Heterogeneity and Productivity*

Quamrul Ashraf Oded Galor Marc Klemp

May 12, 2015

Abstract

This research explores the effects of within-group heterogeneity on group-level productivity within a novel geo-referenced dataset of observed genetic diversity across the globe. It establishes that observed genetic diversity of 230 worldwide ethnic groups, as well as predicted genetic diversity of 1,331 ethnic groups, has a hump-shaped effect on economic prosperity, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. Moreover, the study demonstrates that variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country.

Keywords Heterogeneity, Regional Development, Out-of-Africa Hypothesis, Comparative Development, Genetic Diversity, Nighttime Light Intensity

JEL Classification Codes L25, M14, O10, O40, Z10

*The research of Ashraf and Galor is supported by NSF grant SES-1338426. The research of Klemp is funded by the Carlsberg Foundation and by the Danish Research Council reference no. 1329-00093 and reference no. 1327-00245. Quamrul Ashraf (Quamrul.H.Ashraf@williams.edu): Department of Economics, Williams College, 24 Hopkins Hall Drive, Williamstown, MA 01267, USA. Oded Galor (oded_galor@brown.edu) Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA. Marc Klemp (marc_klemp@brown.edu): Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA, and Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, DK-1353 Copenhagen K, Denmark.

1 Introduction

Conventional wisdom suggests that within-group heterogeneity is a major determinant of the productivity of group members. While diversity diminishes trust, cooperation, and coordination, adversely affecting group-level productivity, complementarities across diverse productive traits stimulate innovations and group-level performance. Thus, in an environment characterized by diminishing marginal returns to diversity and homogeneity, aggregate productivity of groups characterized by an intermediate level of diversity will be higher than that generated by excessively homogenous or heterogeneous groups.

This research explores the effects of within-group heterogeneity on group-level productivity. Exploiting an exogenous source of variation in genetic diversity among ethnic groups across the globe, the study establishes that genetic diversity has a hump-shaped effect on economic prosperity, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. Moreover, the study establishes that variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country.

The analysis is performed using a newly constructed geo-referenced dataset on within-ethnic-group genetic diversity for a large sample of 230 ethnic groups across the globe, greatly surpassing the Human Genome Diversity Project sample of 53 ethnic groups. In particular, the study constructs a novel dataset mapping the observed genetic diversity for 230 ethnic groups across the globe (Pemberton et al., 2013) to the geographical attributes that characterizes their historical homelands as well as the ethnographic characteristics of 145 of these ethnic groups. Moreover, the analysis is further performed on an extended sample of 1,331 ethnic groups for which the study constructs a measure of projected genetic diversity, based on migratory distance of these ethnic groups from East Africa, linking it to the geographical attributes of the historical homelands and ethnographic characteristics of these ethnic groups.

The first part of the empirical analysis establishes that observed genetic diversity within 230 ethnic groups has a hump-shaped relationship with economic prosperity, as captured by luminosity. Using expected heterozygosity as an index of genetic diversity that captures the probability that two individuals selected at random from a given ethnic group differ from one another in a given locus of their genome, that analysis confirms the hypothesis that an intermediate level of diversity is associated with greater group productivity.[1] In particular that analysis suggests that the degree

[1]The measure of expected heterozygosity for prehistorically indigenous ethnic groups is constructed by population geneticists using data on allelic frequencies for a particular class of DNA loci called microsattelites, residing in nonprotein-coding or neutral regions of the human genome (i.e., regions that do not directly result in phenotypic expression). This measure therefore possesses the advantage of not being tainted by the differential forces of natural selection that may have operated on these populations since their prehistoric exodus from Africa. Critically, however, as argued and empirically established by Ashraf and Galor (2013a,b), the observed socioeconomic influence of expected heterozygosity in microsattelites can indeed reflect the latent impact of heterogeneity in phenotypically and cognitively expressed genomic material, in light of mounting evidence from the fields of physical and cognitive anthropology on the existence of a serial founder effect on the observed worldwide patterns in various forms of intragroup phenotypic and cognitive diversity, including phonemic diversity(Atkinson, 2011) as well as interpersonal diversity in skeletal features pertaining to cranial characteristics (Manica et al., 2007; von Cramon-Taubadel and Lycett, 2008; Betti et al., 2009), dental attributes (Hanihara, 2008), and pelvic traits (Betti et al., 2013).

of genetic diversity that is associated with the maximal level of productivity is 0.67, while observed diversity among the ethnic groups in the sample ranges from 0.56 to 0.77.

Furthermore, the analysis establishes that the association between diversity and group productivity is sizable economically. In particular, it implies that productivity of the most genetically homogenous ethnicity in the sample (i.e., the Karitiana in Brazil with a genetic diversity of 0.56) is 23.6 percent lower than the maximal level of productivity (associated with a genetic diversity of 0.67). Similarly, the productivity of the most heterogonous ethnicity in the sample (i.e., the Turu in Tanzania with a genetic diversity of 0.77) is 22.0 percent lower than the maximal level of productivity (associated with a genetic diversity of 0.67).

The estimated effect of genetic diversity on economic prosperity may be subjected to concerns about endogeneity. In particular, the potential effect of economic prosperity on group formation, migration, and conflict may contribute to the observed degree of diversity. Moreover, omitted geographical, institutional, and human capital characteristics may co-determine diversity as well as prosperity. In light of these potential endogeneity concerns, the study exploits an identification strategy that is well rooted in two major hypotheses in the field of population genetics: the Serial Founder Effect and the Out-of-Africa hypothesis. Accordingly, genetic diversity declines with migratory distance from East Africa (Ramachandran et al., 2005). Thus, the analysis exploits migratory distance from East Africa as an exogenous source of variation in observed genetic diversity within ethnic groups and establishes the hump-shaped effect of genetic diversity on economic prosperity. In particular the IV estimation suggests that the degree of genetic diversity that is associated with the maximal level of productivity is 0.66.

The analysis further establishes that if one exploits variations in within-ethnic-group genetic diversity across ethnic groups within a country, rather than in the world as a whole, an intermediate level of diversity is still associated with greater group productivity. In particular, that analysis suggests that, accounting for country fixed effects, the degree of genetic diversity that is associated with the maximal level of productivity is 0.68. Hence, the study establishes that variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country.

Finally, exploiting an extended sample of 1,331 ethnic groups for which the study constructs measures of projected genetic diversity, based on their migratory distance from East Africa, the research confirms the robustness of the hump-shaped effect of group-level heterogeneity on productivity. In particular that estimation suggests that the degree of genetic diversity that is associated with the maximal level of productivity is 0.64.

The research makes four distinct contributions. First, the study constructs a novel geo-referenced dataset mapping observed genetic diversity for ethnic groups across the globe to the geographical attributes that characterizes their historical homelands as well as the ethnographic characteristics of these ethnic groups. Second, the study explores the effects of within-group heterogeneity on group-level productivity, establishing the significance of moderately diverse group for productivity. Thus, the study provides the first ethnic level analysis that captures the trade-off between the beneficial and the detrimental effects of diversity on productivity. Third, the study establishes that variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country, providing a novel angle for the understanding of the origins of regional inequality within nations.

Finally, the research explores the effect of migratory distance from East Africa and genetic diversity at the sub-national level on comparative economic development. In contrast to the cross country analysis of Ashraf and Galor (2013a) that could have been potentially affected by omitted country specific characteristics, the ethnic level analysis establishes that the hump-shaped effect of genetic diversity on economic development is present at the sub-national level as well, implying that the beneficial effects of intermediate level of diversity are present across groups that significantly smaller than countries and robust to the inclusion of country fixed-effects. The findings enhances our understanding 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.

2 Data

The proposed hypothesis, that genetic diversity within ethnic group has hump-shaped effect on economic development, as measured by nighttime light intensity, is examined empirically based on a novel geo-referenced dataset that maps observed and predicted genetic diversity for ethnic groups across the globe to the geographical attributes that characterizes their historical homelands as well as the ethnographic characteristics of these ethnic groups as provided by the Ethnographic Atlas.

2.1 Dependent Variable: Average Luminosity Per Capita

In the absence of a comprehensive data on the income per capita of ethnic groups across the globe, the study uses the mean light intensity per capita over the area of the ethnic group averaged over the period 1992–2013, as a proxy for the standard of living for each ethnic group. The validity of this increasingly used proxy for the standard of living reflects the strong and highly significant positive correlation between luminosity and GDP per capita (Chen and Nordhaus, 2011; Henderson et al., 2012; Ashraf et al., 2014).[2]

Satellite-captured images of global night-light emission are available for each year in the period 1992–2013, for 30 arc second grids, spanning −180 to 180 degrees longitude and −65 to 75 degrees latitude. The average yearly luminosity for each cell over this 22-year period is depicted in Figure A.1 in the appendix. Each cell of approximately one square km (as measured at equator), is

[2] Research has shown that contemporary and present homelands are correlated. In particular, based on individual-level evidence from Africa, Nunn and Wantchekon (2011) found a correlation of 0.75 between the location of individuals in 2005 and their historical homeland, as given by Murdock's ethnographic atlas from 1959. To the extent that the contemporary homelands of the ethnic groups in the data is the same as their historical homelands described by the polygons in our data, the analysis capture the contemporary association between genetic diversity and luminosity. Furthermore, the measurement error introduced by the fact that some ethnic groups were affected by migration over the period of analysis will likely bias our analysis against finding a hump-shaped association.

assigned an integer ranging from 0 to 63 representing its yearly luminosity. The mean luminosity per capita in a given year for an ethnic group is therefore the mean luminosity over all cells within the boundaries of the ethnicity area divided by the population residing in this area.[3]

The light data 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 light emission at the ethnicity-area level which typically consists of a continuum of a large number of pixels and since it reflects the average night-light emission based on a 22-year period. These potential measurement errors are further mitigated by the inclusion of a wide range of confounding geographic characteristics, such as absolute latitude, ruggedness, and regional fixed effects.

The population that resides within the area of each ethnic group is derived from the GPWv3 gridded population counts in the year 2000 (adjusted to match UN totals).[4] The GPWv3 data is based on a large number of count tabulations from worldwide populations matched to geographical boundaries (census or administrative units).[5] In particular, the population counts permit the omission of those who reside in sub-areas characterized by gas flaring.

Lights emitted from gas flaring used in petroleum refineries, chemical plants, natural gas processing plants, and oil and gas wells distorts the accuracy of the proxy and is therefore omitted by the exclusion of sub-areas characterized by gas flaring.[6] Furthermore, since urban areas are often ethnically diverse, the analysis examines the robustness of the results when the emission of light in urban areas is excluded (using the World Urban Areas map, provided by Environmental Systems Research Institute and based on the DeLorme World Base Map).

2.2 Independent Variables: Genetic Diversity and Predicted Genetic Diversity

The analysis is performed using a newly constructed dataset on within-ethnic-group genetic diversity for a large sample of 230 ethnic groups across the globe, greatly surpassing the Human Genome Diversity Project sample of 53 ethnic groups. In particular, the study constructs a novel geo-referenced dataset mapping the observed genetic diversity for 230 ethnic groups across the

[3]The data comes from two separate satellites in 1994 and in the period 1997–2007, resulting in 34 yearly data points for each cell, which are averaged to obtain the average yearly light intensity over this 22-year period. Given the fact that there are years for which luminosity observed by more than one satellite is available, the data is weighted such that data for each year has equal weights.

[4]Gridded Population of the World, Version 3 (GPWv3), dataset (Center for International Earth Science Information Network - CIESIN - Columbia University, United Nations Food and Agriculture Programme - FAO, and Centro Internacional de Agricultura Tropical - CIAT, 2005).

