The Origins of the Division of Labor in Pre-modern Times

Depetris-Chauvin, Emilio and Özak, Ömer

14 February 2018

The Origins of the Division of Labor in Pre-modern Times[*]

Emilio Depetris-Chauvin[†]and Ömer Özak[‡]

May 8, 2020

Abstract

This research explores the historical roots of the division of labor in pre-modern societies. Exploiting a variety of identification strategies and a novel ethnic level dataset combining geocoded ethnographic, linguistic and genetic data, it shows that higher levels of intra-ethnic diversity were conducive to economic specialization in the pre-modern era. The findings are robust to a host of geographical, institutional, cultural and historical confounders, and suggest that variation in intra-ethnic diversity is a key predictor of the division of labor in pre-modern times.

1 Introduction

At least since Adam Smith, the presence of individuals exclusively engaged in specific occupations has been considered fundamental to comparative economic development.[1] The importance of this division of labor is attributed to its essential role in the accumulation of production specific human capital and the potential benefits of learning by doing, all of which may be conducive to increasing productivity, innovation, trade, and economic development.[2] Moreover, these consequences of the division of labor may in turn have provided the fertile ground for the emergence and subsequent evolution of complex social organizations.[3] Indeed, a complex division of labor has been prevalent since pre-modern times, with some types of division of labor, at least along sexual lines within a family, being present in almost every society (Nolan and Lenski, 2011).[4] Despite the prevalence of the division of labor since pre-modern times and its suggested fundamental role for the progression of societies, little, if anything, is known about its deep-rooted determinants. This research aims to fill this void by exploring the deep historical roots of the division of labor in pre-modern societies.[5] Specifically, we explore the role of diversity across individuals in a society on its division of labor, a fundamental hypothesized force underlying the division of labor in the classic literature on the subject (Smith, 1776; Ricardo, 1817; Marx, 1867; Durkheim, 1893).

The role of within-population diversity in economic development has been explored by a recent vibrant literature spurred by the seminal work of Ashraf and Galor (2013b), which established that within-population diversity has a hump-shaped effect on economic development. This relation has been hypothesized to reflect a fundamental interplay between beneficial and detrimental effects of diversity on productivity at the societal level. Further work has explored different potential mechanisms underlying the aforementioned hump-shaped relationship.[6] Interestingly, while Ashraf and Galor (2013b) suggest that higher within-population diversity should be beneficial for productivity due to larger complementarities between individuals, the empirical evidence on the mechanisms underlying these benefits is sparse. Indeed, most of the existing empirical evidence suggests that diversity adversely affects contemporary social cohesiveness, trust, autocratic forms of governance, and violence (Arbatli et al., 2018; Galor and Klemp, 2017). Our research identifies a potential mechanism through which within-population diversity can positively affect productivity and long-run development: the emergence

---

[1]The idea presented by Smith (1776) has been shared by many philosophers and political economists across the ages, including Plato, Xenphon, Aristotle, Kuan Chung, Mencius, Hsün Tzu, al-Ghazali, Ibn Khaldün, Thomas Aquinas, David Hume, Karl Marx, Emilé Durkheim, among others (Sun, 2012).

[2]There exists a large theoretical literature on the relation between division of labor and economic outcomes (Stigler, 1951; Houthakker, 1956; Romer, 1987; Kiyotaki and Wright, 1989; Yang and Borland, 1991; Borland and Yang, 1992; Henrich and Boyd, 2008; Yang and Sachs, 2008).

[3]This relation has also been previously hypothesized in other social sciences outside economics (Claessen and Skalník, 1978; Brumfiel and Earle, 1987; Childe and Wailes, 1996; Polanyi, 2001; Henrich and Boyd, 2008).

[4]Furthermore, a high degree of specialization of labor, tasks, and other functions within different specific groups of people has been linked to societal advancement and prosperity (Durkheim, 1893). E.g., Trigger (1983) argues that the archeological evidence from the Gerzean period in Egypt (ca. 3500BCE) supports the view that the appearance of occupational specialization, such as the existence of craft specialists producing ornaments of gold, silver, cast copper, and lapis lazuli, was accompanied by the rise of complex social and economic institutions.

[5]We use the term pre-modern to highlight the fact that the ethnicities in our analysis are reflecting historical conditions in earlier stages of development before the industrial revolution.

[6]See Ashraf and Galor (2018) for a recent overview of this literature.

of the division of labor.

Our research establishes that population diversity was conducive to the division of labor in pre-modern times across ethnic groups.[7] The underlying idea is simple: more diverse ethnic groups experienced larger variations across individuals in intergenerationally transmitted traits (e.g., preferences, skills, human capital), which increased the complementarities between individuals and between individuals and their environment.[8] Specifically, differences across individuals in intergenerationally transmitted traits, allowed them to better exploit the potential gains of increased occupational specialization.[9] Therefore, during the pre-modern era economic specialization should emerge and be more prevalent among diverse ethnic groups. Additionally, since diverse geographical environments should allow diverse ethnic groups to generate larger complementarities, diverse ethnicities inhabiting diverse environments should have a larger division of labor.

To empirically test the aforementioned predictions, this research combines geocoded ethnographic, linguistic and genetic data to construct a novel dataset of measures of intra-ethnic diversity and economic specialization for pre-modern ethnic groups. In particular, for over 1100 ethnicities, the research constructs novel measures of economic specialization, based on the number of economic activities in which specialization existed in the pre-modern era. By performing the analysis at the *ethnic level*, the research sidesteps potential pitfalls from the aggregation of data (e.g., to the country level). Thus, the analysis focuses on the effects of *intra-ethnic* diversity, mitigating the potential confounding effects of *country-level inter-ethnic* diversity, which have been widely explored in the literature.

A major challenge for the analysis is the measurement of diversity within an ethnic group in the past. The lack of direct measures of skills, preferences or other relevant intergenerationally transmitted traits for individuals within an ethnicity, rules out the direct measurement of historical intra-ethnic diversity in those specific traits. To address this problem, Ashraf and Galor (2013b) have suggested using measures of genetic diversity as proxies for population diversity, which have been widely used in the ensuing literature. Our research generalizes this approach and addresses this challenge by additionally introducing novel measures of intra-ethnic linguistic diversity.[10] In particular, language and genetic material are also intergenerationally transmitted, and as shown below, share with other intergenerationally transmitted traits the common historical source of exogenous variation identified by Ashraf and Galor (2013b). Based on this fact, our analysis employs these measures of genetic and linguistic dirversity as proxies of historical intra-ethnic diversity in intergenerationally transmitted traits. Moreover, it shows that even if genetic and linguistic diversity do not have a direct effect on

---

[7]The analysis follows the approach in the literature and identifies pre-modern societies by their ethnicity, and uses these two terms interchangeably (Gennaioli and Rainer, 2007; Michalopoulos, 2012; Alesina et al., 2013).

[8]While most of the previous literature has focused on the Ricardian comparative advantages generated within the population due to differences in abilities (Ricardo, 1817; Ashraf and Galor, 2013b), it is clear that the same effects can be generated by diversity of preferences (Yang and Sachs, 2008). Thus, diversity in either the supply or demand side may underlie the division of labor.

[9]I.e., the emergence within a society of individuals exclusively engaged in specific occupations, e.g., a baker, a butcher, or a metalworker. Importantly, the lack of economic specialization does not imply the lack of knowledge about an activity. E.g., members of the *Aché* tribe of Paraguay, while having the knowledge to produce arrows, bows, huts, among other goods, were not specialized.

[10]Unlike the previous literature that has employed cross-country measures of linguistic fractionalization or polarization, which reflect the variation in the number of languages, these linguistic diversity measures reflect diversity within a language.

the division of labor, they still capture the effect of diversity in relevant traits of the population. Thus, these measures of genetic and linguistic diversity capture general aspects of historical intra-ethnic diversity, above and beyond genes and phonemes.[11]

The research explores the deep historical roots of the division of labor, by analyzing the effect of intra-ethnic diversity on the prevalence of economic specialization in pre-modern times across ethnic groups in various steps. First, using ordinary least squares and a restricted sample of ethnicities for which ethnographic, genetic and linguistic data exist, the empirical analysis documents the robust positive statistically and economically significant relation between diversity and economic specialization. Clearly, these statistical associations do not necessarily imply causality and could arise from omitted confounders, such as heterogeneity in environmental factors, or as a result of reverse causality from, for instance, the emergence of institutions on the composition of populations.

In order to mitigate these potential concerns, the research follows several strategies. First, it accounts for the confounding effect of a large set of geographical controls of the ethnic homeland. Second, it shows that the main results are not driven by other plausible sources for the emergence of economic specialization such as variation in agricultural suitability, ecological diversity, spatial and intertemporal temperature volatility, pre-1500CE caloric suitability and mobility costs.[12] Third, it follows the instrumental variable approach introduced by Ashraf and Galor (2013b), and exploits the exogenous variation in intra-ethnic diversity generated by a fundamental statistical process brought about by historical migratory patterns known as a serial founder effect (SFE).[13] Indeed, as could be expected in an era when populations were small and knowledge and culture, among others, were passed orally between generations, the decrease in diversity along historical migratory routes has been documented for various intergenerationally transmitted traits. In particular, the diversity in the shape and size of arrow heads and handaxes, cultural memes, and phenotypes, as well as the proxies of intra-ethnic diversity based on genetic and linguistic data used in this paper, have been shown to follow a serial founder effect (Ramachandran et al., 2005; Lycett and von Cramon-Taubadel, 2008; Lycett, 2008; Hamilton and Buchanan, 2009; Betti et al., 2009; Rogers et al., 2009; Atkinson, 2011; Creanza et al., 2015). Reassuringly, after exploiting these three strategies jointly, the research finds a positive robust relation between intra-ethnic diversity and the prevalence of economic specialization for the restricted sample of ethnic groups. Additionally, it proposes conditions under which the estimated coefficient provides a lower bound for the true causal effect, suggesting an economically significant positive effect of intra-ethnic diversity.

---

[11] Our theory does not need to take a stand on which specific trait underlies the effect of intra-ethnic diversity on the division of labor, since any intergenerationally transmitted trait, such as preferences or skills, that leads to larger complementarities, should have qualitatively similar effects (Yang and Borland, 1991; Yang and Sachs, 2008). Indeed, e.g., in Yang and Borland (1991) the model can be interpreted so that diversity in skills or in preferences or both generate economic specialization. Moreover, this suggests that the mechanism of transmission of these traits should not affect the effect of intra-ethnic diversity on economic specialization. Specifically, the effects of intra-ethnic diversity should not depend on whether traits are culturally or genetically transmitted across generations.

[12] While this paper focuses on the effect of intra-ethnic diversity and its interaction with environmental diversity, the analysis also sheds light on the role of geographical factors on the emergence of the division of labor, as well as their relative importance compared to intra-ethnic diversity. In particular, it explores the effect of geographical determinants of market size on the emergence of the division of labor.

[13] As established in section 3.2, SFE generated exogenous variation in the proxies of intra-ethnic diversity employed in this research.

In a second stage of the analysis, the research exploits the implications of the serial founder effect and a two-step econometric method to generate predicted intra-ethnic diversity measures (Murphy and Topel, 2002; Ashraf and Galor, 2013b). This allows the analysis to be performed on a sample of more than 900 ethnicities. By increasing the sample size, the analysis mitigates potential concerns regarding geographical coverage and representativeness of the restricted sample. Moreover, the increased sample size permits the relation between intra-ethnic diversity and specialization to be estimated more precisely. Reassuringly, and in line with the proposed hypothesis, the estimated relation between intra-ethnic diversity and the prevalence of economic specialization is positive, statistically and economically significant. Moreover, the research finds a positive complementarity between an ethnicity's diversity and the diversity of the environment it inhabits on the prevalence of economic specialization. These results are robust to large battery of tests, including different measures of economic specialization, accounting for a large set of historical and cultural confounders, econometric specifications, multiple hypothesis testing, among many others.

The analysis also explores the relative importance of intra-ethnic diversity as a determinant of economic specialization. Indeed, it documents that the variation uniquely associated with intra-ethnic diversity is larger than the one associated with all other determinants included in the analysis. These findings provide evidence that is consistent with a fundamental role of intra-ethnic diversity as a determinant of economic specialization across ethnicities. Furthermore, the analysis explores the relation between intra-ethnic diversity, economic specialization and development outcomes. Specifically, it provides evidence that intra-ethnic diversity has a large direct effect on economic specialization, which is not mediated by other development outcomes. Additionally, it suggests that economic specialization plays a fundamental mediating role in the link between population diversity and development.

This research is the first attempt to identify the deep-rooted historical factors behind the prevalence of economic specialization in pre-modern times across ethnic groups. Moreover, it is the first to document the robust positive effect of (i) intra-ethnic diversity and (ii) the complementarity between the heterogeneity of both population and environment on economic specialization. Additionally, it is the first paper to use measures of linguistic diversity (within a language) to proxy for intra-population diversity. Also, it provides a conceptual empirical framework to rationalize the use of these measures of genetic and linguistic diversity as proxies for population diversity in general. In doing so, this research contributes to various strands of literature.

First, this research contributes to the literature studying societal attributes in the past (Ahlerup and Olsson, 2012; Michalopoulos, 2012; Giuliano and Nunn, 2013; Fenske, 2014). In particular, it sheds light on the origins of a fundamental driver of economic development in the pre-industrial and contemporary eras (Smith, 1776). Given the persistence of culture, institutions, human capital and technology, and their effect on development, the findings suggest a novel potential source of economic development (Diamond, 1997; Glaeser et al., 2004; Acemoglu et al., 2005; Nunn, 2008; Alesina et al., 2013; Ashraf and Galor, 2013b; Michalopoulos and Papaioannou, 2013; Spolaore and Wacziarg, 2013; Depetris-Chauvin, 2014; Galor and Özak, 2016; Andersen et al., 2016).[14] Additionally, by unveiling

---

[14]Indeed, Depetris-Chauvin and Özak (2015, 2016) explore the long-run consequences of the pre-modern division of labor in more detail. They provide evidence that there exists a strong positive association between an ethnicity's pre-modern level of economic specialization and its contemporary level of economic development.

the complementarities between population and the environment the analysis bridges the gap between the literature that focuses on their independent roles in shaping long-run development (Michalopoulos, 2012; Alesina et al., 2013; Ashraf and Galor, 2013b; Galor and Özak, 2016; Giuliano and Nunn, 2016).

Second, this research contributes to the literature on the effects of diversity on economic development, which has previously been explored using various country-level measures of genetic, ethnic, cultural, and religious diversity (Easterly and Levine, 1997; Alesina et al., 2003; Desmet et al., 2012, 2015; Ashraf and Galor, 2013a,b; Arbatlı et al., 2020; Cook, 2015; Alesina et al., 2016). While most of the existing empirical evidence, suggests that diversity adversely affects contemporary social cohesiveness and development, this paper documents a positive effect of intra-ethnic diversity on a key driver of economic development. By doing so, it provides prima-facie empirical evidence consistent with the upward sloping side of the hump-shaped relationship between diversity and economic development established by Ashraf and Galor (2013b).

Finally, this research contributes to the emerging literature that uses genetic diversity to understand the deep-rooted determinants of modern comparative development (Ashraf and Galor, 2013b; Arbatlı et al., 2020; Ashraf and Galor, 2013a; Cook, 2015). By additionally using linguistic data, this is the first paper to exploit non-genetic historical proxies of intra-ethnic diversity in the literature. This analysis sheds new light on the role of these proxies of intra-ethnic diversity as drivers of comparative development across the globe. Specifically, they provide support the view that neutral genetic diversity reflects a wider range of non-neutral human characteristics (Ashraf and Galor, 2018).

The remainder of the paper is organized as follows. Section 2 discusses the empirical strategy and describes the data. Section 3 analyzes the impact of intra-ethnic diversity on economic specialization. Section 5 explores the robustness of the results. Section 6 concludes.

2   Data and Empirical Strategy

This section develops the empirical strategy and introduces measures of pre-modern economic specialization, historical intra-ethnic diversity, and geographical controls at the ethnic level required to implement the empirical strategy. The empirical strategy builds upon the empirical method pioneered by Ashraf and Galor (2013b) and provides an econometric framework to understand the role of neutral measures of population diversity in the estimation of the effect of diversity on economic specialization in particular, and development more generally.

2.1   Empirical Strategy

The proposed theory suggests that the empirical relation between economic specialization and intra-ethnic diversity is given by

$$s_i = \alpha_0 + \alpha_1 d_i + \sum_{k=1}^{K} \alpha_{2k} x_{ik} + \epsilon_i \tag{1}$$

where s_i measures economic specialization, d_i is a weighted average of intra-ethnic diversity measures across various intergenerationally transmitted traits, x_{ik} is the level of geographical characteristic k and ϵ_i is the error term, all for ethnicity i.[15] The main prediction of the theory, which will be tested below, is that $\alpha_1 > 0$. Let $d_i = \sum_{j=1}^{J} \alpha_{1j} d_{ij}$, where d_{ij} is the level of intra-ethnic diversity in intergenerationally transmitted trait $j = 1, \ldots, J$, $\alpha_{1j} > 0$ denotes its importance in the effect of intra-ethnic diversity and $\sum_j \alpha_{1j} = 1$. The identification of the effect of intra-ethnic diversity on economic specialization based on equation (1) poses two types of challenges: measurement and causal identification.

First, as discussed in the introduction, the lack of direct measures of historical intra-ethnic diversity across intergenerationally transmitted traits, prevents the construction of the measure d_i. Nonetheless, as previously discussed, it has been shown that different intergenerationally transmitted traits share a common underlying exogenous determinant: the serial founder effect generated by historical migratory patterns (Ashraf and Galor, 2018).

A serial founder effect (SFE) implies that successive divisions of an original population into various subpopulations generates a loss of diversity in intergenerationally transmitted characteristics such as genes, phonemes, cultural traits, preferences, knowledge, skills, etc. Of particular interest is the SFE generated by the dispersal of anatomically modern humans out of East Africa more than $60,000$ years ago (Ramachandran et al., 2005).[16] In particular, according to the Out-of-Africa hypothesis, which posits the African origin of modern humans, the SFE implies that diversity decreases along migratory routes from East Africa. Importantly, it has been established that genetic and linguistic diversity decrease with the migratory distance from East Africa (Ramachandran et al., 2005; Manica et al., 2007; Atkinson, 2011). Moreover, as could be expected in an era when knowledge and culture, among others, were passed orally between generations, the decrease in diversity along historical migratory routes has also been documented for non-genetic traits such as arrow heads, handaxes, cultural memes, and phenotypes (Lycett and von Cramon-Taubadel, 2008; Lycett, 2008; Hamilton and Buchanan, 2009; Betti et al., 2009; Rogers et al., 2009; Atkinson, 2011).[17] While all intergenerationally transmitted traits are theoretically affected by this SFE, human and cultural evolution may have altered intra-ethnic diversity in these traits and obscured the effect of this exogenous variation. A major advantage of the two proxies of intra-ethnic diversity used in the analysis, is that they are mostly unaffected by the forces of human and cultural evolution, especially the measure of genetic diversity (Ashraf and Galor, 2013b, 2018).

---

[15]Appendix A establishes similar results for the case when specialization is affected by intra-ethnic diversity in specific traits instead of a weighted average of intra-ethnic diversity across various traits. It shows that the estimated effect provides a lower bound for the total effect of intra-ethnic diversity among all traits affected by a SFE. In particular, it shows the robustness of this result to the potential negative effect of specific intergenerationally transmitted traits on economic specialization.

[16]While the East African origins of humans is the most widely accepted theory, there is an ongoing debate on the specific location within Africa. In particular, Chan et al. (2019) among others have suggested a Southern African origin (López et al., 2015, see also).

[17]These effects have been found in both human and non-human species (Baker and Jenkins, 1987). Moreover, the decrease in diversity due to migration and serial founder effects has been found in later migratory processes within continents (Wang et al., 2007; Friedlaender et al., 2008; Lao et al., 2008; Myres et al., 2011; Pinhasi et al., 2012).

Thus, ethnicity i's historical diversity in intergenerationally transmitted trait j is determined by

$$d_{ij} = \beta_{0j} + \beta_{1j}D_i + \sum_{k=1}^{K} \beta_{2jk}x_{ik} + \eta_{ij}, \tag{2}$$

where η_{ij} is the error term, $\beta_{1j} < 0$ and D_i is the historical migratory distance from its homeland to East Africa. This implies that:[18]

Proposition 1. *If some trait p is observable, so that diversity in p can be measured, then d_{ip} serves as a proxy for all other measures of diversity. In particular, for $j \neq p$,*

$$d_{ij} = \gamma_{0j} + \gamma_{1j}d_{ip} + \sum_{k=1}^{K} \gamma_{2jk}x_{ik} + \zeta_{ij}, \tag{3}$$

where $\gamma_{1j} = \frac{\beta_{1j}}{\beta_{1p}} > 0$ for all $j \neq p$.

Using the definition of d_i and equation (3), equation (1) can be rewritten in terms of the proxy of intra-ethnic diversity d_{ip} as

$$s_i = \delta_0 + \delta_1 d_{ip} + \sum_{k=1}^{K} \delta_{2k}x_{ik} + \varepsilon_i, \tag{4}$$

where

$$\delta_1 = \alpha_1 \left(\alpha_{1p} + \sum_{j \neq p} \alpha_{1j}\gamma_{1j} \right). \tag{5}$$

Clearly,

Proposition 2. *The proxy d_{ip} has a positive association with specialization, if and only if, intra-ethnic diversity has a positive effect on it, i.e., $\delta_1 > 0$ if, and only if, $\alpha_1 > 0$.*

This implies that the estimation of δ_{ip} provides a way to identify the sign of the effect of intra-ethnic diversity. Moreover,

Proposition 3. *Assume all variables in the analysis have been standardized to have a variance of 1 in order to allow comparison of coeffcients. If $\alpha_1 > 0$ and the distance D_i has the largest effect on the proxy p, i.e., $\beta_{1p} < \beta_{1j}$ for all $j = 1, \ldots, J$, then δ_1 provides a lower bound to the effect of intra-ethnic diversity on economic specialization, i.e.*

$$\delta_1 < \alpha_1. \tag{6}$$

Thus, equation (4) can be used to identify the sign and provide a lower bound of the effect of intra-ethnic diversity on economic specialization. Moreover, equation (5) implies that the proxy d_{ip}

---

[18]Appendix A provides the proofs, the relation between the various parameters, and all the intermediate steps to obtain the results presented in this section.

can be used to identify this lower bound of the causal effect, even if the proxy itself has no effect on economic specialization, i.e., if $\alpha_{1p} = 0$.

These results suggest that the measurement and identification problems require finding appropriate proxies. As explained in section 2.3, the analysis constructs measures of genetic and linguistic diversity that satisfy the above conditions for various samples of ethnicities. In particular, section 3.2 empirically shows that both types of proxies are affected by a serial founder effect. Moreover, the measure of genetic diversity used in the analysis has been shown to be mainly affected by the serial founder effect of the Out-of-Africa migration of anatomically modern humans and has not been affected by other factors (Ramachandran et al., 2005; Betti et al., 2009; Creanza et al., 2015). On the other hand, while this serial founder effect should have operated on many other intergenerationally transmitted traits, their diversity may have been influenced by many other factors. Thus, the effect of the distance to East Africa is expected to have the largest negative effect on genetic diversity compared to its effect on the diversity of these other traits, i.e., $\beta_{1genetic} < \beta_{1j}$ for any trait j, inclusive linguistic diversity. Indeed, section 3.2 provides supportive evidence for this conjecture. Thus, the analysis employs these measures as proxies for intra-ethnic diversity of all types of intergenerationally transmitted traits in order to estimate this lower bound of the causal effect. Importantly, these proxies capture factors that should be "neutral" to human behavior. Thus, although it can be expected that neither proxy has a direct effect on economic specialization, i.e., $\alpha_{1genetic} = 0$ and $\alpha_{1linguistic} = 0$, they can be used to provide a lower bound for the causal effect of intra-ethnic diversity on economic specialization δ_1. Clearly, the estimation of δ_1 is subject to various additional identification challenges.

The analysis surmounts significant hurdles in the estimation of δ_1, i.e. the lower bound of the effect of intra-ethnic diversity on the division of labor. First, the results may be biased by omitted geographical, institutional, cultural, or human characteristics that might have determined economic specialization and are correlated with the proxy of intra-ethnic diversity. Thus, several strategies are employed to mitigate this concern: (i) The analysis accounts for a large set of confounding geographical characteristics (e.g., absolute latitude, area of the ethnic homeland, average elevation, terrain ruggedness, accessibility to navigable water, average temperature and precipitation). (ii) It accounts for other plausible sources for the emergence of economic specialization such as variation in agricultural suitability, ecological diversity, spatial and intertemporal temperature volatility, pre-1500CE caloric suitability and mobility costs. (iii) It accounts for continental fixed effects, capturing unobserved time-invariant heterogeneity at the continental level. (iv) It conducts within language phylum analyses of the effect of intra-ethnic diversity on the division of labor, accounting for language phylum fixed effects and thus unobserved time-invariant language-phylum-specific factors and common cultural history. (v) It accounts for the potential confounding effects of other historical processes like the adoption of agriculture or the continuity of human presence on economic specialization. (vi) It accounts for the geographical isolation and proximity to other ethnicities in order to account for the confounding effects of potential historical and spatial dependence generated by sharing common cultural ancestry or by the level of potential interaction with other ethnicities.

Second, the results may be biased due to reverse causality, measurement errors or some remaining hard to account omitted factors. In order to mitigate these concerns, the analysis employs an instru-

mental variable approach to estimate δ_1. In particular, it exploits differences in the distance to East Africa as a source of exogenous variation following Ashraf and Galor (2013b). Indeed, as established in section 3.2, this distance is a major determinant of the proxies of intra-ethnic diversity employed in the analysis, thus satisfying the first condition for a good instrument (i.e., relevance). Thus, the main requirement for the identification of the lower bound of the causal effect, δ_1 in equation (4), is the exogeneity assumption $E(D_i\varepsilon_i \mid (x_{ik})_{k=1}^K) = 0$. While this condition cannot be tested, the results in appendix G show that the reduced form estimates in the regression of economic specialization on the distance to East Africa are very stable across many specifications. Since accounting for different sets of confounders affects these estimates remarkably little, this analysis does not seem to be subject to selection on unobservables, suggesting this exogeneity condition may be satisfied in practice. Moreover, Table E.16 provides further support for this exogeneity assumption by accounting for other distances and historical processes, suggesting additionally that the exclusion restriction may hold.

These results can be summarized as follows:

(i) Genetic and linguistic diversity can be used as proxies for intra-ethnic diversity of intergenerationally transmitted traits (affected by similar serial founder effects).

(ii) The distance to East Africa is a plausible instrument for intra-ethnic diversity.

(iii) The estimated effect of the proxies of intra-ethnic diversity presented in the body of the paper captures the combined effect of intra-ethnic diversity in all relevant intergenerationally transmitted traits and provides a lower bound to the true causal effect.

(iv) The estimated effect of the proxies of intra-ethnic diversity does not necessarily imply that genetics or language are the fundamental mechanism behind the effect of intra-ethnic diversity on the division of labor.

2.2 Dependent Variables: Pre-Modern Economic Specialization

The analysis employs the two main sources for ethnic level data currently available, namely the Ethnographic Atlas (EA) and the Standard Cross-Cultural Sample (SCCS). Both datasets have been widely used in anthropology and economics for the study of pre-modern societies and the long-term effects of pre-modern culture and institutions (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Alesina et al., 2013; Fenske, 2014). The Ethnographic Atlas (Murdock, 1967) includes information on 115 characteristics for 1267 ethnicities around the globe. On the other hand, the Standard Cross-Cultural Sample (Murdock and White, 1969) expands the set of characteristics to over 2000 for a subsample of 180 independent ethnicities. By combining both datasets the analysis overcomes the restriction in terms of thematic coverage of the EA and ethnic/geographic coverage of the SCCS.[19] Figure 1 depicts the location of the full sample of ethnicities used in the main analysis. Additionally,

---

[19]The main reason behind the construction of the SCCS was to overcome Galton's independence problem, i.e., the difficulties of drawing inferences from cross-cultural data due to spatial auto-correlation and historical dependence. The sample of ethnicities in the SCCS were chosen so as to minimize this problem (Murdock and White, 1969).
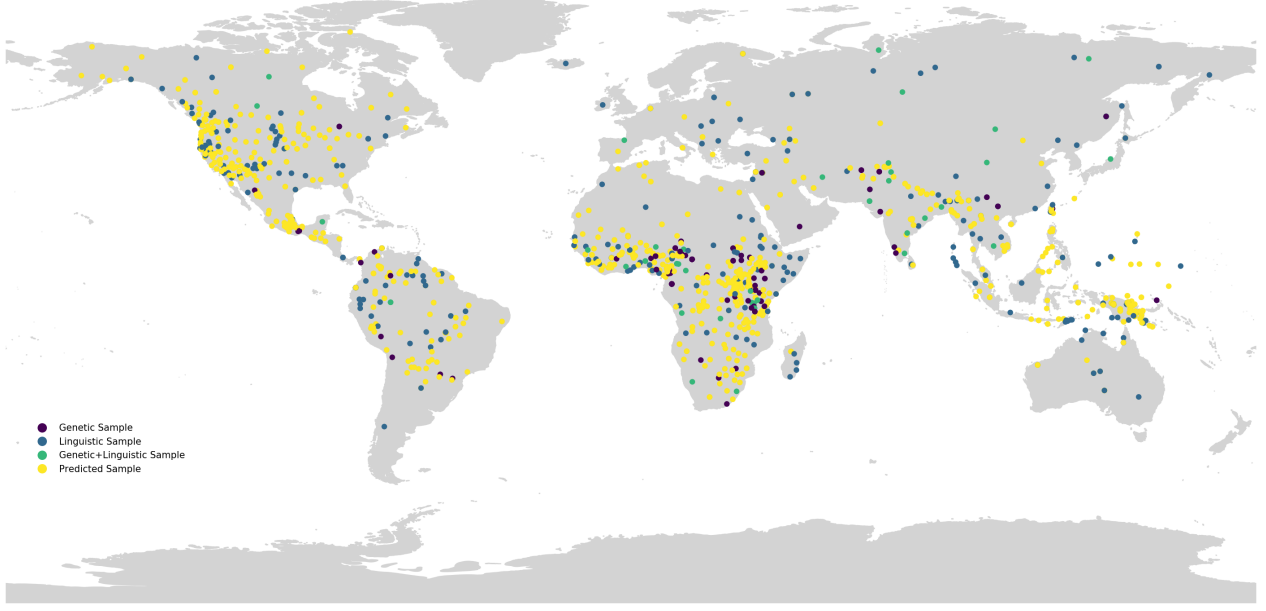
Figure 1: Location of Ethnicities employed in the Analysis (Full and Restricted Samples)

it highlights the ethnicities for which (genetic or linguistic) intra-ethnic diversity data is available and those for which it is predicted as explained below.[20]

In order to explore the hypothesis proposed in this paper, the analysis constructs various novel measures of economic specialization of labor at the ethnic level using data from both the EA and SCCS. In particular, both datasets include variables on the existence of "age or occupational specialization" for metal working (v55), weaving (v56), leather working (v57), pottery making (v58), boat building (v59), house construction (v60), gathering (v61), hunting (v62), fishing (v63), animal husbandry (v64), and agriculture (v65). For each of these activities, the EA and SCCS assess if the ethnic group had craft, industrial or age specialization or if the activity was absent or no specialization occurred. These variables allow the identification of ethnicities in which specialization existed in the pre-modern era. On the other hand, these variables do not allow for the differentiation of ethnicities where no specialization occurred from those in which the activity was absent, thus confounding the lack of specialization with the lack of the activity. In order to overcome this problem, the analysis uses additional information from variables v44-v54 in order to assess, for the same activities, whether the activity was absent or unimportant or present.

