The Out of Africa Hypothesis of Comparative Development Reflected by Nighttime Light Intensity[*]

Quamrul Ashraf   Oded Galor   Marc Klemp

April 26, 2014

Abstract

This research establishes that migratory distance from the cradle of anatomically modern humans in East Africa and its effect on the distribution of genetic diversity across countries has a hump-shaped effect on nighttime light intensity per capita as observed by satellites, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. The finding lends further credence to the hypothesis that a significant portion of the variation in the standard of living across the globe can be attributed to factors that were determined in the distant past.

Keywords Nighttime light intensity, Out of Africa Hypothesis of Comparative Development, Genetic Diversity, Comparative Development, migratory distance from Africa

JEL Classification Codes N10, N30, N50, O10, O50, Z10

1 Introduction

An emerging body of evidence suggests that deeply-rooted factors, determined tens of thousands of years ago, have significantly affected the level of diversity and the course of comparative economic development from the dawn of human civilization to the contemporary era. In particular, Ashraf and Galor (2013) 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. They establish that 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.

The level of economic development in the contemporary period is captured in this line of research by income per capita. This proxy for the standard of living, however, may be systematically mis-measured. Moreover, these mis-measurements could be correlated with genetic diversity and economic development. In particular, in less developed economies, a significant part of economic activities is conducted in the shadow economy and thus reported income per capita may systematically underestimate the standard of living across parts of the globe. Furthermore, given the established adverse effect of genetic diversity on social cohesion, the prevalence of under-reporting of the standard of living may be positively correlated with the degree of genetic diversity. Thus, the established hump-shaped relationship between reported income per capita and genetic diversity could be an artifact of this systematic mis-measurement and the actual standard of living could potentially be monotonically decreasing in the degree of genetic homogeneity.

Using variation in night-light intensity observed by satellites as a proxy for the overall level of economic activity, including that in the informal economy, this research establishes that genetic diversity has a robust hump-shaped effect on living standards across the globe.

2 Data and Empirical Specification

The empirical analysis examines the reduced-form effects of predicted genetic diversity on satellite-measures of yearly light intensity per capita and per area in each country across the globe.

2.1 Dependent Variables

The main dependent variables are the average standard of living in each country over the period 1992–2012, as captured by the average (i) light intensity per capita, (ii) light intensity per square kilometer, and (iii) income per capita, over this time period.[1]

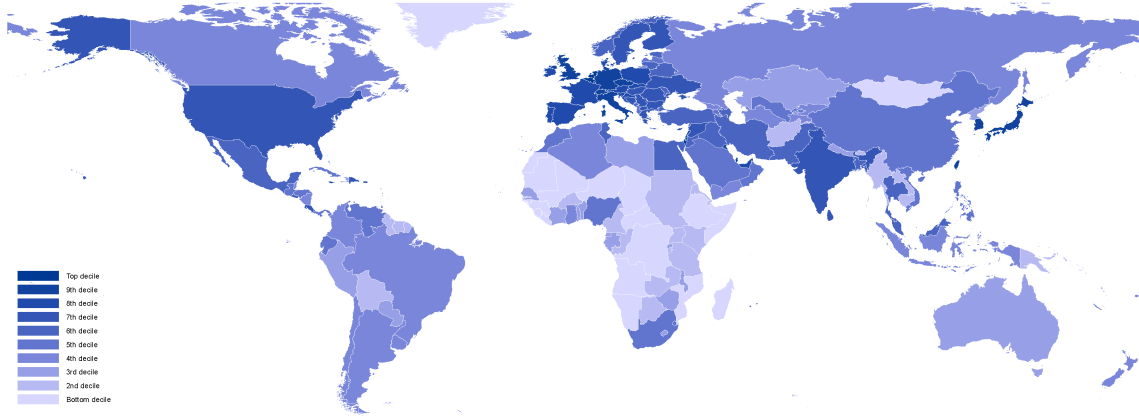[1]Data description and sources can be found in the Appendix.

Figure 1: The distribution of nighttime light intensity per area across countries.

Satellite-captured images of global night-light emission are available for each year in the period 1992–2012, for 30 arc second grids, spanning -180 to 180 degrees longitude and -65 to 75 degrees latitude. Each cell of approximately one square km (as measured at equator), is assigned an integer ranging from 0 to 63 representing its yearly luminosity.[2] The average yearly luminosity for each cell over this 21-year period is depicted in Figure A.1 in the Appendix.[3] The average yearly luminosity over this 21-year period for a country is therefore the aggregate luminosity over all cells within the boundaries of the country.[4] The average light intensity per capita in each country is depicted in Figure 1, and the average light intensity per square kilometer in Figure 2.

The dataset is potentially affected by measurement errors for several reasons. Cells at the extreme bounds of night-light emission (i.e., those with values of 0 or 63) may be bottom or top-censored. Moreover, some cells may be affected by overglow, (i.e., light emitted within one pixel might "spillover" to nearby pixels) and blooming, (i.e., artificial light emission may be magnified over certain terrains, such as water and snow). These sources of measurement errors, however, are unlikely to affect the analysis since the measure of light intensity used is based on the aggregate light emission at the national levels (rather than at the pixel level) and since it reflects the average night-light emission based on a 21-year period. In addition, the empirical analysis accounts for a wide range of confounding geographic characteristics, such as absolute latitude, proximity to water, and regional fixed effects, and thus the potential effect of blooming is further mitigated.

---

[2]Lights emitted from gas flaring used in petroleum refineries, chemical plants, natural gas processing plants, and oil and gas wells. This exclusion makes use of the Global Gas Flaring Shapefiles provided by NOAA-NGDC and found at http://ngdc.noaa.gov/eog/interest/gas_flares_countries_shapefiles.html. Due to problems related to the shape files provided by NOAA-NGDC, this correction is not imposed for the three countries Côte d'Ivoire, Ghana, and Mauritius.

