



Munich Personal RePEc Archive

## **Human Capital and Genetic Diversity**

Sequeira, Tiago and Santos, Marcelo and Ferreira-Lopes,  
Alexandra

Univ. Beira Interior, Univ. Beira Interior, ISCTE-IUL

19 August 2013

Online at <https://mpra.ub.uni-muenchen.de/49135/>  
MPRA Paper No. 49135, posted 21 Aug 2013 19:54 UTC

# Human Capital and Genetic Diversity\*

Tiago Neves Sequeira<sup>†</sup>

*Univ. Beira Interior and CEFAGE-UBI*

Marcelo Santos<sup>‡</sup>

*Univ. Beira Interior and CEFAGE-UBI*

Alexandra Ferreira-Lopes<sup>§</sup>

*ISCTE-IUL, BRU-IUL and CEFAGE-UBI*

## Abstract

The determinants of human capital have been studied sparsely in the literature. Although there is a huge literature on the determinants of schooling linked with the quality of schooling, there are not many contributions that explore the deep determinants of investment in, quantity and quality of human capital. This paper investigates the relationship between human capital and the ancestral genetic diversity of populations. It highlights a strong hump-shaped relationship between genetic diversity and human capital. This means that some of the human capital achievements nowadays may root to the genetic diversity mostly determined many centuries ago. Results are robust to the introduction of several controls, to a consideration of a proxy for human capital in 1500 and to IV estimation.

**Keywords:** human capital; genetic diversity; determinants of development; determinants of human capital.

**JEL Codes:** I25, N10, N30, O10, O50, Z10.

## 1 Introduction

The determinants of human capital have been studied sparsely in the literature. Although there is a huge literature on the determinants of schooling linked with quality of schooling, resulting on a weak causality between inputs to schools and their results, there are not many contributions that explore the deep determinants of investment, quantity and quality of human capital. Exceptions are the significant effect that increased country-risk has in education

---

\*We thank financial support from FCT.

<sup>†</sup>Departamento de Gestão e Economia and CEFAGE-UBI. Universidade da Beira Interior. Estrada do Sineiro. 6200-209 Covilhã, Portugal. Corresponding author. e-mail: sequeira@ubi.pt.

<sup>‡</sup>Departamento de Gestão e Economia and CEFAGE-UBI. Universidade da Beira Interior. Estrada do Sineiro. 6200-209 Covilhã, Portugal.

<sup>§</sup>Instituto Universitário de Lisboa (ISCTE-IUL), ISCTE Business School Economics Department, BRU IUL (Business Research Unit), Lisboa, Portugal, and CEFAGE UBI.

measures, highlighted by Sequeira and Ferraz (2009) and Sequeira (2009). However, the literature on the deep determinants of development has experienced great development in the last decade. Hall and Jones (1999) demonstrated the importance of social infrastructure, a composed measure of law of rule and other institutional measures, in which is followed by Glaeser et. al. (2004). It is interesting to note that this paper highlights the role of human capital as a more significant source of growth than institutions but do not show any causal relationship between institutions and human capital. Rodrick et al. (2004) showed evidence for the supremacy of institutions related to geography and integration as determinants of economic development. In related literature on institutions, Engerman and Sokolo (2000) and Acemoglu, Johnson and Robinson (2005) have stressed the role of colonialism, while the effects of ethnolinguistic fractionalization were examined by Easterly and Levine (1997) and Alesina et al. (2003). Moreover, the historical impact of sociocultural factors has been highlighted by Barro and McCleary (2003), Tabellini (2008), and Guiso et. al. (2009).

Ashraf and Galor (2013) is probably one of the most influential recent papers in the field of economic development.<sup>1</sup> This paper found significant relationship between genetic diversity determined ancestrally and economic development nowadays. The hump-shaped relationship exposed by the authors builds on two opposite effects of genetic diversity. First, an increase in diversity enhances production possibilities due to complementarities between skills or abilities of populations with different genetic roots. In fact, also some competition for survival, as the natural selection explains, also increases adaptability and improving the society capacity of introducing successfully superior technologies. However, after a certain level of genetic diversity, further increases would increase the scope for disarray and mistrust, increasing the probability of conflict. While most of the reasoning due to which there may be a (non-linear) causality between genetic diversity and development are due to accumulation of skills, there is no attempt to explain present education outcomes with genetic diversity. This is the attempt of this paper. In fact, increases in genetic diversity from relatively low levels increase struggle for success which may enhance schooling investment and outcomes; also eventually after some level of genetic diversity, the risk of conflict increases and incentives to accumulate knowledge decrease.<sup>2</sup> Our paper describes exhaustively the relationship between several measures of human capital and genetic diversity and tests its robustness in a cross-section of countries.

In Section 2, we describe the data. In Section 3, we present the main results, subject to a number of extensions. In Section 4 we discuss results on robustness tests linked with the introduction of continent dummies and instrumental variables estimations. Section 5 concludes.

## 2 Data and Sources

In this section we describe the variables and data sources for this work. Our dependent variable is human capital, for which we use different measures - enrolments, attainments, scores in international tests, and also measures of quality-adjusted stocks of human capital

---

<sup>1</sup>The paper was a lead article in American Economic Review and was the Science's Editor Choice in September 2012.

<sup>2</sup>The micro causality between genes and education has also been subject to very recent debate from which, the recent paper from Rietvelt et. al. (2013), in Science, is worth mentioning. The paper presents a statistical significant relationship between genes and educational attainment.

(the product of scores and quantities of human capital). These alternative measures of human capital were taken from Cohen and Soto (2007) - for measures of quantity of human capital and from Hanushek and Woessmann (2012) - for measures of the quality of human capital. We also use a measure of the relational capacity of human capital, using trust as a weight for human capital. Trust were taken from Ashraf and Galor (2013), which used data from the World Values Survey conducted during the period 1981-2008.

Explanatory variables rely on the database from Ashraf and Galor (2013), which include variables that measure genetic diversity, both taken as adjusted to migratory movements and ancestrally adjusted. For our benchmark analysis we use the ancestrally adjusted (to 2000) predicted genetic diversity. As genetic diversity measured by this variable has been adjusted to account for the 2000 composition of populations that can trace their ancestral origins to different source countries in the year 1500, this is the appropriate measure to relate to the distribution of human capital world wide after the year 2000. Other explanatory variables will be introduced as controls later on.

Table 1 summarizes the used dependent variables, which intend to measure quantity and quality of human capital, as well as quality-adjusted measures of years of schooling. The variables `school`, `m_sec15c`, `m_tyr15` and `m_tyr1565` measure quantity of schooling (years of schooling or attainment). `kskh` measure schooling adjusted for social capital. This variable intends to be a proxy of the relational capability of the existing human capital. Next four variables (`cognit`, `lows`, `basic`, `top`) measure quality of schooling, available as tests scores and share of students reaching certain levels of quality in international tests. and the final four variables measure human capital (quantity) weighted by quality (scores). We have also tested other quality-weighted human capital variables, in which we substituted `School` by `m_sec15c`, `m_tyr15` and `m_tyr1565`, alternatively. As results are quite similar to those obtained when using the `School` variable, we choose not to report them. These results are available upon request. Table 2 presents descriptive statistics for the dependent variables. The explanatory variables used - Predicted genetic diversity ancestry adjusted and Mobility index-predicted genetic diversity ancestry adjusted, are `pdiv_aa` e `pdivhmi_aa`, respectively and measure, as explained above, genetic diversity for 2000.

