. Specifically, the range of values for predicted genetic diversity in the extended sample falls within the range of values for observed diversity in the limited sample. This is particularly reassuring because it demonstrates that the methodology used to generate the predicted genetic diversity variable did not project values beyond what is actually observed, indicating that the HGDP-CEPH collection of ethnic groups is indeed a representative sample for the worldwide variation in within-country genetic diversity. Moreover, the fact that the finite-sample moments of log population density in 1500 CE are not significantly different between the limited and extended cross-country samples foreshadows the encouraging similarity of the regression results that are obtained under observed and predicted values of genetic diversity.

The analysis of comparative development in the precolonial era is expanded upon in Section A of the appendix to demonstrate the robustness of the diversity channel with respect to (i) explaining comparative development in earlier historical periods, specifically log population density in 1000 CE and 1 CE, (ii) the technology diffusion hypothesis that postulates a beneficial effect on development arising from spatial proximity to regional technological frontiers, (iii) controls for microgeographic factors including the degree of variation in terrain and access to waterways, and, finally, (iv) controls for the exogenous geographic and biogeographic factors favoring an earlier onset of agriculture in the Diamond channel.

4.1 Results from the Limited Sample

The initial investigation of the proposed genetic diversity hypothesis using the limited sample of countries is of fundamental importance for the subsequent empirical analyses, performed using the extended sample, in three critical dimensions. First, since the limited sample contains observed values of genetic diversity whereas the extended sample comprises values predicted by migratory distance from East Africa, similarity in the results obtained from the two samples would lend credence to the main empirical findings associated with predicted genetic diversity in the extended sample of countries. Second, the fact that migratory distance from East Africa and observed genetic diversity are not perfectly correlated with each other makes it possible to test, using the limited sample of countries, the assertion that migratory distance affects economic development through genetic diversity only and is therefore appropriate for generating predicted genetic diversity in the extended sample of countries.[30] Finally, having verified the above assertion, the limited sample permits an instrumental variables regression analysis of the proposed hypothesis with migratory distance employed as an instrument for genetic diversity. This then constitutes a more direct and accurate test of the genetic diversity channel given possible concerns regarding the endogeneity between genetic diversity and economic development. As will become evident, the results obtained from the limited sample are reassuring on all three aforementioned fronts.

4.1.1 Explaining Comparative Development in 1500 CE

Table 1 presents the limited-sample results from regressions explaining log population density in 1500 CE.[31] In particular, a number of specifications comprising different subsets of the explanatory variables in equation (7) are estimated to examine the independent and combined effects of the genetic diversity, transition timing, and land productivity channels.

[30] The fact that migratory distance from East Africa may be correlated with other potential geographical determinants of genetic diversity, particularly factors like the dispersion of land suitability for agriculture and the dispersion of elevation that have been shown to give rise to ethnic diversity (Michalopoulos, 2011), raises the possibility that migratory distance may not be the only source of exogenous variation in genetic diversity. However, Table D1 in Section D of the appendix indicates that these other factors have little or no explanatory power for the cross-country variation in actual genetic diversity beyond that accounted for by migratory distance via the serial founder effect. Specifically, the OLS coefficient as well as the partial R^2 associated with migratory distance remain both quantitatively and qualitatively robust when the regression is augmented with these geographical controls, all of which are statistically insignificant in explaining genetic diversity. The reader is referred to Section F of the appendix for detailed definitions of the additional control variables used by the analysis in Table D1.

[31] Corresponding to Tables 1 and 2 in the text, Tables D2 and D3 in Section D of the appendix present results with standard errors and 2SLS point estimates corrected for spatial autocorrelation across observations.

TABLE 1: Observed Diversity and Economic Development in 1500 CE

	(1)	(2)	(3)	(4)	(5)
	Dependent variable is log population density in 1500 CE				
Observed diversity	413.504***			225.440***	203.814*
	(97.320)			(73.781)	(97.637)
Observed diversity square	-302.647***			-161.158**	-145.717*
	(73.344)			(56.155)	(80.414)
Log Neolithic transition timing		2.396***		1.214***	1.135
		(0.272)		(0.373)	(0.658)
Log percentage of arable land			0.730**	0.516***	0.545*
			(0.281)	(0.165)	(0.262)
Log absolute latitude			0.145	-0.162	-0.129
			(0.178)	(0.130)	(0.174)
Log land suitability for agriculture			0.734*	0.571*	0.587
			(0.381)	(0.294)	(0.328)
Optimal diversity	0.683***			0.699***	0.699***
	(0.008)			(0.015)	(0.055)
Continent fixed effects	No	No	No	No	Yes
Observations	21	21	21	21	21
R^2	0.42	0.54	0.57	0.89	0.90

Notes: This table establishes the significant hump-shaped relationship between observed genetic diversity and log population density in 1500 CE in the limited 21-country sample while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

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 percent level, imply that a 1 percentage point increase in genetic diversity for the most homogenous society in the regression sample would be associated with a rise in its population density in 1500 CE by 114 percent, whereas a 1 percentage point decrease in diversity for the most diverse society would be associated with a rise in its population density by 64 percent. In addition, the coefficients also indicate that a 1 percentage point change in diversity in either direction at the predicted optimum of 0.683 would be associated with a decline in population density by 3 percent.[32] Furthermore, based on the R^2 coefficient of the regression, the genetic diversity channel appears to explain 42 percent of the variation in log population density in 1500 CE across the limited sample of countries. The quadratic relationship implied by the OLS coefficients reported in Column 1 is depicted together with a nonparametric local polynomial regression line in Figure 3.[33] Reassuringly,

[32] The magnitude of these effects can be derived directly from the estimated linear and quadratic coefficients associated with genetic diversity. Specifically, letting $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the estimated coefficients on genetic diversity and genetic diversity square, equation (7) 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: $\Delta P/P = \exp\{\Delta G(\hat{\beta}_1 + 2\hat{\beta}_2\bar{G} + \bar{G}\Delta G)\} - 1$.

[33] For consistency with Figure 1, which depicts the negative effect of increasing migratory distance from East Africa on genetic diversity, the horizontal axes in Figures 3–5 represent genetic homogeneity (i.e., 1 minus genetic diversity) so as to reflect increasing as opposed to decreasing migratory distance from East Africa.

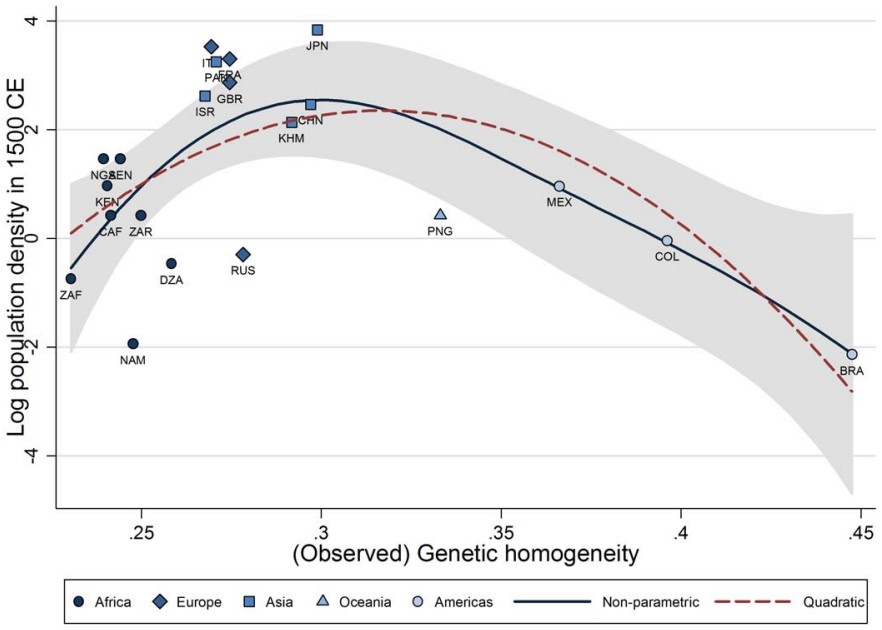


FIGURE 3: Observed Genetic Diversity and Population Density in 1500 CE

Notes: This figure depicts the unconditional hump-shaped relationship, estimated using either a least-squares quadratic fit or a nonparametric regression, between observed genetic homogeneity (i.e., 1 minus observed genetic diversity) and log population density in 1500 CE in the limited 21-country sample. The nonparametric regression line is estimated using local 2nd-degree polynomial smoothing based on a Gaussian kernel function and a kernel bandwidth of 0.06. The shaded area represents the 95 percent confidence interval band associated with the nonparametric regression line.

as illustrated therein, the estimated quadratic falls within the 95 percent confidence interval band of the nonparametric relationship.[34]

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 percent increase in the number of years elapsed since the transition to agriculture increases population density in 1500 CE by 2.4 percent, an effect that is also significant at the 1 percent level. Similarly, consistent with the predictions of the land productivity channel, population density in 1500 CE possesses statistically significant positive elasticities with respect to both the percentage of arable land as well as the index gauging the suitability of land for agriculture. Moreover, the agricultural transition timing and land productivity channels independently explain 54 percent and 57 percent of the limited cross-country sample variation in log population density in 1500 CE.

Column 4 presents the results obtained from exploiting the combined explanatory power of all three channels for log population density in the year 1500 CE. Not surprisingly, given the small sample size as well as the pairwise correlations between covariates reported in Table G2 in Section G

[34] Correspondingly with Figure 3, Figure C1 in Section C of the appendix compares the quadratic fit with a restricted cubic spline (as opposed to nonparametric) regression. The figure indicates that the quadratic fit falls within the 95 percent confidence interval band of the cubic spline regression in much the same way as it does with respect to the nonparametric relationship. The notes to Figures 3 and C1 provide additional details on the estimation procedures underlying the nonparametric and cubic spline regressions respectively.

of the appendix, the estimated conditional effects are sizeably reduced in magnitude in comparison to their unconditional estimates presented in earlier columns. Nonetheless, the OLS coefficients associated with all channels retain their expected signs and continue to remain highly statistically significant. To interpret the conditional effects of the genetic diversity channel, the estimated linear and quadratic coefficients associated with genetic diversity imply that, accounting for the influence of the transition timing and land productivity channels, a 1 percentage point increase in genetic diversity for the most homogenous society in the regression sample would be associated with a rise in its population density in 1500 CE by 58 percent, whereas a 1 percentage point decrease in diversity for the most diverse society would be associated with a rise in its population density by 23 percent. Further, a 1 percentage point change in diversity in either direction at the predicted optimum of 0.699 would be associated with a decline in population density by 1.6 percent. Additionally, by exploiting the combined explanatory power of all three channels, the estimated model explains an impressive 89 percent of the limited-sample cross-country variation in log population density.

Finally, the results from estimating the regression model in equation (7) are reported in Column 5, which indicates that the results from previous columns were not simply reflecting the possible influence of some unobserved continent-specific attributes. In spite of the sample size limitations and the smaller variability of covariates within continents in comparison to that across continents, genetic diversity continues to possess a significant hump-shaped relationship with economic development in a manner consistent with theoretical predictions. Reassuringly, the estimated average within-continent relationship of diversity with log population density in 1500 CE is very similar to the cross-continental relationship reported in Column 4, and the implied optimal level of diversity remains intact, lending credence to the assertion that the hump-shaped relationship between diversity and development is not reflective of unobserved continental characteristics.[35]

To summarize, the limited-sample results presented in Table 1 demonstrate that genetic diversity has a statistically significant hump-shaped relationship with log population density in the year 1500 CE. The analysis, however, also reveals significant effects associated with the Neolithic transition timing and land productivity channels. Indeed, the nonmonotonic relationship of diversity with log population density prevails under controls for these other explanatory channels, and it 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 4.2 below. This similarity provides further assurance regarding the validity of the inferences

[35] Despite controls for continent fixed effects, the fact that (i) the hump-shaped relationship between genetic diversity and economic development appears to be, in part, identified by a relatively smaller number of observations from the Americas on the downward-sloping side of the relationship, coupled with the fact that (ii) equations (7)–(9) a priori impose a quadratic relationship between genetic diversity and economic development, could potentially raise concerns that the empirical models being estimated in this paper are misspecified in that the true relationship between diversity and development is logarithmic rather than quadratic in nature. If the relationship is indeed logarithmic then, upon re-estimating the baseline specifications using logged diversity and the square of logged diversity, one should not expect the latter quadratic term to survive in the regressions. Table D4 in Section D of the appendix presents the results from such an analysis, demonstrating that empirical model misspecification need not be a source of concern. In particular, the results indicate that the baseline findings from both the limited- and extended-sample variants of the historical analysis, as well as those from the contemporary analysis, are qualitatively unaltered when quadratic specifications using logged genetic diversity are employed to examine the impact of diversity on development.

made with the main empirical findings that are associated with predicted as opposed to observed values of genetic diversity.

4.1.2 Establishing the Exogeneity of Migratory Distance

As already mentioned, the fact that the limited cross-country sample comprises observed genetic diversity, which is strongly but not perfectly correlated with migratory distance from East Africa, permits a formal examination of whether migratory distance influences population density solely via the serial founder effect on genetic diversity. This is a particularly important test since, if migratory distance from East Africa actually affects economic development either directly or via some other unobserved channels, then the main empirical analysis conducted using predicted values of diversity would be attributing this latent influence to the genetic diversity channel.[36]

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

The results of the "horse race" regression in Column 2 are perhaps even more striking given the prior that genetic diversity, as opposed to migratory distance, is likely to be afflicted by larger measurement errors. Nevertheless, since migratory distance is measured as the sum of aerial distances between intercontinental waypoints, it may also be viewed as a noisy proxy of the distance along actual migration routes taken by prehistoric humans during their exodus out of Africa. In order to test whether genetic diversity survives a horse race with a less noisy measure of migratory distance from East Africa, Columns 3–4 repeat the preceding analysis using migratory distance based on the index of human mobility employed previously by Ashraf, Galor and Özak (2010). This index captures the average distance from Addis Ababa to the HGDP ethnic groups located within a given country, along "optimal" land-restricted routes that minimize the time cost of movement on the surface of the Earth

[36] Figures C6(a)–C6(c) in Section C of the appendix illustrate that, unlike the significant impact of migratory distance from East Africa on genetic diversity, migratory distance has no systematic relationship with a number of observed physiological characteristics of populations, including average skin reflectance, average height, and average weight, conditional on geographical factors such as the intensity of ultraviolet exposure, absolute latitude, the percentage of arable land, the shares of land in tropical and temperate zones, elevation, access to waterways, and continent fixed effects. Since the physiological characteristics examined in Figures C6(a)–C6(c) represent their averages for contemporary national populations, the migratory distance measure is adjusted to account for the modern ethnic compositions of these populations resulting from cross-country migrations in the post-1500 era.

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

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	2SLS (5)	2SLS (6)
	Dependent variable is log population density in 1500 CE					
Observed diversity		255.219**		361.421**	285.190***	243.110***
		(100.586)		(121.429)	(88.064)	(63.880)
Observed diversity square		-209.808**		-268.514***	-206.576***	-179.581***
		(73.814)		(87.342)	(66.852)	(52.038)
Migratory distance	0.505***	0.070				
	(0.148)	(0.184)				
Migratory distance square	-0.023***	-0.014				
	(0.006)	(0.009)				
Mobility index			0.353**	0.051		
			(0.127)	(0.154)		
Mobility index square			-0.012***	-0.003		
			(0.004)	(0.006)		
Log Neolithic transition timing					1.014***	1.119**
					(0.361)	(0.487)
Log percentage of arable land					0.608***	0.634***
					(0.188)	(0.211)
Log absolute latitude					-0.209*	-0.133
					(0.121)	(0.127)
Log land suitability for agriculture					0.494**	0.549**
					(0.233)	(0.253)
Continent fixed effects	No	No	No	No	No	Yes
Observations	21	21	18	18	21	21
R^2	0.34	0.46	0.30	0.43	–	–
P-value for joint significance of linear and quadratic terms in:						
Observed diversity		0.023		0.027		
Migratory distance		0.235				
Mobility index				0.905		

Notes: This table (i) performs an informal identification test, which demonstrates that the significant unconditional hump-shaped impact of migratory distance from East Africa on log population density in 1500 CE in the limited 21-country sample almost entirely reflects the latent influence of observed genetic diversity, and (ii) employs migratory distance as an excluded instrument for observed genetic diversity to establish the causal hump-shaped effect of genetic diversity on log population density in 1500 CE in the limited 21-country sample while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects. For the 2SLS regressions in Columns 5–6, since the second stage is quadratic in the endogenous regressor, it is necessary to instrument for both genetic diversity *and its squared term* in order for the system to be exactly identified. Thus, following Wooldridge (2010, pp. 267–268), the instrumentation technique introduces a zeroth stage to the analysis where genetic diversity is first regressed on migratory distance and all the second-stage controls to obtain predicted (i.e., fitted) values of diversity. The *predicted* genetic diversity from the zeroth stage is squared, and this squared term is then used as an excluded instrument in the second stage along with migratory distance. The zeroth- and first-stage results of these 2SLS regressions are collected in Table D6 in Section D of the appendix. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

in the absence of steam-powered transportation technologies. The index thus accounts for natural impediments to human mobility, including various meteorological and topographical conditions, and incorporates information on the time cost of travelling under such conditions. Reassuringly, as revealed in Columns 3–4, while distance from East Africa based on the mobility index possesses a significant hump-shaped correlation with log population density, this unconditional relationship virtually disappears once genetic diversity is accounted for by the analysis, lending further support to the claim that distance along prehistoric human migration routes from East Africa confers an effect on development outcomes through genetic diversity alone.[37]

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 instrumented by migratory distance. The results from a similar analysis that also accounts for continent fixed effects are reported in Column 6. Interestingly, in comparison to their OLS counterparts in Table 1, the estimated 2SLS coefficients associated with the diversity channel remain relatively stable in magnitude, suggesting that the potential endogeneity between genetic diversity and economic development need not be a source of concern, conditional on controls for the transition timing and land productivity channels. Overall, the results uncovered here provide support for the inferences made with predicted genetic diversity in the main empirical analysis to follow.

4.2 The Baseline Results from the Extended Sample

This section establishes the hump-shaped impact of genetic diversity, predicted by migratory distance from East Africa, on log population density in 1500 CE, using the extended sample of 145 countries. To reveal the independent and combined effects of the genetic diversity, transition timing, and land productivity channels, Table 3 presents the results from estimating a number of specifications spanning relevant subsets of the explanatory variables in equation (8).

The unconditional hump-shaped relationship between genetic diversity and log population density in 1500 CE is reported in Column 1.[38] In particular, the estimated linear and quadratic coefficients, both statistically significant at the 1 percent 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 59 percent, whereas a 1 percentage point decrease in genetic diversity for the most diverse

[37] The difference in the number of observations between Columns 1–2 (21 obs.) and Columns 3–4 (18 obs.) arises due to the fact that the mobility index cannot be calculated for countries that can only be accessed from Addis Ababa by crossing at least one body of water. Restricting the sample used in Columns 1–2 to that in Columns 3–4 does not qualitatively alter the findings. In addition, the unavailability of the mobility index measure for several countries (due to the aforementioned strict land-accessibility constraint) makes this measure less suitable, in comparison to the baseline migratory distance measure of Ramachandran et al. (2005), to predict genetic diversity in the extended cross-country sample. Nevertheless, Table D5 in Section D of the appendix demonstrates that the main findings from both the extended-sample historical analysis and the contemporary analysis remain qualitatively robust to using genetic diversity predicted by the more sophisticated mobility index, rather than by the baseline waypoints-restricted migratory distance measure of Ramachandran et al. (2005).

[38] This quadratic relationship is depicted in Figures C2(a) and C2(b) in Section C of the appendix along with a nonparametric local polynomial regression line and a restricted cubic spline regression line respectively. As in the limited-sample historical analysis, the estimated quadratic falls within the 95 percent confidence interval bands of both the nonparametric and cubic spline relationships. The figure notes provide additional details on the estimation procedures underlying the nonparametric and cubic spline regressions.

TABLE 3: Predicted Diversity and Economic Development in 1500 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is log population density in 1500 CE					
Predicted diversity	250.986***		213.537***	203.017***	195.416***	199.727**
	(66.314)		(61.739)	(60.085)	(55.916)	(80.281)
Predicted diversity square	-177.399***		-152.107***	-141.980***	-137.977***	-146.167***
	(48.847)		(45.414)	(44.157)	(40.773)	(56.251)
Log Neolithic transition timing		1.287***	1.047***		1.160***	1.235***
		(0.170)	(0.188)		(0.143)	(0.243)
Log percentage of arable land				0.523***	0.401***	0.393***
				(0.117)	(0.096)	(0.103)
Log absolute latitude				-0.167*	-0.342***	-0.417***
				(0.093)	(0.096)	(0.124)
Log land suitability for agriculture				0.189	0.305***	0.257***
				(0.124)	(0.094)	(0.096)
Optimal diversity	0.707***		0.702***	0.715***	0.708***	0.683***
	(0.021)		(0.025)	(0.110)	(0.051)	(0.110)
Continent fixed effects	No	No	No	No	No	Yes
Observations	145	145	145	145	145	145
R^2	0.22	0.26	0.38	0.50	0.67	0.69

Notes: This table establishes the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log population density in 1500 CE in the extended 145-country sample while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

society would raise its population density by 24 percent.[39] Further, population density in 1500 CE is unconditionally predicted by the regression to be maximized at an expected heterozygosity value of about 0.707, which roughly corresponds to the diversity predicted (by migratory distance from East Africa) for southern China. Indeed, a 1 percentage point change in genetic diversity in either direction at the predicted optimum lowers population density by 1.8 percent. Moreover, based on the R^2 of the regression, the cross-country variation in genetic diversity alone explains 22 percent 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 percent increase in the number of years elapsed since the Neolithic transition to agriculture is associated with a 1.3 percent increase in population density, an effect that is also statistically significant at the 1 percent level. Furthermore, 26 percent 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 G4 in Section G of the appendix, the unconditional effects of

[39] Following the earlier discussion regarding the expected heterozygosity index, these effects are therefore associated with a 0.01 change in the probability that two randomly selected individuals from a given population are genetically different from one another. See Footnote 32 for details on how these effects may be computed based on the estimated linear and quadratic coefficients associated with genetic diversity.

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 percent level with the combined cross-country variation in genetic diversity and transition timing explaining 38 percent 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. 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 in Column 1. More importantly, the coefficients remain highly statistically significant and also rather stable in magnitude relative to those estimated while controlling for the timing of the Neolithic transition. In addition, the overall significance of the land productivity channel is also confirmed, particularly by the estimated coefficients on the log percentage of arable land and log absolute latitude variables, which indeed appear to possess their expected signs.[40] Nonetheless, these estimates continue to reflect some amount of omitted variable bias resulting from the exclusion of the transition timing channel. For instance, the fact that log agricultural transition timing has a sample correlation of 0.28 with genetic diversity and one of 0.32 with log absolute latitude implies that the estimated effects of these variables on log population density in Column 4 may be partially capturing the latent influence of the excluded Neolithic transition timing channel.

Column 5 presents the results from exploiting the explanatory power of all three identified channels for log population density in 1500 CE. In line with the theoretical predictions of each hypothesis, the coefficient estimates possess their expected signs and are all statistically significant at the 1 percent level. Moreover, in comparison to their estimates in Columns 3 and 4, the linear and quadratic coefficients associated with the diversity channel remain largely stable. In particular, the estimated coefficients of interest imply that, controlling for the influence of land productivity and the timing of the Neolithic Revolution, a 1 percentage point increase in genetic diversity for the least diverse society in the sample would raise its population density in 1500 CE by 43 percent, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 18 percent. Further, population density in 1500 CE is predicted to be maximized at an expected heterozygosity value of 0.708, where a 1 percentage point change in diversity in either direction would lower population density by 1.4 percent. Overall, based on the R^2 of the regression, the cross-country variations in genetic diversity, agricultural transition timing, and land productivity together explain 67 percent of the cross-country variation in population density in 1500 CE.

Finally, Column 6 reports the results from estimating the baseline regression model, specified in equation (8), which allows the analysis to capture unobserved continent-specific attributes that could potentially have an influence on population density. Despite the more modest cross-country variation

[40] To interpret the coefficients associated with the land productivity channel, a 1 percent increase in the fraction of arable land and in absolute latitude corresponds, respectively, to a 0.5 percent increase and a 0.2 percent 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, Johnson and Robinson (2005).

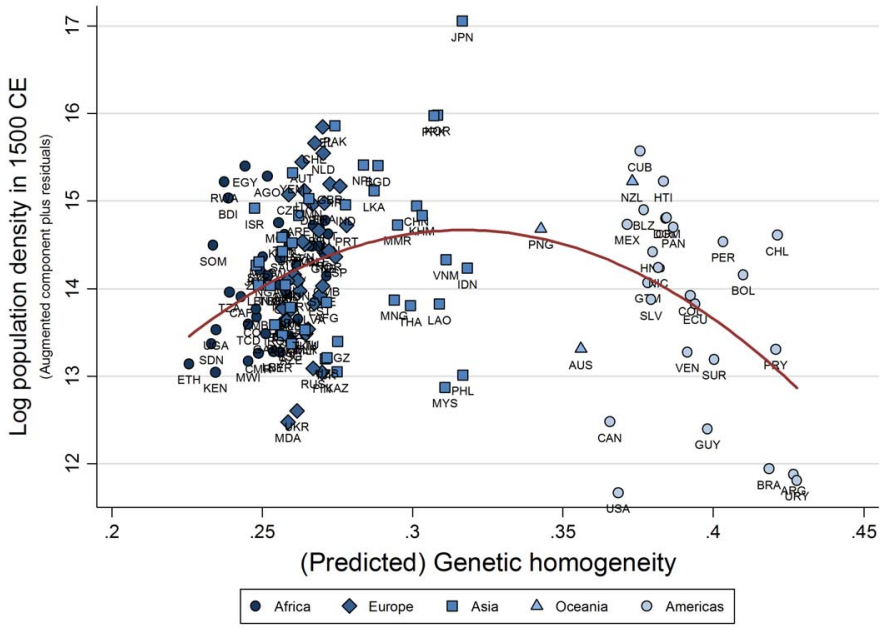


FIGURE 4: Predicted Genetic Diversity and Population Density in 1500 CE

Notes: This figure depicts the hump-shaped effect, estimated using a least-squares quadratic fit, of predicted genetic homogeneity (i.e., 1 minus genetic diversity as predicted by migratory distance from East Africa) on log population density in 1500 CE in the extended 145-country sample, conditional on the timing of the Neolithic Revolution, land productivity, and continent fixed effects. This figure is an augmented component-plus-residual plot rather than the typical added-variable plot of residuals against residuals. Specifically, the vertical axis represents fitted values (as predicted by genetic homogeneity and its square) of log population density plus the residuals from the full regression model. The horizontal axis, on the other hand, represents genetic homogeneity rather than the residuals obtained from regressing homogeneity on the control variables in the model. This methodology permits the illustration of the overall nonmonotonic effect of genetic homogeneity in one scatter plot.

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 continent fixed effects, although the statistical significance of the linear coefficient drops to the 5 percent level.[41] Specifically, the coefficients associated with the diversity channel indicate that, controlling

[41] Table D7 in Section D of the appendix demonstrates that the baseline findings for genetic diversity from both the limited- and extended-sample variants of the historical analysis, as well as those from the contemporary analysis, remain qualitatively intact under alternative regression specifications that control for some of the individual ecological components of the land suitability index, including temperature, precipitation, and soil fertility, in lieu of the baseline control for the overall suitability of land for cultivation.

In addition, consistent with the fact that the baseline control for the timing of the Neolithic Revolution should be expected to capture the contemporaneous effect of the mode of subsistence on population density in the precolonial Malthusian era, Table D8 establishes that augmenting the baseline specifications for examining population density in the years 1500 CE and 1000 CE with an explicit control for the mode of subsistence, while diminishing somewhat the explanatory power of the transition timing channel, does not affect the main findings for genetic diversity. Note that, given underlying data availability constraints on constructing a proxy for the mode of subsistence prevalent in the year 1500 CE, coupled with the fact that cross-country subsistence patterns in 1000 CE should be expected to be highly correlated with those existing in 1500 CE, the analysis in Table D8 controls only for the mode of subsistence prevalent in the year 1000 CE in augmented regressions explaining population density in both time periods. For detailed definitions of additional control variables used by the robustness analyses in Tables D7 and D8, the interested reader is referred to Section F of the appendix.

for the influence of land productivity, the timing of the Neolithic Revolution, and continent 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 percent, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 29 percent. In addition, a 1 percentage point change in genetic diversity in either direction at the predicted optimum level of 0.683, which roughly corresponds to the diversity predicted (by migratory distance from East Africa) for Japan, would lower population density by 1.4 percent. Reassuringly, the optimal level of predicted diversity in the extended sample is quite similar to that obtained for observed diversity in the limited 21-country sample.

To place the worldwide effect of the diversity channel into perspective, the coefficients reported in Column 6 imply that increasing the expected heterozygosity of the most homogenous *native* South American populations by 11 percentage points to the predicted optimum would have raised their population density in 1500 CE by a factor of 6.1. On the other hand, decreasing the expected heterozygosity of the most heterogenous East African populations by 9.1 percentage points to the optimum would have raised their population density by a factor of 3.4. The hump-shaped effect of genetic diversity on log population density in 1500 CE, conditional on the timing of the Neolithic transition, land productivity, and continent fixed effects, is depicted in Figure 4.[42] Moreover, the partial R^2 associated with genetic diversity suggests that residual genetic diversity explains about 7 percent of the cross-country variation in residual log population density in 1500 CE, conditional on the covariates from the baseline regression model.