Based on this information, the analysis constructs three measures of specialization. The first measure of the level of specialization in ethnicity e, s_e^1, counts the number of specialized activities, i.e. $s_e^1 = \sum_a s_{ea}$, where s_{ea} equals 1 if the activity was present *and* specialized in ethnicity e and zero otherwise. The second measure of the level of specialization in ethnicity e, s_e^2, is the share of activities present that were specialized, i.e. $s_e^2 = s_e^1/n_e$, where s_e^1 is the first measure and n_e is the number of activities available in ethnicity e. Finally, the third measure of the level of specialization in ethnicity e, s_e^3, is a score given by $s_e^3 = \sum_a \tilde{s}_{ea}$, where $\tilde{s}_{ea}$ equals 0 if the activity a is not present, 1 if it is present

---

[20]Figures B.1-B.6 map the different subsamples individually to improve visibility.

but it is not specialized, and 3 if the activity is present and specialized in ethnicity e.[21] Importantly, as shown below, the main results in the paper do not depend on the measure of specialization employed in the analysis. However, given the ease of interpretation and space limitations, the analysis focuses mainly on the number of specialized activities in an ethnicity, s_e^1. Figure B.7 depicts the spatial distribution and frequencies of the three measures. Importantly, the new measures of economic specialization correlate strongly among themselves.[22]

2.3 Independent variables: Intra-Ethnic Diversity

This research constructs a novel dataset on georeferenced intra-ethnic diversity at the ethnicity level using two types of proxies, namely genetic and linguistic diversity measures. It is important to note that both measures capture intra-ethnic diversity as opposed to inter-ethnic diversity, which has been widely used in the existing literature that analyzes cross-country differences in population diversity. As previously explained, an essential feature of these diversity measures is that the main source of their variation was caused exogenously by a serial founder effect (SFE).

The analysis constructs a novel dataset on georeferenced genetic diversity at the ethnicity level using the most comprehensive genomic data set on human micro-satellite variation to date (Pemberton et al., 2013). In particular, Pemberton et al. (2013) combine eight previous population-genetic datasets and analyze them following a standardized procedure, which ensures all the data is produced following a uniform method, ensuring comparability across populations and samples. This dataset contains information on 645 common single-nucleotide protein (SNP) loci for 5435 individuals from 267 independent ethnicities. There are two main advantages of using this data. First, it is based on predominantly indigenous populations, which ensures the population inhabited the same location for a prolonged period of time and lowers a potential concern generated by a possible admixture of populations (Pemberton et al., 2013). Second, the SNP's included in the analysis are "neutral" to selection, i.e. they are not involved in processes that encode proteins and thus are not subject to natural selection (Kimura, 1983).

Based on this data, this research constructs for each ethnicity a measure of genetic diversity based on what population geneticists call the expected heterozygosity within a population. In particular, the genetic diversity or expected heterozygosity of a population measures the average probability that two randomly chosen individuals in the population do not share the same allele of a gene, i.e. that they do not have the same variant form of the gene.[23] In order to ensure comparability across populations, the

---

[21]The analysis assigns a higher value to specialization in order to differentiate the effect of specialization from technological development. Reassuringly, using a value of 2 for specialization does not alter the main results.

[22]Moreover, given the theoretical association between division of labor and trade within and among economies, these novel measures are associated with intra-ethnic trade related measures available in the SCCS. In particular, the new measures are positively associated with trade among communities of the same ethnic group, the existence and type of money (media of exchange) and credit, the type of credit source, and the existence of writing and records (Tables C.1-C.4), suggesting that the new measures indeed capture the phenomenon under study. A major concern with the SCCS data is that it is only available for a small subset of ethnicities, especially once the availability of intra-ethnic diversity measures is taken into account.

[23]The literature on diversity has measured this population attribute using various characteristics like religion, language, ethnicity, or genetics. Diversity within a population is usually defined as the probability that two random individuals in a population do not share the same characteristic. For example, religious, linguistic or ethnic diversity/fractionalization estimate the probability that two random individuals in a population do not share the same religion, speak the same

analysis constrains the construction of the genetic diversity measure to the set of 619 common SNP loci for which information exists for all ethnic groups.[24]

Out of the 267 ethnicities for which genetic data is available, this research is able to match a subset of 149 ethnicities to the Ethnographic Atlas (EA). This maps the genetic diversity data to the EA, and thus, to all the cultural, institutional and geographic data contained in the EA or to other datasets to which the EA can be mapped. In particular, and as discussed below, ethnicities can be mapped to the geographical characteristics of their historical homelands.

Additionally, the research uses measures of intra-ethnic linguistic diversity, i.e., diversity of the language spoken by an ethnic group,[25] as alternative proxies of intra-ethnic diversity. In particular, the analysis employs measures of consonant inventories, vowel quality inventories, and the number of genders as identified by linguists in the World Atlas of Language Structures - WALS (Dryer, 2013). WALS is the most comprehensive, authoritative and widely used database of language structures available. Linguists have suggested the three measures employed in this analysis capture plausibly neutral elements of intra-ethnic (language) diversity, which have been determined by historical migratory processes (Rogers et al., 2009; Atkinson, 2011; Creanza et al., 2015). The analysis employs the mapping between WALS and EA/SCCS created by Galor et al. (2016) in order to link these alternative measures of intra-ethnic diversity to the ethnographic and geographic data. This results in 3 different additional samples of ethnicities with intra-ethnic diversity data: a sample of 299 ethnicities based on consonant inventories, another with 301 ethnicities based on vowel quality inventories, and one with 130 ethnicities based on the number of genders. Unlike the genetic diversity measure, these proxies of intra-ethnic diversity were potentially more affected by evolutionary processes (Creanza et al., 2015; Galor et al., 2016), decreasing the variation that could be explained by a serial founder effect. For this reason, the analysis focuses mostly on genetic diversity, but shows that qualitatively similar results are obtained when using these alternative proxies. Indeed, exploiting the predicted difference in the effect of the serial founder effect on these various proxies provides supporting evidence for the interpretation of the estimated effect of the genetic diversity as a lower bound of the true effect of intra-ethnic diversity. Moreover, given that only 48 ethnicities belong jointly to the genetic and the large linguistic diversity samples, while only 23 belong jointly to all samples with intra-ethnic diversity data, the robustness of the results to the measure of intra-ethnic diversity employed in the analysis suggests that sampling biases are not driving the results, and provide somewhat independent evidence for the effect of intra-ethnic diversity on economic specialization.

In order to expand the sample, the analysis generates predicted levels of intra-ethnic diversity for

---

language or have the same ethnic background. Similarly, genetic diversity or expected heterozygosity measure the expected genetic similarity between any two individuals in a population. It is important to note that *all* these measures capture diversity and do not measure any innate superiority of a certain type of characteristic over another. For example, a population in which there exists only one religion, language, ethnicity, or blood type, will be less diverse than one in which there are many, but the measures of diversity do not and cannot be used to identify if one *specific* religion, language, ethnicity or blood type is better than others.

[24]The genetic diversity on the full set of 645 loci is almost perfectly correlated with the measure used in the paper for the 267 original ethnicities in Pemberton et al. (2013). Their correlation is 0.99 ($p < 0.01$).

[25]This approach contrasts with the usual approach employed in the literature which exploits variations in the *number* of languages or ethnic groups within a region. Thus, our analysis captures within ethnic group diversity as opposed to inter-ethnic diversity.

the full sample of 1265 ethnicities available in the EA. In particular, the analysis exploits the variations in the pre-historical migratory distance to East Africa (Addis Ababa) in order to generate the predicted intra-ethnic diversity for the full sample of ethnicities available in the EA. More specifically, the analysis uses the empirical relation between the proxies of intra-ethnic diversity and the migratory distance to East Africa (in the restricted subsamples) to construct an out-of-sample predicted intra-ethnic diversity measure. Additionally, the analysis employs bootstrapped standard errors to address the generated regressor bias in the estimation of standard errors (Murphy and Topel, 2002; Ashraf and Galor, 2013b).

2.4 Geographical and Ethnic Controls

An ethnicity's pattern of economic activities, opportunities to trade, as well as its genetic and linguistic diversity may be confounded with the geographical characteristics of the ethnicity's homeland. Thus, the analysis accounts for a large set of geographical controls in order to attenuate concerns about omitted variable bias. In particular, using the mapping between geographic information systems (GIS) geometries of ethnic homelands and the EA and SCCS generated by Fenske (2014), the analysis constructs for each ethnicity a large set of geographical characteristics of its homeland. Tables B.1-B.2 show the list of all variables and their summary statistics for the various samples used in the paper.

In order to explore the robustness of the potential role of pre-modern economic development in the relationship between intra-ethnic diversity and economic specialization, the analysis further employs various measures from the SCCS and EA. Specifically, it uses measures of mean size of local communities, the levels of jurisdictional hierarchy of local community, class stratification, and statehood.

3 Origins of Economic Specialization

This section explores the deep historical origins of the division of labor. In particular, it exploits the plausible exogenous variation in intra-ethnic diversity generated by serial founder effects (during the migration out of Africa) to analyze the potential effect of intra-ethnic diversity, as measured by intra-ethnic genetic and linguistic diversity, on economic specialization of labor.[26] Although the analysis focuses on the effect of intra-ethnic diversity, it also presents evidence for other potential drivers of economic specialization like environmental diversity, geographically based market potential, and the effect of other geographical endowments.

3.1 Intra-Ethnic Diversity and Economic Specialization (Ordinary Least Squares Analysis)

This subsection explores the statistical relationship between intra-ethnic diversity and economic specialization at the ethnic level. It focuses on 116 ethnic groups for which both genetic and ethnographic data to construct the proposed measure of economic specialization is available. Figure 2(a) shows for

---

[26]The main analysis focuses on genetic diversity as a proxy of intra-ethnic diversity in order to economize space and ease the presentation. Moreover, it should provide, under the identification assumptions discussed in section 2.1, the lower bound on the causal effect of intra-ethnic diversity. Robustness to the proxy of intra-ethnic diversity are included in various parts of the main text and appendices.

these 116 ethnicities the distribution of intra-ethnic diversity for groups above and below the mean economic specialization. Clearly, more specialized groups also have higher intra-ethnic diversity. An illustrative example of the link between diversity and division of labor is given by two societies in the sample: the Konso people of South-Western Ethiopia and the Aché people of Eastern Paraguay. These two ethnic groups are located on both extremes of the sample distribution of the proxies of intra-ethnic diversity, separated by more than five standard deviations from each other. Due to their proximity to the Ethiopian rift valley, Konso's intra-ethnic diversity is among the highest in the world; while the Aché is the less diverse group in the sample of societies analyzed in this research. For thousands of years, both groups inhabited remote locations with little influence from outsiders (Hill and Hurtado, 1996; Hallpike, 2008). The ecological environment for both societies was hard and not particularly rich (Hill and Hurtado, 1996; Freeman and Pankhurst, 2003). The difference in diversity between these two groups maps into differences in their economic specialization of labor. In particular, according to the Ethnographic Atlas, the Konso had labor specialization in 5 activities, whereas the Aché had none.

(a) High/Low Specialization and Diversity

(b) Conditional Association (Binned Scatterplot)
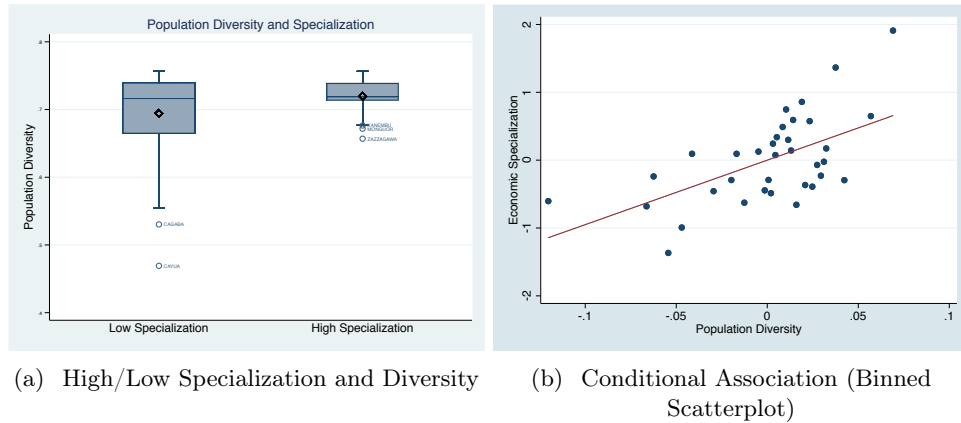
Figure 2: Intra-Ethnic Diversity and Economic Specialization

In order to analyze this relation more systematically, the analysis implements the empirical strategy presented in section 2.1 by exploring variations in equation (4) to estimate δ_1, which as explained above could be interpreted as the lower bound for the causal effect of intra-ethnic diversity if the estimated coefficient on the proxy of intra-ethnic diversity is positive.[27],[28] In order to simplify the exposition, the analysis below refers to δ_1 as the effect of intra-ethnic diversity.

Table 1 analyzes the association between economic specialization and intra-ethnic diversity accounting for a basic set of geographic characteristics of ethnicities' homelands using OLS. In particular, column 1 shows the unconditional relationship between intra-ethnic diversity and economic specialization. The estimated coefficient is statistically significant at the 1 percent level and is consistent with an economically significant effect of intra-ethnic diversity. In particular, a one standard deviation increase in

---

[27]Given space constraints, the results in the body of the paper focus on economic specialization measured by the number of activities that are specialized, i.e., s^1. Section E.1 in the appendix establishes that all results presented in the main body of the paper are robust to the measure of economic specialization employed.

[28]In order to ease the interpretation of the results and compare them across the different specifications presented in this paper, all tables report standardized coefficients. The standard coefficients report the number of standard deviation changes in the dependent variable for a one-standard deviation change in the independent variable.

intra-ethnic diversity is associated with a 0.27 standard deviation increase in economic specialization.

Table 1: Intra-Ethnic Diversity and Economic Specialization

	Economic Specialization						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intra-Ethnic Diversity	0.27***	0.31***	0.27***	0.27***	0.25***	0.27***	0.36***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.09)
Absolute Latitude		0.15					0.80***
		(0.09)					(0.30)
Area			0.01				0.01
			(0.04)				(0.05)
Elevation (Avg.)				-0.03			0.31*
				(0.11)			(0.16)
Precipitation (Avg.)					-0.08		0.13
					(0.09)		(0.16)
Temperature (Avg.)						0.04	0.73***
						(0.08)	(0.25)
Adjusted-R^2	0.07	0.08	0.06	0.06	0.06	0.06	0.15
Observations	116	116	116	116	116	116	116

Notes: This table shows the positive statistically and economically significant correlation between economic specialization and intra-ethnic diversity as measured by expected heterozygosity (i.e., average probability that two randomly chosen individuals in an ethnicity do not share the same allele of a neutral gene) after accounting for a set of basic geographical controls. Economic specialization counts the number of specialized activities present in an ethnicity. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

A potential concern is that intra-ethnic diversity might be capturing the effect of absolute latitude. In particular, technologies and institutions have historically spread more easily across similar latitudes, where climate and the duration of days were not drastically different. Furthermore, the positive high correlation between absolute latitude and development, which has been widely documented in the economic growth and development literature (Spolaore and Wacziarg, 2013), might confound the effect of intra-ethnic diversity. In order to mitigate this potential concern column 2 accounts for the effect of absolute latitude. Although absolute latitude enters positively (albeit statistically insignificantly) in this specification, the effect of intra-ethnic diversity remains highly statistically significant and increases by 10 percent. Column 3 accounts for the total area of the ethnic homeland, since all else equal, larger areas may contain a more diverse population by construction. In particular, cultural assimilation may be more difficult in large territories, thus, contributing to cultural diversity. Additionally, total area may confound the effect of market potential, which is a potential driver of economic specialization.[29] Nonetheless, the inclusion of this control does not affect the estimated effect of intra-ethnic diversity. Column 4 accounts for the effect of mean elevation, which has been suggested may affect the diversity of production specific human capital within groups without affecting the results.

Another potential concern is that intra-ethnic diversity correlates with precipitation and temperature. In particular, it has been shown that both species and cultural diversity are positively correlated with precipitation and net primary productivity, which in turn depends on temperature (Moore et al.,

---

[29]It is worth noting that total area is determined by ethnic homeland borders, which can be arguably endogenous to both diversity and economic specialization.

2002; Nettle, 1998). Furthermore, precipitation and temperature might directly affect economic activities and specialization. Thus, omission of precipitation and temperature might bias the results. Columns 5 and 6 mitigate this potential concern by accounting for average precipitation and average temperature, respectively. As shown in the table, the estimated coefficients on both these controls are negative and not statistically nor economically significant. On the other hand, the coefficient on intra-ethnic diversity remains positive statistically and economically significant. Finally, column 7 accounts for the joint effect of all these basic geographic controls. The statistical relationship between intra-ethnic diversity and economic specialization is statistically significant at the 1 percent level and is consistent with an economically significant effect of intra-ethnic diversity. In particular, an increase of one standard deviation in intra-ethnic diversity is associated with a 0.36 standard deviation increase in economic specialization.

While these results support the proposed hypothesis, the estimated coefficient on intra-ethnic diversity might be biased due to the omission of other variables. In order to mitigate this potential concern, Table 2 accounts for other possible sources of economic specialization and intra-ethnic diversity. In particular, and although the unit of analysis in this paper is different, it is feasible that the geographical covariates suggested by the literature on the emergence of (cross-country) inter-ethnic diversity may affect intra-ethnic diversity through similar mechanisms. Therefore, the analysis accounts for these characteristics, which may have also directly affected economic specialization.

In order to compare with the previous results, column 1 of Table 2 replicates the last result in Table 1. A potential concern is that higher intra-ethnic diversity may be a result of a hostile disease environment. For example, Birchenall (2014) argues that pathogen stress influenced pre-colonial inter-ethnic diversity. Furthermore, a "bad" disease environment can also negatively affect economic activities. Thus, column 2 additionally considers the potential confounding effect of the disease environment by accounting for the ecology of malaria (Kiszewski et al., 2004). As expected, malaria ecology negatively correlates with economic specialization. Given the positive correlation between the disease environment and population diversity, the inclusion of malaria ecology increases the size and statistical significance of the point estimate for intra-ethnic diversity.

Column 3 accounts for the diversity of the ecological environment, which could potentially affect specialization directly (Fenske, 2014) and be correlated with linguistic and cultural diversity (Michalopoulos, 2012; Moore et al., 2002). Reassuringly, although ecological diversity correlates strongly with economic specialization, the point estimate for intra-ethnic diversity is virtually unaltered.[30]

Columns 4 and 5 account for the potentially confounding effects of agricultural and caloric suitability. In particular, Michalopoulos (2012) suggests that variation in soil quality correlates with diversity in production specific human capital, which could foster economic exchange. Moreover, variation in soil quality could potentially be conducive to specialization directly. Additionally, Galor and Özak (2015, 2016) show that pre-industrial population (density) levels are highly correlated with their Caloric Suitability Index (CSI).[31] Since population (density) potentially affects market size and thus specialization

---

[30]Following Fenske (2014), ecological diversity is a Herfindahl index of the shares of each ethnic homeland's area occupied by each ecological type (Olson et al., 2001).

[31]The Caloric Suitability Index (CSI) measures for each cell of 10 kms × 10 kms in the world, the average number of calories that could be potentially produced given the climatic conditions in that cell and the crops available in the

Table 2: Intra-Ethnic Diversity and Economic Specialization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Economic Specialization				
Intra-Ethnic Diversity	0.36***	0.41***	0.36***	0.37***	0.34***	0.33***	0.31***	0.40***	0.31***
	(0.09)	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)	(0.09)	(0.11)	(0.10)
Malaria Ecology		-0.36***							-0.41***
		(0.12)							(0.12)
Ecological Diversity			0.26***						0.20*
			(0.10)						(0.11)
Agricultural Suitability (avg.)				0.00					0.13
				(0.13)					(0.10)
Agricultural Suitability (std.)				0.22*					0.32**
				(0.13)					(0.14)
Caloric Suitability (Pre-1500 ,avg.)					-0.24*				-0.34**
					(0.14)				(0.14)
Caloric Suitability (Pre-1500 ,std.)					0.30**				0.07
					(0.11)				(0.14)
Temperature (Spatial Corr., Avg.)						0.01			-0.05
						(0.09)			(0.08)
Temperature (Volatility, Avg)						-0.58***			-0.11
						(0.19)			(0.20)
Pct. Area within 100kms of Sea							0.02		-0.16
							(0.10)		(0.11)
Coast Length							0.49**		0.60***
							(0.22)		(0.20)
Ruggedness (Avg.)								-0.22	0.07
								(0.22)	(0.18)
Pre-Industrial Mobility (avg.)								0.81*	1.06**
								(0.41)	(0.46)
Pre-Industrial Mobility (std.)								-0.04	-0.36**
								(0.12)	(0.16)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.20	0.27	0.26	0.23	0.24	0.24	0.26	0.23	0.50
Adjusted-R^2	0.15	0.22	0.21	0.17	0.18	0.18	0.20	0.17	0.40
Observations	116	116	116	116	116	116	116	116	116

Notes: This table shows the positive statistically and economically significant correlation between economic specialization and intra-ethnic diversity as measured by measured by expected heterozygosity (i.e., average probability that two randomly chosen individuals in an ethnicity do not share the same allele of a neutral gene) after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

(Smith, 1776), including the mean and the standard deviation of the CSI accounts for this potential confounding channel. Reassuringly, the qualitative results remain unaltered.

Column 6 controls for the confounding effects of both the spatial correlation and the intertemporal volatility of temperature. In particular, Dean et al. (1985) argue that trade alliances among communities were common in regions with high spatial variability in climate. In addition, pre-modern societies could have mitigated the negative impact of climatic variation by extending the set of subsistence activities. Additionally, Ahlerup and Olsson (2012) show that temperature variation predicts inter-ethnic

pre-1500CE period.

diversity. Accounting for these potential confounders does not alter the results.

Columns 7 and 8 account for a potential concern that ethnicities' isolation and access to the sea might jointly affect their intra-ethnic diversity and their economic specialization. In particular, proximity and access to the sea may ease contact with other societies, thus affecting intra-ethnic diversity and facilitating trade. Similarly, isolated ethnicities may be forced to specialize and also be less diverse. However, accounting for the fraction of the ethnic homeland located within 100 kilometers from the sea as well as the length of the ethnic homeland's coastline (Column 7), and for the average ruggedness of the terrain, the average and the standard deviation of the pre-industrial mobility index developed by Özak (2018, 2010) does not alter the qualitative results.

Finally, column 9 accounts for the joint effect of all the previous confounders. The estimated coefficient on intra-ethnic diversity remains positive statistically and economically significant. In particular, a one standard deviation increase in intra-ethnic diversity is associated with an increase of 0.31 standard deviations in economic specialization. Figure 2(b) depicts this conditional association using a binned scatterplot. While these results support the proposed hypothesis, suggesting that intra-ethnic diversity has a positive effect on economic specialization, the point estimates reported so far may still be biased due to unobservable factors that correlate with both the proxy of intra-ethnic diversity and economic specialization. To mitigate this concern, the next sections follow the instrumental variable approach discussed in section 2.1.

3.2 Intra-Ethnic Diversity and Distance to East Africa

This section establishes the negative statistically and economically significant predictive power of the migratory distance from East Africa for intra-ethnic diversity as proxied by genetic and linguistic diversity. In particular, the Out-of-Africa theory predicts that intra-ethnic diversity decreases along the different migratory routes that humans followed out of East Africa (Ashraf and Galor, 2013b). Thus, as suggested by the empirical strategy in section 2.1, the analysis estimates the effect of the migratory distance to East Africa (β_{ij}, $j = genetic, linguistic$ in equation (2)) on the proxies of intra-ethnic diversity. Moreover, it empirically establishes that $\beta_{1genetic} < \beta_{1linguistic}$ providing supporting evidence for the interpretation of δ_1 as a lower bound for the effect of intra-ethnic diversity.

The analysis estimates the pre-industrial migratory distance to East Africa by finding the minimal travel times to East Africa (Addis Ababa) using the Human Mobility Index with Seafaring - HMISea (Özak, 2010, 2018).[32] HMISea estimates the time (in weeks) required to walk across each square kilometer of land, accounting for the topographic, climatic, terrain conditions, and human biological abilities, as well as the time required to cross major seas with pre-industrial technologies. Figure 3 shows the potential migratory routes out of East Africa to the historical ethnic homelands that minimize the travel time according to HMISea. To mitigate potential concerns of endogeneity of the actual historical patterns of migration, the analysis employs the pre-industrial travel time to the ethnic homeland based on HMISea as an instrument for the proxies of intra-ethnic diversity.

---

[32]Appendix F.5 shows the robustness of the analysis to using geodesic or great circle distances. Nonetheless, as established there, migratory distances constructed using HMISea are more fundamental. Specifically, they have larger explanatory power and when accounting for HMISea, the other measures become statistically insignificant.
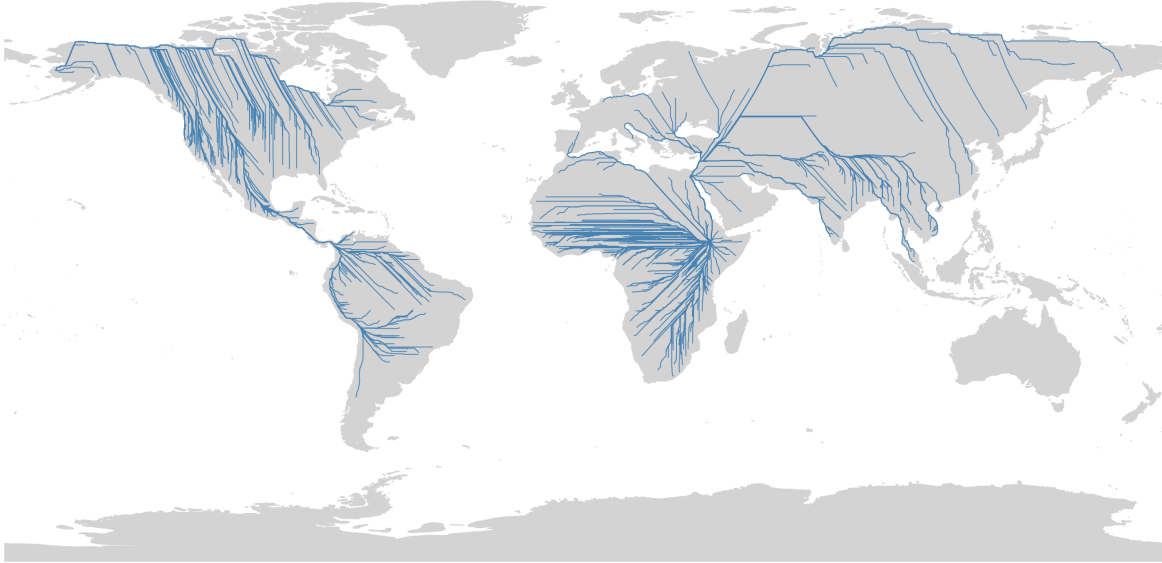
Figure 3: Potential Migratory Routes Out of East Africa

Table 3 explores the relationship between migratory distance to East Africa and genetic diversity (as measured by expected heterozygosity) for 144 ethnic groups for which geo-coded genetic and ethnographic data is available.[33] Two facts stand out from the results in Table 3: (i) migratory distance to East Africa alone explains 72 percent of the variation in intra-ethnic diversity (column 1); and (ii) accounting for the potential confounding effects of all the controls included in Tables 1 and 2, both individually and jointly, affects remarkably little the point estimates for pre-industrial migratory distance to East Africa. Furthermore, as shown in column 9, these results hold also for the restricted sample of 116 ethnic groups from previous section.

The importance of effect of the distance to East Africa on genetic diversity is further confirmed by its semi-partial R^2. In particular, the distance to East Africa has the largest semi-partial R^2 in the analysis. As shown in Table 3, the semi-partial R^2 of the distance to East Africa, is even larger than the sum of the semi-partial R^2's of all other controls combined. Specifically, the results in column (8) imply that the variation that is uniquely related to the distance to East Africa, explains 30% of the total variation in genetic diversity, while the combined variation that is specific to the each of other variables explains less that 5% of the total variation in genetic diversity. Thus, the distance to East Africa is the main explanatory variable of genetic diversity in our analysis.

Finally, the negative relation between the pre-industrial distance to East Africa and intra-ethnic diversity is further confirmed in Table 4, which shows the relation between this distance and 3 measures of linguistic diversity. Figure 4 depicts the unconditional and conditional strong negative relationship between all the proxies of intra-ethnic diversity employed in the analysis and the pre-industrial migratory distance to East Africa. As hypothesized, $\beta_{1genetic} < \beta_{1linguistic}$, thus providing support for

---

[33]Similar results are obtained in the full sample of 267 ethnicities for which genetic data alone is available. The analysis omits islands for which the HMISea does not provide travel speed estimates. Still, the results are robust to imputation based on geodesic distances or by using the HMIOcean measure, which includes more advanced navigation technologies available before the invention of the steam engine.