[5]Unlike previous versions of the dataset, the GPWv3 accounts for the island-levels population counts for countries that are comprised of island chains which are present in our dataset.

[6]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.

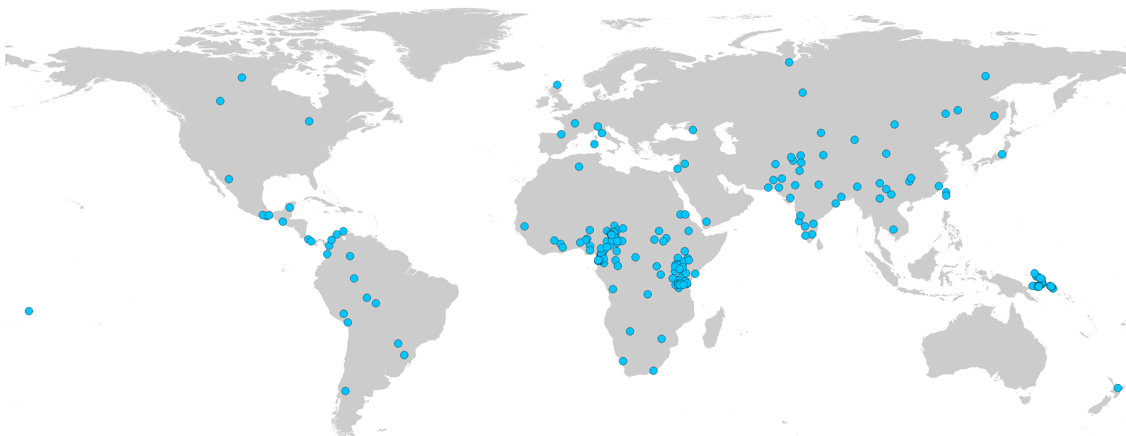

Figure 1: This figure depicts the interior centroids of the historical homelands of ethnic groups in the data with observed genetic diversity.

globe, unified and standardized by Pemberton et al. (2013), to the geographical attributes that characterizes their historical homelands as well as the ethnographic characteristics of 145 of these ethnic groups as provided by the Ethnographic Atlas. Moreover, the analysis is further performed on an extended sample of 1,331 ethnic groups for which the study constructs a measure of projected genetic diversity, based on their migratory distance from East Africa, and maps it to the geographical attributes of their historical homelands as well as the ethnographic characteristics of 1,242 of these groups. The geographical distribution of the ethnic groups homelands is depicted in Figure 1.

The study exploits one of the largest existing dataset of human population-genetic variation, presented in Pemberton et al. (2013) and containing estimates of expected heterozygosity for 239 ethnic groups across the globed. This dataset combines eight human genetic diversity datasets at their 645 shared loci, including the HGDP-CEPH Human Genome Diversity Cell Line Panel (used by Ashraf and Galor, 2013b, in their country-level analysis). Specifically, genetic diversity is captured by expected heterozygosity. It is constructed 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, a gene-specific heterozygosity statistic (i.e., the probability that two randomly selected individuals differ with respect to a given gene) is calculated, which when averaged over multiple genes or DNA loci yields the overall expected heterozygosity for the relevant population.

In light of the insights from two major hypotheses in the field of population genetics: the Serial Founder Effect and the Out-of-Africa hypothesis, according to which, genetic diversity declines with migratory distance from East Africa (Ramachandran et al., 2005; Pemberton et al., 2013), a measure of projected genetic diversity for an extended sample of 1,331 ethnic groups is constructed based on their migratory distance from East Africa. This measure is used for the examination of the hypothesis in the extended sample of ethnic group as well as an instrumental variable that is designed to resolve concerns about endogeneity in the observed genetic diversity sample of 230 ethnic groups.

In particular, migratory distance alone explains more than 84 percent of the cross-group variation in within-group diversity. In addition, the estimated OLS coefficient is highly statistically significant ($p < 6.71 \times 10^{-93}$). It suggests that expected heterozygosity falls by 6.7 percentage points for every 10,000 km increase in migratory distance from East Africa (from 0.767 in Addis Abba to 0.561 in South America). Hence, the Predicted genetic diversity for ethnicity with a migratory distance D_i (measured in 10,000 km) from East Africa is $0.767 - 0.067D_i$.[7]

2.3 Independent Variables: Control Variables

The analysis also accounts for the potentially confounding effects of a large vector of geographical, climatic factors.[8] In particular, in light of the negative association between ecological biodiversity and distance from the equator as well as the potential effect of geographic and climatic characteristics on the degree of diversity, the analysis accounts for the potentially confounding effects of absolute latitude, climatic suitability for agriculture (Ramankutty et al., 2002; Michalopoulos, 2012), ruggedness (Nunn and Puga, 2012), average temperatures, diurnal temperature range, precipitation, wet day frequency, and frost day frequency over the period 1901–2012 (CRU TS3.21 database; Harris et al., 2014), and landmass type, i.e., "primary land", "very small island", "small

[7]In estimating the migratory distance from Addis Ababa (East Africa) for each of the ethnic groups in the data, the shortest traversable paths from Addis Ababa to the interior centroid of each ethnic group computed. Given the limited ability of human to travel across large bodies of water, the traversable area included bodies of water at a distance of 100km from land mass (excluding migration from Africa into Europe via Italy or Spain). Furthermore, for ethnicities that reside in a distance that exceed 100km from the traversable area connected to Addis Ababa, the distance was computed in the following way. A point file was created by clipping the extended traversable area to world boundaries and aggregating it to a resolution of $2,096,707$ pixels which was then converted into points. For each ethnicity centroid, the nearest four distance points were identified and the great circle distance from the ethnicity centroid to those points were calculated. These distances was then added to the migratory distance from Addis Ababa at the distance point to obtain the total migratory distance from the ethnicity centroid from Addis Ababa to each of these four points. The point with the shortest total migratory distance from Addis Ababa was selected to represent the total migratory distance for the ethnicity.

[8]The geographical characteristics of areas characterized by gas flaring or urban settlement are included in the calculation of the control variables in the main analysis. However, as established in the appendix the results are robust to the exclusion of gas flaring areas as well as urban areas.

island", "medium island", and "large island" (provided by Environmental Systems Research Institute).[9]

Furthermore, in light of the potential effect diversity on the emergence of institutions (Galor and Klemp, 2015) as well as the potential effect of institutions on the standard of living, the analysis accounts for the institutional characteristics of each ethnic group as reported in the Ethnographic Atlas, including the historical level of jurisdictional hierarchy beyond the local community level, and the historical type of class stratification. Finally, since the scale of each ethnic group may affect its genetic diversity and economic performances, the analysis account for the potentially confounding effect of the historical mean size of local communities.

3 Empirical Analysis

3.1 Identification Strategy

The first part of the empirical analysis examines the hypothesis that observed genetic diversity within 230 ethnic groups has a hump-shaped relationship with economic prosperity, as captured by luminosity. The analysis confirms the hypothesis that an intermediate level of diversity is associated with greater group productivity.

Nevertheless, the estimated effect of genetic diversity on economic prosperity may be subjected to concerns about endogeneity. In particular, the potential effect of economic prosperity on group formation, migration, and conflict may contribute to the observed degree of diversity. Moreover, omitted geographical, institutional, and human capital characteristics may co-determine diversity as well as prosperity. In light of these potential endogeneity concerns, the study exploits an identification strategy is well rooted in two major hypotheses in the field of population genetics: the Serial Founder Effect and the Out-of-Africa hypothesis. Accordingly, genetic diversity declines with migratory distance from East Africa (Figure 2). Thus, the analysis exploits migratory distance from East Africa as an exogenous source of variation in observed genetic diversity within ethnic groups and establishes the hump-shaped effect of genetic diversity on economic prosperity.

The analysis further establishes that if one exploits variations in within-ethnic-group genetic diversity across ethnic groups within a country, rather than in the world as a whole, an intermediate level of diversity is still associated with greater group productivity. Hence, the study establishes that variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country. Finally, exploiting an extended sample of 1,331 ethnic groups for which the study constructs measures of projected genetic diversity, based on their migratory distance from East Africa, the research confirms the hump-shaped effect of group-level heterogeneity on productivity.

[9] Since soil quality is not reported for all areas on the globe, the analysis accounts for soil quality by controlling for ten dummy variables, indicating each decile of the variable, imputing the average level of soil quality for the ethnic groups with unknown soil quality.

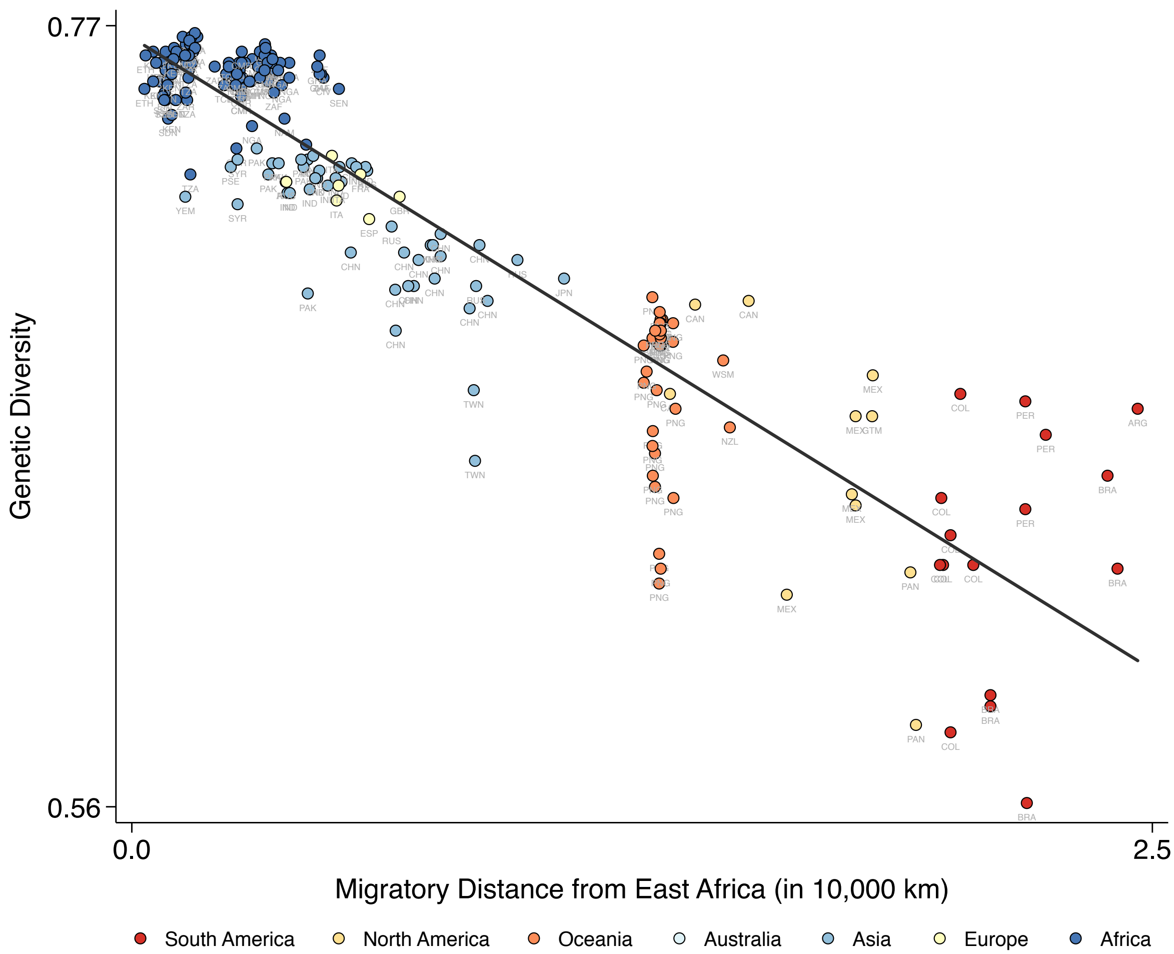

Figure 2: This figure depicts the negative impact of migratory distance from East Africa on genetic diversity across the 230 ethnic groups in the sample. Each ethnic group is represented by a point and a country code (following the World Bank) of the country in which the ethnic group resides.

