

Within-Group Heterogeneity in a Multi-Ethnic Society*

Miriam Artiles[†]

October 10, 2020

JOB MARKET PAPER

[[Click here for the most updated version](#)]

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 development. I collect new data on a natural experiment from Peru’s colonial history: the forced resettlement of native populations in the 16th century. This intervention forced together various ethnic groups in new jurisdictions. Where these groups were composed of more heterogeneous subpopulations, working in different ecological zones of the Andes prior to colonization, ethnic diversity has systematically lower costs and may even become advantageous. Cultural transmission is one likely channel. Specifically, where different ethnic groups were composed of more heterogeneous subpopulations, they engage in more reciprocal behavior and exhibit more open attitudes toward out-group members.

JEL Classification: J15, N16, O10, O12, Q56, Z10

Keywords: Comparative Development, Ethnic Diversity, Within-Group Heterogeneity

*I am grateful to Marta Reynal-Querol and Luigi Pascali for their guidance and support throughout this project. I am also grateful to Joachim Voth and Gianmarco León-Ciliotta for their advice. I sincerely thank Jenny Guardado, Jonas Hjort, Miguel Ángel Carpio, Andreu Arenas, Fernando Fernández Bazán, Damián Romero, and seminar participants. I would also like to thank the Peruvian Ministerio del Ambiente and INEI for providing data.

[†]Universitat Pompeu Fabra and IPEG (e-mail: miriam.artiles@upf.edu; website: www.miriam-artiles.com).

1 Introduction

Some of the wealthiest economies in the world are ethnically diverse. Belgium, Canada, and Switzerland are examples of successful countries that are historically multi-ethnic (The Economist 2004). Other rich economies such as the United States and the United Kingdom attract immigrants from all over the world. At the same time, there are examples of ethnically diverse societies that are extremely poor, especially in Africa and Latin America.¹ Following initial work by Easterly and Levine (1997), a large body of literature has examined the costs and benefits of ethnic diversity (Alesina and La Ferrara 2005). If individuals within ethnic groups are homogeneous, and groups differ in preferences toward policies or public goods, then conflicting preferences can lead to inefficiencies in public good provision or to policy choices that may not benefit the entire society. Inter-group tensions can also result in civil conflicts or exacerbate mistrust and lack of cooperation. However, if ethnic groups differ in subsistence activities or skills, then complementary specializations can generate economic gains, stimulate innovation, and promote inter-group trade. 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 comparative economic development, and when is it bad?

In this paper, I argue that the effect of ethnic diversity on long-run comparative development depends on the heterogeneity of subpopulations within ethnic groups. Underlying previous literature is the assumption that individuals within ethnic groups tend to be homogeneous. However, ethnicities are not necessarily homogeneous entities. Subpopulations within ethnic groups may differ in many dimensions, including preferences, economic activities or skills, as well as cultural, genetic, and linguistic traits (Ashraf and Galor 2013; Desmet et al. 2017; Depetris-Chauvin and Özak 2020). I focus on the degree of heterogeneity in economic specializations within ethnic groups and ask whether having been part of a more heterogeneous ethnic group is a pre-condition that contributes to overcome the costs of ethnic diversity. I show that where ethnic groups are composed of more heterogeneous subpopulations, ethnic diversity has systematically lower costs and may even become advantageous.

I exploit a natural experiment of history: the 16th-century Spanish intervention in Peru. Two features of the study setting allow me to examine whether the consequences of ethnic diversity depend on previous exposure to within-group heterogeneity. The first is variation among the subpopulations of pre-colonial ethnic groups, which were engaged in a variety of specialized production niches *within* their respective ethnic groups. According to Andean ethnohistory, this variation emerged as a result of pre-colonial adaptation to the mountain environment—an environment characterized by multiple ecological zones created by differences in elevation (Murra 1975). The second feature is quasi-random variation in ethnic diversity across colonial

¹While some ethnically diverse countries in Africa have experienced progress, others continue to be afflicted by poverty and civil conflict (United Nations 2010). Social tensions are also present in Latin America, where colonization resulted in more complex societies along ethnic and racial lines (Fuenzalida et al. 1970).

jurisdictions. I analyze a forced resettlement of native populations that occurred after the Spanish conquest, which led to variation in ethnic diversity as a consequence of a mismatch between the pre-colonial strategy of adaptation and the Spanish notion of jurisdiction. Importantly, the degree of within-group heterogeneity to which native populations were exposed before colonization was inherent to reshuffled populations.

Are more heterogeneous ethnic groups better able to integrate in a multi-ethnic society? Although trust may be higher among members of the same ethnicity, individuals belonging to ethnic groups with more heterogeneous subpopulations are traditionally used to operate in diverse environments. They may also exhibit greater “openness to experience”, a personality trait defined as the 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.² In contemporary societies where multiple ethnicities coexist (e.g., as a consequence of voluntary migrations in an increasingly globalized world or due to events of forced displacement), understanding whether the consequences of ethnic diversity depend on previous exposure to within-group heterogeneity is important for policy discussions.³

The following quote by the anthropologist John Murra summarizes the subsistence strategy of pre-colonial Andean groups:

“In a territory so broken up by altitude, aridity, and brusque alternations ..., 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)

During the pre-colonial period, the settlement pattern of native groups was largely influenced by the mountain environment. Societal organization relied on simultaneous access to ecologically specialized production zones. Specifically, according to anthropological studies, the group guaranteed self-sufficiency by sending subpopulations to settle vertically distributed production zones, rather than relying on inter-group exchanges (Murra 1975).⁴ Subsistence was then

²For social psychology studies on openness to experience, see McCrae (1996), McCrae and Costa (1997), Flynn (2005) and Sibley and Duckitt (2008), among others.

³Hatton (2020) documents the global increase in both international migrants and refugees since 1960.

⁴Murra’s work describes Andean groups before the expansion of the Inca empire (see Section 2.1). Although information on their institutions is not abundant, to my knowledge there is no evidence on formal states or other institutions that could serve as a basis for sustained inter-group cooperation during the pre-Inca period. In my setting, group adaptation to the mountain environment resulted in within-group heterogeneity in specialized production zones. This pattern of specialization is consistent with findings that variation in geographic characteristics, such as regional land quality and elevation, may lead to specialization through the formation of region-specific human capital (Michalopoulos 2012). With respect to societal structure, a more tightly-knit extended family network is expected to strengthen within-group cooperation and to discourage interactions with out-group members. In turn, adverse environments may be related to stronger kinship networks (Enke 2019; Moscona et al. forthcoming).

achieved by exchanging resources between individuals from different ecologies within the same ethnic group.

This vertical and dispersed settlement pattern made tribute collection and religious indoctrination a difficult task after the Spanish conquest (1532). By the end of the 16th century, Spanish officials had consequently carried out a forced reorganization of native populations. The aim of the resettlement policy was to concentrate dispersed populations into small jurisdictions with a delimited and continuous space, which went against a key characteristic of the pre-colonial subsistence strategy: the exchange of resources between subpopulations settled at different elevations (Pease 1989). Colonial officials concentrated populations into such jurisdictions without an awareness of the spatial distribution of subpopulations within ethnic groups, that is, without considering that individuals from such a wide variety of elevations could belong to the same ethnic group. As a result of the tension 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, new jurisdictions did not respect initial ethnic divisions (Pease 1989, 1992; Wachtel 1976, 2002).

I use variation in the location of parishes (the jurisdictional unit of analysis) with respect to ethnic borders as a source of quasi-random variation in ethnic diversity. I start by defining ethnic diversity as an indicator for whether the parish capital was located close to an ethnic border. For this task, I use a map of the approximate extent of native groups at the moment of the Spanish conquest (Rowe 1946). I furthermore implement a novel approach to validate historical ethnic borders using information on surnames from colonial baptism records (1605-1780). In certain contexts, provided that surnames are inherited, measures based on surname commonality between individuals can provide information on common ancestry.⁵ In this setting, the pre-Hispanic practice of endogamy and the recent introduction of family names by the Catholic Church in the 16th century allow me to focus on the early common origin of native surnames as markers of common ancestry through the male line. After identifying native surnames based on their linguistic roots, the results for the subsample of parishes with available information suggest that, on average, surname heterogeneity among native populations was significantly higher in parishes located close to ethnic borders (*with ethnic diversity*), as compared to those located at the interior of ethnic homelands (*without ethnic diversity*).

To measure within-group heterogeneity, I use a diversity index that sums over production zones within the land inhabited by each ethnic group. Ecological or production zones are defined according to elevation intervals following a well-established classification for the region of interest (Pulgar Vidal 1941).⁶ In order to capture the average level of within-group heterogeneity to which the native populations in a parish were exposed before colonization, I then compute a weighted average of within-group heterogeneity for each parish. Specifically, I

⁵See Lasker (1980, 1985) and Colantonio et al. (2003) for a review on isonymy methods and the use of surnames in human population biology. See subsection 3.2 for further details.

⁶The homelands of most groups include all production zones, but in different proportions.

use the area share of each ethnic group in the parish as weight.

The empirical analysis intends to capture whether *forced* ethnic diversity has a differential effect on economic development depending on previous exposure to within-group heterogeneity. Results from balance tests show that local officials did not systematically concentrate populations of mixed ethnicity in locations that were inherently different in terms of geography or initial wealth. Furthermore, I document that the correlation between ethnic diversity and average within-group heterogeneity in specializations is not statistically significant, supporting that parishes located close to ethnic borders did not systematically concentrate populations from more heterogeneous ethnic groups.

The first result in the paper documents the direct effect of ethnic diversity, which I benchmark against previous results in the literature. I find that ethnic diversity is robustly associated with lower living standards in the long run. Specifically, I explore a variety of outcomes that capture contemporary living standards. As proxies for local economic activity, I use light intensity per capita (2000-2003) and a measure of non-subsistence agriculture from the agricultural census of 1994.⁷ On access to public infrastructure, I use data from the 1993 population census on access to public sanitation and to the public network of water supply. This result is in line with the literature on the costs of ethnic diversity, though it also highlights the persistent consequences of forced diversity at the local level.

I then examine the effect of ethnic diversity and average within-group heterogeneity. The results exhibit a robust pattern: the coefficient on ethnic diversity is negative, but its interaction with the average level of within-group heterogeneity in specializations is positive. I show that this pattern remains statistically significant when controlling for geographic and initial socioeconomic characteristics, as well as when accounting for the religious order in charge of the parish, colonial bishopric, and administrative province. Looking at the average effect size across contemporary outcomes (Kling et al. 2004; Clingingsmith et al. 2009), the results show that the costs of ethnic diversity tend to be overcome among populations from more heterogeneous ethnic groups. On average, after baseline controls and fixed effects, the negative association between ethnic diversity and contemporary living standards decreases –from -0.47 to -0.13 standard deviations– as average within-group heterogeneity reaches the median value and turns positive –from 0.21 to 0.32 standard deviations– for parishes above the 90th percentile.