Table 3: Intra-Ethnic Diversity and Distance to East Africa

	Intra-Ethnic Diversity (Genetic)								
	Full Sample								Specia-lization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance to East Africa	-0.85*** (0.07)	-0.80*** (0.10)	-0.80*** (0.09)	-0.81*** (0.09)	-0.80*** (0.10)	-0.79*** (0.10)	-0.80*** (0.10)	-0.82*** (0.09)	-0.85*** (0.09)
Malaria Ecology			0.17*** (0.05)					0.20*** (0.07)	0.16** (0.08)
Agricultural Suitability (avg.)				-0.00 (0.06)				-0.01 (0.07)	-0.01 (0.07)
Agricultural Suitability (std.)				0.08 (0.06)				0.13** (0.06)	0.14* (0.07)
Caloric Suitability Index (Pre-1500CE)					0.02 (0.07)			0.07 (0.07)	0.09 (0.09)
Caloric Suitability (Pre-1500 ,std.)					-0.08 (0.05)			-0.13** (0.06)	-0.13* (0.07)
Pct. Area within 100kms of Sea						-0.00 (0.06)		0.13** (0.06)	0.14** (0.07)
Coast Length						0.03 (0.06)		0.04 (0.05)	0.01 (0.07)
Ruggedness (Avg.)							-0.03 (0.12)	-0.19 (0.13)	-0.19 (0.16)
Pre-Industrial Mobility (avg.)							0.05 (0.20)	0.12 (0.23)	0.13 (0.25)
Pre-Industrial Mobility (std.)							-0.13 (0.08)	-0.07 (0.11)	-0.07 (0.12)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semi-partial R^2 of Pre-Industrial Distance	0.72	0.40	0.39	0.40	0.36	0.38	0.36	0.28	0.33
Sum of Semi-partial R^2 of All Other Controls	0.00	0.01	0.02	0.02	0.02	0.01	0.00	0.05	0.05
Adjusted-R^2	0.72	0.72	0.74	0.72	0.72	0.72	0.73	0.75	0.73
Observations	144	144	144	144	144	144	144	144	116

Notes: This table shows the negative statistically and economically significant relation between intra-ethnic diversity, measured by expected heterozygosity (i.e., average probability that two randomly chosen individuals in an ethnicity do not share the same allele of a neutral gene), and the distance to East Africa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

the interpretation of the effect of genetic diversity, δ_1, as a lower bound of the effect of intra-ethnic diversity (Proposition 3).

3.3 Intra-Ethnic Diversity and Economic Specialization (Instrumental Variable Analysis)

This section provides further supportive evidence for the hypothesized positive effect of intra-ethnic diversity on economic specialization by exploiting an instrumental variable strategy based on the migratory distance to East Africa (Ashraf and Galor, 2013b). As shown in the previous section, the

Table 4: Intra-Ethnic Diversity and Distance to East Africa

	Intra-Ethnic Diversity (Linguistic)								
	Consonant Inventory			Vowel Quality Inventory			Number of Genders		
	Full Sample		Specialization	Full Sample		Specialization	Full Sample		Specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance to East Africa	-0.42***	-0.37***	-0.36***	-0.30***	-0.31***	-0.33***	-0.25***	-0.28***	-0.30***
	(0.05)	(0.06)	(0.07)	(0.06)	(0.07)	(0.08)	(0.06)	(0.09)	(0.08)
Main Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Additional Geographical Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Adjusted-R^2	0.17	0.27	0.27	0.08	0.19	0.22	0.05	0.20	0.23
Observations	299	299	255	301	301	256	152	152	131

Notes: This table shows the negative statistically and economically significant relation between measures of linguistic diversity and the distance to East Africa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Each column includes the same set of controls as the same column in Table 3. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

(a) Genetic Diversity (Unconditional) (b) Genetic Diversity (Conditional) (c) Linguistic Diversity (Consonant, Unconditional) (d) Linguistic Diversity (Consonant, Conditional)

(e) Linguistic Diversity (Vowel, Unconditional (f) Linguistic Diversity (Vowel, Conditional) (g) Linguistic Diversity (Genders, Unconditional) (h) Linguistic Diversity (Genders, Conditional)

Figure 4: Pre-Industrial Distance to East Africa and Intra-Ethnic Diversity

migratory distance to East Africa is a plausible instrument for various proxies of intra-ethnic diversity, since (i) it is the main predictor of intra-ethnic diversity, due to the serial founder effect and the Out-of-Africa theory, and (ii) it plausibly only affects economic outcomes through its effect on diversity.[34]

Table 5 presents the results of this instrumental variables (IV) analysis, in which intra-ethnic diversity as proxied by genetic diversity is instrumented by the migratory distance to East Africa for

---

[34]Sections 4-5 present additional evidence in support of the plausibility of the exclusion restriction.

Table 5: Intra-Ethnic Diversity and Economic Specialization (IV)

	OLS	IV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intra-Ethnic Diversity	0.36***	0.51***	0.53***	0.54***	0.54***	0.45***	0.49***	0.44***	0.56***	0.46***
	(0.09)	(0.12)	(0.12)	(0.12)	(0.13)	(0.14)	(0.13)	(0.11)	(0.14)	(0.14)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All
First-stage F-statistic		56.99	59.31	59.04	65.63	52.61	55.27	53.29	63.44	81.54
Adjusted-R^2	0.15	0.14	0.21	0.19	0.15	0.17	0.16	0.19	0.15	0.39
Observations	116	116	116	116	116	116	116	116	116	116

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity and economic specialization, by instrumenting intra-ethnic diversity (measured by expected heterozygosity i.e., average probability that two randomly chosen individuals in an ethnicity do not share the same allele of a neutral gene) with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

the set of 116 ethnicities for which genetic, ethnographic and geographic data exists. In order to facilitate comparison with the OLS results, column 1 replicates the analysis of column 5 in Table 1 by accounting for the effect of the set of basic geographic controls. Columns 2 through 10 use this IV strategy to explore the effect of intra-ethnic diversity on economic specialization, accounting for the set of controls of Table 2. The estimated effect is 22-55% larger than in the OLS analysis, and ranges between 0.44 and 0.56, suggesting an economically significant effect of intra-ethnic diversity on economic specialization.[35] In particular, after accounting for all the confounders analyzed in table 2, a one standard deviation increase in intra-ethnic diversity is associated with about half a standard deviation increase in economic specialization. Importantly, these results are not subject to a weak instrument problem, since the Kleibergen-Paap F-statistics for the first stage, reported at the bottom of the table, are all larger than the critical values suggested by Stock-Yogo.

Additionally, Table 6 establishes the robustness of the results to the measure of intra-ethnic diversity by replicating the main results of Table 5 for each of the three linguistic proxies of intra-ethnic diversity.[36] In particular, proxying intra-ethnic diversity with linguistic diversity as measured by consonant inventory, vowel quality inventory and the number of genders generates qualitatively similar results. The results of Tables 5 and 6 show some noteworthy patterns. First, regardless of the intra-ethnic diversity measure employed, the IV estimates are larger than the OLS estimates, suggesting a potential downward bias due to measurement error, omitted factors or reverse causality. Also, the IV point estimates may reflect a local average treatment effect due to the existence of imperfect compliers of the instrument (Arbatlı et al., 2020). Second, as could be expected, the ratio of IV to OLS estimates

---

[35]While the point estimates are different, we cannot reject the null hypothesis that the OLS and IV estimates are equal to each other under standard levels of confidence. In particular, comparing the OLS estimates in column 1 of Table 2 to the equivalent (i.e., same specification) IV estimates in column 2 in Table 5 generates a Chi-square statistic of 2.03 with a p-value of 0.1545. Similarly, if we compare the OLS estimate for our most demanding specification (column 9 in Table 2) to the equivalent IV (column 10 in Table 5), we get a Chi-square statistic of 0.93 with a p-value of 0.3340.

[36]Tables H.5-H.8 fully replicate Table 5 for each of the linguistic proxies of intra-ethnic diversity.

Table 6: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity

	Economic Specialization (Count)								
	Linguistic Diversity (Consonant Inventory)			Linguistic Diversity (Vowel Quality Inventory)			Linguistic Diversity (Number of Genders)		
	OLS	IV		OLS	IV		OLS	IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intra-Ethnic Diversity Diversity	0.20***	1.15***	1.25***	0.39***	1.16***	1.30***	0.13*	1.22***	1.00***
	(0.06)	(0.20)	(0.29)	(0.06)	(0.25)	(0.34)	(0.08)	(0.28)	(0.37)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Geographical Controls	No	No	Yes	No	No	Yes	No	No	Yes
First-stage F-statistic		45.92	26.47		22.75	18.43		29.40	14.47
Observations	255	255	255	256	256	256	131	131	131

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity and economic specialization, by instrumenting intra-ethnic diversity (as proxied by linguistic characteristics of the language spoken by an ethnicity) with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.
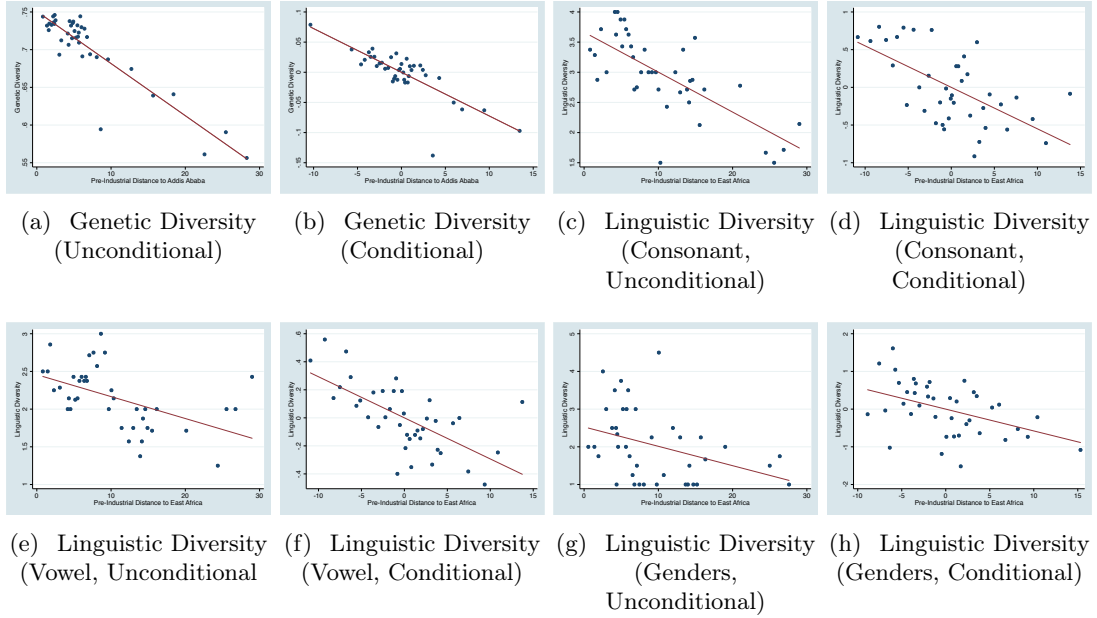
is larger for the linguistic measures of intra-ethnic diversity, since measurement error and factors other than the serial founder effect, such as cultural evolution (Galor et al., 2016), may affect these measures more. And third, the standardized beta from the IV estimates for the linguistic measures are larger than the one based on genetic diversity. These results and the ones in the previous section, provide further support to the interpretation of the estimated effect of genetic diversity as a lower bound of the true effect of intra-ethnic diversity.

Finally, the results are robust to the measure of economic specialization used (see section 2.2 for the construction of the different measures). In particular, employing the alternative measures of economic specialization generates qualitatively identical results and imply a positive effect of intra-ethnic diversity on economic specialization (Tables E.2 and E.1). Moreover, varying both the measure of economic specialization and the proxy of intra-ethnic diversity does not affect the results either (Tables H.9 and H.10).

3.4   Predicted Intra-Ethnic Diversity and Economic Specialization

This section provides additional support for the positive effect of intra-ethnic diversity on economic specialization. In particular, a potential concern with the previous analysis is that it is based on samples of ethnicities, for which both intra-ethnic diversity (either genetic or linguistic) and economic specialization data is available, which could be a potential source of bias. In order to mitigate this potential concern and further explore this effect, this section employs a two-step econometric model to generate a measure of intra-ethnic diversity as predicted by the pre-industrial migratory distance to East Africa (Murphy and Topel, 2002; Ashraf and Galor, 2013b).[37] In particular, based on the estimated

---

[37]Two-step econometric procedures yield consistent estimates of second stage parameters, although the second-step standard error estimates may be incorrect, if they do not account for the additional uncertainty due to the two-step

relation between the migratory distance to East Africa and intra-ethnic diversity in the subsamples of ethnicities analyzed in section 3.2, the analysis predicts intra-ethnic diversity for all ethnicities in the Ethnographic Atlas.[38] This strategy expands roughly eightfold the sample of ethnicities for which the main proxy of intra-ethnic diversity and specialization data is available. Moreover, it allows the analysis to be performed on additional ethnographic data. Finally, as in the case of the previous IV approach, the estimated coefficient on predicted intra-ethnic diversity can be given a causal interpretation, since by construction it captures only the exogenous variation in diversity generated by the serial founder effect and the Out-of-Africa theory. Since this analysis exploits a generated regressor, standard errors are computed using a bootstrapping procedure.[39]

Table 7: Predicted Intra-Ethnic Diversity and Economic Specialization

	Economic Specialization										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Predicted Intra-Ethnic	0.44***	0.42***	0.49***	0.42***	0.41***	0.46***	0.40***	0.42***	0.42***	0.53***	0.59***
Diversity	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.07)	(0.21)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All	All
Continental FE	No	No	No	No	No	No	No	No	No	No	Yes
Semi-partial R^2 of Intra-Ethnic Diversity											
	0.19	0.15	0.16	0.15	0.13	0.16	0.12	0.14	0.14	0.14	0.13
Sum of Semi-partial R^2 of All Other Controls											
	0.00	0.04	0.06	0.05	0.04	0.05	0.05	0.04	0.03	0.08	0.08
Adjusted-R^2	0.19	0.22	0.23	0.24	0.22	0.24	0.22	0.21	0.23	0.29	0.32
Observations	934	934	934	934	934	934	934	934	934	934	934

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity, as predicted by the distance to East Africa (see section 3.2), and economic specialization. These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Bootstrapped standard error estimates in parenthesis; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Based on this extended sample, the analysis replicates in columns 1 to 10 of Table 7 the main econometric specifications of Tables 1, 2, and 5. The strong positive association between intra-ethnic diversity and economic specialization remains statistically and economically significant. Furthermore, the point estimates are remarkably stable across specifications, supporting the view that the potential effect of predicted intra-ethnic diversity is not biased by omitted factors (i.e., selection on unobservables). Moreover, the size of the estimated coefficient on intra-ethnic diversity in this expanded sample lies between the OLS and IV estimates of the reduced sample (see Tables 2 and 5).

procedure (Murphy and Topel, 2002). In order to address this issue, the analysis employs a bootstrapping procedure to correctly estimate standard errors.

[38]Given that the analysis exploits a unique source of variation, the predicted intra-ethnic diversity measures generated based on genetic or linguistic diversity are perfectly correlated. Indeed, they reflect a change in the scale of the measure of intra-ethnic diversity. Thus, the analysis is performed based on the predicted intra-ethnic (genetic) diversity, given the stronger predictive power of the SFE for this measure. The results are similar if instead the other proxies are used.

[39]In particular, a random sample of ethnicities with both diversity and migratory distance data is drawn with replacement out of the original sample. Then equation (2) is re-estimated, accounting for the same set of controls as in the second-stage. Using these new estimates intra-ethnic diversity is predicted again and equation (4) is re-estimated. This procedure is repeated 1001 times and the distribution of the bootstrapped coefficients is used to compute the standard errors. A similar procedure was proposed in Ashraf and Galor (2013b).

Column 11 establishes that the documented relationship is robust to accounting for unobserved time-invariant continent-specific attributes. Indeed, if anything, the inclusion of continental fixed effects increases the estimated coefficient on diversity.[40] In particular, the estimates in columns 10 and 11 imply that a standard deviation increase in predicted intra-ethnic diversity is associated with more than a half of a standard deviation increase in economic specialization.[41] Figure 5 depicts the conditional relations in columns (10) and (11) using binned scatter plots.

(a) Without Continental FE   (b) With Continental FE
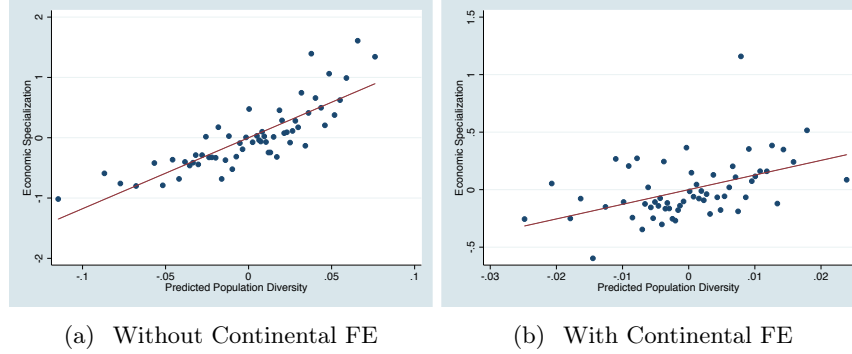
Figure 5: Predicted Intra-Ethnic Diversity and Economic Specialization
Conditional Relations Accounting for All Geographical Controls

Table 7 also establishes the relative importance of intra-ethnic diversity as a predictor of economic specialization. Indeed, it shows that among all the covariates studied, predicted intra-ethnic diversity has the highest semi-partial R^2. Thus, the variation uniquely associated with intra-ethnic diversity is larger than the one associated with any other single predictor of economic specialization. Furthermore, it establishes that all other predictors jointly have a lower explanatory power than intra-ethnic diversity alone. This provides additional supportive evidence for the fundamental role of intra-ethnic diversity as a determinant of economic specialization.

3.5 Complementary Effect of Intra-Ethnic and Geographical Diversity on Economic Specialization

This section explores whether, as hypothesized, diverse populations enjoy complementarities with diverse geographical and ecological endowments. In particular, the hypothesized effect of intra-ethnic diversity on economic specialization may be larger in locations with diverse geography, given that diverse preferences or skills could potentially allow diverse endowments and ecologies to be exploited better and, thus, generate higher levels of economic specialization.

Table 8 analyzes the potential complementarity between population and various measures of geographical diversity by analyzing the heterogenous association of intra-ethnic diversity with economic

---

[40]Table F.1 shows the robustness of these results to accounting for continental fixed effects in all columns. It establishes that the coefficients are larger in all specifications when accounting for continental fixed effects.

[41]Table G.2 shows the point estimates of the reduced form economic specialization-distance to East Africa for all the specifications in Table 7. The point estimates for pre-industrial distance to East Africa are remarkably stable and strongly statistically significant. Indeed, the stability of the point estimates suggests that selection on unobservables is unlikely to drive the results, thus providing supportive evidence for the plausible exogeneity of the instrument.

Table 8: Predicted Intra-Ethnic Diversity and Economic Specialization
Complementarity with Heterogeneous Environments

	Economic Specialization					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Intra-Ethnic Diversity	0.59***	0.49***	0.45***	0.46***	0.55***	0.54***
	(0.25)	(0.26)	(0.26)	(0.30)	(0.26)	(0.28)
Predicted Intra-Ethnic Diversity $\times$ Ecological Diversity		0.80*** (0.38)				
Predicted Intra-Ethnic Diversity $\times$ Precipitation (Volatility, Std.)			1.16** (0.57)			
Predicted Intra-Ethnic Diversity $\times$ Temperature (Spatial Corr., Std.)				0.70* (0.65)		
Predicted Intra-Ethnic Diversity $\times$ Precipitation (Spatial Corr., Std.)					0.77** (0.43)	
Predicted Intra-Ethnic Diversity $\times$ Ruggedness (Avg.)						1.08** (0.59)
Main Controls & Main Effects	Yes	Yes	Yes	Yes	Yes	Yes
All Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.32	0.33	0.33	0.34	0.34	0.33
Observations	934	934	934	934	934	934

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity, as predicted by the distance to East Africa (see section 3.2), and economic specialization. Additionally, it establishes the heterogeneity of association, suggesting a complementarity between intra-ethnic diversity and variations in environmental and geographical factors. These results are robust to accounting for the set of geographical controls of Table 7, continental fixed effects and an extended set of geographical diversity measures (i.e., main/level effects and interactions). Standardized coefficients. Bootstrapped standard error estimates in parenthesis; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

specialization. In particular, it shows the main potential effect of intra-ethnic diversity and its interaction with ecological diversity, the standard deviation of agricultural suitability, temperature volatility, the standard deviation of ruggedness of the terrain, and the standard deviation of pre-industrial mobility.[42] As can be seen there, all main hypothesized effects and interactions of intra-ethnic diversity are positive and highly statistically and economically significant. The estimates imply that the more diverse a population and the more diverse the geography in which it lives, the higher the level of economic specialization.

This result lends support for the hypothesis that diverse populations leverage diverse geographical endowments, potentially generating larger complementarities and thus increasing economic specialization. Additionally, it provides a link between the seemingly contradictory theories based on the composition of the population (Ashraf and Galor, 2013a,b) and those based on geographical factors (Galor and Özak, 2015, 2016; Galor et al., 2016). In particular, it gives an explanation as to why economies with similar populations or environments might have different economic outcomes.

---

[42]The estimated coefficients are again reported as standardized betas, which simplifies the comparison of the main effects across tables. Of course, this makes the interpretation of the interactions difficult, but given that both main effects and interactions are positive, the qualitative nature of the effects is directly observable from the table.

4 Intra-Ethnic Diversity, Economic Specialization and Pre-modern Economic Development

In light of the positive effects of intra-ethnic diversity on economic specialization and the beneficial effects of the latter on economic development (Depetris-Chauvin and Özak, 2015, 2016), as well as the the links between measures of genetic diversity and development outcomes featured in previous research (Ashraf and Galor, 2013b; Arbatli et al., 2018; Ashraf and Galor, 2018), it is important to analyze the potential interrelations among these effects of intra-ethnic diversity. Specifically, this section delves deeper into the potential relations between intra-ethnic diversity and economic specialization and various development outcomes. We are especially interested in exploring whether the estimated effect of intra-ethnic diversity on economic specialization reflects a true direct effect of diversity on specialization or instead some indirect causal path between these variables (Figure D.1).

If economic specialization and other development outcomes (e.g., population density, urbanization, levels of hierarchy and state) had no causal effect on each other (Figure D.1(a)), our previous estimates would reflect the true direct effect of diversity on economic specialization. Of course, given our theoretical hypothesis and previous theoretical and empirical results in the literature, such a scenario is very unlikely. A second possibility is that economic specialization is caused by other development outcomes (Figure D.1(b)), but not the other way around. In such a scenario, a potential concern with our main analysis is that intra-ethnic diversity may have promoted economic development and through this channel affected economic specialization indirectly, while having no direct impact on it. Although this scenario would not affect the econometric validity of our analysis, which would estimate the correct reduced form relation between intra-ethnic diversity and economic specialization, since intra-ethnic diversity would still be the underlying determinant of economic specialization, the estimates would reflect the effect of diversity on development and of development on specialization.

In order to mitigate this concern that intra-ethnic diversity has no direct effect on economic specialization, in Table 9 we explore the robustness of the estimated association between intra-ethnic diversity and economic specialization to accounting for the potential mediating effect of economic development. Specifically, we account for an ethnicity's level of population density (column 1), the mean size of its local communities (column 2), its level of jurisdictional hierarchy at the local level (column 3), its level of class stratification (column 4), or its level of statehood (column 5), which have been previously used to proxy for an ethnicity's level pre-modern economic development. In Panel A we replicate our main specification on the subsample of ethnicities for which we have data on these development outcomes. The estimated average total effect of diversity (Pearl, 2001; Lange et al., 2012; Acharya et al., 2016), which captures the direct and indirect effects of diversity on specialization, remains economically and statistically significant in these subsamples. In Panel B, we estimate the average natural direct effect of diversity (Pearl, 2001; Lange et al., 2012; Acharya et al., 2016), which reflects only the direct impact of diversity on specialization after accounting for the mediating effect of development. The average natural direct effect of diversity is estimated by accounting for these measures of development in our regression. The estimated average natural direct effect of diversity is quite large, suggesting that diversity does in fact have a direct impact on specialization even after accounting for the mediating effect

Table 9: Predicted Intra-Ethnic Diversity, Pre-Modern Development and Economic Specialization

	Economic Specialization				
	Log[Popula-tion Density, 1500CE]	Mean Size of Local Communities	Jurisdictional Hierarchy of Local Community	Class Stratifica-tion	Statehood Level
	(1)	(2)	(3)	(4)	(5)
	Panel A: Average Total Effect of Diversity				
Predicted Intra-Ethnic Diversity	0.56***	0.83***	0.59***	0.68***	0.62***
	(0.18)	(0.22)	(0.17)	(0.18)	(0.17)
Adjusted-R^2	0.33	0.34	0.33	0.33	0.33
	Panel B: Average Natural Direct Effect of Diversity				
Predicted Intra-Ethnic Diversity	0.56***	0.61***	0.71***	0.57***	0.39***
	(0.18)	(0.19)	(0.18)	(0.17)	(0.15)
Adjusted-R^2	0.33	0.51	0.34	0.36	0.48
	Panel C: Average Controlled Direct Effect of Diversity				
Predicted Intra-Ethnic Diversity	0.68***	0.76***	0.88***	0.71***	0.48***
	(0.22)	(0.23)	(0.21)	(0.21)	(0.18)
Adjusted-R^2	0.31	0.24	0.31	0.27	0.20
Main Controls	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes
Development Control	Yes	Yes	Yes	Yes	Yes
Observations	910	509	923	879	912

Notes: This table explores the direct and indirect effects of intra-ethnic diversity on economic specialization. Panel A shows the average total effect of intra-ethnic diversity on economic specialization, which is based on our preferred specification in Table 7, for the subsamples for which pre-modern development measures are available. Panel B estimates the average natural direct effect of intra-ethnic diversity, which is the estimated effect after accounting for the potential mediating development channel (Pearl, 2001; Lange et al., 2012). Panel C estimates the average controlled direct effect of intra-ethnic diversity, which is the estimated effect of diversity on economic specialization after partialling out the potential mediating development channel (Pearl, 2001; Acharya et al., 2016). Each column explores one potential development channel. These results account for the full set of controls in Table 7. Standardized coefficients. Bootstrapped standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

of development. In fact, the estimated coefficients in columns (1)-(4) are quantitatively similar to the estimate in the main analysis.

A potential concern with the estimated natural direct effects is that they may be subject to post-treatment bias due to the inclusion of bad controls (Angrist and Pischke, 2009). Specifically, it is known that simply conditioning on mediating factors (development outcomes) that themselves are affected by the treatment (intra-ethnic diversity) or some other confounder that affects both, may lead to substantial biases in the estimation of direct effects (Angrist and Pischke, 2009; Acharya et al., 2016). In order to address this concern, in Panel C of Table 9 we follow Acharya et al. (2016) and use their two-stage econometric procedure to estimate the average controlled direct effect of diversity, which represents the (direct) causal effect of diversity on economic specialization when the mediating development outcome is held fixed at a particular level.[43] Reassuringly, the estimated average controlled

---

[43]Acharya et al. (2016) prove that their sequential g-estimation method eliminates post-treatment bias. To estimate

direct effect of diversity on specialization remains statistically and economically significant and is always larger than its average natural direct effect. Moreover, in some cases the average controlled direct effect is estimated to be larger (although not always statistically different) from the average total effect.[44] These results suggest that economic specialization has a large direct impact on economic specialization, with only a small fraction of its total effect being potentially mediated by these development outcomes.

While these results support our hypothesis in the case when diversity causes both specialization and development, and the latter causes the former (Figure D.1(b)), they may be biased if instead diversity causes both specialization and development, and the former causes the latter (Figure D.1(c)). Specifically, in the this last case, accounting for development outcomes, as is done in the estimation of average natural direct effects in Panel B, would potentially generate collider bias (Pearl, 2000). Nonetheless, in this case our main results would already reflect the direct effects of diversity on specialization, making the analyses in Table 9 unnecessary.

Thus, these analyses suggest that intra-ethnic diversity has a large positive direct impact on economic specialization, if specialization and these development outcomes do not cause each other. But, it is feasible that causality actually runs simultaneously in both directions. If that were the case, then these development outcomes would be econometrically endogenous and the estimates in Panels B and C in Table 9 may be biased. To explore this possibility further, Panels A-C in Table D.2 replicate the analysis, but focus on estimating the direct effect of intra-ethnic diversity and the potential mediating effect of economic specialization on these various development outcomes. By exploring these reverse causal paths in the cases depicted in Figures D.1(a)-D.1(c) we can gauge the feasibility of the potential mutual causation between specialization and development. The results are quite mixed across these measures of development. First, the total and direct effects of intra-ethnic diversity on these development outcomes have different signs depending on the measure of development considered. This can be expected due to the different effects diversity may have on development (Ashraf and Galor, 2013b). It also provides supporting evidence for the mechanism underlying the fact that the total effect of diversity on specialization is smaller than its direct effect (as suggested above). Second, there is a significant direct effect of diversity on three of these development outcomes, and in most cases a substantial part of the effect seems to be mediated by economic specialization. These results provide support for our previous results, but also suggest that although causality may seem to run mostly from specialization to development, there still remains some concern due to dual causation.

In order to address this concern, one could use an additional instrument in order to provide exogenous variation in one of these endogenous variables. Nonetheless, it is difficult to find an instrument based on theoretical arguments, which affects economic specialization *without* having a potential direct effect on development or the other way around. So, in order to explore this issue in further detail, we follow Depetris-Chauvin and Özak (2015, 2016) and exploit a second instrumental variable strat-

the average controlled direct effect Acharya et al. (2016) suggest the following two-step procedure: First, estimate the same regression used to estimate the average natural effect. Then, demidiate the outcome by subtracting the estimated effect of the bad control in this regression from the outcome variable. Finally, estimate the average controlled direct effect by using the new demidiated outcome variable on the basic set of controls without the mediator.

[44]The average controlled direct effect can be larger than the average total effect if either the association between treatment (diversity) and mediator (development outcome) or between mediator (development outcome) and outcome (specialization) has the opposite sign of the association between treatment (diversity) and outcome (specialization).

egy based on "atheoretical" instrumental variables suggested by Lewbel (2012). This strategy exploits the moment conditions in the cross section of ethnicities to identify the structural parameters in the absence of traditional identifying information.[45] Panels D and E of Tables D.1 and D.2 estimate the average natural and controlled direct effects using this strategy. The results suggest that even in the case when specialization and development cause each other, intra-ethnic diversity has a large direct effect on economic specialization. Moreover, the results provide suggestive evidence that economic specialization is a crucial mechanism linking intra-ethnic diversity and economic development in the pre-modern era.