3.2 Empirical Model

The research estimates the relationship between actual genetic diversity, as well as predicted genetic diversity, and the standard of living as captured by satellite-measured light intensity in ethnicity areas across the world as a whole, within regions of the world, and within countries of the world. In light of the trade-off associated with genetic diversity and the process of development, the study estimates the quadratic relationship between genetic diversity and the mean light intensity per capita.

First, the study explores the effect of observed genetic diversity on luminosity, estimating the regression model

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

where dependent variable, Y_i, is the log of mean light intensity per capita over the area of ethnic group i, averaged over the period 1992–2013.[10] The independent variables, G_i is the observed genetic diversity for ethnicity i; A_i is the log absolute latitude of ethnicity i's area, R_i are regional dummy variables for ethnicity i's area; L_i is a vector of land suitability landmass type dummy variables for ethnicity i's area; T_i is a vector of additional geographical controls for ethnicity i's area; C_i is a vector of institutional controls for ethnicity i; and, ε_i is an ethnicity-specific disturbance term. Furthermore, to account for potential reverse causality or omitted variables, migratory distance is used as an instrument for observed genetic diversity in a 2SLS regression analysis.[11]

Moreover, considering the remarkably strong predictive power of migratory distance from East Africa for genetic diversity, the following regression specification employed to test the effect of predicted genetic diversity in an extended sample of ethnicity.

$$Y_i = \beta_0 + \beta_1 \hat{G}_i + \beta_2 \hat{G}_i^2 + \beta_3 A_i + R_i\beta_4 + C_i\beta_5 + T_i\beta_6 + \varepsilon_i,$$

where $\hat{G}_i$ is predicted genetic diversity for ethnicity i.

In addition, the research investigates the association between genetic diversity of ethnic groups and the standard of living within countries. In particular, the baseline regression specification is estimated by the use of fixed-effects models, employing the within regression estimator, to account for country-specific fixed effects.

[10] Since light intensity is zero for some ethnic groups, 0.01 is added to average luminosity.

[11] Following Wooldridge (2010), pp. 267–268, a zeroth stage is introduced to the analysis, where genetic diversity is first regressed on the migratory distance from East Africa and all the second-stage controls to obtain predicted values of genetic diversity. The predicted genetic diversity from the zeroth stage is squared and the predicted genetic diversity and the squared predicted genetic diversity are then used as excluded instruments in the second stage.

3.3 Baseline Results

This section establishes the hump-shaped impact of observed genetic diversity on luminosity using the sample of 230 ethnicities.[12]

Table 1 presents the results of regression analyses investigating the relationship between observed genetic diversity and log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). The unconditional hump-shaped relationship between genetic diversity and log luminosity is reported in column 1. 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 highly significant hump-shaped relationship ($p < 0.001$).[13] The adjusted R^2 reveals that observed genetic diversity can explain more than 24 percent of the variation in log luminosity across the 230 ethnicities. The analysis suggests that the degree of genetic diversity that is associated with the maximal level of productivity is 0.67, while observed diversity among the ethnic groups in the sample ranges from 0.56 to 0.77.

Reassuringly, column 2–5 reports that the statistically significantly hump-shaped relationship between genetic diversity and light emission per area is robust to the gradual inclusion of controls for log absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects. In all specifications, 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 6 establishes that the finding is robust to the inclusion of the terrain ruggedness measure over the ethnicity areas. In light of possibly direct effects of average temperature and diurnal temperature range on economic performance as well as biodiversity and thus genetic diversity, column 7 reports the results from a regression subjected to these additional control variables. It establishes that the hump-shaped relationship between genetic diversity and luminosity is robust to controlling for the average temperature and the average diurnal temperature range over the ethnicity areas. Additionally, column 8 establishes that the hump-shaped relationship between genetic diversity and luminosity is robust to controlling for precipitation, wet days frequency, and frost day frequency, over the

[12]The analysis omits observation that are not marked as clean in Pemberton et al. (2013)'s data. These omitted observations either do not reflect genetic diversity of a single ethnic group but rather a large geographical region (e.g., Pategonia), or they reflect ethnicities that were subjected to significant admixture. This results in the omission of only two observations for which geographical matching was established. Furthermore, the analysis excludes two ethnicities that are largely viewed as extreme outliers in terms of genetic diversity (e.g. Wang et al., 2007) – the Surui and the Ache of South America. This omission of ethnicities is not particular to our study. The influential research on the Out-of-Africa Hypothesis conducted by Ramachandran et al. (2005), omits the Surui, being "an extreme outlier in a variety of previous analyses", and did not include the Ache as well. In particular, these ethnicities have the lowest levels of genetic diversity in the clean sample and the largest residuals of an OLS regression of genetic diversity on migratory distance from Addis Ababa. Including these observations, nevertheless, does not affect the qualitative analysis.

[13]See Lind and Mehlum (2010).

Table 1: Genetic Diversity and Luminosity

	Outcome Variable: Log Luminosity										
	OLS										2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Expected Heterozygosity	271.813***	285.446***	350.316***	333.404***	304.475***	311.130***	330.430***	358.234***	405.309***	390.219***	1111.474***
	(63.726)	(76.832)	(107.224)	(109.154)	(109.857)	(111.948)	(115.014)	(126.638)	(121.923)	(128.251)	(392.903)
Expected Heterozygosity squared	-208.643***	-218.611***	-270.203***	-257.319***	-237.916***	-243.408***	-255.834***	-271.092***	-303.557***	-292.822***	-840.206***
	(46.553)	(56.478)	(82.325)	(83.687)	(83.828)	(85.619)	(88.109)	(96.413)	(92.755)	(97.439)	(299.072)
Absolute Latitude		-0.004	-0.033*	-0.032*	-0.021	-0.020	0.012	0.042	0.028	0.032	0.002
		(0.014)	(0.018)	(0.019)	(0.020)	(0.020)	(0.026)	(0.029)	(0.027)	(0.028)	(0.027)
Terrain Ruggedness Index						-0.000	-0.000	0.000	-0.000	0.000	-0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.040	0.076	0.086	0.084	0.031
							(0.037)	(0.068)	(0.067)	(0.073)	(0.081)
Diurnal Temperature Range							-0.185**	-0.098	-0.074	-0.060	-0.114
							(0.078)	(0.079)	(0.074)	(0.080)	(0.081)
Precipitation								-0.002	-0.004	-0.004	-0.008**
								(0.004)	(0.004)	(0.004)	(0.004)
Wet Day Frequency								0.116**	0.124**	0.141**	0.166***
								(0.059)	(0.062)	(0.066)	(0.059)
Frost Day Frequency								0.002	0.019	0.022	0.002
								(0.093)	(0.090)	(0.097)	(0.092)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Number of Observations	230	230	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.241	0.238	0.310	0.290	0.312	0.310	0.336	0.352	0.414	0.408	
Expected Heterozygosity at peak	0.651	0.653	0.648	0.648	0.640	0.639	0.646	0.661	0.668	0.666	0.661
95% CI Min	0.622	0.622	0.617	0.611	0.581	0.583	0.605	0.629	0.645	0.641	0.629
95% CI Max	0.665	0.665	0.678	0.681	0.673	0.671	0.677	0.703	0.703	0.705	0.704
Significance of hump-shape	0.001	0.002	0.002	0.004	0.015	0.014	0.006	0.006	0.003	0.005	0.006
1st Stage F-statistic (Kleibergen-Paap)											11.253
Significance of Endogenous Regressors (Anderson-Rubin)											0.000

This table establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

ethnicity areas. In particular, average light intensity per capita is predicted to be maximized at an expected heterozygosity value of 0.661, and an additional test for the presence of a hump-shape rejects the null hypothesis of no hump-shape at the 1 percent significance level.

Columns 9–10 of Table 1 reveal a significant hump-shaped relationship between genetic diversity and light intensity per area and per capita, accounting for the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). The estimated quadratic expressions as well as the additional test for a hump-shaped relationship remain highly significant. In particular, luminosity is predicted to be maximized at an expected heterozygosity value of 0.666, and an additional test for the presence of a hump-shape rejects the null hypothesis of no hump-shape at the 1 percent significance level. The estimated linear and quadratic coefficients in column 10 imply that productivity of the most genetically homogenous ethnicity in the sample (i.e., the Karitiana in Brazil with a genetic diversity of 0.56) is 23.6 percent lower than the maximal level of productivity (associated with a genetic diversity of 0.67). Similarly, the productivity of the most heterogonous ethnicity in the sample (i.e., the Turu in Tanzania with a genetic diversity of 0.77) is 22.0 percent lower than the maximal level of productivity (associated with a genetic diversity of 0.67).

Turning to the 2SLS regression analysis using migratory distance from East Africa as an instrument for observed genetic diversity, the coefficient estimates of column 11 confirms the existence of a hump-shaped relationship. The regression is illustrated in Figure 3. The cluster-robust first-stage F-statistic (Kleibergen and Paap, 2006) is above 11, indicating that the regression is not affected by weak instruments. Reassuringly, 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. Furthermore, the estimated level of genetic diversity that maximizes log luminosity remains stable, highly significant, and is estimated to be 0.661.

Tables A.1a–A.1d in the appendix establish that the findings are robust to alternative specifications wherein a) all geographical variables, i.e., both dependent and independent, are calculated on the basis of the entire ethnic areas, i.e. without omission of any sub-areas due to gas flaring or urban areas (Table A.1a), b) all geographical variables, i.e., both dependent and independent variables, are calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring (Table A.1b), c) the dependent variable is calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring or urban areas while the independent variables are calculated on the basis of the entire ethnic areas (Table A.1c), and d) all geographical variables, i.e., both depen-

[14] 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 genetic diversity, instrumented by the migratory distance from East Africa) of log luminosity 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 genetic diversity rather than the residuals obtained from regressing homogeneity on the control variables in the model. This methodology permits the illustration of the overall non-monotonic effect of genetic diversity in one scatter plot.

