MPRA
Munich Personal RePEc Archive

Within-Group Heterogeneity in a Multi-Ethnic Society

Artiles, Miriam

Pontificia Universidad Católica de Chile

16 April 2022

Online at https://mpra.ub.uni-muenchen.de/112782/
MPRA Paper No. 112782, posted 20 Apr 2022 13:17 UTC

Within-Group Heterogeneity in a Multi-Ethnic Society*

Miriam Artiles†

April 16, 2022

Abstract

Is ethnic diversity good or bad for economic development? Most empirical studies find corrosive effects. In this paper, I show that ethnic diversity need not spell poor development outcomes—a history of within-group heterogeneity can turn ethnic diversity into an advantage for long-run development. I collect new data from a natural experiment regarding Peru's colonial history: the forced resettlement of native populations in the 16th century. This intervention forced together various ethnic groups into new jurisdictions. In those jurisdictions where colonial officials concentrated individuals with a history of within-group heterogeneity, who, prior to colonization, worked in complementary climates of the Andes, ethnic diversity results in systematically lower costs and may even become advantageous. Neither precolonial groups' political complexity nor their degree of economic development explain this result. The transmission of prosocial behavior is one likely channel. I also find evidence consistent with a positive role of economic complementarities between ethnic groups. *JEL* Codes: J15, N16, O10, O12, Q56, Z10.

*I am grateful to Gianmarco León-Ciliotta, Luigi Pascali, Marta Reynal-Querol, and Joachim Voth for their guidance and support throughout this project. I am also grateful to Miguel Ángel Carpio, Emilio Depetris-Chauvin, Martín Fernández-Sánchez, Felipe González, Jenny Guardado, Jonas Hjort, Stelios Michalopoulos, Omer Moav, Ömer Özak, Elias Papaioannou, Paul Schaudt, Felipe Valencia Caicedo, Yanos Zylberberg, and participants of various seminars and workshops for their comments and suggestions. I thank Gshan Irigoin and Diana Tadeo Ruesta for providing excellent research assistance.

†Pontificia Universidad Católica de Chile, Instituto de Economía, e-mail: miriam.artiles@uc.cl

1 Introduction

The effect of ethnic diversity on economic growth and development is a question of long-standing interest in economics. Following the initial work by Easterly and Levine (1997) and Alesina and Glaeser (2004), a large body of literature has examined the costs and benefits of ethnic diversity.[1] Most empirical studies find corrosive effects. When individuals within ethnic groups are homogeneous and groups differ in their preferences for policies or public goods, conflicting preferences can lead to inefficiencies in the provision of public goods or to policy choices that may not benefit the entire society (e.g., Alesina, Baqir and Easterly 1999; Miguel and Gugerty 2005). Inter-group tensions can also result in civil conflicts or exacerbate mistrust and lack of cooperation (e.g., Alesina and La Ferrara 2000; Fearon and Laitin 2003). On the other side, some studies find that if ethnic groups differ in their occupational activities or skills, then complementary specializations can sustain coexistence, facilitate inter-group trade, and generate economic gains (Jha 2013; Becker and Pascali 2019; Jedwab, Johnson and Koyama 2019; Montalvo and Reynal-Querol 2021). While there is a general understanding that diversity brings opportunities and challenges, there is scarce evidence on which factors determine its positive or negative consequences. When is ethnic diversity good for economic development, and when is it bad?

In this paper, I study whether the long-run effect of ethnic diversity on comparative development depends on within-group heterogeneity. Underlying previous literature on the effects of ethnic diversity is the assumption that individuals within ethnic groups tend to be homogeneous. However, ethnicities are not necessarily homogeneous entities. Individuals within ethnic groups may differ along many dimensions, including preferences; economic activities or skills; and cultural, genetic, and linguistic traits (e.g., Horowitz 1998; Desmet, Ortuño-Ortín and Wacziarg 2017; Depetris-Chauvin and Özak 2020). Recent empirical research has emphasized the role of within-group heterogeneity in explaining comparative development. In a pioneering study, Ashraf and Galor (2013) focus on population heterogeneity, as proxied by the degree of genetic diversity across individuals within the same society, and

[1] See Alesina and La Ferrara (2005) for a survey of the initial literature.

show its influence on comparative development in both premodern and modern times.

The main contribution of this paper is to show that exposure to within-group heterogeneity matters for understanding the long-run economic consequences of ethnic diversity. My focus is on skill heterogeneity within ethnic groups. I analyze new data from a natural experiment regarding Peru's colonial history—the forced resettlement of native populations in the 16th century. Unintentionally on the part of the colonizers, this intervention forced together various ethnic groups into small-scale jurisdictions. In those jurisdictions where colonial officials concentrated individuals with a history of within-group heterogeneity, who, prior to colonization, worked in complementary climatic conditions of the Andes, ethnic diversity results in systematically lower costs and may even become advantageous.

Several features of the study setting allow me to examine whether the consequences of ethnic diversity depend on exposure to within-group heterogeneity. First, I consider quasi-random variation in ethnic diversity.[2] Ethnohistoric accounts suggest that geographic proximity to ethnic boundaries created quasi-random variation in the ethnic composition of new jurisdictions (Wachtel 1976; Pease 1989). This happened as a result of a mismatch between the precolonial settlement pattern and the colonial notion of jurisdiction. The second is skill heterogeneity within ethnic groups. Before the Spanish conquest, individuals from the same ethnic group lived at different altitudes in an attempt to maximize the economic base. The anthropologist John Murra writes the following:

> "In a territory so broken up by altitude ..., we should expect wide differences between ecological or production zones ... Access to the productivity of contrasting zones becomes indispensable. This could have been achieved by maintaining a series of markets at different altitudes, run by the ethnic groups inhabiting each separate ecological niche. However, this was not the Andean solution. They opted for the simultaneous access of a given ethnic group to the productivity of many microclimates." (Murra 1995, p. 60-61)

Different disciplines have documented the self-organization of Andean ethnic groups.[3] These

[2]Throughout the paper, I use the term "ethnic group," introduced by Murra (1975) in this context, to refer to the groups that coexisted in the Andean highlands before the Spanish conquest. I refer to the issue of ethnic identity in Section 2.1.

[3]See Brush (1976), Pease (1989), Stanish (1989), and Aldenderfer (1993), among others, for perspectives from human ecology, history, anthropology, and archaeology, respectively.

groups controlled complementary resource zones by coordinating subpopulations with altitude-specific skills. The internal organization of ethnic groups, however, has received little attention in economics—see the recent work by Moscona, Nunn and Robinson (2020). The specialization pattern is in line with Michalopoulos (2012)'s idea that variation in geographic characteristics, such as land quality and elevation, may lead to specialization through the formation of zone-specific skills, applying the idea to specialization *within* ethnic groups.

Were individuals from more heterogeneous ethnic groups better able to function in multi-ethnic societies? The answer is not obvious. Small-scale jurisdictions did not exist before colonization. Recent research shows that the benefits of ethnic diversity tend to flourish at the local level (Montalvo and Reynal-Querol 2021). The literature has also emphasized the positive role of local interactions (Desmet, Gomes and Ortuño-Ortín 2020) and complementarities (Jha 2013; Becker and Pascali 2019) between ethnic groups. If, after being resettled, individuals with a history of within-group heterogeneity were more willing to engage with other ethnicities, mutually beneficial exchange from local inter-ethnic interactions might have become more frequent. Although trust tends to be higher among coethnics, individuals from more heterogeneous ethnic groups were already used to operating in diverse settings; they may have been better able to integrate with other ethnic groups.[4]

The first result of the paper documents the direct effect of ethnic diversity, which I benchmark against previous results in the literature. Guided by the historical evidence, I identify parishes (colonial jurisdictions) that were accidentally created close to spatial boundaries between ethnic groups. The analysis relies on the assumption that Spanish officials were not fully aware of the vertical distribution of coethnic individuals over space. Given the vertical settlement pattern, parishes created close to ethnic boundaries concentrated populations of different ethnic origins (Pease 1989, 1992).[5] I use a subsample of parishes for which surnames from colonial

[4]Individuals from more heterogeneous ethnic groups may exhibit greater "openness to experience," a personality trait defined as a preference for novelty and variety. This trait has been associated with lower levels of prejudice and more favorable attitudes toward out-group members. More open individuals also tend to be less risk averse and more creative when looking for potential solutions. For social psychology studies on openness to experience, see McCrae (1996), McCrae and Costa (1997), Flynn (2005) and Sibley and Duckitt (2008), among others.

[5]I provide empirical support for the historical narrative that colonial officials did not systematically consider ethnic boundaries for the location of parishes. On average, they did not select locations differently (i.e., in a

baptism records are available to validate ethnic diversity. When comparing contemporary living standards between parishes whose initial populations were ethnically diverse and those with an ethnically homogeneous founding population, the results show a robust pattern. On average, parishes with ethnic diversity tend to exhibit lower living standards in the long run, as captured by measures of local economic activity and access to public facilities. This result is in line with the literature on the costs of ethnic diversity, although it also highlights the persistent consequences of forced diversity at the local level.

I then examine whether the average effect of ethnic diversity on living standards differs depending on precolonial exposure to skill heterogeneity within ethnic groups.[6] I am interested in the interaction effect of ethnic diversity and past exposure to within-group heterogeneity across parishes. Specifically, I consider average exposure among the ethnic groups concentrated in each parish. Before presenting the main result, I provide evidence that more heterogeneous ethnic groups did not manipulate the locations of parishes at the time of the policy—parishes without ethnic diversity did not systematically concentrate populations from more heterogeneous ethnic groups. When I examine the interaction effect, the estimates show a negative coefficient on ethnic diversity and a positive coefficient on its interaction with within-group heterogeneity. On average, ethnically diverse parishes whose initial populations were exposed to higher within-group heterogeneity tend to perform better in the long run. I interpret this result as evidence that a history of within-group heterogeneity contributes to mitigating and eventually overcoming the economic cost of ethnic diversity.

The estimated interaction effect persists when controlling for initial prosperity and geography at the parish level. It is also robust to considering the administrative province and ecclesiastical jurisdiction in charge of the given parish. However, apart from belonging to a more complex internal organization, individuals from more heterogeneous ethnic groups were likely exposed to greater economic and institutional development, among many other ethnic traits. I collect data from archaeological sources to explore the correlates of within-group heterogeneity

way that resulted in systematic differences in proximity to ethnic boundaries) depending on the characteristics of native populations or geography.

[6]To proxy for skill heterogeneity, I construct spatial data on the distribution of resource-producing zones within the homelands of precolonial ethnic groups.

before colonization. Estimates controlling for potential correlates and their interactions with ethnic diversity show that the main result is remarkably robust. Overall, robustness checks and supplementary analysis applied to assess measurement error suggest that the estimated interaction effect is plausibly causal.

To understand the evolution of these long-run effects, I use data from the 1876 population census. The results show that the documented pattern of development was accompanied by a structural shift from agriculture toward the tertiary sector and by improved literacy rates. The change in the structure of economic activity was notably strong for women. When examining potential mechanisms, I find evidence consistent with cultural transmission. Historical studies suggest that the self-organization of precolonial groups was sustained through reciprocity and cooperation (Wachtel 1976; Stern 1995). I explore whether the transmission of prosocial behavior facilitated social interactions between ethnic groups and, over the course of generations, contributed to sustaining long-run development. Data on voluntary associations are consistent with this interpretation. Using data from colonial records, I furthermore provide suggestive evidence that a history of within-group heterogeneity favored inter-group unions during the colonial period.

Recent papers have emphasized the role of local interactions in understanding the effects of ethnic diversity.[7] These papers relate to studies on the positive role of inter-group complementarities at the local level. In particular, the theoretical framework developed in Jha (2013, 2018) establishes that peaceful inter-group coexistence can be sustained through the specialization of ethnic groups into complementary activities that are costly to replicate and expropriate.[8] In line with this literature, the estimates show an economic advantage for parishes where the skills of the ethnic minority likely complemented those of ethnic groups

[7] Desmet, Gomes and Ortuño-Ortín (2020) provide cross-country evidence that the costs of ethnic diversity for public good provision tend to be weakened by local inter-group interactions. Montalvo and Reynal-Querol (2021) focus on the size of the unit of analysis, finding that ethnic diversity has a positive effect on economic growth at low levels of geographic aggregation. The authors argue that a potential explanation in the context of Africa is the increase in trade close to spatial ethnic boundaries, which may suggest ethnic specialization into complementary activities (Michalopoulos 2012).

[8] Jha (2013) provides consistent empirical evidence on tolerance toward Muslims in Hindu societies. Other studies have focused on anti-Semitism, finding consistent empirical results; see Becker and Pascali (2019) for evidence in the context of the Protestant Reformation in Germany and Jedwab, Johnson and Koyama (2019) for evidence from the Black Death in Europe.

that were relatively homogeneous. This result holds in a setting where the identity of the ethnic minority varies depending on the geographic location of the parish.

This study contributes to a large body of literature on development and political economy that focuses on the consequences of ethnic diversity. To the best of my knowledge, this is the first empirical paper to explore the long-run effects of ethnic diversity in a setting with variation in exposure to within-group heterogeneity. The initial literature tended to emphasize the costs of ethnic diversity without considering heterogeneity across individuals within ethnic groups. Studies in this literature have been conducted at different levels of analysis. Across countries and US localities, ethnic diversity has been associated with lower levels of economic growth, public good provision, quality of government, and social capital, and with greater political instability and civil conflict.[9] Using micro-level data, Miguel and Gugerty (2005) show that ethnic diversity is associated with lower public good provision in Kenya. More recently, Hjort (2014) has focused on the private sector. The author provides causal evidence for the effect of ethnic diversity on team productivity at a flower plant in Kenya. The results show that teams of ethnically diverse workers are, on average, less productive than homogeneous teams. Evidence on potential mechanisms points toward a taste for discrimination against coworkers of different ethnic origin.

Following Gennaioli and Rainer (2007), Michalopoulos and Papaioannou (2013), and Alesina, Giuliano and Nunn (2013), the socioeconomic characteristics of precolonial societies or ethnic groups have received increasing attention.[10] However, the heterogeneity of individuals within these groups has been less explored. Ashraf and Galor (2013) performed the first empirical study to explore the role of this dimension in explaining comparative development.[11] The authors find that population heterogeneity (from a genetic perspective) has a persistent hump-shaped effect on economic development. Underlying the positive side of this effect is the

[9]See Easterly and Levine (1997), Alesina, Baqir and Easterly (1999), La Porta et al. (1999), Alesina and La Ferrara (2000), Alesina, Glaeser and Sacerdote (2001), Alesina et al. (2003), Fearon and Laitin (2003), Alesina and Glaeser (2004), Montalvo and Reynal-Querol (2005), and Desmet, Weber and Ortuño-Ortín (2009), among many others.

[10]This literature has focused on ethnographic data, predominantly from Murdock (1967)'s *Ethnographic Atlas*, using the terms ethnic group and society interchangeably.

[11]See Esteban and Ray (2011) for a theoretical model of ethnic conflict in which individuals within ethnic groups differ in attitudes and income.

idea that the more heterogeneous a society, the larger the set of potentially different traits among its individuals becomes. If this heterogeneity in traits translates into different specializations, there will be gains from those that are complementary.[12] In this paper, I focus on a genetically homogeneous region—highland Peru—and study complementary altitude-specific skills within ethnic groups.[13] I provide empirical evidence that exposure to within-group heterogeneity contributes to overcoming the detrimental effect of ethnic diversity on economic development, triggering beneficial inter-group engagement at the local level.

The evidence on cultural transmission adds to the literature on the long-run effects of cultural traits (e.g., Nunn and Wantchekon 2011; Voigtländer and Voth 2012; Alesina, Giuliano and Nunn 2013; Guiso, Sapienza and Zingales 2016). The results are consistent with the idea that exposure to within-group heterogeneity favored the formation of a culture of cooperation and more favorable attitudes toward out-group members. Other studies have emphasized the role of climate and geography in shaping culture. For example, Buggle and Durante (2021) find that European regions exposed to higher environmental risk during the premodern era exhibit higher levels of inter-personal trust today. The study argues that, in face of variability in temperature and precipitation, farmers developed cooperative strategies that contributed to the emergence of more trusting attitudes. Nunn and Puga (2012) document the indirect effect that geographic ruggedness had on the development of African countries by allowing protection from slave traders. Separately, a culture of mistrust has been shown to persist among the descendants of individuals affected by the slave trade (Nunn and Wantchekon 2011).

Finally, this paper contributes to a growing body of literature on the long-run consequences of forced displacements throughout history (Becker 2022). The results help to understand the effects of the displacement of indigenous populations as a result of colonization, a research topic with scarce evidence so far; see Valencia Caicedo (2019). In contemporary societies where multiple ethnicities coexist (e.g., due to forced displacements or as a consequence of voluntary migrations in an increasingly globalized world), understanding whether the

[12] See Depetris-Chauvin and Özak (2020) for empirical evidence consistent with the link between population heterogeneity and specialization into different economic activities across premodern societies.

[13] For genetic studies on precolonial populations in the region of analysis, see, for example, Nakatsuka et al. (2020), Valverde et al. (2016), and Kemp, Tung and Summar (2009).

consequences of ethnic diversity depend on exposure to within-group heterogeneity is also important for policy discussions.

The remainder of the paper is organized as follows: Section 2 summarizes the historical background, Section 3 describes the data construction process, Section 4 presents the empirical strategy, Section 5 describes the main results, Section 6 describes potential mechanisms, and Section 7 concludes.

2 Historical Background

2.1 Precolonial Context

Ethnic groups. By the time Spanish conquerors arrived (1532), the Andean civilization comprised several coexisting groups that had been incorporated over the previous century (1438-1525) into the Inca empire, e.g. *Chocorvos*, *Lucanas*, *Soras*, *Chancas*, *Quichuas*, *Caviñas*, *Chilques*, and *Aymaraes*, among others (see Tello 1939; Rowe 1946; Dulanto 2008). The groups in my study region coexisted after the disintegration of the *Wari* culture (ca. 1000) and before being conquered by the Spanish.

The term "ethnic" was introduced by Murra (1975) to refer to these groups. However, the issue of ethnic identity is vaguely discussed in Andean research, primarily because the central Andes has long been perceived as a culturally homogeneous region despite ethnic and linguistic differences being present at the time of Spanish contact.[14] The most common view is that these groups were polities with diverse linguistic dialects (e.g., Rowe 1946; Murra 1975) and differentiated material cultures in both domestic and nondomestic contexts (e.g., Stanish 1989).[15] Group identity seems to have been reinforced by the absence of inter-group marriage. Specifically, the social unit is generally described as an endogamous group of several extended families with descent through the male line (Rowe 1946).

[14] Stanish (2001) points toward the interest of Inca and Spanish powers in promoting cultural unity via state propaganda.

[15] Language tended to be homogenized after 1475 with the spread of Quechua, the language of the Incas. However, early chronicles suggest that linguistic differences coexisted with Quechua for some time; see, for instance, Garcilaso de la Vega (1960)[1609].

Self-organization in a mountain environment. In the human ecology literature, the mountain environment of the Andean highlands is described as a vertical resource system (Brush 1976). Differences in elevation give rise to various microclimates within short distances, and each microclimate is in turn suited to a different assortment of natural resources and crops.

After the pioneering ethnohistoric work of Murra (1975), studies across different disciplines applied his ideas to understand subsistence patterns in precolonial times (e.g., Brush 1976; Pease 1989; Stanish 1989; Aldenderfer 1993; Nash 2009). The settlement pattern of a given ethnic group is described as a vertical archipelago. Specifically, ethnohistoric accounts suggest that subsistence was based on the simultaneous control of different elevation zones. According to Murra's model, the group tried to maximize the economic base by sending individuals to settle vertically arranged resource zones (Murra 1975, 1995, 2002a,b).

Murra's model is often described as a zonal complementarity model (e.g., Stanish 1989; Aldenderfer 1993; Isbell and Silverman 2002b). Since different altitudes provide different resources, they can be interpreted as complements; by having individuals living at different altitudes, the group strengthened access to resources and hence increased total output at the ethnic group level. The idea of specialization is supported by emphasizing that individuals formed permanent settlements and were devoted to obtaining zone-specific resources (Murra 2002a; Pease 1989). Individuals settled at different altitudes can thus be envisioned as subunits or subpopulations from the same ethnic group with zone-specific skills. Ties to the extended family and the rest of the group seem to have been retained to take advantage of complementarities (Pease 1989; Stanish 1989, 2005; Murra 2002a; Nash 2009).[16]

Continuity after the Inca expansion. According to Murra's and subsequent research, self-organization was already in place during the pre-Inca period (Murra 1956, 1975). The Inca expansion (1438–1525) was achieved through the gradual conquest of pre-existing groups. The dominant view is that this led to a dynamic process of state formation whereby differentiated regions or provinces were sequentially created based on ethnic identity (Rowe

[16] This subsistence strategy has been particularly supported for the central and southern Andes. It is important to note, however, that it is unclear how the model applied to coastal societies (e.g., Rostworowski 1977), and, hence, this paper focuses on highland Peru.

1946). Ethnohistoric research suggests that the Inca government was indirect in the sense that each region was governed by the ruler of the corresponding ethnic group (Murra 1975, 2002b). This is a crucial characteristic of Inca rule because it supports the notion that ethnic traits were preserved during this period. At the same time, ethnic rulers were pushed to continue with the control of vertical zones in their respective regions to sustain the empire (Murra 1956, 1975). Archaeologist and anthropologist John H. Rowe (1946) mapped the approximate extent of the groups at the time of the Spanish conquest; see Figure 1.