The analysis in Table 9 suggests that an ethnicity's level of statehood seems to play a major mediating role, lowering the estimated effect of diversity on economic specialization between 20-35%, which may reflect the hypothesized role of states in the emergence of economic specialization. Nonetheless, examples of highly centralized societies without division of labor are virtually absent in the anthropological, archeological and historical literature on pre-modern societies. On the contrary, several examples of stateless pre-modern societies having a noticeable division of labor suggest that statehood was not a necessary precondition for economic specialization. In particular, the Konso people of Ethiopia have a high degree of specialization without any level of jurisdictional hierarchy above the local level. Similarly, the Karen people, who are a culturally and linguistically diverse and historically stateless society that have traditionally traded cotton, forest products, and domestic animals (Hinton, 1979).

Table 10 further explores whether the existence of centralized institutions underlies the estimated association between intra-ethnic diversity and economic specialization. In particular, the table replicates the analysis for the sample of ethnicities with and without a centralized state (Fenske, 2014). This ethnographic measure has been considered the main indicator of the strength and importance of institutions in pre-colonial times (Michalopoulos and Papaioannou, 2013; Gennaioli and Rainer, 2007; Fenske, 2014). As can be seen in Table 10 the positive effect of intra-ethnic diversity on economic specialization does not seem to be mediated by the existence of a State. Moreover, the fact that intra-ethnic diversity generates economic specialization in the sample of ethnicities that do not have a state, suggests that the existence of a state may not have been a necessary precondition for the emergence of economic specialization.

5 Robustness

The results in the previous section lend credence to the hypothesis that intra-ethnic diversity is conducive to the emergence and prevalence of pre-modern division of labor across ethnic groups. Nevertheless, the results conceivably might be biased due to the measure of economic specialization and the econometric method employed in the analysis, as well as due to omitted historical confounders, the

---

[45]This instrumental variable strategy follows in the spirit of Arellano and Bond (1991) and Blundell and Bond (1998), who also generate "atheoretical" instruments using moment conditions in a dynamic panel data setting. While the identification in Arellano and Bond (1991) and Blundell and Bond (1998) comes from temporal variations, Lewbel (2012) bases the identification on the heteroskedastic structure of residuals obtained in an auxiliary regression of the one endogenous variable on the set of exogenous covariates included in the model. See Lewbel (2012) and Depetris-Chauvin and Özak (2016) for details.

Table 10: Predicted Intra-Ethnic Diversity and Economic Specialization
Ethnicities with and without Centralized States

| | Economic Specialization | | | | | |
| | No Centralized State | | | Any Centralized State | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Intra-Ethnic Diversity	0.46***	0.39***	0.50***	0.30***	0.36***	0.40***
	(0.03)	(0.04)	(0.10)	(0.04)	(0.05)	(0.08)
Main Controls	No	Yes	Yes	No	Yes	Yes
Additonal Controls	No	No	Yes	No	No	Yes
Adjusted-R^2	0.21	0.24	0.30	0.09	0.14	0.24
Observations	433	433	433	479	479	479

Notes: This table suggests that the positive statistically and economically significant association between intra-ethnic diversity, as predicted by the distance to East Africa (see section 3.2), and economic specialization is not mediated by the existence of a (pre-modern) State. These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Bootstrapped standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

existence of spatial or cultural dependence, or the presence of inter-ethnic interactions. This section explores the robustness of the results to these and other potential concerns.

5.1 Alternative Measures of Specialization, Econometric Specification and Multiple Hypotheses Testing

This section explores the robustness of the estimated association between predicted intra-ethnic diversity and economic specialization to the measure of specialization used, the estimation method employed and multiple hypotheses testing. In particular, Table E.1 establishes that the main results presented so far hold for all three measures of economic specialization constructed in section 2 (see also Table E.2). Interestingly, intra-ethnic diversity's implied effect on economic specialization is stronger when the measure of economic specialization is based on the share of activities that are specialized. Since this measure should be the less affected by any potential confounding effect of development, this result suggests that other sources of development are not likely driving the estimated effect of intra-ethnic diversity.

Another potential concern with these results is that they may be driven by the empirical specification chosen. In particular, the main measure of economic specialization is a count variable, while another is a share, which may cause OLS estimates to be biased. On the other hand, the results may be biased given the large number of societies that do not have economic specialization. In order to mitigate these concerns, the research replicates the analysis employing Poisson, Negative Binomial, and fractional regression methods, as well as their zero-inflated variants and zero-inflated beta regressions (see Appendix E.2). Reassuringly, the results are robust to the empirical specification chosen and the estimated effect of predicted intra-ethnic diversity on economic specialization remains positive, statistically and economically significant.

Finally, we explore the effect of multiple hypotheses testing using the methods introduced by Benjamini and Hochberg (1995) and Benjamini et al. (2006). Specifically, Benjamini and Hochberg (1995) and Benjamini et al. (2006) introduced False Discovery Rates (FDR) corrected q-values, which have similar interpretations as p-values and give the probability threshold at which the Null hypothesis would be rejected after accounting for multiple hypotheses testing (Anderson, 2008). We performed the correction of multiple hypotheses testing including all variables available in the Ethnographic Atlas and SCCS, which had an effective sample size of over 50 ethnicities to avoid inflating the number of not statistically significant associations. Reassuringly, our main results remain unchanged (Tables E.5 and E.6).

5.2 Spatial Auto-correlation, Cultural Dependence and Inter-Ethnic Interactions

A potential concern with the previous results is that migration and splitting of ancestral groups may have created historical, cultural and spatial correlation among ethnic outcomes. Moreover, the results may be biased due to interactions between an ethnicity and its neighbors, since close contact with neighboring ethnicities may affect information diffusion and trade opportunities, and thus potentially the prevalence of economic specialization. In order to mitigate these potential concerns, the analysis follows various strategies: (i) it accounts for the potential spatial dependence of observations by allowing the error terms to be spatially auto-correlated (see in particular Appendix E.4, Tables E.8, E.10-E.12). (ii) It accounts for the potential historical and cultural dependence of observations by allowing the error terms to be correlated within language phyla (i.e., a group of phylogenetically related languages inside a linguistic family) (Tables E.8-E.12). (iii) It accounts for the potential interaction between an ethnic group and neighboring ethnicities as reflected by geographical proximity (E.13-E.15). Reassuringly, our results remain qualitatively unchanged. (iv) It constrains the analysis to the sample of ethnicities in the Standard Cross-Cultural Sample (SCCS), which was selected by ethnographers to minimize cultural and spatial dependence across observations (Table E.16). (v) It accounts for language phylum fixed-effects, and thus for omitted heterogeneity across culturally and historically related ethnic groups (Table E.16).

5.3 Historical Confounders

Another potential concern is that intra-ethnic diversity is capturing the effect of factors like the transition to agriculture, the history of settlement or technological diffusion on economic specialization. Reassuringly, our results are robust to accounting for major historical confounders (Table E.16). In particular, we analyze the potential confounding effect of (i) the long-lasting influence of the Neolithic Revolution, by accounting for the pre-industrial distance to the closest Neolithic frontier (i.e., the closest location of animal or plant domestication).[46],[47] (ii) the pre-industrial distance from the closest

---

[46]The analysis estimates the minimal travel paths based on HMISea from the centroid of each ethnic homeland to the closest Neolithic frontier. The location of Neolithic frontiers is taken from various sources (Diamond, 1997; Smith, 1997; Benz, 2001; Denham et al., 2003; Pinhasi et al., 2005; Smith, 2006; Dillehay et al., 2007; Lu et al., 2009; Manning et al., 2011; Linseele, 2013).

[47]Alternatively, accounting for the degree of subsistence dependence on agriculture, as measured in the Ethnographic Atlas (v5), does not alter the results either.

technological frontier in the year 1500CE as identified by Ashraf and Galor (2011).[48] (iii) the duration of human settlements since prehistoric times, which estimates the date since the first uninterrupted settlement by anatomically modern humans (Ahlerup and Olsson, 2012).[49]

5.4 Robustness to Origin of Serial Founder Effect

While the East African origins of humans is the most widely accepted theory, there is an ongoing debate on the specific location within Africa where the expansion of modern humans started. In particular, Chan et al. (2019) among others have suggested a Southern African origin (López et al., 2015, see also). In section F.4 the analysis explores the robustness to the origin of the Serial Founder Effect (SFE). Specifically, it considers alternative origins located in South Africa, as well as in the location of the most genetically diverse ethnicity within the sample. Tables F.6 and H.12 establish that the effect of distance from the origin of the SFE on intra-ethnic diversity is practically identical for all origins. Additionally, Tables F.7, F.8, H.13 establish the robustness of the main results to these alternative origins.

6 Concluding Remarks

This research is the first attempt to identify the deep-rooted historical factors behind pre-modern economic specialization of labor across ethnic groups. It advances and empirically tests the hypothesis that intra-ethnic diversity was conducive to the division of labor in pre-modern times across ethnic groups. To empirically test this hypothesis, this research combines geocoded ethnographic, linguistic and genetic data to construct a novel dataset of measures of intra-ethnic diversity and economic specialization for pre-modern societies. In particular, for over 1100 ethnicities, the research constructs novel measures of economic specialization, based on the number of economic activities in which specialization existed in the pre-modern era. Additionally, it constructs various proxies of historical intra-ethnic diversity based on genetic and linguistic data at the ethnic level. This allows the analysis to explore the potential effects of intra-ethnic diversity (as opposed to country-level inter-ethnic diversity) on economic specialization across ethnic groups.

The analysis presents supportive empirical evidence consistent for the role of pre-modern intra-ethnic diversity as a fundamental driver of the division of labor in pre-modern times. Moreover, it provides evidence that the positive association between intra-ethnic diversity and economic specialization is reinforced for populations inhabiting diverse geographical environments. Using various robustness checks, the analysis suggests that these findings are not confounding the effect of geographical, cultural or institutional factors, nor other historical processes. Furthermore, the results suggest that the implied economic effect of intra-ethnic diversity is large and more important than alternative potential drivers of the division of labor.

---

[48]The technological frontiers are London and Paris in Europe, Fez and Cairo in Africa, Constantinople and Peking in Asia, and Tenochtitlan and Cuzco in the Americas.

[49]Given that the original data is available at the country level, the analysis follows the literature and constructs ethnic level measures by creating population-weighted averages (Alesina et al., 2013; Giuliano and Nunn, 2013; Gennaioli and Rainer, 2007).

This research is one of the few to empirically document that population diversity is conducive to economic development. Moreover, it provides evidence for the fundamental role played by economic specialization in the advancement of societies. In particular, it suggests the division of labor plays an instrumental mediating role in the relation between population diversity and development. Furthermore, the analysis provides prima-facie empirical evidence consistent with the upward sloping side of the hump-shaped relationship between diversity and economic development established by Ashraf and Galor (2013b).

References

Acemoglu, D., Johnson, S. and Robinson, J. (2005). Institutions as a fundamental cause of long-run growth, *Handbook of economic growth* **1**: 385–472.

Acharya, A., Blackwell, M. and Sen, M. (2016). Explaining causal findings without bias: Detecting and assessing direct effects, *American Political Science Review* **110**(3): 512–529.

Ahlerup, P. and Olsson, O. (2012). The roots of ethnic diversity, *Journal of Economic Growth* **17**(2): 71–102.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. T. (2003). Fractionalization, *Journal of Economic Growth* **8**(2): 155–194.

Alesina, A., Giuliano, P. and Nunn, N. (2013). On the origins of gender roles: Women and the plough, *The Quarterly Journal of Economics* **128**(2): 469–530.

Alesina, A., Harnoss, J. and Rapoport, H. (2016). Birthplace diversity and economic prosperity, *Journal of Economic Growth* **21**(2): 101–138.

Andersen, T. B., Dalgaard, C.-J. and Selaya, P. (2016). Climate and the emergence of global income differences, *The Review of Economic Studies* **83**(4): 1334–1363.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects, *Journal of the American statistical Association* **103**(484): 1481–1495.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press, Princeton.

Arbatli, C. E., Ashraf, Q. H., Galor, O. and Klemp, M. (2018). Diversity and Conflict, *Working Papers 2018-6*, Brown University, Department of Economics.

Arbatlı, C. E., Ashraf, Q. H., Galor, O. and Klemp, M. (2020). Diversity and conflict, *Econometrica* **88**(2): 727–797.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations, *The review of economic studies* **58**(2): 277–297.

Ashraf, Q. and Galor, O. (2011). Dynamics and stagnation in the malthusian epoch, *The American Economic Review* **101**(5): 2003–2041.

Ashraf, Q. and Galor, O. (2013a). Genetic diversity and the origins of cultural fragmentation, *The American Economic Review* **103**(3): 528–533.

Ashraf, Q. and Galor, O. (2013b). The out of africa hypothesis, human genetic diversity, and comparative economic development, *The American Economic Review* **103**(1): 1–46.

Ashraf, Q., Galor, O. and Özak, Ö. (2010). Isolation and development, *Journal of the European Economic Association* **8**(2-3): 401–412.

Ashraf, Q. H. and Galor, O. (2018). The macrogenoeconomics of comparative development, *Journal of Economic Literature* **56**(3): 1119–55.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from africa, *Science* **332**(6027): 346–349.

Baker, A. J. and Jenkins, P. F. (1987). Founder effect and cultural evolution of songs in an isolated population of chaffinches, fringilla coelebs, in the chatham islands, *Animal Behaviour* **35**(6): 1793–1803.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal statistical society: series B (Methodological)* **57**(1): 289–300.

Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate, *Biometrika* **93**(3): 491–507.

Benz, B. F. (2001). Archaeological evidence of teosinte domestication from guilá naquitz, oaxaca, *Proceedings of the National Academy of Sciences* **98**(4): 2104–2106.

Betti, L., Balloux, F., Amos, W., Hanihara, T. and Manica, A. (2009). Distance from africa, not climate, explains within-population phenotypic diversity in humans, *Proceedings of the Royal Society of London B: Biological Sciences* **276**(1658): 809–814.

Birchenall, J. A. (2014). Disease and diversity in africa's long-term economic development, *Technical report*, University of California Santa Barbara.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models, *Journal of econometrics* **87**(1): 115–143.

Boix, C. (2015). *Political order and inequality: their foundations and their consequences for human welfare*, Cambridge studies in comparative politics, Cambridge University Press.

Borland, J. and Yang, X. (1992). Specialization and a new approach to economic organization and growth, *The American Economic Review* **82**(2): 386–391.

Brumfiel, E. M. and Earle, T. K. (1987). *Specialization, exchange, and complex societies*, Cambridge University Press, Cambridge.

Chan, E. K., Timmermann, A., Baldi, B. F., Moore, A. E., Lyons, R. J., Lee, S.-S., Kalsbeek, A. M., Petersen, D. C., Rautenbach, H., Förtsch, H. E. et al. (2019). Human origins in a southern african palaeo-wetland and first migrations, *Nature* **575**(7781): 185–189.

Childe, V. G. and Wailes, B. (1996). *Craft specialization and social evolution: in memory of V. Gordon Childe*, Vol. 93, University of Pennsylvania Museum of Archaeology and Anthroplogy, Philadelphia.

Claessen, H. J. M. and Skalník, P. (1978). *The Early state*, Mouton, The Hague.

Conley, T. (1999). Gmm estimation with cross sectional dependence, *Journal of Econometrics* **92**(1): 1–45.

Cook, C. J. (2015). The natural selection of infectious disease resistance and its effect on contemporary health, *Review of Economics and Statistics* **97**(4): 742–757.

Creanza, N., Ruhlen, M., Pemberton, T. J., Rosenberg, N. A., Feldman, M. W. and Ramachandran, S. (2015). A comparison of worldwide phonemic and genetic variation in human populations, *Proceedings of the National Academy of Sciences* **112**(5): 1265–1272.

Dean, J. S., Euler, R. C., Gumerman, G. J., Plog, F., Hevly, R. H. and Karlstrom, T. N. (1985). Human behavior, demography, and paleoenvironment on the colorado plateaus, *American Antiquity* pp. 537–554.

Denham, T. P., Haberle, S. G., Lentfer, C., Fullagar, R., Field, J., Therin, M., Porch, N. and Winsborough, B. (2003). Origins of agriculture at kuk swamp in the highlands of new guinea, *Science* **301**(5630): 189–193.

Depetris-Chauvin, E. (2014). State history and contemporary conflict: Evidence from sub-saharan africa, *Technical report*, Brown University.

Depetris-Chauvin, E. and Özak, Ö. (2015). Population diversity, division of labor and the emergence of trade and state, *Technical Report 1506*, Southern Methodist University, Department of Economics.

Depetris-Chauvin, E. and Özak, Ö. (2016). The origins and long-run consequences of the division of labor, *Technical Report 1610*, Southern Methodist University, Department of Economics.

Desmet, K., Ortuño-Ortín, I. and Wacziarg, R. (2012). The political economy of linguistic cleavages, *Journal of development Economics* **97**(2): 322–338.

Desmet, K., Ortuño-Ortín, I. and Wacziarg, R. (2015). Culture, ethnicity and diversity, *Technical report*, National Bureau of Economic Research.

Diamond, J. M. (1997). *Guns, germs, and steel: the fates of human societies*, 1st ed edn, W.W. Norton & Co., New York.

Dillehay, T. D., Rossen, J., Andres, T. C. and Williams, D. E. (2007). Preceramic adoption of peanut, squash, and cotton in northern peru, *Science* **316**(5833): 1890–1893.

Dryer, Matthew S & Haspelmath, M. e. (2013). *The World Atlas of of Language Structures Online.*, Leipzig: Max Planck Institute for Evolutionary Anthropology.

Durkheim, É. (1893). *De la division du travail social*, 2. éd edn, F. Alcan, Paris.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions, *The Quarterly Journal of Economics* pp. 1203–1250.

Fenske, J. (2014). Ecology, trade, and states in pre-colonial africa, *Journal of the European Economic Association* **12**(3): 612–640.

Freeman, D. and Pankhurst, A. (2003). *Peripheral people: the excluded minorities of Ethiopia*, Red Sea Press, Lawrenceville, NJ.

Friedlaender, J. S., Friedlaender, F. R., Reed, F. A., Kidd, K. K., Kidd, J. R., Chambers, G. K., Lea, R. A., Loo, J.-H., Koki, G., Hodgson, J. A. et al. (2008). The genetic structure of pacific islanders, *PLoS Genet* **4**(1): e19.

Galor, O. and Klemp, M. (2017). Roots of autocracy, *Working Paper 23301*, National Bureau of Economic Research.

Galor, O. and Özak, Ö. (2015). Land productivity and economic development: Caloric suitability vs. agricultural suitability, *SMU Working Paper Series* .

Galor, O. and Özak, Ö. (2016). The agricultural origins of time preference, *American Economic Review* **106**(10).

Galor, O., Özak, Ö. and Sarid, A. (2016). Geographical origins and economic consequences of language structures, *SMU Working Paper Series* .

Gennaioli, N. and Rainer, I. (2007). The modern impact of precolonial centralization in africa, *Journal of Economic Growth* **12**(3): 185–234.

Giuliano, P. and Nunn, N. (2013). The transmission of democracy: from the village to the nation-state, *The American Economic Review* **103**(3): 86–92.

Giuliano, P. and Nunn, N. (2016). Understanding cultural persistence and change.

Glaeser, E. L., La Porta, R., Lopez-de Silanes, F. and Shleifer, A. (2004). Do institutions cause growth?, *Journal of economic Growth* **9**(3): 271–303.

Hallpike, C. R. (2008). *The Konso of Ethiopia: a study of the values of a Cushitic people*, AuthorHouse, Central Milton Keynes.

Hamilton, M. J. and Buchanan, B. (2009). The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: quantifying clovis evolutionary dynamics, *Journal of Anthropological Archaeology* **28**(1): 55–69.

Henrich, J. and Boyd, R. (2008). Division of labor, economic specialization, and the evolution of social stratification, *Current Anthropology* **49**(4): 715–724.

Hill, K. and Hurtado, A. M. (1996). *Aché life history: the ecology and demography of a foraging people*, Aldine de Gruyter, New York.

Hinton, P. (1979). The karen, millennialism, and the politics of accommodation to lowland states, *in* C. F. Keyes (ed.), *Ethnic Adaptation and Identity: The Karen on the Thai Frontier with Burma*, Institute for the Study of Human Issues, pp. 81–98.

Houthakker, H. S. (1956). Economics and biology: specialization and speciation, *Kyklos* **9**(2): 181–189.

Kimura, M. (1983). *The neutral theory of molecular evolution*, Cambridge University Press, Cambridge.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E. and Sachs, J. (2004). A global index representing the stability of malaria transmission, *The American Journal of Tropical Medicine and Hygiene* **70**(5): 486–498.

Kiyotaki, N. and Wright, R. (1989). On money as a medium of exchange, *The Journal of Political Economy* pp. 927–954.

Lange, T., Vansteelandt, S. and Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects, *American journal of epidemiology* **176**(3): 190–195.

Lao, O., Lu, T. T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L. A., Comas, D. et al. (2008). Correlation between genetic and geographic structure in europe, *Current Biology* **18**(16): 1241–1248.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models, *Journal of Business & Economic Statistics* **30**(1): 67–80.

Linseele, V. (2013). Early stock keeping in northeastern africa. near eastern influences and local developments, *in* N. Shirai (ed.), *Neolithisation of Northeastern Africa*, Ex Oriente.

López, S., Van Dorp, L. and Hellenthal, G. (2015). Human dispersal out of africa: a lasting debate, *Evolutionary Bioinformatics* **11**: EBO–S33489.

Lu, H., Zhang, J., Liu, K.-b., Wu, N., Li, Y., Zhou, K., Ye, M., Zhang, T., Zhang, H., Yang, X. et al. (2009). Earliest domestication of common millet (panicum miliaceum) in east asia extended to 10,000 years ago, *Proceedings of the National Academy of Sciences* **106**(18): 7367–7372.

Lycett, S. J. (2008). Acheulean variation and selection: does handaxe symmetry fit neutral expectations?, *Journal of Archaeological Science* **35**(9): 2640–2648.

Lycett, S. J. and von Cramon-Taubadel, N. (2008). Acheulean variability and hominin dispersals: a model-bound approach, *Journal of Archaeological Science* **35**(3): 553–562.

Manica, A., Amos, W., Balloux, F. and Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation, *Nature* **448**(7151): 346–348.

Manning, K., Pelling, R., Higham, T., Schwenniger, J.-L. and Fuller, D. Q. (2011). 4500-year old domesticated pearl millet (pennisetum glaucum) from the tilemsi valley, mali: new insights into an alternative cereal domestication pathway, *Journal of Archaeological Science* **38**(2): 312–322.

Marx, K. (1867). *Das kapital: Kritik der politischen oekonomie.*, O. Meissner, Hamburg.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity, *The American Economic Review* **102**(4): 1508.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development, *Econometrica* **81**(1): 113–152.

Moore, J. L., Manne, L., Brooks, T., Burgess, N. D., Davies, R., Rahbek, C., Williams, P. and Balmford, A. (2002). The distribution of cultural and biological diversity in africa, *Proceedings of the Royal Society of London B: Biological Sciences* **269**(1501): 1645–1653.

Murdock, G. P. (1967). Ethnographic atlas: a summary, *Ethnology* pp. 109–236.

Murdock, G. P. and White, D. R. (1969). Standard cross-cultural sample, *Ethnology* pp. 329–369.

Murphy, K. M. and Topel, R. H. (2002). Estimation and inference in two-step econometric models, *Journal of Business & Economic Statistics* **20**(1): 88–97.

Myres, N. M., Rootsi, S., Lin, A. A., Järve, M., King, R. J., Kutuev, I., Cabrera, V. M., Khusnutdinova, E. K., Pshenichnov, A., Yunusbayev, B. et al. (2011). A major y-chromosome haplogroup r1b holocene era founder effect in central and western europe, *European Journal of Human Genetics* **19**(1): 95–101.

Nettle, D. (1998). Explaining global patterns of language diversity, *Journal of anthropological archaeology* **17**(4): 354–374.

Nolan, P. and Lenski, G. E. (2011). *Human societies: an introduction to macrosociology*, 11th ed edn, Paradigm Publishers, Boulder.

Nunn, N. (2008). The long-term effects of africa's slave trades, *The Quarterly Journal of Economics* **123**(1): 139–176.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C. et al. (2001). Terrestrial ecoregions of the world: A new map of life on earth a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, *BioScience* **51**(11): 933–938.

Özak, Ö. (2010). The voyage of homo-economicus: Some economic measures of distance.

Özak, Ö. (2018). Distance to the pre-industrial technological frontier and economic development, *Journal of Economic Growth* **23**(2): 175–221.

Pearl, J. (2000). *Causality: models, reasoning, and inference*, Cambridge University Press, Cambridge, U.K.

Pearl, J. (2001). Direct and indirect effects, *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 411–420.

Pemberton, T. J., DeGiorgio, M. and Rosenberg, N. A. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation, *G3: Genes| Genomes| Genetics* **3**(5): 891–907.

Pinhasi, R., Fort, J. and Ammerman, A. J. (2005). Tracing the origin and spread of agriculture in europe, *PLoS biology* **3**(12): e410.

Pinhasi, R., Thomas, M. G., Hofreiter, M., Currat, M. and Burger, J. (2012). The genetic history of europeans, *Trends in Genetics* **28**(10): 496–505.

Polanyi, K. (2001). *The great transformation: the political and economic origins of our time*, 2nd beacon paperback ed edn, Beacon Press, Boston, MA.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W. and Cavalli-Sforza, L. L. (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 of the United States of America* **102**(44): 15942–15947.

Ricardo, D. (1817). *On the principles of political economy and taxation*, J. Murray, London.

Rogers, D. S., Feldman, M. W. and Ehrlich, P. R. (2009). Inferring population histories using cultural data, *Proceedings of the Royal Society of London B: Biological Sciences* **276**(1674): 3835–3843.

Romer, P. M. (1987). Growth based on increasing returns due to specialization, *The American Economic Review* **77**(2): 56–62.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*, Whitestone, Dublin.

Smith, B. D. (1997). The initial domestication of cucurbita pepo in the americas 10,000 years ago, *Science* **276**(5314): 932–934.

Smith, B. D. (2006). Eastern north america as an independent center of plant domestication, *Proceedings of the National Academy of Sciences* **103**(33): 12223–12228.

Spolaore, E. and Wacziarg, R. (2013). How deep are the roots of economic development?, *Journal of Economic Literature* **51**(2): 325–69.

Stigler, G. J. (1951). The division of labor is limited by the extent of the market, *The Journal of Political Economy* pp. 185–193.

Sun, G.-Z. (2012). *The division of labour in economics: a history*, Vol. 142 of *Routledge studies in the history of economics*, Routledge, London.

Trigger, B. G. (1983). The rise of egyptian civilization, *in* B. G. Trigger, B. J. Kemp, D. O'Connor and A. B. Lloyd (eds), *Ancient Egypt: a social history*, Cambridge University Press Cambridge, pp. 1–70.

Wang, S., Lewis Jr, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M. V., Molina, J. A., Gallo, C. et al. (2007). Genetic variation and population structure in native americans, *PLoS Genet* **3**(11): e185.

Yang, X. and Borland, J. (1991). A microeconomic mechanism for economic growth, *Journal of political economy* pp. 460–482.

Yang, X. and Sachs, J. D. (2008). *Economic development and the division of labor*, John Wiley & Sons.

Online Appendix (Not for publication)

Additional Results and Supporting Material

A Intra-Ethnic Diversity and Division of Labor: A Structural Model

This section presents a structural econometric model for the analysis of the effect of intra-ethnic diversity on economic specialization. In particular, it shows that if the level of intra-ethnic diversity of various traits has a common factor, then one measure of intra-ethnic diversity can be used as a proxy for *all* these other types of diversity. Additionally, it establishes that if the underlying factor is exogenous for an observable trait's diversity in the estimation of the effect on the division of labor, then it can be used as an instrument for intra-ethnic diversity. Moreover, the instrumental variable estimate of the effect of intra-ethnic diversity in the observed trait on specialization provides a lower bound to the effect of intra-ethnic diversity in all these traits. Finally, it establishes that although this estimated effect provides an unbiased and consistent estimate of the lower bound of the combined effect of *all intra-ethnic diversity*, it cannot identify which specific trait drives this effect.

Assume the structural equation for the relation between economic specialization and intra-ethnic diversity is

$$s_i = \alpha_0 + \sum_{j=1}^{J} \alpha_{1j} d_{ij} + \sum_{k=1}^{K} \alpha_{2k} x_{ik} + \epsilon_i \tag{7}$$

where s_i measures economic specialization, d_{ij} is the level of intra-ethnic diversity in trait $j = 1, \ldots, J$, x_{ik} is the level of geographical characteristic k and ϵ_i is the error term, all for ethnicity i. Additionally, assume that an ethnicity's diversity in trait k is determined by

$$d_{ij} = \beta_{0j} + \beta_{1j} D_i + \sum_{k=1}^{K} \beta_{2jk} x_{ik} + \eta_{ij}. \tag{8}$$

where D_i is the historical migratory distance from an ethnicity's homeland to the ancestral origin, i.e., the distance to the source that generates the serial founder effect (which in the case of this paper is East Africa).

Assumption 1. Let $x_i = (x_{ik})_{k=1}^{K}$ and assume that for all $j = 1, \ldots, J$:

(i) $\beta_{1j} < 0$,

(ii) $E(\epsilon_i \mid D_i, x_i) = E(\eta_i \mid D_i, x_i) = 0$,

(iii) $E(D_i \epsilon_{ij} \mid x_i) = E(D_i \eta_{ij} \mid x_i) = 0$,

i.e., D_i is exogenous for all measures of diversity.

Additionally, assume that some trait p is observable and so that diversity in p can be measured. Clearly,

Proposition 4. d_{ip} serves as a proxy for all other measures of diversity.

Proof. Notice that

$$D_i = \frac{d_{ip}}{\beta_{1p}} - \frac{\beta_{0p} + \sum_{k=1}^{K} \beta_{2pk} x_{ik} + \eta_{ip}}{\beta_{1p}} \tag{9}$$

and for all $j \neq l$,

$$d_{ij} = \left(\beta_{0j} - \frac{\beta_{0p}}{\beta_{1l}} \right) + \frac{\beta_{1j}}{\beta_{1p}} d_{ip} + \sum_{k=1}^{K} \left(\beta_{2jk} - \frac{\beta_{2pk}}{\beta_{1p}} \right) x_{ik} + \left(\eta_{ij} - \frac{\eta_{ip}}{\beta_{1p}} \right)$$

$$= \gamma_{0j} + \gamma_{1j} d_{ip} + \sum_{k=1}^{K} \gamma_{2jk} x_{ik} + \zeta_{ij}, \tag{10}$$

where $\gamma_{1j} \neq 0$ for all $j \neq p$. $\square$

Notice that this is precisely equation (3). Replacing it into (7), it follows that

$$s_i = \left(\alpha_0 + \sum_{j \neq p} \alpha_{1j} \gamma_{0j} \right) + \left(\alpha_{1p} + \sum_{j \neq p} \alpha_{1j} \gamma_{1j} \right) d_{ip} + \sum_{k=1}^{K} \left(\alpha_{2k} + \sum_{j \neq p} \alpha_{1j} \gamma_{2jk} \right) x_{ik} + \left(\epsilon_i + \sum_{j \neq p} \alpha_{1j} \zeta_{ij} \right)$$

$$= \delta_0 + \delta_1 d_{ip} + \sum_{k=1}^{K} \delta_{2k} x_{ik} + \varepsilon_i, \tag{11}$$

which is equation (4).