To summarize the results reported in Table 3, genetic diversity as predicted by migratory distance from East Africa is found to have a highly statistically significant nonmonotonic 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 continent fixed effects as well.

4.3 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 nonmonotonic 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 nonmonotonic

[42] Plots depicting the partial regression lines associated with the first- and second-order effects of genetic homogeneity on log population density in 1500 CE are presented in Figures C3(a)–C3(b) in Section C of the appendix.

TABLE 4: Robustness to Alternative Distances

Distance from:	Addis Ababa (1)	Addis Ababa (2)	London (3)	Tokyo (4)	Mexico City (5)
	Dependent variable is log population density in 1500 CE				
Migratory distance	0.138**		-0.040	0.052	-0.063
	(0.061)		(0.063)	(0.145)	(0.099)
Migratory distance square	-0.008***		-0.002	-0.006	0.005
	(0.002)		(0.002)	(0.007)	(0.004)
Aerial distance		-0.008			
		(0.106)			
Aerial distance square		-0.005			
		(0.006)			
Log Neolithic transition timing	1.160***	1.158***	1.003***	1.047***	1.619***
	(0.144)	(0.138)	(0.164)	(0.225)	(0.277)
Log percentage of arable land	0.401***	0.488***	0.357***	0.532***	0.493***
	(0.091)	(0.102)	(0.092)	(0.089)	(0.094)
Log absolute latitude	-0.342***	-0.263***	-0.358***	-0.334***	-0.239***
	(0.091)	(0.097)	(0.112)	(0.099)	(0.083)
Log land suitability for agriculture	0.305***	0.254**	0.344***	0.178**	0.261***
	(0.091)	(0.102)	(0.092)	(0.080)	(0.092)
Observations	145	145	145	145	145
R^2	0.67	0.59	0.67	0.59	0.63

Notes: This table establishes that, unlike migratory distance from East Africa, alternative concepts of distance, including aerial distance from East Africa and migratory distances from placebo points of origin in other continents across the globe, do not possess any systematic relationship, hump-shaped or otherwise, with log population density in 1500 CE while controlling for the timing of the Neolithic Revolution and land productivity. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

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 East Africa 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 East Africa (i.e., 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 East Africa.[43] As revealed in Table G4 in Section G of the appendix, with the exception of migratory distance from Tokyo, these other distances are rather strongly correlated with migratory distance from East Africa. Despite

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

some of these high correlations, however, the results presented in Table 4 indicate that migratory distance from East Africa is the only concept of distance that confers a significant nonmonotonic 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 East Africa 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 East Africa, 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 prehistoric migration of humans out of Africa have had a lasting impact on comparative development. Given the high correlations between migratory distance from East Africa 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 East Africa 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.

5 The Contemporary Analysis: Data and Empirical Strategy

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

5.1 The Index of Contemporary National Population Diversity

The index of genetic diversity for contemporary national populations accounts for their ethnic compositions resulting from population flows amongst countries in the post-1500 era, the genetic diversity of the precolonial ancestral population of each component ethnic group, and the genetic distances between these ancestral populations. Specifically, given the genetic diversity of the ancestral populations of the source countries, data on post-1500 population flows can be used to construct a weighted average expected heterozygosity measure for the national population of each country in the contemporary period.[44] 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 subnational 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 the field of population genetics. Details regarding the construction of this ancestry-adjusted measure of genetic diversity,

[44] The data on ethnic compositions are obtained from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2010) who compile, for each country in their data set, the share of the country's population in 2000 CE that is descended from the population of every other country in 1500 CE.

TABLE 5: Adjusted versus Unadjusted Diversity

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is log income per capita in 2000 CE					
Predicted diversity	541.792***	248.699***			524.240***	374.297**
(ancestry adjusted)	(130.250)	(86.798)			(172.284)	(189.015)
Predicted diversity square	-387.026***	-172.552***			-370.660***	-264.700*
(ancestry adjusted)	(91.148)	(61.446)			(123.664)	(137.333)
Predicted diversity			140.903***	10.152	-1.063	-67.278
(unadjusted)			(51.614)	(52.732)	(74.681)	(84.783)
Predicted diversity square			-107.686***	-7.418	-2.002	52.844
(unadjusted)			(38.133)	(38.000)	(57.317)	(67.248)
Continent fixed effects	No	Yes	No	Yes	No	Yes
Observations	143	143	143	143	143	143
R^2	0.13	0.47	0.08	0.45	0.14	0.48
P-value for joint significance of linear and quadratic terms in:						
Adjusted diversity					0.010	0.039
Unadjusted diversity					0.419	0.748

Notes: This table establishes that, when explaining log income per capita in 2000 CE, the ancestry-adjusted measure of genetic diversity outperforms the unadjusted measure in terms of (i) the qualitative robustness of the hump-shaped effect to continent fixed effects and (ii) maintaining explanatory power in regressions that perform a horse race between the two measures of diversity. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

that also accounts for the diversity arising from differences between subnational ethnic groups, are presented in Section B of the appendix.

Reassuringly, the ancestry-adjusted measure of genetic diversity dominates the unadjusted measure in predicting economic development in the contemporary period.[45] In line with the diversity hypothesis, Column 1 in Table 5 reveals a significant unconditional hump-shaped relationship between the adjusted measure of diversity and income per capita in the year 2000 CE.[46]

Column 2 establishes that the unconditional quadratic relationship from Column 1 remains qualitatively intact when conditioned for the impact of continent fixed effects. As revealed in Columns 3 and 4, however, while the unadjusted measure also possesses a significant unconditional hump-shaped relationship with income per capita across countries, the relationship disappears once

[45] Table D9 in Section D of the appendix establishes that migratory distance from East Africa, adjusted to reflect the weighted average of migratory distances of the precolonial ancestral populations of a country today, is the only distance concept that confers a significant hump-shaped effect on income per capita in 2000 CE. As shown in the table, the other distance concepts, including (i) the unadjusted measure of migratory distance from East Africa (used in the historical analysis), (ii) the aerial distance from East Africa, and (iii) the ancestry-adjusted aerial distance from East Africa, do not confer any systematic nonmonotonic effect on income per capita in 2000 CE, given that the ancestry-adjusted migratory distance measure is accounted for by the regression.

[46] This quadratic relationship is depicted in Figures C4(a) and C4(b) in Section C of the appendix along with a nonparametric local polynomial regression line and a restricted cubic spline regression line respectively. As in the preceding historical analysis, the estimated quadratic falls within the 95 percent confidence interval bands of both the nonparametric and cubic spline relationships. The figure notes provide additional details on the estimation procedures underlying the nonparametric and cubic spline regressions.

the regression is augmented to account for continent fixed effects. Moreover, examining jointly the explanatory powers of the ancestry-adjusted and unadjusted measures of genetic diversity for income per capita, Columns 5 and 6 demonstrate the superior relative performance of the adjusted measure, regardless of whether continent fixed effects are accounted for, lending further credence ex post to the methodology employed in constructing the index of contemporary population diversity.

5.2 The Empirical Model

Maintaining symmetry with the earlier historical analysis, a regression specification similar to that employed for the historical regressions is adopted initially to examine the impact of genetic diversity on log income per capita in the year 2000 CE while controlling for 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 through contemporary cultural and institutional factors.

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

$$\ln y_i = \gamma_0 + \gamma_1 \hat{G}_i + \gamma_2 \hat{G}_i^2 + \gamma_3 \ln T_i + \boldsymbol{\gamma}_4' \ln \mathbf{X}_i + \boldsymbol{\gamma}_5' \ln \boldsymbol{\Lambda}_i + \boldsymbol{\gamma}_6' \ln \boldsymbol{\Gamma}_i + \eta_i, \tag{9}$$

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, T_i and $\mathbf{X}_i$ are the Neolithic transition timing and land productivity controls for country i, $\boldsymbol{\Lambda}_i$ is a vector of institutional and cultural controls for country i, $\boldsymbol{\Gamma}_i$ is a vector of additional geographical controls for country i, and, finally, η_i is a country-specific disturbance term.[47]

6 The Contemporary Analysis: Empirical Findings

6.1 Results for Comparative Development

The empirical findings indicate that the highly significant hump-shaped effect of genetic diversity on macroeconomic outcomes in the preindustrial period is present in the contemporary period as well. Furthermore, the persistent hump-shaped impact of genetic diversity on the pattern of comparative economic development is a direct effect that is not captured by contemporary geographical, institutional, and cultural factors.[48]

Using a sample of 143 countries for which data are available for the entire set of control variables used in the baseline regression for the year 1500 CE, Column 1 of Table 6 reveals a significant hump-

[47] The data on income per capita are from the *Penn World Table*, version 6.2. The institutional and cultural controls include the social infrastructure index of Hall and Jones (1999), the share of the population of European descent based on the *World Migration Matrix, 1500–2000* of Putterman and Weil (2010), legal origin dummies and the shares of the population affiliated with major world religions from the data set of La Porta et al. (1999), as well as the ethnic fractionalization index of Alesina et al. (2003). The additional geographical controls include the share of the population at risk of contracting falciparum malaria from Gallup and Sachs (2001), as well as the share of the population living

TABLE 6: Diversity and Economic Development in 2000 CE and 1500 CE

	(1)	(2)	(3)	(4)
	Dependent variable is:			
	Log income per capita in 2000 CE			Log population density in 1500 CE
Predicted diversity (ancestry adjusted)	203.443**	235.409***	242.886***	
	(83.368)	(83.493)	(81.773)	
Predicted diversity square (ancestry adjusted)	-142.663**	-165.293***	-169.960***	
	(59.037)	(59.393)	(58.252)	
Predicted diversity (unadjusted)				198.587**
				(79.225)
Predicted diversity square (unadjusted)				-145.320***
				(55.438)
Log Neolithic transition timing (ancestry adjusted)		0.062	0.005	
		(0.263)	(0.306)	
Log Neolithic transition timing (unadjusted)	-0.151			1.238***
	(0.197)			(0.241)
Log percentage of arable land	-0.112	-0.122	-0.140	0.378***
	(0.103)	(0.108)	(0.112)	(0.108)
Log absolute latitude	0.163	0.171	0.191	-0.423***
	(0.117)	(0.119)	(0.143)	(0.122)
Log land suitability for agriculture	-0.192**	-0.176*	-0.187*	0.264***
	(0.096)	(0.102)	(0.102)	(0.095)
Log population density in 1500 CE			0.047	
			(0.097)	
Optimal diversity	0.713***	0.712***	0.715***	0.683***
	(0.225)	(0.033)	(0.043)	(0.095)
Continent fixed effects	Yes	Yes	Yes	Yes
Observations	143	143	143	143
R^2	0.57	0.57	0.57	0.68

Notes: This table (i) establishes the significant hump-shaped effect of ancestry-adjusted genetic diversity on log income per capita in 2000 CE in a 143-country sample while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, and (ii) demonstrates that the hump-shaped effect of genetic diversity is robust to (a) adjusting the control for the timing of the Neolithic Revolution to incorporate information on post-1500 population flows and (b) accounting for historical inertia in the effect of genetic diversity by way of controlling for log poppulation density in 1500 CE. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

in Köppen-Geiger tropical zones and distance from the nearest coast or sea-navigable river, both from the data set of Gallup, Sachs and Mellinger (1999). See Section F of the appendix for further details.

[48] Since (i) genetic diversity for contemporary national populations is partly based on the F_{st} genetic distances between their precolonial ancestral groups, and (ii) F_{st} genetic distances and relative expected heterozygosities, even amongst populations in the precolonial era, are in part codetermined by migratory distances, it is necessary to ensure that the observed hump-shaped effect of genetic diversity on comparative development, in both the precolonial and contemporary eras, is not reflecting the latent impact of genetic distance to either the cradle of humankind or the world technological frontier, via channels related to the diffusion of development (Spolaore and Wacziarg, 2009). Table D10 in Section D of the appendix demonstrates that the baseline findings for genetic diversity from both the limited- and extended-sample variants of the historical analysis, as well as those from the contemporary analysis, are virtually unaffected when the regression specifications are augmented to account for appropriate measures of F_{st} genetic distance to Ethiopia and to the world technology frontier relevant for the time period being examined. For detailed definitions

shaped effect of genetic diversity on income per capita in 2000 CE, accounting for the set of baseline controls employed in the historical analysis, i.e., the logs of the timing of the Neolithic transition, the percentage of arable land, absolute latitude, and the suitability of land for agriculture, as well as continent fixed effects.[49] Further, consistent with the notion that the optimal level of diversity increased in the process of industrialization, as the beneficial forces associated with greater diversity became intensified in an environment characterized by more rapid technological progress, the findings indicate that the optimal level of diversity with respect to the modern world income distribution is indeed higher than that obtained with respect to population density in the precolonial Malthusian era. Specifically, while the estimate for the optimal level in 1500 CE is 0.683 (Column 4), the estimated optimum in 2000 CE, under the same set of controls, is 0.713.

Column 2 shows that the hump-shaped effect of diversity on income per capita remains virtually intact when the control for the Neolithic transition is adjusted to capture the average time elapsed since the precolonial ancestral populations of each country today experienced the transition to agriculture (i.e., traits that are embodied in the country's population today, rather than the country's geographical attributes). In particular, the estimated linear and quadratic coefficients on genetic diversity are both statistically significant at the 1 percent level. They imply that increasing the diversity of the most genetically homogenous country in the sample (Bolivia) by 1 percentage point would raise its income per capita in 2000 CE by 30 percent, whereas decreasing the diversity of the most genetically diverse country in the sample (Ethiopia) by 1 percentage point would raise its income per capita by 21 percent. Further, a 1 percentage point change in diversity (in either direction) at the optimum level of 0.712 would lower income per capita by 1.6 percent.[50]

Importantly, the hump-shaped effect of genetic diversity on income per capita in 2000 CE does not reflect an inertia originating from its effect on technology and thus on population density in 1500 CE. As established in Column 3, the results are essentially unchanged if the regression accounts for the potentially confounding effect of population density in 1500 CE. Namely, the effect of genetic diversity on income per capita in 2000 CE does not operate though its impact on population density in the year 1500 CE.

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

Using a sample of 109 countries for which data are available for the institutional and cultural controls that are employed in the examination, Column 1 of Table 7 demonstrates that genetic diversity has a hump-shaped effect on income per capita in the year 2000 CE, accounting for the set of baseline controls employed in the historical analysis, i.e., the logs of the weighted timing of the

---

of the various genetic distance controls used by the robustness analysis in Table D10, the reader is referred to Section F of the appendix.

[49] Tables G5–G6 in Section G of the appendix present the relevant descriptive statistics for this 143-country sample. The difference in sample size with the 145-country sample used in the historical analysis arises from the fact that there exist observations for which data are unavailable for both income per capita in 2000 CE and population density in 1500 CE.

[50] Table D13 in Section D of the appendix reports the standardized beta coefficient and partial R^2 associated with each regressor in the baseline regressions for both the historical and contemporary analyses.

TABLE 7: Diversity and Other Determinants of Economic Development in 2000 CE

	Full sample (1)	Full sample (2)	Full sample (3)	Full sample (4)	Full sample (5)	Full sample (6)	Schooling sample (7)	Schooling sample (8)
	Dependent variable is log income per capita in 2000 CE							
Predicted diversity (ancestry adjusted)	311.865***	232.786***	227.325***	251.522***	281.173***	264.464***	271.262***	219.955***
	(83.997)	(64.965)	(65.074)	(68.815)	(70.459)	(83.157)	(72.343)	(64.696)
Predicted diversity square (ancestry adjusted)	-218.579***	-160.645***	-156.930***	-174.234***	-195.010***	-183.405***	-188.845***	-153.875***
	(59.364)	(46.006)	(46.002)	(48.474)	(49.764)	(58.195)	(51.761)	(46.072)
Log Neolithic transition timing (ancestry adjusted)	-0.270	-0.082	0.048	0.366	0.417*	0.374	0.125	-0.037
	(0.271)	(0.202)	(0.189)	(0.242)	(0.231)	(0.240)	(0.233)	(0.210)
Log percentage of arable land	-0.221***	-0.162***	-0.175***	-0.216***	-0.189***	-0.190***	-0.091	-0.089
	(0.061)	(0.049)	(0.048)	(0.047)	(0.049)	(0.049)	(0.056)	(0.056)
Log absolute latitude	0.121	0.078	0.031	0.112	0.003	-0.003	0.001	-0.012
	(0.123)	(0.100)	(0.102)	(0.095)	(0.105)	(0.105)	(0.096)	(0.086)
Social infrastructure		2.375***	2.407***	2.109***	1.867***	1.812***	1.515***	0.908**
		(0.269)	(0.254)	(0.371)	(0.404)	(0.419)	(0.420)	(0.420)
Ethnic fractionalization			-0.701**	-0.522*	-0.351	-0.365	0.027	-0.134
			(0.275)	(0.316)	(0.276)	(0.289)	(0.274)	(0.262)
Percentage of population at risk of contracting malaria					-0.504	-0.483	-0.723*	-0.715**
					(0.344)	(0.369)	(0.383)	(0.349)
Percentage of population living in tropical zones					-0.314	-0.291	-0.410**	-0.192
					(0.199)	(0.204)	(0.208)	(0.197)
Mean distance to nearest waterway					-0.373**	-0.372**	-0.189	-0.067
					(0.184)	(0.186)	(0.215)	(0.189)
Percentage of population of European descent						0.211		
						(0.618)		
Years of schooling								0.134***
								(0.043)
Optimal diversity	0.713***	0.725***	0.724***	0.722***	0.721***	0.721***	0.718***	0.715***
	(0.013)	(0.044)	(0.028)	(0.013)	(0.016)	(0.055)	(0.099)	(0.032)
OPEC fixed effect	No	No	No	No	Yes	Yes	Yes	Yes
Legal origin fixed effects	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^2	0.74	0.84	0.85	0.87	0.90	0.90	0.91	0.93

Notes: This table establishes, using a feasible 109-country sample, that the significant hump-shaped effect of ancestry-adjusted genetic diversity on log income per capita in 2000 CE, while controlling for the ancestry-adjusted timing of the Neolithic Revolution, land productivity, and continent fixed effects, is robust to additional controls for institutional, cultural, and geographical determinants of development, including social infrastructure, ethnic fractionalization, legal origins, major religion shares, tropical disease environments, access to waterways, and natural resource endowments, as well as controls for the share of the population of European descent and human capital formation. All regressions include Sub-Saharan Africa and continent fixed effects. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Neolithic transition, the percentage of arable land, and absolute latitude, as well as continent fixed effects.[51] The estimated linear and quadratic coefficients associated with the diversity channel are both statistically significant at the 1 percent level, and the estimate for the optimal level of diversity is 0.713. The regression in Column 2 examines the robustness of the results to the inclusion of a measure of institutional quality, as captured by the social infrastructure index of Hall and Jones (1999). The estimated hump-shaped effect of genetic diversity remains highly statistically significant and rather stable, while the optimal level of diversity increases to 0.725.[52]

The regression in Column 3 is designed to examine whether the effect of genetic diversity operates via ethnic fractionalization. It demonstrates that the effect of genetic diversity is virtually unaffected by the potentially confounding impact of ethnic fractionalization.[53] While, as established earlier in the literature, ethnic fractionalization does indeed confer a significant adverse effect on income per capita in the year 2000 CE, the hump-shaped impact of genetic diversity remains highly statistically significant. Moreover, the estimate for the optimal level of diversity, 0.724, is effectively unchanged in comparison to earlier columns.

Column 4 demonstrates the robustness of the hump-shaped effect of genetic diversity to the inclusion of additional cultural and institutional controls (i.e., legal origins and the fraction of the population affiliated with major religions). The coefficients associated with genetic diversity remain highly significant statistically and rather stable in magnitude, while the estimated optimal level of diversity, 0.722, remains virtually intact.

Column 5 establishes the robustness of the results to the inclusion of controls for the health environment (i.e., percentage of the population at risk of contracting malaria and percentage of the population in tropical zones), additional geographical controls gauging access to waterways, and an OPEC dummy. The results in this column, which reflects the baseline specification for examining the impact of diversity on development in the modern world, therefore reveal the direct effect of genetic diversity, once institutional, cultural, and geographical factors are accounted for.[54] The direct hump-shaped impact of genetic diversity on log income per capita in 2000 CE, as established

[51] The agricultural suitability index was not found to enter significantly in any of the specifications examined in Table 7 and is therefore dropped from the analysis. Tables G7–G8 in Section G of the appendix present the relevant descriptive statistics for the 109-country sample employed in Tables 7–8.

[52] The inclusion of measures from the *Polity IV* data set, reflecting the extent of democracy or the degree of constraints on the power of chief executives, as additional controls for institutional quality do not affect the results for genetic diversity. Moreover, because these measures enter insignificantly in the regression once social infrastructure has been controlled for, they are excluded from the analysis in Table 7.

[53] Results (not shown) from estimating a similar specification that included ethnic fractionalization square as an additional explanatory variable did not reveal any discernible nonmonotonic relationship between ethnic fractionalization and income per capita in 2000 CE. Importantly, the regression coefficients associated with genetic diversity, as well as the estimate for the optimal level of diversity, were unaffected.

[54] As established by Table D11 in Section D of the appendix, the baseline results for diversity in the contemporary analysis are qualitatively robust to (i) controls for region (rather than continent) fixed effects (Column 1), (ii) dropping observations (Columns 2–4) associated with Sub-Saharan Africa and Latin America that, given their laggard development in comparison to other regions, along with their relatively higher and lower levels of genetic diversity respectively, may a priori be considered to be influential for generating the worldwide hump-shaped relationship between diversity and development, and (iii) restricting the regression sample (Column 5) to only countries in the potentially influential Sub-Saharan Africa and Latin America regional clusters. These results demonstrate that the direct hump-shaped effect of genetic diversity on income per capita in the modern world is not simply a reflection of worldwide cross-regional variations in diversity and economic development.

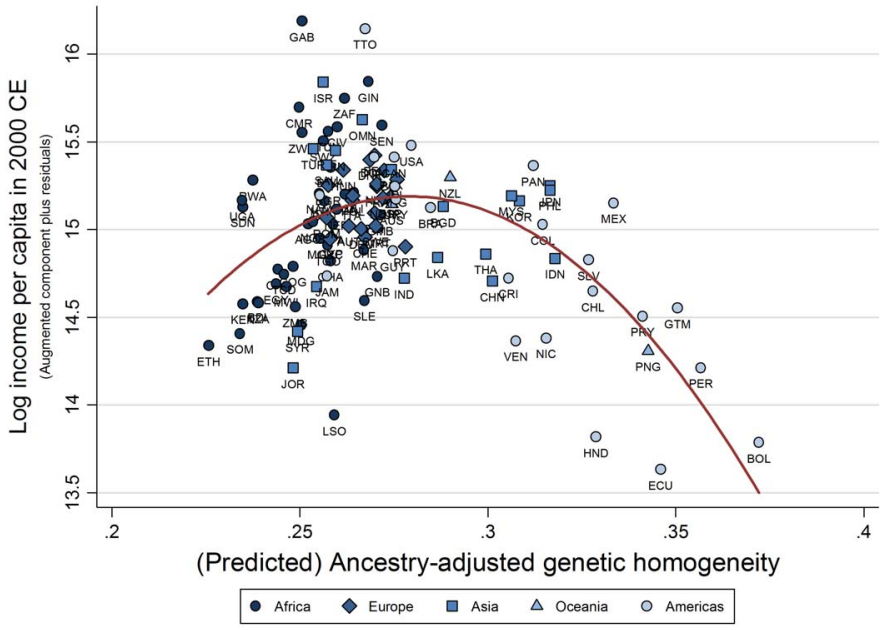


FIGURE 5: Ancestry-Adjusted Genetic Diversity and Income Per Capita in 2000 CE

Notes: This figure depicts the hump-shaped effect, estimated using a least-squares quadratic fit, of ancestry-adjusted genetic homogeneity (i.e., 1 minus ancestry-adjusted genetic diversity) on log income per capita in 2000 CE in a 109-country sample, conditional on the ancestry-adjusted timing of the Neolithic Revolution, land productivity, a vector of institutional, cultural, and geographical determinants of development, and continent fixed effects. This figure is an augmented component-plus-residual plot rather than the typical added-variable plot of residuals against residuals. Specifically, the vertical axis represents fitted values (as predicted by ancestry-adjusted genetic homogeneity and its square) of log income per capita plus the residuals from the full regression model. The horizontal axis, on the other hand, represents ancestry-adjusted genetic homogeneity rather than the residuals obtained from regressing homogeneity on the control variables in the model. This methodology permits the illustration of the overall nonmonotonic effect of genetic homogeneity in one scatter plot.

in Column 5, is depicted in Figure 5.[55] Moreover, the partial R^2 associated with genetic diversity suggests that residual genetic diversity explains about 16 percent of the cross-country variation in residual log income per capita in 2000 CE, conditional on the covariates from the baseline regression model.

The coefficients associated with the diversity channel in Column 5 imply that: (i) increasing the diversity of the most homogenous country in the sample (Bolivia) by 1 percentage point would raise its income per capita in the year 2000 CE by 41 percent, (ii) decreasing the diversity of the most diverse country in the sample (Ethiopia) by 1 percentage point would raise its income per capita by 21 percent, (iii) a 1 percentage point change in genetic diversity (in either direction) at the optimum level of 0.721 (that most closely resembles the diversity level of the U.S.) would lower income per capita by 1.9 percent, (iv) increasing the diversity of Bolivia to the level prevalent in the U.S. would increase Bolivia's per capita income by a factor of 5.4, closing the income gap between the two countries from a ratio of 12:1 to 2.2:1, and (v) decreasing the diversity of Ethiopia to the

[55] Plots depicting the partial regression lines associated with the first- and second-order effects of genetic homogeneity on log income per capita in 2000 CE are presented in Figures C5(a)–C5(b) in Section C of the appendix.

TABLE 8: Addressing Endogenous Post-1500 Migrations

	Full sample (1)	Omit OECD countries (2)	Omit Neo-European countries (3)	Omit Latin American countries (4)	Omit Sub-Saharan African countries (5)	Countries with population at least 97 percent indigenous (6)
	Dependent variable is log income per capita in 2000 CE					
Predicted diversity (ancestry adjusted)	281.173***	266.494***	266.198***	412.214***	267.379***	304.735**
	(70.459)	(84.381)	(72.371)	(148.805)	(103.063)	(111.588)
Predicted diversity square (ancestry adjusted)	-195.010***	-185.223***	-185.113***	-287.058***	-185.698**	-213.389**
	(49.764)	(59.302)	(50.915)	(102.036)	(74.434)	(77.255)
Log Neolithic transition timing (ancestry adjusted)	0.417*	0.415	0.376	0.517*	0.085	0.448*
	(0.231)	(0.283)	(0.231)	(0.298)	(0.442)	(0.254)
Log percentage of arable land	-0.189***	-0.238***	-0.205***	-0.189***	-0.215**	-0.104
	(0.049)	(0.060)	(0.053)	(0.050)	(0.093)	(0.061)
Log absolute latitude	0.003	-0.021	-0.029	-0.139	0.202	-0.074
	(0.105)	(0.119)	(0.109)	(0.126)	(0.227)	(0.130)
Social infrastructure	1.867***	1.373**	1.485***	2.044***	1.618***	1.311*
	(0.404)	(0.575)	(0.501)	(0.545)	(0.479)	(0.716)
Ethnic fractionalization	-0.351	-0.450	-0.404	-0.752**	0.069	-0.044
	(0.276)	(0.375)	(0.297)	(0.348)	(0.404)	(0.412)
Percentage of population at risk of contracting malaria	-0.504	-0.593	-0.585	-0.308	-0.434	-0.153
	(0.344)	(0.376)	(0.363)	(0.486)	(0.547)	(0.434)
Percentage of population living in tropical zones	-0.314	-0.199	-0.300	-0.520**	-0.520	-0.339
	(0.199)	(0.239)	(0.215)	(0.252)	(0.332)	(0.312)
Mean distance to nearest waterway	-0.373**	-0.389*	-0.450**	-0.494***	-0.725	-0.367*
	(0.184)	(0.222)	(0.208)	(0.186)	(0.481)	(0.201)
Optimal diversity	0.721***	0.719***	0.719***	0.718***	0.720***	0.714***
	(0.016)	(0.069)	(0.014)	(0.023)	(0.154)	(0.012)
Observations	109	83	105	87	71	37
R^2	0.90	0.82	0.89	0.93	0.87	0.98

Notes: This table addresses concerns regarding the endogeneity, arising from endogenous post-1500 population flows, of ancestry-adjusted genetic diversity to economic development by establishing that the significant hump-shaped effect of ancestry-adjusted genetic diversity on log income per capita in 2000 CE is robust to restricting the sample to (i) non-OECD economies (i.e., economies that were less attractive to migrants), (ii) non-Neo-European countries (i.e., excluding the U.S., Canada, Australia, and New Zealand), (iii) non-Latin American countries, (iv) non-Sub-Saharan African countries, and (v) countries whose indigenous population is larger than 97 percent of the entire population (i.e., under conditions that virtually eliminate the role of migration in contributing to diversity) while controlling for the ancestry-adjusted timing of the Neolithic Revolution, land productivity, a vector of institutional, cultural, and geographical determinants of development, and continent fixed effects. All regressions include controls for major religion shares as well as OPEC, legal origin, and continent fixed effects. All regressions with the exception of Column 5 also include a fixed effect for Sub-Saharan Africa. Bootstrap standard errors are reported in parentheses in Columns 1–5. Heteroskedasticity robust standard errors are reported in parentheses in Column 6. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

level prevalent in the U.S. would increase Ethiopia's per capita income by a factor of 1.7 and thus close the income gap between the two countries from a ratio of 47:1 to 27:1.