[Figure 1 about here]

2.2 The Spanish Intervention

The contemporary administrative division of Peru has its origin in the initial colonial period. When Viceroy Francisco de Toledo first disembarked in Peru (1569), native populations followed the Andean pattern, living scattered along mountain slopes. This vertical and dispersed settlement pattern was an obstacle for the Spanish administration. In the words of Spanish official Juan de Matienzo, "the *indios*, for being isolated in *huaycos* and ravines, do not live in right order, and this is the main obstacle to be indoctrinated" (in Medina 1974a, p. 155).

To facilitate tribute collection and religious indoctrination, Toledo ordered the forced reorganization of native populations into residential (*reducciones*) and religious (*doctrinas*) jurisdictions. Between 1570 and 1575, colonial officials arranged the division of populations from all discovered lands in the Viceroyalty of Peru into *reducciones*.[17] In turn, several *reducciones* were under the jurisdiction of a single *doctrina* or *parroquia*, a parish served either by the regular or secular clergy.[18]

Within four decades of the conquest of the Inca empire, the Spanish administration had undertaken a complete reorganization of native populations. It is important to note that the

[17]See the original design in the form of a grid system in Appendix A.

[18]The regular clergy included priests of several religious orders (Santo Domingo, La Merced, San Francisco, San Agustín, and Compañía de Jesús), but secular priests who were not members of any order were also present; see de Armas Medina (1953).

resettlement created a new paradigm for these populations. The new model limited population movement, thus pointing against a key feature of the precolonial economy: the exchange of resources between different elevation zones (Pease 1989). The intention of the resettlement was not to create sustainable jurisdictions but to concentrate dispersed populations in a way more consistent with the Spanish view of the world (i.e., small-scale, continuous, and delimited jurisdictions; Medina 1974a,b, 1993).

Notably, there was tension at the time of the policy between the pre-existing settlement pattern, which was a native response to the mountain environment, and the Spanish notion of jurisdiction, a feature of a more horizontal world. Ethnohistoric accounts suggest that colonial officials did not consider that individuals from such a wide range of elevations could belong to the same ethnic group (Murra 1975).[19] Given the vertical settlement pattern, the new jurisdictions did not always respect pre-existing ethnic divisions (Pease 1989, 1992; Wachtel 1976, 2002).

Historical studies note that, in practice, the limitation of movement was effective at the parish level (Saignes 1991; Medina 1974a,b, 1993). In fact, this system was maintained throughout the entire colonial period, and, at the time of independence from Spain, parishes were called districts, forming the basis for what is currently the third-level administrative division of the country.[20] I hence use the parish as the unit of analysis in empirical exercises.

3 Data Construction

3.1 Sample

This study focuses on the Peruvian territory conquered by the Inca empire that remained in the Viceroyalty of Peru for the entire colonial period (1532-1810). The census prepared from 1791 to 1795 under the administration of Viceroy Gil de Taboada y Lemos lists all parishes created in this territory (see Appendix A for details).

[19] Appendix A provides anecdotal evidence of other circumstances in which 16th-century Spanish observers appeared to be not always familiar with the mountain environment and native practices.

[20] For details on the transition from parishes to districts, see *Guía Política, Eclesiástica y Militar del Virreynato del Perú, para el Año de 1793* and *Calendario y Guía de Forasteros para el Año de 1834*.

I assign geographic coordinates to each parish. In particular, I start by matching each parish to a modern district using the district's name and year of creation. I then assign coordinates to each parish capital using a map from the Peruvian Ministerio del Ambiente (MINAM) that provides the name and coordinates of all existing population centers within each district. In most cases, the old parish capital remains the district capital. For districts where this is not the case (i.e., where the capital was changed after independence from Spanish rule), I assign coordinates corresponding to the parish capital.[21]

Following the historical literature on Murra's model, I focus on parishes located in highland Peru (i.e., more than 500 meters above sea level). I exclude the two capital parishes of Cuzco and Arequipa,[22] as well as six parishes that now form part of Chile. The final sample consists of 336 parishes; see Panel (a) of Figure 1.

3.2 Measuring Ethnic Diversity

3.2.1 A Measure of Ethnic Diversity

A measure of ethnic diversity would ideally be based on administrative data detailing the ethnicity of relocated individuals. Unfortunately, such colonial data do not exist.[23] Guided by the historical literature, I consider geographic proximity to ethnic boundaries. I start by defining $Ethnic\ div_p$ as a dummy variable indicating that parish p was created close to an ethnic boundary. Specifically, $Ethnic\ div_p$ takes value 1 for the presence of an ethnic boundary within a 10-km buffer from the parish capital, and 0 otherwise (see Figure A.1).[24] The spatial

[21]I check for priests in charge of religious indoctrination in each parish using a combination of historical sources; see Lissón Chávez (1943), de Armas Medina (1953), de Córdoba Salinas (1957)[1651] and García (1997). In many cases, it is possible to track the names of the priests and whether they were part of the regular or secular clergy.

[22]Based on new data on ancient DNA for the region of analysis, Cuzco is also an exception to the trend of genetic homogeneity during the Inca period, which evidences population mobility in the administrative center of the Inca empire (Nakatsuka et al. 2020).

[23]The colonial administration did not register the ethnicity or race of individuals beyond the words "indio" or "tributario" in most of the region.

[24]According to Paz Soldán (1877), 10 km corresponds to approximately 3 *leguas* (a colonial measure of distance) in the 16th century. Colonial descriptions commonly lie between 2 and 3 *leguas*; see, for example, Jiménez de la Espada (1881). When the distance between the capitals of two parishes is less than 10 km, I use equidistant boundaries to ensure that the buffers do not overlap. Finally, an ethnic group is counted as part of the buffer only if its homeland occupies at least 1 percent of the buffer's area. This exercise ensures that the ethnic

exercise aims to capture whether colonial officials accidentally concentrated populations from different ethnic origins within the same parish.

For this exercise, I rely on Rowe (1946)'s mapping of the approximate extent of precolonial ethnic groups. Panel (a) of Figure 1 shows the spatial distribution of parishes: those with an ethnic border within the 10-km buffer are displayed in yellow (35 percent of parishes in the sample), while those located further inside ethnic homelands are displayed in blue. In robustness checks, I consider a measure of ethnic fractionalization based on the Herfindahl index ($Ethnic\ frac_p = 1 - \sum_e w_{pe}^2$). However, since w_{pe} represents the area share of ethnic group e within the buffer of parish p rather than the exact population share, this measure may arguably be more affected by measurement error than $Ethnic\ div_p$.

In the next section, I examine the extent to which $Ethnic\ div_p$ captures differences in ethnicity at the parish level. For this validation exercise, I use a subsample of parishes with information on surnames from colonial baptism records. In particular, I explore whether surname heterogeneity among individuals with native surnames was significantly higher in parishes created close to ethnic boundaries than in parishes created in the interior of ethnic homelands.

3.2.2 Validating Ethnic Diversity

Isonymy methods. In certain contexts, measures based on the frequency distribution of surnames can shed light on the biological relationships between human populations. Provided that surnames are inherited, the underlying premise of this approach is that surname commonality between individuals (isonymy) can be used to trace common ancestry (Lasker 1980, 1985; Colantonio et al. 2003). Two main diversity indices have been applied to surnames:

$$D = 1 - \sum_{k=1}^{K} p_k^2, \qquad S = -\sum_{k=1}^{K} p_k \ln(p_k)$$

where p_k represents the proportion of individuals with surname k in the population and K is the total number of different surnames. The first index, $D \in [0, 1]$, is a standard measure

group has at least one grid cell of 1 km × 1 km inside the buffer.

of diversity based on the Simpson or Herfindahl index. The second index, $S \in [0, \ln(K)]$, takes its theoretical basis from information theory (Shannon 1948).[25] As long as any two individuals with the same surname inherited the surname from a common ancestor, S can be interpreted as the average uncertainty in predicting ancestry: if each surname has the same relative frequency in the population (surnames are evenly distributed across individuals), the uncertainty in predicting the most probable ancestor of a randomly selected individual will be high; in contrast, a more uneven distribution in which a few surnames are shared by a large portion of the population (e.g., an isolated community characterized by endogamous marriages) implies less uncertainty in predicting ancestry.

Isonymy methods make a strong assumption (i.e., that surname commonality directly translates into common ancestry).[26] Are these methods appropriate for this application? Two contextual features are worth noting. The first is precolonial endogamy with ancestry traced through the male line (Rowe 1946); the second is that no system of family names existed prior to the Spanish conquest, but first names related to mythical ancestors did. The Catholic Church introduced the Hispanic system of family names for the purpose of religious indoctrination. While the potential adoption of Spanish surnames over time represents a limitation, qualitative evidence suggests that the common practice during the early colonial period was for priests to choose a Spanish first name, with the mythical first names of the individual's parents adopted as surnames (see Appendix B and Carpio and Guerrero (2021) for further details).

I focus on the early common origin of native surnames representing common ancestry through the male line. Using baptism records from the colonial period (1605–1780), I created a dataset of 112,340 individuals with native paternal surnames. Each baptism record, accessed via FamilySearch.org (Genealogical Society of Utah), includes the full name of the individual, name of the parish, and date of baptism. The dataset provides information for 65 parishes, of which 20 percent have an ethnic border within the 10-km buffer. To identify native surnames,

[25]The Shannon index has also been applied to measure genetic diversity (Lewontin 1972) and species diversity (Magurran 2004).

[26]This assumption does not hold in contexts where one surname has multiple origins (e.g., unrelated individuals with common surnames due to their ancestors sharing the same occupation) or in contexts where surname changes are permitted for nongenetic reasons (e.g., illegitimacy or adoption).

I constructed a dictionary of indigenous linguistic roots and looked for the occurrence of these roots within surnames; see Appendix B.

Empirical results. Since it can be reasonably assumed that not all historical records have been preserved, the results should be interpreted with caution.[27] Table 1 presents OLS estimates from regressing surname diversity measures on $Ethnic\ div_p$. In each column, the dependent variable is either the S index or the D index, constructed using individuals with native paternal surnames.

Panel A shows the baseline results. For each surname diversity index, the first column shows the unconditional correlation; the second column controls for the log number of individuals found in the records of the parish and the share of individuals with nonnative surnames; the third column accounts for potential differences in the mean and standard deviation of elevation, mean and standard deviation of land caloric suitability during the precolonial period (Galor and Özak 2016a), longitude, latitude, and log distance to perennial rivers; the last column includes ecclesiastical jurisdiction fixed effects, accounting for potential differences in the administration of baptism across five colonial bishoprics (Lima, Arequipa, Huamanga, Trujillo, and Cuzco). Panel B shows the results obtained after dropping individuals whose surnames occur only once in the dataset.[28] In Panel C, I show the results obtained from using groups of similar surnames (instead of raw surnames) to compute surname diversity indices. This approach takes into account potential changes in the writing of surnames over time.[29]

The results for the subsample of parishes with information suggest that, on average, parishes located close to ethnic boundaries exhibit higher levels of surname diversity among individuals with native surnames (between 0.41 and 0.56 standard deviations) than do parishes located in the interior of ethnic homelands. Henceforth, I refer to $Ethnic\ div_p$ as the ethnic diversity dummy.

[27]The number of parishes with information varies by year. The mean parish comprises 1,726 individuals with native paternal surnames, of which 857 are men, relative to a sample mean of 1,627 individuals per parish according to the census of 1791–1795 (of which 769 are men). See Appendix B.

[28]This causes the sample size to decrease from 112,340 to 106,124 individuals.

[29]Specifically, I group surnames if deletion, insertion, or substitution of only one character is required to transform one surname into another (i.e., the surnames have a Levenshtein distance equal to one).

[Table 1 about here]

3.3 Measuring Within-Group Heterogeneity

3.3.1 A Measure of Within-Group Heterogeneity

According to ethnohistoric evidence, individuals that were settled in different resource zones during the precolonial period developed zone-specific skills. However, comprehensive data on the number of individuals settled in each zone prior to the Spanish conquest do not exist. In this paper, I construct spatial data on the distribution of resource-producing zones to compute a proxy for skill heterogeneity within ethnic groups.

For this task, I build on the work of Pulgar Vidal (1941), a well-known Peruvian geographer who distinguished five natural resource zones in my study region: Yunga (500–2,300 m], Quechua (2,300–3,500 m], Suni or Jalca (3,500–4,000 m], Puna (4,000–4,800 m], and Janca (4,800–6,768 m], where figures in parentheses refer to elevation in meters above sea level. His work combines local knowledge of geography with native folklore, offering a well-established account of the mountain environment in Peru. Each zone was known for specific crops. For example, the natural limit of maize cultivation was the Quechua zone; grains such as quinua and cañihua and lupins such as tarhui proliferated in the zone known as Suni or Jalca; and several varieties of potato, a tuber that can provide more carbohydrates per hectare than maize at high altitudes, grew exceptionally well in the Puna (Sandweiss and Richardson 2008; Burger and Merwe 1990).

I map the spatial distribution of the zones using elevation data from version 1.2 of the Harmonized World Soil Database (FAO). Specifically, I assign each grid cell in the FAO data (approximately 1 km $\times$ 1 km at the equator) to a particular zone based on median elevation. The resulting map is shown in Panel (b) of Figure 1. Spanish officials established capitals of colonial parishes in all of the zones (approximately 23.53, 34.57, 44.23, and 28.57 percent of capitals established in the Yunga, Quechua, Suni, and Puna zones, respectively, had an ethnic border within the 10-km buffer; see Table A.1).[30]

[30] I exclude the Janca zone from the analysis (3.43 percent of the total territory) because it cannot be

Using this map, I compute a measure of heterogeneity in resource zones at the ethnic group level (47 groups). I start by considering the reciprocal of the Simpson or Herfindahl index, a common measure of diversity in ecological studies (Magurran 2004):[31]

$$H_e = \frac{1}{\sum_j s_{ej}^2}$$

where s_{ej} is the area share of zone j within the homeland of ethnic group e. The index increases as resource zones become more diverse. I normalize the index to 1 for the group with the highest value to facilitate the interpretation. Figure A.2 shows the density of the normalized index at the ethnic group level. Approximately 23 percent of the groups have an index value of below 0.5, while the index for the remaining 77 percent ranges from 0.5 to 1, with similar mean and median values (0.661 and 0.682, respectively).

3.3.2 Validating Resource-Producing Zones

This section explores the extent to which resource zones are meaningful. Specifically, I explore whether the measure of heterogeneity in resource zones (H) does indeed explain crop variety in the data. For this exercise, one would ideally use land suitability data for all crops available before 1500. In the absence of these data, I use information on native crops from modern sources. I rely on the 2012 agricultural census, which provides geo-referenced data for an extensive set of native crops.[32] I explore the determinants of crop variety across grid cells of different sizes, as well as across ethnic groups, based on the number of native crops reported by farmers at the time of the census.

I start by computing H at the grid-cell level for 25 km × 25 km grid cells covering the entire study region. In Column 1 of Table 2, I regress the log of the number of crops on this measure. The estimated beta coefficient is positive (0.523) and statistically significant.

permanently inhabited due to oxygen constraints (e.g., Sandweiss and Richardson 2008). Pasture for camelids was the primary resource of this zone.

[31] This index has also been used in urban studies to measure diversity in sectors of economic activity; see, for instance, Duranton and Puga (2000). In robustness exercises, I consider $\widetilde{H}_e = 1 - \sum_j s_{ej}^2$.

[32] I follow the classification of native crops in Tapia (2013), who identifies 41 main native crops in the region. The census includes information on 38 crops.

The coefficient remains stable in magnitude and statistical significance when I control for an alternative potential predictor of crop variety—variation in elevation (Column 3).[33] The same pattern arises when including fixed effects that account for differences across hydrographic basins (Column 4),[34] log area (Column 5), and mean elevation (Column 6). Column 7 shows the same pattern when using the number of crops without being logarithmically transformed as an outcome variable.

According to unconditional OLS estimates, a one standard deviation increase in H is associated with a 0.523 standard deviation increase in log crop variety (Column 1). In the case of variation in elevation, the associated standard deviation increase in log crop variety is 0.398 (Column 2). However, this coefficient becomes small (0.007) and statistically insignificant after I control for H (Column 3). In Column 8, I substitute H with dummy variables indicating the number of resource zones within the grid cell. The magnitudes of the estimated coefficients increase with the number of zones (relative to those for grid cells with only one resource zone). Similar results are found across 50 km $\times$ 50 km grid cells (Table A.2) and ethnic groups (Table A.3).

[Table 2 about here]

3.3.3 Correlates of Within-Group Heterogeneity

In Table 3, I explore the precolonial correlates of within-group heterogeneity at the ethnic group level (H_e). Columns 1-3 show that mean elevation, land caloric suitability, and river density are not significantly correlated with within-group heterogeneity. Interestingly, there is no statistically significant correlation between within-group heterogeneity and the size of the ethnic group, as measured by land area (Column 4) and approximate population (Column 5) before colonization.[35] Note that, under autarky, group size coincides with market size, which could create incentives for specialization and innovation (Smith 1776). I do find that within-group heterogeneity is positively correlated with approximate population density

[33] I follow Michalopoulos (2012) in using the standard deviation of raw elevation as a measure of variation. The results are robust to considering measures of terrain ruggedness (available upon request).

[34] The hydrographic system in the study region is composed of 62 basins.

[35] Population is available for 46 groups (see Appendix C for data sources).

(Column 6), which could reflect economic prosperity during the precolonial period (Ashraf and Galor 2011, 2013; Maloney and Valencia Caicedo 2016).

Columns 7-9 explore precolonial socioeconomic and institutional characteristics. In the absence of systematized ethnographic data, I collect information from archaeological sources (see Appendix C).[36] In line with Column 6, the data suggest that within-group heterogeneity is positively correlated with urbanization, as measured by a dummy for the presence of towns and urban centers within the ethnic group's homeland (Column 7). Column 8 shows evidence consistent with the idea that incentives for internal exchange may lead to political centralization (Fenske 2014). In particular, I create a dummy for any material indicator that could evince political complexity (i.e., administrative centers and monumental architecture—public buildings and communal spaces, including temples, palaces, and complex mound platforms; Stanish 2001) and find a positive correlation with within-group heterogeneity. Column 9 shows no correlation with the presence of elite residences, nonetheless.[37] Overall, the results highlight the importance of examining robustness to the inclusion of these correlates in equation 2.

[Table 3 about here]

4 Empirical Strategy

4.1 The Effect of Ethnic Diversity

I first explore whether the ethnic composition of colonial parishes influenced comparative economic development in the long run:

$$y_p = \beta_0 + \beta_1 Ethnic\ div_p + X_p^{'}\gamma + \upsilon_p \tag{1}$$

[36] Although archaeological data are subject to geographic coverage issues, the Peruvian Andes are relatively well-studied archaeologically (Isbell and Silverman 2002a, 2008). Furthermore, site-level archaeological data are receiving increasing attention in economics (Matranga and Pascali 2021).

[37] In Table A.4, I explore the role of technology. Unsurprisingly, within-group heterogeneity is positively correlated with the presence of terraces. I find no correlation with food storage structures, however. This may be due to the fact that some crops, such as potatoes, were sometimes simply spread on the ground or placed underground to freeze, leaving no archaeological record.

where y_p is a contemporary development outcome for parish p, $Ethnic\ div_p$ is a dummy variable indicating whether the parish was created close to an ethnic boundary, and X_p is a vector of parish-level control variables measured at baseline.[38] The historical evidence suggests that geographic proximity to ethnic boundaries created quasi-random variation in ethnic diversity across parishes. The analysis relies on the assumption that colonial officials were not fully aware of the vertical distribution of coethnic individuals over space (Murra 1975). Thus, they were unlikely to have systematically considered ethnic boundaries when deciding where to establish new parishes. Given the vertical settlement pattern, parishes created close to ethnic boundaries concentrated populations of different ethnic origins (Pease 1989, 1992).

In the next section, I explore whether there is empirical support for this historical narrative in the sample. Spanish officials could have followed recommendations for the location of parishes or avoided locations where they suspected it would be easier for native populations to escape (e.g., plains or lower elevations). I explore whether factors that could influence the location of parishes and affect post-resettlement economic development varied significantly with proximity to ethnic boundaries.

4.1.1 Balance Tests for Ethnic Diversity

The colonial regulation (1569–1570) described three desirable location characteristics (see, for example, Jiménez de la Espada 1881). The first was land quality and abundance. Enough land was needed to be worked by native families following their own rules of crop rotation. Land plots were thought to be the primary means for paying tribute.[39] The second characteristic was access to water. Proximity to surface water, which in this context meant access to the system of river basins, was a key advantage for the irrigation of land and the possibility of sustaining populations that depended mainly on subsistence agriculture. Finally, to facilitate religious indoctrination, the locations would ideally be far from *huacas*, sacred native shrines

[38] Throughout the paper, I report heteroskedasticity-robust standard errors, standard errors adjusted for spatial autocorrelation, and standard errors clustered at the level of the colonial province in specifications with province fixed effects.

[39] Colonial tribute took the form of a personal tax paid by all native men aged 18 to 50.

that generally honored nature. Local officials were also tasked with destroying the houses where native families used to live prior to resettlement. Shortly after the creation of new jurisdictions, families refusing to relocate were to be punished and forced to move.