Proposition 5. *Clearly, for any two proxies p and p',*

$$\frac{\delta_1^p}{\delta_1^{p'}} = \frac{\beta_{1p'}}{\beta_{1p}}, \tag{12}$$

i.e., the effect of intra-ethnic diversity as proxied by p is identical to the effect of intra-ethnic diversity as proxied by p', once one accounts for the differential effect of serial effect on both proxies.

Proof. Since

$$\delta_1^p = \alpha_{1p} + \sum_{j \neq p} \alpha_{1j} \gamma_{1j}^p, \qquad \delta_1^{p'} = \alpha_{1p'} + \sum_{j \neq p'} \alpha_{1j} \gamma_{1j}^{p'}, \qquad \Longrightarrow \qquad \beta_{1p} \delta_1^p = \beta_{1p'} \delta_1^{p'}. \qquad \square$$

Proposition 6. *If assumption 1 holds, D_i is a valid instrumental variable for d_{ip}. Thus, δ_1 can be consistently estimated.*

Proof. By assumption, $E(D_i \varepsilon_i \mid x_i) = 0$ and $\beta_{1p} \neq 0$. Thus, D_i satisfies both the exogeneity and relevance conditions required for a valid instrument. Additionally, notice that it also satisfies the exclusion restriction, since D_i can only affect s_i via intra-ethnic diversity. $\square$

Finally, notice that

Proposition 7. *If $\delta_1 > 0$, then $\alpha_{1j} > 0$ for at least one $j = 1, \ldots, J$. Moreover, if $\alpha_{1p} = 0$ and $\delta_1 > 0$, then $\alpha_{1j} > 0$ for some $j = 1, \ldots, J$. Additionally, if $\alpha_{1j} = 0$ and $\alpha_{1j'} > 0$ for two subsets of $\{1, \ldots, J\}$, then $\delta_1 > 0$.*

Proof. Follows directly from equation (11). $\square$

Moreover,

Proposition 8. *Assume that $\beta_{1p} < \beta_{1j}$ and $\alpha_{1j} \geq 0$ for all $j = 1, \ldots, J$. Then,*

$$\delta_1 = \alpha_{1p} + \sum_{j \neq p} \alpha_{1j} \gamma_{1j} < \sum_j \alpha_{1j}, \tag{13}$$

40

i.e., δ_1 *provides a lower bound to the effect of all types of intergenerationally transmitted intra-ethnic diversity on economic specialization. Similarly, if* $\beta_{1p} < \beta_{1j}$ *and* $\alpha_{1j} \leq 0$ *for all* $j = 1, \ldots, J$. *Then,*

$$\delta_1 = \alpha_{1p} + \sum_{j \neq p} \alpha_{1j}\gamma_{1j} > \sum_j \alpha_{1j}, \tag{14}$$

i.e., δ_1 *provides an upper bound to the effect of all types of intergenerationally transmitted intra-ethnic diversity on economic specialization.*

Proof. Follows directly from the assumption and the definition of γ_{1j}. $\qquad\square$

These results imply that:

(i) Genetic and linguistic diversity can be used as proxies for intra-ethnic diversity of intergenerationally transmitted traits (affected by similar serial founder effects).

(ii) The distance to East Africa is a valid instrument for intra-ethnic diversity. The main concern being the exogeneity assumption $E(D_i\varepsilon_i) = 0$. The results in appendix G show that the reduced form estimates are very stable across specifications, suggesting that this condition may be satisfied in practice. Further supportive evidence in favor of this assumption is provided in Table E.16.

(iii) The estimated effect of the proxies of intra-ethnic diversity presented in the body of the paper capture the combined effect of intra-ethnic diversity in all relevant intergenerationally transmitted traits.

(iv) The estimated effect of the proxies of intra-ethnic diversity do not necessarily imply that genetics or language are the fundamental mechanism behind the effect of intra-ethnic diversity on the division of labor.

B   Sample Properties

B.1   Sample Distribution
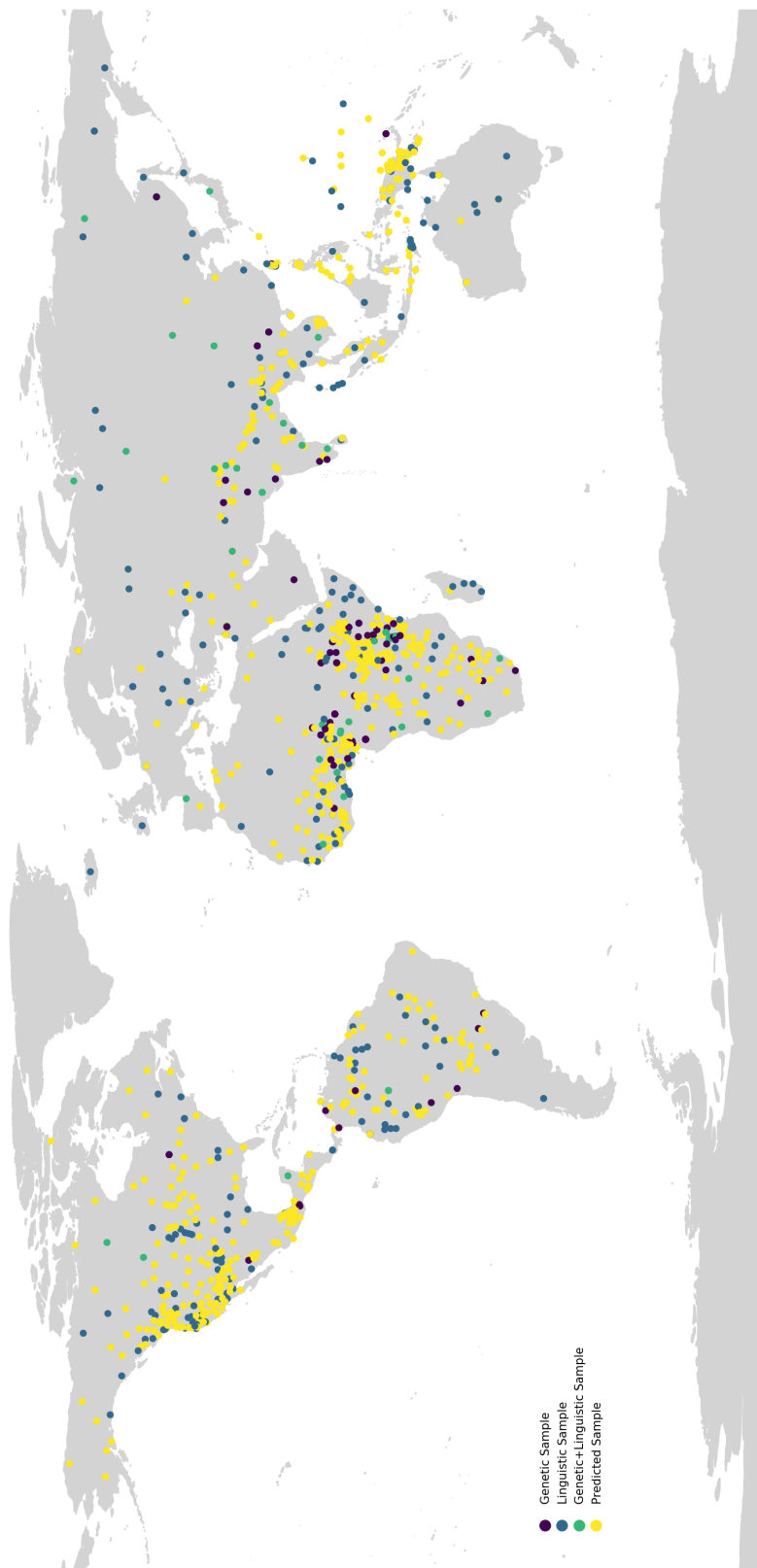
Figure B.1: Location of Ethnicities employed in the Analysis (Full and Restricted Samples)
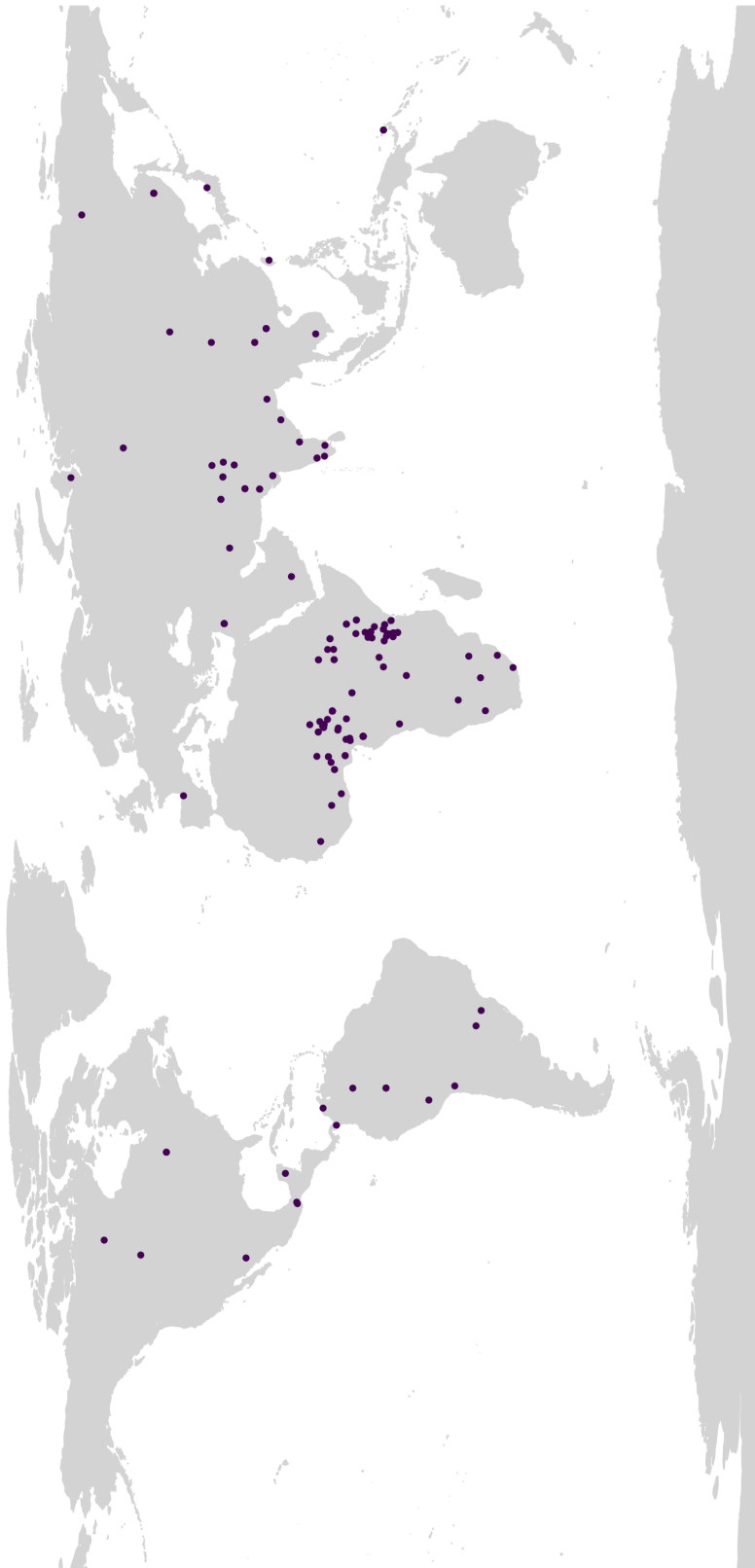
Figure B.2: Genetic Sample
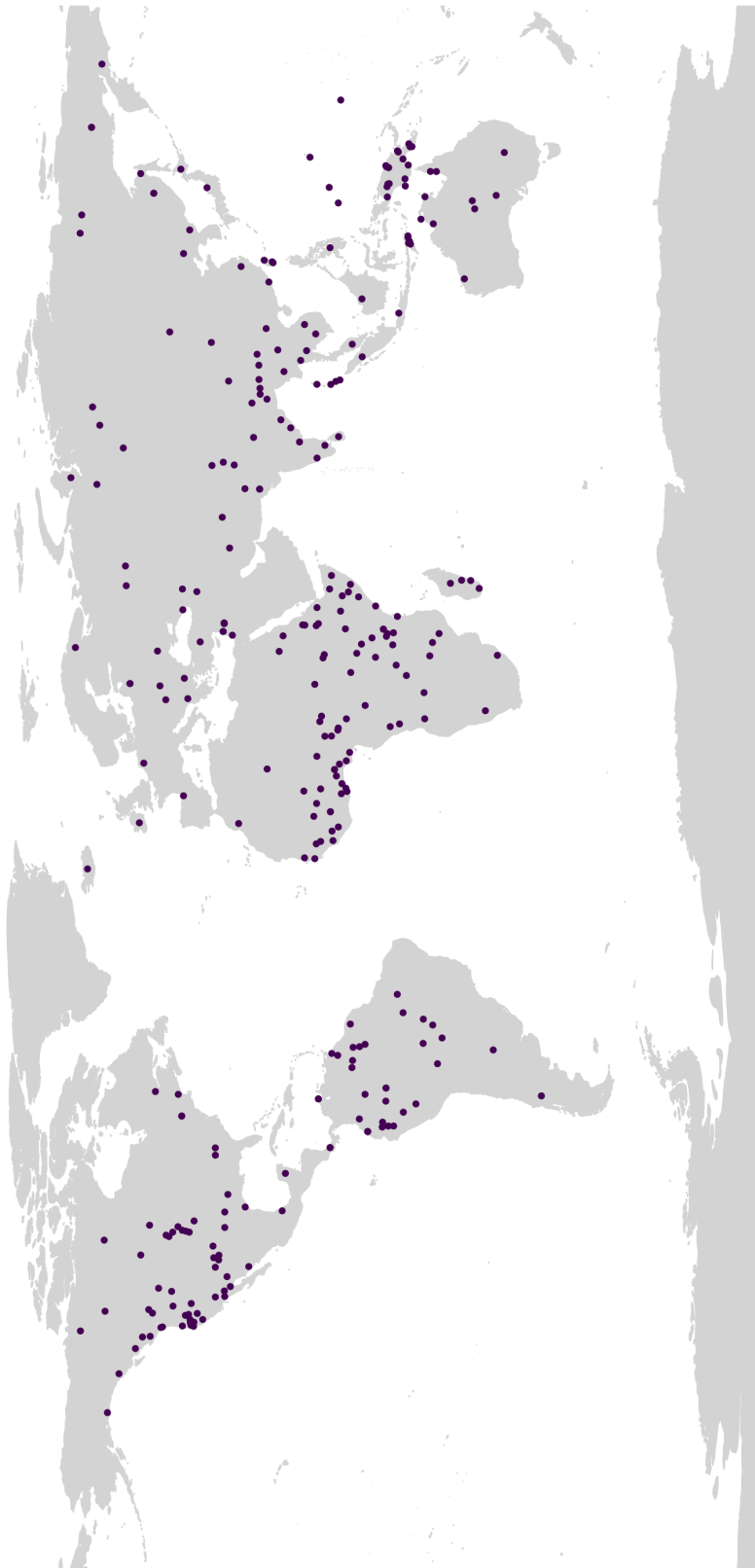
Figure B.3: Linguistic Sample

45
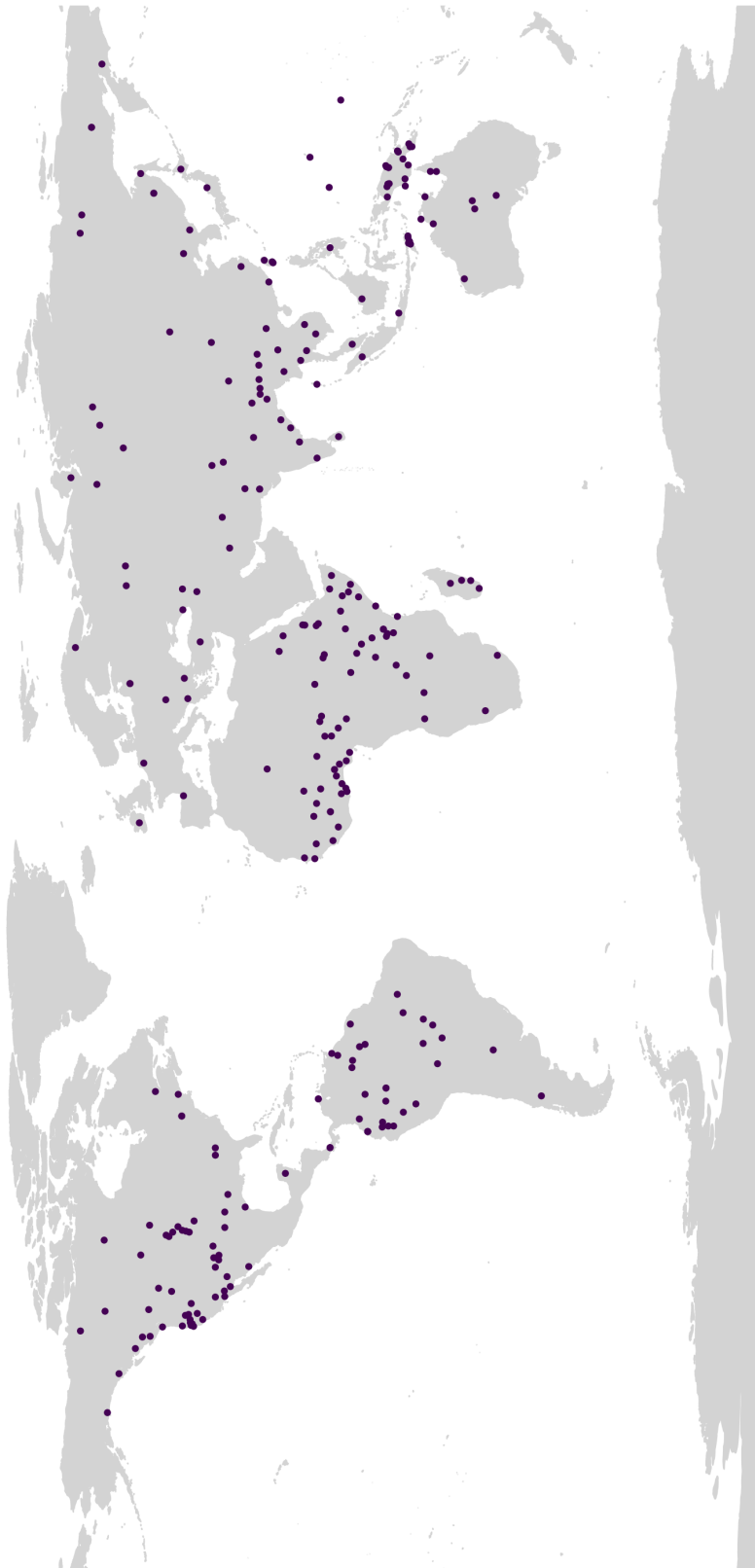
Figure B.4: Linguistic Sample 1 (Consonant Inventory)
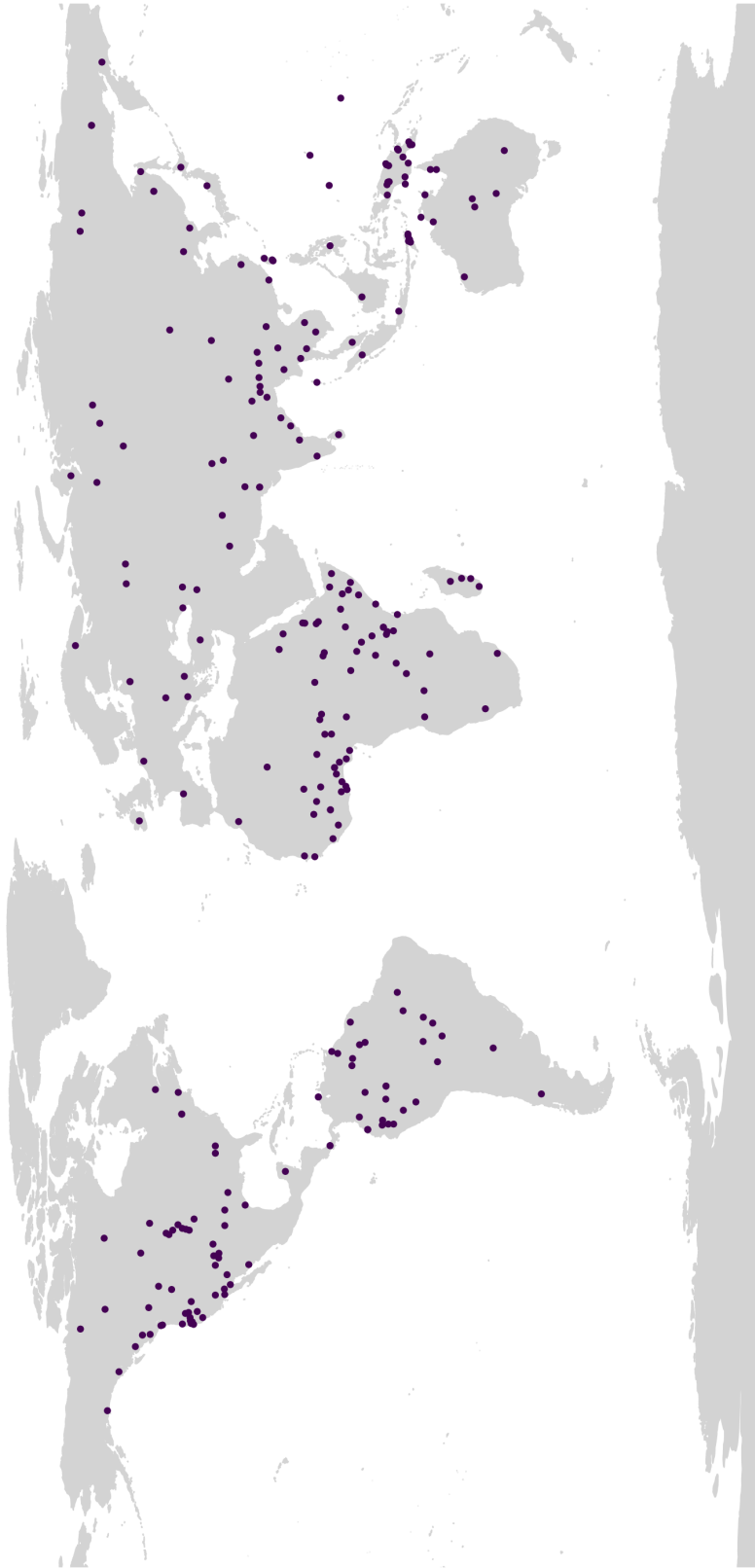
Figure B.5: Linguistic Sample 2 (Vowel Quality Inventory)
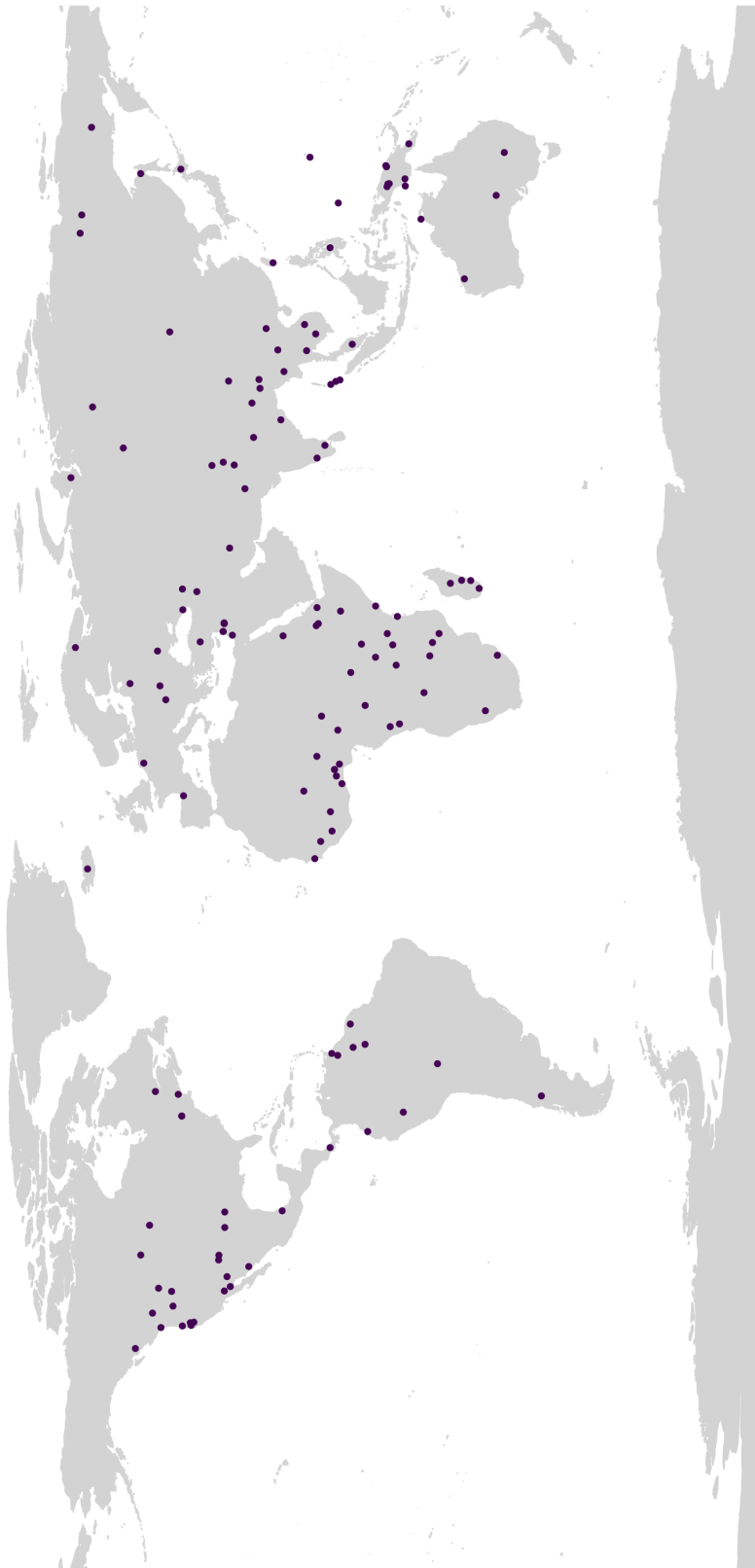
Figure B.6: Linguistic Sample 3 (Number of Genders)

B.2 Summary Statistics

(a) Number of specialized activities (s_e^1)

(b) Number of specialized activities (s_e^1)

(c) Share of activities that are specialized (s_e^2)

(d) Share of activities that are specialized (s_e^2)

(e) Score of economic specialization (s_e^3)

(f) Score of economic specialization (s_e^3)
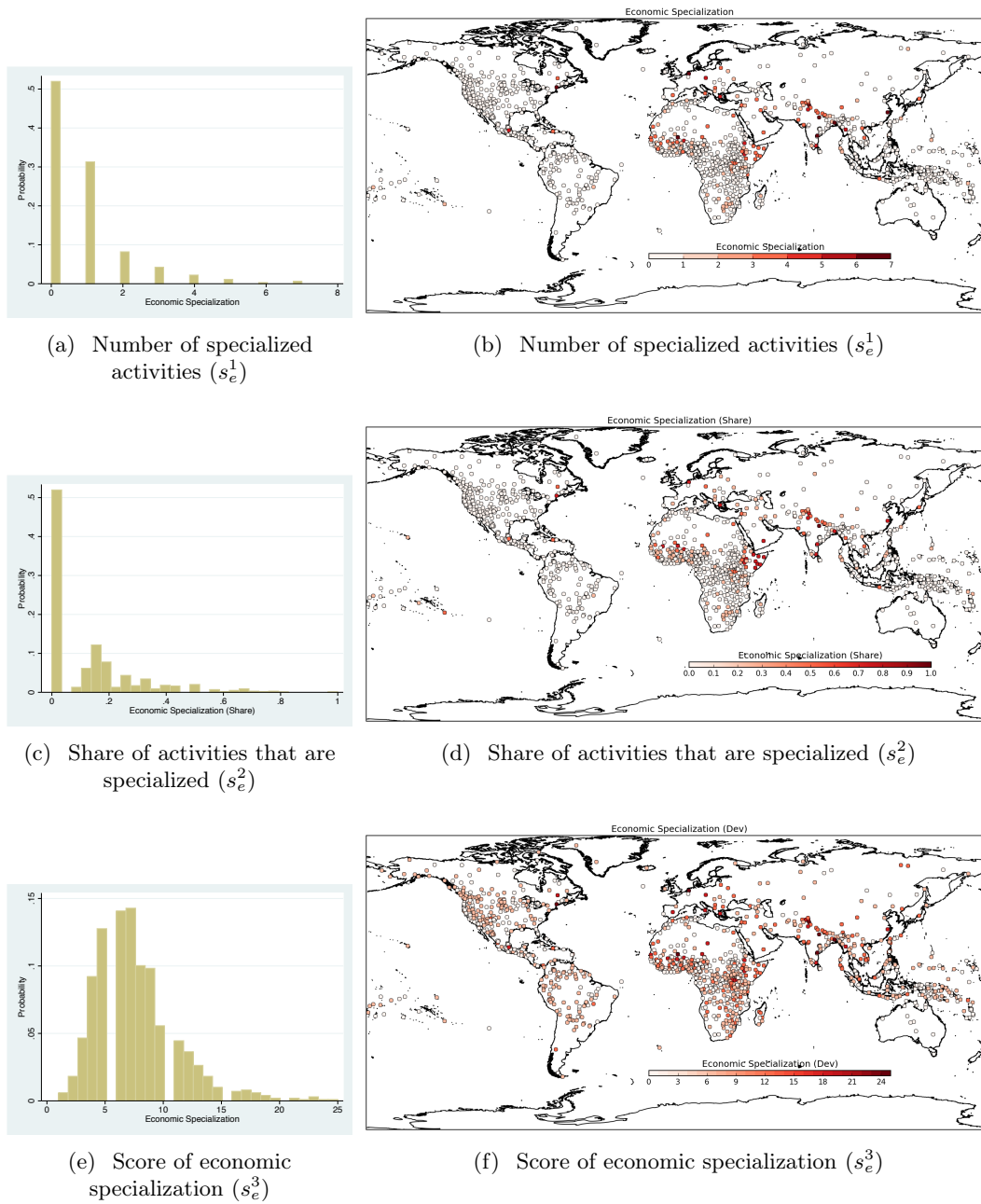
Figure B.7: Distribution of Economic Specialization Measures