[3]The data is originated from separate two satellites in 1994 and the period 1997–2007, resulting in 33 yearly data points for each cell, which are averaged to obtain the average yearly light intensity over this 21-year period.

[4]The country borders are defined by the World Borders Dataset, found at http://thematicmapping.org/downloads/world_borders.php.
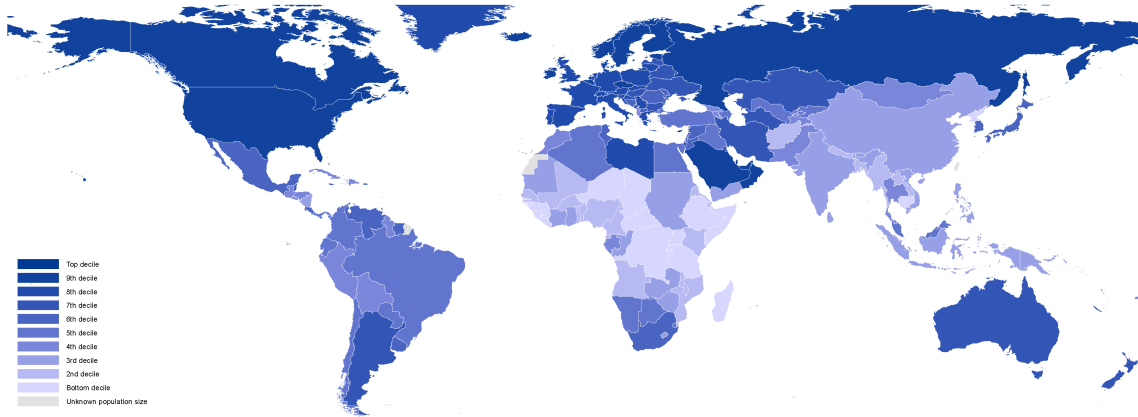
Figure 2: The distribution of nighttime light intensity per capita across countries.

Night-light measures of the standard of living and GDP per capita are positively correlated, in accordance with earlier findings (Henderson et al., 2012). In particular, log light intensity per capita, and log light intensity per square kilometer have a significant and positive correlation with the log of GDP per capita as depicted in Figure 3. For comparability, since the lights-based measures of living standards reflect the average over the period 1992–2012, GDP per capita (PPP-adjusted, constant 2005 international dollars) is averaged over the same 21-year period.[5]

2.2   Independent Variable: Genetic Diversity

The independent variable is the measure of contemporary genetic diversity for national populations as constructed by Ashraf and Galor (2013). Population geneticists measure the extent of diversity in genetic material across individuals within a given population (e.g., an ethnic group) using an index called expected heterozygosity. Like most other measures of diversity, this index may be interpreted 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. Given allelic frequencies for a particular gene or DNA locus, a gene-specific heterozygosity statistic may be computed (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.

While existing data on genetic diversity pertain only to ethnic groups, a large number of national populations today are composed of multiple ethnicities, some of which may not be indigenous to their current geographical locations. Hence, the index of genetic diversity for contemporary national

---

[5]It should be noted, that for 8 countries, the GDP per capita for some years in the period 1992–2012 is not given by the World Bank. In these cases, the average GDP per capita over the period 1992–2012 is calculated based on the years for which GDP per capita is known. The quadratic expression uncovered in Table 3 remains highly significant in the presence of all the relevant control variables if these 8 observations are excluded from the data.
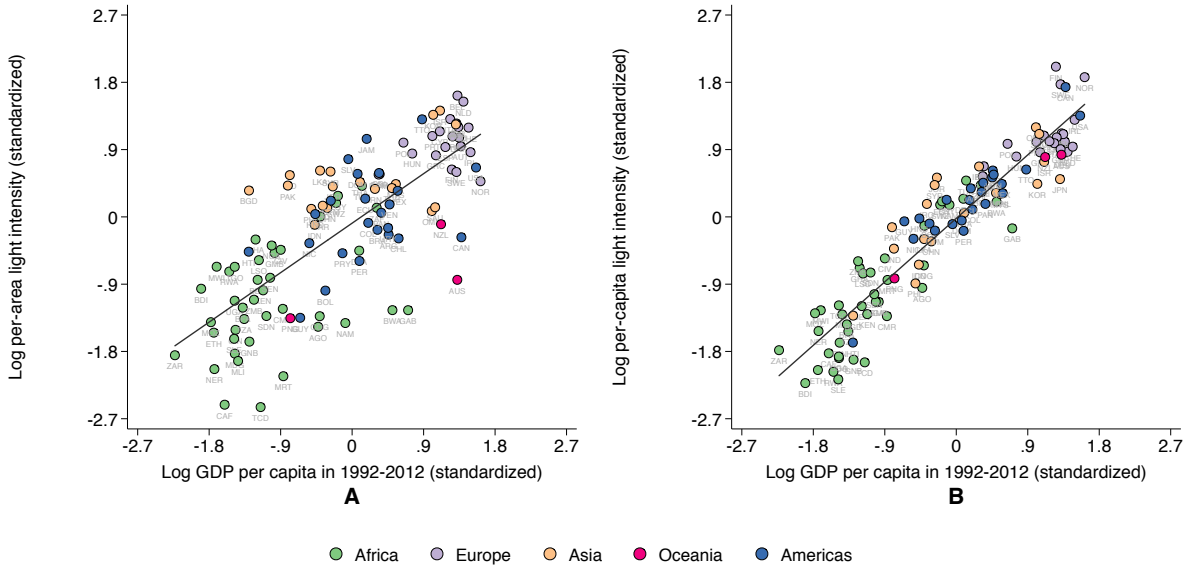
Figure 3: This figure depicts the association between log light intensity per capita in standard deviations from the mean and log GDP per capita in standard deviations from the mean (A), and the association between log per-capita light intensity in standard deviations from the mean and log per-capita GDP in standard deviations from the mean (B), for the sample of 111 countries corresponding to Table 3, column 5 and 6.

populations, as constructed by Ashraf and Galor (2013), accounts for the expected heterozygosity of each sub-national group as well as the additional component of diversity at the country level that arises from the pairwise genetic distances amongst its precolonial ancestral populations.