The extent to which Spanish officials applied the recommendations is unclear (Pease 1989). Nonetheless, the results in Table 4 show that, on average, there are no statistically significant differences between ethnically diverse and nondiverse parishes in baseline factors. I start by exploring the mean and standard deviation of terrain characteristics, such as elevation and land caloric suitability during the precolonial period. There are no significant differences in these characteristics, which alleviates concerns regarding the possibility that colonial officials selected locations differently (i.e., in a way that resulted in systematic differences in proximity to ethnic boundaries) at different elevations or in plains as opposed to more rugged terrain.

Log distance to perennial rivers is also balanced between parishes with and without ethnic diversity. I collected data on the locations of precolonial shrines to explore the third recommendation. On average, ethnically diverse and nondiverse parishes do not differ significantly in log distance to native shrines. Finally, I consider log distance to *mita* mines (Dell 2010)[40] and local prosperity at the time of the policy, as proxied by the value of expected tribute.[41] These characteristics are also balanced.

[Table 4 about here]

Table A.5 provides evidence on statistical balance for additional characteristics. To proxy for the threat of native attack at the time of the policy, I geo-referenced data on precolonial defensive sites (e.g., fortresses, walled sites, and *pukaras*).[42] The table shows statistical balance between ethnically diverse and nondiverse parishes in log distance to defensive sites. It also shows statistical balance in log distance to precolonial administrative centers and in log distance to the Inca road network. Colonial officials could also have been interested in specific crops, such as maize or potatoes (Brush 1976). The table shows no statistically significant differences in average land suitability for these crops during the precolonial period.

[40]The legal requirement to send native populations to mines subjected to the *mita* started in 1573.

[41]The expected tribute was based on colonial officials' assessment of the number of individuals present at the time of the policy (Cook 1982; Puente Brunke 1991).

[42]See, for example, Arkush and Tung (2013) for further information on pre-colonial Andean defenses.

4.2 The Effect of Ethnic Diversity and Within-Group Heterogeneity

I am interested in exploring whether the average effect of ethnic diversity on economic development differed depending on past exposure to skill heterogeneity within ethnic groups ($\beta_3 \neq 0$):

$$y_p = \beta_0 + \beta_1 Ethnic\ div_p + \beta_2 \overline{H}_p + \beta_3 \left(Ethnic\ div_p \times \overline{H}_p \right) + X_p' \gamma + \epsilon_p \tag{2}$$

where $\overline{H}_p$ is a measure of average exposure to within-group heterogeneity before colonization. In particular, I consider the average exposure among the ethnic groups concentrated in each parish: $\overline{H}_p = \sum_e w_{pe} H_e$, where w_{pe} is the area share of ethnic group e within the buffer of parish p (i.e., a weight accounting for the relative size of the ethnic group in the parish) and H_e is the ethnic-level measure of within-group heterogeneity prior to resettlement.

To introduce the interaction term in equation 2, ethnic diversity should not be determined by $\overline{H}_p$. Considering the quasi-random nature of the variation in ethnic diversity, $Ethnic\ div_p$ should be orthogonal to $\overline{H}_p$. However, one may be concerned about the possibility that, for example, more heterogeneous ethnic groups negotiated locations in the interior of ethnic homelands, in which case parishes without ethnic diversity would systematically concentrate populations from more heterogeneous ethnic groups. Before presenting the main results, I document that $Ethnic\ div_p$ is uncorrelated with $\overline{H}_p$ in the data. I also provide supporting evidence that more heterogeneous ethnic groups did not manipulate locations at the time of resettlement.

In robustness checks, I furthermore address the concern that β_3 captures any differential effect of ethnic diversity due to correlates of within-group heterogeneity that may have been relevant for post-resettlement economic development (i.e., omitted variable bias). The correlations in Section 3.3.3 suggest that, apart from belonging to a more complex internal organization, individuals from more heterogeneous ethnic groups belonged to more developed polities, thus being exposed to more complex economic and political institutions. I extend equation 2 to control for the average of each ethnic-level correlate ($\overline{G}_p = \sum_e w_{pe} G_e$) and its interaction with ethnic diversity.

4.2.1 Within-Group Heterogeneity and Colonial Locations

I examine whether there is evidence that more heterogeneous ethnic groups tried to manipulate colonial locations, thus inducing changes in the ethnic composition of parishes. I start by aggregating the number of parishes at the ethnic group level. Columns 1 and 2 of Table A.6 show that within-group heterogeneity (H_e) is not significantly correlated with the total number of parishes in which the ethnic group was concentrated, nor is it correlated with the proportion of parishes located close to spatial boundaries between ethnic groups, on average. Column 3 shows that the average distance from a parish to the closest ethnic border is also uncorrelated with within-group heterogeneity. This evidence suggests that ethnic groups with more heterogeneous subpopulations did not influence distance to ethnic borders at the time of the policy.

Consistently, in Table A.7, I document that ethnic diversity is uncorrelated with average exposure to within-group heterogeneity ($\overline{H}_p$) across parishes. The fact that this correlation is not statistically significant suggests that, on average, parishes without ethnic diversity did not systematically concentrate populations from more heterogeneous ethnic groups.

5 Results

5.1 Main Results

I start by comparing contemporary living standards between parishes whose initial populations were ethnically diverse and those with an ethnically homogeneous founding population (equation 1). To capture overall living standards, I explore different measures of local economic activity and access to public facilities. First, I follow the empirical literature in using luminosity data from satellite images at night to proxy for local economic activity (Michalopoulos and Papaioannou 2013, 2018). The second proxy I consider is nonsubsistence agriculture. Subsistence farming has traditionally been a widespread practice in the Andean highlands (Mayer 2002). The Peruvian agricultural census asks farmers whether they dedicate most of their harvest to self-consumption or, instead, to sale or trade in local markets. Finally, the population census provides data on access to public sanitation and to the public water network.[43] The previous outcomes are measured in different years of 1990–2000 (see Appendix C and Table A.8 for the data sources and summary statistics, respectively).

Table 5 presents the first statistical examination of the data. To analyze the overall effect of ethnic diversity on the previous outcomes, I follow the methodology in Kling et al. (2004) and Clingingsmith, Khwaja and Kremer (2009). Specifically, I report the standardized average effect size (AES), which accounts for the covariance across underlying individual effects, jointly with heteroskedasticity-robust standard errors. On average, colonial parishes built on ethnically diverse populations tend to exhibit lower living standards in the long run.[44] The estimated unconditional effect of ethnic diversity is negative (-0.20 standard deviations) and statistically significant. Neither parish-level baseline characteristics nor ecclesiastical jurisdiction fixed effects fully explain this result (Columns 2 and 3, respectively).[45]

Column 4 shows that precolonial exposure to within-group heterogeneity ($\overline{H}_p$) is positively correlated with contemporary living standards. This correlation is stronger among parishes with ethnic diversity (Column 5) than among parishes where only one ethnic group was concentrated (Column 6). All regressions include parish-level baseline characteristics and ecclesiastical jurisdiction fixed effects. In Column 7, I present the results obtained from estimating the interaction effect of ethnic diversity and average exposure to within-group heterogeneity for the whole sample (equation 2). The coefficient on ethnic diversity is negative, but its interaction with $\overline{H}_p$ is positive. This pattern persists when I include fixed effects that account for the colonial administrative province rather than the ecclesiastical jurisdiction (Column 8).[46]

[Table 5 about here]

Table 6 explores this interaction effect in more detail. Panel A shows the results for 1990–2000, while Panel B explores living standards for 2010–2020 (based on different waves

[43]Rural access to water and sanitation in Peru was among the lowest in Latin America during 1990–2000. See the 2000 report from the WHO and UNICEF Joint Monitoring Programme for Water Supply and Sanitation.

[44]Figure A.3 plots the different outcome variables as a function of the number of ethnic groups.

[45]The vector of baseline controls includes all variables in Table 4, as well as longitude and latitude.

[46]The ecclesiastical jurisdiction varies at the province level (44 provinces).

of the same censuses). All columns except the first, presented for reference, include parish-level baseline characteristics and colonial province fixed effects. I report robust standard errors clustered at the province level in brackets.[47] The results show a similar pattern in the two decades. On average, ethnic diversity is strongly associated with lower living standards. However, the positive coefficient on the interaction term suggests that parishes whose initial populations were exposed to higher within-group heterogeneity perform relatively better in the long run. Separating the effects on local economic activity and access to public facilities shows that much of the overall effect is driven by differences in economic activity (Columns 2, 4, and 6).[48] Columns 3, 5, and 7 show similar results when controlling for log population density and a rural dummy variable in each decade (Michalopoulos and Papaioannou 2013, 2014).

[Table 6 about here]

Figure 2 plots the AES of ethnic diversity and 95 percent confidence intervals. To understand the magnitude, consider that $\overline{H}_p$ ranges from 0.301 to 1 with a median of 0.675. For example, the estimates for overall living standards for 2010–2020 imply that the negative effect of ethnic diversity decreases from -0.881 to -0.023 as $\overline{H}_p$ reaches the median. The effect becomes positive and statistically significant for $\overline{H}_p$ above 0.830 (corresponding to approximately 16.7 percent of parishes in the sample), which is consistent with the raw data (Figure A.4). I interpret this result as evidence that a history of within-group heterogeneity contributes to mitigating (and eventually overcoming) the economic cost of ethnic diversity. The following section addresses potential omitted variable bias and measurement error.

[Figure 2 about here]

5.2 Supplementary Analyses

This section uses Column 2 of Table 6 (estimates for 2010-2020 living standards after inclusion of baseline controls and colonial province fixed effects) as the baseline specification for all robustness and sensitivity checks.

[47] Table A.9 reports OLS estimates separately for each outcome variable and standard errors adjusted for spatial autocorrelation (Colella et al. 2019), as well as the R-squared.

[48] Note, however, that nighttime luminosity data may also capture public lighting (Hodler and Raschky 2014).

5.2.1 Precolonial Characteristics of Ethnic Groups

In Table 7, I explore whether the main result persists when I control for potential correlates of within-group heterogeneity and their interactions with ethnic diversity. Individuals from more heterogeneous ethnic groups were likely exposed to stronger economic and institutional development (see Section 3.3.3). I consider all precolonial ethnic-level characteristics analyzed in Table 3 and compute the average of each characteristic among the ethnic groups concentrated in the parish ($\overline{G}_p$). The first column shows the baseline specification for reference. The main result is remarkably robust, alleviating concerns that relevant socioeconomic or institutional characteristics of ethnic groups could be driving the entire result.

[Table 7 about here]

5.2.2 Precolonial Land Occupation and Placebos for H

The estimates in Tables 5 and 6 are likely affected by nonclassical measurement error. A potential source of error is the underlying assumption that individuals were uniformly distributed over space during the precolonial period (i.e., H_e is based on the area share of each resource-producing zone rather than the exact population share). In Table 8, I assess how estimates change when alternative scenarios are considered. In the absence of historical data on the spatial distribution of the population, I follow recent archaeological studies in using precolonial site records as evidence of land occupation (Morrison et al. 2021). Columns 1 and 2 show estimates of equation 2 obtained after restricting the total land area in the region of analysis to distances of 20 km and 10 km around precolonial archaeological sites, respectively.[49] Point estimates are slightly higher, although the estimated effect of ethnic diversity echoes the baseline results ($-0.616 + 0.879 \times \overline{H}_p$).

The empirical analysis also relies on the mapping of historical ethnic boundaries (Figure 1). Given the historical nature of the map, it is important to assess how precise the delineated

[49]This corresponds to a reduction in the total land area under analysis of 7.35 and 31.33 percent, respectively. For this exercise, I combine an official database of precolonial archaeological sites (*Catastro de Monumentos Arqueológicos Prehispánicos*, Ministerio de Cultura, Perú) with my own survey of published archaeological studies based on inventories and handbooks for the region of analysis (see Ravines Sánchez 1985, Ramos Giraldo 2001, and Isbell and Silverman 2002a, 2008).

boundaries are. In Column 3 of Table 8, I follow the approach in Alesina, Michalopoulos and Papaioannou (2016). In particular, I perform the analysis using artificial ethnic boundaries (from Thiessen polygons) rather than historical ones.[50] Compared to baseline results, the estimates are smaller and not statistically significant, which suggests that artificial boundaries are no more precise than historical ones. In Column 4, I instead use the approximate boundaries of the first administrative demarcations created after the Spanish conquest (*corregimientos*). Close correspondence between the spatial boundaries of precolonial ethnic groups and *corregimientos* would suggest that the Spanish administration created the latter based on prior knowledge of the spatial distribution of the groups. The fact that the point estimates are not statistically different from zero suggests that this was not the case.

[Table 8 about here]

5.2.3 Sensitivity Analyses and Additional Robustness Checks

In the appendix, I report estimates from using alternative definitions to measure the key explanatory variables. Table A.10 shows the results obtained after using $Ethnic\ frac_p = 1 - \sum_e w_{pe}^2$ to measure ethnic diversity. Table A.11 employs index $\widetilde{H}_e = 1 - \sum_j s_{ej}^2$ to compute average within-group heterogeneity. If anything, the positive coefficient on the interaction term is higher in magnitude. Figure A.5 alleviates concerns about influential observations. The figure displays point estimates and confidence intervals from baseline regressions that exclude one parish at a time.

Finally, it is important to note that any potential effect of ethnic diversity after resettlement is conditional on the survival of ethnic groups. The decline in native populations after European contact has been well documented by historical studies (Cook 1982; Denevan 1992). To the extent that all groups were similarly affected by disease and abuse, the estimates should be interpreted as the effect of ethnic diversity among the descendants of survivors.[51] Table A.12 shows that the results are robust to the inclusion of parish-level variables that aim to capture the

[50]Thiessen polygons are created using the centroids of historical ethnic homelands as input.

[51]The results are robust to controlling for the distance of each ethnic group to the closest outbreak of smallpox—an infectious disease caused by the variola virus—known to have occurred before resettlement and for the distance to the Inca road system, which may have increased transmission risk (available upon request).

structure of the population by the late colonial period. Specifically, Columns 2 and 3 control for the log of the "indigenous" population and the percentage of the "mestizo" population, defined according to the 1791–1795 census. The results are also robust to controlling for the log number of priests per capita and to the inclusion of religious order fixed effects (Columns 4 and 5, respectively).[52]

5.3 Medium-Term Outcomes

This section provides evidence that the documented pattern of development was accompanied by a shift in the structure of economic activity from agriculture toward the tertiary sector and by improved literacy rates. The 1876 population census provides data on occupations and literacy for 286 parishes. I classify the different occupations in my sample by sector of economic activity and then compute the share of the population employed in each sector. Most parishes continued to be predominantly agricultural by the late 19th century, with 70 percent of the population employed in the primary sector, on average. This population was predominantly illiterate: on average, only 10 percent of a parish's population could read or write, with considerable gender differences (5 percent of women and 16 percent of men).

Table 9 reports the results obtained from estimating equation 2 using the literacy rate and the share of employment in the secondary and tertiary sectors as outcome variables. The results are displayed for the entire population and by gender. All regressions include baseline controls and colonial province fixed effects. The results for literacy and tertiary-sector employment are consistent with the main result. Exposure to within-group heterogeneity attenuates the detrimental effect of ethnic diversity on literacy rates. From the distribution of occupations, compared to parishes built on a single ethnic group, those built on ethnically diverse populations tend to be more oriented toward tertiary-sector activities the higher the average level of exposure becomes. The estimates suggest that the change in the structure of economic activity happened to the detriment of agriculture and was notably strong for women.

[Table 9 about here]

[52]Only one parish in the sample was administered by the Jesuit order, which has been shown to positively influence long-run human capital and economic development (Valencia Caicedo 2019).

6 Potential Mechanisms

6.1 Technology Adoption in Agriculture

Structural transformation may have been accompanied by technology adoption in agriculture (Valencia Caicedo 2019). I explore this potential mechanism using data from the 1994 agricultural census, which asked farmers about insecticides, improved seeds, fertilizers, and their knowledge of biological control. Table 10 reports OLS estimates for equation 2 using dummies for the corresponding technology as outcomes (i.e., whether the share of farmers employing the technology is above the median). All regressions control for the log of the total number of farmers.

Overall, technology adoption does not seem to be the main mechanism involved. For insecticides, biological control, and improved seeds (Columns 1-3), the signs of the coefficients on $Ethnic\ div_p$ and its interaction with $\overline{H}_p$ are consistent with the main result. However, the estimates are only slightly significant for improved seeds. In Columns 4 and 5, I compare the use of chemical fertilizer with a traditional alternative—organic fertilizer, used by 64.5 percent of farmers, on average. The estimates are not statistically significant, although, consistently, the coefficients of interest have the opposite sign in the case of organic fertilizer.

[Table 10 about here]

6.2 Cultural Transmission

6.2.1 Evidence on Contemporary Associations

In Table 11, I explore cultural transmission. Historical studies suggest that precolonial exchange between coethnics with zone-specific skills was sustained through reciprocity and cooperation (see Stern 1995, p. 76, Murra 1975, p. 27-28, and Wachtel 1976, p. 96-97). I analyze data on different types of voluntary associations to explore the potential transmission of prosocial behavior.

In Column 1, the outcome variable is a dummy for the presence of neighborhood associations

(2002 registry of municipalities). In Column 2, the outcome is a dummy indicating whether the share of farmers participating in agricultural associations is above the median (1994 agricultural census). I estimate the baseline specification by OLS, controlling for the log of the total population and the log of the total number of farmers in Columns 1 and 2, respectively. The estimates are in line with the main result, suggesting that the transmission of a culture of cooperation, as one dimension of prosocial behavior, is a plausible channel. In Columns 3-6, I analyze individual-level data from yearly waves of a Peruvian household survey (ENAHO). All regressions include individual controls (gender, age, age squared, years of schooling, civil status, and mother tongue) and year fixed effects, with standard errors clustered at the parish level. The results show similar estimates for participation in neighborhood associations, professional associations, labor unions, and sports clubs.[53]

[Table 11 about here]

While previous studies find that ethnic diversity tends to be associated with lower levels of social engagement (e.g., Alesina and La Ferrara 2000), the results suggest that it may depend on exposure to within-group heterogeneity. One possibility is that precolonial internal relations established to reach a common goal—maximize the economic base—contributed to the formation of cooperative behavior. When individuals from different ethnic groups were forced to reside together and pay tribute on a collective basis, the transmission of prosocial behavior may have facilitated social interactions between ethnic groups.[54]

6.2.2 Evidence on Inter-Group Unions (1605-1780)

Inter-group contact during the colonial period is a necessary step for the previous result. Did exposure to within-group heterogeneity favor inter-group contact? This section provides suggestive evidence. Marriage between different ethnic groups is often used as a proxy for cultural assimilation and integration (Gordon 1964). In turn, it has been linked to inter-group

[53]Unfortunately, the ENAHO survey does not include questions on generalized or inter-group trust, and Latinobarometer surveys do not cover most of the sample.

[54]Although the amount of tribute to be paid was assigned individually, responsibility for its payment fell collectively on the families of native men (Wachtel 1976; Sánchez-Albornoz 1978).

contact in the context of voluntary migration (Bazzi et al. 2019). In the Peruvian case, colonial marriage records from digital genealogical sources are limited in quantity and geographic coverage. However, comparing the first and second surnames of each individual in the sample of baptism records (1605–1780) offers the possibility of exploring inter-group unions.

Each individual inherits two surnames in the Hispanic system of family names. The first surname corresponds to the paternal surname of the father, while the second corresponds the paternal surname of the mother. The sample includes 17,411 individuals with native first and second surnames distributed in 41 parishes, of which 10 have ethnic diversity. A second challenge is the fact that I do not observe ethnicity in the data but only surnames. I use a measure of dissimilarity between the two surnames of each individual to detect potential inter-group unions. In particular, I use the Levenshtein distance (L), which is defined as the minimum number of spelling changes required to transform one surname into another. I divide L by the length of the longest surname to interpret the measure as the percentage of dissimilarity between surnames.[55]

Figure 3 presents a graphical summary of the data for the sample of unions between individuals with native paternal surnames. I compute the percentage of dissimilarity for each union and then obtain the average dissimilarity at the parish level. The left graph of Panel (a) suggests a positive correlation with exposure to within-group heterogeneity in the subsample of parishes with ethnic diversity. The right graph shows a similar pattern when using the share of unions with dissimilarity levels of above 50 percent on the y-axis. The two graphs account for the log of the total number of individuals found in the records of the parish. Despite the small sample size, I run OLS regressions and find that the previous correlations are statistically significant and also hold at the union level (Table A.13). Panel (b) replicates the exercise for the subsample of parishes without ethnic diversity (see also Columns 3 and 11 of Table A.13). Reassuringly, the results suggest that no correlation exists in this subsample.

[Figure 3 about here]

[55] See Dickens (2021) for an application of this measure used to compute distance between languages.