Table B.1: Summary Statistics on Base Sample

	Mean	Std	Min	Max	N
Economic Specialization	1.34	(1.41)	0.00	7.00	116
Economic Specialization (Share)	0.20	(0.19)	0.00	0.80	116
Economic Specialization (Dev)	9.02	(3.96)	2.00	25.00	116
Intra-Ethnic Diversity	0.70	(0.05)	0.47	0.76	116
Absolute Latitude	15.95	(15.22)	0.04	68.67	116
Area	0.18	(0.85)	0.00	8.97	116
Elevation (Avg.)	823.71	(727.51)	27.79	3581.35	116
Precipitation (Avg.)	91.00	(57.54)	11.77	334.73	116
Temperature (Avg.)	20.69	(8.43)	-13.44	28.27	116
Malaria Ecology	7.88	(9.07)	0.00	29.36	116
Ecological Diversity	0.26	(0.22)	0.00	0.67	116
Agricultural Suitability (avg.)	0.76	(0.33)	0.00	1.00	116
Agricultural Suitability (std.)	0.08	(0.11)	0.00	0.45	116
Caloric Suitability Index (Pre-1500CE)	2699.11	(1040.20)	0.00	5030.97	116
Caloric Suitability (Pre-1500 ,std.)	418.27	(360.47)	0.00	1520.41	116
Temperature (Spatial Corr., Avg.)	0.93	(0.17)	0.00	1.00	116
Temperature (Volatility, Avg)	0.84	(0.48)	0.27	2.87	116
Pct. Area within 100 kms of Sea	0.19	(0.33)	0.00	1.00	116
Coast Length	0.49	(2.16)	0.00	19.65	116
Ruggedness (Avg.)	110.62	(149.48)	1.27	1076.01	116
Pre-Industrial Mobility (avg.)	0.27	(0.06)	0.07	0.37	116
Pre-Industrial Mobility (std.)	0.05	(0.04)	0.01	0.25	116

Table B.2: Summary Statistics on Full Sample

	Mean	Std	Min	Max	N
Economic Specialization	0.85	(1.20)	0.00	7.00	934
Economic Specialization (Share)	0.13	(0.17)	0.00	1.00	934
Economic Specialization (Dev)	7.74	(3.59)	1.00	25.00	934
Predicted Intra-Ethnic Diversity	0.68	(0.05)	0.54	0.76	934
Absolute Latitude	20.77	(16.59)	0.02	71.22	934
Area	0.07	(0.37)	0.00	8.97	934
Elevation (Avg.)	755.14	(676.82)	1.06	4417.96	934
Precipitation (Avg.)	105.83	(71.13)	0.00	499.24	934
Temperature (Avg.)	19.09	(8.60)	-15.31	29.58	934
Malaria Ecology	5.58	(8.05)	0.00	33.95	934
Ecological Diversity	0.19	(0.21)	0.00	0.82	934
Agricultural Suitability (avg.)	0.76	(0.34)	0.00	1.00	934
Agricultural Suitability (std.)	0.07	(0.10)	0.00	0.47	934
Caloric Suitability Index (Pre-1500CE)	2673.34	(1282.61)	0.00	6955.56	934
Caloric Suitability (Pre-1500 ,std.)	362.60	(333.18)	0.00	2436.89	934
Temperature (Spatial Corr., Avg.)	0.86	(0.28)	0.00	1.00	934
Temperature (Volatility, Avg)	0.98	(0.57)	0.00	3.08	934
Pct. Area within 100 kms of Sea	0.30	(0.41)	0.00	1.00	934
Coast Length	0.34	(2.97)	0.00	81.92	934
Ruggedness (Avg.)	137.45	(160.05)	0.05	1137.67	934
Pre-Industrial Mobility (avg.)	0.27	(0.07)	0.06	0.47	934
Pre-Industrial Mobility (std.)	0.06	(0.05)	0.00	0.27	934

C Relation between Economic Specialization and Other Intra-Ethnic Exchange Related Measures

Table C.1: Correlation of Intra-Ethnic Economic Specialization and Intra-Ethnic Trade Measures

	Intra-Ethnic Economic Specialization and Intra-Ethnic Trade Measures									
	Economic Specialization	Economic Specialization (Share)	Economic Specialization (Dev)	Importance of Trade	Importance of Intercommunity Trade as Food Source	Money	Credit Source	Writing and Records	Technological Specialization	Complexity
Pre-modern Economic Specialization	1.00									
Pre-modern Economic Specialization (Share)	0.94***	1.00								
Pre-modern Economic Specialization (Dev)	0.87***	0.74***	1.00							
Importance of Trade	0.20***	0.19**	0.13*	1.00						
Intercommunity Trade as Food Source	0.32***	0.32***	0.25***	0.77***	1.00					
Money	0.35***	0.38***	0.30***	0.29***	0.43***	1.00				
Credit Source	0.25***	0.21***	0.26***	0.28***	0.31***	0.35***	1.00			
Writing and Records	0.55***	0.55***	0.47***	0.24***	0.31***	0.38***	0.38***	1.00		
Technological Specialization	0.60***	0.58***	0.67***	0.18**	0.28***	0.37***	0.26***	0.49***	1.00	
Complexity	0.61***	0.59***	0.63***	0.27***	0.36***	0.55***	0.39***	0.70***	0.73***	1.00

Notes: This Table shows the pairwise correlation between different measures of economic specialization and measures related to economic exchange and technologies that facilitate it; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.2: Pre-Modern Economic Specialization and Other Intra-Ethnic Trade Related Outcomes

	Pre-Modern Measures of Intra-Ethnic Trade				
	Importance of Trade	Intercommu-nity Trade as Food Source	Money	Credit	Writing and Records
	(1)	(2)	(3)	(4)	(5)
Pre-modern Economic Specialization	0.16*	0.22***	0.22**	0.49***	0.51***
	(0.09)	(0.07)	(0.09)	(0.08)	(0.07)
Main Controls	Yes	Yes	Yes	Yes	Yes
All Additonal Controls	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.07	0.22	0.19	0.29	0.49
Observations	177	174	174	162	177

Notes: This Table establishes the positive economically and statistically positive association between pre-modern economic specialization and other pre-modern inter-community trade related outcomes at the ethnic level. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.3: Pre-Modern Economic Specialization (Share) and Other Intra-Ethnic Trade Related Outcomes

	Pre-Modern Measures of Intra-Ethnic Trade				
	Importance of Trade	Intercommu-nity Trade as Food Source	Money	Credit	Writing and Records
	(1)	(2)	(3)	(4)	(5)
Pre-Modern Economic Specialization (Share)	0.15*	0.24***	0.28***	0.49***	0.54***
	(0.09)	(0.07)	(0.08)	(0.08)	(0.07)
Main Controls	Yes	Yes	Yes	Yes	Yes
All Additonal Controls	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.07	0.23	0.21	0.28	0.51
Observations	177	174	174	162	177

Notes: This Table establishes the positive economically and statistically positive association between pre-modern economic specialization and other pre-modern intra-community trade related outcomes at the ethnic level. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C.4: Pre-Modern Economic Specialization (Dev) and Other Intra-Ethnic Trade Related Outcomes

	Pre-Modern Measures of Intra-Ethnic Trade				
	Importance of Trade	Intercommu- nity Trade as Food Source	Money	Credit	Writing and Records
	(1)	(2)	(3)	(4)	(5)
Pre-Modern Economic Specialization (Dev)	0.06	0.15*	0.17**	0.38***	0.38***
	(0.09)	(0.08)	(0.09)	(0.07)	(0.06)
Main Controls	Yes	Yes	Yes	Yes	Yes
All Additonal Controls	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.06	0.20	0.18	0.25	0.43
Observations	177	174	174	162	177

Notes: This Table establishes the positive economically and statistically positive association between pre-modern economic specialization and other pre-modern intra-community trade related outcomes at the ethnic level. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

D Intra-Ethnic Diversity, Economic Specialization and Pre-modern Economic Development
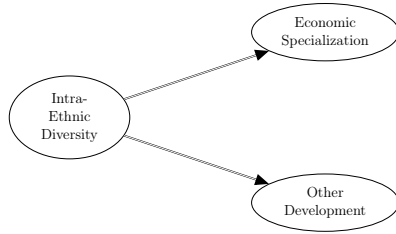
(a) Population Diversity causes Specialization and Other Development outcomes. Specialization and Other Development Outcomes do not cause each other.
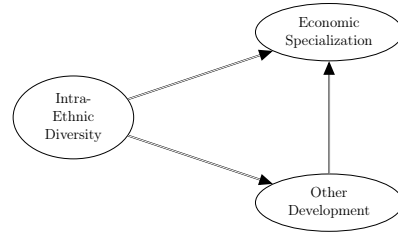
(b) Population Diversity causes Specialization and Other Development outcomes. Other Development Outcomes causes Specialization, but not the other way around.
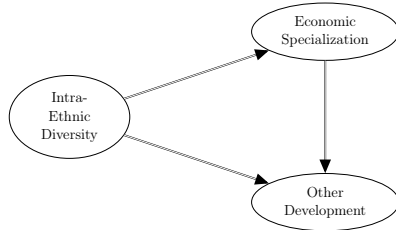
(c) Population Diversity causes Specialization and Other Development outcomes. Specialization causes Other Development Outcomes, but not the other way around.
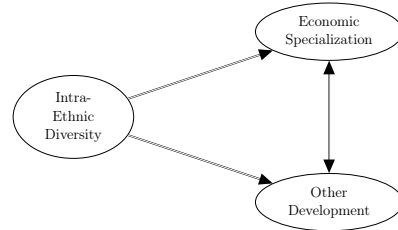
(d) Population Diversity causes Specialization and Other Development outcomes. Specialization and Other Development Outcomes cause each other.

Figure D.1: Potential Causal Relationships Between Intra-Ethnic Diversity, Economic Specialization and Development

Table D.1: Predicted Intra-Ethnic Diversity, Pre-Modern Development and Economic Specialization

	Economic Specialization				
	Log[Population Density, 1500CE]	Mean Size of Local Communities	Jurisdictional Hierarchy of Local Community	Class Stratification	Statehood Level
	(1)	(2)	(3)	(4)	(5)
	Panel A: Average Total Effect of Diversity				
Predicted Intra-Ethnic Diversity	0.56***	0.83***	0.59***	0.68***	0.62***
	(0.18)	(0.22)	(0.17)	(0.18)	(0.17)
Adjusted-R^2	0.33	0.34	0.33	0.33	0.33
	Panel B: Average Natural Direct Effect of Diversity				
Predicted Intra-Ethnic Diversity	0.56***	0.61***	0.71***	0.57***	0.39***
	(0.18)	(0.19)	(0.18)	(0.17)	(0.15)
Adjusted-R^2	0.33	0.51	0.34	0.36	0.48
	Panel C: Average Controlled Direct Effect of Diversity				
Predicted Intra-Ethnic Diversity	0.68***	0.76***	0.88***	0.71***	0.48***
	(0.22)	(0.23)	(0.21)	(0.21)	(0.18)
Adjusted-R^2	0.31	0.24	0.31	0.27	0.20
	Panel D: Average Natural Direct Effect of Diversity (Lewbel IV)				
Predicted Intra-Ethnic Diversity	0.55***	0.61***	0.74***	0.56***	0.40***
	(0.17)	(0.19)	(0.19)	(0.17)	(0.15)
First-stage F-statistic	8.88	24.25	3.66	5.77	5.48
Hansen's J-statistic	20.11	163.35	72.84	60.15	121.20
J-stat p-value	0.58	0.29	0.00	0.00	0.01
Adjusted-R^2	0.32	0.51	0.34	0.36	0.48
	Panel E: Average Controlled Direct Effect of Diversity (Lewbel IV)				
Predicted Intra-Ethnic Diversity	0.67***	0.76***	0.92***	0.70***	0.49***
	(0.22)	(0.23)	(0.21)	(0.21)	(0.18)
First-stage F-statistic	8.88	24.25	3.66	5.77	5.48
Hansen's J-statistic	20.11	163.35	72.84	60.15	121.20
J-stat p-value	0.58	0.29	0.00	0.00	0.01
Adjusted-R^2	0.28	0.23	0.31	0.27	0.21
Main Controls	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes
Development Control	Yes	Yes	Yes	Yes	Yes
Observations	910	509	923	879	912

Notes: This table explores the direct and indirect effects of intra-ethnic diversity on economic specialization. Panel A shows the average total effect of intra-ethnic diversity on economic specialization, which is based on our preferred specification in Table 7, for the subsamples for which pre-modern development measures are available. Panel B estimates the average natural direct effect of intra-ethnic diversity, which is the estimated effect after accounting for the potential mediating development channel (Pearl, 2001; Lange et al., 2012). Panel C estimates the average controlled direct effect of intra-ethnic diversity, which is the estimated effect of diversity on economic specialization after partialling out the potential mediating development channel (Pearl, 2001; Acharya et al., 2016). Panels D andE replicate the analyses in panels B and C, but exploit the Lewbel's atheoretical instrumental variable approach to instrument for the potential mediating development channel, addressing potential concerns due to endogeneity (Lewbel, 2012). Each column explores one potential development channel. These results account for the full set of controls in Table 7. Standardized coefficients. Bootstrapped standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table D.2: Predicted Intra-Ethnic Diversity, Pre-Modern Development and Economic Specialization
Exploring the Opposite Causal Relation

	Log[Population Density, 1500CE]	Mean Size of Local Communities	Jurisdictional Hierarchy of Local Community	Class Stratification	Statehood Level
	(1)	(2)	(3)	(4)	(5)
Panel A: Average Total Effect of Diversity					
Predicted Intra-Ethnic Diversity	0.01	-0.04	-0.92***	0.52***	0.42***
	(0.18)	(0.20)	(0.17)	(0.17)	(0.14)
Adjusted-R^2	0.33	0.36	0.18	0.29	0.36
Panel B: Average Natural Direct Effect of Diversity					
Predicted Intra-Ethnic Diversity	-0.02	-0.37*	-1.01***	0.38**	0.16
	(0.17)	(0.19)	(0.17)	(0.16)	(0.12)
Economic Specialization	0.05	0.40***	0.15***	0.21***	0.43***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
Adjusted-R^2	0.33	0.46	0.19	0.32	0.48
Panel C: Average Controlled Direct Effect of Diversity					
Predicted Intra-Ethnic Diversity	-0.04	-0.87*	-0.74***	0.21**	0.20
	(0.45)	(0.44)	(0.12)	(0.09)	(0.15)
Adjusted-R^2	0.32	0.25	0.16	0.23	0.22
Panel D: Average Natural Direct Effect of Diversity (Lewbel IV)					
Predicted Intra-Ethnic Diversity	0.02	-0.68*	-0.60***	0.16**	0.16
	(0.37)	(0.35)	(0.10)	(0.07)	(0.12)
Economic Specialization	-0.00	0.72***	0.09***	0.11***	0.42***
	(0.09)	(0.08)	(0.03)	(0.02)	(0.04)
First-stage F-statistic	54.72	48.92	54.81	51.25	54.47
Hansen's J-statistic	22.92	31.96	39.55	35.35	28.25
J-stat p-value	0.41	0.08	0.01	0.04	0.17
Adjusted-R^2	0.33	0.46	0.19	0.32	0.48
Panel E: Average Controlled Direct Effect of Diversity (Lewbel IV)					
Predicted Intra-Ethnic Diversity	0.03	-0.85*	-0.74***	0.20**	0.20
	(0.45)	(0.44)	(0.12)	(0.09)	(0.15)
Economic Specialization	-0.00	0.72***	0.09***	0.11***	0.42***
	(0.09)	(0.08)	(0.03)	(0.02)	(0.04)
First-stage F-statistic	54.72	48.92	54.81	51.25	54.47
Hansen's J-statistic	22.92	31.96	39.55	35.35	28.25
J-stat p-value	0.41	0.08	0.01	0.04	0.17
Adjusted-R^2	0.33	0.26	0.16	0.22	0.22
Main Controls	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes
Development Control	Yes	Yes	Yes	Yes	Yes
Observations	910	509	923	879	912

Notes: This table explores the direct and indirect effects of intra-ethnic diversity on pre-modern development and the potential mediating role of economic specialization. Panel A shows the average total effect of intra-ethnic diversity on pre-modern economic development, which is based on our preferred specification in Table 7, for the subsamples for which pre-modern development measures are available. Panel B estimates the average natural direct effect of intra-ethnic diversity, which is the estimated effect after accounting for the potential mediating effect of economic specialization (Pearl, 2001; Lange et al., 2012). Panel C estimates the average controlled direct effect of intra-ethnic diversity, which is the estimated effect of diversity on pre-modern development after partialling out the potential mediating effect of economic specialization (Pearl, 2001; Acharya et al., 2016). Panels D and E replicate the analyses in panels B and C, but exploit the Lewbel's atheoretical instrumental variable approach to instrument for the potential mediating effect of economic specialization, addressing potential concerns due to endogeneity (Lewbel, 2012). Each column explores one potential development channel. These results account for the full set of controls in Table 7. Standardized coefficients. Bootstrapped standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

E   Robustness

The results in the main text lend credence to the hypothesis that intra-ethnic diversity is conducive to the emergence and prevalence of pre-modern division of labor across ethnic groups. Nevertheless, the results conceivably might be biased due to the measure of economic specialization and the econometric method employed in the analysis, as well as due to omitted historical confounders, the existence of spatial or cultural dependence, or the presence of inter-ethnic interactions. This section explores the robustness of the results to these and other potential concerns.

E.1   Robustness to Alternative Measures of Specialization

This section explores the robustness of the estimated association between predicted intra-ethnic diversity and economic specialization to the measure of specialization used, the estimation method employed and multiple hypotheses testing. In particular, Table E.1 establishes that the main results presented so far hold for all three measures of economic specialization constructed in section 2 (see also Table E.2). Interestingly, intra-ethnic diversity's implied effect on economic specialization is stronger when the measure of economic specialization is based on the share of activities that are specialized. Since this measure should be the less affected by any potential confounding effect of development, this result suggests that other sources of development are not likely driving the estimated effect of intra-ethnic diversity.

Table E.1: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Measurement of Specialization

	Economic Specialization Measures								
	Count			Share			Score		
	OLS	IV	Full	OLS	IV	Full	OLS	IV	Full
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intra-Ethnic Diversity	0.27***	0.46***	0.59***	0.33***	0.37***	0.73***	0.13**	0.31**	0.41**
	(0.05)	(0.14)	(0.21)	(0.05)	(0.12)	(0.17)	(0.06)	(0.14)	(0.18)
Main Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
All Additional Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Continental FE	No	No	Yes	No	No	Yes	No	No	Yes
First-stage F-statistic		81.54			81.54			81.54	
R^2	0.08	0.49	0.34	0.11	0.49	0.40	0.02	0.46	0.25
Adjusted-R^2	0.07	0.39	0.32	0.10	0.39	0.39	0.01	0.35	0.23
Observations	116	116	934	116	116	934	116	116	934

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity, as predicted by the distance to East Africa (see section 3.2), and various measures of economic specialization. These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Bootstrapped standard error estimates in parenthesis; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table E.2: Intra-Ethnic Diversity and Economic Specialization:
Robustness to Specialization Measure

	Economic Specialization Measures					
	Count		Share		Score	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Intra-Ethnic Diversity	0.27***	0.46***	0.33***	0.37***	0.13**	0.31**
	(0.05)	(0.14)	(0.05)	(0.12)	(0.06)	(0.14)
Main Controls	No	Yes	No	Yes	No	Yes
All Additional Controls	No	Yes	No	Yes	No	Yes
First-stage F-statistic		81.54		81.54		81.54
R^2	0.08	0.49	0.11	0.49	0.02	0.46
Adjusted-R^2	0.07	0.39	0.10	0.39	0.01	0.35
Observations	116	116	116	116	116	116

Notes: This Table establishes the robustness of the OLS and IV main results in Tables 1 and 5 to the election of the specialization measure. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

E.2 Robustness to Empirical Specification

Another potential concern with these results is that they may be driven by the empirical specification chosen. In particular, the main measure of economic specialization is a count variable, while another is a share, which may cause OLS estimates to be biased. On the other hand, the results may be biased given the large number of societies that do not have economic specialization. In order to mitigate these concerns, the research replicates the analysis employing Poisson, Negative Binomial, and fractional regression methods, as well as their zero-inflated variants and zero-inflated beta regressions (see Appendix E.2). Reassuringly, the results are robust to the empirical specification chosen and the estimated effect of predicted intra-ethnic diversity on economic specialization remains positive, statistically and economically significant.

Table E.3: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Estimation Method

| | Economic Specialization | | | |
	Poisson	Negative Binomial	Zero-inflated Poisson	Zero-inflated Negative Binomial
	(1)	(2)	(3)	(4)
	Panel A: Economic Specialization			
Predicted Intra-Ethnic Diversity	0.59***	0.59***	0.62***	0.62***
	(0.22)	(0.23)	(0.22)	(0.23)
	Panel B: Probability Economic Specialization is always equal to Zero			
Predicted Intra-Ethnic Diversity			-10.54**	-10.69**
			(4.72)	(4.91)
Continental FE	Yes	Yes	Yes	Yes
Main Controls	Yes	Yes	Yes	Yes
Additonal Controls	Yes	Yes	Yes	Yes
Pseudo-R^2	0.25	0.21		
Observations	934	934	934	934
α		0.05		0.04
Log-likelihood	-932.93	-932.07	-911.76	-911.08
BIC	2043.70	2048.81	2049.23	2054.70
AIC	1917.87	1918.14	1889.53	1890.16

Notes: This Table establishes the robustness of the results to estimation method. In particular, the results employ estimation method better suited for count variables. Column 1 shows the results of a Poisson regression, column 2 of a Negative-Binomial, and columns 3 and 4 the results of zero-inflated Poisson and Negative Binomial regressions respectively. Panel A establishes the positive effect of predicted population on economic specialization (conditional on having economic specialization). Additionally, Panel B establishes the negative effect of intra-ethnic diversity on the probability of not having any economic specialization. Coefficients show effect of increasing predicted intra-ethnic diversity by 1 standard deviation. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

(a) Poisson

(b) Negative Binomial

(c) Zero-Inflated Poisson

(d) Zero-Inflated Negative Binomial

Figure E.1: Robustness to Estimation Method
Observed and Predicted Probabilities in Count Regressions

Table E.4: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Estimation Method (Share)

	Economic Specialization (Share)		
	Fractional Regression		
	Logit	Probit	Zero-inflated Beta
	(1)	(2)	(3)
	Panel A: Economic Specialization		
Predicted Intra-Ethnic Diversity	0.88***	0.49***	0.42*
	(0.28)	(0.15)	(0.22)
Std-β	0.09***	0.09***	0.11***
	(0.03)	(0.03)	(0.02)
	Panel B: Probability Economic Specialization is always equal to Zero		
Predicted Intra-Ethnic Diversity			-1.85***
			(0.19)
Continental FE	Yes	Yes	Yes
Main Controls	Yes	Yes	Yes
Additonal Controls	Yes	Yes	Yes
Observations	934	934	934
Log-likelihood	-303.26	-303.19	-68.34
BIC	784.36	784.20	369.22
AIC	658.53	658.37	204.68

Notes: This Table establishes the robustness of the results to estimation method. In particular, the results employ estimation method better suited for fractional outcome variables. Column 1 shows the results of a fractional Logit regression, column 2 of a fractional Probit regression, and columns 3 the results of a Beta regression. Panel A establishes the positive effect of predicted population on economic specialization (conditional on having economic specialization). Additionally, Panel B establishes the negative effect of intra-ethnic diversity on the probability of not having any economic specialization. Coefficients show effect of increasing predicted intra-ethnic diversity by 1 standard deviation. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

E.3   Robustness to Multiple-Hypothesis Testing and False Discovery rates

This section explores the effect of multiple hypotheses testing using the methods introduced by Benjamini and Hochberg (1995) and Benjamini et al. (2006). Specifically, Benjamini and Hochberg (1995) and Benjamini et al. (2006) introduced False Discovery Rates (FDR) corrected q-values, which have similar interpretations as p-values and give the probability threshold at which the Null hypothesis would be rejected after accounting for multiple hypotheses testing (Anderson, 2008). We performed the correction of multiple hypotheses testing including all variables available in the Ethnographic Atlas and SCCS, which had an effective sample size of over 50 ethnicities to avoid inflating the number of not statistically significant associations. Reassuringly, our main results remain unchanged (Tables E.5 and E.6).

Table E.5: Robustness to Multiple Hypothesis Testing
Accounting for False Discovery Rates

	OLS-EA	OLS-EA	IV-EA	IV-EA	IV-EA-SCCS
	(1)	(2)	(3)	(4)	(5)
Coefficient	9.973	8.614	12.536	12.725	12.725
Uncorrected p-value	0.000	0.003	0.001	0.001	0.001
BH FDR corrected q-value	0.002	0.016	0.008	0.004	0.024
BKY sharpened FDR corrected q-value	0.002	0.013	0.007	0.004	0.024
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes
Extended Geographical Controls	No	Yes	Yes	Yes	Yes
Continental FE	No	No	No	Yes	Yes
Predicted Sample	No	No	No	Yes	Yes
Number of Hypotheses Tested	60	60	60	74	468

Notes: This Table presents (non-standardized) estimated coefficients and p-values based on the most complete specifications in Tables 1 (column 1), 2 (column 2), 5 (column 3), and 7 (columns 4-5). Intra-ethnic diversity is measured by genetic diversity. FDR correction based on running alternative regressions on all variables of the Ethnographic Atlas and SCCS that satisfied the following conditions: (i) the variable was not used in the construction of our specialization measure, (ii) the sample in the regression included at least 50 ethnicities, and (iii) there was any variation in the sample. Given these restrictions, SCCS can only be used in the predicted sample. Uncorrected and corrected p-values are based on heteroskedasticity robust standard errors. False Discovery Rates (FDR) corrected q-values based on the methods presented in Benjamini and Hochberg (1995) (BH) and Benjamini et al. (2006) using the implementation of Anderson (2008). q-values have similar interpretation to p-values and give the probability threshold at which the Null hypothesis would be rejected after accounting for multiple hypothesis testing (see Anderson (2008)).

Table E.6: Robustness to Multiple Hypothesis Testing
Accounting for False Discovery Rates

	Economic Specialization								
	Linguistic Diversity (Consonant Inventory)			Linguistic Diversity (Vowel Quality Inventory)			Linguistic Diversity (Number of Genders)		
	OLS	IV	IV	OLS	IV	IV	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Coefficient	0.264	1.493	1.620	0.835	2.519	2.820	0.157	1.452	1.188
Uncorrected p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.089	0.000	0.006
BH FDR corrected q-value	0.010	0.001	0.001	0.001	0.001	0.001	0.377	0.001	0.035
BKY sharpened FDR corrected q-value	0.010	0.001	0.001	0.001	0.001	0.001	0.427	0.001	0.033
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Geographical Controls	No	No	Yes	No	No	Yes	No	No	Yes
Number of Hypotheses Tested	74	74	74	74	74	74	68	68	68

Notes: This Table presents (non-standardized) estimated coefficients and p-values based on the most complete specifications in Tables 1 (column 1), 2 (column 2), 5 (column 3), and 7 (columns 4-5). Intra-ethnic diversity is measured by linguistic diversity. FDR correction based on running alternative regressions on all variables of the Ethnographic Atlas and SCCS that satisfied the following conditions: (i) the variable was not used in the construction of our specialization measure, (ii) the sample in the regression included at least 50 ethnicities, and (iii) there was any variation in the sample. Given these restrictions, SCCS can only be used in the predicted sample. Uncorrected and corrected p-values are based on heteroskedasticity robust standard errors. False Discovery Rates (FDR) corrected q-values based on the methods presented in Benjamini and Hochberg (1995) (BH) and Benjamini et al. (2006) using the implementation of Anderson (2008). q-values have similar interpretation to p-values and give the probability threshold at which the Null hypothesis would be rejected after accounting for multiple hypothesis testing (see Anderson (2008)).

E.4 Robustness to Spatial Auto-correlation and Cultural Dependence

This section explores the robustness of the results to spatial auto-correlation and cultural dependence. In particular, a potential concern with the previous results is that migration and splitting of ancestral groups may have created historical, cultural and spatial correlation among ethnic outcomes. In order to mitigate these potential concerns, the analysis follows various strategies: (i) it accounts for the potential spatial dependence of observations by allowing the error terms to be spatially auto-correlated. (ii) It accounts for the potential historical and cultural dependence of observations by allowing the error terms to be correlated within language phyla (i.e., a group of phylogenetically related languages inside a linguistic family). (iii) It constrains the analysis to the sample of ethnicities in the Standard Cross-Cultural Sample (SCCS), which was selected by ethnographers to minimize cultural and spatial dependence across observations. (iv) It accounts for language phylum fixed-effects, and thus for omitted heterogeneity across culturally and historically related ethnic groups.

Reassuringly, accounting for spatial auto-correlation and cultural dependence does not qualitatively alter the results (see in particular Appendix E.4). Specifically, Tables E.8, E.10-E.12 show that accounting for spatial auto-correlation does not affect the statistical significance of the results. Additionally, Tables E.8-E.12 establish that clustering standard errors at the language phyla (i.e., a group of phylogenetically related languages inside a linguistic family) level does not affect the statistical significance of the results. Finally, Table E.16 shows that constraining the sample to ethnicities selected to minimize cultural and historical dependence or accounting for language phylum fixed-effects, and thus for omitted heterogeneity across culturally related ethnic groups does not affect the qualitative nature of the results.