The regression in Column 6 examines the robustness of the baseline results to accounting for the potentially confounding effect of the share of the population of European descent. The results indicate that the coefficients associated with genetic diversity remain highly statistically significant and reassuringly stable in magnitude, while the estimated optimal level of diversity is virtually unaffected.[56]

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

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

In particular, as established in Table 8, the hump-shaped effect of genetic diversity remains highly significant and the optimal diversity estimate remains virtually intact if the sample is restricted to (i) non-OECD economies (i.e., economies that were less attractive to migrants) in Column 2, (ii) non-Neo-European countries (i.e., excluding the U.S., Canada, Australia, and New Zealand) in Column 3, (iii) non-Latin American countries in Column 4, (iv) non-Sub-Saharan African countries in Column 5, and (v) countries whose indigenous population is larger than 97 percent of the entire population (i.e., under conditions that virtually eliminate the role of migration in contributing to diversity over the last 500 years) in Column 6.[57]

6.2 The Costs and Benefits of Genetic Diversity

This section presents empirical evidence on some of the channels through which genetic diversity confers a hump-shaped effect on income per capita across countries in the modern world. In line with the theory that diversity should be expected to confer costs on productivity, in the form of lower social

[56] Since the share of the population of European descent is expected to be highly correlated with unobserved fixed effects associated with European countries as well as the Neo-Europes (i.e., the U.S., Canada, Australia, and New Zealand), Table D12 in Section D of the appendix repeats the analysis from Columns 5–6 in Table 7 on different cuts of the cross-country sample, focusing primarily on countries where the share of Europeans in the population is not confoundingly close to 1. Importantly, the findings therein suggest that the baseline results are robust to controlling for the share of Europeans in the population even when the sample is restricted to non-OECD countries (Columns 3–4), non-Neo-European countries (Columns 5–6), and non-European countries (Columns 7–8).

[57] This result reflects the well-known fact from the field of population genetics that the overwhelming majority of genetic diversity in human populations stems from the diversity *within* groups, as opposed to the diversity *between* groups (see, e.g., Barbujani et al., 1997).

capital, and benefits, in the form of more rapid knowledge creation, it establishes that countries with greater diversity are also characterized, on average, by a lower prevalence of interpersonal trust and a higher intensity of scientific knowledge creation. Specifically, exploiting cross-country variations in the degree of interpersonal trust and the annual average number of scientific articles per capita in the 1981–2000 time period, the analysis demonstrates that genetic diversity has a statistically significant negative relationship with the prevalence of trust but a positive one with scientific productivity, conditional on a similar set of baseline controls employed in the preceding analysis of contemporary comparative development.[58]

Using a sample of 58 countries for which trust data as well as data on all baseline controls employed by the analysis are available, Column 1 of Table 9 shows that genetic diversity has a statistically significant negative effect on the prevalence of trust, accounting for the Neolithic transition timing and land productivity channels, as well as contemporary cultural, geographical, and institutional factors. The coefficient corresponding to the diversity channel indicates that a 1 percentage point increase in genetic diversity is associated with a 2 percentage point decrease in the prevalence of trust. Moreover, as demonstrated in Column 2, this adverse impact of diversity on trust remains intact when the regression specification is augmented to account for the effect of average years of schooling in the population.

In contrast, using a sample of 93 countries for which data on scientific productivity as well as data on the full set of baseline controls are available, Column 4 shows that genetic diversity has a statistically significant positive effect on the average annual number of published scientific articles per capita. Specifically, the coefficient of interest indicates that a 1 percentage point increase in genetic diversity is associated with an increase in the annual number of scientific articles per capita of about 0.02, conditional on the influence of Neolithic transition timing, land productivity, and contemporary cultural, geographical, and institutional factors. In addition, as established by Column 5, accounting for the effect of average years of schooling in the population does not qualitatively alter the beneficial impact of diversity on scientific productivity in the 1981–2000 time period.

Finally, Columns 3 and 6 demonstrate that the conditional detrimental effect of genetic diversity on trust as well as its beneficial effect on scientific productivity both remain fully intact when the relevant regressions are performed on a common sample of countries, thereby lending further support to the theoretical assertion that diversity confers conflicting effects on productivity, generating inefficiencies in the production process while, at the same time, fostering the expansion of society's production possibility frontier.

[58] Consistently with empirical findings in the existing literature, Table D14 in Section D of the appendix demonstrates that the prevalence of trust and the intensity of scientific knowledge creation both individually possess statistically significant positive correlations with log income per capita in 2000 CE. These correlations hold both unconditionally (Columns 1 and 4) and conditional on either the baseline set of controls from the historical analysis (Columns 2 and 5) or a more comprehensive set of institutional, cultural, and geographical controls from the contemporary analysis (Columns 3 and 6). For details on the definitions and data sources of the variables gauging the prevalence of trust and the intensity of scientific knowledge creation, the reader is referred to Section F of the appendix.

TABLE 9: The Costs and Benefits of Diversity

	Full sample (1)	Full sample (2)	Common sample (3)	Full sample (4)	Full sample (5)	Common sample (6)
	Dependent variable is:					
	Degree of interpersonal trust			Scientific articles per capita per year 1981–2000		
Predicted diversity (ancestry adjusted)	-1.880**	-2.226**	-1.920**	2.484***	1.860***	3.023**
	(0.829)	(0.862)	(0.940)	(0.566)	(0.550)	(1.222)
Log Neolithic transition timing (ancestry adjusted)	0.069	0.089	0.091	-0.085*	-0.080*	-0.189**
	(0.062)	(0.063)	(0.060)	(0.047)	(0.046)	(0.085)
Log percentage of arable land	0.004	-0.002	-0.014	0.008	0.005	-0.005
	(0.019)	(0.018)	(0.019)	(0.016)	(0.015)	(0.037)
Log absolute latitude	-0.003	0.003	-0.008	0.046*	0.055**	0.079
	(0.027)	(0.028)	(0.031)	(0.024)	(0.023)	(0.073)
Social infrastructure	0.200***	0.146*	0.149*	0.702***	0.547***	0.627***
	(0.069)	(0.079)	(0.081)	(0.109)	(0.125)	(0.197)
Ethnic fractionalization	0.060	0.048	0.054	0.091	0.073	0.088
	(0.065)	(0.064)	(0.067)	(0.094)	(0.093)	(0.155)
Percentage of population at risk of contracting malaria	-0.063	-0.033	-0.048	0.093*	0.135***	0.026
	(0.090)	(0.092)	(0.100)	(0.054)	(0.048)	(0.120)
Percentage of population living in tropical zones	-0.074	-0.040	-0.046	0.020	0.051	0.113
	(0.053)	(0.059)	(0.062)	(0.057)	(0.055)	(0.181)
Mean distance to nearest waterway	0.094	0.098	0.077	0.104***	0.115***	0.159**
	(0.064)	(0.060)	(0.058)	(0.038)	(0.035)	(0.067)
Years of schooling		0.013	0.017*		0.031***	0.021
		(0.010)	(0.010)		(0.008)	(0.023)
Observations	58	58	56	93	93	56
R^2	0.79	0.79	0.81	0.80	0.82	0.82

Notes: This table establishes that ancestry-adjusted genetic diversity has a significant negative effect on social capital, as reflected by the prevalence of interpersonal trust amongst individual respondents of the *World Values Survey* in the 1981–2008 time period, and a significant positive effect on the intensity of scientific knowledge creation, as reflected by the annual average number of scientific articles per capita in the 1981–2000 time period, while controlling for the ancestry-adjusted timing of the Neolithic Revolution, land productivity, a vector of institutional, cultural, and geographical determinants of development, human capital formation, and continent fixed effects. All regressions include additional geographical controls for terrain characteristics as well as OPEC, legal origin, Sub-Saharan Africa, and continent fixed effects. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

7 Concluding Remarks

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

The level of genetic diversity within a society is found to have a hump-shaped effect on development outcomes in the precolonial era, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. Moreover, the level of genetic diversity in each country today (as determined by the genetic diversities and genetic distances amongst its ancestral populations) has a nonmonotonic effect on income per capita in the modern world. While the intermediate level of genetic diversity prevalent among Asian and European populations has been conducive for development, the high degree of diversity among African populations and the low degree of diversity among Native American populations have been a detrimental force in the development of these regions.

Finally, this research contributes to the understanding of the role of European colonialism in reshaping comparative development across countries over the last 500 years. Specifically, the results suggest that the cross-country migrations that occurred during the course of European colonization significantly altered the genetic diversity and, hence, the composition of human capital in colonized countries. In particular, the level of diversity that existed in these locations during the precolonial era changed substantially, towards the optimal level for development, in the post-1500 time period. Moreover, consistently with documented patterns of European colonization, the change in diversity was larger in those locations where initial population density was lower. Thus, the reversal of fortunes in comparative development over the last 500 years can potentially be traced to a larger change in the genetic diversity of countries that were less developed during the preindustrial era.

Appendix

This appendix (i) discusses empirical results from additional robustness checks conducted for the historical analysis (Section A), (ii) presents the methodology underlying the construction of the ancestry-adjusted measure of genetic diversity for contemporary national populations (Section B), (iii) collects supplementary figures (Section C) and tables (Section D) of empirical results referenced in the paper, (iv) presents details on the 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel (Section E), (v) provides detailed definitions and data sources of all the variables employed by the empirical analyses in the present study (Section F), (vi) collects descriptive statistics of the cross-country samples employed by the baseline regressions in both the limited- and extended-sample variants of the historical analysis as well as the contemporary analysis (Section G), and, (vii) discusses experimental evidence from scientific studies in the field of evolutionary biology on the costs and benefits of genetic diversity (Section H).

A Additional Robustness Checks for the Historical Analysis

A.1 Results for Earlier Historical Periods

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

The results from replicating the analysis in Section IV.B of the paper to explain log population density in 1000 CE and 1 CE are presented in Tables A.1 and A.2 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 aforementioned explanatory channels, are composed of 140 countries for the 1000 CE regressions and 126 countries for the analysis in 1 CE. Despite more constrained sample sizes, however, the empirical findings once again reveal a highly statistically significant hump-shaped effect of genetic diversity, predicted by migratory distance from East Africa, on 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 continent fixed effects.

In a pattern similar to that observed in Table 3 of the paper, the unconditional effects of genetic diversity in Tables A.1 and A.2 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

TABLE A.1: Predicted Diversity and Economic Development in 1000 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is log population density in 1000 CE					
Predicted diversity	219.722***		158.631**	179.523***	154.913**	201.239**
	(68.108)		(63.604)	(65.981)	(61.467)	(97.612)
Predicted diversity square	-155.442***		-113.110**	-126.147***	-109.806**	-145.894**
	(50.379)		(46.858)	(48.643)	(44.967)	(68.252)
Log Neolithic transition timing		1.393***	1.228***		1.374***	1.603***
		(0.170)	(0.180)		(0.151)	(0.259)
Log percentage of arable land				0.546***	0.371***	0.370***
				(0.140)	(0.106)	(0.114)
Log absolute latitude				-0.151	-0.380***	-0.373***
				(0.103)	(0.110)	(0.137)
Log land suitability for agriculture				0.043	0.211**	0.190*
				(0.135)	(0.104)	(0.106)
Optimal diversity	0.707***		0.701***	0.712***	0.705**	0.690**
	(0.039)		(0.127)	(0.146)	(0.108)	(0.293)
Continent fixed effects	No	No	No	No	No	Yes
Observations	140	140	140	140	140	140
R^2	0.15	0.32	0.38	0.36	0.61	0.62

Notes: This table establishes the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log population density in 1000 CE in an extended 140-country sample while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

previously, these unconditional estimates certainly reflect some amount of omitted variable bias resulting from the exclusion of the transition timing and land productivity channels in Malthusian economic development. On the other hand, unlike the pattern in Table 3 of the paper, the coefficients associated with the diversity channel also weaken moderately in statistical significance, dropping to the 5 percent level when controlling for transition timing in the 1000 CE analysis and to the 10 percent level under controls for the land productivity channel in the 1 CE analysis. Nonetheless, these reductions in statistical significance are 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 A.1 and A.2 reveals the results from exploiting the combined explanatory power of the genetic diversity, transition timing, and land productivity channels for log population density in 1000 CE and 1 CE. Interestingly, in each case, the linear and quadratic coefficients associated with diversity remain rather stable when compared to the corresponding estimates obtained under a partial set of controls in earlier columns. In comparison to the corresponding results for population density in 1500 CE from Table 3 of the paper, the coefficients of the diversity channel uncovered here are statistically significant at the 5 percent as opposed to the 1 percent level, a

TABLE A.2: Predicted Diversity and Economic Development in 1 CE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is log population density in 1 CE					
Predicted diversity	227.826***		183.142***	129.180*	134.767**	231.689**
	(72.281)		(57.772)	(66.952)	(59.772)	(113.162)
Predicted diversity square	-160.351***		-132.373***	-88.040*	-96.253**	-166.859**
	(53.169)		(42.177)	(49.519)	(43.718)	(79.175)
Log Neolithic transition timing		1.793***	1.636***		1.662***	2.127***
		(0.217)	(0.207)		(0.209)	(0.430)
Log percentage of arable land				0.377**	0.314**	0.348***
				(0.158)	(0.125)	(0.134)
Log absolute latitude				0.190	-0.121	-0.115
				(0.125)	(0.119)	(0.135)
Log land suitability for agriculture				0.160	0.238*	0.210*
				(0.173)	(0.124)	(0.125)
Optimal diversity	0.710***		0.692***	0.734**	0.700***	0.694***
	(0.052)		(0.027)	(0.347)	(0.188)	(0.194)
Continent fixed effects	No	No	No	No	No	Yes
Observations	126	126	126	126	126	126
R^2	0.16	0.42	0.46	0.32	0.59	0.61

Notes: This table establishes the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log population density in 1 CE in an extended 126-country sample while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

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 continent fixed effects, demonstrating that the coefficients associated with the diversity channel in each historical period maintain significance in spite of the lower average degree of cross-country variation in diversity within each continent as compared to that observed worldwide. Moreover, the magnitudes of the diversity coefficients remain rather stable, particularly in the 1000 CE analysis, and increase somewhat in the 1 CE analysis despite the smaller sample size and, hence, even lower within-continent variation in diversity exploited by the latter regression. Further, the estimated optimal levels of diversity in the two periods are relatively stable in comparison to that obtained under the baseline regression for the year 1500 CE. The coefficients associated with diversity from the 1000 CE analysis suggest that, accounting for land productivity, the timing of the Neolithic transition, and continent fixed effects, a 1 percentage point increase in genetic diversity for the least diverse society in the sample would raise its population density by 38 percent, whereas a 1 percentage point decrease in diversity for the most diverse society would raise its population density by 26 percent. On the other hand, for the 1 CE analysis, a similar increase in genetic diversity for the least diverse society would raise its population density by 47 percent, whereas a similar decrease in diversity for the most

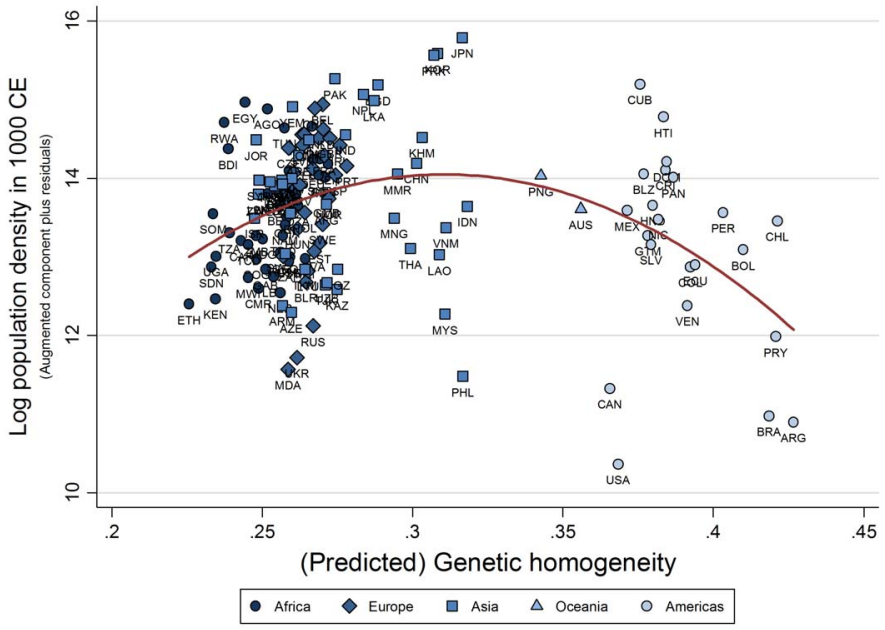


FIGURE A.1: Predicted Genetic Diversity and Population Density in 1000 CE

Notes: This figure depicts the hump-shaped effect, estimated using a least-squares quadratic fit, of predicted genetic homogeneity (i.e., 1 minus genetic diversity as predicted by migratory distance from East Africa) on log population density in 1000 CE in an extended 140-country sample, conditional on the timing of the Neolithic Revolution, land productivity, and continent fixed effects. This figure is an augmented component-plus-residual plot rather than the typical added-variable plot of residuals against residuals. Specifically, the vertical axis represents fitted values (as predicted by genetic homogeneity and its square) of log population density plus the residuals from the full regression model. The horizontal axis, on the other hand, represents genetic homogeneity rather than the residuals obtained from regressing homogeneity on the control variables in the model. This methodology permits the illustration of the overall nonmonotonic effect of genetic homogeneity in one scatter plot.

diverse society would raise its population density by 28 percent.[A.1] The hump-shaped effects, implied by these coefficients, of genetic diversity on log population density in the years 1000 CE and 1 CE are depicted in Figures A.1 and A.2.[A.2]

In sum, the results presented in Tables A.1 and A.2 suggest that, consistent with the predictions of the proposed diversity channel, genetic diversity has indeed been a significant determinant of Malthusian economic development in earlier historical periods as well. The overall nonmonotonic effect of diversity on population density in the years 1000 CE and 1 CE is robust, in terms of both magnitude and statistical significance, to controls for the timing of the agricultural transition, the natural productivity of land for agriculture, and other unobserved continent-specific geographical and socioeconomic characteristics. More fundamentally, the analysis demonstrates the persistence

[A.1] These effects are calculated directly via the methodology outlined in Footnote 31 of the paper, along with the sample minimum and maximum genetic diversity values of 0.573 and 0.774, respectively, in both the 1000 CE and 1 CE regression samples.

[A.2] For consistency with Figure 1 of the paper, which depicts the negative effect of increasing migratory distance from East Africa on genetic diversity, the horizontal axes in Figures A.1–A.2 represent genetic homogeneity (i.e., 1 minus genetic diversity) so as to reflect increasing as opposed to decreasing migratory distance from East Africa.

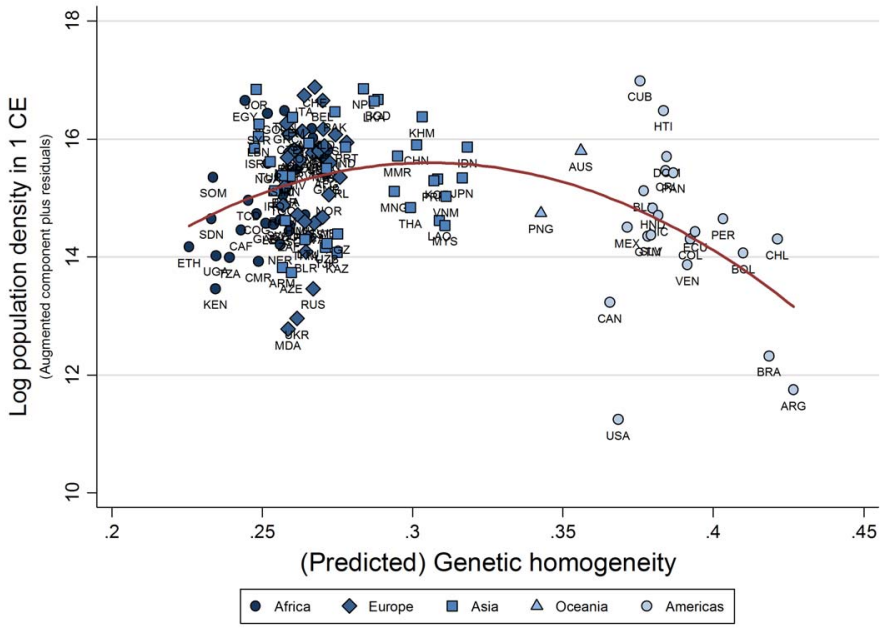


FIGURE A.2: Predicted Genetic Diversity and Population Density in 1 CE

Notes: This figure depicts the hump-shaped effect, estimated using a least-squares quadratic fit, of predicted genetic homogeneity (i.e., 1 minus genetic diversity as predicted by migratory distance from East Africa) on log population density in 1 CE in an extended 126-country sample, conditional on the timing of the Neolithic Revolution, land productivity, and continent fixed effects. This figure is an augmented component-plus-residual plot rather than the typical added-variable plot of residuals against residuals. Specifically, the vertical axis represents fitted values (as predicted by genetic homogeneity and its square) of log population density plus the residuals from the full regression model. The horizontal axis, on the other hand, represents genetic homogeneity rather than the residuals obtained from regressing homogeneity on the control variables in the model. This methodology permits the illustration of the overall nonmonotonic effect of genetic homogeneity in one scatter plot.

of the diversity channel, along with the optimal level of diversity, over a long expanse of time during the agricultural stage of development.

A.2 Robustness to the Technology Diffusion Hypothesis

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

TABLE A.3: The Regional Technological Frontiers of the World, 1–1500 CE

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, U.K.	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

The purpose of the current investigation is to ensure that the analyses conducted in Section IV.B of the paper and in the preceding appendix section 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 prehistoric migratory routes from East Africa (such as Cairo and Istanbul) as origins of spatial technology diffusion during the precolonial 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 earlier regression specifications. As will become evident, however, while technology diffusion is indeed found to have been a significant determinant of comparative development in the precolonial era, the baseline findings for genetic diversity remain robust to controls for this additional influential hypothesis.

To account for the technology diffusion channel, the current analysis constructs, for each historical period examined, a control variable measuring the great circle distance from the closest regional technological frontier in that period. Following the well-accepted notion that the process of preindustrial urban development was typically more pronounced in societies that enjoyed higher agricultural surpluses, the analysis adopts historical city population size as an appropriate metric to

identify the period-specific sets of regional technological frontiers. Specifically, based on historical urban population data from Chandler (1987) and Modelski (2003), the procedure commences with assembling, for each period, a set of regional frontiers comprising the two largest cities, belonging to different civilizations or disparate sociopolitical entities, from each of Africa, Europe, Asia, and the Americas.[A.3] 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 set of regional frontiers obtained for each period as shown in Table A.3.[A.4] 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 smallest of the great circle distances from the regional frontiers to the country's capital city.

To anticipate the robustness of the baseline results for genetic diversity, predicted by migratory distance from East Africa, 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 only weakly correlated with distance from the closest regional technological frontier in each period, with the respective correlation coefficients being -0.04 and 0.03.[A.5] These encouragingly low sample correlations are indicative of the fact that the estimated baseline regression specifications for the historical analysis were, indeed, not simply attributing to genetic diversity the effects possibly arising from the technology diffusion channel.

Column 1 of Table A.4 reports the results from estimating the baseline specification for log population density in 1500 CE while controlling for technology diffusion as originating from the regional frontiers identified for this period. In comparison to the baseline estimates revealed in Column 6 of Table 3 in the paper, the regression coefficients associated with the genetic diversity channel remain relatively stable, decreasing only moderately in magnitude and statistical significance. Some similar robustness characteristics may be noted for the transition timing and land productivity channels as well. Importantly, however, the estimate for the optimal level of diversity remains virtually unchanged and highly statistically significant. Interestingly, the results also establish the technology diffusion channel as a significant determinant of comparative development in the precolonial Malthusian era. In particular, a 1 percent increase in distance from the closest regional

[A.3] The exclusion of Oceania from the list of continents employed is not a methodological restriction but a natural result arising from the fact that evidence of urbanization does not appear in the historical record of this continent until after European colonization. Moreover, the consideration of the Americas as a single unit is consistent with the historical evidence that this landmass only harbored two distinct major civilizational sequences – one in Mesoamerica and the other in the Andean region of South America. Indeed, the imposition of the criteria that the selected cities in each continent (or landmass) should belong to different sociopolitical units is meant to capture the notion that technology diffusion historically occurred due to civilizational influence, broadly defined, as opposed to the influence of only major urban centers that were developed by these relatively advanced societies.

[A.4] Note that, for the year 1 CE, there are four cities appearing within the territories of the Roman Empire, which a priori seems to violate the criterion that the regional frontiers selected should belong to different sociopolitical entities. This, however, 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.

[A.5] These correlations differ slightly from those presented in Table G.4 in Section G of this appendix, where the correlations are presented for the entire 145-country sample used in the regressions for 1500 CE.

TABLE A.4: Robustness to the Technology Diffusion Hypothesis

	(1)	(2)	(3)
	Dependent variable is log population density in:		
	1500 CE	1000 CE	1 CE
Predicted diversity	156.736**	183.771**	215.858**
	(75.572)	(88.577)	(105.286)
Predicted diversity square	-114.626**	-134.609**	-157.724**
	(52.904)	(61.718)	(73.681)
Log Neolithic transition timing	0.909***	1.253***	1.676***
	(0.254)	(0.339)	(0.434)
Log percentage of arable land	0.363***	0.323***	0.342***
	(0.104)	(0.121)	(0.131)
Log absolute latitude	-0.492***	-0.454***	-0.212
	(0.134)	(0.149)	(0.142)
Log land suitability for agriculture	0.275***	0.239**	0.191
	(0.090)	(0.105)	(0.120)
Log distance to regional frontier in 1500 CE	-0.187***		
	(0.070)		
Log distance to regional frontier in 1000 CE		-0.230*	
		(0.121)	
Log distance to regional frontier in 1 CE			-0.297***
			(0.102)
Optimal diversity	0.684***	0.683***	0.684**
	(0.169)	(0.218)	(0.266)
Continent fixed effects	Yes	Yes	Yes
Observations	145	140	126
R^2	0.72	0.64	0.66

Notes: This table establishes, using the extended cross-country sample for each historical time period, that the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log population density in the years 1500 CE, 1000 CE, and 1 CE, conditional on the timing of the Neolithic Revolution, land productivity, and continent fixed effects, is robust to controlling for distance to the closest regional technological frontier in each historical time period. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

frontier is associated with a decrease in population density by 0.2 percent, an effect that is statistically significant at the 1 percent level.

Columns 2 and 3 establish the robustness of the hump-shaped effect of genetic diversity on economic development in 1000 CE and 1 CE to controls for technology diffusion arising from the technological frontiers identified for these earlier historical periods. Specifically, comparing the regression for 1000 CE in Column 2 with its relevant baseline (i.e., Column 6 of Table A.1), the linear and quadratic coefficients associated with genetic diversity remain largely stable under controls for technology diffusion, decreasing slightly in magnitude but maintaining statistical significance. A similar stability pattern also emerges for the coefficients capturing the influence of the diversity channel across the 1 CE regressions. Indeed, the estimates for optimal diversity in these earlier periods remain rather stable relative to their respective baselines in Tables A.1 and A.2. Finally, in

line with the predictions of the technology diffusion hypothesis, a statistically significant negative effect of distance from the closest regional frontier on economic development is observed for these earlier historical periods as well.

The results uncovered herein demonstrate the persistence of the significant hump-shaped effect of genetic diversity on comparative development over the period 1–1500 CE, despite controls for the apparently influential role of technology diffusion arising from the technological frontiers that were relevant during this period of world history. Indeed, these findings lend further credence to the proposed diversity channel by demonstrating that the historical analysis, based on genetic diversity predicted by migratory distance from East Africa, has not been ascribing to genetic diversity the explanatory power that should otherwise be attributed to the impact of spatial technology diffusion.

A.3 Robustness to Microgeographic Factors

This section addresses concerns regarding the possibility that the hump-shaped effect of genetic diversity on precolonial comparative development 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. On the other hand, by geographically isolating population subgroups, a rugged landscape could also have nurtured their ethnic differentiation over time (Michalopoulos, 2011), and may thus confer an adverse effect on society's aggregate productivity via the increased likelihood of ethnic conflict. Similarly, while proximity to waterways can directly affect crop yields by making beneficial practices such as irrigation possible, it may also have augmented productivity indirectly by lowering transportation costs and, thereby, fostering urban development, trade, and technology diffusion.[A.6]

To ensure that the significant hump-shaped effect of genetic diversity on log population density in 1500 CE, as revealed in Section IV.B of the paper, is not simply reflecting the latent influence of microgeographic factors, the current analysis examines variants of the relevant baseline regression specification augmented with controls for terrain quality and proximity to waterways. In particular, the controls for terrain quality are derived from the *G-ECON* data set compiled by Nordhaus (2006), and they include mean elevation and a measure of terrain roughness, both aggregated up to the country level from grid-level data at a granularity of 1-degree latitude by 1-degree longitude. Moreover, in light of the possibility that the impact of terrain undulation could be nonmonotonic, the specifications examined also control for the squared term of the terrain roughness index. The control variables gauging access to waterways, obtained from the data set of Gallup, Sachs and Mellinger (1999), include the expected distance from any point within a country to the nearest coast or sea-

[A.6] Indeed, a significant positive relationship between proximity to waterways and contemporary population density has been demonstrated by Gallup, Sachs and Mellinger (1999).