### 3 Results

In this section, we document the fact that there is an hump-shaped relationship between different variables linked with human capital (quantity, quality and quality-adjusted measures) and the genetic diversity of countries. We do this estimating OLS regressions. In the robustness Section we disregard the potential endogeneity problem. We present figures (Figures 1 and 2) and a table (Table 3). As become evident from the figures and the table there is a non-linear relationship (in a hump-shaped form or of inverted-U) between human capital variables and genetic diversity. This is common to the different measures of human capital that we used. Moreover, interestingly this rough regression through different dependent variables predict incredibly similar maximum values for the genetic diversity above which human capital tends to decrease, which is around 0.70, which is slightly below the median value for these variables (see Table 2). This means that there is a quite realistic value for genetic diversity below which human capital increases with diversity and above which it decreases. Moreover this is

Table 1: Human Capital Variables

Variables for Human Capital (HC)	Name	Measure (years and source)
Years of schooling of population+25	School	1960-2010 (Ashraf and Galor, 2013)
Social Capabilities of HC	kskh	Trust*School (Ashraf and Galor, 2013)
% of population + 15 with secondary education	m_sec15c	1960-2010 (Cohen and Soto, 2007)
Years of schooling of population +15, inc. students	m_tyr15	1960-2010 (Cohen and Soto, 2007)
Years of schooling of population 15-64	m_tyr1564	1960-2010 (Cohen and Soto, 2007)
Average test score in math and science	cognitive	primary through secondary, all years 1964-2003 (Hanushek and Woessmann, 2012)
Average test score in math and science	lows	lower secondary, all years 1964-2003 (Hanushek and Woessmann, 2012)
Share of students reaching basic literacy	basic	average test scores in math and science, primary through secondary, all years 1964-2003 (Hanushek and Woessmann, 2012)
Share of top-performing students	top	based on average test scores in math and science, primary to secondary school, all years 1964-2003 (Hanushek and Woessmann, 2012)
Interaction between School and Cognitive	Schoolxcognitive	—
Interaction between School and lows	Schoolxlows	—
Interaction between School and basic	Schoolxbasic	—
Interaction between School and top	Schoolxtop	—

Table 2: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
(a) pdiv_aa	.7267	.0269	.6279	.7743
(b) pdivhmi_aa	.7229	.02904	.6178	.7826
(1) school	4.8623	2.8126	.4089	10.8622
(2) kskh	1.8560	1.4717	.0661	5.7926
(3) m_sec15c	.1528	.1303	.0088	.5103
(4) m_tyr15	5.6471	3.0635	.5183	11.7767
(5) m_tyr1564	5.8586	3.2381	.5417	12.2517
(6) cognitive	4.5429	.5709	3.0893	5.3376
(7) lowsec	4.5369	.6120	2.6830	5.5116
(8) basic	.7569	.2045	.1817	.9738
(9) top	.0592	.05268	.0001	.2043
(10) schoolxcognitive	30.6408	13.7546	7.8473	53.2535
(11) schoolxlowsec	30.6199	13.8527	7.9422	53.5439
(12) schoolxbasic	5.3417	2.7852	.9131	9.9741
(13) schoolxtop	.46392	.4049	.0013	1.3676

Notes: (a) Predicted genetic diversity ancestry adjusted; (b) Mobility index-predicted genetic diversity ancestry adjusted. Ancestry adjustment is made to make variables consistent to time-measurement in 2000. Details are given in the Appendix F of Ashraf and Galor (2013).; (1) Years of schooling; (2) Years of schooling\*Interpersonal Trust; (3) % of population aged 15 or over with complete secondary education; (4) Years of schooling of population 15 and over, whether studying or not; (5) Years of schooling of population 15-64 who is not studying; (6) Average test score in math and science, primary through end of secondary school, all years (scaled to PISA scale divided by 100); (7) Average test score in math and science, only lower secondary, all years (scaled to PISA scale divided by 100); (8) Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school, all years); (9) Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years); (10) Years of schooling\*Average test score in math and science, primary through end of secondary school; (11) Years of schooling\*Average test score in math and science, only lower secondary; (12) Years of schooling\*Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school); (13) Years of schooling\*Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years).

a very similar result to that obtained by Ashraf and Galor (2013: Table 1), when testing the relationship between genetic diversity and development (using the log of population density as dependent variable).

According to our estimates, a 1 percentage point increase in genetic diversity for the country with least genetic diversity in the sample would imply an increase in schooling of around 1.64 years (or 0.83 points in cognitive score) and the same increase for the country with highest genetic diversity roughly implies a reduction in 1.53 year of school (or 0.65 points in cognitive score). These are relevant quantitative effects, representing almost 1/3 of the schooling average value and near 15% of the tests scores average. Moreover, the positive effect of increased genetic diversity for lower levels is relatively higher than the negative effect of increased genetic diversity for higher levels. The inclusion of continent dummies would cause in our case a reduction of the positive effect of genetic diversity. If this would be the case, a 1 percentage point increase for the least diverse would cause an increase of nearly 0.6 years in year of school (or 0.34 in cognitive score) and would cause a decrease of almost 0.2 year of school (or 0.28 in cognitive score) for the most diverse. We will further discuss the effect of introducing continent dummies in the robustness section.

### 3.1 Other Controls

We will now test the robustness of the empirical relationship we described earlier to the introduction of several different controls. We proceed as follows. First, we introduced in a (not-shown) regression for School all the covariates which Ashraf and Galor (2013) tested for income (see their Table 7). From those we selected the statistical significant coefficients (at the 10% level).<sup>3</sup> Selected covariates were social infrastructure and the percentage of population in risk of contracting malaria. These are in fact the most related variables to the expropriation risk of returns from human capital. Firstly, social infrastructure intends to measure wedges between private and social returns (this variable was originally due to Hall and Jones 1999). The higher the wedges, the less incentives to accumulate human capital. Social Infrastructure is measured by an index of government anti-diversion policies (GADP) created from data assembled by a firm that specializes in providing assessments of risk to international investors, Political Risk Services. Specifically it is a composed indicator of law and order, bureaucratic quality, corruption, risk of expropriation and government repudiation of contracts. The second element of Social Infrastructure captures the extent to which a country is open to international trade. This may also be related to human capital investment as it can measure the extent to which human capital is subject to competition from human capital producing good abroad. Moreover, the risk of contracting malaria (originally from Gallup and Sachs, 2001) is directly related to disability and death thus providing a direct effect of expropriation of potential returns from human capital.

In Table 4 we note that despite of these two new covariates being highly statistically significant in distribution of human capital across countries, with a positive effect of Social Infrastructure and a negative effect of the percentage of population in risk of contracting malaria, the hump-shaped relationship between the different measures of human capital and genetic diversity is maintained with high statistical significance for most of human capital

---

<sup>3</sup>Not shown results are available upon request. We adopted this strategy for parsimoniousness. However the inclusion of non-significant variables would not change our main results.