Table E.7: Expected Heterozygosity and Economic Specialization

	Economic Specialization						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expected	0.27***	0.31***	0.27***	0.27***	0.25**	0.27***	0.36**
Heterozygosity	(0.05)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)	(0.14)
Absolute Latitude		0.15					0.80
		(0.13)					(0.47)
Area			0.01				0.01
			(0.05)				(0.06)
Elevation (Avg.)				-0.03			0.31*
				(0.07)			(0.17)
Precipitation (Avg.)					-0.08		0.13
					(0.07)		(0.21)
Temperature (Avg.)						0.04	0.73*
						(0.08)	(0.37)
Adjusted-R^2	0.07	0.08	0.06	0.06	0.06	0.06	0.15
Observations	116	116	116	116	116	116	116

Notes: This Table establishes the robustness of the OLS results in Table 1 to correlation in the error term within a language phylum level (i.e., a group of phylogenetically related languages inside a linguistic family). Standardized coefficients. Heteroskedasticity robust standard error estimates clustered at the language phylum level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.
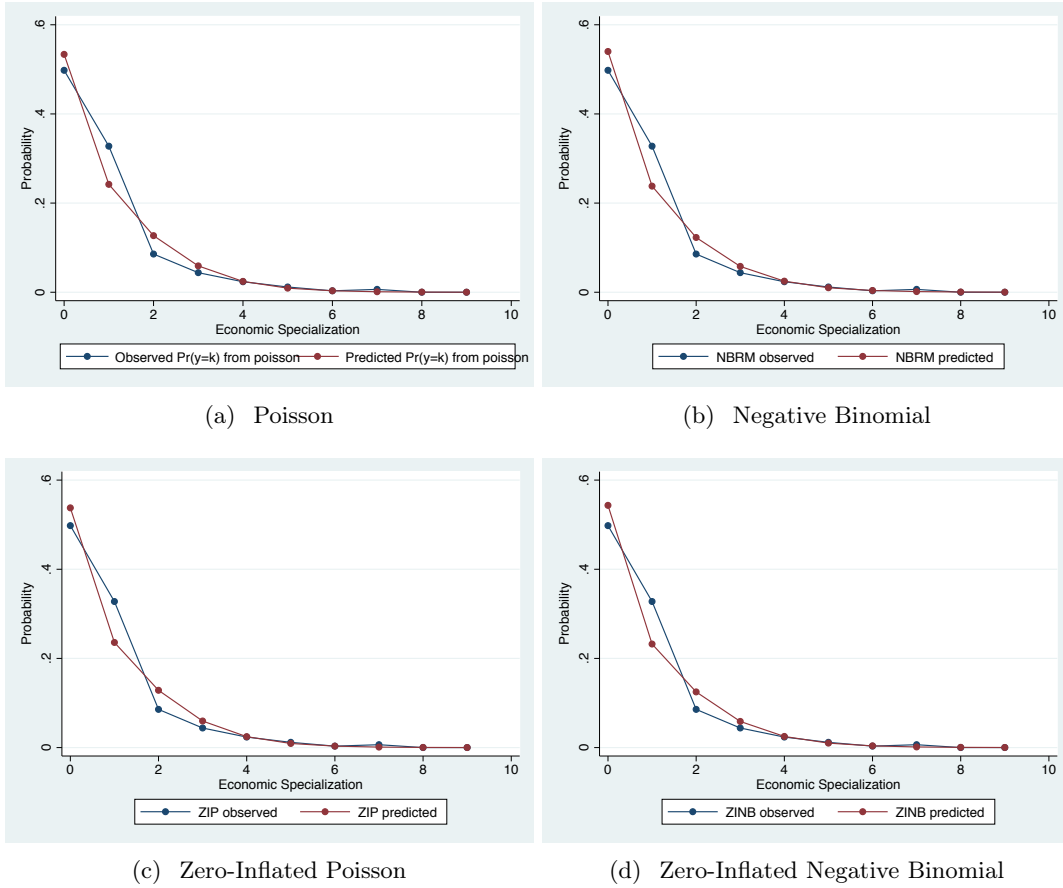
Table E.8: Intra-Ethnic Diversity and Economic Specialization
(Robustness to Clustering and Spatial Auto-Correlation)

	Economic Specialization								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intra-Ethnic Diversity	0.36***	0.41***	0.36***	0.37***	0.34***	0.33***	0.31***	0.40***	0.31***
	(0.09)	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)	(0.09)	(0.11)	(0.10)
	([0.14])	([0.14])	([0.15])	([0.13])	([0.16])	([0.15])	([0.13])	([0.16])	([0.12])
	[0.11]	[0.10]	[0.11]	[0.10]	[0.12]	[0.11]	[0.10]	[0.12]	[0.09]
	{0.10}	{0.10}	{0.09}	{0.10}	{0.10}	{0.10}	{0.10}	{0.10}	{0.09}
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.15	0.22	0.21	0.17	0.18	0.18	0.20	0.17	0.40
Observations	116	116	116	116	116	116	116	116	116

Notes: This Table establishes the robustness of the OLS results to clustering by language phylum and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table 2. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum (i.e., a group of phylogenetically related languages inside a linguistic family) in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table E.9: Intra-Ethnic Diversity and Economic Specialization (IV)

	OLS	IV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intra-Ethnic Diversity	0.36**	0.51***	0.53***	0.54***	0.54***	0.45***	0.49***	0.44***	0.56***	0.46***
	(0.14)	(0.15)	(0.15)	(0.15)	(0.15)	(0.17)	(0.15)	(0.13)	(0.17)	(0.16)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All
First-stage F-statistic		37.90	40.37	39.65	46.73	32.32	36.32	35.80	47.55	69.10
Adjusted-R^2	0.15	0.14	0.21	0.19	0.15	0.17	0.16	0.19	0.15	0.39
Observations	116	116	116	116	116	116	116	116	116	116

Notes: This Table establishes the robustness of the IV results to clustering by language phylum (i.e., a group of phylogenetically related languages inside a linguistic family). The additional controls in each column are the ones of the same column in Table 5. Standardized coefficients. Heteroskedasticity robust standard error estimates clustered at the linguistic phyllum level (i.e., a group of phylogenetically related languages inside a linguistic family) are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table E.10: Intra-Ethnic Diversity and Distance to East Africa
(Robustness to Clustering and Spatial Auto-Correlation)

	Intra-Ethnic Diversity								
	Full Sample								Specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial	-0.85***	-0.80***	-0.80***	-0.81***	-0.80***	-0.79***	-0.80***	-0.82***	-0.85***
Distance	(0.07)	(0.10)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)
to East Africa	([0.09])	([0.12])	([0.11])	([0.11])	([0.13])	([0.12])	([0.12])	([0.10])	([0.10])
	[0.08]	[0.10]	[0.10]	[0.09]	[0.11]	[0.11]	[0.10]	[0.08]	[0.08]
	{0.04}	{0.05}	{0.05}	{0.05}	{0.06}	{0.06}	{0.06}	{0.06}	{0.10}
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Agr. Suit.	CSI	Volatility	Sea	Mobility	All
Adjusted-R^2	0.72	0.72	0.74	0.72	0.72	0.72	0.73	0.75	0.73
R^2	0.72	0.74	0.75	0.74	0.74	0.74	0.75	0.78	0.76
Observations	144	144	144	144	144	144	144	144	116

Notes: This Table establishes the robustness of the negative effect of the migratory distance on intra-ethnic diversity to clustering by language phylum(i.e., a group of phylogenetically related languages inside a linguistic family) and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table 3. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum (i.e., a group of phylogenetically related languages inside a linguistic family) in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table E.11: Intra-Ethnic Diversity and Economic Specialization (Reduced Form)
(Robustness to Clustering and Spatial Auto-Correlation)

	Economic Specialization								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial	-0.41***	-0.43***	-0.44***	-0.45***	-0.37***	-0.40***	-0.36***	-0.46***	-0.39***
Distance to Addis	(0.10)	(0.10)	(0.10)	(0.09)	(0.12)	(0.10)	(0.09)	(0.11)	(0.13)
Ababa	([0.12])	([0.13])	([0.12])	([0.11])	([0.15])	([0.13])	([0.11])	([0.15])	([0.15])
	[0.09]	[0.10]	[0.10]	[0.09]	[0.11]	[0.10]	[0.09]	[0.10]	[0.10]
	{0.10}	{0.10}	{0.10}	{0.10}	{0.11}	{0.10}	{0.10}	{0.10}	{0.10}
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All
Adjusted-R^2	0.17	0.23	0.25	0.21	0.19	0.21	0.22	0.19	0.42
Observations	116	116	116	116	116	116	116	116	116

Notes: This Table establishes the robustness of the OLS results to clustering by language phylum (i.e., a group of phyloge- netically related languages inside a linguistic family) and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table G.1. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum (i.e., a group of phylogenetically related languages inside a linguistic family) in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table E.12: Intra-Ethnic Diversity and Economic Specialization (Reduced Form)
(Robustness to Clustering and Spatial Auto-Correlation)

	Economic Specialization										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Pre-Industrial	-0.45***	-0.42***	-0.45***	-0.42***	-0.41***	-0.45***	-0.40***	-0.42***	-0.42***	-0.47***	-0.36***
Distance to	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.10)
East Africa	([0.08])	([0.08])	([0.09])	([0.07])	([0.07])	([0.09])	([0.07])	([0.08])	([0.08])	([0.09])	([0.20])
	[0.07]	[0.06]	[0.06]	[0.06]	[0.06]	[0.07]	[0.06]	[0.06]	[0.06]	[0.06]	[0.18]
	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.09}
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All	All
Continental FE	No	No	No	No	No	No	No	No	No	No	Yes
Adjusted-R^2	0.20	0.22	0.23	0.24	0.24	0.24	0.24	0.22	0.23	0.29	0.32
Observations	932	932	932	932	932	932	932	932	932	932	932

Notes: This Table establishes the robustness of the OLS results to clustering by language phylum (i.e., a group of phyloge-netically related languages inside a linguistic family) and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table G.1. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum (i.e., a group of phylogenetically related languages inside a linguistic family) in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

E.5 Robustness to Inter-Ethnic Interactions

This section explores the robustness of the results to the potential presence of inter-ethnic interactions. Specifically, the results may be biased due to interactions between an ethnicity and its neighbors, since close contact with neighboring ethnicities may affect information diffusion and trade opportunities, and thus potentially the prevalence of economic specialization. Moreover, inter-ethnic interactions may have also affected the levels of intra-ethnic diversity. Furthermore, the historical migration process out of Africa may not only have affected intra-ethnic diversity, but also the potential for interaction among ethnicities. Thus, the results may be reflecting the effect of inter-ethnic relations as opposed to intra-ethnic diversity.

Table E.13: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Distances to Other Ethnicities

	Economic Specialization				
	(1)	(2)	(3)	(4)	(5)
Predicted Intra-Ethnic Diversity	0.57***	0.63***	0.64***	0.66***	0.70***
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
Pre-Industrial Distance to Closest 5 Ethnicities		0.13***			
		(0.04)			
Pre-Industrial Distance to Closest 10 Ethnicities			0.12***		
			(0.04)		
Pre-Industrial Distance to Closest 25 Ethnicities				0.12***	
				(0.04)	
Pre-Industrial Distance to Closest 50 Ethnicities					0.13***
					(0.04)
Main Controls	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.33	0.34	0.34	0.33	0.33
Observations	930	930	930	930	930

Notes: This table shows the robustness of the positive statistically and economically significant association between intra-ethnic diversity and economic specialization to its pre-industrial distance to other ethnicities. Standardized coefficients. Heteroskedasticity robust standard error estimates clustered at the language phylum level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Tables E.13 and E.15 explore the potential bias due to inter-ethnic interactions. In particular, Table E.13 accounts for the potential interaction between an ethnic group and neighboring ethnicities as reflected by geographical proximity. Reassuringly, the inclusion of migratory distances to various sets of neighboring ethnicities does not affect the qualitatively nature of the results. Additionally, Table E.15 explores the effect of potential inter-ethnic interactions at a global scale. It accounts for the centrality of an ethnicity in the global potential network generated by the minimum spanning tree based on the migratory paths among ethnicities, which should reflect the underlying information and trade networks among ethnic groups. The results suggest that the centrality of an ethnic group in this global migratory network does not affect the qualitative results. Thus, intra-ethnic diversity remains positively and statistically associated with economic specialization. Moreover, these results provide additional supporting evidence that the migratory distance from East Africa may indeed satisfy the exclusion restriction.

Table E.14: Robustness to Distances to Other Ethnicities

	Economic Specialization				
	(1)	(2)	(3)	(4)	(5)
Predicted Intra-Ethnic Diversity	0.57***	0.61***	0.54***	0.59***	0.65***
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
Pre-Industrial Distance to Ethnicities within 1 weeks		-0.10***			
		(0.03)			
Pre-Industrial Distance to Ethnicities within 2 weeks			0.03		
			(0.03)		
Pre-Industrial Distance to Ethnicities within 5 weeks				0.15***	
				(0.04)	
Pre-Industrial Distance to Ethnicities within 10 weeks					0.19***
					(0.06)
Main Controls	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.33	0.34	0.33	0.34	0.34
Observations	932	932	932	932	932

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table E.15: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Centrality in Ethnic Network

	Economic Specialization						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Intra-Ethnic Diversity	0.59***	0.56***	0.59***	0.59***	0.59***	0.62***	0.59***
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
Betweenness Network Centrality	0.01						
	(0.03)						
Closeness Network Centrality		0.05*					
		(0.03)					
Closeness Vitality Network Centrality			0.01				
			(0.03)				
Degree Network Centrality				-0.00			
				(0.03)			
Eigenvector Network Centrality					-0.01*		
					(0.01)		
Katz Network Centrality						0.05*	
						(0.03)	
Load Network Centrality							0.01
							(0.03)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.32	0.33	0.32	0.32	0.32	0.33	0.32
Observations	934	934	934	934	934	934	934

Notes: This table shows the robustness of the positive statistically and economically significant association between intra-ethnic diversity and economic specialization to measures of an ethnicity's network centrality. The network of ethnicities is the minimum spanning tree of the network that connects all ethnic groups by their minimum travel time paths. Betweenness centrality of an ethnicity e is the sum of the fraction of all-pairs of ethnicities' shortest paths that pass through e. Closeness centrality of an ethnicity e is the reciprocal of the sum of the shortest path distances from e to all ethnicities. Closeness vitality of an ethnicity e is the change in the sum of distances between all pairs of ethnicities when excluding that e. The degree centrality for an ethnicity is the fraction of ethnicities it is connected to. Eigenvector centrality computes the centrality for an ethnicity based on the centrality of its neighbors. Katz centrality computes the centrality for a node based on the centrality of its neighbors. The load centrality of an ethnicity e is the fraction of all shortest paths that pass through e. All columns account for the set of basic geographical controls of Table 1, the extended set of confounders from Table 2, and continental fixed effects. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

E.6 Robustness to Historical Confounders

This section establishes the robustness of the estimated positive relation between intra-ethnic diversity and economic specialization to accounting for potential historical confounders. Thus, mitigating the potential concern that intra-ethnic diversity is capturing the effect of factors like the transition to agriculture, the history of settlement or technological diffusion on economic specialization. Moreover, it further explores the potential confounding effects of common ancestry or historical interaction with other ethnicities. Additionally, it also explores the robustness of these results to accounting for cultural, historical and spatial dependence.

Table E.16: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Historical Confounders

| | Economic Specialization | | | | | | | |
| | Full Sample | | | | | | | SCCS |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted Intra-Ethnic Diversity	0.35**	0.35**	0.34**	0.34**	0.35**	0.36**	0.28**	0.51***
	(0.16)	(0.16)	(0.16)	(0.14)	(0.16)	(0.17)	(0.12)	(0.11)
Pre-Industrial Distance to Neolithic Frontier		-0.07**						
		(0.03)						
Pre-Industrial Distance to Technological Frontier (1500CE)			-0.12**					
			(0.05)					
Pre-Industrial Isolation from All Other Ethnicities				-0.08				
				(0.10)				
Pre-Industrial Distance to Closest 5 Ethnicities					0.07			
					(0.06)			
Pre-Industrial Distance to Closest 25 Ethnicities						0.03		
						(0.06)		
Duration of Continuous Human Presence							0.18	
							(0.14)	
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Language Phylum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.45	0.46	0.46	0.46	0.46	0.45	0.46	0.48
Observations	921	921	921	921	921	921	921	166

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity, as predicted by the distance to East Africa (see section 3.2), and economic specialization after accounting for other potential historical sources of specialization and development. All columns account for the set of geographical controls of Table 7 and language phylum fixed effects. Standardized coefficients. Heteroskedasticity robust standard error estimates clustered at the language phylum level in parenthesis; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table E.16 explores the robustness of the results to historical confounders. In particular, column 1 explores the effect of accounting for language phylum fixed effects and additionally clustering standard errors estimates at the same level. By exploiting only within-philum variation, the analysis mitigates potential biases due to cultural, historical or spatial dependence among ethnicities. Reassuringly, the results remain qualitatively unchanged. Column 2 additionally analyzes the potential confounding effect of the long-lasting influence of the Neolithic Revolution, by accounting for the pre-industrial

distance to the closest Neolithic frontier (i.e., the closest location of animal or plant domestication).[50] As argued by Diamond (1997), an earlier transition from hunting and gathering practices to agriculture provided an initial advantage to some societies, which later translated into a persistent technological superiority. Moreover, it has been suggested that an earlier transition to agriculture allowed the creation of an economic surplus and the emergence of economic specialization (Boix, 2015). Additionally, country-level precolonial development has been positively associated with the time since the Neolithic Revolution (Ashraf and Galor, 2011). In line with these findings, column 2 shows that the pre-industrial distance to the closest Neolithic frontier does have a negative effect on economic specialization. Still, the estimated effect of predicted intra-ethnic diversity on economic specialization remains positive strongly statistically and economically significant, suggesting that the omission of the Neolithic transition was not spuriously driving the main results.[51]

Additionally, column 3 analyzes the potential confounding effect of the pre-modern distance from the closest technological frontier in the year 1500CE as identified by Ashraf and Galor (2011).[52] In particular, if technology diffuses from a technological frontier, one can expect ethnicities close to the frontier to acquire more technologies and develop economically, all of which might potentially be conducive to economic specialization. Indeed, the estimated association between the distance from the frontier and economic specialization is negative statistically and economically significant, in line with this prediction. Nonetheless, the estimated positive association between predicted intra-ethnic diversity and economic specialization remains statistically and economically significant.

Columns 4-6 explore the potential confounding effect of an ethnicity's location relative to other ethnicities. In particular, relative isolation from other ethnic groups may affect inter-ethnic trade and conflict, as well as innovation (Ashraf et al., 2010; Özak, 2018). Thus, the analysis explores the potential effect of accounting for an ethnicity's level of pre-industrial isolation from all other ethnicities (i.e., its average pre-industrial distance to all other ethnicities), and its average distance to the closest 5 or 25 ethnic groups. Accounting for these average distances has no effect on the results.[53]

Column 7 includes an indicator of the duration of human settlements since prehistoric times, which estimates the date since the first uninterrupted settlement by anatomically modern humans (Ahlerup and Olsson, 2012).[54] Clearly, this measure should be highly correlated with migratory distance to East Africa and intra-ethnic diversity, since the closer a location is to East Africa, the earlier it could have been populated by anatomically modern humans. Thus, the omission of the duration of continuous human presence may bias the estimated effect of intra-ethnic diversity documented above, if a longer history of uninterrupted settlement facilitated the division of labor via, for example, a greater chance for the emergence of social stratification or a dominant elite.[55] However, the results in column 7 reveal that accounting for the duration of continuous human presence has no major impact on the estimated coefficient for predicted intra-ethnic diversity.

Column 8 replicates the analysis on the subsample of ethnic groups that belong to the Standard Cross Cultural Sample (SCCS). As explained in section 2.2, the SCCS sample was selected by ethno-

---

[50]The analysis estimates the minimal travel paths based on HMISea from the centroid of each ethnic homeland to the closest Neolithic frontier. The location of Neolithic frontiers is taken from various sources (Diamond, 1997; Smith, 1997; Benz, 2001; Denham et al., 2003; Pinhasi et al., 2005; Smith, 2006; Dillehay et al., 2007; Lu et al., 2009; Manning et al., 2011; Linseele, 2013).

[51]Alternatively, accounting for the degree of subsistence dependence on agriculture, as measured in the Ethnographic Atlas (v5), does not alter the results either.

[52]The technological frontiers are London and Paris in Europe, Fez and Cairo in Africa, Constantinople and Peking in Asia, and Tenochtitlan and Cuzco in the Americas.

[53]Similar results are obtained if one accounts for the closest 10 or 50 groups.

[54]Given that the original data is available at the country level, the analysis follows the literature and constructs ethnic level measures by creating population-weighted averages (Alesina et al., 2013; Giuliano and Nunn, 2013; Gennaioli and Rainer, 2007).

[55]In fact, Ahlerup and Olsson (2012) show that the historical duration of human settlements is a strong predictor of ethnolinguistic fractionalization.

graphers in order to minimize the potential spatial and historical dependence among ethnic groups in order to overcome Galton's independence problem. Encouragingly, the qualitative results remain unchanged, although the coefficient increases by almost 50%.

F Genetic Diversity and Economic Specialization

F.1 Robustness to Continental Fixed Effects

Table F.1: Predicted Intra-Ethnic Diversity and Economic Specialization
(Robustness to Continental Fixed Effects)

	Economic Specialization									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Predicted Intra-Ethnic Diversity	0.73***	0.57***	0.58***	0.51***	0.51***	0.53***	0.59***	0.55***	0.57***	0.58***
	(0.15)	(0.15)	(0.15)	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)	(0.15)	(0.17)
Malaria Ecology			-0.13***							-0.09*
			(0.05)							(0.05)
Ecological Diversity				0.12***						0.09***
				(0.03)						(0.03)
Agricultural Suitability (avg.)					-0.07*					-0.16***
					(0.04)					(0.04)
Agricultural Suitability (std.)					0.04					0.01
					(0.05)					(0.05)
Caloric Suitability Index (Pre-1500CE)						0.06*				0.11***
						(0.03)				(0.03)
Caloric Suitability (Pre-1500 ,std.)						0.08*				0.07
						(0.04)				(0.05)
Temperature (Spatial Corr., Avg.)							0.01			0.03
							(0.03)			(0.04)
Temperature (Volatility, Avg)							-0.25***			-0.14*
							(0.06)			(0.08)
Pct. Area within 100kms of Sea								0.05		-0.02
								(0.03)		(0.04)
Coast Length								0.01		0.01
								(0.03)		(0.03)
Ruggedness (Avg.)									0.10	0.08
									(0.06)	(0.07)
Pre-Industrial Mobility (avg.)									0.15	0.28**
									(0.10)	(0.11)
Pre-Industrial Mobility (std.)									-0.04	-0.14*
									(0.08)	(0.08)
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.26	0.29	0.29	0.30	0.29	0.29	0.29	0.29	0.29	0.32
Observations	934	934	934	934	934	934	934	934	934	934

Notes: This Table establishes the robustness of the results in 7 to the inclusion of continental fixed effects. It establishes the positive statistically and economically significant effect of intra-ethnic (genetic) diversity as predicted by the distance to East Africa (see section 3.2) on economic specialization. These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2.Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

F.2 Robustness Observation Period

Table F.2: Intra-Ethnic Diversity and Economic Specialization
Robustness to accounting for Observation Date

	Economic Specialization					
	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Intra-Ethnic Diversity	0.27***	0.36***	0.32***	0.44***	0.52***	0.59***
	(0.04)	(0.09)	(0.10)	(0.14)	(0.04)	(0.17)
Observation Date	Yes	Yes	Yes	Yes	Yes	Yes
Main Controls	No	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes	Yes	Yes
Continental FE	No	No	No	No	No	Yes
Adjusted-R^2	0.07	0.16	0.42	0.41	0.28	0.33
R^2	0.08	0.21	0.52	0.51	0.29	0.34
Observations	116	116	116	116	933	933

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

F.3 Robustness to Continental Fixed Effects

Table F.3: Intra-Ethnic Diversity and Distance to Addis-Ababa
Robustness to Continental Fixed Effects

	Population Diversity								
	Full Sample								Specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance	-0.53***	-0.41**	-0.48**	-0.41**	-0.42**	-0.40**	-0.42**	-0.48***	-0.51***
to East Africa	(0.12)	(0.18)	(0.18)	(0.18)	(0.18)	(0.19)	(0.18)	(0.18)	(0.19)
Malaria Ecology			0.10					0.12	0.06
			(0.06)					(0.07)	(0.08)
Agricultural Suitability (avg.)				0.00				0.01	-0.00
				(0.08)				(0.11)	(0.13)
Agricultural Suitability (std.)				0.12*				0.17**	0.19**
				(0.07)				(0.08)	(0.09)
Caloric Suitability (Pre-1500 ,avg.)					0.03			0.05	0.08
					(0.08)			(0.10)	(0.11)
Caloric Suitability (Pre-1500 ,std.)					-0.05			-0.12*	-0.14*
					(0.05)			(0.07)	(0.08)
Pct. Area within 100kms of Sea						0.04		0.13**	0.15**
						(0.07)		(0.06)	(0.07)
Coast Length						0.04		0.07	0.04
						(0.06)		(0.05)	(0.06)
Ruggedness (Avg.)							-0.13	-0.23	-0.25
							(0.14)	(0.15)	(0.17)
Pre-Industrial Mobility (avg.)							0.14	0.26	0.35
							(0.20)	(0.23)	(0.25)
Pre-Industrial Mobility (std.)							-0.07	-0.08	-0.08
							(0.09)	(0.11)	(0.12)
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.75	0.74	0.74	0.74	0.74	0.74	0.74	0.76	0.74
R^2	0.76	0.76	0.76	0.77	0.76	0.76	0.77	0.79	0.78
Observations	144	144	144	144	144	144	144	144	116

Notes: This table shows robustness of the negative statistically and economically significant relation between intra-ethnic diversity and the distance to East Africa, shown in Table 3, to accounting for continental fixed effects. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table F.4: Intra-Ethnic Diversity and Economic Specialization (IV)
Robustness to Continental Fixed Effects

| | Economic Specialization | | | | | | | | | |
| | OLS | IV | | | | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intra-Ethnic Diversity	0.30**	0.71	0.54	0.62	0.70	0.65	0.98	0.73	0.79	0.68
	(0.13)	(0.70)	(0.64)	(0.71)	(0.73)	(0.68)	(0.72)	(0.68)	(0.70)	(0.59)
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All
First-stage F-statistic		5.06	5.65	4.82	5.19	4.95	5.93	4.94	5.54	6.52
Adjusted-R^2	0.31	0.26	0.30	0.30	0.26	0.28	0.17	0.28	0.22	0.33
Observations	116	116	116	116	116	116	116	116	116	116

Notes: This table explores the robustness of IV results presented in Table 5 to accounting for continental fixed effects. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table F.5: Intra-Ethnic Diversity and Economic Specialization (Reduced Form)
Robustness to Continental FE

| | Economic Specialization | | | | | | | | |
| | OLS | IV | | | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance to East Africa	-0.32	-0.26	-0.27	-0.31	-0.30	-0.53	-0.33	-0.37	-0.41
	(0.33)	(0.33)	(0.32)	(0.35)	(0.33)	(0.38)	(0.32)	(0.35)	(0.38)
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All
Adjusted-R^2	0.29	0.30	0.33	0.31	0.29	0.30	0.32	0.28	0.40
Observations	116	116	116	116	116	116	116	116	116

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

F.4 Robustness to Alternative Origins of Serial Founder Effect

Table F.6: Intra-Ethnic Diversity and Serial Founder Effect
Robustness to Origin of Serial Founder Effect

	Intra-Ethnic Diversity					
	East Africa		Southern Origin		Most Diverse	
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Industrial Distance to East Africa	-0.85***	-0.82***				
	(0.07)	(0.09)				
Pre-Industrial Distance to South-West Africa			-0.85***	-0.84***		
			(0.07)	(0.09)		
Pre-Industrial Distance to Most Diverse Ethnicity					-0.84***	-0.82***
					(0.07)	(0.09)
Main Controls	No	Yes	No	Yes	No	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
Adjusted-R^2	0.72	0.75	0.72	0.74	0.71	0.74
Observations	144	144	144	144	144	144

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table F.7: Intra-Ethnic Diversity and Economic Specialization (IV)
Robustness to Origin of Serial Founder Effect

	Economic Specialization					
	East Africa		Southern Origin		Most Diverse	
	(1)	(2)	(3)	(4)	(5)	(6)
Intra-Ethnic Diversity	0.51***	0.46***	0.37***	0.37**	0.47***	0.42***
	(0.12)	(0.14)	(0.11)	(0.15)	(0.12)	(0.15)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
First-stage F-statistic	56.99	81.54	59.05	64.32	53.08	70.63
Adjusted-R^2	0.14	0.39	0.15	0.40	0.14	0.40
Observations	116	116	116	116	116	116

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table F.8: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Origin

| | Economic Specialization | | | | | |
| | East Africa | | Southern Origin | | Most Diverse | |
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Intra-Ethnic Diversity	0.44***	0.52***	0.42***	0.52***	0.41***	0.52***
	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)	(0.04)
Main Controls	No	Yes	No	Yes	No	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
Adjusted-R^2	0.19	0.28	0.17	0.28	0.17	0.28
Observations	934	934	934	934	934	934

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

F.5 Robustness to Other Distances

Table F.9: Predicted Intra-Ethnic Diversity and Economic Specialization
Robustness to Distances to Some Capitals

	Economic Specialization			
	(1)	(2)	(3)	(4)
Predicted Intra-Ethnic Diversity	0.52***	0.41***	0.55***	0.35***
	(0.04)	(0.04)	(0.05)	(0.05)
Distance to London		-0.31***		
		(0.04)		
Distance to Tokyo			0.06	
			(0.05)	
Distance to Mexico City				0.20***
				(0.04)
Main Controls	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes
Adjusted-R^2	0.28	0.33	0.28	0.29
Observations	934	934	934	934

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table F.10: Intra-Ethnic Diversity and Distance to Addis-Ababa
Robustness to Distance Measure

	Intra-Ethnic Diversity								
	Full Sample								Specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Great Circle Distance to East Africa	-0.86***	-0.87***	-0.87***	-0.88***	-0.90***	-0.87***	-0.89***	-0.95***	-0.96***
	(0.07)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.11)	(0.10)	(0.11)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.74	0.73	0.73	0.73	0.73	0.73	0.73	0.74	0.71
Observations	144	144	144	144	144	144	144	144	116

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table F.11: Intra-Ethnic Diversity and Distance to Addis-Ababa
Robustness to Distance Measure

	Intra-Ethnic Diversity								
	Full Sample								Specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Geodesic Distance to East Africa	-0.86***	-0.87***	-0.87***	-0.88***	-0.90***	-0.87***	-0.89***	-0.95***	-0.96***
	(0.07)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.11)	(0.10)	(0.11)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.74	0.73	0.73	0.73	0.73	0.73	0.73	0.74	0.71
Observations	144	144	144	144	144	144	144	144	116

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table F.12: Intra-Ethnic Diversity and Distance to Addis-Ababa
Robustness to Distance Measure