TABLE A.5: Robustness to Microgeographic Factors

	(1)	(2)	(3)
	Dependent variable is log population density in 1500 CE		
Predicted diversity	160.346**	157.073**	157.059**
	(78.958)	(79.071)	(69.876)
Predicted diversity square	-118.716**	-112.780**	-114.994**
	(55.345)	(55.694)	(48.981)
Log Neolithic transition timing	1.131***	1.211***	1.215***
	(0.225)	(0.201)	(0.197)
Log percentage of arable land	0.397***	0.348***	0.374***
	(0.099)	(0.099)	(0.087)
Log absolute latitude	-0.358***	-0.354***	-0.352***
	(0.124)	(0.132)	(0.122)
Log land suitability for agriculture	0.188*	0.248***	0.160**
	(0.101)	(0.082)	(0.081)
Mean elevation	-0.404		0.502*
	(0.251)		(0.273)
Terrain roughness	5.938***		4.076**
	(1.870)		(1.840)
Terrain roughness square	-7.332**		-7.627***
	(2.922)		(2.906)
Mean distance to nearest waterway		-0.437**	-0.390**
		(0.178)	(0.181)
Percentage of land near a waterway		0.731**	1.175***
		(0.310)	(0.294)
Optimal diversity	0.675***	0.696***	0.683***
	(0.224)	(0.188)	(0.083)
Continent fixed effects	Yes	Yes	Yes
Observations	145	145	145
R^2	0.72	0.75	0.78

Notes: This table establishes, using the extended 145-country sample, that the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log population density in 1500 CE, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, is robust to additional controls for microgeographic factors, including terrain characteristics and access to waterways. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

navigable river and the percentage of a country's land area located near (i.e., within 100 km of) a coast or sea-navigable river.[A.7] Foreshadowing the robustness of the baseline results, mean elevation, terrain roughness, and terrain roughness square possess only moderate correlation coefficients of -0.11, 0.16, and 0.09, respectively, with migratory distance from East Africa. Moreover, migratory distance is also only moderately correlated with the measures of proximity to waterways, possessing sample correlations of -0.20 and 0.19 with the distance and land area variables described above.

[A.7] For completeness, specifications controlling for the squared terms of the other microgeographic factors were also examined. The results from these additional regressions, however, did not reveal any significant nonlinear effects and are therefore not reported.

The results from estimating augmented regression specifications for explaining log population density in 1500 CE, incorporating controls for either terrain quality or access to waterways, are shown in Columns 1 and 2 of Table A.5. In each case, the coefficients associated with the diversity channel remain statistically significant and relatively stable, experiencing only a moderate decrease in magnitude, when compared to the baseline results reported in Column 6 of Table 3 in the paper. Interestingly, the control variables for terrain quality in Column 1 and those gauging access to waterways in Column 2 appear to confer statistically significant effects on population density in 1500 CE, mostly in directions consistent with priors. The results suggest that terrain roughness does indeed have a nonmonotonic impact on aggregate productivity, with the beneficial effects dominating at relatively lower levels of terrain roughness and the detrimental effects dominating at higher levels. Further, regions with greater access to waterways are found to support higher population densities.

The final column of Table A.5 examines the influence of the genetic diversity channel under 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 hump-shaped effect on log population density in 1500 CE, without exhibiting any drastic reductions in the magnitude of its impact. Moreover, the estimate for the optimal level of diversity remains fully intact in comparison to the baseline estimate from Column 6 of Table 3 in the paper. The results uncovered here therefore suggest that the significant nonmonotonic impact of genetic diversity, predicted by migratory distance from East Africa, on log population density in 1500 CE is indeed not a spurious relationship arising from the omission of microgeographic factors as explanatory variables in the baseline regression specification.

A.4 Robustness to Exogenous Factors in the Diamond Hypothesis

This section demonstrates the robustness of the hump-shaped effect of genetic diversity, predicted by migratory distance from East Africa, on precolonial comparative development to additional controls for the Neolithic transition timing channel. In particular, the analysis is intended to alleviate concerns that the significant nonmonotonic impact of genetic diversity presented in Section IV.B of the paper, although estimated while controlling for the timing of the Neolithic Revolution, may still capture some latent influence of this other explanatory channel if correlations exist between migratory distance from East Africa and exogenous factors governing the timing of the Neolithic transition. The results from estimating some extended regression specifications for log population density in 1500 CE, reflecting variants of the baseline specification in equation (8) of the paper that additionally account for the ultimate determinants in the Diamond hypothesis, are presented in Table A.6.

Following the discussion from Section III.C of the paper on the geographic and biogeographic determinants of the Neolithic Revolution, the additional control variables employed by the current analysis include (i) climate, measured as a discrete index with higher integer values assigned to countries in Köppen-Geiger climatic zones that are more favorable to agriculture, (ii) the orientation of the continental axis, measured as the ratio of the largest longitudinal distance to the largest latitudinal distance of the continent or landmass to which a country belongs, (iii) the size of the continent, measured as the total land area of a 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 prehistorically native to the region in which a country belongs.[A.8] These variables are obtained from the data set of Olsson and Hibbs (2005).

Column 1 of Table A.6 presents the results from estimating the baseline specification for log population density in 1500 CE using the restricted 96-country sample of Olsson and Hibbs (2005). Reassuringly, the highly significant coefficients associated with diversity and the other explanatory channels remain rather stable in magnitude relative to their estimates obtained with the unrestricted sample from Column 5 of Table 3 in the paper, 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.[A.9]

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 geographic determinants or the set of biogeographic determinants or both. The results indicate that the coefficients associated with diversity continue to remain highly statistically significant and relatively stable in magnitude in comparison to their baseline estimates from Column 1. Interestingly, when controlling for only the geographic antecedents of the Neolithic Revolution in Column 2, climate alone is significant amongst these additional factors. Likewise, when only the biogeographic antecedents are controlled for in Column 3, the number of domesticable animals rather than plants is significant. In addition, none of the ultimate factors in the Diamond channel possess statistical significance when both geographic and biogeographic determinants are controlled for in Column 4, a result that possibly reflects the high correlations amongst these control variables. Regardless of these tangential issues, however, genetic diversity, as already mentioned, continues to exert a significant hump-shaped effect on precolonial comparative development.

The final column in Table A.6 establishes the robustness of the hump-shaped effect of genetic diversity on log population density 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 determinants of comparative development in the Diamond hypothesis. More importantly,

[A.8] While the influence of the number of domesticable species of plants and animals on the likelihood of the emergence of agriculture is evident, the role of the geographic antecedents of the Neolithic Revolution requires some elaboration. A larger size of the continent or landmass implied greater biodiversity and, hence, a greater likelihood that at least some species suitable for domestication would exist. In addition, a more pronounced East-West (relative to North-South) orientation of the major continental axis meant an easier diffusion of agricultural practices within the landmass, particularly among regions sharing similar latitudes and, hence, similar environments suitable for agriculture. This orientation factor is argued by Diamond (1997) to have played a pivotal role in comparative economic development by favoring the early rise of complex agricultural civilizations on the Eurasian landmass. Finally, certain climates are known to be more beneficial for agriculture than others. For instance, moderate zones encompassing the Mediterranean and Marine 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 influence of these various geographic and biogeographic factors on the timing of the Neolithic Revolution has been established empirically by Olsson and Hibbs (2005) and Putterman (2008).

[A.9] Note that the specifications estimated in the current analysis do not incorporate continent dummies since a sizeable portion of unobserved continent-specific effects are captured by most of the (bio)geographic variables in the Diamond channel that are measured at either the continental or the macro-regional levels. Augmenting the specifications with continent fixed effects, however, does not significantly alter the results for genetic diversity.

TABLE A.6: Robustness to Ultimate Determinants in the Diamond Hypothesis

	(1)	(2)	(3)	(4)	(5)
	Dependent variable is log population density in 1500 CE				
Predicted diversity	216.847***	252.076***	174.414***	212.123***	274.916***
	(62.764)	(71.098)	(62.505)	(70.247)	(73.197)
Predicted diversity square	-154.750***	-180.650***	-125.137***	-151.579***	-197.120***
	(45.680)	(52.120)	(45.568)	(51.463)	(53.186)
Log Neolithic transition timing	1.300***				1.160***
	(0.153)				(0.298)
Log percentage of arable land	0.437***	0.431***	0.441***	0.411***	0.365***
	(0.116)	(0.119)	(0.111)	(0.116)	(0.112)
Log absolute latitude	-0.212**	-0.426***	-0.496***	-0.487***	-0.332**
	(0.102)	(0.131)	(0.154)	(0.163)	(0.145)
Log land suitability for agriculture	0.288**	0.184	0.297**	0.242*	0.280**
	(0.135)	(0.143)	(0.146)	(0.146)	(0.122)
Climate		0.622***		0.419	0.374*
		(0.137)		(0.268)	(0.225)
Orientation of continental axis		0.281		0.040	-0.169
		(0.332)		(0.294)	(0.255)
Size of continent		-0.007		-0.005	-0.006
		(0.015)		(0.013)	(0.012)
Domesticable plants			0.015	-0.005	0.003
			(0.019)	(0.023)	(0.021)
Domesticable animals			0.154**	0.121	-0.013
			(0.063)	(0.074)	(0.073)
Optimal diversity	0.701***	0.698***	0.697***	0.700***	0.697***
	(0.021)	(0.019)	(0.051)	(0.078)	(0.020)
Observations	96	96	96	96	96
R^2	0.74	0.70	0.70	0.72	0.78

Notes: This table establishes, using a feasible 96-country sample, that the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log population density in 1500 CE, while controlling for the timing of the Neolithic Revolution and land productivity, is robust to additional controls for the geographic and biogeographic antecedents of the Neolithic Revolution, including climate, the orientation of the continental axis, the size of the continent, and the numbers of prehistoric domesticable species of plants and animals. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

the linear and quadratic coefficients associated with the diversity channel maintain relative stability, increasing moderately in magnitude when compared to their baseline estimates, but remaining highly statistically significant. Overall, the results in Table A.6 suggest that the baseline estimate of the hump-shaped impact of genetic diversity, presented in Section IV.B of the paper, is indeed not reflecting additional latent effects of the influential agricultural transition timing channel in precolonial comparative development.

B The Index of Contemporary Population Diversity

This section discusses the methodology applied to construct the index of genetic diversity for contemporary national populations such that it additionally accounts for the between-group component of diversity. To this effect, the index makes use of the concept of F_{st} genetic distance from the field of population genetics.

Specifically, for any subpopulation pair, the F_{st} genetic distance between the two subpopulations captures the proportion of their combined genetic diversity that is unexplained by the weighted average of their respective genetic diversities. Consider, for instance, a population comprised of two ethnic groups or subpopulations, A and B. The F_{st} genetic distance between A and B would then be defined as

$$F_{st}^{AB} = 1 - \frac{\theta_A H_{\text{exp}}^A + \theta_B H_{\text{exp}}^B}{H_{\text{exp}}^{AB}}, \tag{B.1}$$

where θ_A and θ_B are the shares of groups A and B, respectively, in the combined population, H_{exp}^A and H_{exp}^B are their respective expected heterozygosities, and H_{exp}^{AB} is the expected heterozygosity of the combined population. Thus, given (i) genetic distance, F_{st}^{AB}, (ii) the expected heterozygosities of the component subpopulations, H_{exp}^A and H_{exp}^B, and (iii) their respective shares in the overall population, θ_A and θ_B, the overall diversity of the combined population is

$$H_{\text{exp}}^{AB} = \frac{\theta_A H_{\text{exp}}^A + \theta_B H_{\text{exp}}^B}{\left(1 - F_{st}^{AB}\right)}. \tag{B.2}$$

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

To surmount this issue, the current analysis appeals to a second prediction of the serial founder effect regarding the genetic differentiation of populations through isolation by geographical distance. Accordingly, in the process of the initial stepwise diffusion of the human species from Africa into the rest of the world, offshoot colonies residing at greater geographical distances from parental ones would also be more genetically differentiated from them. This would arise due to the larger number intervening migration steps and the concomitantly larger number of genetic diversity subsampling events that are associated with offshoots residing at locations farther away from parental colonies. Indeed, this second prediction of the serial founder effect is bourne out in the data as well. Based on data from Ramachandran et al. (2005), Figure B.1 shows the strong positive effect of pairwise

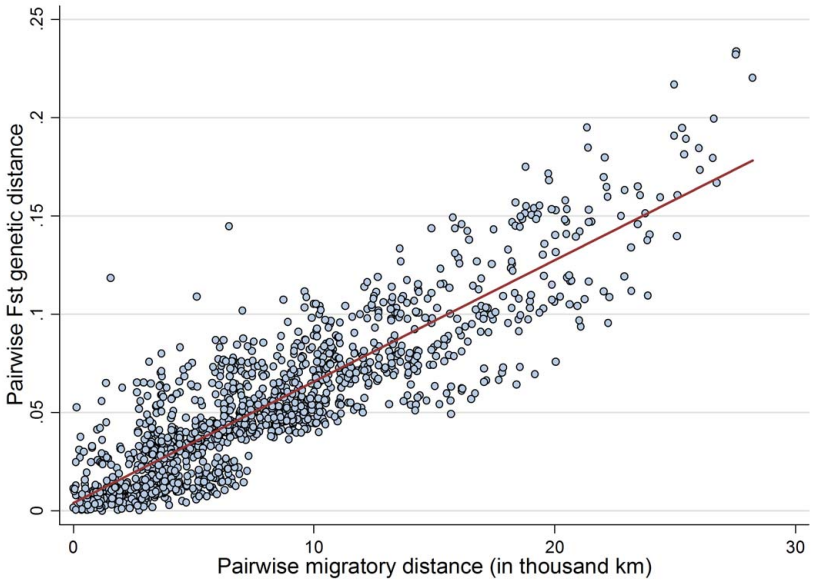


FIGURE B.1: Pairwise F_{st} Genetic Distance and Pairwise Migratory Distance

Notes: This figure depicts the positive impact of pairwise migratory distance on pairwise F_{st} genetic distance across all 1,378 ethnic group pairs from the set of 53 ethnic groups that constitute the HGDP-CEPH Human Genome Diversity Cell Line Panel.

migratory distance on pairwise genetic distance across all pairs of ethnic groups in the HGDP-CEPH sample. Specifically, according to the regression, variation in migratory distance explains 78 percent of the variation in F_{st} genetic distance across the 1,378 ethnic group pairs. Moreover, the estimated OLS coefficient is highly statistically significant, possessing a t-statistic of 53.62, and suggests that pairwise F_{st} genetic distance falls by 0.062 percentage points for every 10,000 km increase in pairwise migratory distance. The construction of the index of genetic diversity for contemporary national populations thus employs F_{st} genetic distance values *predicted* by pairwise migratory distances.

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

$$\hat{H}^{AB}_{\text{exp}} = \frac{\theta_A \hat{H}^{A}_{\text{exp}}(d_A) + \theta_B \hat{H}^{B}_{\text{exp}}(d_B)}{\left[1 - \hat{F}^{AB}_{st}(d_{AB})\right]}, \tag{B.3}$$

where, for $i \in \{A, B\}$, $\hat{H}^{i}_{\text{exp}}(d_i)$ denotes the expected heterozygosity predicted by the migratory distance, d_i, of country i from East Africa (i.e., the predicted genetic diversity of country i in the historical analysis), and θ_i is the contribution of country i, as a result of post-1500 migrations, to the combined population being considered. Moreover, $\hat{F}^{AB}_{st}(d_{AB})$ is the genetic distance predicted by the migratory distance between countries A and B, obtained by applying the coefficients associated with the regression line depicted in Figure B.1. In practice, since contemporary national populations are typically composed of more than two ethnic groups, the procedure outlined in equation (B.3) is

applied recursively in order to incorporate a larger number of component ethnic groups in modern populations.

C Supplementary Figures

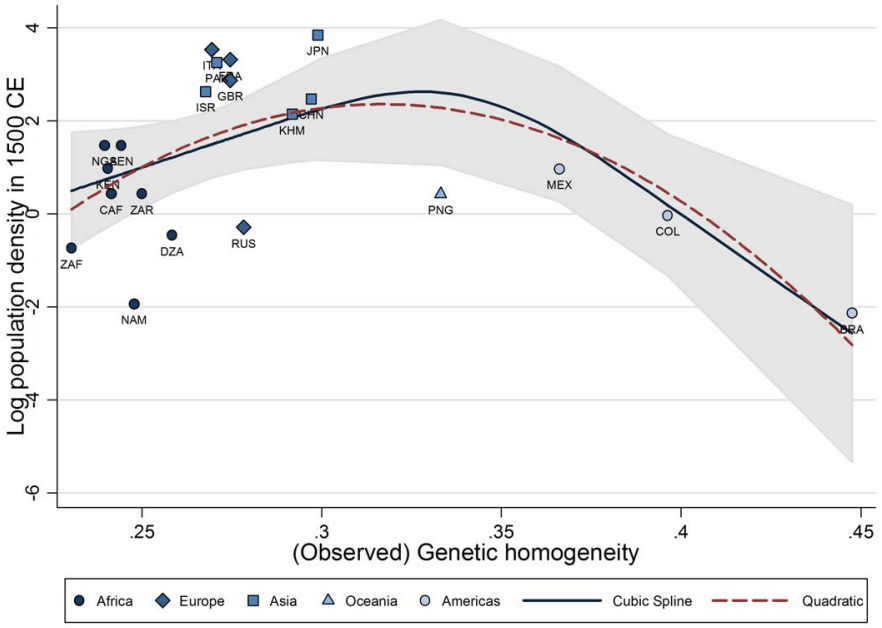
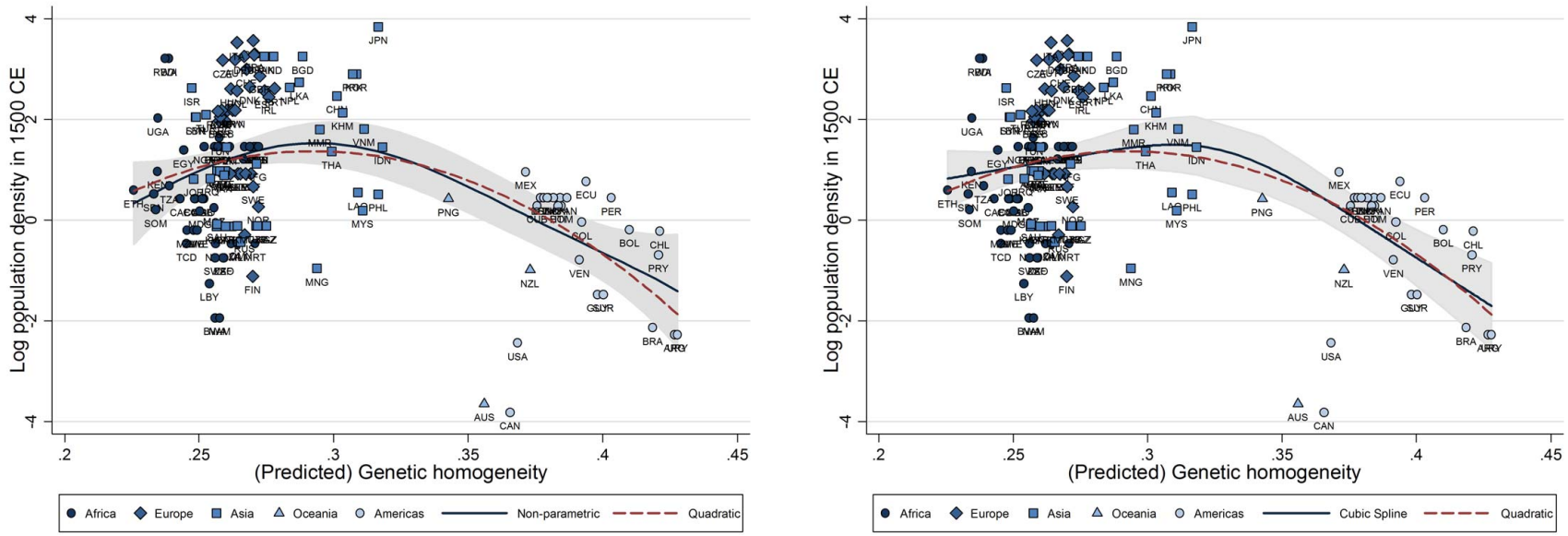


FIGURE C.1: Observed Genetic Diversity and Population Density in 1500 CE – The Unconditional Quadratic and Cubic Spline Relationships

Notes: This figure depicts the unconditional hump-shaped relationship, estimated using either a least-squares quadratic fit or a restricted cubic spline regression, between observed genetic homogeneity (i.e., 1 minus observed genetic diversity) and log population density in 1500 CE in the limited 21-country sample. The restricted cubic spline regression line is estimated using three equally-spaced knots on the domain of observed genetic homogeneity values. The shaded area represents the 95 percent confidence interval band associated with the cubic spline regression line.

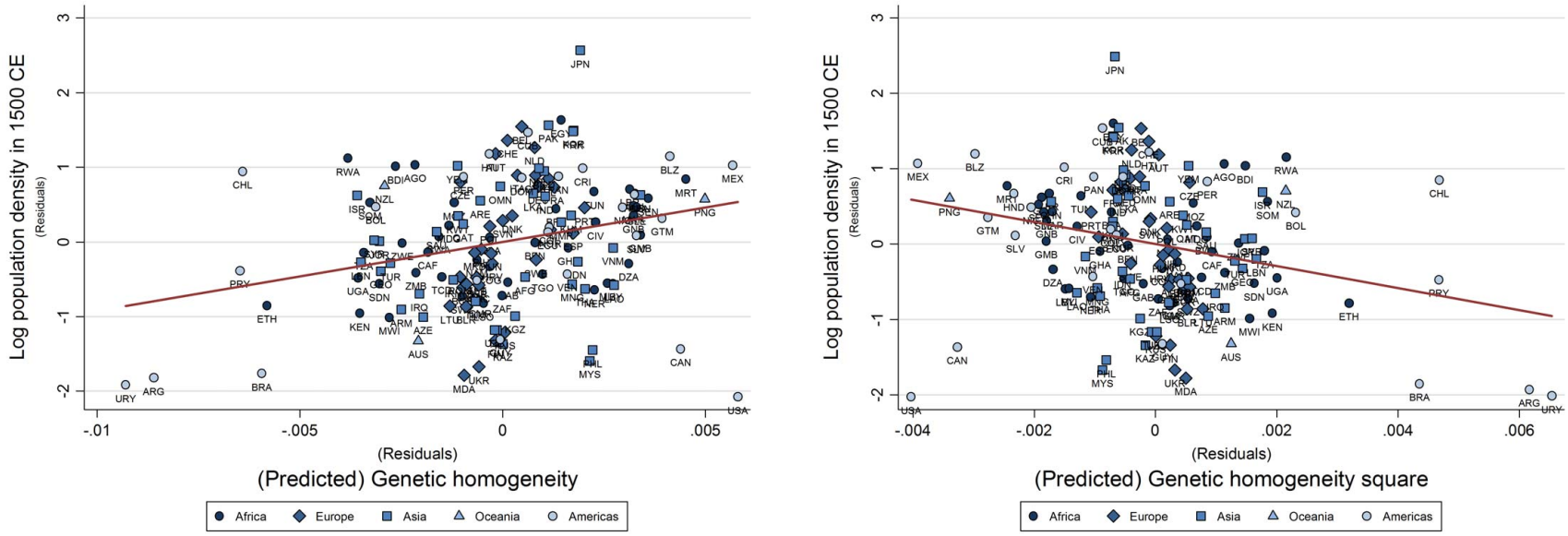


(a) Quadratic vs. Nonparametric

(b) Quadratic vs. Cubic Spline

FIGURE C.2: Predicted Genetic Diversity and Population Density in 1500 CE – The Unconditional Quadratic, Nonparametric, and Cubic Spline Relationships

Notes: This figure depicts the unconditional hump-shaped relationship, based on three different estimation techniques, between predicted genetic homogeneity (i.e., 1 minus genetic diversity as predicted by migratory distance from East Africa) and log population density in 1500 CE in the extended 145-country sample. The relationship is estimated using (i) a least-squares quadratic fit (both panels), (ii) a nonparametric regression (Panel (a)), and (iii) a restricted cubic spline regression (Panel (b)). The nonparametric regression line in Panel (a) is estimated using local 2nd-degree polynomial smoothing based on a Gaussian kernel function and a kernel bandwidth of 0.06. The restricted cubic spline regression line in Panel (b) is estimated using three equally-spaced knots on the domain of predicted genetic homogeneity values. The shaded areas in Panels (a) and (b) represent the 95 percent confidence interval bands associated with the nonparametric and cubic spline regression lines respectively.

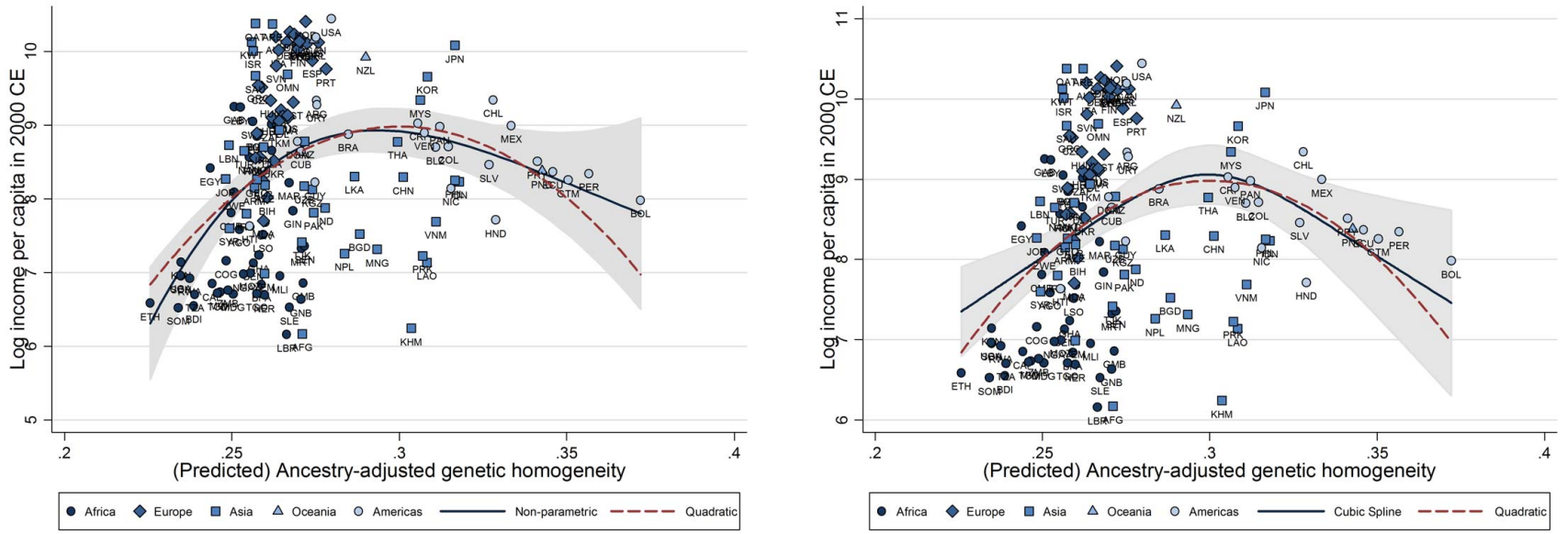


(a) First-Order Partial Effect

(b) Second-Order Partial Effect

FIGURE C.3: Predicted Genetic Diversity and Population Density in 1500 CE – The First- and Second-Order Partial Effects

Notes: This figure depicts the positive first-order partial effect (Panel (a)) and the negative second-order partial effect (Panel (b)) of predicted genetic homogeneity (i.e., 1 minus genetic diversity as predicted by migratory distance from East Africa) on log population density in 1500 CE in the extended 145-country sample, conditional on the timing of the Neolithic Revolution, land productivity, and continent fixed effects. Each panel shows an added-variable plot of residuals against residuals. Thus, the x- and y-axes in Panel (a) plot the residuals obtained from regressing the linear term in genetic homogeneity and log population density, respectively, on the quadratic term in genetic homogeneity as well as the aforementioned set of covariates. Conversely, the x- and y-axes in Panel (b) plot the residuals obtained from regressing the quadratic term in genetic homogeneity and log population density, respectively, on the linear term in genetic homogeneity as well as the aforementioned set of covariates.