Moreover, since genetic diversity in the Human Genome Diversity Project is limited to only 53 ethnic groups, the highly significant negative linear relationship between genetic diversity of indigenous populations and migratory distance from the cradle of humankind (reflecting the Out of Africa Hypothesis and the Serial Founder Effect) is exploited in order to project genetic diversity for each ancestral population of each sub-national group based on its migratory distance from Africa.[6] Similarly, the highly significant positive linear relationship between pairwise genetic distances amongst indigenous populations and their pairwise migratory distances is exploited to project pairwise genetic distances amongst precolonial ancestral populations. Furthermore, the use of projected, rather than actual, diversity alleviates potential concerns about the endogeneity of the levels of genetic diversity and the level of economic development.

---

[6]According to the Out of Africa Hypothesis, the human species, having evolved to its modern form in East Africa some 150,000 years ago, embarked on populating the entire globe in a stepwise migration process beginning about 70,000–90,000 BP. Furthermore, the contemporary distribution of diversity across indigenous ethnic groups reflects a serial founder effect originating in East Africa. Accordingly, since the populating of the world occurred in a series of stages where sub-groups left initial colonies to create new colonies further away, carrying with them only a portion of the genetic diversity of their parental colonies, contemporary genetic diversity in human populations decrease with increasing distance along prehistoric migratory paths from East Africa (e.g., Ramachandran et al., 2005; Prugnolle et al., 2005).

2.3 Control Variables

The analysis accounts for a wide range of potentially confounding factors that have been associated with the vast differences in living standards across the globe. The baseline analysis accounts for a set of exogenous geographical controls. In addition, the analysis accounts two alternative channels, the Neolithic channel, and the institutional cultural and fractionalization measures through which genetic diversity may operate.

In particular, since the measure of genetic diversity exploited in the analysis is highly correlated with migratory distance from East Africa, geographical attributes that may be correlated with migratory distance could have a confounding effect on economic outcome through channels unrelated to intra-population diversity. To ensure that migratory from Africa and thus genetic diversity is not simply capturing an indirect effect of distance from the equator on the standard of living (e.g., through a climate and institutional related channel), absolute latitude is included in the regressions as one of the baseline geographical controls. Similarly, mean distance to the nearest waterway that may be correlated with migratory distance from Africa is introduced as a control variable to account for its potentially confounding effect on trade. In addition, the analysis accounts for the potential confounding effect of the percentage of arable land.

Moreover, in the process of the prehistoric demic expansion of anatomically modern humans from Africa to the rest of the world, population bottlenecks that occurred at specific inter-continental crossings (e.g., the Bering Strait) led to discrete spatial differences in observed genetic diversity across these crossings. Therefore, variation in genetic diversity may be correlated with cross-continental variation in unobserved (or observed or but imprecisely measured) continent-specific characteristics. To ensure that the estimated effect of genetic diversity on the standard of living is not simply reflecting the latent influence of these time-invariant continent-specific cultural, institutional, and geographical factors, a complete set of regional dummies is included as a standard control throughout the analysis.[7]

3   Correlation between Lights-Based Measures of Living Standards and GDP per Capita

Table 1 establishes a highly significant positive linear association between (i) log light intensity per square kilometer and GDP per capita and (ii) log the light intensity per capita and log GDP per capita for the period 1992–2012. Furthermore, the table establishes that these associations are robust to the inclusion or the exclusion of cells associated with gas flaring and to controlling for the potential confounding effects of log absolute latitude as well as regional fixed effects.

---

[7]This issue is perhaps more pronounced for cultural and institutional factors because of the possibility of systematic measurement errors at the continent level in variables reflecting cultural and institutional characteristics. The most credible way of addressing these concerns is the inclusion of regional fixed effects throughout the analysis, in order to mitigate the effect of both omitted variable bias and the bias arising from systematic measurement error in potentially correlated controls.

Table 1: The Association Between Lights-Based Measures of Development and GDP per Capita, 1992–2012

Dependent variable (in logs):	Light intensity per area where data assoc. with gas flaring are				Light intensity per capita where data assoc. with gas flaring are			
	Excluded			Included	Excluded			Included
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log average per-capita GDP (1992–2012)	1.198*** (.0792)	1.130*** (.0864)	.991*** (.101)	1.022*** (.0979)	1.066*** (.0523)	.926*** (.0518)	.735*** (.0547)	.773*** (.0555)
Log absolute latitude		.189* (.100)	.000329 (.115)	-.0472 (.112)		.392*** (.0601)	.181*** (.0624)	.130** (.0634)
Regional fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Number of observations	182	182	182	182	182	182	182	182
Adjusted R^2	.557	.564	.602	.615	.696	.753	.816	.811

This table establishes the significant linear association between (i) log mean light intensity and log per capita GDP (columns 1–4) and (ii) log per capita light intensity and log per capita GDP (columns 5–8) while controlling for log absolute latitude and regional fixed effects. The measures of light intensity excludes light emission from gas flaring in columns 1–3 and 6–8 and includes light emission from gas flaring in columns 4 and 8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The left segment of Table 1 (columns 1–4) establishes that over the period 1992–2012, for a sample of 182 countries, the association between log GDP per capita and log light intensity per square kilometer is positive and highly significant. In particular, in the absence of any controls, a one percent increase in the average GDP per capita is associated with a 1.20 percent increase in light intensity per square kilometer (column 1). This association is depicted in panel A of Figure 3, for the sample of 111 countries that constitute the baseline regression sample in Table 2. Columns 2 and 3 establish that the positive linear association remains highly significant statistically when on accounts for the confounding effects of log absolute latitude (column 2) as well as regional fixed effects (column 3). A one percent increase in the average GDP per capita is associated with a 0.99 percent increase in light intensity per square kilometer. Finally, column 4 establishes that the association reported in column 3, is robust to the inclusion of cells associated with gas flaring.