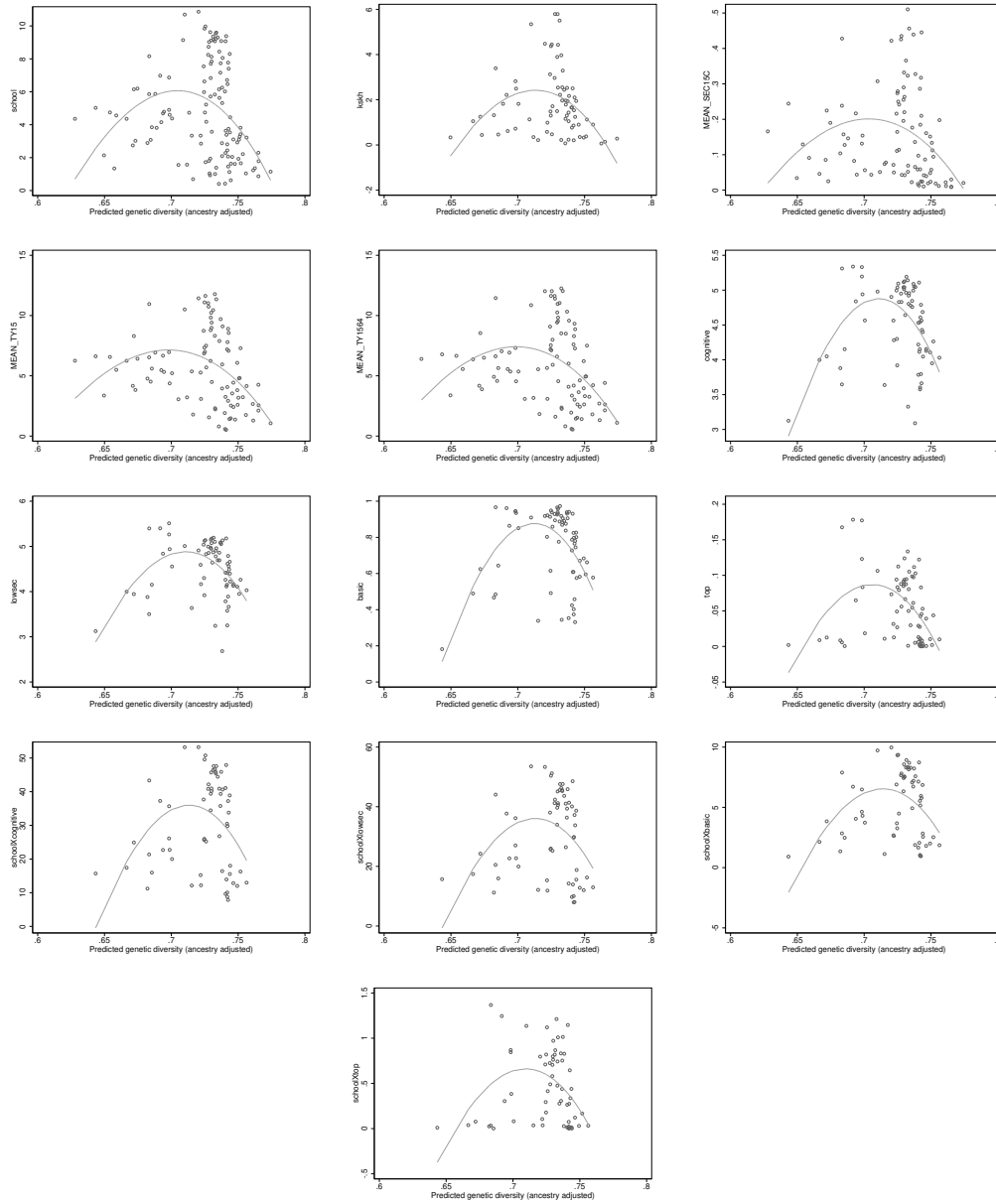


Figure 1: The hump-shaped Relationship between Human Capital and Genetic Diversity

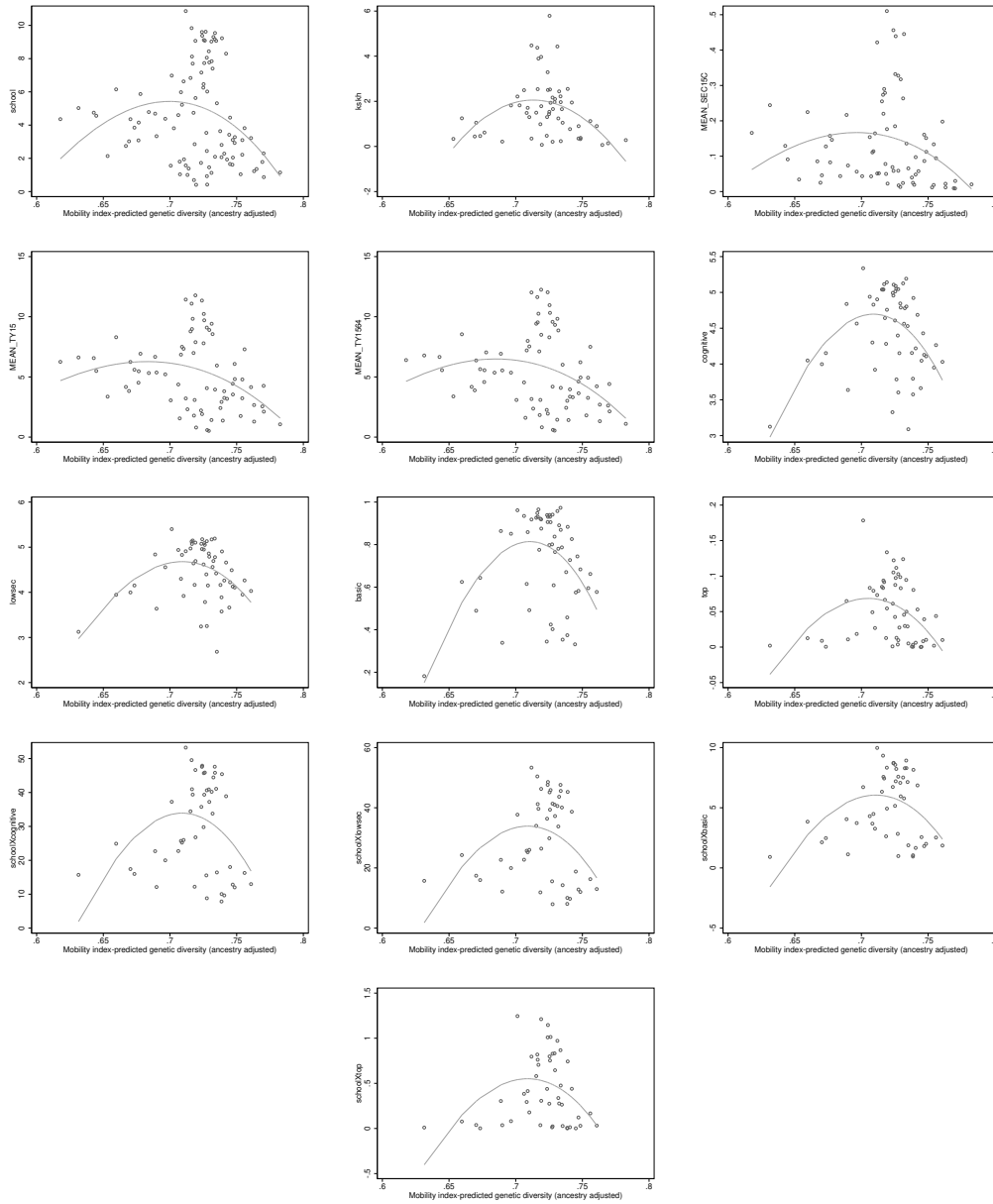


Figure 2: The hump-shaped Relationship between Human Capital and Genetic Diversity (Mobility Adjusted)



Table 3: Restricted Regressions

Dep. Var.	Diversity Var. ( $x$ )	Coef. for $x$ ( $s.e$ )	Coef. for $x^2$ ( $s.e$ )	$R^2$	N
(1) school	pdiv_aa	1390,8*** (329,7)	-989,5*** (231,0)	0.1291	123
(2) kskh	pdiv_aa	1128,9*** (236,3)	-793,0*** (165,7)	0.1770	73
(3) m_sec15c	pdiv_aa	47,9*** (15,4)	-34,2*** (10,8)	0.0943	94
(4) m_tyr15	pdiv_aa	1265,3*** (333,0)	-909,2*** (233,7)	0.1525	94
(5) m_tyr1564	pdiv_aa	1350,9*** (356,6)	-969,4*** (250,2)	0.1482	94
(6) cognitive	pdiv_aa	654,2*** (90,7)	-460,9*** (64,3)	0.2747	71
(7) lowsec	pdiv_aa	667,8*** (96,7)	-470,7*** (68,6)	0.2476	71
(8) basic	pdiv_aa	243,1*** (33,4)	-170,9*** (23,7)	0.2933	71
(9) top	pdiv_aa	47,5*** (10,6)	-33,7*** (7,5)	0.2235	71
(10) schoolxcognitive	pdiv_aa	10805,5*** (3882,7)	-7585,1*** (2735,6)	0.1436	63
(11) schoolxlowsec	pdiv_aa	10938,9*** (3909,9)	-7680,8*** (2754,4)	0.1442	63
(12) schoolxbasic	pdiv_aa	2508,7*** (733,9)	-1759,0*** (517,1)	0.1950	63
(13) schoolxtop	pdiv_aa	350,5*** (97,9)	-247,4*** (69,1)	0.1620	63
(1) school	pdivhmi_aa	790,1*** (216,4)	-566,8*** (152,3)	0.0874	96
(2) kskh	pdivhmi_aa	820,6*** (163,6)	-575,0*** (114,7)	0.2163	58
(3) m_sec15c	pdivhmi_aa	25,7** (10,9)	-18,5** (7,7)	0.0695	73
(4) m_tyr15	pdivhmi_aa	593,7*** (222,0)	-436,1*** (156,8)	0.1141	73
(5) m_tyr1564	pdivhmi_aa	649,9*** (237,7)	-475,9*** (167,8)	0.1106	73
(6) cognitive	pdivhmi_aa	434,8*** (61,4)	-307,4*** (44,1)	0.2617	57
(7) lowsec	pdivhmi_aa	430,1*** (63,9)	-303,9*** (45,9)	0.2263	57
(8) basic	pdivhmi_aa	160,7*** (22,6)	-113,3*** (16,3)	0.2669	57
(9) top	pdivhmi_aa	29,8*** (7,6)	-21,2*** (5,4)	0.2074	57
(10) schoolxcognitive	pdivhmi_aa	7721,6*** (2543,8)	-5456,0*** (1802,8)	0.1512	49
(11) schoolxlowsec	pdivhmi_aa	7762,2*** (2558,7)	-5485,1*** (1812,8)	0.1508	49
(12) schoolxbasic	pdivhmi_aa	1776,2*** (478,6)	-1251,9*** (339,5)	0.2035	49
(13) schoolxtop	pdivhmi_aa	233,7*** (73,3)	-165,2*** (51,7)	0.1784	49