I conduct a series of robustness checks to make sure that the previous pattern is not driven by omitted group characteristics or due to specific variable definitions. I first show that the observed pattern is robust to accounting for the interaction of ethnic diversity and the main correlates of within-group heterogeneity (i.e., variation in elevation and in pre-colonial caloric suitability within the ethnic homeland). Furthermore, the same pattern is observed when

⁷Since high quality data on income per capita is not available at the local level, I follow the empirical literature in using luminosity data from satellite images at night (Henderson et al. 2012; Hodler and Raschky 2014; Michalopoulos and Papaioannou 2013, 2014).

accounting for the interaction effect of ethnic diversity and other characteristics of ethnic groups, such as surface area and approximate population density at the time of the Spanish conquest. I then check that the results exhibit the same pattern when controlling for an extended set of (uninteracted) ethnic characteristics, as well as when including fixed effects for the majority ethnic group in the parish and when excluding the largest groups in terms of surface area from the analysis. Additional results show that the same pattern arises if I use a robust version of the ethnic diversity dummy variable that accounts for imprecise ethnic borders, or if I use a standard measure of ethnic diversity based on the Herfindahl index. Using an alternative diversity index to measure within-group heterogeneity in specializations yields the same pattern. Moreover, estimates from a placebo analysis using borders of colonial provinces instead of ethnic borders are small and not statistically significant. Overall, these results support that, beyond the direct effect of geography, having belonged to an ethnic group with more heterogeneous subpopulations plays an important role in explaining the long-term effects of ethnic diversity.

To understand the evolution of these long-term effects, I use data from the 1876 population census. I show that the documented pattern was accompanied by structural change and improved literacy rates. While the estimated coefficient on ethnic diversity is negative, its interaction with the average level of within-group heterogeneity is again positive. I find that parishes built on ethnically diverse populations, as compared to those built on a single ethnic group, tend to be more oriented toward manufacturing, retail, and services, to the detriment of the agricultural sector, as average within-group heterogeneity increases. By 1876, ethnically diverse parishes were also associated with better literacy rates, relative to those without ethnic diversity, the higher the average within-group heterogeneity among their ancestors.

Finally, I explore the mechanism underlying these findings. Although it is difficult to identify the precise mechanism at work, the historical literature suggests that pre-colonial exchanges were sustained by engaging in reciprocities between individuals of the same ethnic group. One possibility is that the subsistence strategy of Andean groups, which relied on within-group interactions between individuals with different specializations, shaped individual attitudes toward diverse people and ideas. The transmission of reciprocal behavior and more open attitudes toward out-group members may have favored inter-group contact and beneficial interplays between individuals of different ethnic origin during the colonial period. To examine cultural transmission, I start by looking at contemporary neighborhood and agricultural associations. Previous literature has shown that ethnic diversity is associated with lower social engagement (e.g., Alesina and La Ferrara 2000). Consistent with the transmission of cultural traits, I find that this negative association decreases the higher the average within-group heterogeneity to which native populations were exposed. I then provide evidence supporting that exposure to within-group heterogeneity favored inter-group contact and societal integration during Spanish rule. In particular, I use information on the parents of baptized individuals and focus on inter-ethnic unions as a proxy for integration.⁸ Lastly,

⁸The use of inter-ethnic marriages as a proxy for integration is well established in sociology (Gordon 1964).

I explore economic complementarities between ethnic groups. In line with recent papers emphasizing the positive role of inter-group complementarities for the integration of minority groups (Jha 2013; Becker and Pascali 2019), the results suggest that an economic advantage emerges where a majority with low within-group heterogeneity enjoys any complementary skill arising from a minority that comes from a highly heterogeneous ethnic group.

To my knowledge, this is the first paper to explore the effects of ethnic diversity in a setting with variation in within-group heterogeneity. By doing so, this study contributes to a large literature on the consequences of ethnic diversity. Studies in this literature have been conducted at different levels of analysis, obtaining mixed results while implicitly assuming that individuals within ethnic groups tend to be homogeneous. Across countries and US localities, ethnic heterogeneity has been associated with lower levels of economic growth, public good provision, quality of government, and social capital, as well as with higher 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. Focusing instead on the private sector, Hjort (2014) provides causal evidence for the effect of ethnic diversity on team productivity at a Kenyan flower plant, showing that teams of ethnically diverse workers are, on average, less productive than homogeneous teams. The underlying mechanism appears to be a taste for discrimination against co-workers of different ethnic origin.

More recently, Montalvo and Reynal-Querol (forthcoming) have focused on the size of the unit of analysis, finding that ethnic diversity has a positive effect on economic growth if we look at small units. They argue that a potential explanation in the case of Africa is the increase in trade close to ethnic borders, suggesting ethnic specialization into complementary activities. This result links with studies on the positive role of inter-group complementarities. The theoretical framework in Jha (2013, 2018) establishes that peaceful inter-ethnic coexistence can be sustained through the specialization of ethnic groups into complementary activities that are costly to replicate and to expropriate. Becker and Pascali (2019) provide empirical evidence in the context of anti-Semitism in Germany.¹⁰ In sum, while previous studies have found a variety of results, this paper provides evidence that the medium- and long-term effects of ethnic diversity depend on previous exposure to within-group heterogeneity.

The diversity of subpopulations within societies or ethnic groups has received little attention in the literature. Ashraf and Galor (2013), the first paper to consider interpersonal diversity *within* populations, show that the degree of genetic diversity within societies has influenced their comparative development in both pre-colonial and modern times. Desmet et al. (2017) focus on the degree of cultural diversity and document that most of the diversity in contemporary norms and attitudes takes place between individuals of the same ethnolinguistic

For recent applications in economics see, for instance, Bazzi et al. (2019).

⁹See Easterly and Levine (1997), La Porta et al. (1999), Alesina et al. (1999), Alesina and La Ferrara (2000), Alesina et al. (2003), Alesina and La Ferrara (2005), Fearon and Laitin (2003), Montalvo and Reynal-Querol (2005), Desmet et al. (2009), and Desmet et al. (2012), among many others.

¹⁰In the context of firms, see Lazear (1999) on the positive role of skill complementarities.

group rather than between groups. Depetris-Chauvin and Özak (2020) find that diversity in genetic and linguistic traits within pre-modern societies was conducive to labor division and specialization. I focus on the degree of heterogeneity in specialized production zones within pre-colonial ethnic groups and find that exposure to within-group heterogeneity is a pre-condition that contributes to overcome the negative effects of ethnic diversity.¹¹

The cultural transmission mechanism adds to the literature on the long-term effects of cultural traits (Nunn and Wantchekon 2011; Voigtländer and Voth 2012; Alesina et al. 2013; Guiso et al. 2016, among others). Specifically, the results in this paper support that within-group heterogeneity in specializations, which emerged as a result of native adaptation to the mountain environment, favored the formation of reciprocal behavior and more open attitudes toward out-group members. This in turn contributed to a greater ability to engage in multi-ethnic societies. I provide evidence consistent with the idea that having belonged to an ethnic group with more heterogeneous subpopulations facilitated inter-group contact during the colonial period and, over the course of generations, contributed to sustain long-run performance.¹²

The remainder of the paper is organized as follows: Section 2 provides a summary of the pre-colonial setting and the Spanish intervention, Section 3 explains the data construction process, Section 4 describes the empirical strategy, Section 5 presents the results, and Section 6 concludes.

2 Historical Background

2.1 Pre-Colonial Settlement Pattern

The *Chocorvos*, the *Lucanas*, the *Soras*, the *Chancas*, the *Quichuas*, the *Caviñas*, the *Chilques* and the *Aymaraes*, among others, were native groups under Inca rule at the time of the Spanish conquest (Tello 1939; Rowe 1946).¹³ Specifically, by the time the Spanish arrived (1532), the Andean civilization comprised several coexisting groups that had been incorporated

¹¹While within-group heterogeneity has been less explored, the literature has studied other characteristics of pre-industrial societies or ethnic groups. This literature has mainly focused on Murdock (1967)'s *Ethnographic Atlas*. For example, the hierarchical structure of ethnic institutions, as measured by the number of jurisdictional levels beyond the local community, has received increasing attention (Gennaioli and Rainer 2007; Michalopoulos and Papaioannou 2013). See Alesina et al. (2013) for the practice of plough agriculture and Bentzen et al. (2017) for rules of succession to the office of local headman, among others.

¹²Other papers have emphasized the role of climate and geography in shaping culture. For example, Buggle and Durante (2017) find that European regions exposed to higher environmental risk during the pre-industrial era exhibit higher levels of interpersonal trust today. The study argues that, in face of variability in precipitation and temperature, farmers developed cooperative strategies that contributed to the emergence of more trusting attitudes. Nunn and Puga (2012) document the indirect positive effect of ruggedness 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).

¹³I interchangeably refer to pre-colonial native groups as “tribes” or “ethnic groups”.

during the previous century to the Inca empire. The archaeologist and anthropologist John Rowe maps the approximate extent of these groups by 1530 (Rowe 1946).

The exact stage of development as political units is unclear for most groups. However, there were no formal states before the expansion of the Inca empire. Based on early chronicles, the social unit is generally described as an endogamous group of several extended families with descent through the male line. Studies suggest that leadership was sometimes based on personal prestige, while in other cases it was inherited. The historical and anthropological literature also recognizes language differences, although these disparities tended to be homogenized with the spread of Quechua, the language of the Incas.¹⁴

In the Andean highlands, differences in elevation give rise to a variety of vertically arranged ecological zones. Within a short distance, diverse environmental and soil conditions create specific production niches. The pioneering ethnographic and ethnohistorical work of Murra (1975, 1995, 2002) documents how pre-Inca Andean peoples managed to overcome the complexities of the mountain environment. Murra's model describes the settlement pattern of a given ethnic group as vertical and dispersed (a "vertical archipelago"), where subsistence was based on simultaneous access to ecologically specialized production zones. Specifically, the group guaranteed self-sufficiency by sending subpopulations to settle vertically distributed production zones. These settlements were usually located at a three or four days' walk from the main settlement of the group, around which the system was organized and where individuals maintained ties to their extended family and homeland.¹⁵ Since each zone is characterized by a particular microclimate, they are respectively suited to a different assortment of crops. Rather than relying on inter-group exchanges, ethnohistoric research suggests that group subsistence was achieved by exchanging resources between subpopulations spread across specialized production zones.¹⁶

According to Murra's and subsequent research, this adaptive strategy was already in place during the pre-Inca period. The Inca expansion (1438-1525) was achieved through the gradual incorporation of groups to the empire, which led to the creation of provinces based on ethnic distinctions (Rowe 1946). Inca government was indirect in the sense that provinces were governed by the ethnic rulers of the corresponding groups. This is a key characteristic of Inca rule because it supports that ethnic traits were preserved during this period (Murra 1975, 2002). At the same time, provincial rulers were pushed to continue with their control of different vertical zones in order to sustain the empire (Murra 2002).