The right segment of Table 1 (column 5–8) establishes that over the period, 1992–2012, and for a sample of 182 countries, the association between log GDP per capita and log light intensity per capita is positive and highly significant. In particular, in the absence of any controls, a one percent increase in the average GDP per capita is associated with a 1.07 percent increase in light intensity per capita (column 5). Columns 6 and 7 establish that the positive association remains highly statistically significant while controlling for the log absolute latitude (column 6) and as well

---

[8]Given the quadratic nature of this relationship, the 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 squared value) of log per-capita light intensity plus the residuals from the full regression model, keeping the values of the control variables at zero (given the linear model, holding the control variables at other values, like their medians, would only shift the figure up or down on the second axis). 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 over all non-monotonic effect of genetic homogeneity in one scatter plot.
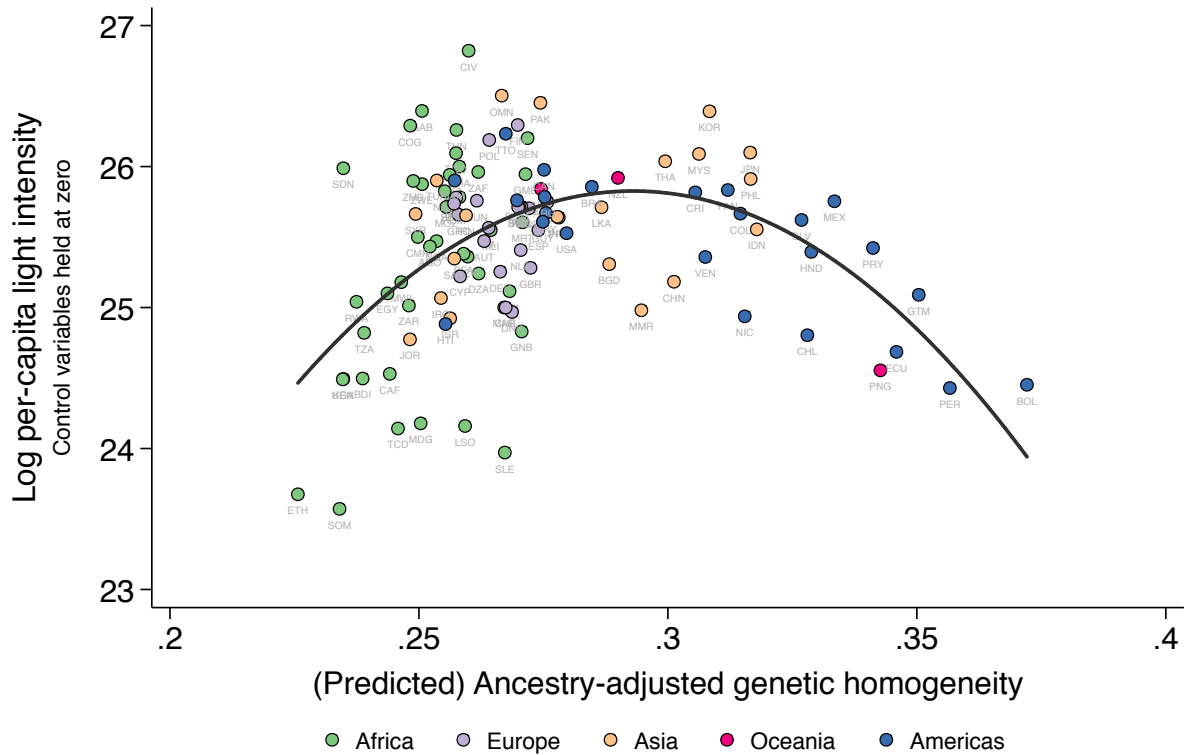
Figure 4: This figure depicts the hump-shaped effect of ancestry-adjusted genetic homogeneity (i.e., one minus ancestry-adjusted genetic diversity) on log per-capita light intensity in 1992–2012 for a sample of 111 countries, conditional on the log ancestry-adjusted timing of the Neolithic Revolution, log arable land area, a vector of institutional, cultural, and geographical determinants of development, and regional fixed effects.[8]

as regional fixed effects (column 7). A one percent increase in the average GDP per capita is associated with a 0.74 percent increase in light intensity per capita. Furthermore, columns 6 and 7 establish that the log absolute latitude is positively and significantly associated with log light intensity per-capita. Finally, column 8 establishes that the baseline association between log light intensity per-capita and per-capita GDP, reported in column 7, is robust to the inclusion of cells associated with gas flaring.

Importantly, as depicted in Figure 3, and as is indicative from the higher R^2 in columns 5–8 in comparison to column 1–4, the association between log light intensity per-capita and GDP per capita is tighter than the association between log light intensity per-area and GDP per capita.

4   Empirical Strategy

The research estimates the relationship between predicted genetic diversity and the standard of living as captured by satellite-measured light intensity based on a cross country analysis. In light of the trade-off associated with genetic diversity and process of development, the study estimates the quadratic effect of predicted genetic diversity on the global distribution of light intensity per capita and per square kilometer. In particular, the study estimates the regression model

$$Y_i = \beta_0 + \beta_1 G_i + \beta_2 G_i^2 + \beta_3 A_i + R_i \beta_4 + C_i \beta_5 + G_i \beta_6 + \varepsilon_i,$$

where Y_i is the dependent variable (i.e., log light intensity per capita, log light intensity per area or log GDP per capita), in country i over the period 1992–2012, G_i is the index of genetic diversity for country i; A_i is the log absolute latitude, R_i are regional dummy variables for country i; C_i is a vector of institutional and cultural controls for country i; G_i is a vector of additional geographical controls for country i; and, ε_i is a country-specific disturbance term.