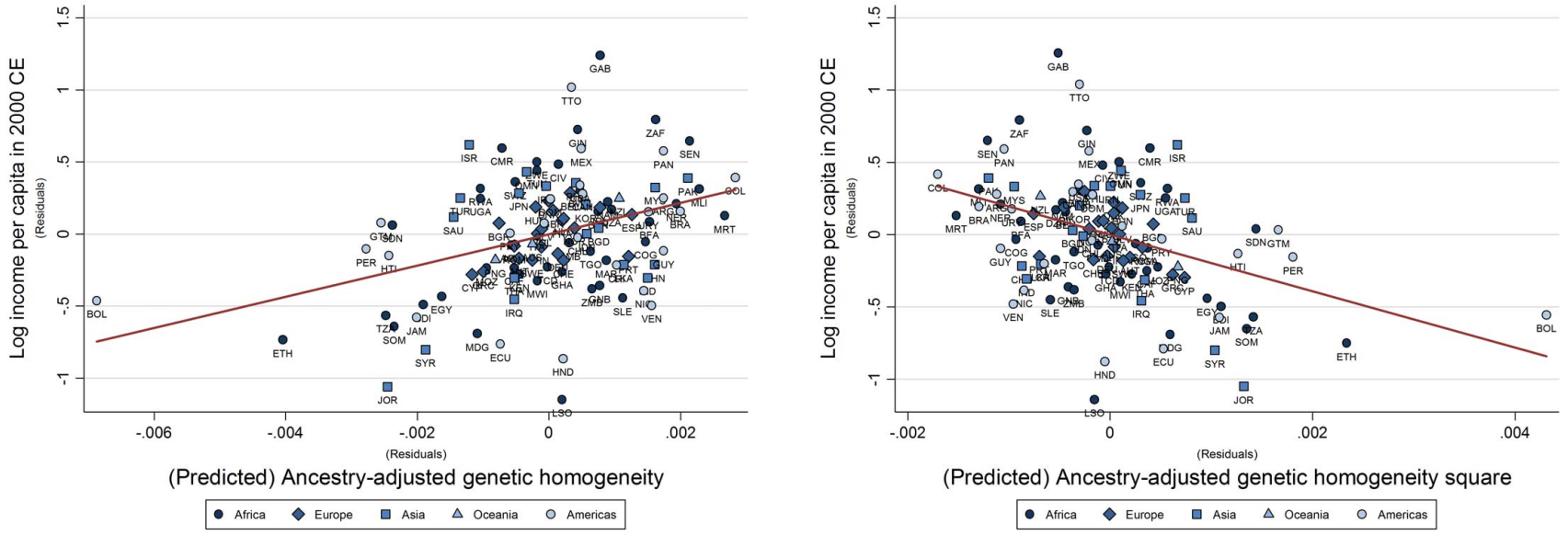


(a) Quadratic vs. Nonparametric

(b) Quadratic vs. Cubic Spline

FIGURE C.4: Ancestry-Adjusted Genetic Diversity and Income Per Capita in 2000 CE – The Unconditional Quadratic, Nonparametric, and Cubic Spline Relationships

Notes: This figure depicts the unconditional hump-shaped relationship, based on three different estimation techniques, between ancestry-adjusted genetic homogeneity (i.e., 1 minus ancestry-adjusted genetic diversity) and log income per capita in 2000 CE in a 143-country sample. The relationship is estimated using (i) a least-squares quadratic fit (both panels), (ii) a nonparametric regression (Panel (a)), and (iii) a restricted cubic spline regression (Panel (b)). The nonparametric regression line in Panel (a) is estimated using local 2nd-degree polynomial smoothing based on a Gaussian kernel function and a kernel bandwidth of 0.06. The restricted cubic spline regression line in Panel (b) is estimated using three equally-spaced knots on the domain of ancestry-adjusted genetic homogeneity values. The shaded areas in Panels (a) and (b) represent the 95 percent confidence interval bands associated with the nonparametric and cubic spline regression lines respectively.

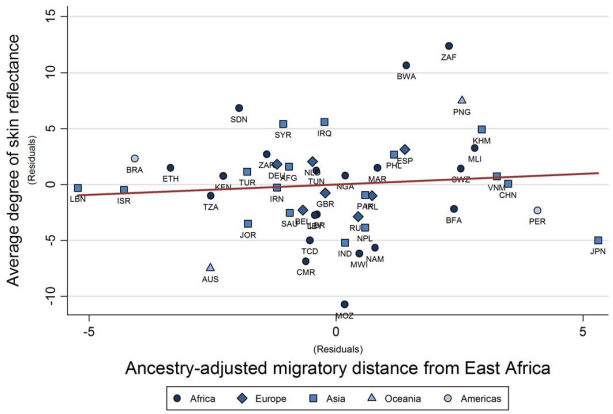


(a) First-Order Partial Effect

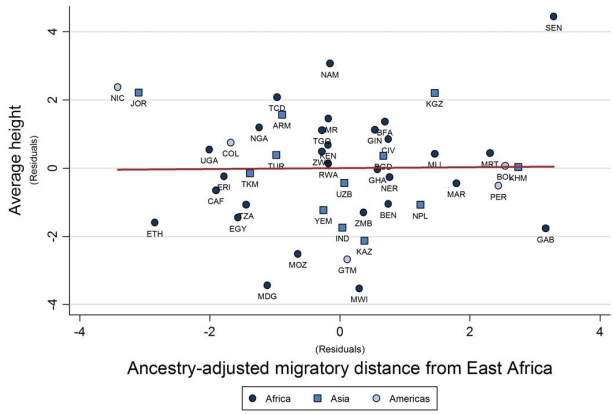
(b) Second-Order Partial Effect

FIGURE C.5: Ancestry-Adjusted Genetic Diversity and Income Per Capita in 2000 CE – The First- and Second-Order Partial Effects

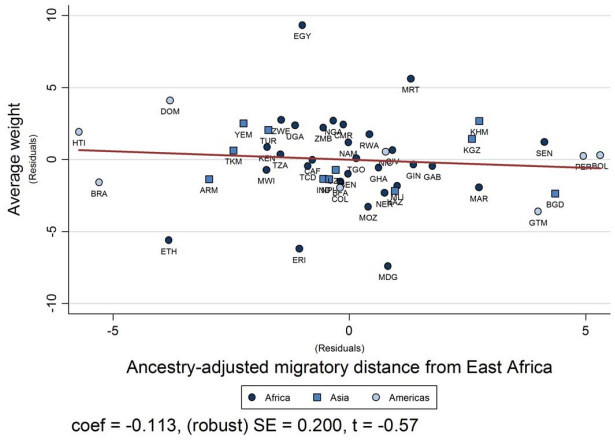
Notes: This figure depicts the positive first-order partial effect (Panel (a)) and the negative second-order partial effect (Panel (b)) of ancestry-adjusted genetic homogeneity (i.e., 1 minus ancestry-adjusted genetic diversity) on log income per capita in 2000 CE in a 109-country sample, conditional on the ancestry-adjusted timing of the Neolithic Revolution, land productivity, a vector of institutional, cultural, and geographical determinants of development, and continent fixed effects. Each panel shows an added-variable plot of residuals against residuals. Thus, the x- and y-axes in Panel (a) plot the residuals obtained from regressing the linear term in genetic homogeneity and log income per capita, respectively, on the quadratic term in genetic homogeneity as well as the aforementioned set of covariates. Conversely, the x- and y-axes in Panel (b) plot the residuals obtained from regressing the quadratic term in genetic homogeneity and log income per capita, respectively, on the linear term in genetic homogeneity as well as the aforementioned set of covariates.



(a) Migratory Distance and Skin Reflectance



(b) Migratory Distance and Height



(c) Migratory Distance and Weight

FIGURE C.6: Ancestry-Adjusted Migratory Distance from East Africa and Some Mean Physiological Characteristics of Contemporary National Populations

Notes: This figure depicts that ancestry-adjusted migratory distance from East Africa has no systematic relationship with some mean physiological characteristics of contemporary national populations, including the average degree of skin reflectance (Panel (a)), average height (Panel (b)), and average weight (Panel (c)), conditional on the intensity of ultraviolet exposure, absolute latitude, the percentage of arable land, the shares of land in tropical (or subtropical) and temperate zones, mean elevation, access to waterways, and continent fixed effects. Each panel shows an added-variable plot of residuals against residuals. Thus, the x- and y-axes in each panel plot the residuals obtained from regressing, respectively, ancestry-adjusted migratory distance and the relevant physiological characteristic (i.e., average skin reflectance, average height, or average weight) on the aforementioned set of covariates.

D Supplementary Results

TABLE D.1: Robustness of the Role of Migratory Distance in the Serial Founder Effect

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is observed genetic diversity					
Migratory distance from East Africa	-0.799***	-0.826***	-0.798***	-0.796***	-0.798***	-0.690***
	(0.054)	(0.062)	(0.066)	(0.072)	(0.089)	(0.148)
Absolute latitude		-0.016	-0.003	-0.004	-0.003	0.074
		(0.015)	(0.017)	(0.015)	(0.022)	(0.045)
Percentage of arable land		-0.015	-0.013	-0.009	-0.010	0.002
		(0.026)	(0.031)	(0.028)	(0.040)	(0.045)
Mean land suitability for agriculture		1.937	-1.244	-0.795	-0.904	1.370
		(1.507)	(5.400)	(4.803)	(6.260)	(5.330)
Range of land suitability			-1.175	-1.594	-1.477	-2.039
			(4.564)	(4.364)	(5.789)	(5.715)
Land suitability Gini			-3.712	-3.767	-3.805	-4.103
			(4.774)	(4.402)	(4.805)	(4.165)
Mean elevation				0.937	0.918	-2.457
				(2.352)	(2.393)	(1.567)
Standard deviation of elevation				-0.129	-0.112	3.418
				(2.284)	(2.288)	(2.137)
Mean distance to nearest waterway					-0.044	0.503
					(1.153)	(0.982)
Continent fixed effects	No	No	No	No	No	Yes
Observations	21	21	21	21	21	21
R^2	0.94	0.95	0.95	0.96	0.96	0.98
Partial R^2 of migratory distance		0.94	0.93	0.93	0.90	0.81

Notes: Using the limited 21-country sample, this table (i) establishes that the significant negative effect of migratory distance from East Africa on observed genetic diversity is robust to controls for geographical factors linked to ethnic diversity (Michalopoulos, 2011), including absolute latitude, the percentage of arable land, the mean, range, and Gini coefficient of the distribution of land suitability for agriculture, the mean and standard deviation of the distribution of elevation, access to waterways, and continent fixed effects, and (ii) demonstrates that these geographical factors have little or no explanatory power for the cross-country variation in observed genetic diversity beyond that accounted for by migratory distance. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.2: The Results of Table 1 with Corrections for Spatial Autocorrelation

	(1)	(2)	(3)	(4)	(5)
	Dependent variable is log population density in 1500 CE				
Observed diversity	413.504***			225.440***	203.814***
	[85.389]			[55.428]	[65.681]
Observed diversity square	-302.647***			-161.158***	-145.717***
	[64.267]			[42.211]	[53.562]
Log Neolithic transition timing		2.396***		1.214***	1.135***
		[0.249]		[0.271]	[0.367]
Log percentage of arable land			0.730***	0.516***	0.545***
			[0.263]	[0.132]	[0.178]
Log absolute latitude			0.145	-0.162*	-0.129
			[0.180]	[0.084]	[0.101]
Log land suitability for agriculture			0.734*	0.571**	0.587**
			[0.376]	[0.240]	[0.233]
Continent fixed effects	No	No	No	No	Yes
Observations	21	21	21	21	21
R^2	0.42	0.54	0.57	0.89	0.90

Notes: This table establishes that the significant hump-shaped relationship between observed genetic diversity and log population density in 1500 CE in the limited 21-country sample, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, is robust to accounting for spatial autocorrelation across observations. Standard errors corrected for spatial autocorrelation, following Conley (1999), are reported in brackets. To perform this correction, the spatial distribution of observations is specified on the Euclidean plane using aerial distances between all pairs of observations in the sample, and the autocorrelation is modeled as declining linearly away from each observation up to a threshold of 5,000 km. This threshold effectively excludes spatial interactions between the Old World and the New World, which is appropriate given the historical period being considered. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.3: The Results of Table 2 with Corrections for Spatial Autocorrelation

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	Spatial GMM (5)	Spatial GMM (6)
	Dependent variable is log population density in 1500 CE					
Observed diversity		255.219***		361.420***	285.191***	243.108***
		[77.933]		[108.692]	[81.538]	[55.325]
Observed diversity square		-209.808***		-268.514***	-206.577***	-179.579***
		[58.315]		[77.740]	[61.906]	[44.271]
Migratory distance	0.505***	0.070				
	[0.110]	[0.138]				
Migratory distance square	-0.023***	-0.014*				
	[0.004]	[0.008]				
Mobility index			0.353***	0.051		
			[0.108]	[0.125]		
Mobility index square			-0.012***	-0.003		
			[0.003]	[0.005]		
Log Neolithic transition timing					1.014***	1.119***
					[0.356]	[0.370]
Log percentage of arable land					0.608***	0.634***
					[0.187]	[0.202]
Log absolute latitude					-0.209**	-0.133
					[0.107]	[0.099]
Log land suitability for agriculture					0.494**	0.549**
					[0.227]	[0.242]
Continent fixed effects	No	No	No	No	No	Yes
Observations	21	21	18	18	21	21
R^2	0.34	0.46	0.30	0.43	–	–

Notes: This table establishes that the results from (i) performing an informal identification test, which demonstrates that the significant unconditional hump-shaped impact of migratory distance from East Africa on log population density in 1500 CE in the limited 21-country sample almost entirely reflects the latent influence of observed genetic diversity, and (ii) employing migratory distance as an excluded instrument for observed genetic diversity to establish the causal hump-shaped effect of genetic diversity on log population density in 1500 CE in the limited 21-country sample, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, are robust to accounting for spatial autocorrelation across observations. Columns 5–6 present the results from estimating the corresponding 2SLS specifications in Table 2 of the paper using Conley's (1999) spatial GMM estimation procedure. Standard errors corrected for spatial autocorrelation, following Conley (1999), are reported in brackets. To perform this correction, the spatial distribution of observations is specified on the Euclidean plane using aerial distances between all pairs of observations in the sample, and the autocorrelation is modeled as declining linearly away from each observation up to a threshold of 5,000 km. This threshold effectively excludes spatial interactions between the Old World and the New World, which is appropriate given the historical period being considered. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.4: Robustness to Log-Log Specifications

	Limited-sample historical analysis		Extended-sample historical analysis			Contemporary analysis
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is:					
	Log population density in 1500 CE		Log population density in 1500 CE	Log population density in 1000 CE	Log population density in 1 CE	Log income per capita in 2000 CE
Log diversity	451.066**	415.864*	423.298**	420.595**	481.771**	511.650***
	(151.797)	(190.651)	(170.963)	(199.565)	(239.630)	(183.412)
Squared log diversity	-425.579**	-396.906*	-407.990**	-402.344**	-458.413**	-476.125***
	(150.883)	(204.655)	(159.515)	(185.258)	(222.306)	(172.285)
Log Neolithic transition timing	1.249***	1.177*	1.241***	1.606***	2.132***	0.062
	(0.370)	(0.618)	(0.232)	(0.265)	(0.425)	(0.260)
Log percentage of arable land	0.510***	0.417*	0.392***	0.370***	0.348***	-0.122
	(0.164)	(0.221)	(0.105)	(0.116)	(0.133)	(0.104)
Log absolute latitude	-0.155	-0.123	-0.413***	-0.368***	-0.110	0.173
	(0.131)	(0.170)	(0.124)	(0.127)	(0.137)	(0.126)
Log land suitability for agriculture	0.571*	0.674	0.256***	0.189*	0.209	-0.176*
	(0.296)	(0.407)	(0.099)	(0.108)	(0.129)	(0.098)
Continent fixed effects	No	Yes	Yes	Yes	Yes	Yes
Observations	21	21	145	140	126	143
R^2	0.89	0.91	0.69	0.62	0.62	0.57

Notes: This table establishes that, in both the limited- and extended-sample variants of the historical analysis as well as in the contemporary analysis, the hump-shaped effect of genetic diversity on economic development, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, remains qualitatively robust under quadratic specifications in logged genetic diversity, thereby demonstrating that the true relationship between genetic diversity and economic development, as examined in the paper, is indeed hump shaped rather than linear in log-transformed variables. The relevant measures of genetic diversity employed by the analysis are observed genetic diversity in Columns 1–2, predicted genetic diversity (i.e., genetic diversity as predicted by migratory distance from East Africa) in Columns 3–5, and ancestry-adjusted genetic diversity in Column 6. The ancestry-adjusted measure of the timing of the Neolithic Revolution is used in Column 6. Heteroskedasticity robust standard errors are reported in parentheses in Columns 1–2. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses in Columns 3–6. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.5: Robustness to Using Genetic Diversity Predicted by the Human Mobility Index

	Extended-sample historical analysis			Contemporary analysis
	(1)	(2)	(3)	(4)
	Dependent variable is:			
	Log population density in 1500 CE	Log population density in 1000 CE	Log population density in 1 CE	Log income per capita in 2000 CE
Diversity	152.163***	146.551**	196.145**	175.130**
	(54.100)	(64.082)	(80.093)	(68.279)
Diversity square	-113.298***	-110.752**	-148.569***	-120.270**
	(37.707)	(44.392)	(55.513)	(47.998)
Log Neolithic transition timing	1.553***	2.052***	3.242***	0.045
	(0.255)	(0.265)	(0.349)	(0.270)
Log percentage of arable land	0.362***	0.314**	0.182	-0.146
	(0.111)	(0.128)	(0.126)	(0.110)
Log absolute latitude	-0.489***	-0.401***	-0.087	0.192
	(0.135)	(0.146)	(0.129)	(0.142)
Log land suitability for agriculture	0.256**	0.221*	0.301***	-0.147
	(0.101)	(0.118)	(0.116)	(0.106)
Continent fixed effects	Yes	Yes	Yes	Yes
Observations	127	123	113	125
R^2	0.71	0.66	0.69	0.55

Notes: This table establishes that, in both the extended-sample historical analysis and the contemporary analysis, the hump-shaped effect of genetic diversity on economic development, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, remains qualitatively robust when the human mobility index, as opposed to the baseline waypoints-restricted migratory distance concept of Ramachandran et al. (2005), is used to predict genetic diversity for the extended-sample historical analysis and to construct the ancestry-adjusted measure of genetic diversity for the contemporary analysis. The relevant measures of genetic diversity employed by the analysis are genetic diversity predicted by the human mobility index in Columns 1–3 and its ancestry-adjusted counterpart in Column 4. The ancestry-adjusted measure of the timing of the Neolithic Revolution is used in Column 4. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. For additional details on the human mobility index and how it is used to either predict genetic diversity for the extended-sample historical analysis or to construct the ancestry-adjusted measure of genetic diversity for the contemporary analysis, the reader is referred to the definitions of these variables in Section F of this appendix. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.6: The Zeroth- and First-Stage Results of the 2SLS Regressions in Table 2

	2SLS regression in Column 5, Table 2			2SLS regression in Column 6, Table 2		
	Zeroth stage	First stage		Zeroth stage	First stage	
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is:			Dependent variable is:		
	Observed diversity	Observed diversity	Observed diversity square	Observed diversity	Observed diversity	Observed diversity square
Excluded instruments:						
Migratory distance from East Africa	-0.814***	-5.143***	-5.841***	-0.657***	-3.992***	-4.234***
	(0.064)	(0.686)	(0.959)	(0.196)	(0.850)	(1.053)
Predicted diversity square (from zeroth stage)	–	-3.937***	-4.327***	–	-3.699***	-3.752***
		(0.628)	(0.880)		(0.916)	(1.145)
Second stage controls:						
Log Neolithic transition timing	-0.017**	-0.119***	-0.139***	-0.013**	-0.082***	-0.088***
	(0.006)	(0.016)	(0.023)	(0.004)	(0.017)	(0.021)
Log percentage of arable land	0.000	0.004**	0.007**	0.002	0.015***	0.017***
	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)	(0.005)
Log absolute latitude	0.003*	0.018***	0.020***	0.005*	0.027***	0.028***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.006)	(0.007)
Log land suitability for agriculture	0.005	0.032***	0.035***	0.005	0.031***	0.032***
	(0.005)	(0.004)	(0.006)	(0.005)	(0.007)	(0.009)
Continent fixed effects	No	No	No	Yes	Yes	Yes
Observations	21	21	21	21	21	21
R^2	0.96	0.99	0.99	0.98	0.99	0.99

Notes: This table presents the zeroth- and first-stage results of the 2SLS regressions from Columns 5–6 of Table 2 in the paper that employ migratory distance as an excluded instrument for observed genetic diversity to establish the causal hump-shaped effect of genetic diversity on log population density in 1500 CE in the limited 21-country sample while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects. Since the second stage of each of these 2SLS regressions is quadratic in the endogenous regressor, it is necessary to instrument for both genetic diversity *and its squared term* in order for the system to be exactly identified. Thus, following Wooldridge (2010, pp. 267–268), the instrumentation technique introduces a zeroth stage to the analysis where genetic diversity is first regressed on migratory distance and all the second-stage controls to obtain predicted (i.e., fitted) values of diversity. The *predicted* genetic diversity from the zeroth stage is squared, and this squared term is then used as an excluded instrument in the second stage along with migratory distance. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.7: Robustness to the Ecological Components of Land Suitability for Agriculture

	Limited-sample historical analysis		Extended-sample historical analysis			Contemporary analysis
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is:					
	Log population density in 1500 CE		Log population density in 1500 CE	Log population density in 1000 CE	Log population density in 1 CE	Log income per capita in 2000 CE
Diversity	236.192**	242.583**	171.881**	181.681**	195.610**	272.915***
	(97.554)	(92.470)	(77.885)	(88.465)	(98.917)	(95.293)
Diversity square	-166.679**	-171.668**	-122.687**	-129.288**	-137.814**	-192.826***
	(73.114)	(74.712)	(55.168)	(62.044)	(69.331)	(67.769)
Log Neolithic transition timing	1.565***	1.689*	1.214***	1.526***	1.839***	-0.024
	(0.378)	(0.866)	(0.225)	(0.244)	(0.357)	(0.286)
Log percentage of arable land	0.547**	0.539	0.382***	0.338***	0.199*	-0.135
	(0.192)	(0.313)	(0.091)	(0.093)	(0.104)	(0.089)
Log absolute latitude	0.183	0.221	-0.137	-0.124	0.292*	0.067
	(0.160)	(0.289)	(0.166)	(0.170)	(0.150)	(0.164)
Log temperature	1.520**	1.423	1.508***	1.971***	3.092***	-0.289
	(0.591)	(0.993)	(0.535)	(0.558)	(0.580)	(0.461)
Log precipitation	0.714***	0.767*	0.358**	0.197	0.176	-0.235
	(0.224)	(0.372)	(0.153)	(0.163)	(0.182)	(0.155)
Log soil fertility	0.162	0.171	0.473	0.541	0.772**	-0.262
	(0.554)	(0.579)	(0.303)	(0.329)	(0.359)	(0.297)
Continent fixed effects	No	Yes	Yes	Yes	Yes	Yes
Observations	21	21	145	140	126	143
R^2	0.93	0.93	0.72	0.67	0.72	0.57

Notes: This table establishes that, in both the limited- and extended-sample variants of the historical analysis as well as in the contemporary analysis, the hump-shaped effect of genetic diversity on economic development, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, remains qualitatively robust when controls for some of the individual ecological components of the suitability of land for agriculture, including temperature, precipitation, and soil fertility, are used in lieu of the baseline control for the overall index of land suitability. The relevant measures of genetic diversity employed by the analysis are observed genetic diversity in Columns 1–2, predicted genetic diversity (i.e., genetic diversity as predicted by migratory distance from East Africa) in Columns 3–5, and ancestry-adjusted genetic diversity in Column 6. The ancestry-adjusted measure of the timing of the Neolithic Revolution is used in Column 6. Heteroskedasticity robust standard errors are reported in parentheses in Columns 1–2. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses in Columns 3–6. For additional details on the individual ecological components of the suitability of land for agriculture, the reader is referred to the definitions of these variables in Section F of this appendix. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.8: Robustness to the Mode of Subsistence Prevalent in 1000 CE

	Limited-sample historical analysis		Extended-sample 1500 CE analysis		Extended-sample 1000 CE analysis	
	Baseline specification (1)	Augmented specification (2)	Baseline specification (3)	Augmented specification (4)	Baseline specification (5)	Augmented specification (6)
	Dependent variable is:					
	Log population density in 1500 CE		Log population density in 1500 CE		Log population density in 1000 CE	
Diversity	186.515**	180.340***	194.141**	196.984***	200.006**	241.019***
	(67.317)	(51.275)	(76.265)	(56.215)	(96.042)	(60.902)
Diversity square	-131.922**	-129.882***	-142.349***	-144.055***	-145.722**	-173.861***
	(51.717)	(39.681)	(53.585)	(40.161)	(66.585)	(42.845)
Log Neolithic transition timing	1.401	0.867	1.107***	1.013***	1.549***	1.388***
	(0.826)	(0.849)	(0.254)	(0.210)	(0.320)	(0.252)
Log subsistence mode in 1000 CE		1.678***		2.030***		2.543**
		(0.465)		(0.730)		(0.996)
Log percentage of arable land	0.504**	0.453**	0.421***	0.458***	0.384***	0.447***
	(0.191)	(0.150)	(0.101)	(0.097)	(0.123)	(0.114)
Log absolute latitude	-0.172	-0.158	-0.420***	-0.332***	-0.369***	-0.288**
	(0.150)	(0.155)	(0.128)	(0.115)	(0.141)	(0.121)
Log land suitability for agriculture	0.706**	0.413	0.226**	0.162*	0.173	0.067
	(0.298)	(0.269)	(0.098)	(0.094)	(0.111)	(0.100)
Continent fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20	20	140	140	135	135
R^2	0.90	0.92	0.67	0.72	0.58	0.67

Notes: This table establishes that, in both the limited-sample historical analysis and the extended-sample analyses for the years 1500 CE and 1000 CE, the hump-shaped effect of genetic diversity on log population density, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, is robust to an additional control for the mode of subsistence (i.e., hunting and gathering versus sedentary agriculture) practiced during the precolonial era. Given underlying data availability constraints on constructing a proxy for the mode of subsistence prevalent in the year 1500 CE, coupled with the fact that cross-country subsistence patterns in 1000 CE should be expected to be highly correlated with those that existed in 1500 CE, the analysis controls for the mode of subsistence prevalent in the year 1000 CE in augmented specifications explaining log population density in both time periods. Moreover, since data on the mode of subsistence prevalent in the year 1000 CE are unavailable for some observations in the baseline regression samples for the limited and extended-sample historical analyses, the table also presents the results from estimating the baseline specifications for these analyses using the corresponding subsistence mode data-restricted samples. This permits a more fair assessment of reductions in omitted variable bias arising from the omission of the subsistence mode control from the baseline specifications. The relevant measures of genetic diversity employed by the analysis are observed genetic diversity in Columns 1–2 and predicted genetic diversity (i.e., genetic diversity as predicted by migratory distance from East Africa) in Columns 3–6. Heteroskedasticity robust standard errors are reported in parentheses in Columns 1–2. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses in Columns 3–6. For additional details on how the proxy for the mode of subsistence prevalent in the year 1000 CE is constructed, the reader is referred to the definition of this variable in Section F of this appendix. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.9: Ancestry-Adjusted Migratory Distance versus Alternative Distances

	(1)	(2)	(3)	(4)
	Dependent variable is log income per capita in 2000 CE			
Migratory distance	0.588***	0.488***	0.502***	0.528**
(ancestry adjusted)	(0.074)	(0.129)	(0.157)	(0.230)
Migratory distance square	-0.029***	-0.025***	-0.026***	-0.026***
(ancestry adjusted)	(0.004)	(0.006)	(0.007)	(0.009)
Migratory distance		0.077		
(unadjusted)		(0.088)		
Migratory distance square		-0.002		
(unadjusted)		(0.003)		
Aerial distance			0.096	
(unadjusted)			(0.198)	
Aerial distance square			-0.004	
(unadjusted)			(0.011)	
Aerial distance				0.097
(ancestry adjusted)				(0.328)
Aerial distance square				-0.006
(ancestry adjusted)				(0.018)
Observations	109	109	109	109
R^2	0.28	0.29	0.29	0.28

Notes: This table (i) establishes a significant unconditional hump-shaped impact of ancestry-adjusted migratory distance from East Africa (i.e., the weighted average of the migratory distances of a country's precolonial ancestral populations) on log income per capita in 2000 CE, (ii) confirms that this nonmonotonic effect is robust to controls for alternative concepts of distance, including (a) the unadjusted measure of migratory distance from East Africa, (b) aerial or "as the crow flies" distance from East Africa, and (c) ancestry-adjusted aerial distance from East Africa, and (iii) demonstrates that these alternative concepts of distance do not possess any systematic relationship, hump-shaped or otherwise, with log income per capita in 2000 CE, conditional on accounting for the nonmonotonic effect of ancestry-adjusted migratory distance. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.10: Robustness to F_{st} Genetic Distances to East Africa and the World Technological Frontier