6.3 Cultural Transmission or Economic Complementarities?

Economic complementarities between the ethnic minority and majority groups could also have sustained beneficial coexistence (Jha 2013, 2018; Becker and Pascali 2019). In my setting, economic complementarities are a potential channel in parishes where the ethnic minority belonged to a highly heterogeneous group but, conversely, the majority was relatively homogeneous (i.e., the skills of the minority will likely complement those of a desperate majority). In parishes where both the minority and majority belonged to highly heterogeneous ethnic groups, cultural transmission may be a more plausible channel (i.e., both will already be accustomed to interpersonal interactions with heterogeneous individuals).[56]

I define the minority group as the one with the lowest share of area in the 10-km buffer. Note that the identity of the ethnic minority is not the same in all parishes, as it depends on the geographical location of the parish.[57] Table 12 presents the results obtained from estimating regressions of the following form for the subsample of parishes with ethnic diversity:

$$y_p = \delta_0 + \delta_1 High\ min_p + \delta_2 High\ maj_p + \delta_3 \left(High\ min_p \times High\ maj_p \right) + X_p^{'} \gamma + \varepsilon_p$$

where y_p refers to contemporary living standards, $High\ min_p$ is a dummy variable indicating whether the minority belonged to a relatively heterogeneous ethnic group (H_e above the 50th percentile in Column 1 and above the 75th percentile in Column 2), and $High\ maj_p$ is an analogous dummy variable for the majority group.

To the limits of the natural experiment, the results suggest that economic complementarities are a plausible channel—parishes in which only the ethnic minority was highly heterogeneous tend to exhibit a long-run economic advantage ($\hat{\delta}_1 > 0$).[58] In Column 3, I explore whether, in line with this interpretation, marketplaces for the exchange of local goods tend to be located in these parishes. In particular, I use a dummy for the presence of local retail markets as the outcome variable. Consistently, the estimated coefficient on $High\ min_p$ is positive

[56]Unfortunately, historical data detailing the skills or elevations of origin of relocated individuals do not exist.

[57]There are 44 different ethnic minorities in the sample according to this definition.

[58]When using the 75th percentile cutoff, they represent 10 percent of the parishes with ethnic diversity compared to the 15 percent in which both the ethnic minority and majority groups were highly heterogeneous.

and statistically significant. This result is not explained by total population (Column 4) or agglomeration (Column 5).

[Table 12 about here]

7 Conclusion

Ethnic groups may not be homogeneous entities. Recent research shows that a deeper understanding of within-group heterogeneity can help shed light on the features that shape comparative economic growth and development (Ashraf and Galor 2013). Yet, the empirical literature on the consequences of ethnic diversity has tended to overlook the role of heterogeneity across individuals within the same ethnic group.

This paper shows that exposure to within-group heterogeneity matters for understanding the long-run effect of ethnic diversity on comparative economic development. The study setting allows exploring the role of within-group heterogeneity in complementary traits. Using a natural experiment from Peru's colonial history, I find that in colonial jurisdictions where the Spanish administration concentrated individuals with a history of complementary specializations within ethnic groups, ethnic diversity results in systematically lower costs and may even become advantageous.

More research is needed on the origins of within-group heterogeneity. In particular, the past and current patterns of ethnic specialization are not yet fully understood in economics (Michalopoulos 2012). The results suggest that research aimed at studying variations in patterns of economic specialization within ethnic groups is important for understanding the process of economic development.

References

Academia Mayor de la Lengua Quechua, *Diccionario Quechua – Español – Quechua, Qheswa – Español – Qheswa: Simi Taqe*, 2nd ed., Cusco, Perú: Gobierno Regional Cusco, 2005.

Aldenderfer, Mark S, *Domestic Architecture, Ethnicity, and Complementarity in the South-Central Andes*, University of Iowa Press, 1993.

Alesina, Alberto and Edward Glaeser, *Fighting Poverty in the US and Europe: A world of Difference*, Oxford University Press, 2004.

— **and Eliana La Ferrara**, "Participation in Heterogeneous Communities," *The Quarterly Journal of Economics*, 2000, *115* (3), 847–904.

— **and** — , "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

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

— **, Edward Glaeser, and Bruce Sacerdote**, "Why Doesn't the US Have a European-Style Welfare System?," 2001. Working Paper 8524, National Bureau of Economic Research.

— **, Paola Giuliano, and Nathan Nunn**, "On the Origins of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, 2013, *128* (2), 469–530.

— **, Reza Baqir, and William Easterly**, "Public Goods and Ethnic Divisions," *The Quarterly Journal of Economics*, 1999, *114* (4), 1243–1284.

— **, Stelios Michalopoulos, and Elias Papaioannou**, "Ethnic Inequality," *Journal of Political Economy*, 2016, *124* (2), 428–488.

Arkush, Elizabeth and Tiffiny A Tung, "Patterns of War in the Andes from the Archaic to the Late Horizon: Insights from Settlement Patterns and Cranial Trauma," *Journal of Archaeological Research*, 2013, *21* (4), 307–369.

Ashraf, Quamrul and Oded Galor, "Dynamics and Stagnation in the Malthusian Epoch," *American Economic Review*, 2011, *101* (5), 2003–41.

_ and _ , "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 2013, *103* (1), 1–46.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong, "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *American Economic Review*, 2019, *109* (11), 3978–4025.

Becker, Sascha O, "Forced Displacement in History: Some Recent Research," *Australian Economic History Review*, 2022, *62* (1), 2–25.

_ and Luigi Pascali, "Religion, Division of Labor, and Conflict: Anti-Semitism in Germany over 600 Years," *American Economic Review*, 2019, *109* (5), 1764–1804.

Bertonio, Ludovico, *Vocabulario de la Lengua Aymara*, La Paz, Bolivia: Instituto de Lenguas y Literaturas Andinas–Amazónicas (ILLA–A), 2011. [1612].

Brush, Stephen B, "Man's Use of an Andean Ecosystem," *Human Ecology*, 1976, *4* (2), 147–166.

Buggle, Johannes C and Ruben Durante, "Climate Risk, Cooperation and the Co-Evolution of Culture and Institutions," *The Economic Journal*, 2021, *131* (637), 1947–1987.

Burger, Richard L and Nikolaas J Van der Merwe, "Maize and the Origin of Highland Chavín Civilization: An Isotopic Perspective," *American Anthropologist*, 1990, *92* (1), 85–95.

Caicedo, Felipe Valencia, "The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America," *The Quarterly Journal of Economics*, 2019, *134* (1), 507–556.

Carpio, Miguel Angel and María Eugenia Guerrero, "Did the Colonial mita Cause a Population Collapse? What Current Surnames Reveal in Peru," *The Journal of Economic History*, 2021, *81* (4), 1015–1051.

Carraffa, Alberto García and Arturo García Carraffa, *Diccionario Heráldico y Genealógico de Apellidos Españoles y Americanos*, Madrid: Imprenta Antonio Marzo, 1920–1963.

Chávez, Emilio Lissón, *La Iglesia de España en el Perú. Colección de Documentos para la Historia de la Iglesia en el Perú.*, Vol. 3, 4, Sevilla: Editorial Católica Española, 1943.

Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer, "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering," *The Quarterly Journal of Economics*, 2009, *124* (3), 1133–1170.

Colantonio, Sonia E, Gabriel W Lasker, Bernice A Kaplan, and Vicente Fuster, "Use of Surname Models in Human Population Biology: A Review of Recent Developments," *Human Biology*, 2003, *75* (6), 785–807.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig, "Inference with Arbitrary Clustering," 2019. IZA Discussion Paper n. 12584.

CONADI, *Diccionario Trilingüe: Aymara-Castellano-Inglés*, Chile: Corporación Nacional de Desarrollo Indígena (CONADI), Dirección Regional Arica y Parinacota, 2011.

Cook, Noble David, "Population Data for Indian Peru: Sixteenth and Seventeenth Centuries," *The Hispanic American Historical Review*, 1982, *62* (1), 73–120.

_ , *La Catástrofe Demográfica Andina: Perú 1520-1620*, Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú, 2010.

de Armas Medina, Fernando, *Cristianización del Perú (1532-1600)*, Sevilla: Escuela de Estudios Hispano-Americanos, 1953.

de Córdoba Salinas, Diego, *Crónica Franciscana de las Provincias del Perú*, Washington, DC: Academy of American Franciscan History, 1957.

de la Espada, Marcos Jiménez, *Relaciones Geográficas de Indias. Perú*, Vol. 1, Madrid: Ministerio de Fomento, 1881.

de la Puente Brunke, José, "Un Documento de Interés en Torno al Tributo Indígena en el Siglo XVI," *Historica*, 1991, *15* (2), 265–313.

de la Vega, Garcilaso, *Obras Completas del Inca Garcilaso de la Vega*, Vol. 133, Ediciones Atlas, 1960. [1609].

de Matienzo, Juan, *Gobierno del Perú. Obra escrita en el siglo XVI por el licenciado Don Juan Matienzo, Oidor de la Real Audiencia de Charcas*, Facultad de Filosofía y Letras de Buenos Aires, Sección de Historia, 1910. [1567].

Dell, Melissa, "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 2010, *78* (6), 1863–1903.

Denevan, William M, *The Native Population of the Americas in 1492*, University of Wisconsin Press, 1992.

Depetris-Chauvin, Emilio and Ömer Özak, "The Origins of the Division of Labor in Pre-Modern Times," *Journal of Economic Growth*, 2020, *25* (3), 297–340.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg, "Culture, Ethnicity, and Diversity," *American Economic Review*, 2017, *107* (9), 2479–2513.

_ **, Joseph Flavian Gomes, and Ignacio Ortuño-Ortín**, "The Geography of Linguistic Diversity and the Provision of Public Goods," *Journal of Development Economics*, 2020, *143*, 102–384.

___ , **Shlomo Weber, and Ignacio Ortuño-Ortín**, "Linguistic Diversity and Redistribution," *Journal of the European Economic Association*, 2009, *7* (6), 1291–1318.

Dickens, Andrew, "Understanding Ethnolinguistic Differences: The Roles of Geography and Trade," *The Economic Journal*, 2021. Forthcoming.

Dulanto, Jalh, "Between Horizons: Diverse Configurations of Society and Power in the Late pre-Hispanic Central Andes," in "The Handbook of South American Archaeology," Springer, 2008, pp. 761–782.

Duranton, Gilles and Diego Puga, "Diversity and Specialisation in Cities: Why, Where and When Does it Matter?," *Urban Studies*, 2000, *37* (3), 533–555.

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

Esteban, Joan and Debraj Ray, "A Model of Ethnic Conflict," *Journal of the European Economic Association*, 2011, *9* (3), 496–521.

Fearon, James D and David D Laitin, "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 2003, *97* (1), 75–90.

Fenske, James, "Ecology, Trade, and States in Pre-Colonial Africa," *Journal of the European Economic Association*, 2014, *12* (3), 612–640.

Flynn, Francis J, "Having an Open Mind: The Impact of Openness to Experience on Interracial Attitudes and Impression Formation," *Journal of Personality and Social Psychology*, 2005, *88* (5), 816–826.

Galor, Oded and Ömer Özak, "The Agricultural Origins of Time Preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

___ **and** ___ , "The Agricultural Origins of Time Preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

García, Juan Carlos, "Oposiciones a Parroquias y Doctrinas. El Catálogo de la Sección Concursos, Archivo Arzobispal de Lima: Siglos XVII-XIX," *Revista Andina*, 1997, *15* (2), 421–491.

Gennaioli, Nicola and Ilia Rainer, "The Modern Impact of Precolonial Centralization in Africa," *Journal of Economic Growth*, 2007, *12* (3), 185–234.

Giraldo, Jesús Ramos, *Contribución para un Primer Inventario General de Sitios Arqueológicos del Perú*, Lima: Instituto Nacional de Cultura, 2001.

Gootenberg, Paul, "Population and Ethnicity in Early Republican Peru: Some Revisions," *Latin American Research Review*, 1991, *26* (3), 109–157.

Gordon, Milton Myron, *Assimilation in American Life: The Role of Race, Religion, and National Origins*, New York: Oxford University Press, 1964.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Long-Term Persistence," *Journal of the European Economic Association*, 2016, *14* (6), 1401–1436.

Hjort, Jonas, "Ethnic Divisions and Production in Firms," *The Quarterly Journal of Economics*, 2014, *129* (4), 1899–1946.

Hodler, Roland and Paul A Raschky, "Regional Favoritism," *The Quarterly Journal of Economics*, 2014, *129* (2), 995–1033.

Holguín, Diego González, *Vocabvlario de la Lengva General de todo el Perv llamada Lengua Qquichua, o del Inca*, Lima: Universidad Nacional Mayor de San Marcos, 1952. [1608].

Horowitz, Donald L, "Structure and Strategy in Ethnic Conflict," in "Annual World Bank Conference on Development Economics," Washington, DC: World Bank, 1998, pp. 345–370.

Isbell, William H and Helaine Silverman, *Andean Archaeology I*, Springer, 2002.

_ and _ , "Theorizing Variations in Andean Sociopolitical Organization," in "Andean Archaeology I," Springer, 2002, pp. 3–11.

— **and** — , *The Handbook of South American Archaeology*, Springer, 2008.

Jedwab, Remi, Noel D Johnson, and Mark Koyama, "Negative Shocks and Mass Persecutions: Evidence from the Black Death," *Journal of Economic Growth*, 2019, *24* (4), 345–395.

Jha, Saumitra, "Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia," *American Political Science Review*, 2013, *107* (4), 806–832.

— , "Trading for Peace," *Economic Policy*, 2018, *33* (95), 485–526.

Kemp, Brian M, Tiffiny A Tung, and Marshall L Summar, "Genetic Continuity after the Collapse of the Wari Empire: Mitochondrial DNA Profiles from Wari and post-Wari Populations in the Ancient Andes," *American Journal of Physical Anthropology*, 2009, *140* (1), 80–91.

Kling, Jeffrey R, Jeffrey B Liebman, Lawrence F Katz, and Lisa Sanbonmatsu, "Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment," 2004. KSG Working Paper No. RWP04-035.

Lasker, G. W., *Surnames and Genetic Structure*, Cambridge University Press, 1985.

Lasker, Gabriel W, "Surnames in the Study of Human Biology," *American Anthropologist*, 1980, *82* (3), 525–538.

Lewontin, R. C., "The Apportionment of Human Diversity," *Evolutionary Biology*, 1972, *6*, 381–398.

Lucca, Manuel De, *Diccionario Aymara-Castellano, Castellano-Aymara*, La Paz, Bolivia: Comisión de Alfabetización y Literatura en Aymara (CALA), 1983.

Lynch, John, *Administración Colonial Española, 1782-1810: El Sistema de Intendencias en el Virreinato del Río de la Plata*, Buenos Aires: Editorial Universitaria de Buenos Aires, 1962.

Magurran, Anne E, *Measuring Biological Diversity*, Oxford, U.K.: Blackwell Science Ltd, 2004.

Maloney, William F and Felipe Valencia Caicedo, "The Persistence of (Subnational) Fortune," *The Economic Journal*, 2016, *126* (598), 2363–2401.

Matranga, Andrea and Luigi Pascali, "The Use of Archaeological Data in Economics," in "The Handbook of Historical Economics," Elsevier, 2021, pp. 125–145.

Mayer, Enrique, *The Articulated Peasant: Household Economies in the Andes*, Westview Press, 2002.

McCrae, Robert R, "Social Consequences of Experiential Openness," *Psychological Bulletin*, 1996, *120* (3), 323–337.

_ **and Paul T Costa**, "Conceptions and Correlates of Openness to Experience," *Handbook of Personality Psychology*, 1997, pp. 825–847.

Medina, Alejandro Málaga, "Las Reducciones en el Perú (1532-1600)," *Historia y Cultura*, 1974, *8*, 141–172. Museo Nacional de Arqueología e Historia del Perú.

_ , "Las Reducciones en el Perú durante el Gobierno del Virrey Francisco de Toledo," *Anuario de Estudios Americanos*, 1974, *31*, 819–842. Escuela de Estudios Hispanoamericanos de Sevilla, Consejo Superior de Investigaciones Científicas.

_ , "Las Reducciones Toledanas en el Perú," in "Pueblos de Indios: Otro Urbanismo en la Región Andina," Quito: Ediciones ABYA-YALA, 1993, pp. 263–316.

Michalopoulos, Stelios, "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 2012, *102* (4), 1508–39.

_ **and Elias Papaioannou**, "Pre-Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 2013, *81* (1), 113–152.

— and — , "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, 2014, *129* (1), 151–213.

— and — , "Spatial Patterns of Development: A Meso Approach," *Annual Review of Economics*, 2018, *10*, 383–410.

Miguel, Edward and Mary Kay Gugerty, "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, 2005, *89* (11-12), 2325–2368.

Montalvo, Jose G and Marta Reynal-Querol, "Ethnic Diversity and Economic Development," *Journal of Development Economics*, 2005, *76* (2), 293–323.

— and — , "Ethnic Diversity and Growth: Revisiting the Evidence," *Review of Economics and Statistics*, 2021.

Morrison, Kathleen D, Emily Hammer, Oliver Boles, Marco Madella, Nicola Whitehouse, Marie-Jose Gaillard, Jennifer Bates, Marc Vander Linden, Stefania Merlo, Alice Yao et al., "Mapping Past Human Land Use Using Archaeological Data: A New Classification for Global Land Use Dynthesis and Data Harmonization," *PloS one*, 2021, *16* (4), e0246662.

Moscona, Jacob, Nathan Nunn, and James A Robinson, "Segmentary Lineage Organization and Conflict in Sub-Saharan Africa," *Econometrica*, 2020, *88* (5), 1999–2036.

Murdock, George P, *Ethnographic Atlas*, University of Pittsburgh Press, 1967.

Murra, J. V., "The Economic Organization of the Inca State," Technical Report, The University of Chicago 1956.

— , *Formaciones Económicas y Políticas del Mundo Andino*, Lima: Instituto de Estudios Peruanos, 1975.

— , "Did Tribute and Markets Prevail in the Andes before the European Invasion?," in "Ethnicity, Markets, and Migration in the Andes: At the Crossroads of History and Anthropology," Duke University Press, 1995, pp. 57–72.

—, *El Mundo Andino: Población, Medio Ambiente y Economía*, Vol. 24, Lima: Instituto de Estudios Peruanos-Fondo Editorial de la Pontificia Universidad Católica del Perú, 2002.

—, "Las Sociedades Andinas antes de 1532," in "América Latina en la Época Colonial," Vol. 1, Barcelona: Crítica, 2002, pp. 56–82.

Nakatsuka, Nathan, Iosif Lazaridis, Chiara Barbieri, Pontus Skoglund, Nadin Rohland, Swapan Mallick, Cosimo Posth, Kelly Harkins-Kinkaid, Matthew Ferry, Éadaoin Harney et al., "A Paleogenomic Reconstruction of the Deep Population History of the Andes," *Cell*, 2020, *181* (5), 1131–1145.

Nash, D. J., "Household Archaeology in the Andes," *Journal of Archaeological Research*, 2009, *17* (3), 205–261.

Nunn, Nathan and Diego Puga, "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

— **and Leonard Wantchekon**, "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 2011, *101* (7), 3221–52.

Pease, Franklin, *Del Tawantinsuyu a la Historia del Perú*, Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú, 1989.

—, *Perú: Hombre e Historia. Entre el siglo XVI y el XVIII*, Lima: Edubanco, Fundación del Banco Continental para el Fomento de la Educación y la Cultura, 1992.

Platt, Lyman D, *Hispanic Surnames and Family History*, Baltimore, MD: Genealogical Publishing Company, 1996.

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

Religiosos Franciscanos Misioneros de los Colegios de Propaganda Fide del Perú, *Vocabulario Políglota Incaico: Comprende más de 12,000 Voces Castellanas y 100,000 de Keshua* *del Cuzco, Ayacucho, Junín, Ancash y Aymará*, Lima, Perú: Ministerio de Educación, 1998. [1905].

RENIEC, *Introducción a un Tesoro de Nombres Quechuas en Apurímac*, Lima, Perú: Registro Nacional de Identificación y Estado Civil (RENIEC), jointly with Terra Nuova (`www.terranuova.org`) and Apurimac ONLUS (`www.apurimac.it`), 2012.

Rostworowski, María, *Etnia y Sociedad. Costa Peruana Prehispánica*, Lima: Instituto de Estudios Peruanos, 1977.

Rowe, John Howland, "Inca Culture at the Time of the Spanish Conquest," in "Handbook of South American Indians," Vol. 2 1946, pp. 183–330.

Saignes, Thierry, "Lobos y Ovejas. Formación y Desarrollo de los Pueblos y Comunidades en el Sur Andino (Siglos XVI-XX)," in "Reproducción y Transformación de las Sociedades Andinas, Siglos XVI-XX," Quito: Ediciones ABYA-YALA, 1991, pp. 91–137.

Sánchez-Albornoz, Nicolás, *Indios y Tributos en el Alto Perú*, Lima: Instituto de Estudios Peruanos, 1978.

Sánchez, Rogger H Ravines, *Inventario de Monumentos Arqueológicos del Perú*, Instituto Nacional de Cultura, 1985.

Sandweiss, Daniel H and James B Richardson, "Central Andean Environments," in "The Handbook of South American Archaeology," Springer, 2008, pp. 93–104.

Shannon, Claude Elwood, "A Mathematical Theory of Communication," *Bell System Technical Journal*, 1948, *27* (3), 379–423.

Sibley, Chris G and John Duckitt, "Personality and Prejudice: A Meta-Analysis and Theoretical Review," *Personality and Social Psychology Review*, 2008, *12* (3), 248–279.

Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations* 1776.

Soldán, Mariano Felipe Paz, *Diccionario Geográfico Estadístico del Perú*, Lima: Imprenta del Estado, 1877.