Given the cross-country migrations in the post-1500 era, the distribution of genetic diversity across some part of the world may have been effected by the distribution of income across countries, and hence the relationship between predicted genetic diversity and the contemporary standard of living, as captured by night-light intensity, may be endogenous. Thus, to estimate the causal effect of genetic diversity on the standard of living, the regression is estimated over sub-samples of countries that that were less affected by the post-1500 migration, and in particular, a subset of the countries where vast majority of the contemporary population is indigenous.

5   Baseline Results

This section establishes the hump-shaped impact of genetic diversity on log light intensity per area as well as log light intensity per capita using the sample of 164 countries for which both variables are available. Table 2 presents the results of the relationship between genetic diversity and log light intensity per area and per-capita light intensity, accounting for the potential confounding effects of

Table 2: The Effect of Genetic Diversity on Lights-Based Measures of Development, 1992–2012

Dependent variable (in logs):	Light intensity per area, cells assoc. with gas flaring are			Light intensity per capita, cells assoc. with gas flaring are				GDP per capita
	Excluded		Included	Excluded		Included	Excluded	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted genetic diversity	1034.6***	547.0***	550.6***	801.6***	420.0***	423.2***	403.5***	352.8***
(ancestry adjusted)	(179.9)	(171.2)	(169.7)	(199.2)	(103.7)	(110.3)	(105.0)	(90.48)
Predicted genetic diversity	-735.7***	-387.7***	-390.4***	-569.3***	-301.8***	-304.3***	-289.3***	-251.8***
(ancestry adjusted) squared	(127.3)	(122.4)	(121.4)	(140.0)	(75.11)	(79.75)	(76.01)	(64.79)
Log absolute latitude		.0148	-.0396		.186**	.128	.207**	.0552
		(.163)	(.172)		(.0875)	(.0986)	(.0884)	(.114)
Regional fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Number of observations	164	164	164	164	164	164	158	158
Adjusted R^2	.124	.488	.484	.112	.695	.669	.703	.493
Joint signific. of genetic div. (p)	.000	.006	.005	.000	.000	.001	.001	.001
Maximizing genetic diversity	.703	.705	.705	.703	.695	.695	.697	.700
Lower limit of 90% CI	.697	.688	.688	.694	.682	.681	.682	.686
Upper limit of 90% CI	.708	.726	.725	.709	.711	.711	.716	.716
Significance of hump-shape (p)	.000	.003	.003	.000	.000	.000	.000	.000

This table establishes the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log light intensity per capita while controlling for log absolute latitude and regional fixed effects. * p <0.10, ** p <0.05, *** p <0.01

absolute latitude and regional fixed effects.[9] The unconditional hump-shaped relationship between genetic diversity and log light intensity per area is reported in column 1. In particular, the first-order element of the quadratic expression is positive and significant at the 1 percent level and the second-order element of the quadratic expression is negative and significant at the 1 percent level. Moreover, an additional test establishes a significant hump-shaped relationship ($p < 0.001$).[10] Furthermore, the estimated linear and quadratic coefficients, imply that a 1 percentage point increase in genetic diversity for the least diverse society in the regression sample would be associated with a light intensity per area that is higher by 111 percent, whereas a 1 percentage point decrease in genetic diversity for the most diverse society would be associated with a light intensity per area that is higher by 105 percent. Further, light intensity per area is predicted to be maximized at an expected heterozygosity value of about 0.703.

Reassuringly, column 2 reports that the statistically significantly hump-shaped relationship between genetic diversity and light emission per area is robust to the inclusion of controls for log absolute latitude as well as regional fixed effects (based on the regional classification by the World Bank). The first and second order elements, as well as the additional test for the presence of a hump-shaped relation, remain highly significant. Furthermore, column 3 establishes that the statistical finding is robust to the inclusion of pixels associated with gas flaring.

Column 4–7 reports a similar hump-shaped relationship between genetic diversity and log light intensity per capita. In particular, column 4 establishes that the estimates of the coefficients on genetic diversity and its squared value have the expected signs and are both highly significant. Fur-

---

[9]Here and elsewhere, heteroskedasticity-robust standard errors are used, unless otherwise stated.
[10]See Lind and Mehlum (2010).

thermore, a test of the significance of a hump-shape is also highly significant ($p < 0.001$). Column 5 establishes that the statistically significantly hump-shaped relationship of genetic diversity on light emission per capita is robust to accounting for the log absolute latitude as well as regional fixed effects. The first and second order elements of the quadratic expression, as well as the additional test for the presence of a hump-shaped relation, remain highly significant. Furthermore, column 6 establishes that the statistical finding is robust to the inclusion of pixels associated with gas flaring.

Based on a sample of 158 countries where GDP per capita is available, column 7 establish that the estimated hump-shaped relationship remains highly significant. Furthermore, there exist a similar hump-shaped association between genetic diversity and GDP per capita is observed in column 8. Interestingly, inferred from the adjusted R^2, the model explain less of the variation in GDP per capita than of the variation in light intensity per capita.

6   Robustness to Controlling for Additional Factors

Using a sample of 148 countries for which additional relevant control variables are available, columns 1–4 of Table 3 reveal a significant hump-shaped relationship between genetic diversity and light intensity per area and per capita, accounting for the potential confounding effects of geographical factors (i.e., logs of the absolute latitude, percentage of arable land, the mean distance to the nearest waterway, and island status), the timing of the Neolithic transition, percentage of population living in tropical zones, ethnic fractionalization, as well as regional fixed effects.