Note: Dependent Variables - (1) Years of schooling; (2) Years of schooling\*Interpersonal Trust; (3) % of population aged 15 or over with complete secondary education; (4) Years of schooling of population 15 and over, whether studying or not; (5) Years of schooling of population 15-64 who is not studying; (6) Average test score in math and science, primary through end of secondary school, all years (scaled to PISA scale divided by 100); (7) Average test score in math and science, only lower secondary, all years (scaled to PISA scale divided by 100); (8) Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school, all years); (9) Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years); (10) Years of schooling\*Average test score in math and science, primary through end of secondary school; (11) Years of schooling\*Average test score in math and science, only lower secondary; (12) Years of schooling\*Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school); (13) Years of schooling\*Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years).

\* means significant at 10% level, \*\* means significant at 5% level and \*\*\* means significant at 1% level. Heteroscedasticity robust standard-errors are reported in parenthesis. pdiv\_aa means Predicted genetic diversity ancestry adjusted and pdivhmi\_aa means Mobility index-predicted genetic diversity ancestry adjusted.

Table 4: Human Capital and Genetic Diversity (other controls)

	Predicted genetic diversity	Predicted genetic diversity square	Social infrastructure	% of population at risk of contracting malaria	$R^2$ /Observations
(1)	364.64** (166.84)	-252.45** (119.16)	5.34*** (0.83)	-2.6*** (0.55)	0.65 100
(2)	316.29* (161.02)	-217.81* (114.55)	3.64*** (0.61)	-0.69** (0.27)	0.54 61
(3)	6.34 (10.69)	-4.06 (7.61)	0.25*** (0.05)	-0.1*** (0.03)	0.51 91
(4)	189.2 (189.09)	-129.92 (135.25)	6.65*** (0.85)	-2.71*** (0.58)	0.71 91
(5)	197.30 (196.92)	-134.29 (140.84)	7.24*** (0.89)	-2.81*** (0.61)	0.72 91
(6)	361.10*** (104.54)	-253.80*** (74.14)	1.17*** (0.28)	-0.41 (0.28)	0.55 57
(7)	371.53*** (108.75)	-261.49*** (77.13)	1.19*** (0.30)	-0.54 (0.35)	0.52 57
(8)	147.11*** (44.96)	-103.07*** (31.94)	0.39*** (0.11)	-0.09 (0.11)	0.51 57
(9)	24.93*** (7.49)	-17.71*** (5.36)	0.08*** (0.02)	-0.05** (0.02)	0.48 57
(10)	4991.90** (2288.12)	-3509.42** (1630.76)	25.90*** (5.88)	-19.98*** (6.58)	0.52 53
(11)	5113.63** (2298.49)	-3597.50** (1637.63)	26.05*** (5.92)	-20.62*** (6.53)	0.53 53
(12)	1256.36*** (440.16)	-879.96*** (313.88)	5.64*** (1.14)	-3.57*** (1.19)	0.58 53
(13)	192.82*** (66.63)	-136.49*** (47.54)	0.72*** (0.18)	-0.55** (0.22)	0.49 53

Note: Dependent Variables - (1) Years of schooling; (2) Years of schooling\*Interpersonal Trust; (3) % of population aged 15 or over with complete secondary education; (4) Years of schooling of population 15 and over, whether studying or not; (5) Years of schooling of population 15-64 who is not studying; (6) Average test score in math and science, primary through end of secondary school, all years (scaled to PISA scale divided by 100); (7) Average test score in math and science, only lower secondary, all years (scaled to PISA scale divided by 100); (8) Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school, all years); (9) Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years); (10) Years of schooling\*Average test score in math and science, primary through end of secondary school; (11) Years of schooling\*Average test score in math and science, only lower secondary; (12) Years of schooling\*Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school); (13) Years of schooling\*Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years).

Level of significance: \*\*\* for p-value<0.01; \*\*for p-value<0.05; \* for p-value<0.1.  
Values between parentheses are standard errors.

variables, with exceptions being % of population with secondary school (`m_sec15c`) and years of schooling in population above 15 years old (`m_tyr15` and `m_tyr1564`). With `m_sec15c` there is a marginally significant linear effect (when `pdiv_aa` squared is dropped). Results that use the alternative measure of genetic diversity (predicted through the human mobility index - `pdivhmi_aa`) are presented in Table A.1 in the Appendix and conclusions are in line with the previous ones. It is interesting to note that the quantitative effect of genetic diversity is now smaller than that emerged from the restricted regressions. In fact, an 1% point increase in the lowest genetic diversity would imply 0.52 years (nearly 6 months) more in schooling (and 0.46 additional points in school variable) and would imply nearly 3.5 months less schooling (and less 0.35 points in ‘cognitive’ score) if the country departs from the highest levels of genetic diversity. The higher effects of increased diversity for lower levels than for higher levels is now slightly increased.

### 3.2 A Proxy for Human Capital in 1500

Ashraf and Galor (2013) focus their paper in explaining development in 1500 due to genetic diversity. Thus their measure of genetic diversity in 1500 restricts attention to the development outcomes in the pre-colonial era when, arguably, regional populations were indigenous to their current geographic location. Specifically, in light of the serial founder effect, the presence of multiple indigenous ethnicities in a given region would have had a negligible impact on the diversity of the regional population during this period (Ashraf and Galor, 2013: 4). In face of our objective of explaining the relationship between human capital and genetic diversity, it would be interesting to see if the hump-shaped relationship we have described above between current levels of human capital and predicted genetic diversity for 2000, would also be verified for 1500. To this end, our main challenge would be to find variables (with extensive coverage around the world) to measure human capital. As Allen (2003) and Reis (2002) show there is not a direct measure for human capital in 1500 and some rough proxies such as wages could be obtained only for a few European countries, as in Allen (2003). Literacy rate (the ability to sign) could be collected by Reis (2002) for 12 European countries but only around 1800. As this author argued “For earlier periods, the data is patchier and less reliable and therefore the long run trends are hard to identify except for in a few countries.” (p.18). Also, most of the schools and education in the Middle-ages based exclusively on religious contents and even literacy (capacity to read) come from the need to learn how to read the Bible and other religious books (see e.g. Houston, 2002). So literacy or schooling measures to proxy human capital in 1500 must be excluded due to lack of data with extensive coverage. Moreover, if it could be collected for few countries, its measure would hardly be matched with nowadays concept of human capital. Firstly, it excludes knowledge directly linked with productive skills, due to apprenticeship, training and experience. Second, it is very directed to religious thought.

Thus, we thought that alternative available proxies may be linked with human capital. As Aiyar et. al. (2008) puts it “the distinction between technology and human capital was much less clear cut in pre-industrial times than it is today. In modern societies it is feasible and often useful to treat the two as conceptually different. Human capital is inherently linked to the human body, whereas technological knowledge blueprints can be stored outside humans, in computers and books. But before the advent of the printing press and widespread literacy, technological knowledge would have to be embodied primarily in humans, and actively

transmitted across generations in order to be preserved in society.”(p. 127). So if we use technology measures in 1500, we may have a good proxy for embodied knowledge, meaning a close definition to what we consider nowadays human capital. Fortunately, Comin et. al. (2010) has come with a database on technologies on different fields that could be used for our purposes. The database covers the technology usage (or not) in each country (in 1500) for agriculture, transportation, communication, industry, military and an average sector (the average of previous sectoral indexes).