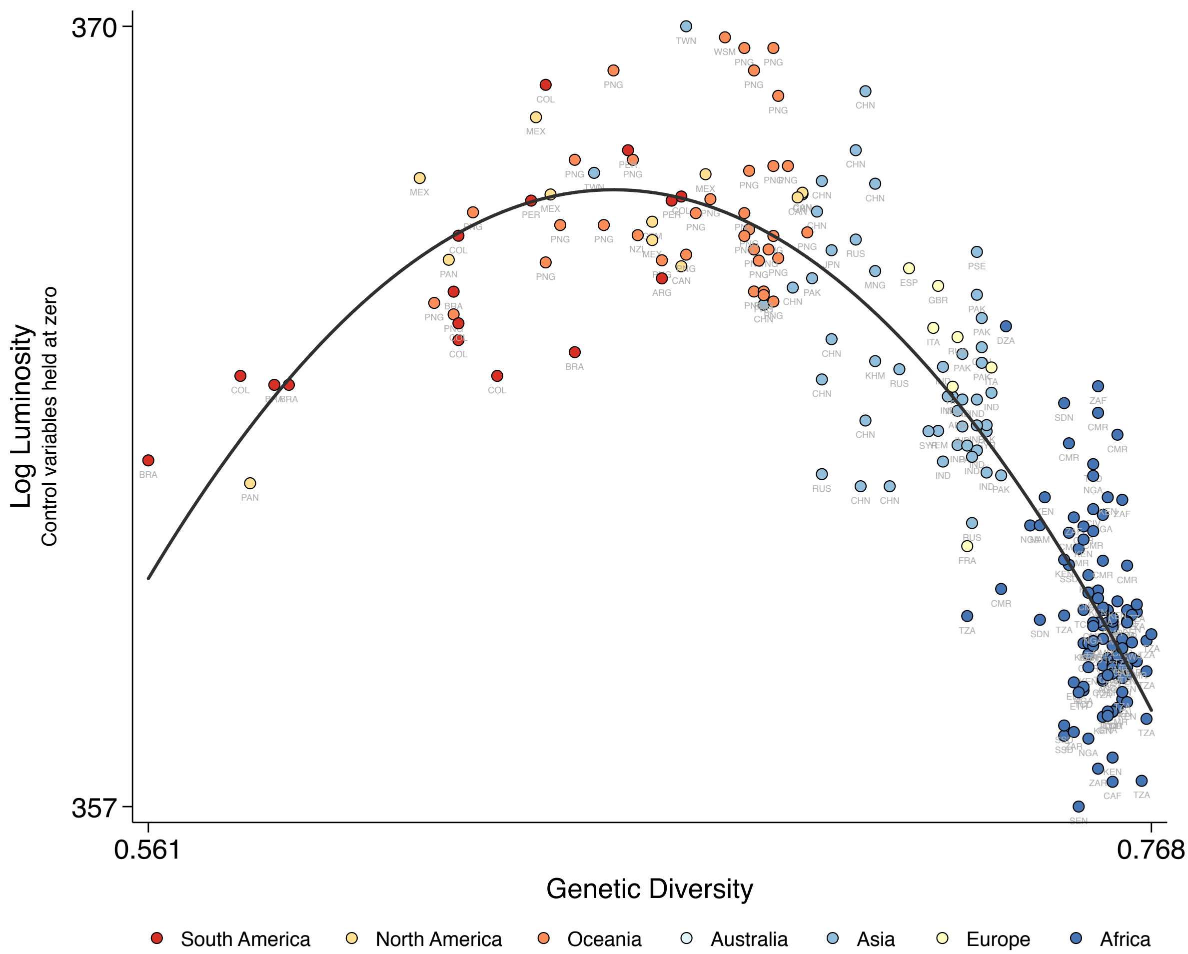

Figure 3: This figure depicts the hump-shaped effect of genetic diversity, instrumented by the migratory distance from East Africa, on log luminosity in 1992–2013 for a sample of 230 ethnic groups, conditional on absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects).[14] Each ethnic group is represented by a point and a country code (following the World Bank) of the country in which the ethnic group resides.

dent and independent variables, are calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring or urban areas (Table A.1d).

3.4 Accounting for Country-Specific Fixed Effects

The analysis further establishes that if one exploits variations in within-ethnic-group genetic diversity across ethnic groups within a country, rather than in the world as a whole, an intermediate level of diversity is still associated with greater group productivity. Table 2 presents the results of regression analyses that account for country-fixed effects in addition to the full set of control variables. As reported in column 1, 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. Further, luminosity is predicted to be maximized at an expected heterozygosity value of 0.679 (95 percent confidence interval: 0.633–0.786).

Reassuringly, columns 2–4 reports that the statistically significant hump-shaped relationship between genetic diversity and luminosity is robust to the gradual inclusion of controls for log absolute latitude, landmass type fixed effects and soil quality fixed effects. In all specifications, the first and the second order elements, as well as the additional test for the presence of a hump-shaped relation, remain highly significant. Furthermore, column 5 establishes that the finding is robust to the inclusion of the terrain ruggedness measure over the ethnicity areas, and column 6 establishes that the hump-shaped relationship between genetic diversity and log luminosity is robust to controlling for the average temperature and the average diurnal temperature range over the ethnicity areas. Additionally, column 7 establishes that the hump-shaped relationship between genetic diversity and log luminosity is robust to controlling for precipitation, wet days frequency, and frost day frequency, over the ethnicity areas. In particular, log luminosity is predicted to be maximized at an expected heterozygosity value of 0.671, and an additional test for the presence of a hump-shape rejects the null hypothesis of no hump-shape.

Finally, columns 8–9 of Table 2 reveal a significant hump-shaped relationship between genetic diversity and log luminosity, accounting for the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). The estimated quadratic expressions as well as the additional test for a hump-shaped relationship remain highly significant. In particular, in column 9, predicted to be maximized at an expected heterozygosity value of 0.676, and an additional test for the presence of a hump-shape rejects the null hypothesis of no hump-shape.

Tables A.2a–A.2d in the appendix establish that the findings are robust to alternative specifications wherein a) all geographical variables, i.e., both dependent and independent, are calculated on the basis of the entire ethnic areas, i.e. without omission of any sub-areas due to gas flaring or urban areas (Table A.2a), b) all geographical variables, i.e., both dependent and independent variables, are calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring (Table

Table 2: Genetic Diversity and Luminosity — Accounting for Country-Specific Fixed Effects

	Outcome Variable: Log Luminosity								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expected Heterozygosity	282.707**	315.278***	295.775***	298.561***	312.796***	356.939***	415.161***	401.710***	377.718***
	(112.090)	(91.313)	(103.644)	(105.264)	(106.676)	(114.048)	(125.000)	(126.257)	(135.209)
Expected Heterozygosity squared	-208.212**	-232.382***	-216.817***	-220.716***	-233.374***	-266.295***	-309.408***	-296.670***	-279.355***
	(84.282)	(68.610)	(78.006)	(79.347)	(81.026)	(87.968)	(96.931)	(97.367)	(103.416)
Absolute Latitude		-0.041**	-0.049**	-0.046**	-0.044**	-0.099***	-0.116***	-0.065**	-0.048
		(0.018)	(0.022)	(0.022)	(0.021)	(0.029)	(0.034)	(0.031)	(0.034)
Terrain Ruggedness Index					-0.000	-0.000***	-0.000***	-0.000***	-0.000***
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature						-0.108**	0.035	0.066	0.095
						(0.043)	(0.115)	(0.124)	(0.121)
Diurnal Temperature Range						-0.011	-0.005	-0.001	0.011
						(0.075)	(0.071)	(0.098)	(0.084)
Precipitation							-0.007***	-0.008***	-0.007***
							(0.002)	(0.001)	(0.002)
Wet Day Frequency							0.080**	0.124***	0.131***
							(0.038)	(0.040)	(0.042)
Frost Day Frequency							0.219	0.182	0.206
							(0.132)	(0.138)	(0.152)
Soil Quality FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	Yes
Number of Observations	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.021	0.032	0.030	0.029	0.033	0.060	0.105	0.220	0.233
Expected Heterozygosity at peak	0.679	0.678	0.682	0.676	0.670	0.670	0.671	0.677	0.676
95% CI Min	0.633	0.646	0.641	0.634	0.634	0.646	0.650	0.657	0.655
95% CI Max	0.786	0.726	0.766	0.752	0.732	0.731	0.728	0.742	0.755
Significance of hump-shape	0.031	0.006	0.024	0.019	0.012	0.012	0.010	0.015	0.021

This table establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, within countries, i.e. accounting for country-fixed effects as well as the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

A.2b), c) the dependent variable is calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring or urban areas while the independent variables are calculated on the basis of the entire ethnic areas (Table A.2c), and d) all geographical variables, i.e., both dependent and independent variables, are calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring or urban areas (Table A.2d).

3.5 Estimation Using Predicted Genetic Diversity on a Larger Sample of Ethnicities

Finally, the study examines the effect of genetic diversity on productivity, exploiting an extended sample of 1,331 ethnic groups for which the study constructs measures of projected genetic diversity, based on their migratory distance from East Africa.

Table 3 reports the results of regressions of the effect of predicted genetic diversity, based on migratory distance from Addis Ababa, on log luminosity. Column 1 establishes that 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 highly significant hump-shaped relationship ($p < 0.0001$). The adjusted R^2 reveal that expected heterozygosity can explain more than 50 percent of the variation in log luminosity across the 1,331 ethnicities. Furthermore, log luminosity is predicted to be maximized at an expected heterozygosity value of 0.657 (95 percent confidence interval: 0.650–0.663).

Reassuringly, columns 2–10 report that the statistically significantly hump-shaped relationship between genetic diversity and log luminosity is robust to the gradual inclusion of controls for log absolute latitude, regional fixed effects, soil quality fixed effects, landmass type fixed effects, terrain ruggedness, average temperature, diurnal temperature range, precipitation, wet day frequency, frost day frequency, century of description in the Ethnographic Atlas fixed effects, mean size of local communities fixed effects, jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects. In all specifications, the first and second order elements, as well as the additional test for the presence of a hump-shaped relation, remain highly significant. In particular, in column 10, log luminosity is predicted to be maximized at an expected heterozygosity value of 0.642, and an additional test for the presence of a hump-shape rejects the null hypothesis of no hump-shape (95 percent confidence interval: 0.594–0.668).

Tables A.3a–A.3d in the appendix establish that the findings are robust to alternative specifications wherein a) all geographical variables, i.e., both dependent and independent, are calculated on the basis of the entire ethnic areas, i.e. without omission of any sub-areas due to gas flaring or urban areas (Table A.3a), b) all geographical variables, i.e., both dependent and independent variables, are calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring (Table A.3b), c) the dependent variable is calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring or urban areas while the independent variables are calculated on the basis of the entire ethnic areas (Table A.3c), and d) all geographical variables, i.e., both depen-

Table 3: Predicted Genetic Diversity and Luminosity

	Outcome Variable: Log Luminosity									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted Expected Heterozygosity	580.527***	473.509***	313.781***	344.817***	359.656***	351.688***	310.834***	311.468***	335.450***	304.746***
	(42.967)	(45.943)	(61.302)	(64.104)	(66.084)	(68.053)	(67.361)	(69.925)	(69.620)	(69.719)
Predicted Expected Heterozygosity squared	-441.486***	-362.447***	-251.163***	-275.148***	-278.897***	-274.613***	-243.549***	-243.836***	-259.365***	-237.195***
	(30.678)	(33.071)	(44.019)	(46.076)	(47.234)	(48.412)	(47.934)	(50.104)	(49.817)	(49.787)
Absolute Latitude		0.024***	-0.000	-0.007	-0.004	-0.006	0.004	0.005	0.012	0.012
		(0.004)	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.012)	(0.011)	(0.011)
Terrain Ruggedness Index						0.000	0.000*	0.000*	0.000	0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.027*	0.031	0.043*	0.037*
							(0.015)	(0.024)	(0.023)	(0.022)
Diurnal Temperature Range							-0.072***	-0.071**	-0.095***	-0.095***
							(0.023)	(0.030)	(0.030)	(0.029)
Precipitation								-0.000	-0.001	-0.002
								(0.002)	(0.002)	(0.002)
Wet Day Frequency								0.007	0.007	0.007
								(0.019)	(0.019)	(0.019)
Frost Day Frequency								0.003	-0.007	-0.009
								(0.027)	(0.026)	(0.026)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes
Number of Observations	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331
Adjusted R^2	0.501	0.520	0.537	0.546	0.578	0.578	0.583	0.582	0.609	0.619
Expected Heterozygosity at peak	0.657	0.653	0.625	0.627	0.645	0.640	0.638	0.639	0.647	0.642
95% CI Min	0.650	0.643	0.582	0.588	0.612	0.601	0.591	0.591	0.607	0.594
95% CI Max	0.663	0.660	0.648	0.649	0.665	0.663	0.664	0.664	0.670	0.668
Significance of hump-shape	0.000	0.000	0.037	0.024	0.003	0.009	0.021	0.020	0.006	0.017

This table establishes the significant hump-shaped relationship between predicted genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

dent and independent variables, are calculated on the basis of the ethnic areas omitting sub-areas characterized by gas flaring or urban areas (Table A.3d). Furthermore, in light of the significant displacement of native American populations in the United States, and thus the disassociation between genetic diversity of the native population of each ethnic enclave and the genetic diversity of the contemporary inhabitants in this territory, Table A.3e establishes that the results are robust to the exclusion of ethnic groups originating in the United States.