¹⁴See the first book by Garcilaso de la Vega (1960)[1609], Rowe (1946) and Murra (1975).

¹⁵The analysis in Pease (1989) for the *Collaguas* native group also suggests that populations located at lower elevations were not excluded from contact with higher elevations.

¹⁶This subsistence strategy is supported for the Andean range of Peru, and especially for the central and southern Andes. However, it is unclear how the model applies in relation to coastal peoples (Murra 1975; Pease 1989). In this paper, I move beyond specific case-study evidence and perform a systematic analysis of Andean groups regarding the extent to which they relied on diverse production zones.

2.2 The Spanish Intervention

The contemporary administrative division of Peru has its origin in the early colonial period. When Viceroy Francisco de Toledo (1569-1581) first disembarked in Peru, native populations followed the Andean pattern, living scattered along mountain slopes.¹⁷ This dispersed settlement pattern was an obstacle for the Spanish administration. In words of the 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” (Medina 1974a, p. 155). In order to facilitate tribute collection and religious indoctrination, Toledo ordered the forced reorganization of native populations into residential (*reducciones*) and religious (*doctrinas*) jurisdictions.

The idea of the *reducciones* (also known as *pueblos de indios*) was to concentrate native populations into small villages with a delimited and continuous space (Medina 1974a,b, 1993). The model of the village was originally designed by de Matienzo as a grid system, interconnected by straight-line streets that formed quadrilaterals and were organized around a main square with a church (see Appendix A). Colonial officials carried out the resettlement policy between 1570 and 1575, arranging the division of native populations from all discovered lands in the Viceroyalty of Peru into such *pueblos de indios*. In turn, several *pueblos de indios* were under the jurisdiction of a single *doctrina* or *parroquia*, a parish served either by the regular or the secular clergy.¹⁸

Within four decades of the conquest of the Inca empire, the Spanish administration had undertaken a complete reorganization of native populations. There is little documentation on how the resettlement policy was carried out. However, the historical literature does note that colonial officials, who were not fully aware of the spatial distribution of subpopulations, concentrated populations without considering that individuals so discontinuously scattered across elevations could belong to the same ethnic group. There was a tension at the moment 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. As a result, colonial jurisdictions did not respect pre-existing ethnic divisions (Pease 1989, 1992; Wachtel 1976, 2002).

It is important to note that the new model limited the movement of native populations, pointing against a key characteristic of the pre-Hispanic subsistence strategy: the exchange of resources

¹⁷At the time, most native populations were under the *encomienda*, a Spanish labor system that rewarded conquerors (*encomendero*) with the services of a particular number of natives. However, the *encomienda* did not imply a title over land. The *encomendero* was only entitled to a share of the product of native labor (Dougnaç Rodríguez 1994, ch. 9). The historical literature suggests that there was no conflict with pre-Hispanic ethnic divisions under this system because the *encomienda* was not based on territory but on population (Pease 1992, p. 180; Murra 2002, p. 62).

¹⁸The 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 religious order were also present, see de Armas Medina (1953).

between subpopulations settled at different elevation zones (Pease 1989). The intention of the resettlement policy was not to create sustainable jurisdictions that maximized access to resources from different ecological zones, but rather to concentrate dispersed populations in a way more consistent with the Spanish view of the world. Importantly, historical studies widely discuss that, in practice, the limitation of movement was effective at the parish level (Saignes 1991; Medina 1974a,b, 1993). Indeed, this structure was maintained over 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 in the country.¹⁹

3 Data

3.1 Explanatory Variables

Unit of analysis. I focus 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 during 1791-95 under the administration of Viceroy Gil de Taboada y Lemos lists all parishes created in this territory. 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*.²⁰

I assign longitude and latitude coordinates to each parish capital. For this, it is important to note that following independence from Spanish rule, provinces were the basis for the second administrative level in the country, and parishes were transformed into districts (third administrative level). I start by matching each parish to a modern district using the name and year of creation of the district, as well as the province to which the latter belonged. Then, I 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), I assign the coordinates corresponding to the parish capital.

Using a combination of historical sources, I check for the presence of priests in charge of religious indoctrination in each parish. In many cases, it is possible to know the names of

¹⁹For details on the correspondence between parishes and 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*.

²⁰The census excludes parishes from the *intendencia* of Puno because it 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 on the case of Puno. 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, Ap., 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 the Census of 1792. It is considered a baseline for the study of population figures just before independence from Spain (Gootenberg 1991).

such priests and whether they were part of the regular or the secular clergy.²¹ Following the historical literature on Murra’s model, the analysis focuses on parish capitals located in the highland region.²² The analysis also excludes the two capital parishes of Cuzco and Arequipa, as well as six parishes that are now part of Chile. The final sample includes 336 parishes (see panel (a) of Figure 1).

Ethnic diversity. The analysis aims to capture the ethnic origin of the populations that were forced to reside within the parish jurisdiction in the 16th century. Colonial descriptions of priests’ walks from the parish capital commonly lie between 2 and 3 *leguas*.²³ Following Paz Soldán (1877), the colonial measure of 1 *legua* corresponded to approximately 3,340 meters during the initial colonial period. I therefore construct a buffer of 10km radius (approximately 3 *leguas*) around each parish capital. When the distance between capitals is less than 10km, I use equidistant boundaries to ensure that buffer polygons do not overlap each other. The mean buffer area is 240km².

To measure ethnic diversity, I create a dummy variable that takes value 1 if there is an ethnic border within the 10km buffer from the parish capital, and 0 otherwise.²⁴ Figure 2 illustrates the exercise: parishes with ethnic diversity (in yellow) are those located close to ethnic borders, while those located further inside ethnic homelands are parishes without ethnic diversity (in blue). Following this definition, approximately 35 percent of parishes in the sample have ethnic diversity, while the remaining 65 percent have only one ethnicity within the 10km buffer. Panel (a) of Figure 1 shows the spatial variation in ethnic diversity. Importantly, the construction of this variable relies on the map of native groups at the time of the Spanish conquest (Rowe 1946). Section 3.2 validates ethnic borders using information on baptism records for the colonial period.

Within-group heterogeneity. Within-group heterogeneity is captured by diversity in ecologically specialized production zones. Ecological or production zones are defined according to elevation intervals following a standard classification for the area of interest (Pulgar Vidal 1941): 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 refer to elevation in meters above the sea level.²⁵ There are 49 ethnic groups in the analysis. Panel (b) of Figure 1 displays the

²¹See Lissón Chávez (1943), de Armas Medina (1953), Córdoba y Salinas (1957)[1651] and García (1997).

²²Excluding coastal parishes (0-500 meters above the sea level) at the same time alleviates concerns regarding potential population resettlements from the north to the south coast of Peru during the Inca period (Bongers et al. 2020). See also robustness checks related with the Inca period (Section 5.3).

²³See, for example, *Relaciones Geográficas de Indias*, Vol 1., compiled by Marcos Jiménez de la Espada, Madrid: Ministerio de Fomento.

²⁴A tribe is counted as part of the 10km buffer only if the area share of the tribe within the buffer is at least 1 percent. This accounts for lack of precision in ethnic borders, ensuring that there is at least one grid cell of 1km × 1km with centroid inside the area of the tribe in the buffer.

²⁵There are different approaches to study the territory of highland Peru. The traditional classification by Pulgar Vidal is preferred in this application because it was developed by taking into account local geographical

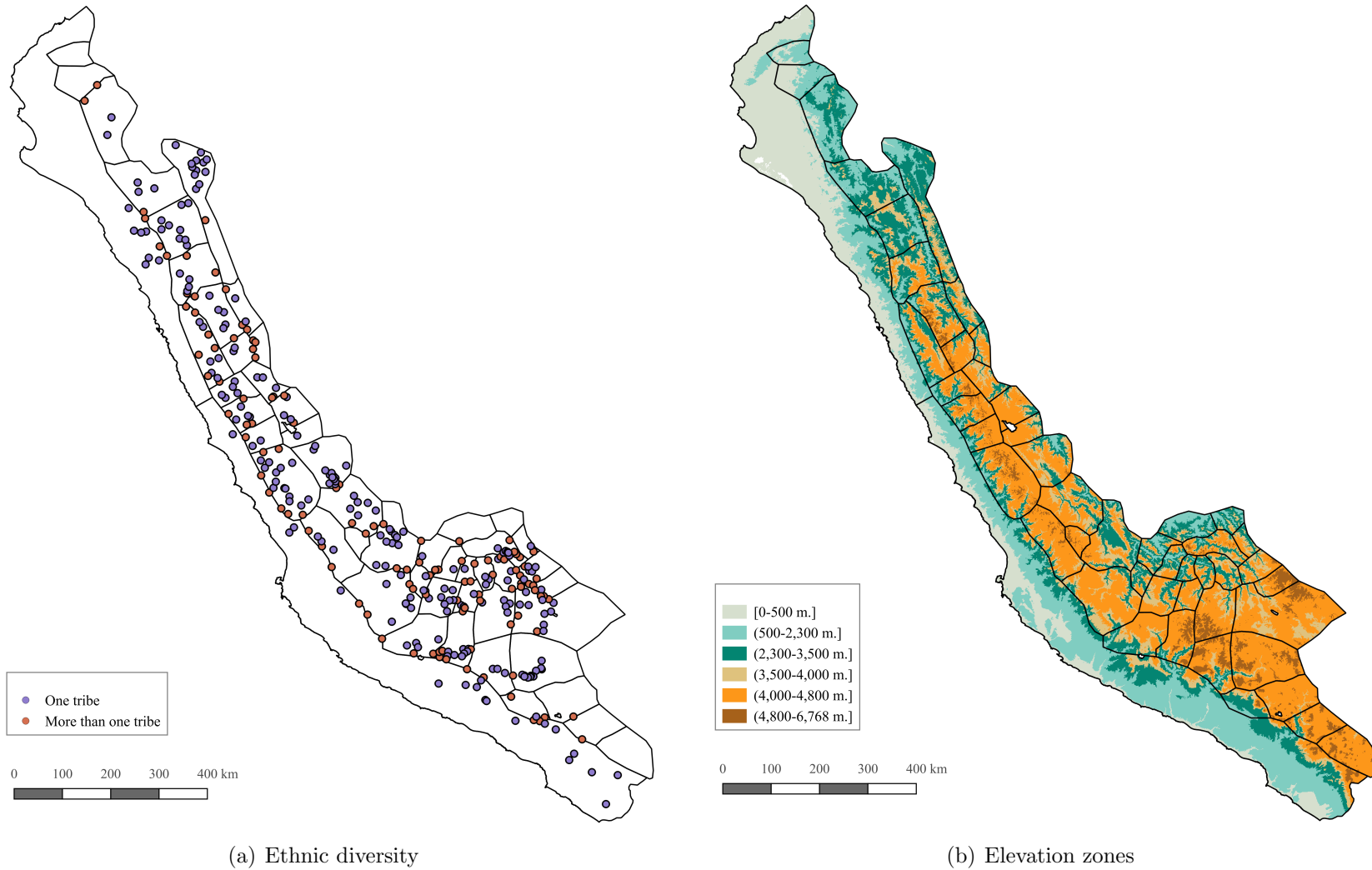


FIGURE 1

Notes. Polygon borders in black represent the approximate extent of tribes under the Inca empire at the time of the Spanish conquest (Rowe 1946). In panel (a), dots represent parish capitals; parishes with more than one tribe are those with an ethnic border within a 10km buffer from the parish capital. Panel (b) displays the different elevation zones within the territory of each tribe following the classification in Pulgar Vidal (1941). 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 (in meters) constructed based on information from the NASA Shuttle Radar Topographic Mission (SRTM).