We relate each of the variables for 1500 (as dependent variables) with genetic diversity, using as controls all the variables used by Ashraf and Galor (2013) in their benchmark regressions. Tables 5 and A.2 show the results. All regressions present an impressive robust relationship between genetic diversity and our proxies for human capital. In particular it also draws the so-called hump-shaped relationship between genetic diversity and human capital. The cutting point above which the marginal effect of increases in genetic diversity begins to be negative is around 0.7, exactly the same as above, with the 2000 data for human capital. An increase in 1 percentage point (0.01) in genetic diversity for countries with the lowest genetic diversity would imply an increase in 0.050 score for agriculture, 0.055 for communications, 0.031 for transportation, 0.083 for industry, 0.038 for military and 0.052 for the average sector. This represents sizeable effects given that the human capital proxies are measured between 0 and 1, corresponding to values between 3.1% (transportation) to 8.3% (industry). For those countries with the highest genetic diversity, a further increase in 0.01 in diversity would cause decreases in human capital from 0.02 (2.3%) in transportation to 0.09 (9.1%) in agriculture. There is evidence of a relatively higher positive effect of diversity on human capital for all sectors, except for agriculture, where the negative effect of increasing genetic diversity is higher for high levels of diversity than the positive effect that occurs in low levels of genetic diversity. The only exceptions for the hump-shaped robust relationship are for transportation and military associated knowledge when genetic diversity is adjusted with the human capital mobility index (Table A.2). However, in both that cases, it can be shown that there is a robust linear relationship between genetic diversity and human capital, meaning that an increase in 1% in genetic diversity (for all levels of diversity) would imply increases in human capital linked with transportation in 1.4% and with military in 1.8%. Most significant variables in determining the distribution of these proxies of human capital around the world, apart from genetic diversity, are the Neolithic transition time and absolute latitude.

Table 5: Regressions for human capital proxies in 1500

	Predicted genetic diversity	Predicted genetic diversity square	Log [Neolithic transition timing]	Log [percentage of arable land]	Log [absolute latitude]	Log [land suitability for agriculture]	$R^2$ / Observations
(1)	59.80*** (16.64)	-44.16*** (12.20)	0.29*** (0.04)	0.06** (0.03)	-0.002 (0.02)	-0.004 (0.03)	0.57 106
(2)	47.10*** (10.67)	-33.51*** (7.98)	0.27*** (0.04)	0.05 (0.03)	0.15*** (0.02)	0.03 (0.03)	0.73 106
(3)	23.94*** (7.25)	-16.84*** (5.44)	0.23*** (0.03)	0.04** (0.02)	0.07*** (0.02)	-0.024 (0.02)	0.68 106
(4)	30.74** (13.80)	-21.63** (10.33)	0.29*** (0.04)	0.05** (0.02)	0.11*** (0.03)	-0.02 (0.03)	0.62 106
(5)	57.28*** (14.10)	-39.43*** (10.45)	0.27*** (0.03)	0.03 (0.03)	0.06*** (0.02)	-0.02 (0.03)	0.76 106
(6)	43.77*** (8.18)	-31.11*** (6.09)	0.27*** (0.02)	0.04** (0.02)	0.08*** (0.02)	-0.01 (0.02)	0.80 106

Note: Dependent Variables - (1) Average human capital in agriculture in period 1500 CE; (2) Average human capital in communications in period 1500 CE; (3) Average human capital in transportation in period 1500 CE; (4) Average human capital in military in period 1500 CE; (5) Average human capital in industry in period 1500 CE; (6) Average of the sectoral human capital indexes in period 1500 CE; Level of significance: \*\*\* for p-value<0.01; \*\*for p-value<0.05; \* for p-value<0.1. Values between parentheses are standard errors.

## 4 Robustness

One of the tests provided on Ashraf and Galor (2013) was the inclusion of continent dummies. These continent dummies can account for other determinants of human capital other than genetic diversity and the other controls introduced in previous regressions, which may be continent-specific. Broadly speaking, we still obtain statistically significant hump-shaped effects of genetic diversity and human capital both for 2000 and 1500.<sup>4</sup> For 2000, we followed the strategy described above to select a meaningful specification. Departing from an initial regression for school, we tested the significance of all the covariates also used by Ashraf and Galor (2013), introducing continent dummies into the regression. Then we selected variables with significant coefficients (at the 10% level), maintaining continent dummies. One of the significant variables is now different from the selected above: The percentage of population in risk of contracting malaria is now replaced by the percentage of population living in tropical zones. Almost all the human capital variables maintain the hump-shaped robust relationship with genetic diversity. In fact, when compared with the conclusions drawn from the results in Table 4, the differences are the following: the years of education for population above 15 variables are now linearly related to genetic diversity, with 5% and 10% statistical significance, meaning a small effect of nearly 2 months of schooling due to an increase of 1% in genetic diversity. Moreover, the hump-shaped robust (and also the linear one) relationship with the some of the adjusted human capital measures (for test scores - cognitive and lowsec - and for social capital, kskh) disappears with the introduction of continent dummies. The results with the human mobility index adjustment for genetic diversity confirm all the results described above but recovers the highly significant (at 5%) hump-shaped relationship between genetic diversity and the social capital adjusted measure of human capital. Generally, a 1% increase in the lowest level of diversity implies an increase in schooling of nearly 7 months (and of 0.34 in test scores) and a 1% diversity increase in the highest level would imply a decrease of 2.6 months in schooling (0.28 in test scores).

To further address robustness of our main result, we may argue that there is a causal relationship between genetic diversity and human capital in the post-1500 era. However, genetic diversity may be also an endogenous outcome of geographic areas with more human capital, as genetic diversity could have been improved through migration from less knowledge-endowed areas to more knowledge-endowed areas (specific examples of those migrations were the barbaric invasions of the Roman Empire or the colonization of Africa). On the contrary, well-endowed areas were able also to construct barriers to deter colonization (e.g. the China's Great Wall or Roman Legions for centuries). The question is if these types of migrations were in fact increasing genetic diversity, so as genetic diversity and human capital may be outcomes of some other determinants of development and thus genetic diversity could not be regarded as exogenous towards human capital nowadays. To minimize the endogeneity problem Ashraf and Galor (2013) constructed a measure of predicted genetic diversity based on "physical" distance from East Africa, which was the variable we have been using in the paper. As the authors put it "given the obvious exogeneity of migratory distance from East Africa with respect to development outcomes in the Common Era, the use of migratory distance to project genetic diversity alleviates concerns regarding the potential endogeneity between observed genetic diversity and economic development (...). Specifically, the identifying assumption being

---

<sup>4</sup>Results are available upon request.

employed here is that distances along prehistoric human migration routes from Africa have no direct effect on economic development during the Common Era.” Ashraf and Galor (2013: 6, 14).

In what follows, following what was done by Ashraf and Galor (2013) to test exogeneity towards development (the log of population density), we tested the exogeneity of *observed* genetic diversity in explaining human capital around the world. Firstly, we have regressed the average sectoral human capital index in 1500 (tr3) on observed genetic diversity (linear and square terms) migratory distances from East Africa (linear and square terms) together with Log Neolithic transition time, Log percentage of arable land, Log absolute latitude, Log land suitability for agriculture and continental dummies. The linear and squared terms of observed genetic diversity are significant at 5% and both terms on distance are non-significant. We repeat the procedure substituting distances with the mobility index (linear and squared terms). This means that when considering genetic diversity and distances or migrations as potentially simultaneous determinants of human capital development, we end out concluding that genetic diversity is always statistically stronger as a candidate to explain human capital. Secondly, we replicated 2SLS regression the authors have presented in their Table 2, column (6) but now with the average sectoral human capital index in 1500 (tr3) we have used earlier, as the dependent variable. This serves as a test for exogeneity of of *observed* genetic diversity for the smaller sample. Given that the *observed* genetic diversity could be established as exogenous, then the predicted genetic diversity would be reasonably considered as exogenous by the arguments exposed earlier. This will be column (1) in our robustness Table 6.