	Limited-sample historical analysis		Extended-sample historical analysis		Contemporary analysis	
	Baseline specification (1)	Augmented specification (2)	Baseline specification (3)	Augmented specification (4)	Baseline specification (5)	Augmented specification (6)
	Dependent variable is:					
	Log population density in 1500 CE		Log population density in 1500 CE		Log income per capita in 2000 CE	
Diversity	184.672**	201.230**	194.681**	194.743**	186.863**	208.896**
	(63.338)	(65.970)	(81.638)	(85.603)	(83.678)	(96.230)
Diversity square	-130.466**	-144.435**	-142.204**	-145.347**	-129.008**	-148.170**
	(48.774)	(50.450)	(57.192)	(60.802)	(59.424)	(68.752)
Log Neolithic transition timing	1.150*	1.044	1.236***	1.025***	-0.220	-0.128
	(0.634)	(0.915)	(0.242)	(0.312)	(0.287)	(0.344)
Log percentage of arable land	0.433**	0.314	0.392***	0.362***	-0.158	-0.137
	(0.172)	(0.396)	(0.099)	(0.107)	(0.100)	(0.102)
Log absolute latitude	-0.138	-0.160	-0.419***	-0.477***	0.066	0.077
	(0.150)	(0.184)	(0.122)	(0.128)	(0.118)	(0.124)
Log land suitability for agriculture	0.636*	0.778	0.259***	0.324***	-0.146	-0.148
	(0.294)	(0.559)	(0.095)	(0.097)	(0.093)	(0.092)
Genetic distance to the U.K. (1500 match)		-0.014		-0.027		
		(0.049)		(0.027)		
Genetic distance to Ethiopia (1500 match)		-0.047		-0.069		
		(0.084)		(0.043)		
Genetic distance to the U.S. (weighted)						0.030
						(0.037)
Genetic distance to Ethiopia (weighted)						-0.076*
						(0.044)
Continent fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20	20	142	142	139	139
R^2	0.91	0.91	0.69	0.71	0.61	0.62

Notes: This table establishes that, in both the limited- and extended-sample variants of the historical analysis as well as in the contemporary analysis, the hump-shaped effect of genetic diversity on economic development, while controlling for the timing of the Neolithic Revolution, land productivity, and continent fixed effects, is robust to additional controls for F_{st} genetic distance to East Africa (i.e., Ethiopia) and to the world technological frontier relevant for the time period being analyzed (i.e., the U.K. in 1500 CE and the U.S. in 2000 CE), thereby demonstrating that the hump-shaped effect of genetic diversity is not reflecting the latent impact of these variables via channels related to the diffusion of development (Spolaore and Wacziarg, 2009). Since data on the relevant genetic distance measures are unavailable for some observations in the baseline regression samples for the limited- and extended-sample variants of the historical analysis as well as the contemporary analysis, the table also presents the results from estimating the baseline specifications for these analyses using the corresponding genetic distance data-restricted samples. This permits a more fair assessment of reductions in omitted variable bias arising from the omission of the genetic distance controls from the baseline specifications. The relevant measures of genetic diversity employed by the analysis are observed genetic diversity in Columns 1–2, predicted genetic diversity (i.e., genetic diversity as predicted by migratory distance from East Africa) in Columns 3–4, and ancestry-adjusted genetic diversity in Columns 5–6. The ancestry-adjusted measure of the timing of the Neolithic Revolution is used in Columns 5–6. Heteroskedasticity robust standard errors are reported in parentheses in Columns 1–2. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses in Columns 3–6. For additional details on the genetic distance measures employed by the analysis, the reader is referred to the definitions of these variables in Section F of this appendix. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.11: Robustness to Region Fixed Effects and Subsampling on Regional Clusters

	Full sample (1)	Omit Sub-Saharan African countries (2)	Omit Latin American countries (3)	Omit Sub-Saharan African and Latin American countries (4)	Include only Sub-Saharan African and Latin American countries (5)
	Dependent variable is log income per capita in 2000 CE				
Predicted diversity	299.280***	285.838***	492.194***	596.432***	250.528***
(ancestry adjusted)	(65.790)	(88.335)	(141.365)	(194.582)	(83.944)
Predicted diversity square	-208.886***	-199.755***	-342.644***	-415.173***	-174.724***
(ancestry adjusted)	(47.250)	(64.406)	(98.355)	(137.698)	(60.568)
Log Neolithic transition	0.305*	0.024	0.352*	-0.035	0.382
timing (ancestry adjusted)	(0.174)	(0.241)	(0.208)	(0.269)	(0.271)
Log percentage of arable	-0.193***	-0.197***	-0.192***	-0.152**	-0.181**
land	(0.047)	(0.070)	(0.043)	(0.067)	(0.069)
Log absolute latitude	-0.049	0.118	-0.189	-0.174	-0.037
	(0.090)	(0.117)	(0.114)	(0.153)	(0.112)
Social infrastructure	1.224***	0.845*	1.525**	0.869	0.654
	(0.455)	(0.434)	(0.601)	(0.609)	(0.940)
Ethnic fractionalization	-0.355	0.112	-0.739**	-0.511*	-0.294
	(0.254)	(0.277)	(0.285)	(0.296)	(0.449)
Percentage of population at	-0.665**	-0.557**	-0.426	0.122	-1.079**
risk of contracting malaria	(0.297)	(0.265)	(0.416)	(0.320)	(0.429)
Percentage of population	-0.306	-0.554***	-0.469**	-0.976***	-0.131
living in tropical zones	(0.186)	(0.190)	(0.219)	(0.265)	(0.218)
Mean distance to nearest	-0.384**	-0.767***	-0.416***	-0.665**	-0.281
waterway	(0.179)	(0.246)	(0.149)	(0.247)	(0.238)
Optimal diversity	0.716***	0.715***	0.718***	0.718***	0.717***
	(0.008)	(0.011)	(0.006)	(0.007)	(0.014)
Observations	109	71	87	49	60
R^2	0.90	0.89	0.93	0.94	0.80

Notes: This table establishes, using the 109-country sample from the contemporary analysis, that the significant hump-shaped effect of ancestry-adjusted genetic diversity on log income per capita in 2000 CE, while controlling for the ancestry-adjusted timing of the Neolithic Revolution, land productivity, and a vector of institutional, cultural, and geographical determinants of development, is qualitatively robust to (i) controls for region (rather than continent) fixed effects, (ii) omitting observations associated with either Sub-Saharan Africa (Column 2) or Latin America (Column 3) or both (Column 4), which, given the laggard comparative development of these regions along with their relatively higher and lower levels of genetic diversity respectively, may a priori be considered to be influential for generating the worldwide hump-shaped relationship between diversity and development, and (iii) restricting the sample to only observations in the potentially influential Sub-Saharan African and Latin American regions (Column 5). All regressions include controls for major religion shares as well as OPEC, legal origin, and region fixed effects. The region fixed effects employed by the analysis include fixed effects for (i) Sub-Saharan Africa (except in Columns 2 and 4), (ii) Middle East and North Africa (except in Column 5), (iii) Europe and Central Asia (except in Column 5), (iv) South Asia (except in Column 5), (v) East Asia and Pacific (except in Column 5), and (vi) Latin America and Caribbean (except in Columns 3–5). Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.12: Robustness to Percentage of the Population of European Descent

	Full sample		Omit OECD countries		Omit Neo-European countries		Omit European countries	
	Baseline Specification (1)	Augmented Specification (2)	Baseline Specification (3)	Augmented Specification (4)	Baseline Specification (5)	Augmented Specification (6)	Baseline Specification (7)	Augmented Specification (8)
	Dependent variable is log income per capita in 2000 CE							
Predicted diversity (ancestry adjusted)	281.173***	264.464***	266.494***	244.273**	266.198***	266.611***	285.466***	268.295**
	(70.459)	(83.157)	(84.381)	(114.359)	(72.371)	(90.847)	(79.049)	(109.024)
Predicted diversity square (ancestry adjusted)	-195.010***	-183.405***	-185.223***	-169.653**	-185.113***	-185.403***	-198.234***	-186.270**
	(49.764)	(58.195)	(59.302)	(79.762)	(50.915)	(63.534)	(55.641)	(75.954)
Log Neolithic transition timing (ancestry adjusted)	0.417*	0.374	0.415	0.368	0.376	0.377	0.505*	0.463
	(0.231)	(0.240)	(0.283)	(0.305)	(0.231)	(0.248)	(0.276)	(0.307)
Log percentage of arable land	-0.189***	-0.190***	-0.238***	-0.237***	-0.205***	-0.205***	-0.204***	-0.205***
	(0.049)	(0.049)	(0.060)	(0.060)	(0.053)	(0.054)	(0.054)	(0.055)
Log absolute latitude	0.003	-0.003	-0.021	-0.026	-0.029	-0.028	0.011	0.003
	(0.105)	(0.105)	(0.119)	(0.122)	(0.109)	(0.111)	(0.110)	(0.111)
Social infrastructure	1.867***	1.812***	1.373**	1.357**	1.485***	1.485***	1.836***	1.776***
	(0.404)	(0.419)	(0.575)	(0.604)	(0.501)	(0.509)	(0.450)	(0.529)
Ethnic fractionalization	-0.351	-0.365	-0.450	-0.473	-0.404	-0.404	-0.406	-0.415
	(0.276)	(0.289)	(0.375)	(0.400)	(0.297)	(0.302)	(0.356)	(0.363)
Percentage of population at risk of contracting malaria	-0.504	-0.483	-0.593	-0.556	-0.585	-0.586	-0.513	-0.495
	(0.344)	(0.369)	(0.376)	(0.437)	(0.363)	(0.403)	(0.374)	(0.401)
Percentage of population living in tropical zones	-0.314	-0.291	-0.199	-0.163	-0.300	-0.301	-0.271	-0.248
	(0.199)	(0.204)	(0.239)	(0.254)	(0.215)	(0.223)	(0.227)	(0.239)
Mean distance to nearest waterway	-0.373**	-0.372**	-0.389*	-0.376	-0.450**	-0.450**	-0.320	-0.321
	(0.184)	(0.186)	(0.222)	(0.229)	(0.208)	(0.212)	(0.204)	(0.207)
Percentage of population of European descent		0.211		0.270		-0.006		0.187
		(0.618)		(0.914)		(0.710)		(0.810)
Optimal diversity	0.721***	0.721***	0.719***	0.720***	0.719***	0.719***	0.720***	0.720***
	(0.016)	(0.055)	(0.069)	(0.108)	(0.014)	(0.019)	(0.016)	(0.031)
Observations	109	109	83	83	105	105	88	88
R^2	0.90	0.90	0.82	0.82	0.89	0.89	0.85	0.85

Notes: This table (i) establishes, using the 109-country sample from the contemporary analysis, that the significant hump-shaped effect of ancestry-adjusted genetic diversity on log income per capita in 2000 CE (as shown in Column 1), while controlling for the ancestry-adjusted timing of the Neolithic Revolution, land productivity, a vector of institutional, cultural, and geographical determinants of development, and continent fixed effects, is robust to the omission of (a) OECD countries (Column 3), (b) Neo-European countries, i.e., the U.S., Canada, Australia, and New Zealand (Column 5), and (c) European countries (Column 7) from the regression sample, and (ii) demonstrates that, in each of the corresponding samples, including the unrestricted one, the conditional hump-shaped effect of ancestry-adjusted genetic diversity is robust to additionally controlling for the percentage of the population of European descent (Columns 2, 4, 6, and 8), thereby confirming that the hump-shaped effect of genetic diversity on contemporary comparative development is not reflecting the latent impact of channels related to the diffusion of Europeans during the era of European colonization. All regressions include controls for major religion shares as well as OPEC, legal origin, Sub-Saharan Africa, and continent fixed effects. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.13: Standardized Beta and Partial R^2 Coefficients for the Baseline Specifications

	Limited-sample historical analysis		Extended-sample historical analysis			Contemporary analysis
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is:					
	Log population density in 1500 CE		Log population density in 1500 CE	Log population density in 1000 CE	Log population density in 1 CE	Log income per capita in 2000 CE
Diversity	7.203 {0.321}	6.512 {0.212}	7.008 {0.072}	6.955 {0.056}	7.540 {0.062}	5.509 {0.033}
Diversity squared	-6.913 {0.299}	-6.251 {0.178}	-6.936 {0.074}	-6.838 {0.058}	-7.364 {0.063}	-5.485 {0.033}
Log Neolithic transition timing	0.372 {0.346}	0.348 {0.237}	0.490 {0.205}	0.653 {0.276}	0.765 {0.338}	0.024 {0.000}
Log percentage of arable land	0.343 {0.288}	0.363 {0.244}	0.315 {0.110}	0.304 {0.082}	0.256 {0.062}	-0.123 {0.012}
Log absolute latitude	-0.109 {0.064}	-0.086 {0.032}	-0.257 {0.101}	-0.236 {0.072}	-0.069 {0.006}	0.135 {0.021}
Log land suitability for agriculture	0.291 {0.323}	0.299 {0.318}	0.225 {0.060}	0.173 {0.028}	0.163 {0.028}	-0.200 {0.031}
Continent fixed effects	No	Yes	Yes	Yes	Yes	Yes
Observations	21	21	145	140	126	143
R^2	0.89	0.90	0.69	0.62	0.61	0.57

Notes: This table reports, for both the limited- and extended-sample variants of the historical analysis as well as for the contemporary analysis, the standardized beta coefficient and the partial R^2 coefficient (in curly brackets) associated with each explanatory variable in the baseline specification corresponding to each analysis. The standardized beta coefficient associated with an explanatory variable in a given regression is to be interpreted as the number of standard deviations by which the dependent variable changes in response to a one standard deviation increase in the explanatory variable. The partial R^2 coefficient associated with an explanatory variable in a given regression is to be interpreted as the fraction of the residual variation in the dependent variable that is explained by the residual variation in the explanatory variable, where residual variation of a (dependent or explanatory) variable refers to the variation in that variable that is unexplained by the variation in all other explanatory variables in the regression. The relevant measures of genetic diversity employed by the analysis are observed genetic diversity in Columns 1–2, predicted genetic diversity (i.e., genetic diversity as predicted by migratory distance from East Africa) in Columns 3–5, and ancestry-adjusted genetic diversity in Column 6. The ancestry-adjusted measure of the timing of the Neolithic Revolution is used in Column 6.

TABLE D.14: Interpersonal Trust, Scientific Productivity, and Economic Development in 2000 CE

	Specifications with interpersonal trust			Specifications with scientific articles		
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is log income per capita in 2000 CE					
Interpersonal trust	3.715***	2.365***	1.729***			
	(0.702)	(0.660)	(0.593)			
Scientific articles				3.391***	1.968***	1.430***
				(0.312)	(0.334)	(0.293)
Log Neolithic transition		-0.230	-0.311		-0.125	-0.032
timing (ancestry adjusted)		(0.289)	(0.290)		(0.268)	(0.189)
Log percentage of arable		-0.222*	-0.225*		-0.089	-0.027
land		(0.119)	(0.117)		(0.099)	(0.074)
Log absolute latitude		0.353**	-0.194		0.166	0.015
		(0.135)	(0.247)		(0.105)	(0.108)
Log land suitability for		0.177	0.122		-0.102	-0.140*
agriculture		(0.110)	(0.090)		(0.112)	(0.081)
Executive constraints			0.163***			0.109**
			(0.051)			(0.044)
Ethnic fractionalization			-0.724**			-0.670**
			(0.349)			(0.333)
Percentage of population at			-1.110***			-1.284***
risk of contracting malaria			(0.406)			(0.267)
Percentage of population			-0.548			0.088
living in tropical zones			(0.548)			(0.253)
Mean distance to nearest			0.065			-0.176
waterway			(0.096)			(0.159)
OPEC fixed effect	No	No	Yes	No	No	Yes
Legal origin fixed effects	No	No	Yes	No	No	Yes
Continent fixed effects	No	Yes	Yes	No	Yes	Yes
Observations	73	73	73	125	125	125
R^2	0.27	0.68	0.84	0.41	0.68	0.80

Notes: This table (i) establishes that the prevalence of interpersonal trust and the annual average number of scientific articles per capita both possess statistically significant positive relationships with log income per capita in 2000 CE, using feasible 73-country and 125-country samples respectively, and (ii) demonstrates that these relationships hold unconditionally (Columns 1 and 4) and conditional on either (a) a baseline set of control variables (Columns 2 and 5), including controls for the ancestry-adjusted timing of the Neolithic Revolution, land productivity, and continent fixed effects, or (b) a more comprehensive set of controls for institutional, cultural, and geographical determinants of development (Columns 3 and 6), including controls for executive constraints, ethnic fractionalization, legal origins, tropical disease environments, access to waterways, and natural resource endowments. All regressions with continent fixed effects also control for a Sub-Saharan Africa fixed effect. Heteroskedasticity robust standard errors are reported in parentheses. For additional details on the prevalence of interpersonal trust and the annual average number of scientific articles per capita, the reader is referred to the definitions of these variables in Section F of this appendix. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.15: Robustness to an Alternative Definition of Neolithic Transition Timing

	Limited-sample analysis				Extended-sample analysis	
	OLS	OLS	2SLS	2SLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable is log population density in 1500 CE					
Panel A. Controlling for logged Neolithic transition timing with respect to 1500 CE						
Diversity	228.020***	198.225**	282.385***	249.033***	207.666***	177.155**
	(73.696)	(85.848)	(88.267)	(72.352)	(54.226)	(83.806)
Diversity square	-162.990**	-142.553*	-204.342***	-184.470***	-147.104***	-128.666**
	(56.121)	(71.792)	(67.057)	(57.795)	(39.551)	(58.497)
Log Neolithic transition timing	1.005***	0.939*	0.850***	0.950**	0.831***	0.793***
	(0.320)	(0.495)	(0.309)	(0.371)	(0.129)	(0.204)
Log percentage of arable land	0.517***	0.420*	0.602***	0.443**	0.419***	0.431***
	(0.170)	(0.217)	(0.193)	(0.196)	(0.095)	(0.101)
Log absolute latitude	-0.143	-0.114	-0.189	-0.113	-0.299***	-0.416***
	(0.131)	(0.171)	(0.125)	(0.130)	(0.094)	(0.126)
Log land suitability for agriculture	0.558*	0.658	0.489**	0.661**	0.280***	0.219**
	(0.304)	(0.403)	(0.241)	(0.312)	(0.095)	(0.099)
Continent fixed effects	No	Yes	No	Yes	No	Yes
Observations	21	21	21	21	144	144
R^2	0.89	0.91			0.65	0.67
Panel B. Controlling for nonlogged Neolithic transition timing with respect to 1500 CE						
Diversity	256.812***	258.914**	311.376***	341.608***	199.715***	225.672***
	(79.383)	(94.992)	(85.804)	(93.208)	(57.660)	(80.558)
Diversity square	-185.560***	-191.630**	-226.889***	-259.555***	-141.308***	-170.633***
	(60.373)	(80.404)	(64.985)	(76.477)	(42.129)	(56.691)
Neolithic transition timing	0.234**	0.227	0.202***	0.258**	0.274***	0.352***
	(0.089)	(0.160)	(0.075)	(0.127)	(0.038)	(0.054)
Log percentage of arable land	0.580***	0.497*	0.655***	0.527**	0.416***	0.372***
	(0.161)	(0.233)	(0.170)	(0.208)	(0.091)	(0.105)
Log absolute latitude	-0.229	-0.202	-0.269**	-0.197*	-0.407***	-0.527***
	(0.134)	(0.167)	(0.117)	(0.119)	(0.101)	(0.124)
Log land suitability for agriculture	0.588**	0.685	0.509**	0.699**	0.306***	0.259***
	(0.272)	(0.430)	(0.218)	(0.336)	(0.095)	(0.098)
Continent fixed effects	No	Yes	No	Yes	No	Yes
Observations	21	21	21	21	145	145
R^2	0.89	0.90			0.64	0.69

Notes: This table establishes that, in both the limited- and extended-sample variants of the historical analysis for the year 1500 CE, the hump-shaped effect of genetic diversity on log population density remains qualitatively robust under an alternative definition of the Neolithic transition timing variable. In this case, the timing of the Neolithic Revolution reflects the number of years elapsed, until the year 1500 CE (as opposed to 2000 CE), since the transition to sedentary agriculture. The analysis employs the logged version of this variable in Panel A and its nonlogged version in Panel B. In Columns 5–6, the higher number of observations in Panel B (relative to Panel A) reflects the inclusion of Australia, which was yet to experience the Neolithic Revolution as of 1500 CE, in the sample. This permits the relevant regressions in Panel B to exploit information on both the realized and unrealized "potential" of countries to experience the Neolithic Revolution as of 1500 CE. The relevant measures of genetic diversity employed by the analysis are observed genetic diversity in Columns 1–4 and predicted genetic diversity (i.e., genetic diversity as predicted by migratory distance from East Africa) in Columns 5–6. In Columns 3–4, genetic diversity and its squared term are instrumented following the methodology implemented in Columns 5–6 of Table 2 of the paper. Heteroskedasticity robust standard errors are reported in parentheses in Columns 1–4. Bootstrap standard errors, accounting for the use of generated regressors, are reported in parentheses in Columns 5–6. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE D.16: Results for Distances Under an Alternative Definition of Neolithic Transition Timing

Distance from:	Addis Ababa (1)	Addis Ababa (2)	London (3)	Tokyo (4)	Mexico City (5)
	Dependent variable is log population density in 1500 CE				
Panel A. Controlling for logged Neolithic transition timing with respect to 1500 CE					
Migratory distance	0.152**		-0.046	0.072	-0.007
	(0.061)		(0.064)	(0.139)	(0.101)
Migratory distance square	-0.008***		-0.002	-0.008	0.003
	(0.002)		(0.002)	(0.007)	(0.004)
Aerial distance		-0.030			
		(0.106)			
Aerial distance square		-0.003			
		(0.006)			
Log Neolithic transition timing	0.831***	0.847***	0.702***	0.669***	1.106***
	(0.125)	(0.127)	(0.140)	(0.165)	(0.248)
Log percentage of arable land	0.419***	0.502***	0.367***	0.530***	0.506***
	(0.094)	(0.105)	(0.095)	(0.088)	(0.098)
Log absolute latitude	-0.299***	-0.211**	-0.320***	-0.290***	-0.195**
	(0.091)	(0.097)	(0.114)	(0.098)	(0.085)
Log land suitability for agriculture	0.280***	0.239**	0.327***	0.169**	0.241**
	(0.096)	(0.106)	(0.094)	(0.081)	(0.096)
Observations	144	144	144	144	144
R^2	0.65	0.56	0.64	0.57	0.60
Panel B. Controlling for nonlogged Neolithic transition timing with respect to 1500 CE					
Migratory distance	0.144**		-0.078	0.017	0.072
	(0.064)		(0.062)	(0.146)	(0.099)
Migratory distance square	-0.008***		-0.000	-0.005	-0.001
	(0.002)		(0.002)	(0.007)	(0.004)
Aerial distance		0.099			
		(0.104)			
Aerial distance square		-0.012*			
		(0.006)			
Neolithic transition timing	0.274***	0.279***	0.223***	0.240***	0.316***
	(0.038)	(0.037)	(0.038)	(0.062)	(0.065)
Log percentage of arable land	0.416***	0.504***	0.357***	0.552***	0.507***
	(0.086)	(0.100)	(0.089)	(0.088)	(0.087)
Log absolute latitude	-0.407***	-0.367***	-0.442***	-0.391***	-0.353***
	(0.095)	(0.103)	(0.118)	(0.107)	(0.096)
Log land suitability for agriculture	0.306***	0.244**	0.349***	0.173**	0.260***
	(0.092)	(0.102)	(0.092)	(0.080)	(0.090)
Observations	145	145	145	145	145
R^2	0.64	0.56	0.64	0.56	0.58

Notes: This table establishes that (i) the hump-shaped effect of migratory distance from East Africa on log population density in 1500 CE and (ii) the absence of a similar effect associated with alternative concepts of distance remain qualitatively robust under an alternative definition of the Neolithic transition timing variable. In this case, the timing of the Neolithic Revolution reflects the number of years elapsed, until the year 1500 CE (as opposed to 2000 CE), since the transition to sedentary agriculture. The analysis employs the logged version of this variable in Panel A and its nonlogged version in Panel B. The higher number of observations in Panel B (relative to Panel A) reflects the inclusion of Australia, which was yet to experience the Neolithic Revolution as of 1500 CE, in the sample. This permits the regressions in Panel B to exploit information on both the realized and unrealized "potential" of countries to experience the Neolithic Revolution as of 1500 CE. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

E The 53 HGDP-CEPH 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

F Variable Definitions and Sources

F.1 Outcome Variables

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

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

Interpersonal trust. The fraction of total respondents within a given country, from five different waves of the *World Values Survey* conducted during the time period 1981–2008, that responded with "Most people can be trusted" (as opposed to "Can't be too careful") when answering the survey question "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"

Scientific articles. The mean, over the period 1981–2000, of the annual number of scientific articles per capita, calculated as the total number of scientific and technical articles published in a given year divided by the total population in that year. The relevant data on the total number of articles and population in a given year are obtained from the World Bank's *World Development Indicators*.

F.2 Genetic Diversity Variables

Observed genetic diversity (for the limited historical sample). The average expected heterozygosity across ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel that are located within a given country. The expected heterozygosities of the ethnic groups are from Ramachandran et al. (2005).

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

Note that for Table D.5 in Section D of this appendix, the migratory distance concept used to predict the genetic diversity of a country's population is the human mobility index, calculated for the journey from

Addis Ababa (Ethiopia) to the country's modern capital city, as opposed to the baseline waypoints-restricted migratory distance concept used elsewhere. For additional details on how the human mobility index is calculated, the interested reader is referred to the definition of this variable further below.

Predicted genetic diversity (ancestry adjusted). The expected heterozygosity (genetic diversity) of a country's population, predicted by migratory distances from East Africa (i.e., Addis Ababa, Ethiopia) to the year 1500 CE locations of the ancestral populations of the country's component ethnic groups in 2000 CE, as well as by pairwise migratory distances between these ancestral populations. The source countries of the year 1500 CE ancestral populations are identified from the *World Migration Matrix, 1500–2000*, discussed in Putterman and Weil (2010), and the modern capital cities of these countries are used to compute the aforementioned migratory distances. The measure of genetic diversity is then calculated by applying (i) the regression coefficients obtained from regressing expected heterozygosity on migratory distance from East Africa at the ethnic group level, using the worldwide sample of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel, (ii) the regression coefficients obtained from regressing pairwise F_{st} genetic distances on pairwise migratory distances between these ethnic groups, and (iii) the ancestry weights representing the fractions of the year 2000 CE population (of the country for which the measure is being computed) that can trace their ancestral origins to different source countries in the year 1500 CE. The construction of this measure is discussed in detail in Section B of this appendix. The expected heterozygosities, geographical coordinates, and pairwise F_{st} genetic distances of the 53 ethnic groups are from Ramachandran et al. (2005). The ancestry weights are from the *World Migration Matrix, 1500–2000*.

Note that, in contrast to the baseline waypoints-restricted migratory distance concept used elsewhere, the migratory distance concept used to predict the ancestry-adjusted genetic diversity of a country's population for Table D.5 in Section D of this appendix is the human mobility index, calculated for the journey from Addis Ababa (Ethiopia) to each of the year 1500 CE locations of the ancestral popul ations of the country's component ethnic groups in 2000 CE, as well as for the journey between each pair of these ancestral populations. For additional details on how the human mobility index is calculated, the interested reader is referred to the definition of this variable further below.

F.3 Distance Variables

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

Migratory distance from East Africa (for the extended historical sample). The great circle distance from Addis Ababa (Ethiopia) to the country's modern capital city along a land-restricted path forced through one or more of five aforementioned intercontinental waypoints. Distances are calculated using the Haversine formula and are measured in units of 1,000 km. The geographical coordinates of the intercontinental waypoints

are from Ramachandran et al. (2005), while those of the modern capital cities are from the CIA's *World Factbook*.

Migratory distance from East Africa (ancestry adjusted). The cross-country weighted average of (the extended sample definition of) migratory distance from East Africa (i.e., Addis Ababa, Ethiopia), where the weight associated with a given country in the calculation represents the fraction of the year 2000 CE population (of the country for which the measure is being computed) that can trace its ancestral origins to the given country in the year 1500 CE. The ancestry weights are obtained from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2010).

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

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

Aerial distance from East Africa (ancestry adjusted). The cross-country weighted average of aerial distance from East Africa (i.e., Addis Ababa, Ethiopia), where the weight associated with a given country in the calculation represents the fraction of the year 2000 CE population (of the country for which the measure is being computed) that can trace its ancestral origins to the given country in the year 1500 CE. The ancestry weights are from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2010).

Distance to regional frontier in 1 CE, 1000 CE, and 1500 CE. The great circle distance from a country's capital city to the closest regional technological frontier for a given year. The year-specific set of regional frontiers comprises the two most populous cities, reported for that year and belonging to different civilizations or sociopolitical entities, from each of Africa, Europe, Asia, and the Americas. Distances are calculated using the Haversine formula and are measured in km. The historical urban population data used to identify the frontiers are obtained from Chandler (1987) and Modelski (2003), and the geographical coordinates of ancient urban centers are obtained using *Wikipedia*.

Human mobility index. The average migratory distance across ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel that are located within a given country. The migratory distance of an ethnic group is the distance from Addis Ababa (Ethiopia) to the location of the group along an "optimal" land-restricted path that minimizes the time cost of travelling on the surface of the Earth in the absence of steam-powered transportation technologies. The optimality of a path is determined by incorporating information on natural impediments to human spatial mobility, such as the meteorological and topographical conditions prevalent along the path, as well as information on the time cost of travelling under such conditions as reported by Hayes (1996). Distances are measured in weeks of travel time. The geographical coordinates of the ethnic

groups are from Ramachandran et al. (2005). The methodology underlying the construction of this index is discussed in greater detail by Ashraf, Galor and Özak (2010) and Özak (2010).