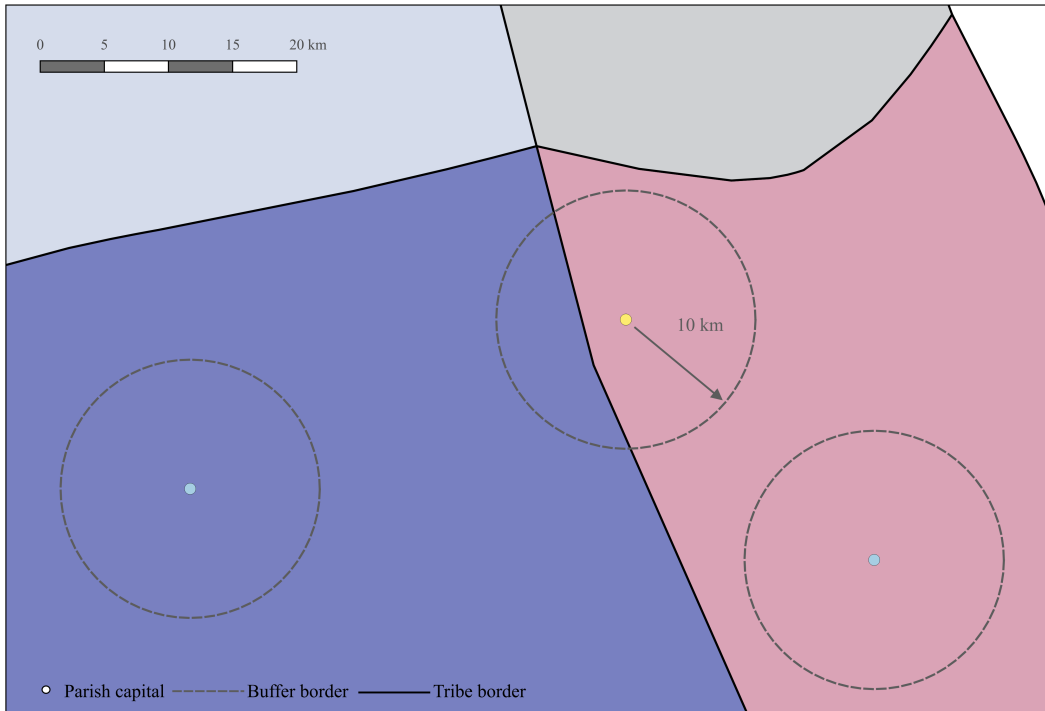


FIGURE 2

Notes. The figure illustrates the construction of a 10km buffer around each parish capital.

composition of elevation zones within each ethnic polygon. The homelands of most groups (63 percent) include all elevation zones, but in different proportions, see Appendix A.²⁶

I use a diversity index that sums over elevation zones. The reciprocal of the Simpson or Herfindahl index is a common diversity measure in ecology (Magurran 2004). It has also been used in urban studies to measure diversity in sectors of economic activity (Duranton and Puga 2000). It takes the following form:

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

where s_{ej} is the area share of elevation zone j within the homeland of ethnic group e . The index increases as the composition of elevation zones within an ethnic group's territory becomes more diverse. I have normalized the index to take value 1 for the group with the highest diversity at the end of the pre-colonial period. Elevation data come from version 1.2 of the Harmonized World Soil Database (FAO), which provides 30 arc-second raster data (grid cells of approximately $1km \times 1km$ at the equator) with median elevation based on information from the NASA Shuttle Radar Topographic Mission (SRTM). After classifying each grid cell knowledge, including native folklore. See Pulgar Vidal (2012, p. 29) and Tapia (2013).

²⁶Table A.1 shows the number of parishes by elevation zone of the parish capital. The proportion of parishes with ethnic diversity ranges from 31 percent in the Yunga region to 79 percent in the Suni region.

with centroid within the ethnic polygon into an elevation zone, I can compute the area share of each zone j within the homeland of that ethnicity. Figure A.1 shows the density of the normalized index at the ethnic group level. Approximately 20 percent of the groups have an index value below 0.5, while the index for the remaining 80 percent ranges from 0.5 to 1, with similar mean (0.666) and median (0.676) values.²⁷

Finally, for each parish, I compute the weighted average of within-group heterogeneity within the $10km$ buffer as:

$$\bar{H}_p = \sum_e w_{pe} H_e$$

where w_{pe} is the area share of ethnic group e within the $10km$ buffer from the parish capital (p). It aims to capture the average level of within-group heterogeneity to which native populations in a parish were exposed before colonization. The mean value of the average level of within-group heterogeneity in the sample of 336 parishes is 0.674.

3.2 Validating Ethnic Borders

Should we trust ethnic borders? Using data from baptism records (1605-1780), I explore whether surname heterogeneity among native populations was significantly higher in parishes with ethnic diversity, as captured by the spatial analysis, compared to parishes with only one ethnicity within the $10km$ buffer.

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

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

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, $D \in [0, 1]$, relies on the Simpson or Herfindahl index. As long as any two individuals with the same surname inherited it from a common ancestor, the index can be interpreted as the probability that two individuals taken at random from the population have different ancestry. The second index, $S \in [0, \ln(K)]$, is known as the Shannon index and takes its theoretical basis from information theory (Shannon

²⁷Using information on the area planted with native crops from the 2012 national agricultural census, Table A.2 shows that the measure of within-group heterogeneity is positively and significantly correlated with crop diversity.

1948).²⁸ In the context of surnames, S can be interpreted as the average uncertainty in predicting ancestry. The idea is that 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 lower uncertainty in predicting ancestry.²⁹

Introduction of surnames in Peru. Isonymy methods make a strong assumption (i.e., surname commonality directly translates into common ancestry). This assumption does not hold in contexts where one surname can have multiple origins (e.g., non-related individuals with a common surname due to their ancestors sharing the same occupation) or in contexts where surname changes are permitted for non-genetic reasons (e.g., illegitimacy or adoption). Are isonymy methods, therefore, appropriate for this application?

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, 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). 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 this approach may be limited by the adoption of Hispanic surnames over time, qualitative evidence suggests that the common practice was for priests to choose a Hispanic first name, with the native first names of the parents adopted as surnames (RENIEC 2012).³⁰ Garcilaso de la Vega (1960)[1609] also suggests that surnames adopted by native populations were initially related to their ethnic origin.³¹ Given the practice of endogamy and the short history of family names, this application focuses on the early common origin of surnames representing common ancestry through the male line.

²⁸The Shannon index is an entropy measure that has also been applied to genetic diversity (Lewontin 1972) and to species diversity (Magurran 2004).

²⁹Appendix B plots each index for different numbers of groups of equal size. Both indices grow with the number of groups, but the S index grows faster than the D index.

³⁰Each individual inherits two surnames in the Hispanic system of family names: the first surname corresponds to the first surname of the father, while the second surname is the first surname of the mother.

³¹See Carpio and Guerrero (2020) for further details on the introduction of surnames in Peru.

Surnames from baptism records. The website FamilySearch.org contains baptism records for colonial Peru. The organization, which seeks to help trace users' ancestry, seeks volunteers from around the world to make indexed genealogical records freely available. For the purpose of this analysis, each baptism record includes information on three key characteristics of the individual: full name, name of the parish, and date of baptism.³² The original handwritten record has also been uploaded in some cases and can be easily accessed.

I created a dataset of 112,340 individuals with native first surname covering the period 1605-1780.³³ The dataset includes information for 66 parishes, of which 20 percent are parishes with ethnic diversity according to the spatial exercise. To identify native surnames, I constructed a dictionary of native linguistic roots and looked for the occurrence of these roots within surnames; see Appendix B for further details. It is important to note that, since not all records have survived, the number of parishes with information varies by year. Furthermore, the number of individuals with surname information in each parish depends on recovered historical records, and thus the results should be interpreted with caution. Appendix B reports the total number of individuals, the number of parishes with information on baptized individuals, and other descriptive statistics for disaggregated time periods. The mean parish in the dataset of baptisms comprises 1,702 individuals with native first surname, of which 846 are males, relative to a sample mean of 1,603 individuals per parish according to the census of 1792 (of which 758 are males). The number of individuals with native first surname per parish has a right-skewed distribution, with a median of 562 individuals. Data from the census of 1792 are less skewed, with a median of 945 individuals.

Results. Table 1 presents the results from regressing surname diversity measures on the ethnic diversity dummy variable. In each column, the dependent variable is either the S index or the D index, constructed using individuals with native first surname. All variables except the indicator for ethnic diversity are standardized to have mean 0 and standard deviation 1. The results for the subsample of parishes with available information suggest that, on average, parishes located close to ethnic borders exhibited higher surname diversity among native populations (between 0.42 and 0.55 standard deviations), compared to parishes located at the interior of ethnic homelands.

Panel A shows baseline results. For each surname diversity index, the first column shows the unconditional correlation. The second column adds the log total number of individuals found in the records of the parish and the share of individuals with non-native first surname as control variables. The third column includes geographic controls and log distance to the closest mining center during the colonial period. The last column includes ecclesiastical jurisdiction fixed effects, accounting for potential differences in the administration of baptism across colonial bishoprics. The coefficient on the ethnic diversity dummy variable is always

³²As an example, Appendix B displays a baptism record from the website.