However, we want to document further that the hump-shaped relationship between human capital and genetic diversity (both in 2000 and 1500) can be regarded as a causal relationship. Thus, we used the geographical aerial distances to East Africa, terrestrial distances to London and Mexico as instruments to the predicted genetic diversity for the larger sample. In fact, geographical distances to given points in the Word, despite being highly correlated with predicted genetic diversity (remember that terrestrial distances to East Africa were used to construct predicted genetic diversity, departing from observed genetic diversity), they are not reasonably assumed to be, by themselves, direct determinants of human capital investment many centuries after prehistoric migration routes. Thus geographical distances are good candidates to instrument the predicted genetic diversity. Columns (2) and (3) present 2SLS regressions for the average sectoral human capital index in 1500 (tr3) and columns (4) and (5) present 2SLS regressions for schooling years (School) and test scores (Cognitive). Note that while in columns for the 1500 dependent variables we introduce a set of covariates that are arguably exogenous, in the 2000 regressions, we avoided the introduction of more controls, because its own exogeneity might be at stake (e.g. the social infrastructure measure).

There are two main issues with validity of IV estimates, one is the possibility of weak instruments, this is, instruments that are not sufficiently correlated with the instrumented variables, the other is adequate instruments, e.g., their potential correlation with the error term. We accounted for both problems in our estimations. Firstly, we have carefully analysed all the first-step regressions for highly significance of regressors. Secondly, we have analysed the result of the Kleibergen-Paap rk LM statistic for under-identification (insufficient instruments), the Cragg-Donald Wald F statistic for weak-identification (weak instruments). We also used the Stock-Wright LM S statistic for the joint significance of endogenous variables in the main regression, which is a test of robust inference, robust to weak instruments. If this test is

rejected this means that, instruments can be used to explain the dependent variable, not only through the instrumented variable. In our case this would mean that distance could be used to explain human capital (see the discussion above about the exclusion of migratory distance as a predictor of human capital development several centuries later). Thirdly, we have analysed the Hansen J-statistic to test for adequate instruments. For a good IV regression, all the tests but the the Stock-Wright LM S statistic and J-statistic should reject. Fourthly, we have analysed an endogeneity test, which indicates whether we can treat genetic diversity as exogenous in the context of the regressed equations. Column (1) show us a regression for the restricted sample (which were not used until now). It used the *observed* genetic diversity (which were used by Ashraf and Galor (2013) to predict genetic diversity in the enlarged country sample). We follow strictly the instrumentation technique followed by the author. We reach a statistically significant hump-shaped influence of genetic diversity into human capital in which all the other predictors of human capital are also statistically significant. The tests indicate that we can reject the hypothesis that instruments are weak.

Columns (2) and (3) present regressions for the enlarged sample in which we used *predicted* genetic diversity to predict human capital in 1500. The results reveal the same robust relationship, the tests indicate that instruments are not weak and are not correlated to the error term. Thus, inference can be done. However, contrary to what happen in column (1) the endogeneity test indicates now that statistically *predicted* genetic diversity cannot be treated as exogenous. Columns (4) and (5) presents examples of regressions for the enlarged sample in which we used *predicted* genetic diversity to predict human capital in 2000. The results reveal the same robust relationship, the tests indicate that instruments are not weak and are not correlated to the error term, which clearly indicates that inference can be done. In fact, in these cases, the endogeneity test indicates that the predicted genetic diversity could indeed have been treated as exogenous. It is also worth noting that quantitatively the coefficients that define the hump-shaped relationship between human capital and genetic diversity are not sizeably different from the coefficients estimated by OLS earlier in the paper, being the greater quantitative difference those of coefficients in Columns (2) and (3). It is striking that the threshold level above which genetic diversity causes human capital to diminish is almost similar as before, 0.7!

We have experimented several specifications for the different human capital variables we have studied above (both for 1500 and 2000) and from all experiments it was possible to demonstrate that genetic diversity can be regarded as *causing* human capital rises for low levels of diversity and as *causing* human capital drops for high values of genetic diversity.<sup>5</sup>

---

<sup>5</sup>We ran several regressions for 1500 and 2000. For 1500, despite in the regression for the average sectoral human capital the predicted diversity does not appear to be statistically exogenous, for the human capital associated with agriculture, communications and industry, the endogeneity test does not reject. For these human capital proxies IV regression are also well-behaved regarding the instruments properties and the hump-shaped relationship is revealed. For 2000, almost all the results suggest that predicted genetic diversity (ancestrally adjusted) can be treated as exogenous within regressions that use the same instrument set as in columns (4) and (5) in Table 6 - exceptions are only for years of schooling for population above 15 years old and for the share of students with top scores in tests (m\_tyr15, m\_tyr1564 and top). Despite that, even for the exceptions, it is possible to find instrument sets that allow for the endogeneity test to fail rejection (accepting that instrumented variables can be treated as exogenous). Moreover, most variables linked with human capital in 2000 are robustly related to the predicted genetic diversity (ancestrally adjusted) - in the typical hump-shaped manner in regressions specified as those in Tables (4) and (5). More important than that, for those variables that could not be statistically treated as exogenous (the three mentioned above), it is possible to specify well-behaved IV



Table 6: Human Capital and Genetic Diversity (2SLS estimates)

Dependent Var.	(1) Average HC 1500	(2) Average HC 1500	(3) Average HC 1500	(4) School	(5) Cognitive ×School
Observed Diversity	42.1** (0.048)	–	–	–	–
Observed Diversity Square	-34.8* (0.051)	–	–	–	–
Predicted Diversity	–	29.0*** (0.001)	28.8*** (0.001)	873.0** (0.014)	8008** (0.041)
Predicted Diversity Square	–	-21.0*** (0.001)	-20.9*** (0.001)	-614.6** (0.013)	-5674** (0.041)
Log Neolithic Transition Time	0.22*** (0.004)	0.22*** (0.000)	0.22*** (0.000)	–	–
Log Percentage Arable Land	0.14*** (0.003)	0.05*** (0.006)	0.05*** (0.006)	–	–
Log Absolute Latitude	0.06*** (0.003)	0.03* (0.056)	0.01* (0.056)	–	–
Log Land Suit. for Agriculture	0.06* (0.051)	0.03* (0.079)	0.02* (0.076)	–	–
Kleibergen-Paap rk LM statistic	5.7** (0.017)	41.2*** (0.000)	41.7*** (0.000)	31.6*** (0.000)	14.7** (0.041)
Cragg-Donald Wald F statistic	4.6 <sup>Δ</sup>	446.8 <sup>†††</sup> (11.04)	353.7 <sup>†††</sup> (13.97)	8.9 <sup>†</sup> (6.20)	9.7 <sup>†</sup> (6.20)
Stock-Wright LM S statistic	2.8 (0.242)	8.8* (0.066)	8.9 (0.112)	12.9 (0.116)	7.8 (0.453)
Hansen J-Statistic	exact. id.	1.57 (0.457)	1.59 (0.662)	6.91 (0.330)	7.25 (0.299)
Endog. Test	3.16 (0.206)	10.68 (0.005)	16.88 (0.000)	2.49 (0.288)	3.40 (0.183)
N	19	106	106	123	63

Note: Excluded Instruments - (1) 2 instruments: First-stage fitted values of observed genetic diversity (square) and migratory distance from East Africa, equation is exactly identified; (2) 4 instruments: aerial distance from East Africa, aerial distance from East Africa (square), terrestrial distance from London and terrestrial distance from London (square); (3) 5 instruments: the same as in column (3) plus Geodesic Centroid Latitude; (4) and (5): 8 instruments: aerial distance from East Africa, aerial distance from East Africa (square), terrestrial distance from London, Tokyo and Mexico, terrestrial distance from London, Tokyo and Mexico (square). Kleibergen-Paap rk LM statistic tests underidentification, under the null of the matrix of reduced form coefficients has rank=K1-1. Cragg-Donald Wald F statistic tests the null under which equation is weakly identified. This is compared with the Stock-Yogo weak ID test critical values, which are reported in parenthesis in that line. Stock-Wright LM S statistic tests the null under which the joint endogenous regressors have null coefficients. Hansen J-Statistic tests the null under which the instruments are valid, i.e., uncorrelated with the error term.