4 Concluding Remarks

This research explores the effects of within-group heterogeneity on group-level productivity. The study provides the first ethnic-level analysis that captures the trade-off between the beneficial and the detrimental effects of diversity on productivity, demonstrating that an intermediate level of diversity is associated with greater group productivity. It exploits an exogenous source of variation in genetic diversity among ethnic groups across the globe, and establishes that genetic diversity has a hump-shaped effect on economic prosperity, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. Furthermore, the study demonstrates that variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country.

The analysis is performed using a newly constructed geo-referenced dataset on within-ethnic-group genetic diversity for a large sample of 230 ethnic groups across the globe, greatly surpassing the Human Genome Diversity Project sample of 53 ethnic groups. In particular, the study constructs a novel geo-referenced dataset mapping the observed genetic diversity for 230 ethnic groups across the globe (Pemberton et al., 2013) to the geographical attributes that characterizes their historical homelands as well as the ethnographic characteristics of 145 of these ethnic groups. Moreover, the analysis is further performed on an extended sample of 1,331 ethnic groups for which the study constructs a measure of projected genetic diversity, based on migratory distance of these ethnic groups from East Africa, linking it to the geographical attributes of the historical homelands and ethnographic characteristics of these ethnic groups.

Exploiting variation in observed genetic diversity within 230 ethnic groups, the empirical analysis establishes that observed genetic diversity within 230 ethnic groups has a hump-shaped relationship with economic prosperity, as captured by luminosity. The analysis suggests that the degree of genetic diversity that is associated with the maximal level of productivity is rather stable and is estimated to be 0.67, while observed diversity among the ethnic groups in the sample ranges from 0.56 to 0.77. Furthermore, the analysis establishes that the association between diversity and group productivity is sizable economically. In particular, it implies that productivity of the most genetically homogenous ethnicity in the sample (i.e., the Karitiana in Brazil with a genetic diversity of 0.56) is 23.6 percent lower than the maximal level of productivity (associated with a genetic diversity of 0.67). Similarly, the productivity of the most heterogonous ethnicity in the sample (i.e., the Turu in Tanzania with a genetic diversity of 0.77) is 22.0 percent lower than the maximal level of productivity (associated with a genetic diversity of 0.67).

In light of potential endogeneity concerns, the analysis exploits migratory distance from East Africa as an exogenous source of variation in observed genetic diversity within ethnic groups and establishes the hump-shaped effect of genetic diversity on economic prosperity. In particular the IV estimation suggests that the degree of genetic diversity that is associated with the maximal level of productivity is rather stable and is estimated to be 0.66.

The study provides a novel angle for the understanding of the origins of regional inequality within nations. It establishes that variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country. In particular, that analysis suggests that, accounting for country fixed effects, the degree of genetic diversity that is associated with the maximal level of productivity is still rather stable and is estimated to be 0.68. Thus, variations in within-ethnic-group genetic diversity across ethnic groups contribute to ethnic and thus regional variation in economic development within a country.

Finally, exploiting an extended sample of 1,331 ethnic groups for which the study constructs measures of projected genetic diversity, based on their migratory distance from East Africa, the research confirms the robustness of the hump-shaped effect of group-level heterogeneity on productivity. In particular that estimation suggests that the degree of genetic diversity that is associated with the maximal level of productivity is estimated to be 0.64.

References

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

Ashraf, Q. and O. Galor (2013b): "Genetic Diversity and the Origins of Cultural Fragmentation," *American Economic Review*, 103, 528–33.

Ashraf, Q., O. Galor, and M. Klemp (2014): "The out of Africa hypothesis of comparative development reflected by nighttime light intensity," *Working Paper, Brown University, Department of Economics*.

Atkinson, Q. D. (2011): "Phonemic diversity supports a serial founder effect model of language expansion from Africa," *Science*, 332, 346–349.

Betti, L., F. Balloux, W. Amos, T. Hanihara, and A. Manica (2009): "Distance from Africa, not climate, explains within-population phenotypic diversity in humans," *Proceedings of the Royal Society B: Biological Sciences*, 276, 809–814.

Betti, L., N. von Cramon-Taubadel, A. Manica, and S. J. Lycett (2013): "Global geometric morphometric analyses of the human pelvis reveal substantial neutral population history effects, even across sexes," *PloS one*, 8, e55909.

Center for International Earth Science Information Network - CIESIN - Columbia University, United Nations Food and Agriculture Programme - FAO, and Centro Internacional de Agricultura Tropical - CIAT (2005): "Gridded Population of the World, Version 3 (GPWv3)," `http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-count`.

Chen, X. and W. D. Nordhaus (2011): "Using luminosity data as a proxy for economic statistics," *Proceedings of the National Academy of Sciences*, 108, 8589–8594.

Galor, O. and M. Klemp (2015): "The Roots of Autocracy," *Mimeo, Brown University, Department of Economics*.

Hanihara, T. (2008): "Morphological variation of major human populations based on nonmetric dental traits," *American journal of physical anthropology*, 136, 169–182.

Harris, I., P. Jones, T. Osborn, and D. Lister (2014): "Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset," *International Journal of Climatology*, 34, 623–642.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012): "Measuring economic growth from outer space," *American economic review*, 102, 994–1028.

Kleibergen, F. and R. Paap (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of econometrics*, 133, 97–126.

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.

Manica, A., W. Amos, F. Balloux, and T. Hanihara (2007): "The effect of ancient population bottlenecks on human phenotypic variation," *Nature*, 448, 346–348.

Michalopoulos, S. (2012): "The origins of ethnolinguistic diversity," *The American economic review*, 102, 1508.

Nunn, N. and D. Puga (2012): "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 94, 20–36.

Nunn, N. and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *The American Economic Review*, 101, 3221–3252.

Pemberton, T. J., M. DeGiorgio, and N. A. Rosenberg (2013): "Population structure in a comprehensive genomic data set on human microsatellite variation," *G3: Genes— Genomes— Genetics*, g3–113.

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.

Ramankutty, N., J. A. Foley, J. Norman, and K. McSweeney (2002): "The global distribution of cultivable lands: current patterns and sensitivity to possible climate change," *Global Ecology and Biogeography*, 11, 377–392.

von Cramon-Taubadel, N. and S. J. Lycett (2008): "Brief communication: human cranial variation fits iterative founder effect model with African origin," *American Journal of Physical Anthropology*, 136, 108–113.

Wang, S., C. M. Lewis Jr, M. Jakobsson, S. Ramachandran, N. Ray, G. Bedoya, W. Rojas, M. V. Parra, J. A. Molina, C. Gallo, et al. (2007): "Genetic variation and population structure in Native Americans," *PLoS Genetics*, 3, e185.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*, MIT press.

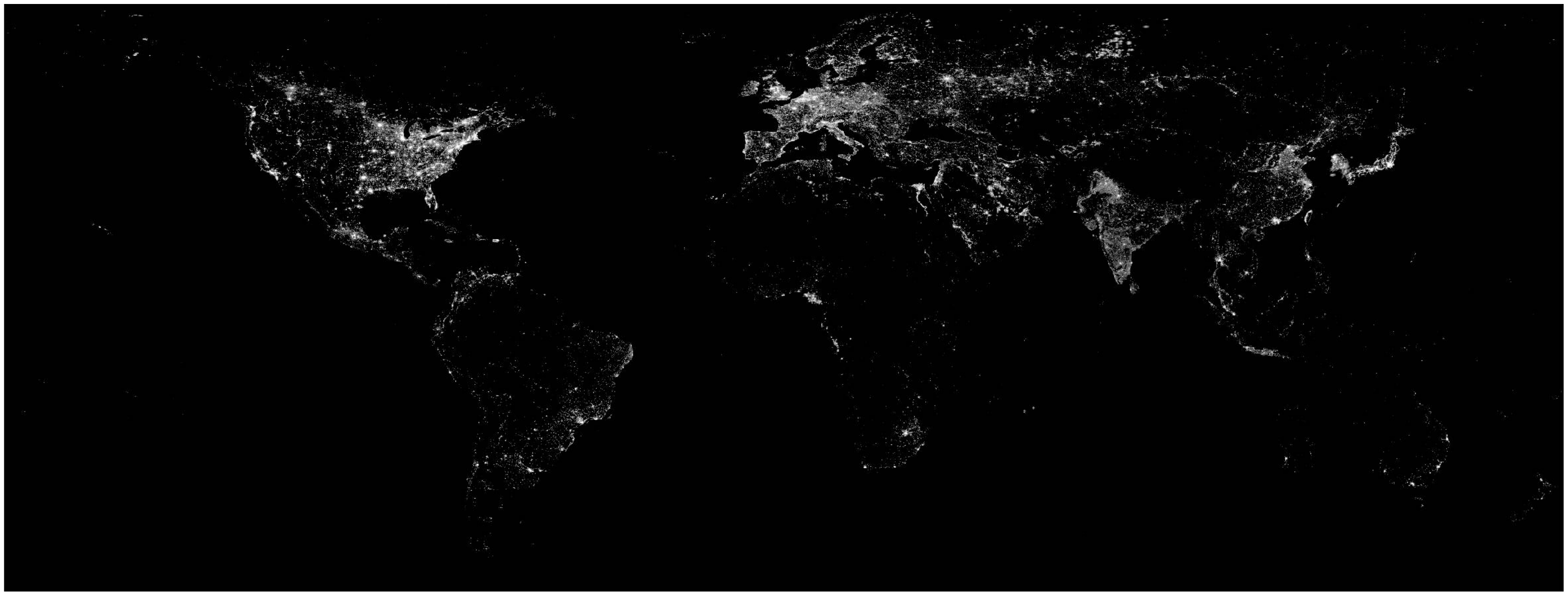

Figure A.1: Significantly downscaled representation of the average nighttime light emission data within country borders, 1992–2013. Data source: NOAA-NGDC.