³³The records also provide information on gender: 50.32 percent of individuals with native first surname are females, while the remaining 49.68 percent are males.

positive and statistically significant. Panel B shows that the results are similar after dropping individuals whose surnames only occur once in the dataset.³⁴ Finally, Panel C shows that the results are robust to using groups of similar surnames (instead of raw surnames) in order to compute surname diversity indices. This approach takes into account potential changes in the writing of surnames over time.³⁵

3.3 Contemporary Outcomes

I use several outcomes that capture contemporary living standards. Specifically, I look at local economic activity and access to public infrastructure. Table A.1 reports summary statistics for the full sample and separately by parishes with and without ethnic diversity.

Since high quality data on income per capita is not available at the local level, I follow the empirical literature in using luminosity data from satellite images at night as a proxy for local economic activity (Michalopoulos and Papaioannou 2018). The NOAA’s National Geophysical Data Center provides yearly raster data on the average intensity of nighttime lights at a resolution of 30 arc-seconds.³⁶ I compute average light intensity per capita for the period 2000-2003 using luminosity data from satellite F15.³⁷ Specifically, the sum of light intensity across all grid cells with centroid within the 10km buffer is divided by the total population within the same buffer. Population data for the year 2000 come from version 4.10 of the Gridded Population of the World (CIESIN). These data are also mapped at a 30 arc-second resolution with population counts based on census data and adjusted to match the United Nation’s estimated population counts at the country level. Following the empirical literature, I log transform the measure of light intensity per capita and add 0.01 before computing the logarithm (Henderson et al. 2012; Hodler and Raschky 2014; Michalopoulos and Papaioannou 2013, 2014).

A second proxy for local economic activity is non-subsistence agriculture. Subsistence farming has traditionally been a widespread practice in the Andean highlands. The 1994 national agricultural census provides district-level data on the number of agricultural producers that devoted most of the harvest to sale or trade in local markets rather than to own consumption. I assign the data of the corresponding contemporary district to each parish. On average, 75 percent of agricultural producers in a district reserved most of their harvests for their own consumption. Using this information, I construct an indicator for non-subsistence agriculture

³⁴Doing so decreases the sample of individuals with native first surname from 112,343 to 106,124 individuals.

³⁵Specifically, I link surnames if only one change (deletion, insertion, or substitution) is required to transform one surname into the other. This results in a decrease in the total number of different surnames (K) in the dataset from 308 to 234.

³⁶After removing observations with clouds, each grid cell is assigned an integer from 0 to 63, with higher values indicating more intense light. Ephemeral events (e.g., fires) are discarded and background noise is set to zero. The objective of the NOAA’s data processing is to capture human-induced lighting (lights from human settlements, towns, and cities). More details on data processing can be found here.

³⁷I try to minimize measurement error by averaging yearly data from the same satellite.

taking value 1 if the share of producers devoting most of the harvest to sale or trade was above the median value in the sample ($P_{50} = 0.03$), and 0 otherwise.

To measure access to public infrastructure, I use district-level data from the 1993 population and housing census on the share of occupied dwellings with access to public sanitation (i.e., having a connection to the public sewer system) and to the public network of water supply. Starting in 1962, a new law on rural access to water and sanitation established that the state would provide the necessary infrastructure, but communal organizations in each district would be responsible for operating and managing the systems. Through communal assemblies, local people managed the repair, cleaning, and disinfection of the infrastructure in order to keep the systems running.³⁸ On average, only 12 percent of occupied dwellings in a district had access to public sanitation.³⁹ The mean share of dwellings with access to the public water supply in a district was relatively higher (24 percent).

4 Empirical Strategy

Location of parishes. I use variation in the location of parishes with respect to ethnic borders as a source of quasi-random variation in ethnic diversity. The analysis relies on the assumption that the location of parishes with respect to ethnic borders, which leads to variation in ethnic diversity across parishes, was not determined by factors related to pre-existing characteristics of native populations or the environment that could influence post-resettlement economic development.

What determined the location of parishes? Although the regulation enacted by Toledo in 1569-1570 described appropriate locations, the extent to which Spanish officials applied the recommendations is unclear. (Pease 1989).⁴⁰ Nonetheless, the regulation included three main characteristics. The first was land abundance. The village had to be surrounded by enough land to be worked by native families following their own rules of crop rotation. Plots of land were thought to be the main source for the payment of tribute, which initially took the form of a tax paid in agricultural output by all native males aged 18 to 50.⁴¹ The second characteristic was access to water sources. Proximity to surface water, which in this context meant access to the system of river basins, was a key advantage that would allow the irrigation of land and the possibility to sustain populations that mainly depended on subsistence agriculture. Finally, in order to facilitate religious indoctrination, villages should be far away from *huacas*, sacred native shrines that generally honored nature. Local officials were also tasked with

³⁸Law 13997 of 1962. Check here for an official report for the implementation of the law.

³⁹Rural access to sanitation facilities in Peru was among the lowest in the world by the year 2000; see the World Bank Natural Rural Water Supply and Sanitation Project for Peru.

⁴⁰See “Instrucción General para los Visitadores” in Lohmann Villena (1986), Appendix III in de La Espada (1881) and Medina (1974b) for details on the 16th century regulation.

⁴¹It was paid on a collective basis. Although the amount to be paid was individual, the responsibility for the payment of tribute fell collectively on the families of native males (Sánchez-Albornoz 1978; Wachtel 1976).

destroying the houses where native families used to live before the resettlement. Shortly after the creation of new jurisdictions, families refusing to relocate were to be punished and forced to move.

In the first part of the empirical analysis (Section 5.1), I examine balance in geographic characteristics between parishes located close to ethnic borders (*with ethnic diversity*) and those located at the interior of ethnic homelands (*without ethnic diversity*). Unfortunately, no comprehensive data on *huacas* exists to test if distance to sacred native shrines varies with proximity to ethnic borders.

Empirical specification. I explore whether the effects of ethnic diversity depend on the degree of within-group heterogeneity to which native populations were exposed before colonization. Specifically, I am interested in the interaction effect of ethnic diversity and the average level of within-group heterogeneity (β_3) in the following specification:

$$y_p = \beta_0 + \beta_1 \text{Ethnic div}_p + \beta_2 \text{Av. within-group } H_p + \beta_3 (\text{Ethnic div}_p \times \text{Av. within-group } H_p) + X_p' \gamma + v_p \quad (1)$$

where y_p is a measure of contemporary living standards in parish p , Ethnic div_p is a dummy variable indicating whether there is an ethnic border within a 10km buffer from the parish capital, and $\text{Av. within-group } H_p$ is the weighted average of within-group heterogeneity within the 10km buffer (\bar{H}). I start with the unconditional relationship and then sequentially add control variables (X_p) to the specification. Geographic controls include mean elevation, variation in elevation, mean pre-colonial caloric suitability, variation in pre-colonial caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude of the parish capital.⁴² Following the previous literature (Michalopoulos and Papaioannou 2013, 2014), I also control for the log of contemporary population density as to capture the effect of ethnic diversity on contemporary living conditions beyond agglomeration. I then add colonial controls. Specifically, I include log tributary population at the time of the resettlement policy (1575), log distance to the closest mining center involving native populations (Dell 2010), and a vector of demographic characteristics reflecting the composition of the population at the end of the colonial period.⁴³ Finally, I add ecclesiastical jurisdiction fixed effects that account for the colonial bishopric to which the parish belonged. I also show results accounting for the religious order in charge of the parish and from an alternative specification using fixed effects at the level of the administrative province instead of the ecclesiastical jurisdiction.⁴⁴ All data

⁴²Elevation and caloric suitability measures refer either to the average or to the standard deviation across grid cells within the corresponding contemporary district. See the data appendix.

⁴³The historical literature has noted the potential decline of native populations in areas under high tributary and mining pressure (Sánchez-Albornoz 1978). The vector of demographic characteristics includes separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792.

⁴⁴The ecclesiastical jurisdictions are Lima, Arequipa, Huamanga, Trujillo, and Cuzco. There are seven categories under the religious order fixed effect: one of the regular orders (Santo Domingo, La Merced, San

sources and definitions are reported in the appendix.

Rather than looking at the effect of ethnic diversity on individual outcomes, I report the standardized average effect size (AES) across contemporary outcomes following the methodology of Kling et al. (2004) and Clingingsmith et al. (2009). The AES averages the standardized individual effects estimated from a seemingly-unrelated regression framework, accounting for the covariance across estimates. Results for individual outcomes are reported in the appendix.

5 Results

Section 5.1 presents the results from balance tests and explores the correlates of within-group heterogeneity. Section 5.2 presents the results from estimating equation 1. Section 5.3 describes robustness checks. Section 5.4 presents the results for mid-term development outcomes and Section 5.5 examines the mechanism.

5.1 Pre-Resettlement Characteristics

Balance tests for ethnic diversity. I start by examining balance in geography, tributary population at the time of the resettlement policy, and distance to colonial mines. Regarding geography, I consider the following characteristics within the $10km$ buffer from the parish capital: mean elevation, variation in elevation, mean pre-colonial caloric suitability, variation in pre-colonial caloric suitability, and log distance to river.

Table 2 presents the results from balance tests. I report robust standard errors in brackets and standard errors corrected for spatial dependence with a distance cutoff of $50km$ in parentheses (Conley 1999). Overall, there are no systematic differences in geography between parishes located at the interior of ethnic homelands and those located close to ethnic borders (columns 1-5). Columns (6) and (7) explore proxies for initial wealth. Column (6) shows that log tributary population at the beginning of the colonial period is balanced, while column (7) shows that there is no statistically significant difference in distance to mines between parishes with and without ethnic diversity.

Relevant characteristics do not vary with proximity to ethnic borders, supporting that colonial officials did not systematically concentrate populations of mixed ethnic origin in places that were inherently different in terms of geography or initial wealth. Moreover, column (8) shows that ethnic diversity, created as a result of the Spanish intervention, is not significantly correlated with the average level of within-group heterogeneity to which native populations were exposed before colonization. This result is consistent with the idea that parishes located

Francisco, San Agustín, and Compañía de Jesús), more than one religious order, and secular clergy if no specific order was in charge of the parish during most of the colonial period. Administrative province fixed effects account for 44 colonial provinces. The ecclesiastical jurisdiction varies at the province level.

close to ethnic borders during the 16th-century resettlement did not systematically concentrate populations from more heterogeneous ethnic groups.

Correlates of within-group heterogeneity. Table 3 explores the correlates of within-group heterogeneity. The analysis is run at the ethnic group level. It presents OLS estimates for individual regressions with the following dependent variables: mean elevation, standard deviation of elevation, mean pre-colonial caloric suitability, standard deviation of pre-colonial caloric suitability, log river density, total land area, and the log of approximate population density at the time of the Spanish conquest.