Level of significance: \*\*\* for p-value<0.01; \*\*for p-value<0.05; \* for p-value<0.1. <sup>Δ</sup> in column (1) 4.58 is the critical value for 15% critical IV size; <sup>†††</sup> in columns (2) and (3), 11.04 is the critical value for relative IV bias of 5% (of the OLS bias); <sup>†</sup> in column (4) and (5), 6.20 is the critical value for relative IV bias of 20% (of the OLS bias).

Values inside parentheses are standard errors, except for the Cragg-Donald Wald F statistic. All regressions include continent dummies.

## 5 Conclusion

We build on Ashraf and Galor (2013) to determine the influence of genetic diversity in one of the most important sources of development: human capital. Thus we constructed a database of human capital variables coming from Cohen and Soto (2007) - for measures of quantity of human capital and from Hanushek and Woessmann, 2012 - for measures of quality of human capital and then merged it with the database of genetic diversity, from Ashraf and Galor (2013). We found a hump-shaped relationship between human capital and genetic diversity. A 1% change in low levels of genetic diversity may imply large effects in schooling that can oscillate between 4 to 13<sup>6</sup> months of schooling (0.34 to 0.46 points in international tests scores) and negative effects when there is high genetic diversity (2 to 10 months of schooling and near 0.3 points in scores).

We have also tested proxies for human capital in 1500. Given the difficulty in obtaining reliable data about literacy or schooling in that year, we argue that the measures of technology adoption from Comin et. al. (2010) are indeed good proxies for human capital given the embodied nature of knowledge in these days. In fact, we found again a statistically robust hump-shaped relationship between human capital in 1500 and genetic diversity. Quantitative effects are also relevant: a 1% increase in low levels of diversity can account for increases in 2.8% to 5.2% in average human capital in 1500 and a similar increase in high levels of diversity account for reductions in human capital that range from near 1.6% to 4.8%. This dismisses the Comin et. al. (2010) conclusion according to technology adoption (our proxy for embodied knowledge or human capital in 1500) is not related to population or genetic characteristics, as we found evidence of a strong relationship between these measures and genetic diversity.

The main measures used to measure genetic diversity were calculated using observed genetic diversity and migratory distances from East Africa. As it is not reasonable to assume that migratory distances from East Africa would have an effect in the after-1500 human capital outcomes, that variable can be regarded as exogenous. Thus given the exogeneity of observed genetic diversity, the exogeneity of predicted genetic diversity (the variable we use in our main regression) follows. Thus, we tested the exogeneity of observed genetic diversity to a regression for 1500 human capital and we show and discuss a number of robustness tests with instrumental variables regressions. The overall conclusion is that the hump-shaped relationship between human capital and genetic diversity (both in 2000 and 1500) can indeed be regarded as a causal relationship. Thus post-1500 human capital outcomes may have been determined centuries ago, when great human migrations shaped the nowadays countries genetic diversity.

---

models that highlight the robust hump-shaped relationship with predicted genetic diversity.

<sup>6</sup>Using the most preferable specifications (Tables 4 and 6 and regressions with controls and continent dummies).

## References

- [1] Acemoglu, D., S. Johnson, and J. A. Robinson (2005). Institutions as a Fundamental Cause of Long-Run Growth. *Handbook of Economic Growth*, **Vol IA.** , ed. Philippe Aghion and Steven N. Durlauf. Amsterdam, The Netherlands:Elsevier North-Holland.
- [2] Allen, R. (2003). Progress and Poverty in Early Modern Europe. *The Economic History Review New Series*, **56(3)**: 403-443
- [3] Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth*, **8(2)**: 155-194.
- [4] Ayiar, S, C-J. Daalgard, and O. Moav (2008). Technological Progress and Regress in Pre-Industrial Times. *Journal of Economic Growth*, **13**: 125-144.
- [5] Ashraf, Q., and O. Galor (2013). The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *American Economic Review*, **103(1)**: 1-46.
- [6] Barro, R. J., and R. M. McCleary (2003). Religion and Economic Growth across Countries. *American Sociological Review*, **68(5)**: 760-781.
- [7] Cohen, D. and M. Soto (2007). Growth and human capital: Good data, good results. *Journal of Economic Growth*, **12**: 51-76.
- [8] Comin, D., W. Easterly, and E. Gong (2010). Was the Wealth of Nations Determined in 1000 BC? *American Economic Journal: Macroeconomics*, **2(3)**: 65-97.
- [9] Easterly, W., and R. Levine (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics*, **112(4)**: 1203-1250.
- [10] Engerman, S. L., and K. L. Sokolo (2000). History Lessons: Institutions, Factor Endowments, and Paths of Development in the NewWorld. *Journal of Economic Perspectives*, **14(3)**: 217-232.
- [11] Glaeser, E. L., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer (2004). Do Institutions Cause Growth? *Journal of Economic Growth*, **9(3)**: 271-303.
- [12] Hall, R. and C. I. Jones (1999). Why Do Some Countries Produce So Much More Output Per Worker Than Others? *Quarterly Journal of Economics*, **114(1)**: 83-116.
- [13] Hanushek, E. and L. Woessmann (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, **17(4)**: 267-321
- [14] Houston, R. (2002). Literacy in Early Modern Europe - Culture and Education 1500-1800, Second Edition. Pearson Education.
- [15] Gallup, J. and J. Sachs (2001). The Economic Burden of Malaria *American Journal of Tropical Medicine and Hygiene*, **64(1-2)**: 85-96.

- [16] Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural Biases in Economic Exchange? *Quarterly Journal of Economics*, **124(3)**: 1095-1131.
- [17] Reis, J. (2002). Human capital, immaterial goods, and the standard of living in pre-industrial Europe. Working-Paper 1/2002. Instituto de Ciencias Sociais. Universidade de Lisboa.
- [18] Rodrik, D., A. Subramanian, and F. Trebbi (2004). Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth*, **9(2)**: 131-165.
- [19] Cornelius A. Rietveld, Sarah E. Medland, Jaime Derringer, Jian Yang, Tnu Esko, Nicolas W. Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal, Eva Albrecht, Behrooz Z. Alizadeh, Najaf Amin, John Barnard, Sebastian E. Baumeister, Kelly S. Benke, Lawrence F. Bielak, Jeffrey A. Boatman, Patricia A. Boyle, Gail Davies, Christiaan de Leeuw, Niina Eklund, Daniel S. Evans, Rudolf Ferhmann, Krista Fischer, Christian Gieger, Hkon K. Gjessing, Sara Hgg, Jennifer R. Harris, Caroline Hayward, Christina Holzapfel, Carla A. Ibrahim-Verbaas, Erik Ingelsson, Bo Jacobsson, Peter K. Joshi, Astanand Jugessur, Marika Kaakinen, Stavroula Kanoni, Juha Karjalainen, Ivana Kolcic, Kati Kristiansson, Zoltn Kutalik, Jari Lahti, Sang H. Lee, Peng Lin, Penelope A. Lind, Yongmei Liu, Kurt Lohman, Marisa Loitfelder, George McMahon, Pedro Marques Vidal, Osorio Meirelles, Lili Milani, Ronny Myhre, Marja-Liisa Nuotio, Christopher J. Oldmeadow, Katja E. Petrovic, Wouter J. Peyrot, Ozren Polaek, Lydia Quaye, Eva Reinmaa, John P. Rice, Thais S. Rizzi, Helena Schmidt, Reinhold Schmidt, Albert V. Smith, Jennifer A. Smith, Toshiko Tanaka, Antonio Terracciano, Matthijs J. H. M. van der Loos, Veronique Vitart, Henry Vlzke, Jrgen Wellmann, Lei Yu, Wei Zhao, Jri Allik, John R. Attia, Stefania Bandinelli, Franois Bastardot, Jonathan Beauchamp, David A. Bennett, Klaus Berger, Laura J. Bierut, Dorret I. Boomsma, Ute Bltman, Harry Campbell, Christopher F. Chabris, Lynn Cherkas, Mina K. Chung, Francesco Cucca, Mariza de Andrade, Philip L. De Jager, Jan-Emmanuel De Neve, Ian J. Deary, George V. Dedoussis, Panos Deloukas, Maria Dimitriou, Gun Eirksdttir, Martin F. Elderson, Johan G. Eriksson, David M. Evans, Jessica D. Faul, Luigi Ferrucci, Melissa E. Garcia, Henrik Grnberg, Vilmundur Gunason, Per Hall, Juliette M. Harris, Tamara B. Harris, Nicholas D. Hastie, Andrew C. Heath, Dena G. Hernandez, Wolfgang Hoffmann, Adriaan Hofman, Rolf Holle, Elizabeth G. Holliday, Jouke-Jan Hottenga, William G. Iacono, Thomas Illig, Marjo-Riitta Jrvelin, Mika Khnen, Jaakko Kaprio, Robert M. Kirkpatrick, Matthew Kowgier, Antti Latvala, Lenore J. Launer, Debbie A. Lawlor, Terho Lehtimki, Jingmei Li, Paul Lichtenstein, Peter Lichtner, David C. Liewald, Pamela A. Madden, Patrik K. E. Magnusson, Tomi E. Mkinen, Marco Masala, Matt McGue, Andres Metspalu, Andreas Mielck, Michael B. Miller, Grant W. Montgomery, Sutapa Mukherjee, Dale R. Nyholt, Ben A. Oostra, Lyle J. Palmer, Aarno Palotie, Brenda W. J. H. Penninx, Markus Perola, Patricia A. Peyser, Martin Preisig, Katri Rikknen, Olli T. Raitakari, Anu Realo, Susan M. Ring, Samuli Ripatti, Fernando Rivadeneira, Igor Rudan, Aldo Rustichini, Veikko Salomaa, Antti-Pekka Sarin, David Schlessinger, Rodney J. Scott, Harold Snieder, Beate St Pourcain, John M. Starr, Jae Hoon Sul, Ida Surakka, Rauli Svento, Alexander Teumer, The LifeLines Cohort Study,