Table A.1a: Genetic Diversity and Luminosity — No Excluded Sub-Areas

	Outcome Variable: Log Luminosity										
	OLS										2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Expected Heterozygosity	266.241***	278.970***	340.946***	322.267***	293.135***	301.113***	318.749***	340.553***	386.048***	370.153***	1094.028***
	(63.758)	(76.733)	(107.131)	(108.773)	(109.496)	(111.511)	(114.283)	(125.427)	(120.263)	(126.456)	(392.879)
Expected Heterozygosity squared	-204.317***	-213.624***	-262.896***	-248.662***	-229.124***	-235.708***	-246.727***	-257.460***	-288.820***	-277.434***	-826.276***
	(46.565)	(56.395)	(82.269)	(83.441)	(83.600)	(85.355)	(87.593)	(95.499)	(91.480)	(95.992)	(298.810)
Absolute Latitude		-0.004	-0.033*	-0.032*	-0.020	-0.019	0.015	0.044	0.030	0.034	0.004
		(0.014)	(0.018)	(0.019)	(0.020)	(0.020)	(0.025)	(0.029)	(0.027)	(0.028)	(0.027)
Terrain Ruggedness Index						-0.000	-0.000	0.000	-0.000	0.000	-0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.045	0.088	0.098	0.097	0.045
							(0.037)	(0.067)	(0.066)	(0.071)	(0.080)
Diurnal Temperature Range							-0.189**	-0.103	-0.079	-0.068	-0.122
							(0.078)	(0.079)	(0.074)	(0.079)	(0.080)
Precipitation								-0.002	-0.003	-0.004	-0.008**
								(0.004)	(0.004)	(0.004)	(0.004)
Wet Day Frequency								0.113*	0.121*	0.136**	0.163***
								(0.059)	(0.062)	(0.065)	(0.058)
Frost Day Frequency								0.018	0.035	0.038	0.018
								(0.092)	(0.088)	(0.095)	(0.091)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Number of Observations	230	230	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.230	0.227	0.301	0.281	0.303	0.302	0.332	0.346	0.414	0.409	
Expected Heterozygosity at peak	0.652	0.653	0.648	0.648	0.640	0.639	0.646	0.661	0.668	0.667	0.662
95% CI Min	0.621	0.620	0.616	0.609	0.575	0.578	0.603	0.627	0.645	0.640	0.629
95% CI Max	0.665	0.665	0.679	0.683	0.674	0.672	0.679	0.709	0.707	0.710	0.707
Significance of hump-shape	0.001	0.002	0.002	0.005	0.018	0.017	0.008	0.008	0.004	0.006	0.007
1st Stage F-statistic (Kleibergen-Paap)											11.253
Significance of Endogenous Regressors (Anderson-Rubin)											0.000

This table establishes the robustness of the results of the regression analyses in Table 1 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.1b: Genetic Diversity and Luminosity — W/o Gasflaring Areas in Outcome and Controls

	Outcome Variable: Log Luminosity										
	OLS										2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Expected Heterozygosity	271.813***	285.446***	350.316***	333.404***	304.475***	311.499***	330.801***	358.749***	406.171***	391.018***	1113.108***
	(63.726)	(76.832)	(107.224)	(109.154)	(109.857)	(111.962)	(115.016)	(126.626)	(122.028)	(128.349)	(393.369)
Expected Heterozygosity squared	-208.643***	-218.611***	-270.203***	-257.319***	-237.916***	-243.696***	-256.130***	-271.474***	-304.228***	-293.447***	-841.403***
	(46.553)	(56.478)	(82.325)	(83.687)	(83.828)	(85.632)	(88.115)	(96.410)	(92.835)	(97.516)	(299.392)
Absolute Latitude		-0.004	-0.033*	-0.032*	-0.021	-0.020	0.012	0.042	0.028	0.032	0.002
		(0.014)	(0.018)	(0.019)	(0.020)	(0.020)	(0.026)	(0.029)	(0.027)	(0.028)	(0.027)
Terrain Ruggedness Index						-0.000	-0.000	0.000	-0.000	0.000	-0.000
(W/o Gas-Flaring Areas)						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.040	0.076	0.085	0.083	0.030
(W/o Gas-Flaring Areas)							(0.037)	(0.068)	(0.067)	(0.073)	(0.081)
Diurnal Temperature Range							-0.185**	-0.098	-0.074	-0.060	-0.114
(W/o Gas-Flaring Areas)							(0.078)	(0.079)	(0.074)	(0.080)	(0.081)
Precipitation								-0.002	-0.004	-0.004	-0.008**
(W/o Gas-Flaring Areas)								(0.004)	(0.004)	(0.004)	(0.004)
Wet Day Frequency								0.117**	0.124**	0.141**	0.166***
(W/o Gas-Flaring Areas)								(0.059)	(0.062)	(0.066)	(0.059)
Frost Day Frequency								0.002	0.019	0.021	0.001
(W/o Gas-Flaring Areas)								(0.093)	(0.090)	(0.097)	(0.092)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Number of Observations	230	230	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.241	0.238	0.310	0.290	0.312	0.310	0.336	0.352	0.414	0.408	
Expected Heterozygosity at peak	0.651	0.653	0.648	0.648	0.640	0.639	0.646	0.661	0.668	0.666	0.661
95% CI Min	0.622	0.622	0.617	0.611	0.581	0.583	0.605	0.629	0.645	0.641	0.629
95% CI Max	0.665	0.665	0.678	0.681	0.673	0.671	0.677	0.703	0.703	0.705	0.704
Significance of hump-shape	0.001	0.002	0.002	0.004	0.015	0.014	0.006	0.006	0.003	0.005	0.006
1st Stage F-statistic (Kleibergen-Paap)											11.198
Significance of Endogenous Regressors (Anderson-Rubin)											0.000

This table establishes the robustness of the results of the regression analyses in Table 1 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.1c: Genetic Diversity and Luminosity — W/o Gasflaring and Urban Areas in Outcome but No Excluded Sub-Areas in Controls

	Outcome Variable: Log Luminosity										
	OLS										2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Expected Heterozygosity	277.358***	290.356***	351.784***	335.481***	306.709***	313.099***	333.874***	362.750***	409.633***	392.647***	1109.713***
	(63.889)	(77.027)	(107.486)	(109.695)	(110.378)	(112.435)	(115.766)	(127.394)	(122.604)	(129.108)	(392.405)
Expected Heterozygosity squared	-212.896***	-222.399***	-271.248***	-258.831***	-239.525***	-244.798***	-258.392***	-274.519***	-306.766***	-294.532***	-838.353***
	(46.677)	(56.626)	(82.514)	(84.090)	(84.219)	(85.974)	(88.704)	(97.003)	(93.296)	(98.099)	(298.751)
Absolute Latitude		-0.004	-0.034*	-0.034*	-0.023	-0.022	0.010	0.039	0.026	0.031	0.001
		(0.014)	(0.018)	(0.019)	(0.020)	(0.020)	(0.026)	(0.029)	(0.027)	(0.029)	(0.027)
Terrain Ruggedness Index						-0.000	-0.000	0.000	-0.000	0.000	-0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.039	0.072	0.083	0.082	0.030
							(0.037)	(0.067)	(0.067)	(0.073)	(0.081)
Diurnal Temperature Range							-0.186**	-0.100	-0.076	-0.064	-0.117
							(0.078)	(0.079)	(0.074)	(0.080)	(0.081)
Precipitation								-0.002	-0.004	-0.004	-0.008**
								(0.004)	(0.004)	(0.004)	(0.004)
Wet Day Frequency								0.116*	0.124**	0.140**	0.166***
								(0.059)	(0.062)	(0.066)	(0.059)
Frost Day Frequency								-0.003	0.015	0.018	-0.002
								(0.093)	(0.090)	(0.097)	(0.093)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Number of Observations	230	230	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.248	0.245	0.318	0.298	0.319	0.317	0.343	0.358	0.421	0.414	
Expected Heterozygosity at peak	0.651	0.653	0.648	0.648	0.640	0.640	0.646	0.661	0.668	0.667	0.662
95% CI Min	0.623	0.622	0.617	0.611	0.582	0.584	0.606	0.629	0.645	0.642	0.630
95% CI Max	0.665	0.665	0.678	0.681	0.673	0.671	0.677	0.702	0.703	0.706	0.706
Significance of hump-shape	0.001	0.002	0.002	0.004	0.015	0.014	0.006	0.005	0.003	0.005	0.006
1st Stage F-statistic (Kleibergen-Paap)											11.253
Significance of Endogenous Regressors (Anderson-Rubin)											0.000

This table establishes the robustness of the results of the regression analyses in Table 1 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.1d: Genetic Diversity and Luminosity — W/o Gasflaring and Urban Areas in Outcome and Controls

	Outcome Variable: Log Luminosity										
	OLS										2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Expected Heterozygosity	277.358***	290.356***	351.784***	335.481***	306.709***	313.440***	334.239***	363.271***	410.474***	393.382***	1111.125***
	(63.889)	(77.027)	(107.486)	(109.695)	(110.378)	(112.444)	(115.759)	(127.378)	(122.717)	(129.212)	(392.881)
Expected Heterozygosity squared	-212.896***	-222.399***	-271.248***	-258.831***	-239.525***	-245.063***	-258.683***	-274.906***	-307.420***	-295.107***	-839.391***
	(46.677)	(56.626)	(82.514)	(84.090)	(84.219)	(85.983)	(88.702)	(96.998)	(93.381)	(98.179)	(299.076)
Absolute Latitude		-0.004	-0.034*	-0.034*	-0.023	-0.022	0.010	0.039	0.026	0.030	0.001
		(0.014)	(0.018)	(0.019)	(0.020)	(0.020)	(0.026)	(0.029)	(0.027)	(0.029)	(0.027)
Terrain Ruggedness Index						-0.000	-0.000	0.000	-0.000	0.000	-0.000
(W/o Gas-Flaring or Urban Areas)						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.038	0.071	0.082	0.081	0.029
(W/o Gas-Flaring or Urban Areas)							(0.037)	(0.067)	(0.067)	(0.073)	(0.081)
Diurnal Temperature Range							-0.186**	-0.100	-0.076	-0.064	-0.116
(W/o Gas-Flaring or Urban Areas)							(0.078)	(0.079)	(0.074)	(0.080)	(0.081)
Precipitation								-0.002	-0.004	-0.004	-0.008**
(W/o Gas-Flaring or Urban Areas)								(0.004)	(0.004)	(0.004)	(0.004)
Wet Day Frequency								0.116*	0.124**	0.140**	0.167***
(W/o Gas-Flaring or Urban Areas)								(0.059)	(0.062)	(0.066)	(0.059)
Frost Day Frequency								-0.003	0.015	0.018	-0.002
(W/o Gas-Flaring or Urban Areas)								(0.093)	(0.090)	(0.097)	(0.093)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes	Yes
Number of Observations	230	230	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.248	0.245	0.318	0.298	0.319	0.317	0.343	0.358	0.421	0.414	
Expected Heterozygosity at peak	0.651	0.653	0.648	0.648	0.640	0.640	0.646	0.661	0.668	0.667	0.662
95% CI Min	0.623	0.622	0.617	0.611	0.582	0.584	0.606	0.629	0.645	0.642	0.630
95% CI Max	0.665	0.665	0.678	0.681	0.673	0.671	0.677	0.702	0.703	0.706	0.706
Significance of hump-shape	0.001	0.002	0.002	0.004	0.015	0.014	0.006	0.005	0.003	0.005	0.006
1st Stage F-statistic (Kleibergen-Paap)											11.197
Significance of Endogenous Regressors (Anderson-Rubin)											0.000

This table establishes the robustness of the results of the regression analyses in Table 1 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.2a: Genetic Diversity and Luminosity — Accounting for Country-Fixed Effects — No Excluded Sub-Areas