Unsurprisingly, since within-group heterogeneity in specializations is measured as a diversity index that sums over elevation zones, it is positively and significantly correlated with variation in elevation and in the pre-colonial caloric suitability of land (columns 1 to 5). The correlation between within-group heterogeneity and total land surface is not statistically significant, and log transforming surface does not change the result (columns 6 and 7). Finally, the correlation with log population density, which reflects comparative development in the period prior to 1500CE (Ashraf and Galor 2011, 2013), is positive but not statistically significant (column 8). In robustness exercises, I account for the main correlates of within-group heterogeneity as well as for other observable characteristics of ethnic groups.

5.2 Main Results

For comparison with the previous literature, I first look at the overall effect of ethnic diversity, comparing average living standards between parishes with ethnic diversity and those with similar baseline characteristics in which only one ethnic group was concentrated (Table 4).

Looking at the standardized AES across contemporary outcomes, I find that ethnic diversity is robustly associated with lower living standards in the long run (between -0.16 and -0.22 standard deviations relative to parishes without ethnic diversity). Column (1) shows that the unconditional AES (-0.21 standard deviations) is statistically significant at the 1 percent level. Columns (2) to (4) show that the magnitude and statistical significance are not affected by the inclusion of geography controls, population density, initial tributary population, distance to colonial mines, and demographic controls that reflect the composition of the population by the end of the colonial period. Including ecclesiastical jurisdiction fixed effects results in a decrease of the AES from -0.21 to -0.16 standard deviations (column 5), though it remains statistically significant at the 1 percent level. For comparison, column (6) reports the estimated effect without accounting for population density, which results in an AES of -0.17 (significant at the 5 percent level). The results are in line with the literature on the costs of ethnic diversity (e.g., Miguel and Gugerty 2005; Hjort 2014). However, they highlight the persistent effects of forced ethnic diversity at the local level: after almost two hundred years of independence from Spanish rule, parishes whose initial populations were ethnically diverse

tend to be worse off than parishes with an ethnically homogeneous founding population.⁴⁵

Table 5 shows the results for the interaction effect of ethnic diversity and average within-group heterogeneity (equation 1). Columns (1) to (5) follow the same structure as Table 4, sequentially adding baseline controls and ecclesiastical jurisdiction fixed effects to the unconditional specification. The results exhibit a robust pattern: while the estimated coefficient on ethnic diversity is negative, its interaction with native groups' average within-group heterogeneity is positive. Both coefficients are statistically significant at the 1 percent level, showing that the costs of ethnic diversity tend to be overcome among populations from more heterogeneous ethnic groups. The estimates are unchanged when including fixed effects that account for the religious order in charge of the parish (column 6). In column (7), I include fixed effects at the level of the colonial province instead of the ecclesiastical jurisdiction, showing that both coefficients remain significant at the 1 percent level when accounting for the administrative province. The last column reports the AES without population density as control variable and shows that the same pattern arises.⁴⁶

Figure 3 plots the estimated AES and 95 percent confidence intervals after baseline controls, religious order fixed effects, and colonial province fixed effects (column 7). On average, the negative association between ethnic diversity and contemporary living standards decreases (from -0.47 to -0.13 standard deviations relative to parishes without ethnic diversity) as average within-group heterogeneity reaches the median value ($\bar{H}_{50} = 0.67$) and turns positive (from 0.21 to 0.32 standard deviations) for parishes above the 90th percentile ($\bar{H}_{90} = 0.88$).⁴⁷ Table A.8 additionally reports the estimated AES of within-group heterogeneity for the whole sample and separately for parishes with and without ethnic diversity. The results show that there is a positive association between exposure to a high level of within-group heterogeneity and contemporary living standards in the subsample of parishes with ethnic diversity. The next section addresses concerns regarding measurement and omitted characteristics of ethnic groups.

5.3 Robustness Checks

I start with a series of robustness checks to make sure that the previous pattern is not driven by omitted group characteristics. All robustness exercises are shown for columns (6) and (7) of Table 5 (i.e., with baseline controls, religious order fixed effects, and either ecclesiastical jurisdiction or colonial province fixed effects). First, is within-group heterogeneity as captured by diversity in specializations or is just variation in any geographic characteristic? Table 6 shows that the pattern of results is robust to accounting for the main correlates of within-group

⁴⁵Tables A.3 and A.4 report the results separately for each contemporary outcome. Table A.5 reports the AES using a standard measure of ethnic diversity based on the Herfindahl index, instead of the ethnic diversity dummy variable.

⁴⁶Tables A.6 and A.7 present the results separately for each contemporary outcome.

⁴⁷Figure A.2 shows the estimated marginal effect of ethnic diversity on each contemporary outcome.

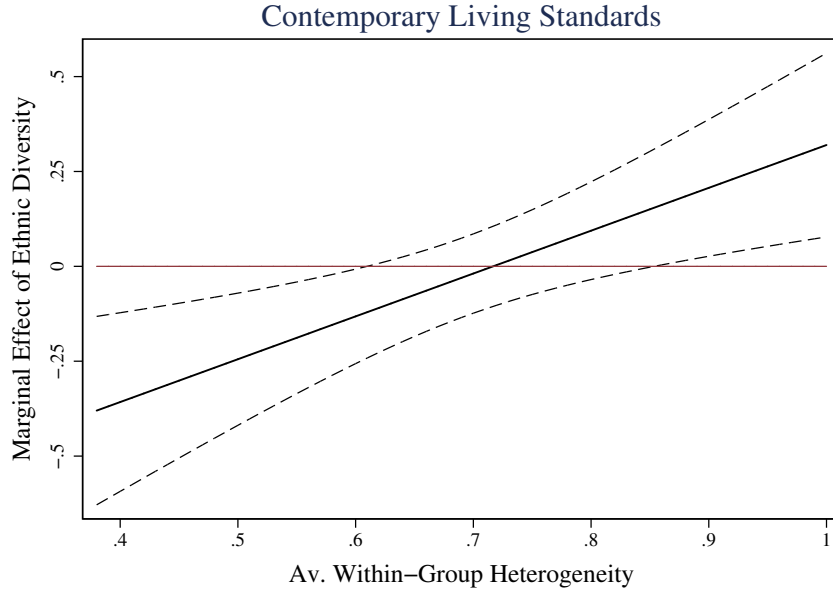


FIGURE 3

Notes. The solid line represents the standardized average effect size (AES) across contemporary outcomes (ln average light intensity per capita –2000-2003–, indicator for non-subsistence agriculture –1994–, share of dwellings with access to public sanitation –1993–, and share of dwellings with access to the public water network –1993–) after baseline controls, religious order fixed effects, and colonial province fixed effects. Dashed lines represent 95% confidence intervals.

heterogeneity (i.e., weighted average of ethnic-level variation in elevation and in pre-colonial caloric suitability within the 10km buffer) and their interaction with ethnic diversity. Thus, although my measure of within-group heterogeneity in specializations is based on elevation zones, it is not capturing the same pattern as variation in geographic elevation or in the caloric suitability of land.

Second, is within-group heterogeneity or other ethnic characteristic? Gains from trade as a consequence of ecological diversity could predict state centralization (Fenske 2014). Although information on pre-Inca institutions is scarce, to the best of my knowledge there is no evidence on complex states in the region of analysis before the expansion of the Inca empire. Table 7 shows that the same pattern of results is observed when accounting for the interaction of ethnic diversity and different group characteristics that could be related with stronger ethnic development, like the average land area of ethnic groups (columns 1 and 2), approximate population density at the time of the Spanish conquest (columns 3 and 4), and land suitability for maize, a cereal grain characterized by its high caloric content and known to be available in Peru prior to the conquest (columns 5 and 6).⁴⁸

Table 8 implements additional robustness checks. Columns (1) and (2) show that the results

⁴⁸The availability of appropriate cereal crops over roots and tubers could predict the emergence of more complex hierarchies (Mayshar et al. 2020). On the caloric content of crops, see, for instance, the USDA Nutrient Database for Standard Reference and Galor and Özak (2016).

exhibit the same pattern when controlling for an extended set of (uninteracted) ethnic characteristics. Specifically, I include the weighted average of ethnic-level land area, mean elevation, variation in elevation, mean pre-colonial caloric suitability, variation in pre-colonial caloric suitability, log river density, and log population density at the time of the Spanish conquest.⁴⁹ Columns (3) and (4) include fixed effects for the majority ethnic group, defined as the group with the highest area share within the 10km buffer.⁵⁰ The coefficients exhibit the same pattern and remain statistically significant at the 1 percent level. Finally, columns (5) and (6) show that the results are not affected by excluding parishes in which the majority ethnic group is a coastal group (approximately 10 percent of parishes). Coastal groups also tend to occupy larger surface areas than groups from the highlands.

The appendix reports the results from using alternative variable definitions. In Table A.9, I use a standard measure of diversity based on the Herfindahl index to quantify ethnic fractionalization ($1 - \sum_e w_{pe}^2$), where w_{pe} is the area share of ethnic group e within a 10km buffer from the parish capital p . In Table A.10, I use the same diversity index to measure within-group heterogeneity in specializations ($1 - \sum_j s_{ej}^2$), where s_{ej} refers to the area share of elevation zone j within the homeland of ethnic group e . The same pattern arises, showing that the results are not sensitive to specific variable definitions. Table A.11 presents estimates from using a robust version of the ethnic diversity dummy variable that accounts for imprecise ethnic borders. Specifically, it requires the area share of each ethnic group within the 10km buffer to be at least 10 percent (approximately 30km²).

I report further robustness checks in the appendix. First, to examine influential observations, Figure A.3 displays point estimates and confidence intervals for the baseline specification after excluding parishes one by one. Second, I report OLS results from using a living standards index as outcome variable (standardized score of the first principal component for the four contemporary outcomes, all with positive factor loadings; see Table A.12). Third, Table A.13 reports standard errors corrected for spatial dependence with different distance cutoffs (Conley 1999). Additional robustness checks related with the Inca period are reported in Table A.14.

Finally, did the formation of administrative provinces after the Spanish conquest rely on previous ethnic borders? A one-to-one correspondence between the homeland of ethnic groups and the jurisdiction of 16th-century Spanish provinces would be inconsistent with the natural experiment. A map of early provinces (*corregimientos* in Cook 1982) suggests that this was not the case, as only 13 percent of parishes have a 10km buffer that intersects both an ethnic border and a *corregimiento* border.⁵¹ Furthermore, excluding such parishes from the analysis does not change the results (Table A.15). To conclude, I report results from a placebo test

⁴⁹The weighted average uses the area share of each ethnic group within the 10km buffer as weight.

⁵⁰Out of the 49 ethnic groups, 46 are represented as a majority group in the analysis. The median and mean number of parishes per majority group are six and seven parishes, respectively. There are four groups with only one parish.

⁵¹There are 44 *corregimientos* in the region of analysis.

using *corregimiento* borders instead of ethnic borders (A.16). Specifically, I run equation 1 using (i) a dummy variable for whether there is a *corregimiento* border within the 10km buffer from the parish capital and (ii) the weighted average of within-*corregimiento* heterogeneity. Estimates are small and not statistically significant.