Stanish, Charles, "Household Archeology: Testing Models of Zonal Complementarity in the South Central Andes," *American Anthropologist*, 1989, *91* (1), 7–24.

_ , "The Origin of State Societies in South America," *Annual Review of Anthropology*, 2001, *30* (1), 41–64.

_ , "Migration, Colonies, and Ethnicity in the South-Central Andes," in "Us and Them: Archaeology and Ethnicity in the Andes," Los Angeles: The Cotsen Institute of Archaeology, UCLA, 2005, pp. 226–232.

Stern, S. J., "The Variety and Ambiguity of Native Andean Intervention in European Colonial Markets," in "Ethnicity, Markets, and Migration in the Andes: At the Crossroads of History and Anthropology," Duke University Press, 1995, pp. 73–100.

Tapia, M., "Diagnóstico de los Ecosistemas de Montañas en el Perú," Technical Report, FAO–MINAM (Ministerio del Ambiente, Perú) 2013.

Tello, Julio César, "Origen y Desarrollo de las Civilizaciones Prehistóricas Andinas," in "Actas y Trabajos. XXVII Congreso Internacional de Americanistas," Vol. 1 1939, pp. 589–720.

Valverde, Guido, María Inés Barreto Romero, Isabel Flores Espinoza, Alan Cooper, Lars Fehren-Schmitz, Bastien Llamas, and Wolfgang Haak, "Ancient DNA Analysis Suggests Negligible Impact of the Wari Empire Expansion in Peru?s Central Coast during the Middle Horizon," *PLoS One*, 2016, *11* (6), e0155508.

Vidal, Javier Pulgar, "Las Ocho Regiones Naturales del Perú," in "Boletín del Museo de Historia Natural "Javier Prado"" number 17 1941, pp. 145–160. Imprenta D. Miranda, Lima.

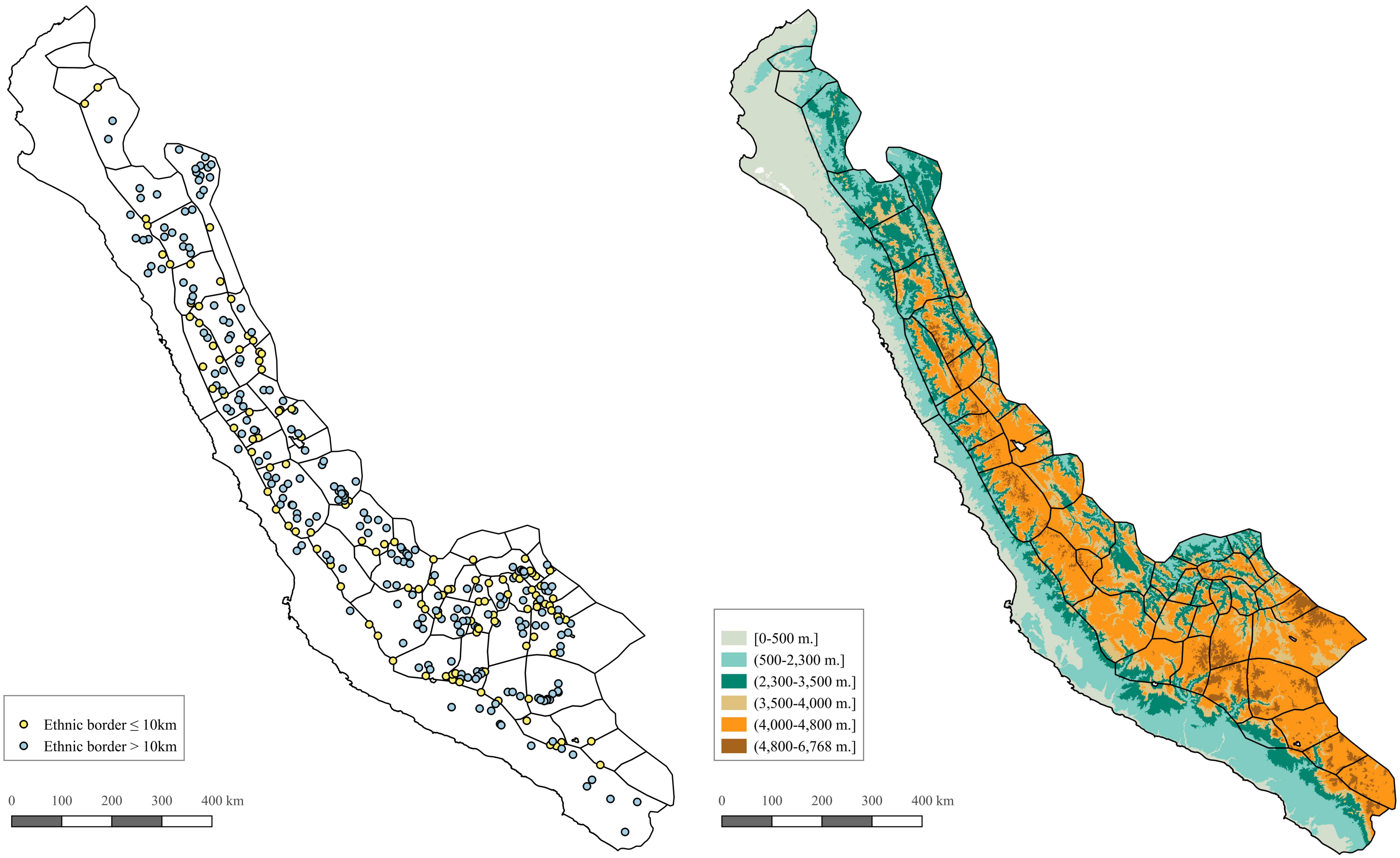
—, *Geografía del Perú. Las Ocho Regiones Naturales del Perú*, 12 ed., Lima: Instituto de Ciencias de la Naturaleza, Territorio y Energías Renovables de la Pontificia Universidad Católica del Perú (INTE-PUCP), 2012.

Voigtländer, Nico and Hans-Joachim Voth, "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany," *The Quarterly Journal of Economics*, 2012, *127* (3), 1339–1392.

Vollmer, Günter, *Bevölkerungspolitik und Bevölkerungsstruktur im Vizekönigreich Peru zu Ende der Kolonialzeit, 1741-1821* 1967.

Wachtel, N., "Los Indios y la Conquista Española," in "América Latina en la Época Colonial," Vol. 1, Barcelona: Crítica, 2002, pp. 153–185.

Wachtel, Nathan, *Los Vencidos. Los Indios del Perú frente a la Conquista Española (1530-1570)*, Madrid: Alianza Editorial, 1976.

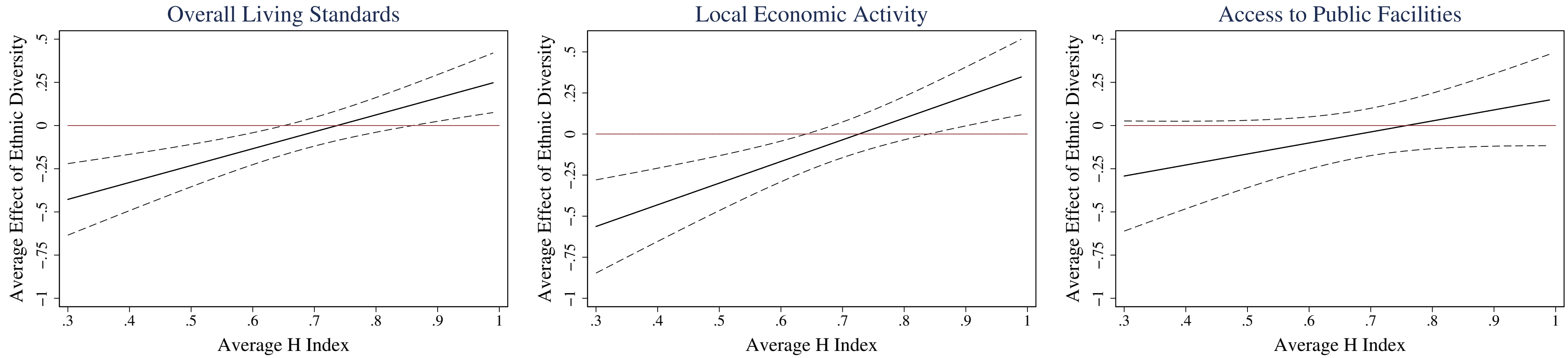


(a) Colonial Parishes

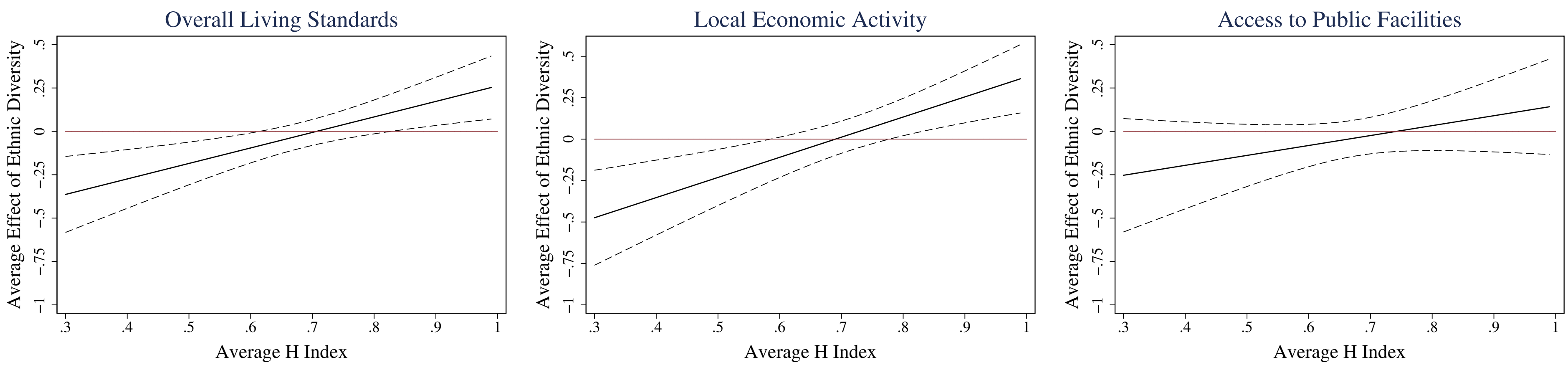
(b) Resource-Producing Zones

FIGURE 1: Colonial Parishes and Resource-Producing Zones

Notes. Lines in black represent the approximate extent of the groups at the time of the Spanish conquest (Rowe 1946). In Panel (a), dots represent colonial parishes: those with an ethnic border within a 10-km buffer from the parish capital are displayed in yellow; the remaining are displayed in blue. Panel (b) displays natural resource-producing zones (Pulgar Vidal 1941). Elevation intervals refer to meters above sea level. For elevation data, I use version 1.2 of the Harmonized World Soil Database (FAO). It provides 30 arc-second raster data with median elevation constructed based on information from the NASA Shuttle Radar Topographic Mission. The maps are displayed using a World Geodetic System projection (WGS 1984).



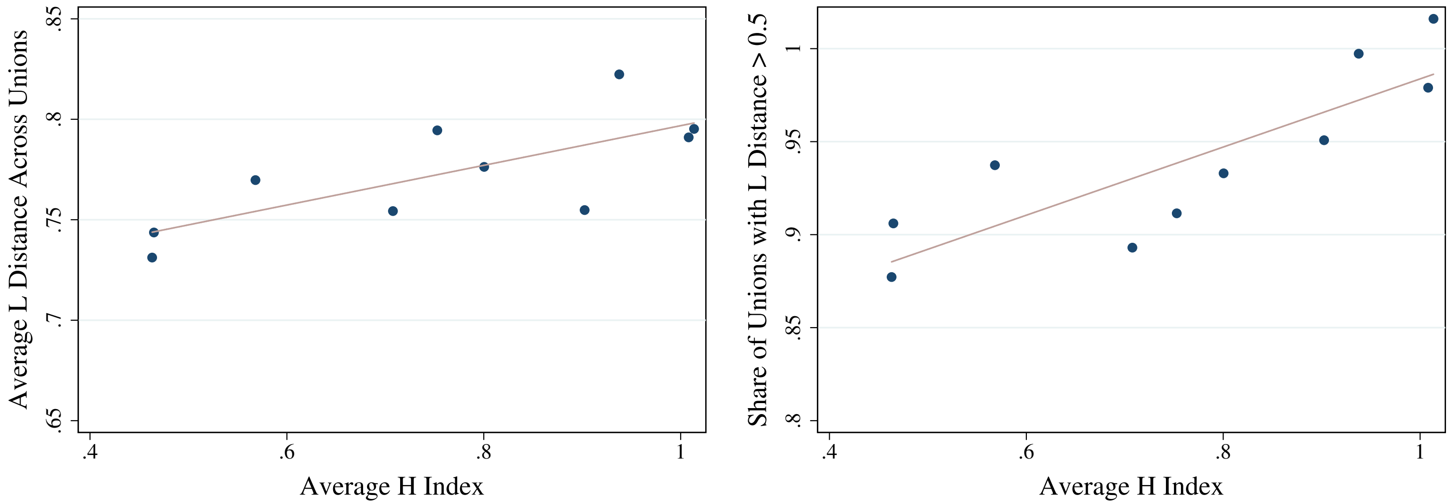
(a) 1990 – 2000



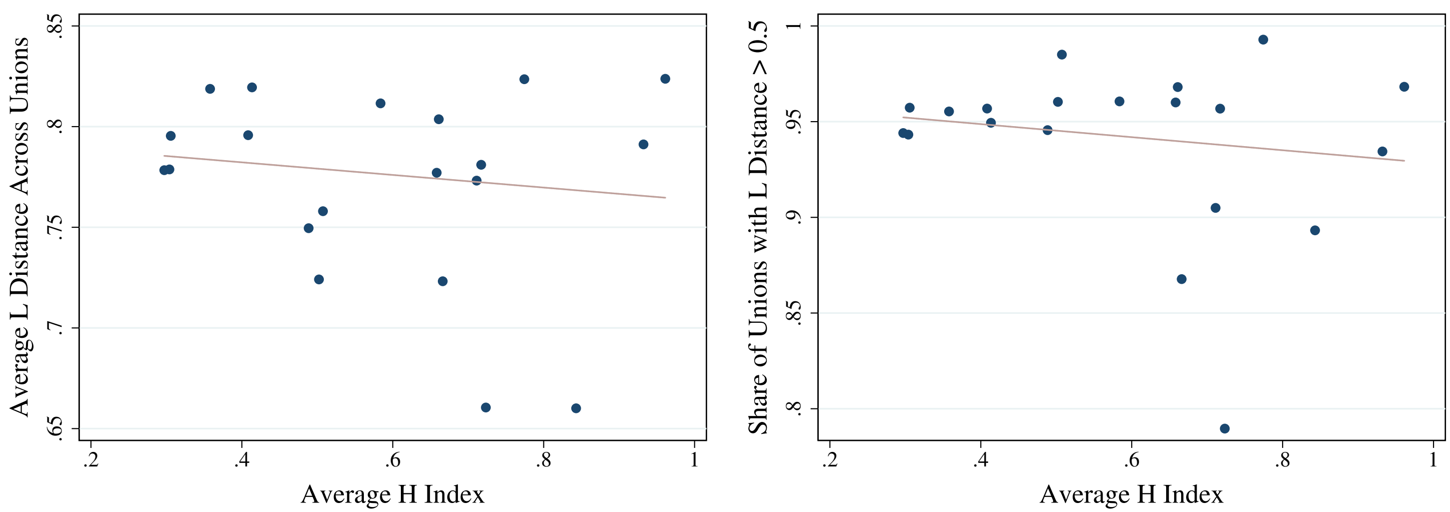
(b) 2010 – 2020

FIGURE 2: Ethnic Diversity and Contemporary Living Standards

Notes. The solid line represents the standardized AES of ethnic diversity after control variables and colonial province fixed effects (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009); see Table 6. The AES for local economic activity refers to the log average light intensity per capita (2000–2003 in Panel a and 2010–2013 in Panel b) and an indicator for nonsubsistence agriculture—a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median (1994 in Panel a and 2012 in Panel b). The AES for access to public facilities refers to the share of dwellings with access to public sanitation (1993 in Panel a and 2017 in Panel b) and the share of dwellings with access to the public water network (1993 in Panel a and 2017 in Panel b). The AES for overall living standards refers to the previous four variables. Dashed lines represent 95 percent confidence intervals.



(a) Parishes with ethnic diversity (10 parishes)



(b) Parishes without ethnic diversity (31 parishes)

FIGURE 3: Within-Group Heterogeneity and Inter-group Unions (1605–1870)

Notes. The figure uses parish-level data. Binned scatterplots controlling for the log number of individuals found in the records of the parish for the period 1605–1870. In the graphs on the left side, the outcome variable is the mean normalized Levenshtein distance across all unions of individuals. In the graphs on the right side, the outcome variable is the share of unions with a normalized Levenshtein distance above 0.5.

Table 1: Validating Ethnic Diversity

	Dependent Variable: Surname Diversity (1605 – 1780)							
	S Index	S Index	S Index	S Index	D Index	D Index	D Index	D Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Baseline							
Ethnic diversity (dummy)	0.512	0.447	0.533	0.558	0.481	0.448	0.477	0.503
	[0.212]**	[0.182]**	[0.198]***	[0.202]***	[0.170]***	[0.182]**	[0.183]**	[0.203]**
	(0.213)**	(0.207)**	(0.220)**	(0.214)***	(0.217)**	(0.201)**	(0.240)**	(0.250)**
	Panel B: Nonunique surnames							
Ethnic diversity (dummy)	0.462	0.408	0.476	0.502	0.454	0.428	0.436	0.456
	[0.209]**	[0.159]**	[0.175]***	[0.185]***	[0.173]**	[0.178]**	[0.170]**	[0.188]**
	(0.208)**	(0.180)**	(0.210)**	(0.212)**	(0.218)**	(0.196)**	(0.230)*	(0.246)*
	Panel C: Grouped surnames							
Ethnic diversity (dummy)	0.482	0.416	0.504	0.525	0.461	0.427	0.454	0.478
	[0.215]**	[0.184]**	[0.199]**	[0.199]**	[0.172]***	[0.181]**	[0.180]**	[0.199]**
	(0.213)**	(0.205)**	(0.214)**	(0.198)***	(0.216)**	(0.199)**	(0.238)*	(0.240)**
Number of parishes	65	65	65	65	65	65	65	65
Ln total individuals (1605–1780)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
% Non-native surnames (1605–1780)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Parish-level controls	No	No	Yes	Yes	No	No	Yes	Yes
Ecclesiastical jurisd. FE	No	No	No	Yes	No	No	No	Yes

Notes. The unit of observation is the parish. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Ethnic diversity takes value 1 if there is an ethnic border within a 10-km buffer from the parish capital, and 0 otherwise. All variables except dummies are standardized to have zero mean and standard deviation equal to one. The vector of parish-level controls includes the mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, and log distance to perennial rivers. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: Validating Resource-Producing Zones

	Dependent Variable:								
	Ln (# Crops)						# Crops	Ln (# Crops)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
H index	0.523		0.518	0.505	0.483	0.498	0.518		
	[0.040]***		[0.042]***	[0.050]***	[0.049]***	[0.050]***	[0.062]***		
	(0.047)***		(0.064)***	(0.048)***	(0.048)***	(0.047)***	(0.063)***		
SD of elevation		0.398	0.007	0.064	0.072	-0.025	-0.017		0.001
		[0.043]***	[0.043]	[0.073]	[0.074]	[0.067]	[0.074]		[0.072]
		(0.061)***	(0.074)	(0.077)	(0.077)	(0.061)	(0.067)		(0.079)
Ln land area					0.124	0.157	0.166	0.172	0.172
					[0.019]***	[0.020]***	[0.030]***	[0.027]***	[0.027]***
					(0.021)***	(0.021)***	(0.034)***	(0.031)***	(0.031)***
Mean elevation						-0.193	-0.161	-0.279	-0.279
						[0.059]***	[0.065]**	[0.081]***	[0.081]***
						(0.062)***	(0.072)**	(0.086)***	(0.090)***
Dummy (number of zones=2)								1.067	1.067
								[0.161]***	[0.166]***
								(0.227)***	(0.225)***
Dummy (number of zones=3)								1.839	1.838
								[0.163]***	[0.167]***
								(0.230)***	(0.215)***
Dummy (number of zones=4)								2.029	2.027
								[0.174]***	[0.205]***
								(0.243)***	(0.246)***
Hydrographic basin FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	526	526	526	526	526	526	526	526	526

Notes. The unit of observation is the 25 km × 25 km grid cell. The table reports OLS estimates. Robust standard errors in brackets; clustered at the FE level in Columns 4-9. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). The dependent variable refers to the number of native crops. All variables except dummies are standardized to have zero mean and standard deviation equal to one. All regressions control for longitude and latitude. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Precolonial Correlates of Within-Group Heterogeneity

	Dependent Variable:								
	Mean Elevation	Mean Caloric Suit.	Ln River Density	Ln Land Area	Ln Population	Ln Population Density	Dummy Urbanization	Dummy Political Complexity	Dummy Elite Residences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
H index	-0.268	0.203	0.041	-0.205	0.119	0.304	0.104	0.155	0.036
	[0.193]	[0.126]	[0.120]	[0.148]	[0.138]	[0.121]**	[0.045]**	[0.050]***	[0.053]
	(0.234)	(0.150)	(0.123)	(0.198)	(0.140)	(0.108)***	(0.048)**	(0.064)**	(0.059)
Observations	47	47	47	47	46	46	47	47	47