	Outcome Variable: Log Luminosity								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expected Heterozygosity	270.433**	304.032***	279.337***	282.099***	296.461***	342.895***	402.803***	391.547***	367.222***
	(108.650)	(87.738)	(100.587)	(102.085)	(103.225)	(111.179)	(122.668)	(124.385)	(131.507)
Expected Heterozygosity squared	-198.589**	-223.521***	-204.018***	-207.863***	-220.635***	-255.423***	-299.869***	-288.795***	-271.108***
	(81.527)	(65.739)	(75.320)	(76.535)	(78.049)	(85.469)	(94.929)	(95.632)	(100.451)
Absolute Latitude		-0.042**	-0.050**	-0.047**	-0.045**	-0.099***	-0.116***	-0.066**	-0.051
		(0.018)	(0.022)	(0.022)	(0.020)	(0.028)	(0.033)	(0.031)	(0.035)
Terrain Ruggedness Index					-0.000	-0.000***	-0.000***	-0.000***	-0.000***
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature						-0.109**	0.041	0.073	0.103
						(0.042)	(0.115)	(0.125)	(0.123)
Diurnal Temperature Range						-0.024	-0.016	-0.011	-0.005
						(0.072)	(0.068)	(0.096)	(0.082)
Precipitation							-0.006***	-0.007***	-0.007***
							(0.002)	(0.001)	(0.002)
Wet Day Frequency							0.079*	0.121***	0.129***
							(0.039)	(0.041)	(0.041)
Frost Day Frequency							0.230*	0.195	0.221
							(0.132)	(0.137)	(0.152)
Soil Quality FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	Yes
Number of Observations	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.020	0.032	0.035	0.034	0.038	0.067	0.113	0.240	0.255
Expected Heterozygosity at peak	0.681	0.680	0.685	0.679	0.672	0.671	0.672	0.678	0.677
95% CI Min	0.634	0.648	0.642	0.633	0.634	0.646	0.650	0.658	0.656
95% CI Max	0.798	0.728	0.774	0.757	0.735	0.733	0.729	0.743	0.757
Significance of hump-shape	0.034	0.006	0.028	0.021	0.013	0.012	0.010	0.015	0.021

This table establishes the robustness of the results of the regression analyses in Table 2 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, within countries, i.e. accounting for country-fixed effects as well as the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.2b: Genetic Diversity and Luminosity — Accounting for Country-Fixed Effects — W/o Gasflaring Areas in Outcome and Controls

	Outcome Variable: Log Luminosity								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expected Heterozygosity	282.707**	315.278***	295.775***	298.561***	312.867***	357.029***	415.219***	401.700***	377.668***
	(112.090)	(91.313)	(103.644)	(105.264)	(106.605)	(113.946)	(124.954)	(126.205)	(135.155)
Expected Heterozygosity squared	-208.212**	-232.382***	-216.817***	-220.716***	-233.452***	-266.381***	-309.464***	-296.671***	-279.322***
	(84.282)	(68.610)	(78.006)	(79.347)	(80.979)	(87.894)	(96.892)	(97.317)	(103.365)
Absolute Latitude		-0.041**	-0.049**	-0.046**	-0.044**	-0.099***	-0.116***	-0.065**	-0.048
		(0.018)	(0.022)	(0.022)	(0.021)	(0.028)	(0.033)	(0.031)	(0.034)
Terrain Ruggedness Index					-0.000	-0.000***	-0.000***	-0.000***	-0.000***
(W/o Gas-Flaring Areas)					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature						-0.108**	0.035	0.065	0.095
(W/o Gas-Flaring Areas)						(0.043)	(0.115)	(0.124)	(0.122)
Diurnal Temperature Range						-0.012	-0.005	-0.001	0.011
(W/o Gas-Flaring Areas)						(0.075)	(0.071)	(0.098)	(0.084)
Precipitation							-0.007***	-0.008***	-0.007***
(W/o Gas-Flaring Areas)							(0.002)	(0.001)	(0.002)
Wet Day Frequency							0.080**	0.124***	0.131***
(W/o Gas-Flaring Areas)							(0.038)	(0.040)	(0.042)
Frost Day Frequency							0.219	0.182	0.206
(W/o Gas-Flaring Areas)							(0.132)	(0.138)	(0.152)
Soil Quality FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	Yes
Number of Observations	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.021	0.032	0.030	0.029	0.033	0.060	0.105	0.221	0.233
Expected Heterozygosity at peak	0.679	0.678	0.682	0.676	0.670	0.670	0.671	0.677	0.676
95% CI Min	0.633	0.646	0.641	0.634	0.634	0.646	0.650	0.657	0.655
95% CI Max	0.786	0.726	0.766	0.752	0.732	0.731	0.728	0.742	0.754
Significance of hump-shape	0.031	0.006	0.024	0.019	0.012	0.012	0.010	0.015	0.021

This table establishes the robustness of the results of the regression analyses in Table 2 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, within countries, i.e. accounting for country-fixed effects as well as the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.2c: Genetic Diversity and Luminosity — Accounting for Country-Fixed Effects — W/o Gasflaring and Urban Areas in Outcome but No Excluded Sub-Areas in Controls

	Outcome Variable: Log Luminosity								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expected Heterozygosity	286.025**	319.941***	302.583***	305.322***	319.896***	364.118***	423.983***	409.956***	384.521***
	(113.078)	(90.969)	(103.610)	(105.266)	(106.703)	(114.100)	(125.275)	(127.041)	(135.641)
Expected Heterozygosity squared	-210.797**	-235.964***	-222.053***	-225.885***	-238.845***	-271.842***	-316.223***	-303.000***	-284.535***
	(85.015)	(68.332)	(78.020)	(79.402)	(81.115)	(88.052)	(97.205)	(97.990)	(103.825)
Absolute Latitude		-0.042**	-0.051**	-0.049**	-0.047**	-0.101***	-0.119***	-0.066**	-0.049
		(0.019)	(0.023)	(0.023)	(0.021)	(0.029)	(0.034)	(0.032)	(0.035)
Terrain Ruggedness Index					-0.000	-0.000***	-0.000***	-0.000***	-0.000***
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature						-0.107**	0.040	0.075	0.105
						(0.044)	(0.116)	(0.126)	(0.125)
Diurnal Temperature Range						-0.013	-0.006	-0.002	0.010
						(0.076)	(0.074)	(0.101)	(0.088)
Precipitation							-0.007***	-0.008***	-0.007***
							(0.002)	(0.001)	(0.002)
Wet Day Frequency							0.080**	0.125***	0.134***
							(0.039)	(0.040)	(0.042)
Frost Day Frequency							0.227*	0.193	0.216
							(0.133)	(0.140)	(0.154)
Soil Quality FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	Yes
Number of Observations	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.021	0.034	0.030	0.028	0.033	0.059	0.106	0.222	0.234
Expected Heterozygosity at peak	0.678	0.678	0.681	0.676	0.670	0.670	0.670	0.676	0.676
95% CI Min	0.633	0.647	0.642	0.635	0.635	0.646	0.649	0.657	0.654
95% CI Max	0.782	0.724	0.760	0.747	0.728	0.728	0.725	0.739	0.751
Significance of hump-shape	0.030	0.005	0.021	0.016	0.010	0.010	0.009	0.013	0.019

This table establishes the robustness of the results of the regression analyses in Table 2 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, within countries, i.e. accounting for country-fixed effects as well as the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.2d: Genetic Diversity and Luminosity — Accounting for Country-Fixed Effects — W/o Gasflaring and Urban Areas in Outcome and Controls

	Outcome Variable: Log Luminosity								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expected Heterozygosity	286.025**	319.941***	302.583***	305.322***	319.935***	364.173***	424.042***	409.860***	384.384***
	(113.078)	(90.969)	(103.610)	(105.266)	(106.625)	(113.989)	(125.217)	(126.973)	(135.582)
Expected Heterozygosity squared	-210.797**	-235.964***	-222.053***	-225.885***	-238.895***	-271.898***	-316.275***	-302.932***	-284.433***
	(85.015)	(68.332)	(78.020)	(79.402)	(81.060)	(87.970)	(97.156)	(97.930)	(103.771)
Absolute Latitude		-0.042**	-0.051**	-0.049**	-0.047**	-0.101***	-0.118***	-0.066**	-0.049
		(0.019)	(0.023)	(0.023)	(0.021)	(0.029)	(0.034)	(0.032)	(0.035)
Terrain Ruggedness Index					-0.000	-0.000***	-0.000***	-0.000***	-0.000***
(W/o Gas-Flaring or Urban Areas)					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature						-0.107**	0.041	0.074	0.104
(W/o Gas-Flaring or Urban Areas)						(0.044)	(0.116)	(0.126)	(0.125)
Diurnal Temperature Range						-0.013	-0.006	-0.002	0.010
(W/o Gas-Flaring or Urban Areas)						(0.076)	(0.074)	(0.101)	(0.088)
Precipitation							-0.007***	-0.008***	-0.007***
(W/o Gas-Flaring or Urban Areas)							(0.002)	(0.001)	(0.002)
Wet Day Frequency							0.080**	0.125***	0.134***
(W/o Gas-Flaring or Urban Areas)							(0.038)	(0.040)	(0.042)
Frost Day Frequency							0.227*	0.193	0.216
(W/o Gas-Flaring or Urban Areas)							(0.133)	(0.140)	(0.155)
Soil Quality FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	Yes
Number of Observations	230	230	230	230	230	230	230	230	230
Adjusted R^2	0.021	0.034	0.030	0.028	0.033	0.059	0.106	0.222	0.234
Expected Heterozygosity at peak	0.678	0.678	0.681	0.676	0.670	0.670	0.670	0.676	0.676
95% CI Min	0.633	0.647	0.642	0.635	0.635	0.646	0.649	0.657	0.654
95% CI Max	0.782	0.724	0.760	0.747	0.728	0.727	0.725	0.739	0.751
Significance of hump-shape	0.030	0.005	0.021	0.016	0.010	0.010	0.009	0.013	0.019

This table establishes the robustness of the results of the regression analyses in Table 2 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between observed genetic diversity and the dependent variable, log luminosity, within countries, i.e. accounting for country-fixed effects as well as the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.3a: Predicted Genetic Diversity and Luminosity — No Excluded Sub-Areas

	Outcome Variable: Log Luminosity									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted Expected Heterozygosity	573.791***	464.429***	305.325***	335.703***	350.550***	343.379***	301.964***	300.773***	325.614***	294.066***
	(43.054)	(45.813)	(61.581)	(64.269)	(66.685)	(68.499)	(67.540)	(70.016)	(69.428)	(69.302)
Predicted Expected Heterozygosity squared	-436.583***	-355.813***	-245.360***	-268.897***	-272.686***	-268.831***	-237.342***	-236.319***	-252.429***	-229.625***
	(30.739)	(32.981)	(44.202)	(46.177)	(47.629)	(48.701)	(48.039)	(50.147)	(49.676)	(49.486)
Absolute Latitude		0.025***	0.000	-0.007	-0.004	-0.006	0.005	0.006	0.012	0.011
		(0.004)	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.012)	(0.011)	(0.011)
Terrain Ruggedness Index						0.000	0.000	0.000	0.000	0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.027*	0.032	0.043*	0.037*
							(0.015)	(0.024)	(0.023)	(0.022)
Diurnal Temperature Range							-0.074***	-0.073**	-0.097***	-0.098***
							(0.023)	(0.030)	(0.030)	(0.029)
Precipitation								-0.000	-0.001	-0.001
								(0.002)	(0.002)	(0.002)
Wet Day Frequency								0.005	0.005	0.005
								(0.019)	(0.019)	(0.019)
Frost Day Frequency								0.005	-0.006	-0.007
								(0.027)	(0.026)	(0.026)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes
Number of Observations	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331
Adjusted R^2	0.499	0.518	0.536	0.545	0.577	0.577	0.582	0.581	0.608	0.619
Expected Heterozygosity at peak	0.657	0.653	0.622	0.624	0.643	0.639	0.636	0.636	0.645	0.640
95% CI Min	0.650	0.643	0.577	0.583	0.607	0.597	0.585	0.585	0.603	0.589
95% CI Max	0.663	0.660	0.647	0.648	0.664	0.662	0.663	0.663	0.669	0.667
Significance of hump-shape	0.000	0.000	0.052	0.036	0.005	0.012	0.029	0.030	0.008	0.024