Henning Tiemeier, Frank J. A. van Rooij, David R. Van Wagoner, Erkki Vartiainen, Jorma Viikari, Peter Vollenweider, Judith M. Vonk, Grard Waeber, David R. Weir, H.-Erich Wichmann, Elisabeth Widen, Gonneke Willemsen, James F. Wilson, Alan F. Wright, Dalton Conley, George Davey-Smith, Lude Franke, Patrick J. F. Groenen, Albert Hofman, Magnus Johannesson, Sharon L. R. Kardia, Robert F. Krueger, David Laibson, Nicholas G. Martin, Michelle N. Meyer, Danielle Posthuma, A. Roy Thurik, Nicholas J. Timpson, Andr G. Uitterlinden, Cornelia M. van Duijn, Peter M. Visscher, Daniel J. Benjamin, David Cesarini, and Philipp D. Koellinger (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science*, **340** (6139): 1467-1471.

[20] Sequeira, T. and N. Ferraz (2009). Is Education prejudiced by Country-Risk? *Economic Record* **85**(270): 307-317.

[21] Sequeira, T. (2009). On the relationship between Education and Country-Risk: A new fact. *Acta Oeconomica* **59**(4): 441-454.

[22] Tabellini, G. (2008). Institutions and Culture. *Journal of the European Economic Association*, **6**(2-3): 255-294.

## A Appendix

Table A.1: Human Capital and Genetic Diversity (other controls)

	Mobility index- predicted genetic diversity	Mobility index- predicted genetic diversity square	Social infrastruc- ture	% of population at risk of contract- ing malaria	$R^2$ / Ob- servations
(1)	227.93** (105.18)	-156.42** (74.28)	5.82*** (0.90)	-2.53*** (0.55)	0.70 78
(2)	248.92*** (86.44)	-168.62*** (61.35)	3.61*** (0.58)	-0.80*** (0.25)	0.70 47
(3)	7.19 (8.04)	-4.70 (5.64)	0.25*** (0.06)	-0.10*** (0.03)	0.52 72
(4)	89.17 (133.93)	-59.07 (95.45)	7.19*** (0.86)	-2.65*** (0.56)	0.74 72
(5)	108.63 (139.96)	-71.69 (99.64)	7.78*** (0.90)	-2.75*** (0.59)	0.75 72
(6)	284.70*** (73.03)	-200.08*** (52.64)	1.08*** (0.27)	-0.36 (0.30)	0.51 44
(7)	279.69*** (75.24)	-196.47*** (54.18)	1.10*** (0.29)	-0.47 (0.39)	0.47 44
(8)	113.78*** (32.74)	-79.84*** (23.65)	0.37*** (0.11)	-0.06 (0.13)	0.46 44
(9)	16.02** (6.27)	-11.26** (4.48)	0.079*** (0.02)	-0.05*** (0.02)	0.49 44
(10)	3856.87** (1598.96)	-2717.58** (1146.60)	28.06*** (5.78)	-23.38*** (6.96)	0.61 40
(11)	3879.79** (1611.17)	-2733.94** (1154.79)	28.19*** (5.83)	-23.82*** (6.88)	0.61 40
(12)	933.54*** (298.61)	-654.01*** (215.11)	6.08*** (1.11)	-4.03*** (1.21)	0.65 40
(13)	116.71** (51.31)	-81.54** (36.55)	0.77*** (0.16)	-0.55** (0.21)	0.55 40

Note: Dependent Variables - (1) Years of schooling; (2) Years of schooling\*Interpersonal Trust; (3) % of population aged 15 or over with complete secondary education; (4) Years of schooling of population 15 and over, whether studying or not; (5) Years of schooling of population 15-64 who is not studying; (6) Average test score in math and science, primary through end of secondary school, all years (scaled to PISA scale divided by 100); (7) Average test score in math and science, only lower secondary, all years (scaled to PISA scale divided by 100); (8) Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school, all years); (9) Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years); (10) Years of schooling\*Average test score in math and science, primary through end of secondary school; (11) Years of schooling\*Average test score in math and science, only lower secondary; (12) Years of schooling\*Share of students reaching basic literacy (based on average test scores in math and science, primary through end of secondary school); (13) Years of schooling\*Share of top-performing students (based on average test scores in math and science, primary through end of secondary school, all years).

Level of significance: \*\*\* for p-value<0.01; \*\*for p-value<0.05;\* for p-value<0.1.  
Values between parentheses are standard errors.

Table A.2: Regressions for human capital proxies in 1500

	Mobility index- predicted genetic diversity	Mobility index- predicted genetic diversity square	Log [Neolithic transition timing]	Log [per- centage of arable land]	Log [absolute latitude]	Log [land suitability for agricul- ture]	$R^2$ / Ob- servations
(1)	35.59** (15.41)	-26.27** (11.39)	0.37*** (0.05)	0.04 (0.04)	-0.01 (0.02)	0.002 (0.03)	0.56 94
(2)	32.02*** (9.72)	-22.57*** (7.33)	0.31*** (0.06)	0.04 (0.04)	0.13*** (0.02)	0.03 (0.03)	0.71 94
(3)	7.73 (6.83)	-4.71 (5.17)	0.30*** (0.03)	0.02 (0.02)	0.06*** (0.02)	-0.01 (0.02)	0.69 94
(4)	12.94 (10.37)	-8.36 (7.84)	0.37*** (0.05)	0.02 (0.03)	0.10*** (0.03)	0.003 (0.03)	0.62 94
(5)	27.12** (11.68)	-17.08* (8.81)	0.34*** (0.04)	-0.01 (0.03)	0.07*** (0.02)	-0.002 (0.02)	0.73 94
(6)	23.08*** (7.74)	-15.80*** (5.79)	0.34*** (0.03)	0.02 (0.02)	0.07*** (0.02)	0.005 (0.02)	0.78 94

Note: Dependent Variables - (1) Average human capital in agriculture in period 1500 CE; (2) Average human capital in communications in period 1500 CE; (3) Average human capital in transportation in period 1500 CE; (4) Average human capital in military in period 1500 CE; (5) Average human capital in industry in period 1500 CE; (6) Average of the sectoral human capital indexes in period 1500 CE; Level of significance: \*\*\* for p-value<0.01; \*\* for p-value<0.05; \* for p-value<0.1. Values between parentheses are standard errors.