Table 3, column 1 replicates Table 2, column 6 for the smaller sample and shows that the hump-shaped effect of genetic diversity on log light intensity per capita remains intact. In particular, the estimated linear and quadratic coefficients on genetic diversity are both statistically significant at the 1 percent level and the additional test of the presence of a hump-shaped relation is also highly significant ($p = 0.002$). In light of possibly direct effects of cultivated land, water and island status of a country on development and thus light emission, column 2 reports the results from a regression in the presence of these control variables. The estimated quadratic expression, as well as the additional test for a hump-shaped relationship, remain highly significant. Furthermore, given the potential influence of the time elapsed since the Neolithic revolution and the percentage of the population living in tropical zones (Ashraf and Galor, 2011; Gallup et al., 1999), column 3 reports the results accounting for the potential confounding effects for these variables. The pronounced estimated quadratic expressions as well as the additional test for a hump-shaped relationship remain highly significant. Furthermore, column 4 establish that the estimated hump-shape remain significant in the presence of control variables accounting for ethnic fractionalization and column 5 establish the robustness towards accounting for percentage of population at risk of contracting malaria. Column 6, based on regressions using a smaller sample of 111 countries for which an additional set of control variables are available, reveal a similar statistically significant hump-shaped relation, conditional on social infrastructure as a proxy for institutions (Hall and Jones, 1999), OPEC membership, legal

Table 3: The Effect of Genetic Diversity on Light Intensity per Capita and GDP per Capita, 1992–2012

Dependent variable (in logs):	Light intensity per capita							GDP per capita
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted genetic diversity (ancestry adjusted)	364.6***	326.9***	421.9***	415.6***	393.8***	425.5***	367.5***	236.9***
	(109.8)	(115.5)	(119.0)	(122.9)	(113.9)	(109.2)	(115.8)	(78.40)
Predicted genetic diversity (ancestry adjusted) squared	-262.1***	-233.4***	-301.0***	-280.3***	-293.7***	-300.9***	-257.5***	-164.6***
	(79.74)	(83.78)	(86.30)	(89.02)	(82.26)	(79.40)	(84.16)	(56.92)
Log absolute latitude	.166*	.196**	-.0333	-.0391	-.0472	.0402	.0188	.0735
	(.0941)	(.0851)	(.0998)	(.103)	(.107)	(.0674)	(.0617)	(.0751)
Log percentage of arable land		-.307***	-.241***	-.254***	-.286***	-.227***	-.250***	-.235***
		(.0706)	(.0731)	(.0748)	(.0720)	(.0678)	(.0649)	(.0629)
Mean distance to nearest waterway		-.386**	-.449**	-.427**	-.432**	-.262	-.538*	-.494**
		(.163)	(.174)	(.173)	(.166)	(.333)	(.298)	(.240)
Log Neolithic transition timing (ancestry adjusted)			-.175	-.129	-.0211	.594**	.597**	.580**
			(.329)	(.315)	(.277)	(.272)	(.267)	(.239)
Percentage of population living in tropical zones			-1.012***	-.961***	-.414	-.663**	-.827***	-.311*
			(.324)	(.335)	(.311)	(.265)	(.238)	(.164)
Ethnic fractionalization				-.331	-.0878	.325	.437	.157
				(.373)	(.318)	(.334)	(.333)	(.280)
Percentage of population at risk of contracting malaria					-1.744***	-1.176***	-1.039***	-.806**
					(.379)	(.389)	(.381)	(.349)
Social infrastructure						1.567***	1.596***	1.390**
						(.582)	(.570)	(.528)
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Island fixed effect	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
OPEC fixed effect	No	No	No	No	No	Yes	Yes	Yes
Legal origin fixed effects	No	No	No	No	No	Yes	Yes	Yes
Major religion shares	No	No	No	No	No	Yes	Yes	Yes
Number of observations	148	148	148	148	148	111	108	108
Adjusted R^2	.693	.726	.745	.745	.789	.879	.887	.870
Joint signific. of genetic div. (p)	.004	.012	.001	.002	.001	.000	.000	.000
Maximizing genetic diversity	.695	.700	.700	.700	.702	.707	.713	.719
Lower limit of 90% CI	.680	.684	.687	.687	.700	.696	.700	.702
Upper limit of 90% CI	.716	.728	.720	.720	.718	.726	.747	.767
Significance of hump-shape (p)	.002	.011	.002	.003	.002	.002	.019	.040

This table establishes the robustness of the significant hump-shaped effect of genetic diversity, as predicted by migratory distance from East Africa, on log light intensity per capita while controlling for log absolute latitude and regional fixed effects (column 1); island status, log percentage of arable land, and mean distance to nearest waterway (column 2); log ancestry-adjusted duration since the Neolithic revolution and percentage of population living in tropical zones (column 3); ethnic fractionalization (column 4); percentage of population at risk of contracting malaria (column 5); and social infrastructure, OPEC membership, legal origin (either British, French, German, Scandinavian, or Socialist), and the percentage of Muslims, Catholics and Protestants (column 6 and 7); and to the use of GDP per capita as the dependent variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

origin (either British, French, German, Scandinavian, or Socialist), and the percentage of Muslims, Catholics and Protestants.[11]

The coefficient estimates in column 6 imply that increasing the diversity of the most genetically homogenous country in the sample (Bolivia) by 1 percentage point would be associated with a rise in light intensity per capita by 48 percent, whereas decreasing the diversity of the most genetically diverse country in the sample (Ethiopia) by 1 percentage point would be associated with a raise its log per-capita light intensity by 41 percent. Moreover, under these specifications of column 6 and 7, the sum of partial R^2 of genetic diversity and genetic diversity squared with light density is nearly 30 percent and with income per capita is 19 percent.

Finally, turning to a sample of 108 countries for which GDP per capita is reported, in addition to the full set of control variables, column 7 establish that the hump-shaped relationship between light intensity per capita and predicted genetic diversity remain highly significant. Furthermore, in the presence of all the control variable, column 8 establish that the hump-shaped relationship between living standards and predicted genetic diversity is robust to the use of GDP per capita as the independent variable.