This table establishes the robustness of the results of the regression analyses in Table 3 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between predicted genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.3b: Predicted Genetic Diversity and Luminosity — W/o Gasflaring Areas in Outcome and Controls

	Outcome Variable: Log Luminosity									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted Expected Heterozygosity	580.527***	473.509***	313.781***	344.817***	359.656***	351.942***	311.152***	311.258***	335.208***	304.578***
	(42.967)	(45.943)	(61.302)	(64.104)	(66.084)	(68.021)	(67.327)	(69.866)	(69.566)	(69.676)
Predicted Expected Heterozygosity squared	-441.486***	-362.447***	-251.163***	-275.148***	-278.897***	-274.775***	-243.760***	-243.667***	-259.180***	-237.066***
	(30.678)	(33.071)	(44.019)	(46.076)	(47.234)	(48.392)	(47.914)	(50.068)	(49.783)	(49.760)
Absolute Latitude		0.024***	-0.000	-0.007	-0.004	-0.006	0.004	0.005	0.012	0.012
		(0.004)	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.012)	(0.011)	(0.011)
Terrain Ruggedness Index						0.000	0.000*	0.000*	0.000	0.000
(W/o Gas-Flaring Areas)						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.027*	0.031	0.042*	0.037
(W/o Gas-Flaring Areas)							(0.015)	(0.024)	(0.023)	(0.022)
Diurnal Temperature Range							-0.072***	-0.071**	-0.095***	-0.095***
(W/o Gas-Flaring Areas)							(0.023)	(0.030)	(0.030)	(0.029)
Precipitation								-0.000	-0.001	-0.001
(W/o Gas-Flaring Areas)								(0.002)	(0.002)	(0.002)
Wet Day Frequency								0.007	0.006	0.007
(W/o Gas-Flaring Areas)								(0.019)	(0.019)	(0.019)
Frost Day Frequency								0.003	-0.008	-0.009
(W/o Gas-Flaring Areas)								(0.027)	(0.026)	(0.026)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes
Number of Observations	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331
Adjusted R^2	0.501	0.520	0.537	0.546	0.578	0.578	0.583	0.582	0.609	0.619
Expected Heterozygosity at peak	0.657	0.653	0.625	0.627	0.645	0.640	0.638	0.639	0.647	0.642
95% CI Min	0.650	0.643	0.582	0.588	0.612	0.601	0.591	0.591	0.607	0.594
95% CI Max	0.663	0.660	0.648	0.649	0.665	0.663	0.664	0.664	0.670	0.668
Significance of hump-shape	0.000	0.000	0.037	0.024	0.003	0.008	0.020	0.020	0.006	0.017

This table establishes the robustness of the results of the regression analyses in Table 3 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between predicted genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.3c: Predicted Genetic Diversity and Luminosity— W/o Gasflaring and Urban Areas in Outcome but No Excluded Sub-Areas in Controls

	Outcome Variable: Log Luminosity									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted Expected Heterozygosity	583.277***	472.155***	302.166***	333.494***	348.441***	340.541***	300.085***	301.723***	326.718***	296.399***
	(42.995)	(46.058)	(61.509)	(64.273)	(66.333)	(68.304)	(67.627)	(70.114)	(69.687)	(69.797)
Predicted Expected Heterozygosity squared	-443.770***	-361.700***	-242.491***	-266.703***	-270.507***	-266.261***	-235.493***	-236.551***	-252.746***	-230.821***
	(30.696)	(33.153)	(44.142)	(46.175)	(47.379)	(48.555)	(48.098)	(50.212)	(49.838)	(49.814)
Absolute Latitude		0.025***	-0.001	-0.007	-0.005	-0.006	0.004	0.005	0.012	0.011
		(0.004)	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.012)	(0.011)	(0.011)
Terrain Ruggedness Index						0.000	0.000*	0.000*	0.000	0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.027*	0.030	0.042*	0.036
							(0.015)	(0.024)	(0.023)	(0.022)
Diurnal Temperature Range							-0.069***	-0.068**	-0.093***	-0.093***
							(0.023)	(0.030)	(0.030)	(0.029)
Precipitation								-0.000	-0.001	-0.002
								(0.002)	(0.002)	(0.002)
Wet Day Frequency								0.007	0.006	0.007
								(0.019)	(0.019)	(0.019)
Frost Day Frequency								0.000	-0.009	-0.011
								(0.027)	(0.026)	(0.026)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes
Number of Observations	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331
Adjusted R^2	0.505	0.525	0.542	0.552	0.584	0.584	0.589	0.588	0.615	0.626
Expected Heterozygosity at peak	0.657	0.653	0.623	0.625	0.644	0.639	0.637	0.638	0.646	0.642
95% CI Min	0.650	0.643	0.577	0.584	0.609	0.598	0.586	0.587	0.605	0.591
95% CI Max	0.663	0.660	0.648	0.649	0.665	0.663	0.664	0.664	0.670	0.668
Significance of hump-shape	0.000	0.000	0.050	0.033	0.004	0.012	0.027	0.026	0.007	0.021

This table establishes the robustness of the results of the regression analyses in Table 3 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between predicted genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.3d: Predicted Genetic Diversity and Luminosity — W/o Gasflaring and Urban Areas in Outcome and Controls

	Outcome Variable: Log Luminosity									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted Expected Heterozygosity	583.277***	472.155***	302.166***	333.494***	348.441***	340.882***	300.582***	301.594***	326.433***	296.190***
	(42.995)	(46.058)	(61.509)	(64.273)	(66.333)	(68.250)	(67.567)	(70.039)	(69.640)	(69.759)
Predicted Expected Heterozygosity squared	-443.770***	-361.700***	-242.491***	-266.703***	-270.507***	-266.469***	-235.816***	-236.420***	-252.524***	-230.653***
	(30.696)	(33.153)	(44.142)	(46.175)	(47.379)	(48.522)	(48.061)	(50.166)	(49.808)	(49.790)
Absolute Latitude		0.025***	-0.001	-0.007	-0.005	-0.006	0.004	0.005	0.012	0.011
		(0.004)	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.012)	(0.011)	(0.011)
Terrain Ruggedness Index						0.000	0.000*	0.000*	0.000	0.000
(W/o Gas-Flaring or Urban Areas)						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.027*	0.029	0.042*	0.035
(W/o Gas-Flaring or Urban Areas)							(0.015)	(0.024)	(0.023)	(0.022)
Diurnal Temperature Range							-0.069***	-0.068**	-0.092***	-0.093***
(W/o Gas-Flaring or Urban Areas)							(0.023)	(0.030)	(0.030)	(0.029)
Precipitation								-0.000	-0.001	-0.001
(W/o Gas-Flaring or Urban Areas)								(0.002)	(0.002)	(0.002)
Wet Day Frequency								0.006	0.005	0.006
(W/o Gas-Flaring or Urban Areas)								(0.019)	(0.019)	(0.019)
Frost Day Frequency								0.000	-0.009	-0.011
(W/o Gas-Flaring or Urban Areas)								(0.027)	(0.026)	(0.026)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes
Number of Observations	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331	1,331
Adjusted R^2	0.505	0.525	0.542	0.552	0.584	0.584	0.589	0.588	0.615	0.626
Expected Heterozygosity at peak	0.657	0.653	0.623	0.625	0.644	0.640	0.637	0.638	0.646	0.642
95% CI Min	0.650	0.643	0.577	0.584	0.609	0.598	0.586	0.587	0.605	0.591
95% CI Max	0.663	0.660	0.648	0.649	0.665	0.663	0.664	0.664	0.670	0.668
Significance of hump-shape	0.000	0.000	0.050	0.033	0.004	0.012	0.027	0.026	0.007	0.021

This table establishes the robustness of the results of the regression analyses in Table 3 to an alternative exclusion rule of sub-areas in calculating the geographically derived variables. It establishes the significant hump-shaped relationship between predicted genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table A.3e: Predicted Genetic Diversity and Luminosity — Excluding Ethnic Groups Originating in the Territory of the United States

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted Expected Heterozygosity	518.484***	454.271***	278.779***	306.918***	325.604***	322.683***	275.980***	289.657***	336.096***	308.482***
	(45.392)	(47.075)	(66.175)	(70.719)	(72.093)	(73.341)	(72.299)	(73.554)	(73.308)	(73.131)
Predicted Expected Heterozygosity squared	-395.464***	-348.215***	-228.603***	-251.477***	-257.434***	-255.915***	-219.529***	-229.611***	-260.565***	-240.622***
	(32.615)	(33.962)	(47.067)	(50.213)	(50.979)	(51.673)	(50.956)	(52.245)	(52.045)	(51.839)
Absolute Latitude		0.021***	-0.008	-0.014	-0.010	-0.011	0.006	0.005	0.009	0.009
		(0.005)	(0.009)	(0.009)	(0.008)	(0.009)	(0.010)	(0.013)	(0.012)	(0.012)
Terrain Ruggedness Index						0.000	0.000	0.000	0.000	0.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average Temperature							0.036**	0.027	0.046*	0.041*
							(0.017)	(0.026)	(0.025)	(0.025)
Diurnal Temperature Range							-0.085***	-0.087**	-0.100***	-0.100***
							(0.030)	(0.034)	(0.033)	(0.033)
Precipitation								-0.001	-0.002	-0.002
								(0.002)	(0.002)	(0.002)
Wet Day Frequency								0.009	0.012	0.013
								(0.021)	(0.021)	(0.021)
Frost Day Frequency								-0.021	-0.018	-0.016
								(0.036)	(0.035)	(0.034)
World Region FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Soil Quality FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landmass Type FE	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Century in Ethnographic Atlas FE	No	No	No	No	No	No	No	No	Yes	Yes
Mean Size of Local Communities FE	No	No	No	No	No	No	No	No	Yes	Yes
Jurisdictional Hierarchy Beyond Local Community FE	No	No	No	No	No	No	No	No	No	Yes
Type of Class Stratification FE	No	No	No	No	No	No	No	No	No	Yes
Number of Observations	1,129	1,129	1,129	1,129	1,129	1,129	1,129	1,129	1,129	1,129
Adjusted R^2	0.417	0.431	0.456	0.471	0.509	0.508	0.515	0.514	0.543	0.555
Expected Heterozygosity at peak	0.656	0.652	0.610	0.610	0.632	0.630	0.629	0.631	0.645	0.641
95% CI Min	0.647	0.642	0.542	0.545	0.579	0.573	0.554	0.564	0.599	0.586
95% CI Max	0.662	0.660	0.641	0.642	0.659	0.659	0.662	0.663	0.671	0.669
Significance of hump-shape	0.000	0.000	0.193	0.184	0.039	0.054	0.091	0.070	0.013	0.027

This table establishes the robustness of the results of the regression analyses in Table 3 to the exclusion of the United States in the analysis using predicted genetic diversity. It establishes the significant hump-shaped relationship between predicted genetic diversity and the dependent variable, log luminosity, accounting for the potential confounding effects of absolute latitude, regional fixed effects, soil quality fixed effects, and landmass type fixed effects, as well as the scale of the ethnic group (i.e., mean size of local communities fixed effects) and the potential mediating effects of institutional characterizes (jurisdictional hierarchy beyond local community fixed effects, and type of class stratification fixed effects). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.