Notes. The unit of observation is the ethnic group. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). All variables except for dummies are standardized to have zero mean and standard deviation equal to one. The dummy variables for urbanization, political complexity, and elite residences take value 1 for 12.77, 21.28, and 21.28 percent of the groups, respectively. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Balance Tests for Ethnic Diversity

	Ethnic Diversity = 1		Ethnic Diversity = 0				
	mean	sd	mean	sd	Diff.	p-value[a]	p-value[b]
(1) Mean elevation	3478.823	528.971	3407.290	735.120	-71.533	[0.306]	[0.497]
(2) SD of elevation	479.749	188.305	447.310	178.408	-32.439	[0.127]	[0.199]
(3) Mean caloric suitability	126.528	275.787	117.825	262.678	-8.703	[0.780]	[0.833]
(4) SD of caloric suitability	139.022	241.878	122.762	219.047	-16.260	[0.545]	[0.637]
(5) Ln dist. to perennial river	0.673	1.054	0.752	1.082	0.079	[0.516]	[0.609]
(6) Ln dist. to native shrine	4.159	0.999	4.309	1.090	0.150	[0.205]	[0.384]
(7) Ln expected tribute (16th c.)	6.516	0.724	6.504	0.646	-0.012	[0.884]	[0.845]
(8) Ln dist. to *mita* mine	5.667	0.744	5.702	0.727	0.035	[0.680]	[0.736]
Number of parishes	117	117	219	219	336	336	336

Notes. The unit of observation is the parish. Ethnic diversity takes value 1 if there is an ethnic border within a 10-km buffer from the parish capital, and 0 otherwise. P-values from OLS regressions of each of the variables listed in the first column on ethnic diversity; (^{a}) with robust standard errors, (^{b}) with standard errors corrected for spatial dependence using a distance cutoff of approximately one degree at the equator (Colella et al. 2019). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development I

	Overall Living Standards (AES, 1990 – 2000)							
	Full Sample				Ethnic Div = 1	Ethnic Div = 0	Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic diversity (dummy)	-0.200***	-0.167**	-0.117**				-0.554**	-0.598***
	[0.070]	[0.066]	[0.060]				[0.222]	[0.231]
Average H index				0.481**	0.834*	0.420*	0.251	0.228
				[0.192]	[0.495]	[0.223]	[0.205]	[0.269]
Ethnic div. × Av. H index							0.655**	0.780**
							[0.329]	[0.346]
Baseline controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ecclesiastical jurisd. FE	No	No	Yes	Yes	Yes	Yes	Yes	No
Colonial province FE	No	No	No	No	No	No	No	Yes
Number of parishes	336	336	336	336	117	219	336	336
Joint significance (p-value)							0.013	0.022

Notes. The unit of observation is the parish. Robust standard errors in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009) across four outcomes: the log average light intensity per capita (2000–2003), an indicator for nonsubsistence agriculture (1994, a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median), the share of dwellings with access to public sanitation (1993), and the share of dwellings with access to the public water network (1993). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. The p-value refers to the joint significance of ethnic diversity terms. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development II

	Overall Living Standards			Local Econ. Activity		Public Facilities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: AES (1990 – 2000)			
Ethnic diversity (dummy)	-0.826***	-0.598***	-0.721***	-0.931***	-0.958***	-0.265	-0.484*
	[0.250]	[0.208]	[0.178]	[0.243]	[0.243]	[0.308]	[0.270]
Ethnic div. × Av. H index	0.938***	0.780**	0.978***	1.276***	1.319***	0.284	0.638*
	[0.358]	[0.306]	[0.252]	[0.339]	[0.344]	[0.476]	[0.381]
Joint significance (p-value)	0.000	0.010	0.000	0.001	0.000	0.480	0.198
				Panel B: AES (2010 – 2020)			
Ethnic diversity (dummy)	-0.842***	-0.616**	-0.632***	-0.825***	-0.839***	-0.407	-0.425
	[0.281]	[0.249]	[0.191]	[0.247]	[0.243]	[0.401]	[0.288]
Ethnic div. × Av. H index	1.036***	0.879**	0.894***	1.193***	1.215***	0.565	0.572
	[0.394]	[0.386]	[0.275]	[0.343]	[0.334]	[0.624]	[0.418]
Joint significance (p-value)	0.005	0.036	0.004	0.002	0.001	0.536	0.309
Baseline controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. den. and rural dummy	No	No	Yes	No	Yes	No	Yes
Number of parishes	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of colonial province in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009) across different outcomes. The AES for local economic activity refers to the log average light intensity per capita (2000–2003 in Panel A and 2010–2013 in Panel B) and an indicator for nonsubsistence agriculture—a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median (1994 in Panel A and 2012 in Panel B). The AES for access to public facilities refers to the share of dwellings with access to public sanitation (1993 in Panel A and 2017 in Panel B) and the share of dwellings with access to the public water network (1993 in Panel A and 2017 in Panel B). The AES for overall living standards refers to the previous four variables. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. The rural dummy variable takes vale 1 if the share of rural population is above the median, and 0 otherwise (1993 in Panel A and 2017 in Panel B). The p-value refers to the joint significance of ethnic diversity terms. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Robustness: Precolonial Characteristics of Ethnic Groups

	Overall Living Standards (AES, 2010 – 2020)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ethnic diversity (dummy)	-0.616**	-0.891***	-0.690***	-0.719***	-0.618**	-0.611**	-0.625**	-0.614***	-0.958**
	[0.249]	[0.335]	[0.257]	[0.262]	[0.274]	[0.259]	[0.245]	[0.223]	[0.446]
Ethnic div. × Av. H index	0.879**	0.851**	1.064***	0.709*	0.903**	0.898**	0.996***	0.900**	1.318***
	[0.386]	[0.358]	[0.401]	[0.363]	[0.397]	[0.390]	[0.385]	[0.367]	[0.448]
Ethnic div. × Av. elevation		0.378							-0.248
		[0.319]							[0.538]
Ethnic div. × Av. caloric suitability			-0.733						-1.423
			[0.526]						[0.989]
Ethnic div. × Av. ln river density				0.210					0.347
				[0.143]					[0.231]
Ethnic div. × Av. ln population density					0.019				-0.006
					[0.113]				[0.135]
Ethnic div. × Av. urbanization						-0.070			0.085
						[0.194]			[0.335]
Ethnic div. × Av. political complexity							-0.133		-0.277
							[0.144]		[0.242]
Ethnic div. × Av. elite residences								0.008	0.266
								[0.113]	[0.216]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	336	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009) across four outcomes: the log average light intensity per capita (2010–2013), an indicator for nonsubsistence agriculture (2012, a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Robustness: Precolonial Land Occupation and Placebos for H

	Overall Living Standards (AES, 2010 – 2020)			
	Precolonial Land Occupation		Placebos	
	(1)	(2)	(3)	(4)
Ethnic diversity (dummy)	-0.648**	-0.779**		
	[0.260]	[0.310]		
Ethnic div. × Av. H index (20km correction)	0.922**			
	[0.401]			
Ethnic div. × Av. H index (10km correction)		1.057**		
		[0.452]		
Dummy 1 (Artificial ethnic border within parish buffer)			-0.267	
			[0.495]	
Dummy 1 × Av. H index (Artificial)			0.388	
			[0.617]	
Dummy 2 (*Corregimiento* border within parish buffer)				0.149
				[0.314]
Dummy 2 × Av. H index (*Corregimiento*)				-0.203
				[0.421]
Baseline controls	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes
Number of parishes	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009) across four outcomes: the log average light intensity per capita (2010–2013), an indicator for nonsubsistence agriculture (2012, a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Medium-Term Outcomes: Structural Change and Literacy Rate

	Dependent Variable:								
	Literacy Rate (1876)			Share of Pop. in Secondary Sector (1876)			Share of Pop. in Tertiary Sector (1876)		
	All	Women	Men	All	Women	Men	All	Women	Men
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ethnic diversity (dummy)	-0.081**	-0.071**	-0.094**	-0.101	-0.041	-0.155	-0.113***	-0.147***	-0.085**
	[0.035]	[0.031]	[0.043]	[0.136]	[0.094]	[0.127]	[0.041]	[0.045]	[0.035]
Ethnic div. × Av. H index	0.093**	0.081**	0.107**	0.151	-0.023	0.213	0.153**	0.218**	0.102**
	[0.039]	[0.036]	[0.051]	[0.206]	[0.161]	[0.159]	[0.067]	[0.087]	[0.046]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	282	282	282	282	282	282	282	282	282
Mean Dep. Var.	0.102	0.051	0.158	0.241	0.445	0.108	0.074	0.112	0.071

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. Regressions are weighted by the square root of the population in 1876. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Mechanisms: Technology Adoption

	Dependent Variable:				
	Dummy Insecticides (1994)	Dummy Improved Seeds (1994)	Dummy Bio. Control (1994)	Dummy Chemical Fertilizer (1994)	Dummy Organic Fertilizer (1994)
	(1)	(2)	(3)	(4)	(5)
Ethnic diversity (dummy)	-0.258	-0.368	-0.118	-0.021	0.097
	[0.221]	[0.201]*	[0.205]	[0.231]	[0.237]
	(0.175)	(0.185)**	(0.195)	(0.225)	(0.150)
Ethnic div. × Av. H index	0.339	0.458	0.203	-0.031	-0.131
	[0.342]	[0.315]	[0.315]	[0.376]	[0.386]
	(0.229)	(0.273)*	(0.357)	(0.346)	(0.218)
Baseline controls	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes
Number of parishes	334	334	334	334	334
Mean Dep. Var.	0.500	0.500	0.500	0.500	0.500

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). All regressions control for the log of the total number of farmers. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: Mechanisms: Cultural Transmission

	Dependent Variable:					
	Dummy Neigh.	Dummy Agr.	Participation Dummy (2010-2017)			
	Assoc. (2002)	Assoc. (1994)	Neigh. Assoc.	Professional Assoc.	Labor Union	Sport Club
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic diversity (dummy)	-0.329	-0.343	-0.096	-0.036	-0.080	-0.009
	[0.145]**	[0.145]**	[0.041]**	[0.018]**	[0.038]**	[0.014]
	(0.129)**	(0.155)**	(0.055)*	(0.022)	(0.046)*	(0.005)*
Ethnic div. × Av. H index	0.571	0.514	0.137	0.054	0.101	0.012
	[0.239]**	[0.220]**	[0.055]**	[0.027]**	[0.048]**	[0.019]
	(0.202)***	(0.240)**	(0.081)*	(0.041)	(0.058)*	(0.003)***
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes
Number of parishes	334	334	262	262	262	262
Number of individuals			36,522	36,522	36,522	36,522
Mean Dep. Var.	0.128	0.500	0.071	0.042	0.054	0.025

Notes. The unit of observation is the parish in Columns 1-2 and the individual in Columns 3-6. In brackets, robust standard errors clustered at the level of the colonial province (Columns 1-2) or parish (Columns 3-6). In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Column 1 controls for the log of the total population and Column 2 for the log of the total number of farmers. In Columns 3-6, the vector of individual-level controls includes gender, age, age squared, years of schooling, civil status, and mother tongue. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 12: Mechanisms: Cultural Transmission or Economic Complementarities?

	Overall Living Standards (AES, 1990 – 2000)		Dummy Retail Market (1993)		
	(1)	(2)	(3)	(4)	(5)
High min_{50}	0.446				
	[0.374]				
High maj_{50}	-0.135				
	[0.296]				
High min_{50} × High maj_{50}	-0.080				
	[0.458]				
High min_{75}		0.957	0.343	0.292	0.320
		[0.346]***	[0.165]**	[0.152]*	[0.175]*
			(0.144)**	(0.136)**	(0.161)**
High maj_{75}		0.393	-0.058	-0.100	-0.071
		[0.263]	[0.073]	[0.083]	[0.077]
			(0.056)	(0.058)*	(0.066)
High min_{75} × High maj_{75}		-0.494	-0.191	-0.195	-0.176
		[0.647]	[0.154]	[0.147]	[0.157]
			(0.142)	(0.134)	(0.148)
Ln population (1993)				0.078	
				[0.039]*	
				(0.034)**	
Ln population density (1993)					0.017
					[0.026]
					(0.023)
Baseline controls	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes
Number of parishes	117	117	117	117	117
Mean Dep. Var.			0.077	0.077	0.077

Notes. The unit of observation is the parish. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). High min_{50} and High min_{75} are dummy variables indicating that the minority group's H index is above the 50th or 75th percentile, respectively. Variables High maj_{50} and High maj_{75} are analogous dummies for the majority group. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to precolonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Online Appendix

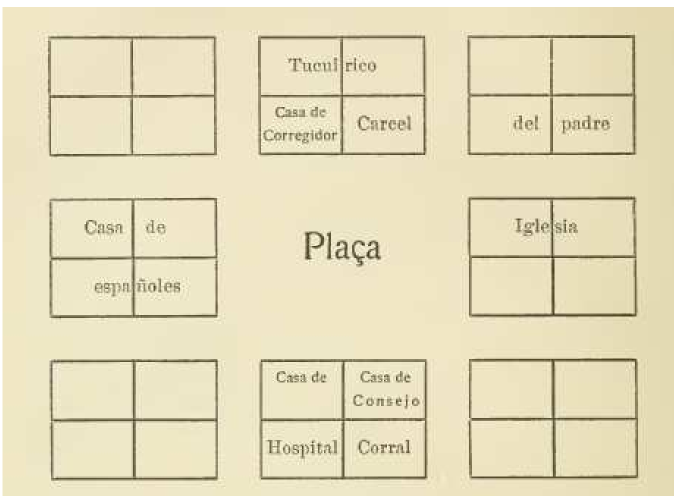
Within-Group Heterogeneity in a Multi-Ethnic Society

Miriam Artiles
PUC-Chile

April 16, 2022

A Appendix - Historical Setting

The Spanish model. Subfigure (a) shows the model of a *reducción*, designed in 1567 by Matienzo (1910)[1567]. Subfigures (b) and (c) show contemporary aerial views of Yanque, Collaguas, created as a result of the resettlement policy in the 16th century (Servicio Aerofotográfico Nacional del Perú, in Medina (1993), and Google Earth imagery, respectively).



(a) Model of *reducción* in Matienzo (1567)

(b) Aerial view of Yanque (in Medina 1993)

(c) Contemporary aerial view of Yanque (Google Earth)

Anecdotal evidence: the Spanish administration in a mountain environment. The Andean environment was different from what Spanish observers of the 16th century had seen before (p. 55, Murra 2002b). Interestingly, as noted by Pulgar Vidal (1941, 2012), Spanish chronicles and colonial documents employed morphological terms used in 16th-century Spain to describe the geography of the Peruvian territory. In particular, they distinguished between three major regions. The word *coast* was used for flat territories with direct access to the ocean, *sierra* for the mountainous territory of the Andes, and *jungle* for the Amazon rainforest. These terms have persisted over time to describe the Peruvian territory in a broad way. However, Pulgar Vidal's studies, which began by analyzing indigenous knowledge of geography embedded in native folklore, support that neither these terms were used by native populations to describe the territory nor a concept of territory divided into three main regions existed before the Spanish colonization. Historical studies also suggest that colonial officials were not fully aware of native practices at the time. For example, it was common for native populations to use different combinations of crop rotation and fallowing in order to increase soil productivity. A common source of conflict with native populations appeared when Spanish officials found uncultivated lands−they generally thought the lands were abandoned when, in fact, they were in a fallow period (Pease 1989).

The census of 1791-95. Parishes are displayed by administrative region (*intendencia*) and province (*partido*). After the Bourbon reforms of 1784-1785, the viceroyalty was divided into *intendencias*, and *intendencias* were, in turn, divided into *partidos*. The census covers the territory under the Viceroyalty of Peru, thus excluding parishes in the *intendencia* of Puno. Puno was under the jurisdiction of the Audiencia of Charcas (modern Bolivia), in the Viceroyalty of Río de la Plata, until 1795 (*Real Cédula* of February 1, 1796); see Lynch (1962, p. 67-68) for more details. A summary of the census was published as an appendix to Manuel Fuentes' *Memorias de los virreyes que han gobernado el Peru* (1859, vol. 6, p. 6-9). The document was signed by José Ignacio de Lequanda and dated January 10, 1796. The whole census with figures at the parish level was later published in Vollmer (1967), where it is referred to as "Census of 1792." The census is considered a baseline for the study of population before independence from Spain (Gootenberg 1991).

Consequently, the ethnic groups in Rowe (1946) whose territories were not under the Viceroyalty of Peru are not part of the analysis. Most of these groups were under the jurisdiction of the Audiencia of Charcas at the time of the census: *Pacasa* or *Pacaje*, *Caranga* or *Caranca*, *Charca*, *Quillaca* or *Quillagua*, *Omasuyo*, *Collahuaya*, *Cochapampa*, *Yamparú*, *Chicha*, *Lipe*, and *Uru* (all in modern Bolivia). The census also excludes the territories of the *Lupaca* and *Colla* (in Puno, modern Peru); and *Tarapacá* (modern Chile). The *Moyopampa* group, in the Amazon region, also lies outside the area of interest, as well as *Tarata* and *Calva*, with no colonial parishes in their territories.

B Appendix - Surnames

Introduction of surnames in Peru. Historical chronicles describe the social unit at the time of the Spanish conquest as an endogamous group of several extended families with ancestry traced through the male line (Rowe 1946). Before the expansion of the Inca empire, the groups claimed descent from a mythical ancestor, usually an animal or some element of nature, which was worshiped and sometimes honored with rites and sacrifices (see Garcilaso de la Vega (1960)[1609], first book). Historical evidence suggests that no system of family names existed prior to the arrival of the Spanish, but rather first names related to the mythical ancestor. The system of family names was introduced by the Catholic Church with the purpose of religious indoctrination. At least since the First Council of Lima in 1551-52, one of the main tasks of Spanish priests was the baptism of children and adults (de Armas Medina 1953, ch. 10). To my knowledge, there were no specific instructions regarding the choice of first names and surnames. While the adoption of Hispanic surnames over time may represent a limitation, qualitative evidence suggests that the common practice was for priests to choose a Hispanic first name, with the mythical first names of the parents adopted as surnames (RENIEC 2012). Garcilaso de la Vega (1960)[1609] also suggests that surnames adopted by native populations were initially related to their ethnic origin. See Carpio and Guerrero (2021) for further details on the introduction of surnames in Peru.

Colonial baptism records. The website FamilySearch.org provides access to baptism records from colonial Peru. The organization, which seeks to help trace users' ancestry, seeks volunteers from around the world to make indexed genealogical records freely available. The results in Table 1 use information from the collection "Perú, bautismos, 1556-1930."[59] Each baptism record includes information on the full name and gender of the individual, name of the parish, and date of baptism. The original handwritten record has also been uploaded in some cases and can be easily accessed.

URL

https://familysearch.org/ark:/61903/1:1:FJFK-J97

Información indexada

Ocultar

Imagen no disponible

Nombre	**Catharina Huaman**
Sexo	Female
Fecha de bautismo	01 May 1743
Lugar de bautismo	San Juan Bautista, Yanahuara, Arequipa, Peru
Fecha de nacimiento	30 Apr 1743
Nombre del padre	Ramon Huaman
Nombre de la madre	Pasquala Machaca

Cita

"Perú, bautismos, 1556-1930," database, *FamilySearch* (https://familysearch.org/ark:/61903/1:1:FJFK-J97 : 10 February 2018), Catharina Huaman, 30 Apr 1743; citing San Juan Bautista, Yanahuara, Arequipa, Peru, reference v 2 p 53, index based upon data collected by the Genealogical Society of Utah, Salt Lake City; FHL microfilm 1,155,316.

Identification of native surnames. I first excluded Hispanic and foreign surnames from the analysis. The main source for the identification of Hispanic surnames is Platt (1996), which includes an index of Hispanic surnames developed in Latin America and the United States. The author writes "the word Hispanic refers to individuals born in Latin America or the United States, whose parents speak Spanish and whose principal cultural background was Spanish." This source includes the list of surnames in Carraffa and Carraffa (1920–1963), the traditional reference for Hispanic surnames.[60] I complement Basque surnames using a list provided by

[59] `https://www.familysearch.org/search/collection/location/1927168?region=Peru`. Accessed in December 2018. Index based upon data collected by the Genealogical Society of Utah, Salt Lake City.

[60] The suggestions of the Biblioteca Nacional de España can be accessed here. The list of surnames in Carraffa and Carraffa (1920–1963) can also be accessed through The Library of Congress.

the Real Academia de la Lengua Vasca.

In order to identify native surnames, I then constructed a dictionary of linguistic roots from the Quechuan and Aymaran language families. There is no unique source for the identification of surnames from these families. The transformation of native surnames over time (*castellanización*), as well as the presence of many regional varieties of Quechua and Aymara, make necessary the combination of different (temporal and regional) sources. For Quechua, the main sources are the classic dictionary by González Holguín (1952)[1608] and a recent dictionary compiled by the Academia Mayor de la Lengua Quechua (2005). I also include the list of names provided by the Peruvian *Registro Nacional de Identificación y Estado Civil* (RENIEC 2012). For Aymara, the main sources are the classic dictionary by Bertonio (2011)[1612], the list of surnames provided by De Lucca (1983), and a recent dictionary compiled by CONADI (2011). I complement the analysis using two additional sources: (1) *Vocabulario Políglota Incaico*, originally compiled by Franciscan missionaries in Peru, which provides an extensive list of words in four dialects of Quechua (varieties of Cuzco, Ayacucho, Junín and Ancash) and Aymara, see Fide (1998)[1905]; and (2) the *An Crúbadán-Corpus Building for Minority Languages* project, which provides downloadable text datasets for different dialects of Quechua and Aymara based on online text resources, including translations of the Bible and the Universal Declaration of Human Rights.