Genetic distance to the U.K. or Ethiopia (1500 match). The F_{st} genetic distance, as reported by Spolaore and Wacziarg (2009), between the year 1500 CE populations of a given country and the U.K. (or Ethiopia), calculated as the genetic distance between the two ethnic groups comprising the largest shares of each country's population in the year 1500 CE.

Genetic distance to the U.S. or Ethiopia (weighted). The F_{st} genetic distance, as reported by Spolaore and Wacziarg (2009), between the contemporary national populations of a given country and the U.S. (or Ethiopia), calculated as the average pairwise genetic distance across all ethnic group pairs, where each pair comprises two distinct ethnic groups, one from each country, and is weighted by the product of the proportional representations of the two groups in their respective national populations.

F.4 Timing of the Neolithic Revolution and Subsistence Mode Variables

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

Note that the historical analysis, as conducted in Section 4 of the paper, employs the Neolithic transition timing variable defined above (i.e., measured as the number of thousand years since the onset of sedentary agriculture as of the year 2000 CE). This results in the inclusion of countries that were yet to experience the onset of sedentary agriculture as of the year 1500 CE in the sample, thereby permitting the relevant regressions to exploit information on both the realized and unrealized "potential" of countries to undergo the Neolithic Revolution. Nevertheless, Tables D.15 and D.16 in Section D of this appendix demonstrate that all the results of the historical analysis are robust under an alternative definition of the Neolithic transition timing variable where this variable reflects the number of years elapsed, until the year 1500 CE (as opposed to 2000 CE), since the transition to agriculture.

Neolithic transition timing (ancestry adjusted). The cross-country weighted average of Neolithic transition timing, where the weight associated with a given country in the calculation represents the fraction of the year 2000 CE population (of the country for which the measure is being computed) that can trace its ancestral origins to the given country in the year 1500 CE. The ancestry weights are obtained from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2010).

Subsistence mode in 1000 CE. An index in the [0,1]-interval that gauges the extent to which sedentary agriculture was practiced, in the year 1000 CE, within a region delineated by a country's modern international borders. This index is constructed using data from Peregrine's (2003) *Atlas of Cultural Evolution*, which reports, amongst other variables, a measure of the mode of subsistence on a 3-point categorical scale at the level of a cultural group (or "archaeological tradition") that existed in the year 1000 CE. Specifically, the

measure is taken to assume a value of 0 in the absence of sedentary agriculture (i.e., if the cultural group exclusively practiced hunting and gathering), a value of 0.5 when agriculture was practiced but only as a secondary mode of subsistence, and a value of 1 when agriculture was practiced as the primary mode of subsistence. Given that the cross-sectional unit of observation in Peregrine's (2003) data set is a cultural group, specific to a given region on the global map, and since spatial delineations of groups, as reported by Peregrine (2003), do not necessarily correspond to contemporary international borders, the measure is aggregated to the country level by averaging across those cultural groups that are reported to appear within the modern borders of a given country. For more details on the underlying data employed to construct this index, the interested reader is referred to Peregrine (2003).

F.5 Geographical Variables

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

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

Land suitability for agriculture. A geospatial index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH. This index was initially reported at a half-degree resolution by Ramankutty et al. (2002). Formally, Ramankutty et al. (2002) calculate the land suitability index, S, as the product of climate suitability, S_{clim}, and soil suitability, S_{soil}, i.e., $S = S_{clim} \times S_{soil}$. The climate suitability component is estimated to be a function of growing degree days, GDD, and a moisture index, α, gauging water availability to plants, calculated as the ratio of actual to potential evapotranspiration, i.e., $S_{clim} = f_1(GDD)f_2(\alpha)$. The soil suitability component, on the other hand, is estimated to be a function of soil carbon density, C_{soil}, and soil pH, pH_{soil}, i.e. $S_{soil} = g_1(C_{soil})g_2(pH_{soil})$. The functions, $f_1(GDD)$, $f_2(\alpha)$, $g_1(C_{soil})$, and $g_2(pH_{soil})$ are chosen by Ramankutty et al. (2002) by empirically fitting functions to the observed relationships between cropland areas, GDD, α, C_{soil}, and pH_{soil}. For more details on the specific functional forms chosen, the interested reader is referred to Ramankutty et al. (2002). Since Ramankutty et al. (2002) report the land suitability index at a half-degree resolution, Michalopoulos (2011) aggregates the index to the country level by averaging land suitability across grid cells within a country. This study employs the country-level aggregate measure reported by Michalopoulos (2011) as the control for land suitability in the baseline regression specifications for both historical population density and contemporary income per capita.

Range of land suitability. The difference between the maximum and minimum values of a land suitability index, reported at a half-degree resolution by Ramankutty et al. (2002), across grid cells within a country. This variable is obtained from the data set of Michalopoulos (2011). For additional details on the land suitability index, the interested reader is referred to the definition of the land suitability variable above.

Land suitability Gini. The Gini coefficient based on the distribution of a land suitability index, reported at a half-degree resolution by Ramankutty et al. (2002), across grid cells within a country. This variable is obtained from the data set of Michalopoulos (2011). For additional details on the land suitability index, the interested reader is referred to the definition of the land suitability variable above.

Soil fertility. The soil suitability component, based on soil carbon density and soil pH, of an index of land suitability for agriculture. The soil suitability data are reported at a half-degree resolution by Ramankutty et al. (2002) and are aggregated to the country level by Michalopoulos (2011) by averaging across grid cells within a country. For additional details on the soil suitability component of the land suitability index, the interested reader is referred to the definition of the land suitability variable above.

Mean elevation. The mean elevation of a country in km above sea level, calculated using geospatial elevation data reported by the *G-ECON* project (Nordhaus, 2006) at a 1-degree resolution, which, in turn, is based on similar but more spatially disaggregated data at a 10-minute resolution from New et al. (2002). The measure is thus the average elevation across the grid cells within a country. The interested reader is referred to the *G-ECON* project web site for additional details.

Standard deviation of elevation. The standard deviation of elevation across the grid cells within a country in km above sea level, calculated using geospatial elevation data reported by the *G-ECON* project (Nordhaus, 2006) at a 1-degree resolution, which, in turn, is based on similar but more spatially disaggregated data at a 10-minute resolution from New et al. (2002). The interested reader is referred to the *G-ECON* project web site for additional details.

Terrain roughness. The degree of terrain roughness of a country, calculated using geospatial surface undulation data reported by the *G-ECON* project (Nordhaus, 2006) at a 1-degree resolution, which is based on more spatially disaggregated elevation data at a 10-minute resolution from New et al. (2002). The measure is thus the average degree of terrain roughness across the grid cells within a country. The interested reader is referred to the *G-ECON* project web site for additional details.

Temperature. The intertemporal average monthly temperature of a country in degrees Celsius per month over the 1961–1990 time period, calculated using geospatial average monthly temperature data for this period reported by the *G-ECON* project (Nordhaus, 2006) at a 1-degree resolution, which, in turn, is based on similar but more spatially disaggregated data at a 10-minute resolution from New et al. (2002). The measure is thus the spatial mean of the intertemporal average monthly temperature across the grid cells within a country. The interested reader is referred to the *G-ECON* project web site for additional details.

Precipitation. The intertemporal average monthly precipitation of a country in mm per month over the 1961–1990 time period, calculated using geospatial average monthly precipitation data for this period reported by the *G-ECON* project (Nordhaus, 2006) at a 1-degree resolution, which, in turn, is based on similar but more spatially disaggregated data at a 10-minute resolution from New et al. (2002). The measure is thus the spatial mean of the intertemporal average monthly precipitation across the grid cells within a country. The interested reader is referred to the *G-ECON* project web site for additional details.

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

Percentage of land near a waterway. The percentage of a country's total land area that is located within 100 km of an ice-free coastline or sea-navigable river. This variable was originally constructed by Gallup,

Sachs and Mellinger (1999) and is part of Harvard University's CID Research Datasets on *General Measures of Geography*.

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

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

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

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

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

Domesticable plants. The number of annual and perennial wild grass species, with a mean kernel weight exceeding 10 mg, that were prehistorically native to the region to which a country belongs. This variable is obtained from the data set of Olsson and Hibbs (2005).

Domesticable animals. The number of domesticable large mammalian species, weighing in excess of 45 kg, that were prehistorically native to the region to which a country belongs. This variable is obtained from the data set of Olsson and Hibbs (2005).

F.6 Institutional, Cultural, and Human Capital Variables

Social infrastructure. An index, calculated by Hall and Jones (1999), that quantifies the wedge between private and social returns to productive activities. To elaborate, this measure is computed as the average of two separate indices. The first is a government anti-diversion policy (GADP) index, based on data from the *International Country Risk Guide*, that represents the average across five categories, each measured as the mean over the 1986–1995 time period: (i) law and order, (ii) bureaucratic quality, (iii) corruption, (iv) risk of expropriation, and (v) government repudiation of contracts. The second is an index of openness, based on Sachs and Warner (1995), that represents the fraction of years in the time period 1950–1994 that the economy was open to trade with other countries, where the criteria for being open in a given year includes: (i) nontariff barriers cover less than 40% of trade, (ii) average tariff rates are less than 40%, (iii) any black market premium was less than 20% during the 1970s and 80s, (iv) the country is not socialist, and (v) the government does not monopolize over major exports.

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

Executive constraints. The 1960–2000 mean of an index, reported annually as a 7-point categorical variable (from 1 to 7) by the *Polity IV* data set, quantifying the extent of institutionalized constraints on the decision-making power of chief executives.

Legal origins. A set of dummy variables, reported by La Porta et al. (1999), that identifies the legal origin of the Company Law or Commercial Code of a country. The five legal origin possibilities are: (i) English Common Law, (ii) French Commercial Code, (iii) German Commercial Code, (iv) Scandinavian Commercial Code, and (v) Socialist or Communist Laws.

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

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

Percentage of population of European descent. The fraction of the year 2000 CE population (of the country for which the measure is being computed) that can trace its ancestral origins to the European continent due to migrations occurring as early as the year 1500 CE. This variable is constructed using data from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2010).

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

G Descriptive Statistics

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

		Obs.	Mean	S.D.	Min.	Max.
(1)	Log population density in 1500 CE	21	1.169	1.756	-2.135	3.842
(2)	Observed genetic diversity	21	0.713	0.056	0.552	0.770
(3)	Migratory distance from East Africa	21	8.238	6.735	1.335	24.177
(4)	Human mobility index	18	10.965	8.124	2.405	31.360
(5)	Log Neolithic transition timing	21	8.342	0.539	7.131	9.259
(6)	Log percentage of arable land	21	2.141	1.168	-0.799	3.512
(7)	Log absolute latitude	21	2.739	1.178	0.000	4.094
(8)	Log land suitability for agriculture	21	-1.391	0.895	-3.219	-0.288

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

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	Log population density in 1500 CE	1.000						
(2)	Observed genetic diversity	0.244	1.000					
(3)	Migratory distance from East Africa	-0.226	-0.968	1.000				
(4)	Human mobility index	-0.273	-0.955	0.987	1.000			
(5)	Log Neolithic transition timing	0.735	-0.117	0.024	0.011	1.000		
(6)	Log percentage of arable land	0.670	0.172	-0.183	-0.032	0.521	1.000	
(7)	Log absolute latitude	0.336	0.055	-0.012	0.044	0.392	0.453	1.000
(8)	Log land suitability for agriculture	0.561	-0.218	0.282	0.245	0.299	0.376	0.049

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

		Obs.	Mean	S.D.	Min.	Max.
(1)	Log population density in 1500 CE	145	0.881	1.500	-3.817	3.842
(2)	Log population density in 1000 CE	140	0.463	1.445	-4.510	2.989
(3)	Log population density in 1 CE	126	-0.070	1.535	-4.510	3.170
(4)	Predicted genetic diversity	145	0.711	0.053	0.572	0.774
(5)	Log Neolithic transition timing	145	8.343	0.595	5.991	9.259
(6)	Log percentage of arable land	145	2.232	1.203	-2.120	4.129
(7)	Log absolute latitude	145	3.003	0.924	0.000	4.159
(8)	Log land suitability for agriculture	145	-1.409	1.313	-5.857	-0.041
(9)	Log distance to regional frontier in 1500 CE	145	7.309	1.587	0.000	9.288
(10)	Log distance to regional frontier in 1000 CE	145	7.406	1.215	0.000	9.258
(11)	Log distance to regional frontier in 1 CE	145	7.389	1.307	0.000	9.261
(12)	Mean elevation	145	0.555	0.481	0.024	2.674
(13)	Terrain roughness	145	0.178	0.135	0.013	0.602
(14)	Mean distance to nearest waterway	145	0.350	0.456	0.014	2.386
(15)	Percentage of land near a waterway	145	0.437	0.368	0.000	1.000
(16)	Climate	96	1.531	1.046	0.000	3.000
(17)	Orientation of continental axis	96	1.521	0.685	0.500	3.000
(18)	Size of continent	96	30.608	13.605	0.065	44.614
(19)	Domesticable plants	96	13.260	13.416	2.000	33.000
(20)	Domesticable animals	96	3.771	4.136	0.000	9.000
(21)	Migratory distance from East Africa	145	8.399	6.970	0.000	26.771
(22)	Aerial distance from East Africa	145	6.003	3.558	0.000	14.420
(23)	Migratory distance from London	145	8.884	7.104	0.000	26.860
(24)	Migratory distance from Tokyo	145	11.076	3.785	0.000	19.310
(25)	Migratory distance from Mexico City	145	15.681	6.185	0.000	25.020

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

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

		(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(13)	Terrain roughness	1.000											
(14)	Mean distance to nearest waterway	-0.002	1.000										
(15)	Percentage of land near a waterway	0.039	-0.665	1.000									
(16)	Climate	0.147	-0.497	0.431	1.000								
(17)	Orientation of continental axis	0.260	-0.236	0.285	0.482	1.000							
(18)	Size of continent	-0.079	0.104	-0.170	0.327	0.668	1.000						
(19)	Domesticable plants	0.074	-0.333	0.334	0.817	0.613	0.487	1.000					
(20)	Domesticable animals	0.132	-0.322	0.349	0.778	0.743	0.516	0.878	1.000				
(21)	Migratory distance from East Africa	0.161	-0.195	0.192	-0.080	-0.159	-0.465	-0.346	-0.249	1.000			
(22)	Aerial distance from East Africa	0.170	-0.220	0.292	-0.074	-0.058	-0.429	-0.281	-0.191	0.934	1.000		
(23)	Migratory distance from London	0.088	-0.081	-0.042	-0.407	-0.455	-0.569	-0.683	-0.612	0.899	0.800	1.000	
(24)	Migratory distance from Tokyo	-0.264	-0.128	-0.118	-0.220	-0.672	-0.310	-0.309	-0.626	0.266	0.172	0.418	1.000
(25)	Migratory distance from Mexico City	-0.284	0.094	-0.194	-0.014	-0.035	0.268	0.148	0.075	-0.822	-0.759	-0.675	-0.025

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

		Obs.	Mean	S.D.	Min.	Max.
(1)	Log income per capita in 2000 CE	143	8.416	1.165	6.158	10.445
(2)	Log population density in 1500 CE	143	0.891	1.496	-3.817	3.842
(3)	Predicted genetic diversity (unadjusted)	143	0.712	0.052	0.572	0.774
(4)	Predicted genetic diversity (ancestry adjusted)	143	0.727	0.027	0.628	0.774
(5)	Log Neolithic transition timing (unadjusted)	143	8.343	0.599	5.991	9.259
(6)	Log Neolithic transition timing (ancestry adjusted)	143	8.495	0.454	7.213	9.250
(7)	Log percentage of arable land	143	2.251	1.180	-2.120	4.129
(8)	Log absolute latitude	143	3.014	0.921	0.000	4.159
(9)	Log land suitability for agriculture	143	-1.416	1.321	-5.857	-0.041

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

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	Log income per capita in 2000 CE	1.000							
(2)	Log population density in 1500 CE	-0.011	1.000						
(3)	Predicted genetic diversity (unadjusted)	-0.212	0.378	1.000					
(4)	Predicted genetic diversity (ancestry adjusted)	-0.160	0.065	0.775	1.000				
(5)	Log Neolithic transition timing (unadjusted)	0.191	0.512	0.276	0.016	1.000			
(6)	Log Neolithic transition timing (ancestry adjusted)	0.438	0.245	-0.150	-0.151	0.762	1.000		
(7)	Log percentage of arable land	-0.008	0.571	0.097	0.083	0.156	0.158	1.000	
(8)	Log absolute latitude	0.502	0.083	0.082	0.043	0.322	0.419	0.248	1.000
(9)	Log land suitability for agriculture	-0.105	0.369	-0.251	-0.243	-0.133	-0.082	0.670	-0.041

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

		Obs.	Mean	S.D.	Min.	Max.
(1)	Log income per capita in 2000 CE	109	8.455	1.189	6.524	10.445
(2)	Predicted genetic diversity (ancestry adjusted)	109	0.726	0.030	0.628	0.774
(3)	Log Neolithic transition timing (ancestry adjusted)	109	8.417	0.478	7.213	9.250
(4)	Log percentage of arable land	109	2.248	1.145	-2.120	4.129
(5)	Log absolute latitude	109	2.841	0.960	0.000	4.159
(6)	Social infrastructure	109	0.453	0.243	0.113	1.000
(7)	Ethnic fractionalization	109	0.460	0.271	0.002	0.930
(8)	Percentage of population at risk of contracting malaria	109	0.369	0.438	0.000	1.000
(9)	Percentage of population living in tropical zones	109	0.361	0.419	0.000	1.000
(10)	Mean distance to nearest waterway	109	0.298	0.323	0.008	1.467
(11)	Percentage of population of European descent	109	0.313	0.404	0.000	1.000
(12)	Years of schooling	94	4.527	2.776	0.409	10.862
(13)	Migratory distance from East Africa (unadjusted)	109	9.081	7.715	0.000	26.771
(14)	Migratory distance from East Africa (ancestry adjusted)	109	6.365	3.929	0.000	19.388
(15)	Aerial distance from East Africa (unadjusted)	109	6.332	3.861	0.000	14.420
(16)	Aerial distance from East Africa (ancestry adjusted)	109	5.178	2.426	0.000	12.180

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

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	Log income per capita in 2000 CE	1.000									
(2)	Predicted genetic diversity (ancestry adjusted)	-0.263	1.000								
(3)	Log Neolithic transition timing (ancestry adjusted)	0.535	-0.232	1.000							
(4)	Log percentage of arable land	0.050	0.094	0.221	1.000						
(5)	Log absolute latitude	0.563	-0.007	0.375	0.242	1.000					
(6)	Social infrastructure	0.808	-0.220	0.362	0.119	0.476	1.000				
(7)	Ethnic fractionalization	-0.644	0.199	-0.406	-0.255	-0.590	-0.468	1.000			
(8)	Percentage of population at risk of contracting malaria	-0.784	0.411	-0.624	-0.204	-0.599	-0.593	0.672	1.000		
(9)	Percentage of population living in tropical zones	-0.418	-0.180	-0.266	-0.001	-0.723	-0.380	0.396	0.414	1.000	
(10)	Mean distance to nearest waterway	-0.444	0.335	-0.276	-0.189	-0.238	-0.309	0.417	0.443	-0.079	1.000
(11)	Percentage of population of European descent	0.732	-0.153	0.464	0.260	0.551	0.674	-0.533	-0.629	-0.385	-0.346
(12)	Years of schooling	0.866	-0.137	0.428	0.138	0.556	0.758	-0.478	-0.665	-0.436	-0.360
(13)	Migratory distance from East Africa (unadjusted)	0.284	-0.749	0.238	-0.099	-0.015	0.099	-0.168	-0.424	0.267	-0.279
(14)	Migratory distance from East Africa (ancestry adjusted)	0.263	-1.000	0.232	-0.094	0.007	0.220	-0.199	-0.411	0.180	-0.335
(15)	Aerial distance from East Africa (unadjusted)	0.335	-0.814	0.188	-0.047	0.064	0.195	-0.229	-0.443	0.239	-0.338
(16)	Aerial distance from East Africa (ancestry adjusted)	0.331	-0.942	0.166	-0.027	0.125	0.299	-0.266	-0.410	0.119	-0.370

		(11)	(12)	(13)	(14)	(15)
(11)	Percentage of population of European descent	1.000				
(12)	Years of schooling	0.773	1.000			
(13)	Migratory distance from East Africa (unadjusted)	0.253	0.191	1.000		
(14)	Migratory distance from East Africa (ancestry adjusted)	0.153	0.137	0.749	1.000	
(15)	Aerial distance from East Africa (unadjusted)	0.283	0.261	0.939	0.814	1.000
(16)	Aerial distance from East Africa (ancestry adjusted)	0.224	0.233	0.677	0.942	0.829

H Evidence from Evolutionary Biology

The proposed diversity hypothesis suggests that there exists a trade-off with respect to genetic diversity in human populations. Specifically, higher diversity generates social benefits by enhancing society's productivity through efficiency gains via complementarities across different productive traits, by increasing society's resilience against negative productivity shocks, and by fostering society's adaptability to a change in the technological environment. Higher diversity also generates social costs, however, by increasing the likelihood of miscoordination and distrust between interacting agents and by inhibiting the emergence and sustainability of cooperative behavior in society. Indeed, the ideas underlying these channels ensue rather naturally from well-established concepts in evolutionary biology.

The following narrative discusses some of the analogous arguments from the field of evolutionary biology and presents supporting evidence from recent scientific studies. These studies typically focus on organisms like ants, bees, wasps, and certain species of spiders and birds that are not only amenable to laboratory experimentation but also display a relatively high degree of social behavior in nature, such as living in task-directed hierarchical societies, characterized by division of labor, or engaging in cooperative rearing of their young. The motivation behind studying such organisms is often related to the work of sociobiologists (e.g., Wilson, 1978; Hölldobler and Wilson, 1990) who have argued that the application of evolutionary principles in explaining the behavior of social insects lends key insights to the understanding of social behavior in more complex organisms like humans.

H.1 Benefits of Genetic Diversity

The notion that genetic diversity within a given population is beneficial for individual reproductive fitness, and thus for the adaptability and survivability of the population as a whole, is one of the central tenets of Darwin's (1859) theory of evolution. In the short term, by reducing the extent of inbreeding, genetic diversity prevents the propagation of potentially deleterious traits in the population across generations (Houle, 1994). In the long term, by permitting the force of natural selection to operate over a wider spectrum of traits, genetic diversity increases the population's capacity to adapt to changing environmental conditions (Frankham et al., 1999).

To elaborate further, the study by Frankham et al. (1999) provides clear experimental evidence for the beneficial effect of genetic diversity in enhancing the survivability of populations under deleterious changes in the environment. In their experiment, populations of the common fruit fly, *Drosophila melanogaster*, were subjected to different rates of inbreeding, and their ability to tolerate increasing concentrations of sodium chloride, or common salt, which is harmful for this species of flies, was compared with that of outbred base populations. Indeed, the less diverse inbred populations became extinct at significantly lower concentrations of sodium chloride than the more genetically diverse base populations.

In related studies, Tarpy (2003) and Seeley and Tarpy (2007) employ the honeybee, *Apis mellifera*, to demonstrate that polyandry, i.e., the practice of mating with multiple male drones by queen bees, may be an adaptive strategy since the resultant increase in genetic diversity increases the colony's resistance to disease. For instance, having created colonies headed by queens that had been

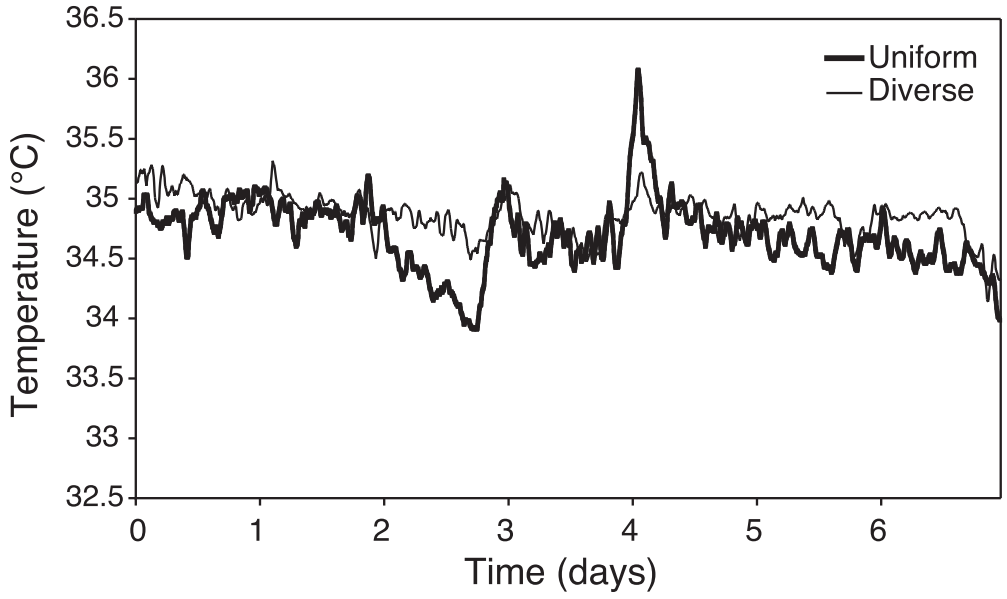


FIGURE H.1: Thermoregulation in Genetically Uniform vs. Diverse Honeybee Colonies

Notes: This figure depicts the results from the experimental study by Jones et al. (2004), illustrating the superior thermoregulation performance, as reflected by lower intertemporal temperature volatility, of a genetically diversity honeybee colony in comparison to a genetically uniform honeybee colony.

Source: Jones et al. (2004).

artificially inseminated by either one or ten drones, Seeley and Tarpy inoculated these colonies with spores of *Paenibacillus larvae*, a bacterium that causes a highly virulent disease in honeybee larvae. The researchers found that, on average, colonies headed by multiple-drone inseminated queens had markedly lower disease intensity and higher colony strength relative to colonies headed by single-drone inseminated queens.

In addition to increasing disease resistance, it has been argued that genetic diversity within honeybee colonies provides them with a system of genetically-based task specialization, thereby enabling them to respond more resiliently to environmental perturbations (Oldroyd and Fewell, 2007). Evidence supporting this viewpoint is provided by the study of Jones et al. (2004). Honeybee colonies need to maintain their brood nest temperature between 32°C and 36°C, and optimally at 35°C, so that the brood develops normally. Workers regulate temperature by fanning hot air out of the nest when the temperature is perceived as being too high and by clustering together and generating metabolic heat when the temperature is perceived to be too low. Ideally, a graded rather than precipitous response is required to ensure that the colony does not constantly oscillate between heating and cooling responses. In their experiment, Jones et al. artificially constructed genetically uniform and diverse honeybee colonies and compared their thermoregulation performances under exposure to ambient temperatures. The researchers found that, over a period of 2 weeks, the within-colony variance in temperatures maintained by the diverse colonies (0.047°C) was less than one-third of the within-colony temperature variance maintained by the uniform ones (0.165°C) and that this difference in thermoregulation performance was statistically significant (F-statistic = 3.5, P-value

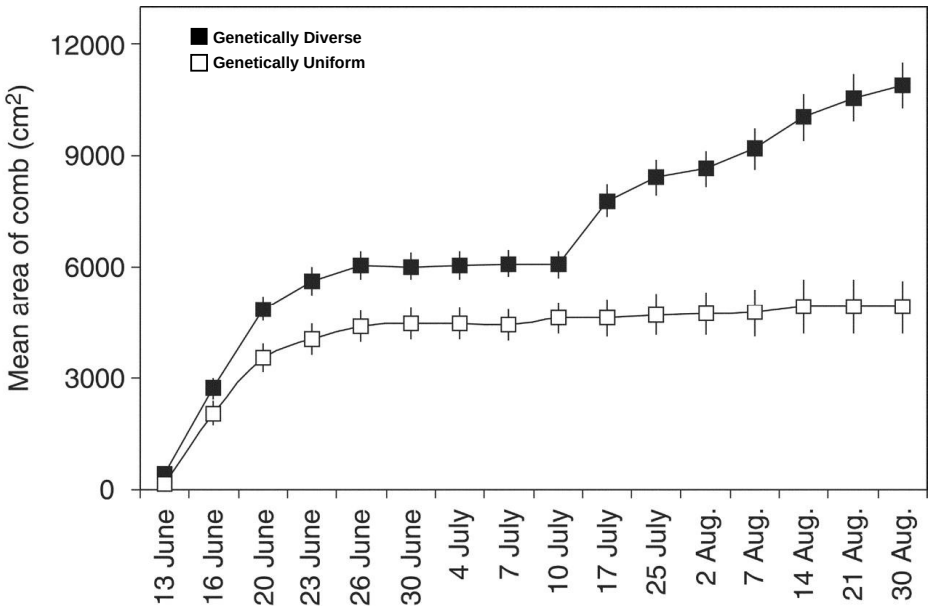


FIGURE H.2: Comb Area Growth in Genetically Diverse versus Uniform Honeybee Colonies

Notes: This figure depicts the results from the experimental study by Mattila and Seeley (2007), illustrating the superior productivity, as reflected by faster mean comb area growth, of genetically diversity honeybee colonies in comparison to genetically uniform honeybee colonies.

Source: Mattila and Seeley (2007).

< 0.001). Figure H.1 illustrates the superior thermoregulation performance of a genetically diverse colony, in comparison to that of a uniform one, in the Jones et al. experiment.