Overall, these robustness checks alleviate concerns regarding potential confounding variation and measurement issues. They support that, beyond the direct effect of geography, having belonged to an ethnic group with more heterogeneous subpopulations prior to the conquest helps overcome the negative effects of ethnic diversity in the long run.

5.4 Mid-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 local manufacturing, retail and services, and by improved literacy rates. The 1876 population census provides detailed data on occupations. I classify the 318 different occupations in my sample by sector of economic activity and then compute the share of male employment in each sector. The data reveal that, overall, parishes continued to be predominantly agricultural in the late 19th century, with 80 percent of the male population employed in the primary sector, on average. The remaining 20 percent of the labor force was composed of local manufacturers working in activities like pottery and carpentry, among other occupations of the secondary sector, followed by a minority employed in retail and services (tertiary sector).

Columns (1) to (6) of Table 9 report the results from estimating equation 1 using the share of male employment in the primary, secondary, and tertiary sectors as dependent variables. For each sector, the first column presents estimates from regressions without control variables, while the second column reports estimates after including baseline controls as well as ecclesiastical jurisdiction and religious order fixed effects. Regressions are weighted by total male population in 1876. I find that parishes built on ethnically diverse populations in the 16th century, compared to those built on a single ethnic group, tend to be more oriented toward secondary and tertiary activities, to the detriment of the agricultural sector, as average within-group heterogeneity increases. Columns (7) and (8) show estimates from using the literacy rate of the male population (those who can read and/or write) as outcome variable. By 1876, ethnically diverse parishes were also associated with better literacy rates, relative to those without ethnic diversity, the higher the average within-group heterogeneity among their ancestors. Figure A.4 shows the estimated marginal effects of ethnic diversity on mid-term outcomes.

5.5 Mechanism

This section explores potential mechanisms. Section 5.5.1 provides evidence on cultural transmission. First, I briefly highlight historical evidence that supports the role of an Andean culture of reciprocity in native societies. I then provide empirical evidence on contemporary associations and examine whether, in line with the cultural transmission mechanism, previous exposure to within-group heterogeneity favored inter-group contact during the colonial period. Section 5.5.2 explores the role of economic complementarities in the subsample of parishes with ethnic diversity.

5.5.1 Cultural Transmission Mechanism

Studies on the Andean culture of reciprocity. Historical evidence suggests that the subsistence strategy based on complementarities between vertical production zones was sustained by engaging in reciprocities between individuals of the same ethnic group. For example, Stern (1995, p. 76) writes that “In general, Andean rules of reciprocity and redistribution served to govern the exchanges” and emphasizes “Andean peoples sought self-sufficiency ... by engaging in reciprocities enabling the collective kin or ethnic group to directly produce diverse goods in scattered ecological zones.”

There are also references to reciprocal behavior under the Inca empire. For example, early chronicles document that during Inca times “if it was necessary for someone to do something else in an emergency, like war or some other urgent matter, the other Indians of the community worked the fields of the absent man without asking or receiving any compensation beyond their food, and, this done, each cultivated his own fields. This assistance which the community rendered to its absent members caused each man to return home willingly when he had finished his job, for he might find on his return after long absence that a harvest which he had neither sown nor reaped was gathered into his house” (in Rowe 1946, p. 266).⁵²

Evidence on contemporary associations. Exposure to within-group heterogeneity may have contributed to the formation of cooperative behavior and more open attitudes toward out-group members. One possibility is that the transmission of this culture favored positive interactions between individuals of different ethnic origin during the colonial period and, transmitted over generations, contributed to sustain long-run performance. Following recent literature on the transmission of cultural traits (e.g., Guiso et al. 2016), Table 10 provides evidence on contemporary associations.⁵³

In columns (1) and (2), the dependent variable is a dummy indicating the presence of

⁵²See also Murra (1975, p. 27-28) and Wachtel (1976, p. 96-97) for further evidence supporting the presence of reciprocal relations between community members before and during the Inca empire, respectively.

⁵³Unfortunately, Peruvian household surveys do not contain questions on generalized trust (e.g., *Would you say that you can trust most people?*) and Latinobarometer surveys only cover 48 of the districts in my the sample.

neighborhood associations (*Registro Nacional de Municipalidades* for the year 2002). In columns (3) and (4), I look at the share of land managed by farmers in agricultural associations (1994 agricultural census). For each outcome, the first column displays unconditional estimates, while the second column displays estimates after including baseline controls as well as ecclesiastical jurisdiction and religious order fixed effects. Previous studies show that ethnic diversity is associated with lower social engagement (e.g., Alesina and La Ferrara 2000). Consistent with cultural transmission, I find that this negative association decreases the higher the average level of native groups’ within-group heterogeneity. Table A.17 shows supporting evidence from household-level survey data using participation in neighborhood associations and unions as dependent variables.

Societal integration during the colonial period. Does exposure to within-group heterogeneity favor inter-group interaction? I explore inter-group contact and integration during the colonial period (1605-1780) using the sample of baptized individuals. Following recent literature, I focus on inter-ethnic unions as a proxy for societal integration (Bazzi et al. 2019). However, since I do not observe ethnicity, but only surnames, I use a measure of linguistic distance between the first surname of each individual’s mother and father in order to detect potential inter-ethnic unions. Specifically, I use a standard measure, called Levenshtein distance (L), equal to the minimum number of changes (i.e., deletions, insertions or substitutions) required to transform one surname into the other.

The sample contains 26,925 individuals with native roots in the first surname of both the mother and the father, covering 61 parishes, of which 13 are parishes with ethnic diversity. Figure 4 shows the scatterplot between the average level of within-group heterogeneity and the share of unions with linguistic distance above the 75th percentile ($L_{75}=7$), separately for parishes with and without ethnic diversity. As an example, the surname *Guaman* has Levenshtein distance 7 with surnames like *Ispilco* or *Chuquili*. Using this threshold, the mean share of unions between linguistically distant individuals in a parish is 0.297. Although the limited coverage of the data remains a concern, the results suggest a positive correlation between within-group heterogeneity and the share of potential inter-ethnic unions, while no relationship is found in the subsample of parishes without ethnic diversity.

Table A.18 presents estimates for equation 1 using the share of unions between linguistically distant individuals as outcome variable. Column (1) shows the unconditional correlation. Column (2) accounts for the log total number of individuals and the mean share of potential partners, defined as people in the parish with whom the individual has L distance above the 75th percentile. Column (3) includes baseline geography controls and log distance to colonial mines. Column (4) adds ecclesiastical jurisdiction and religious order fixed effects and column (5) uses population weights. Consistent with the graphical evidence, the results suggest that having been part of a more heterogeneous ethnic group facilitated integration in parishes with ethnic diversity.

Finally, in line with societal integration, additional results from contemporary survey data

show the same pattern when looking at the formation of identity. When asked *Which group do you identify most with?*, the descendants of populations in parishes with ethnic diversity, as compared to those from parishes built on a single ethnicity, tend to identify more strongly with the administrative region to which they belong than with their own ethnicity, race, or native community, the higher the average within-group heterogeneity among their ancestors (Table A.19).

5.5.2 Cultural Transmission or Economic Complementarities?

Does it matter whether within-group heterogeneity comes from the majority or the minority ethnic group? Previous sections have focused on the average level of within-group heterogeneity in a parish. I now explore the role of the minority and majority ethnic groups in parishes with ethnic diversity. On the one hand, recent papers emphasize the role of economic complementarities for peaceful inter-ethnic coexistence (Jha 2013, 2018; Becker and Pascali 2019). In this setting, if the ethnic majority in the parish comes from an ethnic group of low within-group heterogeneity, the ethnic minority may be able to more easily integrate if they have historically been highly heterogeneous (i.e., the majority may try to enjoy any complementary skill of the minority ethnic group). On the other hand, positive inter-ethnic interactions due to cultural transmission may be more likely in parishes where both the minority and the majority come from ethnic groups with high within-group heterogeneity (i.e., they are already used to social interactions in diverse environments).

I construct a dummy variable indicating whether the majority ethnicity (the one with the highest area share within the 10km buffer) belonged to an ethnic group with high within-group heterogeneity. Specifically, the dummy variable takes value 1 if within-group heterogeneity is above the 75th percentile ($H_{75} = 0.82$), and 0 otherwise. I then construct an analogous dummy variable for high within-group heterogeneity of the ethnic minority.⁵⁴ Looking at the sample of parishes with ethnic diversity, 24 percent have a highly heterogeneous ethnic majority, 17 percent have a highly heterogeneous ethnic minority, and approximately 10 percent are parishes where both the minority and the majority ethnic groups were highly heterogeneous during the pre-colonial period.

I then examine whether the mid-term composition of occupations and contemporary living standards differ depending on which ethnic group drives within-group heterogeneity. Table 11 presents the results from estimating the coefficients on the two dummy variables and their interaction. I first explore occupations among the male population using the 1876 population census. In columns (1) to (6), the outcome variables are the log number of different occupations in the primary (columns 1 and 2), secondary (columns 3 and 4) and tertiary

⁵⁴Most parishes with ethnic diversity (approximately 85 percent) have two ethnic groups within the 10km buffer. For the remaining 15 percent of parishes (all of which have three ethnic groups, except for a single parish with four groups), I focus on the ethnic groups with the highest (ethnic majority) and the lowest (ethnic minority) area shares within the buffer.

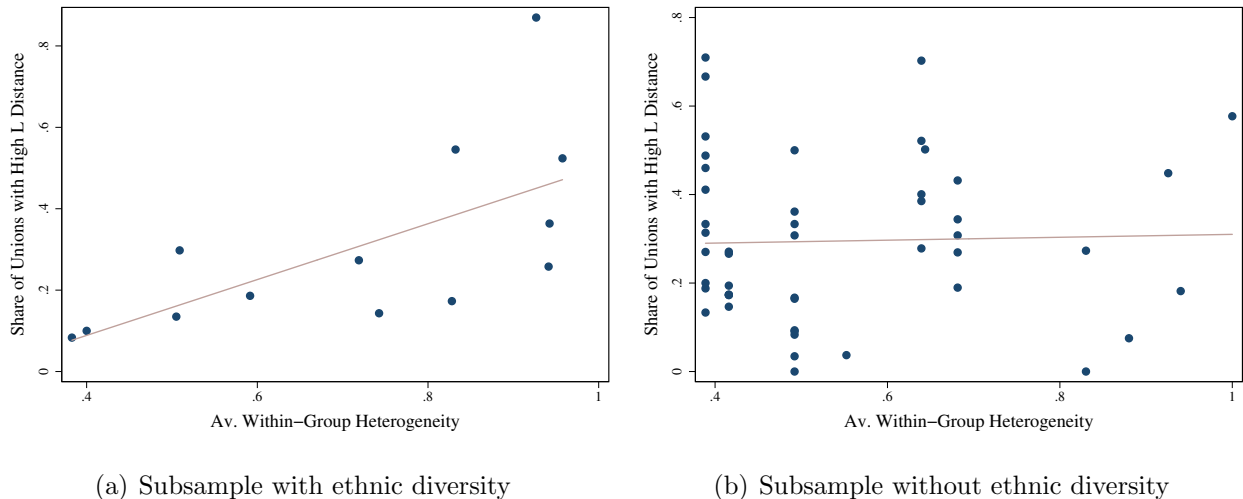


FIGURE 4

Notes. Scatterplot between average within-group heterogeneity and share of unions with Levenshtein distance ≥ 75 th percentile ($L_{75} = 7$), separately for parishes with and without ethnic diversity.