	Intra-Ethnic Diversity			
	Full Sample		Specialization	
	(1)	(2)	(3)	(4)
Pre-Industrial Distance to East Africa	-0.51*	-0.51*	-0.58**	-0.58**
	(0.26)	(0.26)	(0.29)	(0.29)
Great Circle Distance to East Africa	-0.38		-0.33	
	(0.28)		(0.30)	
Geodesic Distance to East Africa		-0.38		-0.33
		(0.28)		(0.30)
Main Controls	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes
Semi-partial R^2 Pre-Industrial Distance	0.01	0.01	0.01	0.01
Semi-partial R^2 Other Distance	0.00	0.00	0.00	0.00
Adjusted-R^2	0.75	0.75	0.73	0.73
Observations	144	144	116	116

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table G.1: Intra-Ethnic Diversity and Economic Specialization (Reduced Form)

	Economic Specialization								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance to East Africa	-0.41***	-0.43***	-0.44***	-0.45***	-0.37***	-0.40***	-0.36***	-0.46***	-0.39***
	(0.10)	(0.10)	(0.10)	(0.09)	(0.12)	(0.10)	(0.09)	(0.11)	(0.13)
Malaria Ecology		-0.31***							-0.34***
		(0.12)							(0.11)
Ecological Diversity			0.30***						0.23**
			(0.10)						(0.10)
Agricultural Suitability (avg.)				0.00					0.06
				(0.13)					(0.10)
Agricultural Suitability (std.)				0.28**					0.36**
				(0.13)					(0.14)
Caloric Suitability (Pre-1500 ,avg.)					-0.18				-0.23
					(0.15)				(0.15)
Caloric Suitability (Pre-1500 ,std.)					0.25**				-0.01
					(0.12)				(0.15)
Temperature (Spatial Corr., Avg.)						0.04			-0.01
						(0.10)			(0.07)
Temperature (Volatility, Avg)						-0.64***			-0.21
						(0.20)			(0.20)
Pct. Area within 100kms of Sea							0.02		-0.12
							(0.10)		(0.11)
Coast Length							0.47**		0.59***
							(0.21)		(0.18)
Ruggedness (Avg.)								-0.32	-0.05
								(0.23)	(0.19)
Pre-Industrial Mobility (avg.)								0.93**	1.15**
								(0.42)	(0.49)
Pre-Industrial Mobility (std.)								-0.08	-0.36**
								(0.13)	(0.16)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.17	0.23	0.25	0.21	0.19	0.21	0.22	0.19	0.42
Observations	116	116	116	116	116	116	116	116	116

Notes: This table shows the point estimates of the reduced form economic specialization-distance to East Africa for all the specifications in the reduced sample. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					Economic Specialization						
Pre-Industrial Dist. to East Africa	-0.45***	-0.42***	-0.45***	-0.42***	-0.41***	-0.45***	-0.40***	-0.42***	-0.42***	-0.47***	-0.36***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.10)
Malaria Ecology			-0.10**							-0.07	-0.06
			(0.04)							(0.05)	(0.05)
Ecological Diversity				0.15***						0.10***	0.09***
				(0.03)						(0.03)	(0.03)
Agricultural Suitability (avg.)					-0.05					-0.17***	-0.14***
					(0.04)					(0.04)	(0.04)
Agricultural Suitability (std.)					0.13***					0.09**	0.07
					(0.04)					(0.04)	(0.04)
Caloric Suitability Index (Pre-1500CE)						0.09***				0.15***	0.11***
						(0.03)				(0.03)	(0.03)
Caloric Suitability (Pre-1500 ,std.)						0.08*				0.06	0.07
						(0.05)				(0.05)	(0.05)
Temperature (Spatial Corr., Avg.)							0.02			0.02	0.03
							(0.02)			(0.04)	(0.04)
Temperature (Volatility, Avg)							-0.32***			-0.23***	-0.14*
							(0.06)			(0.07)	(0.08)
Pct. Area within 100kms of Sea								0.02		-0.05	0.01
								(0.03)		(0.04)	(0.04)
Coast Length								0.04		0.04	0.04
								(0.04)		(0.04)	(0.03)
Ruggedness (Avg.)									0.12*	0.05	0.05
									(0.07)	(0.07)	(0.07)
Pre-Industrial Mobility (avg.)									0.27***	0.41***	0.28**
									(0.10)	(0.11)	(0.11)
Pre-Industrial Mobility (std.)									-0.12	-0.21***	-0.14*
									(0.08)	(0.08)	(0.08)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continental FE	No	No	No	No	No	No	No	No	No	No	Yes
Adjusted-R^2	0.20	0.22	0.23	0.24	0.24	0.24	0.24	0.22	0.23	0.29	0.32
Observations	934	934	934	934	934	934	934	934	934	934	934

Notes: This table shows the point estimates of the reduced form economic specialization-distance to East Africa for all the specifications in the extended sample. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table G.3: Distance to East Africa and Economic Specialization

	Economic Specialization									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pre-Industrial Distance to East Africa	-0.48***	-0.37***	-0.33***	-0.35***	-0.36***	-0.36***	-0.40***	-0.39***	-0.36***	-0.36***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.11)	(0.10)
Malaria Ecology			-0.09*							-0.06
			(0.05)							(0.05)
Ecological Diversity				0.13***						0.09***
				(0.03)						(0.03)
Agricultural Suitability (avg.)					-0.05					-0.14***
					(0.04)					(0.04)
Agricultural Suitability (std.)					0.10**					0.07
					(0.04)					(0.04)
Caloric Suitability Index (Pre-1500CE)						0.05*				0.11***
						(0.03)				(0.03)
Caloric Suitability (Pre-1500 ,std.)						0.09**				0.07
						(0.04)				(0.05)
Temperature (Spatial Corr., Avg.)							0.00			0.03
							(0.03)			(0.04)
Temperature (Volatility, Avg)							-0.25***			-0.14*
							(0.06)			(0.08)
Pct. Area within 100kms of Sea								0.06**		0.01
								(0.03)		(0.04)
Coast Length								0.04		0.04
								(0.03)		(0.03)
Ruggedness (Avg.)									0.10	0.05
									(0.06)	(0.07)
Pre-Industrial Mobility (avg.)									0.16	0.28**
									(0.11)	(0.11)
Pre-Industrial Mobility (std.)									-0.06	-0.14*
									(0.08)	(0.08)
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.26	0.28	0.29	0.30	0.29	0.29	0.29	0.29	0.29	0.32
Observations	934	934	934	934	934	934	934	934	934	934

Notes: This table shows the point estimates of the reduced form economic specialization-distance to East Africa for all the specifications in the extended sample when accounting for continental fixed effects. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table G.4: Heterogeneous Effects of Distance to East Africa on Economic Specialization

	Economic Specialization					
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Industrial Distance	-0.07***	-0.05***	-0.06***	-0.02**	-0.05***	-0.05***
to East Africa	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Ecological Diversity		1.55***				
		(0.32)				
Pre-Industrial Distance to East Africa		-0.07***				
× Ecological Diversity		(0.02)				
Agricultural			3.84***			
Suitability (std.)			(0.91)			
Pre-Industrial Distance to East Africa			-0.24***			
× Agricultural Suitability (std.)			(0.07)			
Temperature				-0.01		
(Volatility, Avg)				(0.20)		
Pre-Industrial Distance to East Africa				-0.05***		
× Temperature (Volatility, Avg)				(0.01)		
Ruggedness (Avg.)					0.00***	
					(0.00)	
Pre-Industrial Distance to East Africa					-0.00***	
× Ruggedness (Avg.)					(0.00)	
Pre-Industrial						6.93***
Mobility (std.)						(1.99)
Pre-Industrial Distance to East Africa						-0.44***
× Pre-Industrial Mobility (std.)						(0.14)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.22	0.25	0.25	0.25	0.26	0.24
Observations	934	934	934	934	934	934

Notes: This table shows the point estimates of the reduced form economic specialization-distance to East Africa for all the specifications in the extended sample when accounting for complementarity with heterogeneous environments. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

H Linguistic Diversity and Economic Specialization

This section explores the relation between economic specialization and another proxy of intra-ethnic diversity, as measured by linguistic diversity. In this paper linguistic diversity refers to a language's diversity in terms of number of genders, consonant inventory, and vowel quality inventory (Dryer, 2013) and not to the number of languages in a location, i.e., in captures diversity *within* a population and *not across* populations.

Table H.1: Linguistic Diversity and Distance to East Africa

	Linguistic Diversity								Speciali-zation
	Full Sample								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Linguistic Diversity (Consonant Inventory)									
Pre-Industrial Distance to East Africa	-0.42***	-0.36***	-0.35***	-0.35***	-0.36***	-0.36***	-0.38***	-0.37***	-0.36***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)
Adjusted-R^2	0.17	0.26	0.26	0.26	0.27	0.27	0.26	0.27	0.27
R^2	0.17	0.28	0.28	0.28	0.29	0.29	0.28	0.31	0.32
Observations	299	299	299	299	299	299	299	299	255
Panel B: Linguistic Diversity (Vowel Quality Inventory)									
Pre-Industrial Distance to East Africa	-0.30***	-0.32***	-0.28***	-0.32***	-0.34***	-0.34***	-0.36***	-0.31***	-0.33***
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Adjusted-R^2	0.08	0.15	0.17	0.14	0.15	0.16	0.15	0.19	0.22
R^2	0.09	0.16	0.19	0.17	0.18	0.19	0.18	0.23	0.27
Observations	301	301	301	301	301	301	301	301	256
Panel C: Linguistic Diversity (Number of Genders)									
Pre-Industrial Distance to East Africa	-0.25***	-0.25***	-0.23***	-0.26***	-0.26***	-0.25***	-0.30***	-0.28***	-0.30***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.09)	(0.08)
Adjusted-R^2	0.05	0.16	0.16	0.18	0.18	0.15	0.17	0.20	0.23
R^2	0.06	0.19	0.20	0.22	0.22	0.20	0.22	0.29	0.32
Observations	152	152	152	152	152	152	152	152	131
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Agr. Suit.	CSI	Sea	Mobility	All	All

Notes: This table shows the negative statistically and economically significant relation between measures of linguistic diversity and the distance to East Africa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Each column includes the same set of controls as the same column in Table 3. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.2: Linguistic Diversity and Distance to East Africa

| | Linguistic Diversity (Consonant Inventory) | | | | | | | | |
| | Full Sample | | | | | | | | Specialization |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance to East Africa	-0.42***	-0.36***	-0.35***	-0.35***	-0.36***	-0.36***	-0.38***	-0.37***	-0.36***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)
Malaria Ecology			0.05					0.01	-0.03
			(0.06)					(0.07)	(0.08)
Agricultural Suitability (avg.)				-0.05				-0.05	-0.05
				(0.05)				(0.08)	(0.08)
Agricultural Suitability (std.)				0.04				0.07	0.04
				(0.05)				(0.06)	(0.06)
Caloric Suitability (Pre-1500 ,avg.)					0.02			0.03	0.04
					(0.06)			(0.07)	(0.08)
Caloric Suitability (Pre-1500 ,std.)					-0.12*			-0.13*	-0.11
					(0.06)			(0.07)	(0.07)
Pct. Area within 100kms of Sea						-0.08		-0.09	-0.11
						(0.06)		(0.06)	(0.07)
Coast Length						0.13		0.14*	0.14
						(0.09)		(0.08)	(0.09)
Ruggedness (Avg.)							0.08	0.08	0.06
							(0.12)	(0.12)	(0.13)
Pre-Industrial Mobility (avg.)							0.32	0.31	0.27
							(0.22)	(0.24)	(0.27)
Pre-Industrial Mobility (std.)							-0.18	-0.14	-0.08
							(0.11)	(0.12)	(0.13)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.17	0.26	0.26	0.26	0.27	0.27	0.26	0.27	0.27
R^2	0.17	0.28	0.28	0.28	0.29	0.29	0.28	0.31	0.32
Observations	299	299	299	299	299	299	299	299	255

Notes: This table shows the negative statistically and economically significant relation between linguistic diversity based on consonant inventory and the distance to East Africa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.3: Linguistic Diversity and Distance to East Africa

	Linguistic Diversity (Vowel Quality Inventory)								
	Full Sample								Specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance to East Africa	-0.30***	-0.32***	-0.28***	-0.32***	-0.34***	-0.34***	-0.36***	-0.31***	-0.33***
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Malaria Ecology			0.19***					0.20***	0.22***
			(0.06)					(0.07)	(0.07)
Agricultural Suitability (avg.)				0.01				-0.15*	-0.18**
				(0.06)				(0.08)	(0.08)
Agricultural Suitability (std.)				0.06				0.04	0.04
				(0.06)				(0.06)	(0.07)
Caloric Suitability (Pre-1500 ,avg.)					0.13**			0.19**	0.25***
					(0.07)			(0.08)	(0.08)
Caloric Suitability (Pre-1500 ,std.)					-0.08			-0.04	-0.05
					(0.07)			(0.07)	(0.07)
Pct. Area within 100kms of Sea						-0.18***		-0.09	-0.11
						(0.06)		(0.07)	(0.07)
Coast Length						0.08		0.10	0.12
						(0.10)		(0.10)	(0.11)
Ruggedness (Avg.)							-0.15	-0.20	-0.17
							(0.12)	(0.12)	(0.13)
Pre-Industrial Mobility (avg.)							0.33	0.32	0.24
							(0.26)	(0.26)	(0.27)
Pre-Industrial Mobility (std.)							-0.15	-0.03	-0.02
							(0.12)	(0.12)	(0.13)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.08	0.15	0.17	0.14	0.15	0.16	0.15	0.19	0.22
R^2	0.09	0.16	0.19	0.17	0.18	0.19	0.18	0.23	0.27
Observations	301	301	301	301	301	301	301	301	256

Notes: This table shows the negative statistically and economically significant relation between linguistic diversity based on vowel quality inventory and the distance to East Africa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.4: Linguistic Diversity and Distance to East Africa

	Linguistic Diversity (Number of Genders)								
	Full Sample								Specialization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial Distance to East Africa	-0.25***	-0.25***	-0.23***	-0.26***	-0.26***	-0.25***	-0.30***	-0.28***	-0.30***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.09)	(0.08)
Malaria Ecology			0.12					0.07	0.08
			(0.12)					(0.14)	(0.15)
Agricultural Suitability (avg.)				0.18**				0.14	0.17*
				(0.07)				(0.09)	(0.09)
Agricultural Suitability (std.)				-0.11				-0.15*	-0.12
				(0.07)				(0.09)	(0.09)
Caloric Suitability (Pre-1500 ,avg.)					0.23**			0.21*	0.16
					(0.09)			(0.12)	(0.10)
Caloric Suitability (Pre-1500 ,std.)					-0.18*			-0.16	-0.11
					(0.11)			(0.12)	(0.11)
Pct. Area within 100kms of Sea						-0.08		-0.04	-0.04
						(0.08)		(0.09)	(0.10)
Coast Length						0.14		0.20	0.24
						(0.15)		(0.16)	(0.16)
Ruggedness (Avg.)							0.20	0.32**	0.36***
							(0.15)	(0.13)	(0.13)
Pre-Industrial Mobility (avg.)							-0.07	-0.55	-0.67*
							(0.27)	(0.34)	(0.36)
Pre-Industrial Mobility (std.)							-0.28*	-0.12	-0.12
							(0.16)	(0.20)	(0.19)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.05	0.16	0.16	0.18	0.18	0.15	0.17	0.20	0.23
R^2	0.06	0.19	0.20	0.22	0.22	0.20	0.22	0.29	0.32
Observations	152	152	152	152	152	152	152	152	131

Notes: This table shows the negative statistically and economically significant relation between linguistic diversity based on number of genders and the distance to East Africa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.5: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity

	OLS	IV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Linguistic Diversity (Consonant Inventory)										
Intra-Ethnic Diversity	0.20***	1.15***	1.18***	1.16***	1.16***	1.21***	1.13***	1.14***	1.15***	1.25***
	(0.06)	(0.20)	(0.22)	(0.20)	(0.21)	(0.21)	(0.21)	(0.20)	(0.22)	(0.29)
First-stage F-statistic		45.92	43.47	45.00	43.12	43.96	43.71	44.47	37.43	26.47
Adjusted-R^2	0.06	-0.69	-0.74	-0.64	-0.70	-0.74	-0.65	-0.67	-0.70	-0.76
Observations	255	255	255	255	255	255	255	255	255	255
Panel B: Linguistic Diversity (Vowel Quality Inventory)										
Intra-Ethnic Diversity	0.39***	1.16***	1.38***	1.16***	1.15***	1.15***	1.03***	1.07***	1.06***	1.30***
	(0.06)	(0.25)	(0.33)	(0.24)	(0.25)	(0.23)	(0.21)	(0.21)	(0.22)	(0.34)
First-stage F-statistic		22.75	16.67	24.99	21.67	26.74	27.39	29.64	26.10	18.43
Adjusted-R^2	0.17	-0.41	-0.70	-0.41	-0.40	-0.37	-0.21	-0.26	-0.27	-0.55
Observations	256	256	256	256	256	256	256	256	256	256
Panel C: Linguistic Diversity (Number of Genders)										
Intra-Ethnic Diversity	0.13*	1.22***	1.35***	1.21***	1.26***	1.27***	1.10***	1.19***	0.95***	1.00***
	(0.08)	(0.28)	(0.35)	(0.28)	(0.30)	(0.30)	(0.26)	(0.29)	(0.23)	(0.37)
First-stage F-statistic		29.40	24.00	28.46	28.32	26.76	30.09	28.52	32.10	14.47
Adjusted-R^2	-0.01	-1.09	-1.33	-1.04	-1.08	-1.18	-0.85	-1.03	-0.57	-0.61
Observations	131	131	131	131	131	131	131	131	131	131
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	-	-	Malaria	Eco. Div.	Agr. Suit.	CSI	Volatility	Sea	Mobility	All

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity (based on linguistic traits) and economic specialization, by instrumenting intra-ethnic diversity with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Each column includes the same set of controls as the same column in Table 5. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.6: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity (Consonant Inventories)

	Economic Specialization									
	OLS	IV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intra-Ethnic Diversity	0.20***	1.15***	1.18***	1.16***	1.16***	1.21***	1.13***	1.14***	1.15***	1.25***
	(0.06)	(0.20)	(0.22)	(0.20)	(0.21)	(0.21)	(0.21)	(0.20)	(0.22)	(0.29)
Malaria Ecology			-0.05							-0.00
			(0.10)							(0.11)
Ecological Diversity				0.26***						0.22**
				(0.09)						(0.11)
Agricultural Suitability (avg.)					0.00					-0.18
					(0.10)					(0.13)
Agricultural Suitability (std.)					0.16					0.01
					(0.11)					(0.13)
Caloric Suitability (Pre-1500 ,avg.)						0.05				0.07
						(0.11)				(0.12)
Caloric Suitability (Pre-1500 ,std.)						0.24**				0.24
						(0.12)				(0.16)
Temperature (Spatial Corr., Avg.)							-0.15*			-0.32**
							(0.09)			(0.14)
Temperature (Volatility, Avg)							0.24			0.38
							(0.23)			(0.29)
Pct. Area within 100kms of Sea								-0.03		-0.15
								(0.08)		(0.12)
Coast Length								0.14		0.15
								(0.21)		(0.19)
Ruggedness (Avg.)									-0.14	-0.16
									(0.18)	(0.20)
Pre-Industrial Mobility (avg.)									-0.04	0.40
									(0.31)	(0.36)
Pre-Industrial Mobility (std.)									0.06	-0.15
									(0.17)	(0.20)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic		45.92	43.47	45.00	43.12	43.96	43.71	44.47	37.43	26.47
Adjusted-R^2	0.06	-0.69	-0.74	-0.64	-0.70	-0.74	-0.65	-0.67	-0.70	-0.76
Observations	255	255	255	255	255	255	255	255	255	255

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity (based on consonant inventories) and economic specialization, by instrumenting intra-ethnic diversity with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.7: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity (Vowel Quality Inventories)

| | OLS | IV | | | | | | | | |
| | | | | | | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Economic Specialization					
Intra-Ethnic Diversity	0.39***	1.16***	1.38***	1.16***	1.15***	1.15***	1.03***	1.07***	1.06***	1.30***
	(0.06)	(0.25)	(0.33)	(0.24)	(0.25)	(0.23)	(0.21)	(0.21)	(0.22)	(0.34)
Malaria Ecology			-0.34**							-0.33*
			(0.14)							(0.17)
Ecological Diversity				-0.05						-0.17
				(0.10)						(0.12)
Agricultural Suitability (avg.)					-0.04					0.00
					(0.09)					(0.14)
Agricultural Suitability (std.)					0.08					0.10
					(0.09)					(0.11)
Caloric Suitability (Pre-1500 ,avg.)						-0.10				-0.10
						(0.09)				(0.14)
Caloric Suitability (Pre-1500 ,std.)						0.20**				0.20*
						(0.10)				(0.11)
Temperature (Spatial Corr., Avg.)							-0.06			-0.03
							(0.08)			(0.12)
Temperature (Volatility, Avg)							-0.42***			-0.26
							(0.16)			(0.23)
Pct. Area within 100kms of Sea								0.11		0.03
								(0.09)		(0.15)
Coast Length								0.19**		0.14
								(0.09)		(0.09)
Ruggedness (Avg.)									0.05	-0.07
									(0.15)	(0.20)
Pre-Industrial Mobility (avg.)									-0.03	0.46
									(0.30)	(0.38)
Pre-Industrial Mobility (std.)									0.13	-0.13
									(0.17)	(0.22)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic		22.75	16.67	24.99	21.67	26.74	27.39	29.64	26.10	18.43
Adjusted-R^2	0.17	-0.41	-0.70	-0.41	-0.40	-0.37	-0.21	-0.26	-0.27	-0.55
Observations	256	256	256	256	256	256	256	256	256	256

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity (based on vowel quality inventories) and economic specialization, by instrumenting intra-ethnic diversity with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.8: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity (Number of Genders)

	OLS	IV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intra-Ethnic Diversity	0.13*	1.22***	1.35***	1.21***	1.26***	1.27***	1.10***	1.19***	0.95***	1.00***
	(0.08)	(0.28)	(0.35)	(0.28)	(0.30)	(0.30)	(0.26)	(0.29)	(0.23)	(0.37)
Malaria Ecology			-0.21							-0.04
			(0.23)							(0.23)
Ecological Diversity				0.22*						0.07
				(0.13)						(0.13)
Agricultural Suitability (avg.)					-0.22*					-0.33**
					(0.13)					(0.16)
Agricultural Suitability (std.)					0.33**					0.23
					(0.13)					(0.17)
Caloric Suitability (Pre-1500 ,avg.)						-0.06				0.13
						(0.16)				(0.17)
Caloric Suitability (Pre-1500 ,std.)						0.25				0.15
						(0.16)				(0.18)
Temperature (Spatial Corr., Avg.)							-0.15			-0.12
							(0.12)			(0.16)
Temperature (Volatility, Avg)							-0.26			-0.34
							(0.30)			(0.32)
Pct. Area within 100kms of Sea								0.09		-0.11
								(0.14)		(0.18)
Coast Length								0.14		0.24
								(0.33)		(0.31)
Ruggedness (Avg.)									-0.44**	-0.57***
									(0.20)	(0.22)
Pre-Industrial Mobility (avg.)									0.10	0.78
									(0.37)	(0.64)
Pre-Industrial Mobility (std.)									0.46***	0.19
									(0.17)	(0.23)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic		29.40	24.00	28.46	28.32	26.76	30.09	28.52	32.10	14.47
Adjusted-R^2	-0.01	-1.09	-1.33	-1.04	-1.08	-1.18	-0.85	-1.03	-0.57	-0.61
Observations	131	131	131	131	131	131	131	131	131	131

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity (based on number of genders) and economic specialization, by instrumenting intra-ethnic diversity with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.9: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity

	Economic Specialization (Share)								
	Linguistic Diversity (Consonant Inventory)			Linguistic Diversity (Vowel Quality Inventory)			Linguistic Diversity (Number of Genders)		
	OLS	IV		OLS	IV		OLS	IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intra-Ethnic Diversity	0.22***	1.32***	1.44***	0.40***	1.34***	1.50***	0.14*	1.43***	1.18***
	(0.06)	(0.22)	(0.32)	(0.06)	(0.27)	(0.36)	(0.08)	(0.32)	(0.40)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	No	No	Yes	No	No	Yes
First-stage F-statistic		45.92	26.47		22.75	18.43		29.40	14.47
Adjusted-R^2	0.10	-0.92	-1.04	0.21	-0.63	-0.85	0.01	-1.49	-0.88
Observations	255	255	255	256	256	256	131	131	131

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity and economic specialization (Share), by instrumenting intra-ethnic diversity (as proxied by linguistic characteristics of the language spoken by an ethnicity) with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Each column includes the same set of controls as the same column in Table 5. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.10: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity

	Economic Specialization (Score)								
	Linguistic Diversity (Consonant Inventory)			Linguistic Diversity (Vowel Quality Inventory)			Linguistic Diversity (Number of Genders)		
	OLS	IV		OLS	IV		OLS	IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intra-Ethnic Diversity	0.16***	0.84***	0.96***	0.31***	0.86***	1.01***	0.14*	1.06***	0.77**
	(0.06)	(0.18)	(0.25)	(0.07)	(0.23)	(0.32)	(0.08)	(0.28)	(0.31)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	No	No	Yes	No	No	Yes
First-stage F-statistic		45.92	26.47		22.75	18.43		29.40	14.47
Adjusted-R^2	0.01	-0.37	-0.40	0.08	-0.21	-0.31	-0.03	-0.80	-0.24
Observations	255	255	255	256	256	256	131	131	131

Notes: This table shows the positive statistically and economically significant association between intra-ethnic diversity and economic specialization (Score), by instrumenting intra-ethnic diversity (as proxied by linguistic characteristics of the language spoken by an ethnicity) with the distance to East Africa (see section 3.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and the extended set of confounders from Table 2. Each column includes the same set of controls as the same column in Table 5. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.11: Intra-Ethnic Diversity and Economic Specialization (IV)
Linguistic Diversity
Robustness to accounting for Observation Date

	Economic Specialization								
	Linguistic Diversity (Consonant Inventory)			Linguistic Diversity (Vowel Quality Inventory)			Linguistic Diversity (Number of Genders)		
	OLS	IV		OLS	IV		OLS	IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intra-Ethnic Diversity	0.19***	1.07***	1.17***	0.37***	1.10***	1.26***	0.12	1.19***	0.98***
	(0.05)	(0.19)	(0.28)	(0.06)	(0.24)	(0.34)	(0.08)	(0.28)	(0.36)
Observation Date	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	No	No	Yes	No	No	Yes
First-stage F-statistic		47.39	27.34		21.13	17.52		28.31	13.69
Adjusted-R^2	0.11	-0.54	-0.60	0.21	-0.31	-0.48	0.01	-1.02	-0.58
Observations	254	254	254	255	255	255	131	131	131

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

H.1 Robustness to Alternative Origins of Serial Founder Effect

Table H.12: Intra-Ethnic Diversity and Serial Founder Effect
Robustness to Origin of Serial Founder Effect

| | Intra-Ethnic Diversity | | | | | |
| | East Africa | | Southern Origin | | Most Diverse | |
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Linguistic Diversity (Consonant Inventory)					
Pre-Industrial Distance to East Africa	-0.42***	-0.37***				
	(0.05)	(0.06)				
Pre-Industrial Distance to South-West Africa			-0.39***	-0.39***		
			(0.05)	(0.06)		
Pre-Industrial Distance to Most Diverse Ethnicity					-0.40***	-0.37***
					(0.05)	(0.06)
Main Controls	No	Yes	No	Yes	No	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
Adjusted-R^2	0.17	0.27	0.15	0.28	0.16	0.27
Observations	299	299	299	299	299	299
	Panel B: Linguistic Diversity (Vowel Quality Inventory)					
Pre-Industrial Distance to East Africa	-0.30***	-0.31***				
	(0.06)	(0.07)				
Pre-Industrial Distance to South-West Africa			-0.31***	-0.28***		
			(0.06)	(0.08)		
Pre-Industrial Distance to Most Diverse Ethnicity					-0.28***	-0.27***
					(0.06)	(0.07)
Main Controls	No	Yes	No	Yes	No	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
Adjusted-R^2	0.08	0.19	0.09	0.17	0.08	0.17
Observations	301	301	301	301	301	301
	Panel C: Linguistic Diversity (Number of Genders)					
Pre-Industrial Distance to East Africa	-0.25***	-0.28***				
	(0.06)	(0.09)				
Pre-Industrial Distance to South-West Africa			-0.35***	-0.36***		
			(0.07)	(0.09)		
Pre-Industrial Distance to Most Diverse Ethnicity					-0.28***	-0.29***
					(0.07)	(0.09)
Main Controls	No	Yes	No	Yes	No	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
Adjusted-R^2	0.05	0.20	0.12	0.24	0.07	0.21
Observations	152	152	152	152	152	152

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table H.13: Intra-Ethnic Diversity and Economic Specialization (IV)
Robustness to Origin of Serial Founder Effect

	Economic Specialization					
	East Africa		Southern Origin		Most Diverse	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Linguistic Diversity (Consonant Inventory)						
Intra-Ethnic Diversity	0.84***	0.57**	0.77***	0.47*	0.77***	0.52**
	(0.21)	(0.25)	(0.20)	(0.25)	(0.20)	(0.24)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
First-stage F-statistic	23.48	20.01	23.47	16.78	23.90	19.22
Adjusted-R^2	-0.36	-0.04	-0.28	0.02	-0.29	-0.01
Observations	103	103	103	103	103	103
Panel B: Linguistic Diversity (Vowel Quality Inventory)						
Intra-Ethnic Diversity	1.15***	1.25***	1.04***	1.13***	1.08***	1.18***
	(0.20)	(0.29)	(0.18)	(0.27)	(0.19)	(0.28)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
First-stage F-statistic	45.92	26.47	45.58	25.60	44.37	25.05
Observations	255	255	255	255	255	255
Panel C: Linguistic Diversity (Vowel Quality Inventory)						
Intra-Ethnic Diversity	1.16***	1.30***	1.14***	1.37***	1.20***	1.35***
	(0.25)	(0.34)	(0.26)	(0.41)	(0.28)	(0.39)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
First-stage F-statistic	22.75	18.43	19.04	13.24	18.41	14.64
Observations	256	256	256	256	256	256
Panel D: Linguistic Diversity (Number of Genders)						
Intra-Ethnic Diversity	1.22***	1.00***	0.82***	0.65**	1.06***	0.84***
	(0.28)	(0.37)	(0.20)	(0.26)	(0.25)	(0.32)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes	No	Yes
First-stage F-statistic	29.40	14.47	33.59	19.80	28.15	15.22
Observations	131	131	131	131	131	131

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.