7   Robustness to Various Samples

Reassuringly, the highly significant hump-shaped effect of genetic diversity on per-capita light emission is not an artifact of postcolonial migration toward 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.

In particular, as established in Table 4, 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., destinations that were less economically attractive to migrants) in column 2 and 7; (ii) non-Neo-European countries (i.e., excluding the United States, Canada, Australia, and New Zealand) in column 3 and 8; (iii) non-Latin American countries in column 4 and 9; and (iv) countries for which the indigenous population is larger than 95 percent

---

[11]The highly significant quadratic terms are robust to estimation by a spatial-autoregressive model with spatial-autoregressive disturbances (using spreg version 1.0.2 [reference] in Stata version 13.1) using the generalized spatial two-stage least squares procedure with a spatial-autoregressive term and a spatial-error term and a spatial-weighting matrix generated from the longitude and latitude of each country in the regression. Furthermore, the highly significant quadratic terms are robust to accounting for the generated nature of the predicted ancestry-adjusted genetic diversity by use of a two-step bootstrapping algorithm to compute the standard errors. 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.

Table 4: The Effect of Genetic Diversity on Light Intensity per Capita, 1992–2012, Addressing Endogenous Post-1500 Migrations

Dependent variable (in logs):	Light intensity per capita									
Countries excluded:	None	OECD	Neo-European	Latin American	Fraction of population > 5% non-indigenous	None	OECD	Neo-European	Latin American	Fraction of population > 5% non-indigenous
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted genetic diversity (ancestry adjusted)	801.6***	534.6***	752.8***	1831.7***	1728.9***	425.5***	457.7***	420.4***	664.3***	932.3***
	(199.2)	(170.9)	(196.9)	(331.5)	(357.2)	(109.2)	(114.3)	(110.2)	(232.6)	(310.1)
Predicted genetic diversity (ancestry adjusted) squared	-569.3***	-381.9***	-534.8***	-1275.4***	-1210.4***	-300.9***	-323.6***	-297.5***	-462.7***	-657.0***
	(140.0)	(120.6)	(138.5)	(228.5)	(246.7)	(79.40)	(82.41)	(80.22)	(161.8)	(212.1)
Log absolute latitude						.0402	.0565	.0319	.00324	-.505*
						(.0674)	(.0775)	(.0706)	(.0807)	(.263)
Log percentage of arable land						-.227***	-.184**	-.221***	-.249***	-.252**
						(.0678)	(.0716)	(.0677)	(.0742)	(.116)
Mean distance to nearest waterway						-.262	-.287	-.317	-.437	-.709
						(.333)	(.359)	(.361)	(.343)	(.460)
Log Neolithic transition timing (ancestry adjusted)						.594**	.711**	.586**	.658**	.117
						(.272)	(.301)	(.268)	(.303)	(.437)
Ethnic fractionalization						.325	.410	.296	-.0821	-.177
						(.334)	(.413)	(.348)	(.389)	(.489)
Percentage of population at risk of contracting malaria						-1.176***	-.872**	-1.140***	-.755	.181
						(.389)	(.382)	(.398)	(.508)	(.804)
Social infrastructure						1.567***	2.083***	1.555**	1.538*	1.980
						(.582)	(.757)	(.617)	(.904)	(1.229)
Regional fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Island fixed effect	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
OPEC fixed effect	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Legal origin fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Major religion shares	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	164	135	160	139	60	111	85	107	89	44
Adjusted R²	.112	.070	.100	.211	.262	.879	.817	.871	.883	.930
Joint signific. of genetic div. (p)	.000	.002	.000	.000	.000	.000	.000	.000	.021	.006
Maximizing genetic diversity	.703	.700	.703	.718	.714	.707	.707	.706	.717	.709
Lower limit of 90% CI	.694	.685	.694	.712	.706	.696	.696	.695	.700	.672
Upper limit of 90% CI	.709	.708	.710	.722	.719	.726	.724	.727	.734	.727
Significance of hump-shape (p)	.000	.003	.000	.000	.000	.002	.001	.003	.004	.014

This table establishes the robustness of the hump-shaped effect of genetic diversity in the presence or the absence of all the control variables towards restricting the sample to non-OECD economies (column 2 and 7); non-Neo-European countries, i.e., excluding the United States, Canada, Australia, and New Zealand, (column 3 and 8); non-Latin American countries (column 4 and 9); and countries for which the indigenous population is larger than 95 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 (column 6 and 10). * p <0.10, ** p <0.05, *** p <0.01

13

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 5 and 10.[12]

8   Conclusion

This research establishes that migratory distance from the cradle of anatomically modern humans in East Africa and its effect on the distribution of genetic diversity across countries has a hump-shaped effect on night-light intensity per capita as observed by satellites, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. The finding lends further credence to the hypothesis that a significant portion of the variation in the standard of living across the globe can be attributed to factors that were determined in the distant past

---

[12]The quadratic terms remain significant in a regression on the sample of 72 countries that excludes the sub-Saharan Africa in the presence of the control variables relevant for column 7–10.

References

ALESINA, A., A. DEVLEESCHAUWER, W. EASTERLY, S. KURLAT, AND R. WACZIARG (2003): "Fractionalization," *Journal of Economic Growth*, 8, 155–194.

ASHRAF, Q. AND O. GALOR (2011): "Dynamics and Stagnation in the Malthusian Epoch," *American Economic Review*, 101, 2003–2041.

ASHRAF, Q. AND O. GALOR (2013): "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103, 1–46.

CANN, H. M., C. DE TOMA, L. CAZES, M.-F. LEGRAND, V. MOREL, L. PIOUFFRE, ET AL. (2002): "A Human Genome Diversity Cell Line Panel," *Science*, 296, 261–262.