A popular hypothesis regarding the benefits of diversity, one that appears most analogous to the arguments raised in this paper, suggests that genetically diverse honeybee colonies may operate more efficiently by performing tasks better as a collective, thereby gaining a fitness advantage over colonies with uniform gene pools (Robinson and Page, 1989). Results from the experimental study by Mattila and Seeley (2007) provide evidence supporting this hypothesis. Since the channel highlighted by this hypothesis is closely related to the idea proposed in the current study, the remainder of this section is devoted to the Mattila and Seeley experiment.

A honeybee colony propagates its genes in two ways: by producing reproductive males (drones) and by producing swarms. Swarming occurs when a reproductive female (queen) and several thousand infertile females (workers) leave their colony to establish a new nest. Swarming is costly and perilous. With limited resources and labor, a swarm must construct new comb, build a food reserve, and begin rearing workers to replace an aging workforce. In temperate climates, newly founded colonies must operate efficiently because there is limited time to acquire the resources to support these activities. Colony founding through swarming is so difficult that only 20% of swarms survive their first year. Most do not gather adequate food to fuel the colony throughout the winter and therefore die of starvation. With the challenges of successful colony founding in mind, Mattila and Seeley conducted a long-term study to compare the development characteristics

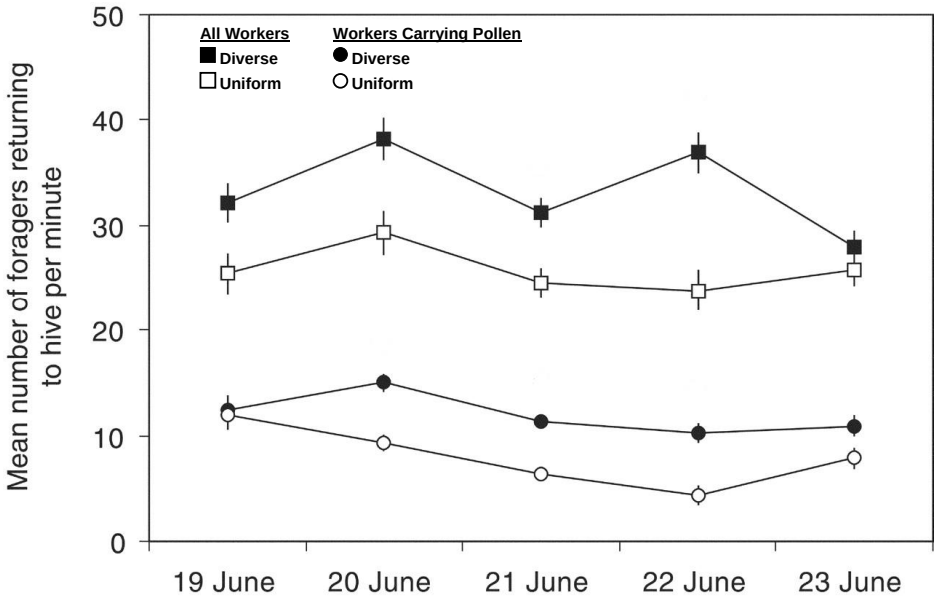


FIGURE H.3: Foraging Rates in Genetically Diverse versus Uniform Honeybee Colonies

Notes: This figure depicts the results from the experimental study by Mattila and Seeley (2007), illustrating the superior productivity, as reflected by a higher mean foraging rate, of genetically diversity honeybee colonies in comparison to genetically uniform honeybee colonies.

Source: Mattila and Seeley (2007).

of genetically diverse and genetically uniform colonies after a swarming event. The researchers began by creating genetically diverse colonies, using queens instrumentally inseminated with semen from multiple drones, and genetically uniform ones, using queens inseminated by one drone. They then generated swarms artificially, selecting from each colony the queen and a random subset of her worker offspring, and allowed these swarms to found new colonies. The observations in the Mattila and Seeley experiment begin on June 11, 2006, when the swarms established their new nest sites. In particular, they document colony development by measuring comb construction, brood rearing, foraging activity, food storage, population size, and mean weight gain at regular intervals.

As depicted in Figure H.2, Mattila and Seeley found that, during the first two weeks of colony development, colonies with genetically diverse worker populations built about 30% more comb than colonies with genetically uniform populations, a difference that was highly statistically significant (F-statistic = 25.7, P-value < 0.001). Furthermore, as illustrated in Figure H.3, during the second week of colony founding, genetically diverse colonies maintained foraging rates (measured as the number returning to hive per minute of either all workers and or only those carrying pollen) that were between 27% and 78% higher than those of genetically uniform colonies. Consequently, after two weeks of inhabiting their nest sites, genetically diverse colonies stockpiled 39% more food than the uniform ones. The researchers also found that production of new workers and brood rearing by existing workers were both significantly higher in the genetically diverse colonies within the first month of colony development. As a result of these various accumulated productivity gains, the

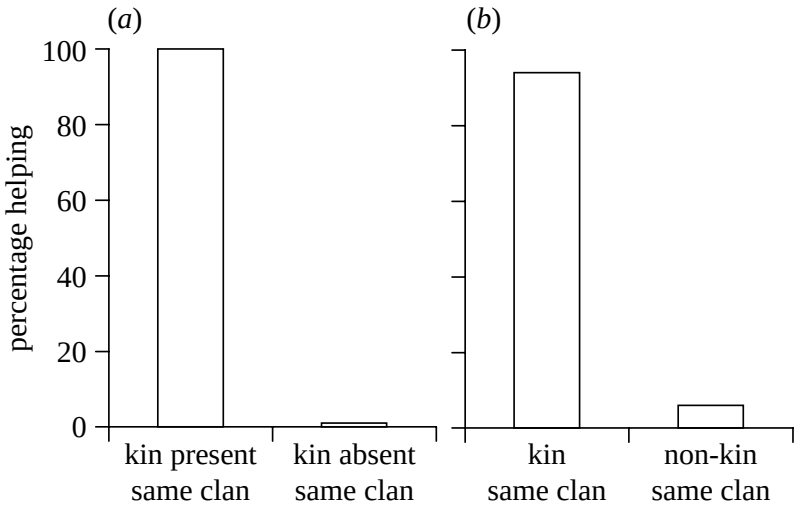


FIGURE H.4: Preferential Bias of Cooperation with Kin in the Long-Tailed Tit

Notes: This figure depicts the results from the experimental study by Russell and Hatchwell (2001), illustrating that, in the long-tailed tit (a species of cooperatively breeding birds), (i) the presence of genetic relatives (kin) within the social unit is a necessary condition for the prevalence of altruistic behavior (Panel (a)) and (ii) altruism is preferentially directed towards genetic relatives when both relatives and non-relatives are present within the same social unit (Panel (b)).

Source: Russell and Hatchwell (2001).

genetically diverse colonies all survived an unusually cold exposure, occurring two months after the establishment of their nest sites, that starved and killed about 50% of the genetically uniform colonies. Based on their findings, the authors conclude that collective productivity and fitness in honeybee colonies is indeed enhanced by intracolonial genetic diversity.

H.2 Benefits of Genetic Relatedness and Homogeneity

The notion that genetic relatedness between individuals, and genetic homogeneity of a group in general, can be collectively beneficial is highlighted in an extension of Darwinian evolutionary theory known as kin selection theory. In particular, the concept of "survival of the fittest" in standard Darwinian theory implies that, over time, the world should be dominated by selfish behavior since natural selection favors genes that increase an organism's ability to survive and reproduce. This implication of evolutionary theory remained at odds with the observed prevalence of altruistic and cooperative behavior in nature until the formalization of kin selection theory by Hamilton (1964) and Maynard Smith (1964). According to this influential theory, the indirect fitness gains of genetic relatives can in some cases more than compensate for the private fitness loss incurred by individuals displaying altruistic or cooperative behavior. Hence, given that relatives are more likely to share common traits, including those responsible for altruism or cooperation, kin selection provides a rationale for the propagation of cooperative behavior in nature.

An immediate implication of kin selection theory is that, when individuals can distinguish relatives from non-relatives (kin recognition), altruists should preferentially direct aid towards their

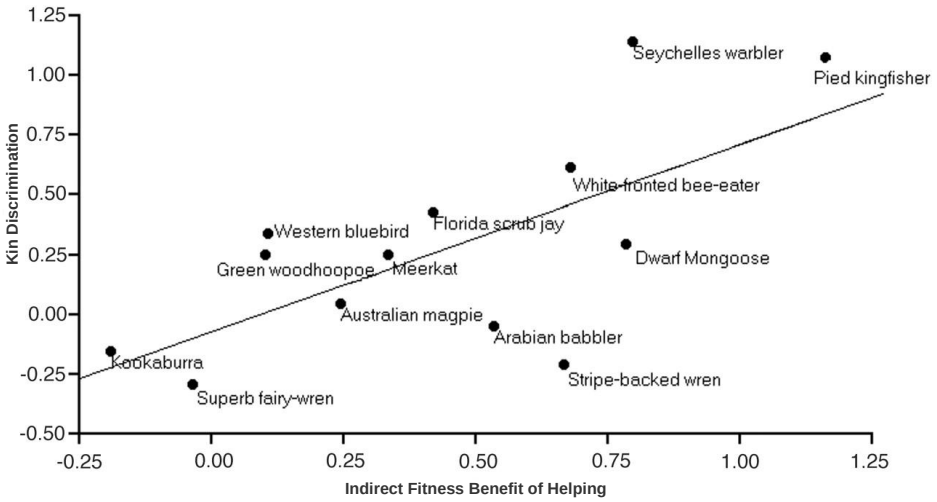


FIGURE H.5: Kin Discrimination and the Indirect Fitness Benefit from Altruism

Notes: This figure depicts the results from the study by Griffin and West (2003), illustrating that the extent of kin discrimination, i.e., the strength of the species-specific correlation between the amount of help in brood rearing and genetic relatedness, is higher in species where there is a larger indirect fitness benefit of altruism, measured in terms of relatives' offspring production and survival.

Source: Griffin and West (2003).

relatives (kin discrimination). The study by Russell and Hatchwell (2001) provides experimental evidence of this phenomenon in *Aegithalos caudatus*, a species of cooperatively breeding birds commonly known as the long-tailed tit. In this species, individuals distinguish between relatives and non-relatives on the basis of vocal contact cues (Sharp et al., 2005), and failed breeders can become potential helpers in rearing the young of successful breeders within the same social unit. In their research, Russell and Hatchwell designed an experiment to investigate whether the presence of kin within the social unit was a necessary condition for altruistic behavior and whether kin were preferred to non-kin when given the choice. As depicted in Figure H.4, the researchers found that failed breeders did not actually become helpers when kin were absent from the social unit (Panel (a)), but when both kin and non-kin were present in the same social unit, the majority of failed breeders provided brood-rearing assistance at the nests of kin (Panel (b)).

Another prediction of kin selection theory is that the extent of altruism should be positively correlated with the degree of genetic relatedness (between potential helpers and beneficiaries) and that this correlation should be stronger the greater the indirect fitness benefit from altruism. Empirical support for this prediction comes from a study by Griffin and West (2003) where relevant data from 18 collectively breeding vertebrate species was used to (i) test the relationship between the amount of help in brood rearing and relatedness and (ii) examine how this correlation varied with the benefit of helping (measured in terms of relatives' offspring production and survival). Specifically, the study exploited variation across social units within each species in genetic relatedness, the amount of help, and the indirect fitness benefit of helping. Consistently with kin selection theory, the researchers

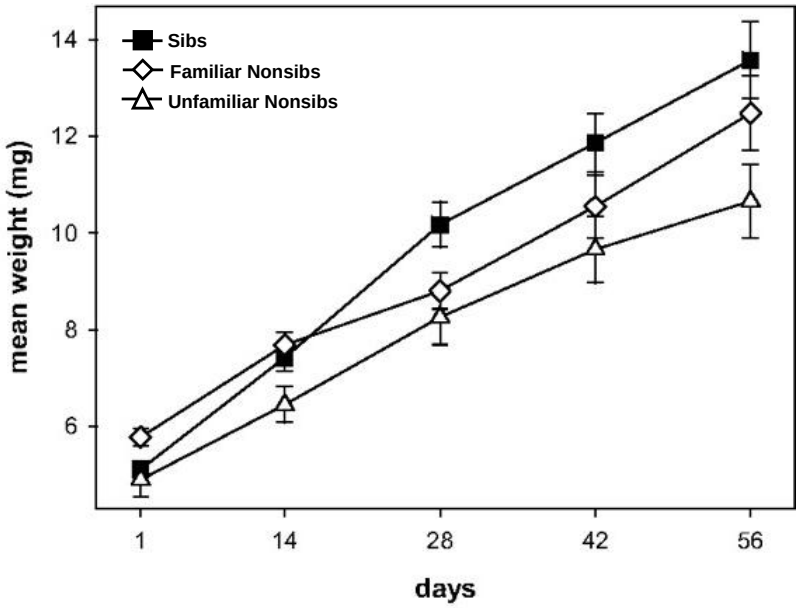


FIGURE H.6: Weight Growth in Kin versus Nonkin Groups of Cooperatively Feeding Spiders

Notes: This figure depicts the results from the experimental study by Schneider and Bilde (2008), illustrating the superior weight gain performance of groups of cooperatively feeding spiders where individuals were genetically related (sibs) in comparison to groups where individuals were either (i) genetically and socially unrelated (unfamiliar nonsibs) or (ii) genetically unrelated but socially related (familiar nonsibs).

Source: Schneider and Bilde (2008).

found that the cross-species average of the species-specific cross-social unit correlation between the amount of help and genetic relatedness was 0.33, a correlation that was statistically significantly larger than zero (P-value < 0.01). Moreover, the study also found that kin discrimination, i.e., the species-specific cross-social unit correlation between the amount of help and relatedness, was higher in species where the indirect fitness benefits from altruism were larger. Figure H.5 depicts the cross-species relationship found by Griffin and West between kin discrimination and the benefit from altruistic behavior.

While the studies discussed thus far provide evidence of a positive correlation between genetic relatedness and altruism, they do not substantiate the effect of relatedness on the other type of social behavior stressed by kin selection theory, that of mutually or collectively beneficial cooperation. This concept is directly associated with solving the problem of public goods provision due to the "tragedy of commons." In particular, cooperation within groups that exploit a finite resource can be prone to cheating whereby the selfish interests of individuals result in disadvantages for all members of the group. While cooperative behavior can be enforced through mechanisms such as reciprocity or punishment, kin selection provides a natural alternative for the resolution of such social dilemmas. Specifically, by helping relatives pass on shared genes to the next generation, cooperation between related individuals can be mutually beneficial. Experimental evidence on the importance of genetic relatedness for cooperative behavior comes from the study by Schneider and Bilde (2008) that

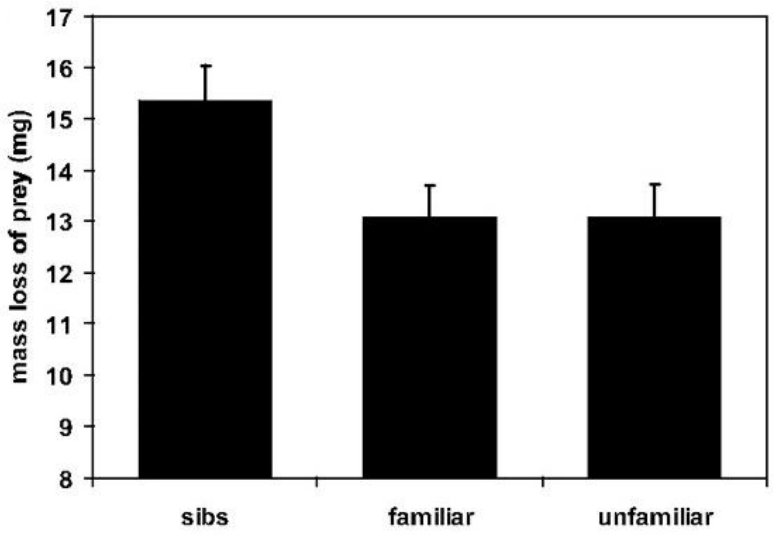


FIGURE H.7: Feeding Efficiency in Kin vs. Nonkin Groups of Cooperatively Feeding Spiders

Notes: This figure depicts the results from the experimental study by Schneider and Bilde (2008), illustrating the superior feeding efficiency of groups of cooperatively feeding spiders where individuals were genetically related (sibs) in comparison to groups where individuals were either (i) genetically and socially unrelated (unfamiliar nonsibs) or (ii) genetically unrelated but socially related (familiar nonsibs).

Source: Schneider and Bilde (2008).

investigates the role of kinship in cooperative feeding amongst the young in *Stegodyphus lineatus*, a species of spider displaying sociality in juvenile stages.

Schneider and Bilde argue that communally feeding spiders are ideal to investigate the costs and benefits of cooperation because of their mode of feeding. These spiders hunt cooperatively by building and sharing a common capture web, but they also share large prey items. Since spiders digest externally by first injecting their digestive enzymes and then extracting the liquidized prey content, communal feeding involves everyone injecting saliva into the same carcass and thus exploiting a common resource that was jointly created. Such a system is especially prone to cheating because each feeder can either invest in the digestion process by contributing enzymes or cheat by extracting the liquidized prey with little prior investment. The outcomes of such conflicts in a collective can thus be quantified by measuring feeding efficiency and weight gain. In this case, kin selection theory predicts that groups with higher mean genetic relatedness should outperform others on these biometrics due to a relatively lower prevalence of such conflicts.

To test this prediction, Schneider and Bilde conducted an experiment with three treatment groups of juvenile spiders: genetically related (sibs), genetically and socially unrelated (unfamiliar nonsibs), and genetically unrelated but socially related (familiar nonsibs). Social, as opposed to genetic, relatedness refers to familiarity gained through learned association as a result of being raised by the same mother (either foster or biological) in pre-juvenile stages. The third treatment group therefore allowed the researchers to control for nongenetic learned associations that could erroneously be interpreted as kin-selected effects. In their experiment, Schneider and Bilde followed two group-level

outcomes over time. They measured growth as weight gained over a period of eight weeks, and they measured feeding efficiency of the groups by quantifying the mass extracted from prey in repeated two-hour assays of cooperative feeding.

As depicted in Figure H.6, consistently with kin selection, sib groups gained significantly more weight than genetically unrelated groups (both familiar and unfamiliar) over the experimental period of 8 weeks (F-statistic $= 9.31$, P-value < 0.01), and while nonsib unfamiliar spider groups had a higher start weight than the two other groups, sib groups overtook them by following a significantly steeper growth trajectory. Indeed, as Figure H.7 illustrates, this growth pattern was due to the higher feeding efficiency of sib groups compared with nonsib groups, the former extracting significantly more mass from their prey during a fixed feeding duration (F-statistic $= 8.91$, P-value < 0.01). Based on these findings, Schneider and Bilde conclude that genetic similarity facilitates cooperation by reducing cheating behavior and, thereby, alleviates the negative social impact of excessive competition.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. "Institutions as a Fundamental Cause of Long-Run Growth." In *Handbook of Economic Growth, Vol IA.* , ed. Philippe Aghion and Steven N. Durlauf. Amsterdam, The Netherlands:Elsevier North-Holland.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature*, 43(3): 762–800.

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

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review*, 101(5): 2003–2041.

Ashraf, Quamrul, Oded Galor, and Ömer Özak. 2010. "Isolation and Development." *Journal of the European Economic Association*, 8(2-3): 401–412.

Bairoch, Paul. 1988. *Cities and Economic Development: From the Dawn of History to the Present.* Trans. by Christopher Braider. Chicago, IL:The University of Chicago Press.

Barbujani, Guido, Arianna Magagni, Eric Minch, and L. Luca Cavalli-Sforza. 1997. "An Apportionment of Human DNA Diversity." *Proceedings of the National Academy of Sciences*, 94(9): 4516–4519.

Barro, Robert J., and Jong-Wha Lee. 2001. "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers*, 53(3): 541–563.

Barro, Robert J., and Rachel M. McCleary. 2003. "Religion and Economic Growth across Countries." *American Sociological Review*, 68(5): 760–781.

Cann, Howard M., Claudia de Toma, Lucien Cazes, Marie-Fernande Legrand, Valerie Morel, Laurence Piouffre, et al. 2002. "A Human Genome Diversity Cell Line Panel." *Science*, 296(5566): 261–262.

Cavalli-Sforza, L. Luca. 2005. "The Human Genome Diversity Project: Past, Present and Future." *Nature Reviews Genetics*, 6(4): 333–340.

Chandler, Tertius. 1987. *Four Thousand Years of Urban Growth: An Historical Census.* Lewiston, NY:The Edwin Mellen Press.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1): 1–45.

Dalgaard, Carl-Johan, and Holger Strulik. 2010. "The Physiological Foundations of the Wealth of Nations." Discussion Paper No. 10-05, University of Copenhagen.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection.* London, UK:John Murray.

Desmet, Klaus, Michel Le Breton, Ignacio Ortuño-Ortín, and Shlomo Weber. 2011. "The Stability and Breakup of Nations: A Quantitative Analysis." *Journal of Economic Growth*, 16(3): 183–213.

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

Diamond, Jared. 2002. "Evolution, Consequences and Future of Plant and Animal Domestication." *Nature*, 418(6898): 700–707.

Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics*, 112(4): 1203–1250.

Engerman, Stanley L., and Kenneth L. Sokoloff. 2000. "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic Perspectives*, 14(3): 217–232.

Frankham, Richard, Kelly Lees, Margaret E. Montgomery, Phillip R. England, Edwin H. Lowe, and David A. Briscoe. 1999. "Do Population Size Bottlenecks Reduce Evolutionary Potential?" *Animal Conservation*, 2(4): 255–260.

Gallup, John L., and Jeffrey D. Sachs. 2001. "The Economic Burden of Malaria." *The American Journal of Tropical Medicine and Hygiene*, 64(1-2): 85–96.

Gallup, John L., Jeffrey D. Sachs, and Andrew D. Mellinger. 1999. "Geography and Economic Development." *International Regional Science Review*, 22(2): 179–232.

Galor, Oded. 2011. *Unified Growth Theory.* Princeton, NJ:Princeton University Press.

Galor, Oded, and Omer Moav. 2002. "Natural Selection and the Origin of Economic Growth." *Quarterly Journal of Economics*, 117(4): 1133–1191.

Galor, Oded, and Omer Moav. 2007. "The Neolithic Revolution and Contemporary Variations in Life Expectancy." Working Paper, Department of Economics, Brown University.

Galor, Oded, and Stelios Michalopoulos. 2011. "Evolution and the Growth Process: Natural Selection of Entrepreneurial Traits." *Journal of Economic Theory*, forthcoming.

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

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review*, 83(3): 525–548.

Griffin, Ashleigh S., and Stuart A. West. 2003. "Kin Discrimination and the Benefit of Helping in Cooperatively Breeding Vertebrates." *Science*, 302(5645): 634–636.

Griffiths, Anthony J. F., Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart. 2000. *An Introduction to Genetic Analysis.* New York, NY:W. H. Freeman & Co.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *Quarterly Journal of Economics*, 124(3): 1095–1131.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114(1): 83–116.

Hamilton, William D. 1964. "The Genetical Evolution of Social Behaviour I and II." *Journal of Theoretical Biology*, 7(1): 1–52.

Hayes, Theodore R. 1996. "Dismounted Infantry Movement Rate Study." U.S. Army Research Institute of Environmental Medicine.

Hölldobler, Bert, and Edward O. Wilson. 1990. *The Ants.* Cambridge, MA:The Belknap Press of Harvard University Press.

Houle, David. 1994. "Adaptive Distance and the Genetic Basis of Heterosis." *Evolution*, 48(4): 1410–1417.

Jones, Eric L. 1981. *The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia.* Cambridge, UK:Cambridge University Press.

Jones, Julia C., Mary R. Myerscough, Sonia Graham, and Benjamin P. Oldroyd. 2004. "Honey Bee Nest Thermoregulation: Diversity Promotes Stability." *Science*, 305(5682): 402–404.

Lagerlöf, Nils-Petter. 2007. "Long-Run Trends in Human Body Mass." *Macroeconomic Dynamics*, 11(3): 367–387.

Landes, David S. 1998. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor.* New York, NY:W. W. Norton & Co.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1999. "The Quality of Government." *Journal of Law, Economics, and Organization*, 15(1): 222–279.

Macaulay, Vincent, Catherine Hill, Alessandro Achilli, Chiara Rengo, Douglas Clarke, William Meehan, et al. 2005. "Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes." *Science*, 308(5724): 1034–1036.

Mattila, Heather R., and Thomas D. Seeley. 2007. "Genetic Diversity in Honey Bee Colonies Enhances Productivity and Fitness." *Science*, 317(5836): 362–364.

Maynard Smith, John. 1964. "Group Selection and Kin Selection." *Nature*, 201(4924): 1145–1147.

McEvedy, Colin, and Richard Jones. 1978. *Atlas of World Population History.* New York, NY:Penguin Books Ltd.

Michalopoulos, Stelios. 2011. "The Origins of Ethnolinguistic Diversity." *American Economic Review*, forthcoming.

Modelski, George. 2003. *World Cities: -3000 to 2000.* Washington, DC:FAROS 2000.

Mokyr, Joel. 1990. *The Lever of Riches: Technological Creativity and Economic Progress.* New York, NY:Oxford University Press.

Murphy, Kevin M., and Robert H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics*, 3(4): 370–379.

New, Mark, David Lister, Mike Hulme, and Ian Makin. 2002. "A High-Resolution Data Set of Surface Climate Over Global Land Areas." *Climate Research*, 21(1): 1–25.

Nordhaus, William D. 2006. "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences*, 103(10): 3510–3517.

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

Oldroyd, Benjamin P., and Jennifer H. Fewell. 2007. "Genetic Diversity Promotes Homeostasis in Insect Colonies." *Trends in Ecology and Evolution*, 22(8): 408–413.

Olsson, Ola, and Douglas A. Hibbs, Jr. 2005. "Biogeography and Long-Run Economic Development." *European Economic Review*, 49(4): 909–938.

Özak, Ömer. 2010. "The Voyage of Homo-Economicus: Some Economic Measures of Distance." Mimeo, Brown University.

Pagan, Adrian. 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors." *International Economic Review*, 25(1): 221–247.

Peregrine, Peter N. 2003. "Atlas of Cultural Evolution." *World Cultures: Journal of Comparative and Cross-Cultural Research*, 14(1): 1–75.

Pomeranz, Kenneth. 2000. *The Great Divergence: Europe, China and the Making of the Modern World Economy.* Princeton, NJ:Princeton University Press.

Prugnolle, Franck, Andrea Manica, and François Balloux. 2005. "Geography Predicts Neutral Genetic Diversity of Human Populations." *Current Biology*, 15(5): R159–R160.

Putterman, Louis. 2008. "Agriculture, Diffusion, and Development: Ripple Effects of the Neolithic Revolution." *Economica*, 75(300): 729–748.

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

Ramachandran, Sohini, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza. 2005. "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*, 102(44): 15942–15947.

Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney. 2002. "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change." *Global Ecology and Biogeography*, 11(5): 377–392.

Robinson, Gene E., and Robert E. Page, Jr. 1989. "The Genetic Basis of Division of Labor in an Insect Society." In *The Genetics of Social Evolution.* , ed. Michael D. Breed and Robert E. Page, Jr. Boulder, CO:Westview Press.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development." *Journal of Economic Growth*, 9(2): 131–165.

Russell, Andrew F., and Ben J. Hatchwell. 2001. "Experimental Evidence for Kin-Biased Helping in a Cooperatively Breeding Vertebrate." *Proceedings of the Royal Society: Biological Sciences*, 268(1481): 2169–2174.

Sachs, Jeffrey D., and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity*, 26(1): 1–118.

Schneider, Jutta M., and Trine Bilde. 2008. "Benefits of Cooperation with Genetic Kin in a Subsocial Spider." *Proceedings of the National Academy of Sciences*, 105(31): 10843–10846.

Seeley, Thomas D., and David R. Tarpy. 2007. "Queen Promiscuity Lowers Disease within Honeybee Colonies." *Proceedings of the Royal Society: Biological Sciences*, 274(1606): 67–72.

Sharp, Stuart P., Andrew McGowan, Matthew J. Wood, and Ben J. Hatchwell. 2005. "Learned Kin Recognition Cues in a Social Bird." *Nature*, 434(7037): 1127–1130.

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

Stringer, Chris B., and Peter Andrews. 1988. "Genetic and Fossil Evidence for the Origin of Modern Humans." *Science*, 239(4845): 1263–1268.

Tabellini, Guido. 2008. "Institutions and Culture." *Journal of the European Economic Association*, 6(2-3): 255–294.

Tarpy, David R. 2003. "Genetic Diversity within Honeybee Colonies Prevents Severe Infections and Promotes Colony Growth." *Proceedings of the Royal Society: Biological Sciences*, 270(1510): 99–103.

Wang, Sijia, Cecil M. Lewis Jr., Mattias Jakobsson, Sohini Ramachandran, Nicolas Ray, Gabriel Bedoya, et al. 2007. "Genetic Variation and Population Structure in Native Americans." *PLoS Genetics*, 3(11): 2049–2067.

Weber, Max. 1905. *The Protestant Ethic and the Spirit of Capitalism.* Trans. by Talcott Parsons and Anthony Giddens, 1930. London, UK:Allen & Unwin.

Wilson, Edward O. 1978. *On Human Nature.* Cambridge, MA:Harvard University Press.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* . 2 ed., Cambridge, MA:The MIT Press.