The following table reports the total number of individuals by time period, as well as the corresponding number of parishes and the number of individuals in the mean and median parish. The statistics refer to individuals with native paternal surname.

Descriptive Statistics - Dataset of Baptisms

By period	# Individuals	# Parishes	Mean # Individuals	Median # Individuals
[1605, 1625]	848	8	106	16.5
(1625, 1650]	5,039	19	265.211	145
(1650, 1675]	8,033	30	267.767	125.5
(1675, 1700]	19,195	40	479.875	209
(1700, 1725]	17,947	49	366.265	197
(1725, 1750]	21,172	46	460.261	205
(1750, 1780]	40,106	63	636.603	184
[1605, 1780]	112,340	65	1,726.754	576

C Appendix - Data

C.1 Parish-Level Variables

Mean elevation. Average elevation across all grid cells with centroid within a 10-km buffer from the parish capital. Source: author's computation using version 1.2 of the Harmonized World Soil Database (FAO). It provides 30 arc-second raster data with median elevation based on information from the NASA Shuttle Radar Topographic Mission (SRTM).

Variation in elevation. Standard deviation of elevation across all grid cells with centroid within a 10-km buffer from the parish capital. Source: see *Mean elevation*.

Mean caloric suitability. Average pre-1500 land caloric suitability across all grid cells with centroid within a 10-km buffer from the parish capital. Source: author's computation using the Caloric Suitability Index constructed by Galor and Özak (2016b), which provides information on potential crop yield given the set of available crops before 1500CE (5 arc-minute raster data).

***Variation in caloric suitability*.** Standard deviation of pre-1500 land caloric suitability across all grid cells with centroid within a 10-km buffer from the parish capital. Source: see *Mean caloric suitability*.

***Mean caloric suitability for maize*.** Average pre-1500 land caloric suitability for maize across all grid cells with centroid within a 10km buffer from the parish capital. Source: See *Mean caloric suitability*.

***Mean caloric suitability for potato*.** Average pre-1500 land caloric suitability for potato across all grid cells with centroid within a 10km buffer from the parish capital. Source: See *Mean caloric suitability*.

***Ln distance to perennial river*.** Natural log of the geodesic distance (km) from the parish capital to the closest perennial river. Source: author's computation using water area features from version 10.0 of the Seamless Digital Chart of the World.

***Ln expected tribute*.** Natural log of the total tribute (*pesos ensayados*) in the 16th century. Source: Cook (1982) and Puente Brunke (1991). The information exists for 117 parishes; for the remaining, it is imputed using the average of the province. The year of the data ranges from 1570 to 1594, depending on the parish. Period: colonial.

***Ln distance to mita mine*.** Natural log of the geodesic distance (km) from the parish capital to the closest mine subjected to the *mita* (as defined in Dell 2010). Source: author's computation. Period: colonial.

***Ecclesiastical jurisdiction*.** Categorical variable indicating the colonial bishopric (Lima, Arequipa, Huamanga, Trujillo, and Cuzco). Source: "Guía PolÃŋtica, Eclesiástica y Militar del Virreinato del Perú para el Año de 1797," Unanue (1797). Period: colonial.

Administrative province. Categorical variable indicating the colonial administrative province (*partido*). Source: census of 1791-95 (Viceroy Gil de Taboada y Lemos). Period: colonial.

Religious order. Categorical variable indicating the religious order in charge of the parish during most of the colonial period (Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, various regular orders, and secular clergy). The last category is assigned if no specific order was in charge of the parish during most of the colonial period. Sources: author's coding using the information in Lissón Chávez (1943), de Armas Medina (1953), de Córdoba Salinas (1957)[1651], and García (1997). Period: colonial.

Ln distance to native shrine. Natural log of the geodesic distance (km) from the parish capital to the closest pre-colonial shrine, according to archaeological records. Source: author's computation; for the main sources of information on pre-colonial archaeological sites, see *Dummy urbanization* (Appendix C.3). Period: pre-colonial.

Ln distance to defensive site. Natural log of the geodesic distance (km) from the parish capital to the closest pre-colonial defensive site, according to archaeological records. Source: author's computation; for the main sources of information on pre-colonial archaeological sites, see *Dummy urbanization* (Appendix C.3). Period: pre-colonial.

Ln distance to administrative site. Natural log of the geodesic distance (km) from the parish capital to the closest pre-colonial administrative site, according to archaeological records. Source: author's computation; for the main sources of information on pre-colonial archaeological sites see *Dummy urbanization* (Appendix C.3). Period: pre-colonial.

Ln distance to Inca road network. Natural log of the geodesic distance (km) from the parish capital to the closest Inca road. Source: author's computation using the map of the Qhapaq Ñan produced by SIGDA (Sistema de Información Geográfica de Arqueología, Ministerio de Cultura, Perú), accessed in March 2021. Period: pre-colonial.

Ln indigenous population. Natural log of the population classified as "indigenous" by colonial authorities. Source: census of 1791-95 (Viceroy Gil de Taboada y Lemos). Period: colonial.

% mestizo population. Percentage of the population classified as "mestizo" by colonial authorities. Source: census of 1791-95 (Viceroy Gil de Taboada y Lemos). Period: colonial.

Ln priests per capita. Natural log of the number of priests divided by "indigenous" population. Source: census of 1791-95 (Viceroy Gil de Taboada y Lemos). Period: colonial.

Ln light intensity per capita. Natural log of 1 plus average light intensity per capita. The average sum of light intensity values across all grid cells with centroid within the 10-km buffer is divided by total population within the same buffer. Source: average cloud free coverages of the DMSP-OLS Nighttime Lights Time Series, produced by the NOAA's National Geophysical Data Center, which provide 30 arc-second yearly raster data. Data from satellites F15 and F18 for the periods 2000-2003 and 2010-2013, respectively (yearly averages from the same satellite). Version 4.10 of the Gridded Population of the World (Center for International Earth Science Information Network−CIESIN) provides 30 arc-second raster data with population counts for the years 2000 and 2010. Population counts are developed through the uniform areal-weighting method using census data adjusted to match the United Nation's population counts at the country level. Period: contemporary.

Non-subsistence agriculture. Dummy variable taking value 1 if the share of agricultural producers devoting most of the harvest to sale or trade in local markets is above the median, and 0 otherwise. Source: 1994 and 2012 national agricultural censuses, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Access to public sanitation. Share of occupied dwellings with access to the public sewer system (inside or outside the dwelling unit). Source: 1993 and 2017 national population and housing censuses, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Access to public water. Share of occupied dwellings with access to the public network of water supply (inside or outside the dwelling unit). Source: 1993 and 2017 national population and housing censuses, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Ln population density. Natural log of total population divided by total land area. Source: author's computation using population data from the 1993 and 2017 national population and housing censuses, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummy rural. Dummy variable taking value 1 if the share of rural population is above the median, and 0 otherwise. Source: 1993 and 2017 national population and housing censuses, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Literacy rate. Share of literate population (those who can read and/or write). Source: 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878). Period: post-independence.

Share of employment in secondary sector. Share of population employed in the secondary sector. Source: author's coding using data on occupations in the 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878). Period: post-independence.

Share of employment in tertiary sector. Share of population employed in the tertiary sector. Source: author's coding using data on occupations in the 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878). Period: post-independence.

Dummy insecticides. Dummy variable taking value 1 if the share of agricultural producers reporting to use insecticides is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummy improved seeds. Dummy variable taking value 1 if the share of agricultural producers reporting to use improved seeds is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummy biological control. Dummy variable taking value 1 if the share of agricultural producers reporting to have knowledge of biological control is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummy chemical fertilizer. Dummy variable taking value 1 if the share of agricultural producers reporting to use chemical fertilizer is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummy organic fertilizer. Dummy variable taking value 1 if the share of agricultural producers reporting to use organic fertilizer is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummy neighborhood association. Dummy variable taking value 1 for the presence of neighborhood associations, and 0 otherwise. Source: 2002 *Registro Nacional de Municipalidades*, provided by the National Institute of Statistics (INEI). Period: contemporary.

Dummy agricultural association. Dummy variable taking value 1 if the share of agricultural producers reporting to participate in agricultural associations is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummy retail market. Dummy variable taking value 1 for the presence of retail markets (*mercados de abastos minoristas*) created before 1993. Source: 2016 CENAMA national census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

C.2 Individual-Level Variables

Participation in voluntary associations. Dummy variables taking value 1 if the individual reports to participate in a voluntary association (separate variables for participation in neighborhood associations, professional associations, labor unions, and sport clubs). Source: 2010-2017 ENAHO surveys, conducted by the National Institute of Statistics (INEI). Period: contemporary.

C.3 Ethnic-Level Variables

Mean elevation. Average elevation across all grid cells with centroid within the ethnic homeland. Source: author's computation; see *Mean elevation* (Appendix C.1).

Mean caloric suitability. Average pre-1500 land caloric suitability across all grid cells with centroid within the ethnic homeland. Source: author's computation; see *Mean caloric suitability* (Appendix C.1).

Ln river density. Natural log of total river length (km, only perennial rivers) divided by total land area (km^2). Source: author's computation; see *Ln distance to perennial river* (Appendix C.1).

Ln land area. Natural log of total land area (km^2) within the ethnic homeland. Source: author's computation.

Ln population. Natural log of approximate population by the time of the Spanish conquest. Source: author's computation using the first estimate of tributary population between 1532 and 1575 for all population centers within the ethnic homeland. Population figures from Cook (1982, 2010).

***Ln population density*.** Natural log of population divided by land area. Source: author's computation; see *Ln population*.

***Dummy urbanization*.** Dummy variable taking value 1 for the presence of pre-colonial towns or urban centers, according to archaeological records. Source: author's coding using information on pre-colonial archaeological sites in Ravines Sánchez (1985), Ramos Giraldo (2001), Isbell and Silverman (2002a, 2008), and an official database of pre-colonial archaeological sites (*Catastro de Monumentos Arqueológicos Prehispánicos*), developed by SIGDA (Sistema de Información Geográfica de Arqueología, Ministerio de Cultura, Perú) and accessed in March 2021.

***Dummy political complexity*.** Dummy variable taking value 1 for the presence of pre-colonial administrative centers and monumental architecture−public buildings and communal spaces, including temples, palaces, and complex mound platforms, according to archaeological records. Source: author's computation; for the main sources of information on pre-colonial archaeological sites see *Dummy urbanization*.

***Dummy elite residences*.** Dummy variable taking value 1 for the presence of elite residences, according to archaeological records. Source: author's computation; for the main sources of information on pre-colonial archaeological sites see *Dummy urbanization*.

D Appendix - Figures

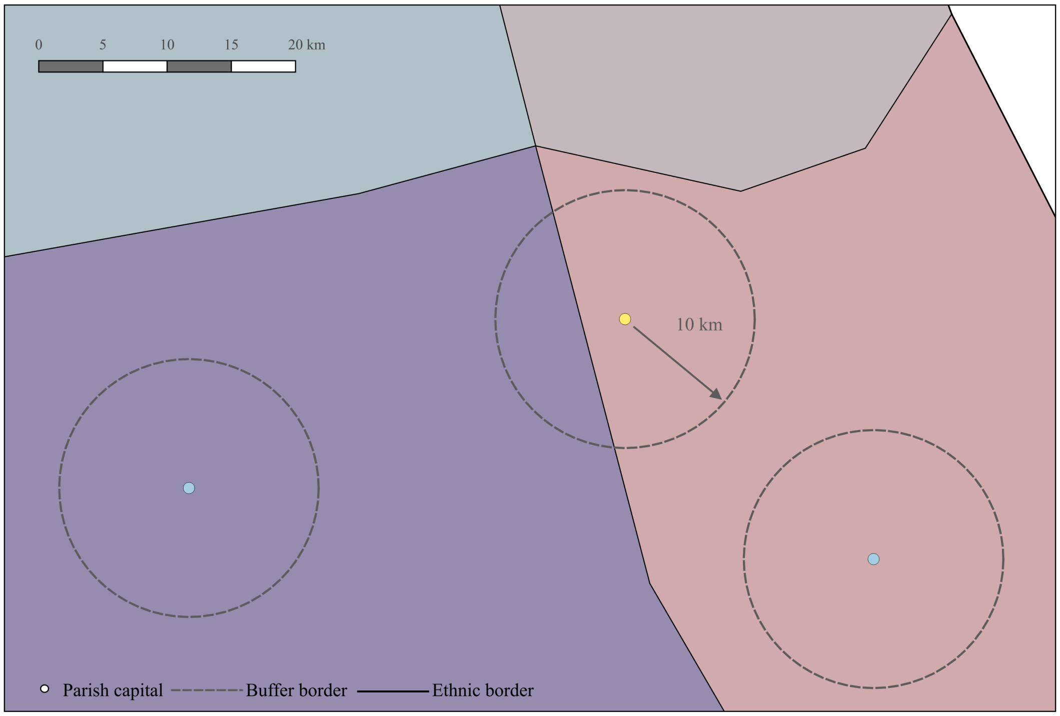


Figure A.1: Buffer Exercise

Notes. Construction of 10-km buffer around each parish capital.

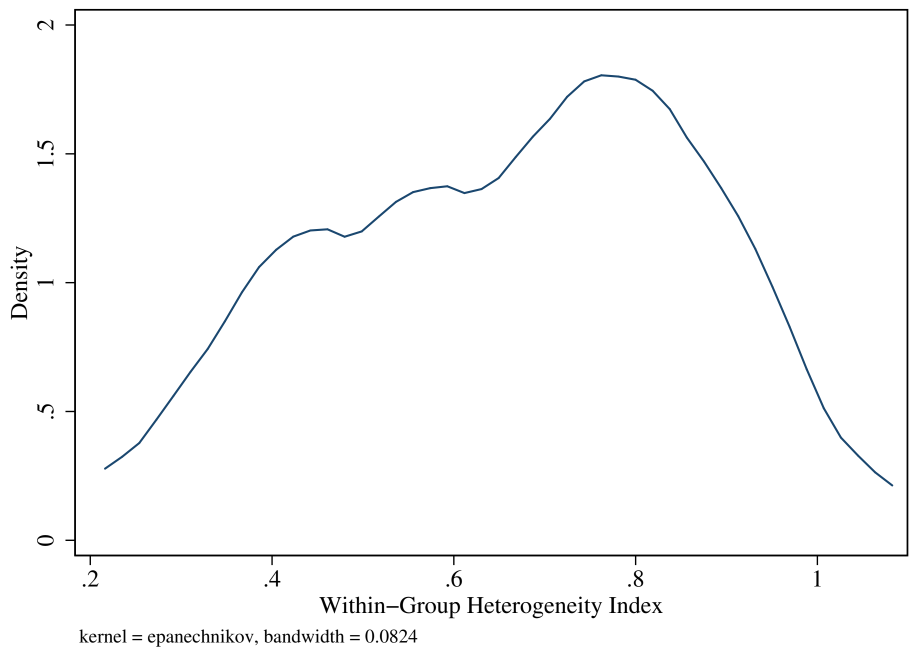


FIGURE A.2: Density of Within-Group Heterogeneity

Notes. Kernel density of within-group heterogeneity at the ethnic group level. Within-group heterogeneity is defined as the the reciprocal of the Herfindahl index: $H_e = 1/\sum_j s_{ej}^2$, where s_{ej} is the area share of resource zone j within the homeland of ethnic group e. The index is normalized to take value 1 for the group with the highest value.

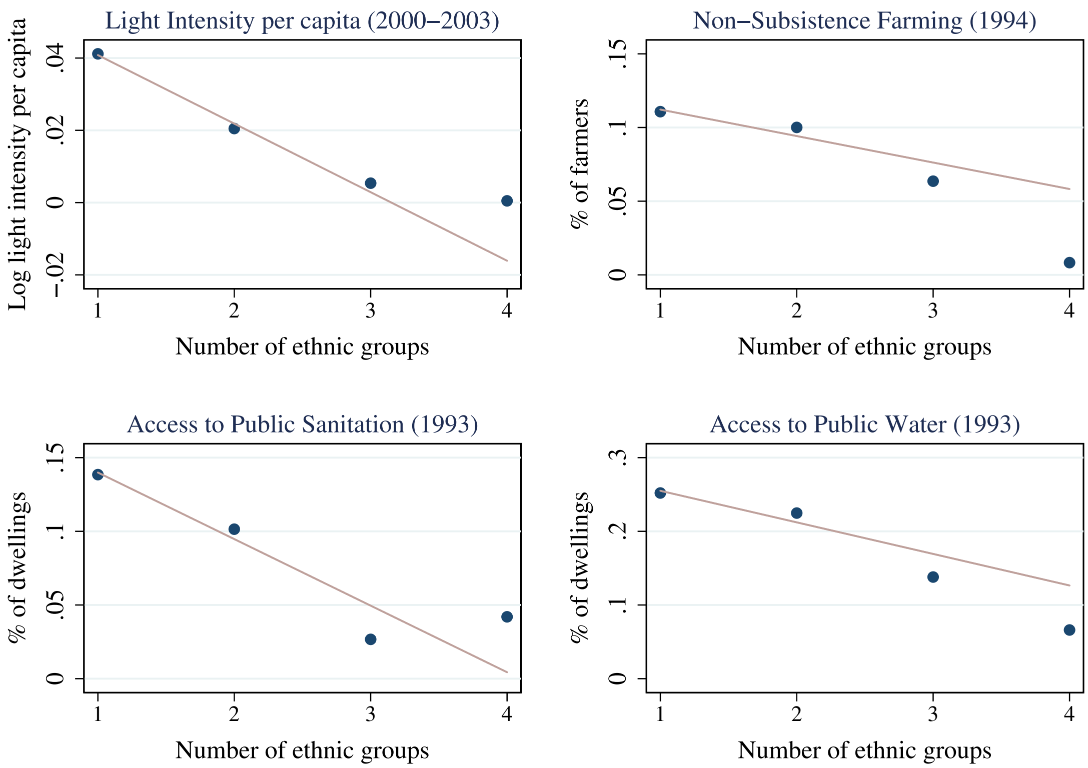


FIGURE A.3: Contemporary Living Standards and Number of Ethnic Groups

Notes. The figure uses parish-level data. Binned scatterplots without control variables. The x-axis refers to the number of ethnic groups concentrated in 16th-century parishes.

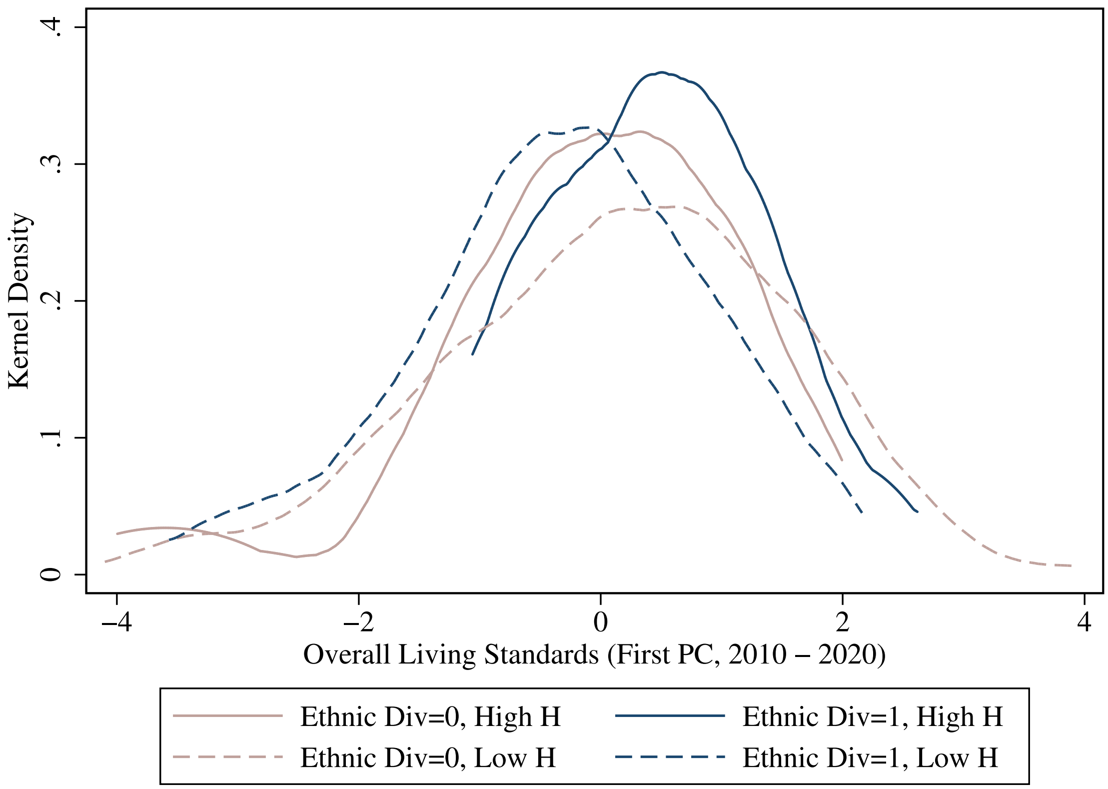


Figure A.4: Kernel Densities

Notes. The figure plots four distributions using parish-level data (Epanechnikov kernel function). High H refers to $\overline{H}_p$ above 0.830. Ethnic diversity takes value 1 if there is an ethnic border within a 10-km buffer from the parish capital, and 0 otherwise. The x-axis refers to the first principal component of the following four outcomes: the log average light intensity per capita (2010–2013), an indicator for nonsubsistence agriculture (2012, a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The first principal component accounts for 54.51 percent of the sample variance.