GALLUP, J. L. AND J. D. SACHS (2001): "The Economic Burden of Malaria," *The American Journal of Tropical Medicine and Hygiene*, 64, 85–96.

GALLUP, J. L., J. D. SACHS, AND A. D. MELLINGER (1999): "Geography and Economic Development," *International Regional Science Review*, 22, 179–232.

HALL, R. E. AND C. I. JONES (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114, 83–116.

HENDERSON, J. V., A. STOREYGARD, AND D. N. WEIL (2012): "Measuring economic growth from outer space," *American economic review*, 102, 994–1028.

LA PORTA, R., F. LOPEZ-DE-SILANES, A. SHLEIFER, AND R. W. VISHNY (1999): "The Quality of Government," *Journal of Law, Economics, and Organization*, 15, 222–279.

LIND, J. T. AND H. MEHLUM (2010): "With or Without U? The Appropriate Test for a U-Shaped Relationship," *Oxford Bulletin of Economics and Statistics*, 72, 109–118.

PRUGNOLLE, F., A. MANICA, AND F. BALLOUX (2005): "Geography Predicts Neutral Genetic Diversity of Human Populations," *Current Biology*, 15, R159–R160.

PUTTERMAN, L. AND D. N. WEIL (2010): "Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125, 1627–1682.

RAMACHANDRAN, S., O. DESHPANDE, C. C. ROSEMAN, N. A. ROSENBERG, M. W. FELDMAN, AND L. L. 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, 15942–15947.
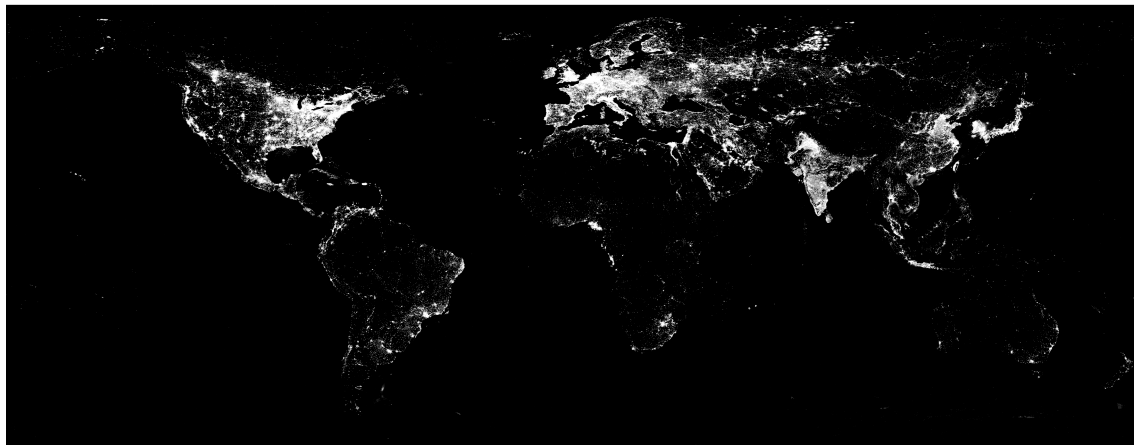
A Global Distribution of Night-Light Intensity

Figure A.1: The distribution of average nighttime light emission within country borders, 1992–2012. Data source: NOAA-NGDC.

B Variable definitions

Light intensity per capita. The data is collected by the U.S. Air Force Defense Meteorological Satellite Program's Operational Linescan System, recorded by low-altitude (830 km) satellites orbiting Earth every 101 minutes, generating a global night time coverage every 24 hours.[13] NOAA-NGDC produces global geographic raster datasets on a per-year and per-satellite level. In this process, the NOAA-NGDC data excludes on an orbit-by-orbit manner, using visual inspection, sunlit data and glare (e.g., from snow coverage) based on the solar elevation angle, moonlit data based on a calculation of lunar illuminance, clouded data based on thermal band data and surface temperature grids, and Northern Lights (aurora borealis). Furthermore, ephemeral events, such as fires have been discarded from the source data.

The lights data is obtained from National Oceanic and Atmospheric Administration's National Geophysical Data Center and processed as described in Section 2.1. The population size measures entering the denominator of the per-capita measure is from the World Bank's World Development Indicators, found at `http://data.worldbank.org/indicator/SP.POP.TOTL`.

Gross Domestic Product (GDP) per capita. The GDP per capita data is purchasing power parity-adjusted GDP per capita in constant 2005 international dollars from the World Bank's World Development Indicators, found at `http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD`.

---

[13]The data are collected and processed by the National Oceanic and Atmospheric Administration's National Geophysical Data Center (NOAA-NGDC). The main purpose of the system is to monitor cloud coverage and navigation conditions. The data can be accessed at `http://ngdc.noaa.gov/eog/download.html`.

Predicted genetic diversity (ancestry adjusted). The expected heterozygosity (genetic diversity) of a country's population as given by (Ashraf and Galor, 2013). The variable is 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 (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,[14] (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 Ashraf and Galor (2013). 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.

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

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.

Island. A country's status of island (a binary variable) is found from the CIA's World Factbook.

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

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

---

[14]For a more detailed description of the HGDP-CEPH Human Genome Diversity Cell Line Panel dataset, the interested reader is referred to Cann et al. (2002).

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.

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 percent of trade, (ii) average tariff rates are less than 40 percent, (iii) any black market premium was less than 20 percent during the 1970s and 80s, (iv) the country is not socialist, and (v) the government does not monopolize over major exports.

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).

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 dataset on malaria.

Organization of the Petroleum Exporting Countries (OPEC) membership. A binary variable indicating membership of OPEC as given by Ashraf and Galor (2013).

Regional dummies. Binary variables indicating each of seven world regions as defined by the World Bank and given by Ashraf and Galor (2013).