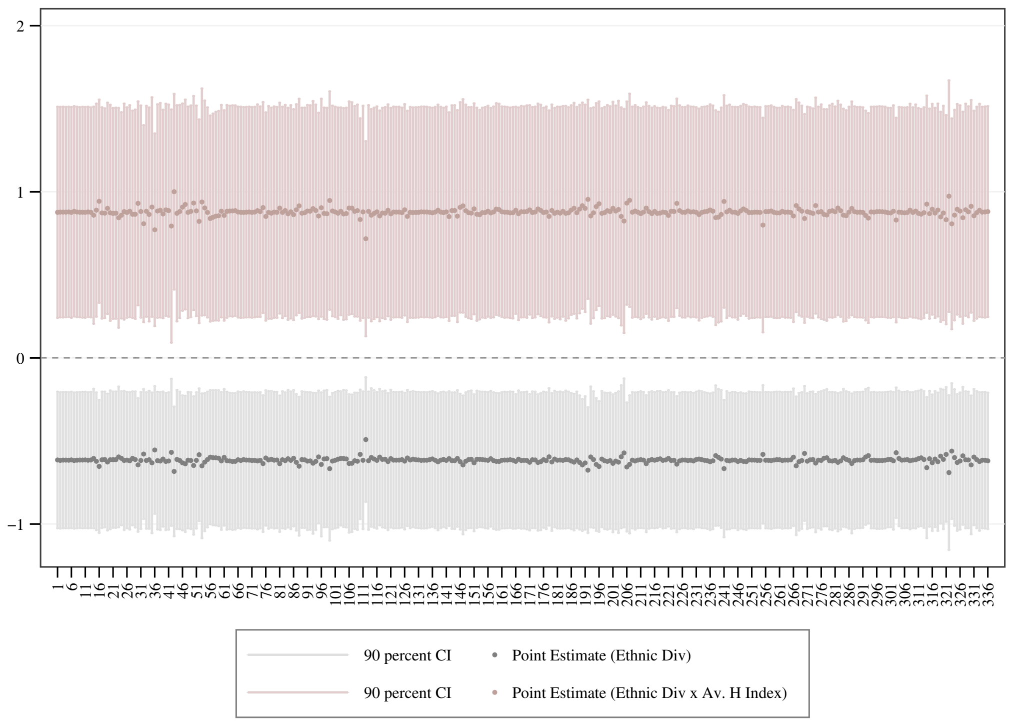


Figure A.5: Examining Influential Observations

Notes. Point estimates (standardized AES, Kling et al. 2004; Clingingsmith et al. 2009) and 90 percent confidence intervals for contemporary outcomes (2010–2020), after control variables and colonial province fixed effects (see Table 6). Each regression excludes one parish (indicated on the x-axis) at a time.

E Appendix - Tables

Table A.1: Frequency of Parishes by Ethnic Diversity and Resource Zone

	Yunga (500-2,300 m]	Quechua (2,300-3,500 m]	Suni or Jalca (3,500-4,000 m]	Puna (4,000-4,800 m]	Total
Ethnic div = 0	26	159	29	5	219
Ethnic div = 1	8	84	23	2	117
Total	34	243	52	7	336

Notes. The table reports the number of parishes by natural resource zone of the parish capital and ethnic diversity. Ethnic diversity takes value 1 if there is an ethnic border within a 10-km buffer from the parish capital, and 0 otherwise.

Table A.2: Validating Resource-Producing Zones − 50 km × 50 km Grid Cells

	Dependent Variable:								
	Ln (# Crops)						# Crops	Ln (# Crops)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
H index	0.593		0.619	0.532	0.458	0.458	0.516		
	[0.060]***		[0.078]***	[0.106]***	[0.088]***	[0.090]***	[0.104]***		
	(0.071)***		(0.114)***	(0.105)***	(0.110)***	(0.112)***	(0.134)***		
SD of elevation		0.469	-0.034	-0.014	0.048	0.047	0.033		0.103
		[0.063]***	[0.069]	[0.101]	[0.097]	[0.072]	[0.086]		[0.093]
		(0.042)***	(0.084)	(0.082)	(0.095)	(0.075)	(0.086)		(0.070)
Ln land area					0.343	0.343	0.360	0.378	0.371
					[0.055]***	[0.062]***	[0.073]***	[0.058]***	[0.054]***
					(0.055)***	(0.055)***	(0.069)***	(0.065)***	(0.058)***
Mean elevation						-0.003	-0.011	-0.216	-0.160
						[0.099]	[0.114]	[0.129]	[0.115]
						(0.067)	(0.082)	(0.100)**	(0.098)
Dummy (number of zones=2)								0.806	0.823
								[0.502]	[0.483]*
								(0.299)	(0.295)*
Dummy (number of zones=3)								1.743	1.697
								[0.489]***	[0.485]***
								(0.500)***	(0.517)***
Dummy (number of zones=4)								2.250	2.096
								[0.496]***	[0.555]***
								(0.500)***	(0.570)***
Hydrographic basin FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	148	148	148	148	148	148	148	148	148

Notes. The unit of observation is the 50 km × 50 km grid cell. The table reports OLS estimates. Robust standard errors in brackets; clustered at the FE level in Columns 4-9. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately two degrees at the equator (Colella et al. 2019). The dependent variable refers to the number of native crops. All variables except for dummies are standardized to have zero mean and standard deviation equal to one. All regressions control for longitude and latitude. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3: Validating Resource-Producing Zones − Ethnic Groups

	Dependent Variable:							
	Ln (# Crops)					# Crops	Ln (# Crops)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
H index	0.482		0.402	0.533	0.527	0.517		
	[0.130]***		[0.128]***	[0.112]***	[0.106]***	[0.103]***		
	(0.091)***		(0.079)***	(0.090)***	(0.086)***	(0.089)***		
SD of elevation		0.409	0.150	0.079	0.135	0.107		0.206
		[0.171]**	[0.167]	[0.133]	[0.141]	[0.137]		[0.168]
		(0.173)**	(0.153)	(0.144)	(0.157)	(0.148)		(0.149)
Ln land area				0.392	0.387	0.400	0.265	0.267
				[0.123]***	[0.126]***	[0.123]***	[0.128]**	[0.123]**
				(0.122)***	(0.124)***	(0.124)***	(0.162)	(0.154)*
Mean elevation					0.134	0.148	-0.206	-0.061
					[0.130]	[0.140]	[0.160]	[0.158]
					(0.106)	(0.114)	(0.170)	(0.172)
Dummy (number of zones=3)							2.178	1.772
							[0.519]***	[0.537]***
							(0.543)***	(0.620)***
Dummy (number of zones=4)							3.076	2.542
							[0.398]***	[0.493]***
							(0.418)***	(0.502)***
Observations	47	47	47	47	47	47	47	47

Notes. The unit of observation is the ethnic group. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). The dependent variable refers to the number of native crops. All variables except for dummies are standardized to have zero mean and standard deviation equal to one. All regressions control for longitude and latitude. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: Precolonial Correlates of H: Technology

	Dependent Variable:	
	Dummy Storage	Dummy Terraces
	(1)	(2)
H index	0.032	0.058
	[0.052]	[0.032]*
	(0.054)	(0.029)**
Observations	47	47

Notes. The unit of observation is the ethnic group. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). The H index is standardized to have zero mean and standard deviation equal to one. The dummy variables for food storage structures and terraces take value 1 for 14.89 and 10.64 percent of the groups, respectively. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Balance Tests−Additional Precolonial Characteristics

	Ethnic Diversity = 1		Ethnic Diversity = 0				
	mean	sd	mean	sd	Diff.	p-value[a]	p-value[b]
(1) Ln dist. to defensive site	4.186	0.705	4.161	0.927	-0.025	[0.783]	[0.856]
(2) Ln dist. to administrative site	3.840	1.008	3.828	1.208	-0.011	[0.927]	[0.815]
(3) Ln dist. to Inca road network	1.409	2.597	0.983	2.458	-0.427	[0.145]	[0.187]
(4) Caloric suitability for maize	357.869	1010.757	304.422	899.243	-53.447	[0.632]	[0.691]
(5) Caloric suitability for potato	595.073	755.261	623.936	800.354	28.863	[0.744]	[0.851]
Number of parishes	117	117	219	219	336	336	336

Notes. The unit of observation is the parish. Ethnic diversity takes value 1 if there is an ethnic border within a 10-km buffer from the parish capital, and 0 otherwise. P-values from OLS regressions of each of the variables listed in the first column on ethnic diversity; ([a]) with robust standard errors, ([b]) with standard errors corrected for spatial dependence using a distance cutoff of approximately one degree at the equator (Colella et al. 2019). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.6: H Index and Ethnic Diversity

	Dependent Variable:		
	Number of Parishes	% Parishes with Ethnic Div.	Mean Ln Dist. to Ethnic Border
	(1)	(2)	(3)
H index	0.154	-0.114	0.021
	[0.145]	[0.169]	[0.191]
	(0.174)	(0.165)	(0.262)
Observations	47	47	47

Notes. The unit of observation is the ethnic group. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). All variables are standardized to have zero mean and standard deviation equal to one. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.7: Average H Index and Ethnic Diversity

	Ethnic Diversity = 1		Ethnic Diversity = 0				
	mean	sd	mean	sd	Diff.	p-value[a]	p-value[b]
Average H index	0.674	0.173	0.652	0.179	-0.022	[0.265]	[0.324]
Number of parishes	117	117	219	219	336	336	336

Notes. The unit of observation is the parish. Ethnic diversity takes value 1 if there is an ethnic border within a 10-km buffer from the parish capital, and 0 otherwise. P-values from OLS regressions of average within-group heterogeneity on ethnic diversity; ([a]) with robust standard errors, ([b]) with standard errors corrected for spatial dependence using a distance cutoff of approximately one degree at the equator (Colella et al. 2019). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.8: Summary Statistics for Contemporary Outcomes

	Min	Mean	Median	Max	SD	# Parishes
Outcome variables (~ 1990-2000)						
Ln light intensity per capita (satellite F15: 2000-2003)	0	0.033	0.013	0.408	0.055	336
Share of farmers practicing nonsubsistence agriculture (1994)	0	0.105	0.031	0.796	0.158	336
Share of dwellings with access to public sanitation (1993)	0	0.122	0.036	0.805	0.169	336
Share of dwellings with access to public water (1993)	0	0.238	0.184	0.838	0.212	336
Outcome variables (~ 2010-2020)						
Ln light intensity per capita (satellite F18: 2010-2013)	0	0.056	0.030	0.603	0.086	336
Share of farmers practicing nonsubsistence agriculture (2012)	0	0.650	0.672	1	0.193	336
Share of dwellings with access to public sanitation (2017)	0.011	0.468	0.471	0.950	0.232	336
Share of dwellings with access to public water (2017)	0.011	0.763	0.819	0.997	0.196	336

Notes. The unit of observation is the parish. All data sources and definitions are reported in Appendix C.

Table A.9: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development−Individual Effects

	Dependent Variable:											
	Non-Subsistence Agriculture			Light Intensity per capita			Public Sanitation			Public Water		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Panel A: 1990 – 2000						
Ethnic diversity (dummy)	-0.333	-0.289	-0.298	-0.107	-0.083	-0.085	-0.094	-0.063	-0.103	-0.102	-0.041	-0.088
	[0.199]	[0.162]*	[0.164]*	[0.026]***	[0.026]***	[0.025]***	[0.066]	[0.052]	[0.042]**	[0.084]	[0.093]	[0.088]
	(0.185)*	(0.132)**	(0.134)**	(0.030)***	(0.028)***	(0.028)***	(0.055)*	(0.043)	(0.040)**	(0.060)*	(0.080)	(0.073)
Ethnic div. × Av. H index	0.494	0.511	0.523	0.128	0.098	0.102	0.067	0.061	0.126	0.090	0.051	0.129
	[0.304]	[0.265]*	[0.266]*	[0.034]***	[0.033]***	[0.032]***	[0.092]	[0.081]	[0.060]**	[0.126]	[0.143]	[0.127]
	(0.287)*	(0.220)**	(0.222)**	(0.036)***	(0.033)***	(0.034)***	(0.068)	(0.058)	(0.053)**	(0.092)	(0.122)	(0.105)
R-Squared	0.033	0.509	0.515	0.137	0.459	0.475	0.023	0.441	0.629	0.011	0.408	0.601
						Panel B: 2010 – 2020						
Ethnic diversity (dummy)	-0.322	-0.292	-0.311	-0.155	-0.108	-0.107	-0.159	-0.092	-0.094	-0.106	-0.085	-0.090
	[0.184]*	[0.158]*	[0.152]**	[0.036]***	[0.034]***	[0.033]***	[0.104]	[0.106]	[0.068]	[0.086]	[0.086]	[0.074]
	(0.126)**	(0.155)*	(0.152)**	(0.049)***	(0.035)***	(0.035)***	(0.094)*	(0.063)	(0.044)**	(0.083)	(0.058)	(0.046)**
Ethnic div. × Av. H index	0.455	0.596	0.629	0.183	0.121	0.119	0.183	0.127	0.125	0.133	0.119	0.123
	[0.289]	[0.242]**	[0.234]**	[0.046]***	[0.042]***	[0.040]***	[0.158]	[0.175]	[0.106]	[0.123]	[0.124]	[0.104]
	(0.217)**	(0.214)***	(0.212)***	(0.059)***	(0.040)***	(0.039)***	(0.107)*	(0.088)	(0.041)***	(0.105)	(0.097)	(0.070)*
R-Squared	0.042	0.550	0.558	0.117	0.415	0.419	0.011	0.322	0.560	0.059	0.250	0.362
Baseline controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Colonial province FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ln pop. den. and rural dummy	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Number of parishes	336	336	336	336	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). The dependent variables are: (1) the log average light intensity per capita (2000-2003 in Panel A and 2010-2013 in Panel B), (2) an indicator for nonsubsistence agriculture—a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median (1994 in Panel A and 2012 in Panel B), (3) the share of dwellings with access to public sanitation (1993 in Panel A and 2017 in Panel B), and (4) the share of dwellings with access to the public water network (1993 in Panel A and 2017 in Panel B). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. The rural dummy variable takes vale 1 if the share of rural population is above the median, and 0 otherwise (1993 in Panel A and 2017 in Panel B). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.10: Robustness: Ethnic $\text{frac}_p = 1 - \sum_e w_{pe}^2$

	AES (2010 – 2020)								
	Overall Living Standards			Local Econ. Activity			Public Facilities		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ethnic frac.	-1.583**	-1.167*	-1.392***	-2.319***	-2.135***	-2.203***	-0.848	-0.200	-0.582
	[0.634]	[0.639]	[0.510]	[0.641]	[0.674]	[0.655]	[0.887]	[0.974]	[0.745]
Ethnic frac. × Av. H index	1.964**	1.788*	2.157***	2.787***	3.098***	3.209***	1.141	0.477	1.104
	[0.881]	[0.929]	[0.714]	[0.911]	[0.939]	[0.911]	[1.241]	[1.456]	[1.054]
Baseline controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Colonial province FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ln pop. den. and rural dummy	No	No	Yes	No	No	Yes	No	No	Yes
Number of parishes	336	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009) across different outcomes. The AES for local economic activity refers to the log average light intensity per capita (2010–2013) and an indicator for nonsubsistence agriculture (2012, a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median). The AES for access to public facilities refers to the share of dwellings with access to public sanitation (2017) and the share of dwellings with access to the public water network (2017). The AES for overall living standards refers to the previous four variables. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. The rural dummy variable takes vale 1 if the share of rural population is above the median, and 0 otherwise (2017). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.11: Robustness: $\widetilde{H}_e = 1 - \sum_j s_{ej}^2$

	AES (2010 – 2020)								
	Overall Living Standards			Local Econ. Activity			Public Facilities		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ethnic diversity (dummy)	-0.955***	-0.668**	-0.696***	-1.170***	-1.031***	-1.050***	-0.739*	-0.304	-0.342
	[0.321]	[0.290]	[0.233]	[0.268]	[0.301]	[0.295]	[0.404]	[0.458]	[0.352]
Ethnic div × Av. $\widetilde{H}$ index	1.440***	1.140**	1.179***	1.770***	1.793***	1.828***	1.110	0.486	0.531
	[0.542]	[0.529]	[0.404]	[0.488]	[0.509]	[0.496]	[0.709]	[0.848]	[0.613]
Baseline controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Colonial province FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Ln pop. den. and rural dummy	No	No	Yes	No	No	Yes	No	No	Yes
Number of parishes	336	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009) across different outcomes. The AES for local economic activity refers to the log average light intensity per capita (2010–2013) and an indicator for nonsubsistence agriculture (2012, a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median). The AES for access to public facilities refers to the share of dwellings with access to public sanitation (2017) and the share of dwellings with access to the public water network (2017). The AES for overall living standards refers to the previous four variables. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. The rural dummy variable takes vale 1 if the share of rural population is above the median, and 0 otherwise (2017). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.12: Robustness: Additional Parish-Level Variables (1791-95)

	Overall Living Standards (AES, 2010 – 2020)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic diversity (dummy)	-0.616**	-0.611**	-0.680***	-0.617**	-0.575**	-0.662***
	[0.249]	[0.245]	[0.244]	[0.249]	[0.255]	[0.240]
Ethnic div. × Av. H index	0.879**	0.853**	0.951***	0.877**	0.841**	0.904***
	[0.386]	[0.379]	[0.366]	[0.385]	[0.386]	[0.349]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes
Ln indigenous population 1791-95	No	Yes	No	No	No	Yes
% mestizo population 1791-95	No	No	Yes	No	No	Yes
Ln priests per capita 1791-95	No	No	No	Yes	No	Yes
Religious order FE	No	No	No	No	Yes	Yes
Number of parishes	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja and Kremer 2009) across four outcomes: the log average light intensity per capita (2010–2013), an indicator for nonsubsistence agriculture (2012, a dummy variable for whether the share of farmers practicing nonsubsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.13: Mechanisms: Cultural Transmission – Inter-group Unions (1605-1780)

	Dependent Variable:										
	Average Normalized L Dist			Share of Unions with:				Dummy Normalized L Dist>0.6			
				Normalized L Dist>0.5		Normalized L Dist>0.6					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Average H index	0.731	0.585	-0.077	0.873	0.862	0.716	0.723	0.027	0.077	0.094	-0.010
	[0.231]**	[0.160]**	[0.134]	[0.132]***	[0.127]***	[0.068]***	[0.081]***	[0.009]**	[0.002]***	[0.024]***	[0.009]
	(0.155)***	(0.095)***	(0.106)	(0.053)***	(0.048)***	(0.047)***	(0.061)***	(0.009)***	(0.003)***	(0.022)***	(0.012)
Ln total individuals		0.451	0.099	-0.043	-0.012	0.266	0.245	0.006	0.067	0.070	-0.016
(1605-1780)		[0.202]*	[0.189]	[0.148]	[0.143]	[0.123]*	[0.159]	[0.010]	[0.003]***	[0.004]***	[0.008]**
		(0.124)***	(0.146)	(0.021)**	(0.050)	(0.076)***	(0.088)***	(0.007)	(0.002)***	(0.002)***	(0.005)***
Average % potential partners					-0.122		0.082				
(1605-1780)					[0.094]		[0.157]				
					(0.050)**		(0.103)				
% Potential partners								0.118	0.128	0.126	0.166
(1605-1780)								[0.007]***	[0.009]***	[0.008]***	[0.011]***
								(0.008)***	(0.008)***	(0.007)***	(0.012)***
Number of parishes	10	10	31	10	10	10	10	10	10	10	31
Number of individuals								3,124	3,124	3,124	14,287
Sample ethnic div = 1	✓	✓		✓	✓	✓	✓	✓	✓	✓	
Sample ethnic div = 0			✓								✓
Colonial province FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Year FE	No	No	No	No	No	No	No	No	No	Yes	Yes

Notes. The unit of observation is the parish in Columns 1-7 and the individual in Columns 8-11. The table reports OLS estimates. In brackets, the table displays robust standard errors (Columns 1-7) or standard errors clustered at the parish level (Columns 8-11). In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). The dependent variables refer to all unions between individuals with native paternal surnames during 1605–1780: the average normalized Levenshtein distance across unions (Columns 1-3), the share of unions with a normalized Levenshtein distance above 0.5 or 0.6 (Columns 4-7), and a dummy variable for whether the normalized Levenshtein distance is above 0.6 (Columns 8-11). All regressions control for the mean and standard deviation of elevation at the parish level. Potential partners are defined as those individuals, different from the true partner, with whom the individual has the same normalized Levenshtein distance. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.