(columns 5 and 6) sectors. For each sector, the first column presents the results without control variables, while the second column includes baseline controls as well as ecclesiastical jurisdiction and religious order fixed effects.

Consistent with the positive role of complementarities, the results suggest that the economic advantage comes from parishes with a majority of low within-group heterogeneity but a highly heterogeneous ethnic minority: (i) having a minority from a highly heterogeneous ethnic group is associated with a greater variety of non-primary occupations among parishes in which the majority ethnic group exhibited low within-group heterogeneity, and (ii) in parishes in which the majority also comes from a highly heterogeneous ethnic group, this effect tends to be offset.⁵⁵ Unfortunately, no data on mid-term income exist to assess whether this variety translated into economic gains. Nonetheless, the AES across contemporary outcomes (columns 7 and 8) provides consistent results.

6 Conclusion

I present evidence that ethnic diversity need not spell poor development outcomes—a history of within-group heterogeneity can turn ethnic diversity into an advantage for development. By showing that the effects of ethnic diversity depend on previous exposure to within-group heterogeneity, this paper’s results provide insights for future research and policy-making. Specifically, the heterogeneity of subpopulations within societies or ethnic groups has received

⁵⁵Table A.20 shows the results controlling for the log of total population in 1876 and using a Herfindahl index to measure diversity in occupations.

little attention. The findings suggest that a deeper understanding of this dimension can help shed light on the features that shape comparative economic growth and development. They also invite us to consider this pre-condition when studying the consequences of voluntary population movements in the context of globalization, as well as the forced displacement of individuals belonging to a certain ethnicity, race, religion, or nationality. I provide evidence that the transmission of reciprocal behavior and more open attitudes toward out-group members is a likely channel when looking at average exposure to within-group heterogeneity. Furthermore, in exploring which ethnic group drives within-group heterogeneity (i.e., the majority or the minority ethnic group), the results suggest that inter-group complementarities can also play an important role.

References

- ACADEMIA MAYOR DE LA LENGUA QUECHUA (2005): *Diccionario Quechua – Español – Quechua, Qheswa – Español – Qheswa: Simi Taje*, Cusco, Perú: Gobierno Regional Cusco, 2nd ed.
- ALESINA, A., R. BAQIR, AND W. EASTERLY (1999): “Public Goods and Ethnic Divisions,” *The Quarterly Journal of Economics*, 114, 1243–1284.
- ALESINA, A., A. DEVLEESCHAUWER, W. EASTERLY, S. KURLAT, AND R. WACZIARG (2003): “Fractionalization,” *Journal of Economic Growth*, 8, 155–194.
- ALESINA, A., P. GIULIANO, AND N. NUNN (2013): “On the Origins of Gender Roles: Women and the Plough,” *The Quarterly Journal of Economics*, 128, 469–530.
- ALESINA, A. AND E. LA FERRARA (2000): “Participation in Heterogeneous Communities,” *The Quarterly Journal of Economics*, 115, 847–904.
- (2005): “Ethnic Diversity and Economic Performance,” *Journal of Economic Literature*, 43, 762–800.
- ASHRAF, Q. AND O. GALOR (2011): “Dynamics and Stagnation in the Malthusian Epoch,” *American Economic Review*, 101, 2003–41.
- (2013): “The “Out of Africa” Hypothesis, Human Genetic Diversity, and Comparative Economic Development,” *American Economic Review*, 103, 1–46.
- BAZZI, S., A. GADUH, A. D. ROTHENBERG, AND M. WONG (2019): “Unity in Diversity? How Intergroup Contact Can Foster Nation Building,” *American Economic Review*, 109, 3978–4025.
- BECKER, S. O. AND L. PASCALI (2019): “Religion, Division of Labor, and Conflict: Anti-Semitism in Germany over 600 Years,” *American Economic Review*, 109, 1764–1804.
- BENTZEN, J., J. G. HARIRI, AND J. A. ROBINSON (2017): “Power and Persistence: The Indigenous Roots of Representative Democracy,” *The Economic Journal*, 129, 678–714.
- BERTONIO, L. (2011): *Vocabulario de la Lengua Aymara*, La Paz, Bolivia: Instituto de Lenguas y Literaturas Andinas–Amazónicas (ILLA–A), [1612].
- BONGERS, J. L., N. NAKATSUKA, C. OSHEA, T. K. HARPER, H. TANTALEÁN, C. STANISH, AND L. FEHREN-SCHMITZ (2020): “Integration of Ancient DNA with Transdisciplinary Dataset Finds Strong Support for Inca Resettlement in the South Peruvian Coast,” *Proceedings of the National Academy of Sciences*.
- BUGGLE, J. AND R. DURANTE (2017): “Climate Risk, Cooperation, and the Co-Evolution of Culture and Institutions,” CEPR Discussion Paper DP12380.

- CARPIO, M. Á. AND M. E. GUERRERO (2020): “Did the Colonial *Mita* Cause a Population Collapse? What Current Surnames Reveal in Peru,” Working Paper.
- CARRAFFA, A. G. AND A. G. CARRAFFA (1920–1963): *Diccionario Heráldico y Genealógico de Apellidos Españoles y Americanos*, Madrid: Imprenta Antonio Marzo.
- CLINGINGSMITH, D., A. I. KHWAJA, AND M. KREMER (2009): “Estimating the Impact of the Hajj: Religion and Tolerance in Islam’s Global Gathering,” *The Quarterly Journal of Economics*, 124, 1133–1170.
- COLANTONIO, S. E., G. W. LASKER, B. A. KAPLAN, AND V. FUSTER (2003): “Use of Surname Models in Human Population Biology: A Review of Recent Developments,” *Human Biology*, 75, 785–807.
- CONADI (2011): *Diccionario Trilingüe: Aymara-Castellano-Inglés*, Chile: Corporación Nacional de Desarrollo Indígena (CONADI), Dirección Regional Arica y Parinacota.
- CONLEY, T. G. (1999): “GMM Estimation with Cross Sectional Dependence,” *Journal of econometrics*, 92, 1–45.
- COOK, N. D. (1982): “Population Data for Indian Peru: Sixteenth and Seventeenth Centuries,” *The Hispanic American Historical Review*, 62, 73–120.
- (2010): *La Catástrofe Demográfica Andina: Perú 1520-1620*, Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- CÓRDOBA Y SALINAS, D. (1957): *Crónica Franciscana de las Provincias del Perú*, Washington, DC: Academy of American Franciscan History.
- DE ARMAS MEDINA, F. (1953): *Cristianización del Perú (1532-1600)*, Sevilla: Escuela de Estudios Hispano-Americanos.
- DE LA ESPADA, M. J. (1881): *Relaciones Geográficas de Indias. Perú*, vol. I, Madrid: Ministerio de Fomento.
- DE LA VEGA, G. (1960): *Obras Completas del Inca Garcilaso de la Vega*, vol. 133, Ediciones Atlas, [1609].
- DE LUCCA, M. (1983): *Diccionario Aymara-Castellano, Castellano-Aymara*, La Paz, Bolivia: Comisión de Alfabetización y Literatura en Aymara (CALA).
- DELL, M. (2010): “The Persistent Effects of Peru’s Mining *Mita*,” *Econometrica*, 78, 1863–1903.
- DEPETRIS-CHAUVIN, E. AND Ö. ÖZAK (2020): “The Origins of the Division of Labor in Pre-Modern Times,” *Journal of Economic Growth*, 25, 297–340, forthcoming.

- DESMET, K., I. ORTUÑO-ORTÍN, AND R. WACZIARG (2012): “The Political Economy of Linguistic Cleavages,” *Journal of Development Economics*, 97, 322–338.
- (2017): “Culture, Ethnicity, and Diversity,” *American Economic Review*, 107, 2479–2513.
- DESMET, K., S. WEBER, AND I. ORTUÑO-ORTÍN (2009): “Linguistic Diversity and Redistribution,” *Journal of the European Economic Association*, 7, 1291–1318.
- DOUGNAC RODRÍGUEZ, A. (1994): *Manual de Historia del Derecho Indiano*, Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas.
- DURANTON, G. AND D. PUGA (2000): “Diversity and Specialisation in Cities: Why, Where and When Does it Matter?” *Urban Studies*, 37, 533–555.
- EASTERLY, W. AND R. LEVINE (1997): “Africa’s Growth Tragedy: Policies and Ethnic Divisions,” *The Quarterly Journal of Economics*, 112, 1203–1250.
- ENKE, B. (2019): “Kinship, Cooperation, and the Evolution of Moral Systems,” *The Quarterly Journal of Economics*, 134, 953–1019.
- FEARON, J. D. AND D. D. LAITIN (2003): “Ethnicity, Insurgency, and Civil War,” *American Political Science Review*, 97, 75–90.
- FENSKE, J. (2014): “Ecology, Trade, and States in Pre-Colonial Africa,” *Journal of the European Economic Association*, 12, 612–640.
- FLYNN, F. J. (2005): “Having an Open Mind: The Impact of Openness to Experience on Interracial Attitudes and Impression Formation,” *Journal of Personality and Social Psychology*, 88, 816–826.
- FUENZALIDA, FERNANDO MAYER, E., G. ESCOBAR, F. BOURRICAUD, AND J. MATOS MAR (1970): *El Indio y el Poder en el Perú*, Lima: Instituto de Estudios Peruanos.
- GALOR, O. AND Ö. ÖZAK (2016): “The Agricultural Origins of Time Preference,” *American Economic Review*, 106, 3064–3103.
- GARCÍA, J. C. (1997): “Oposiciones a Parroquias y Doctrinas. El Catálogo de la Sección Concursos, Archivo Arzobispal de Lima: Siglos XVII-XIX,” *Revista Andina*, 15, 421–491.
- GENNAIOLI, N. AND I. RAINER (2007): “The Modern Impact of Precolonial Centralization in Africa,” *Journal of Economic Growth*, 12, 185–234.
- GONZÁLEZ HOLGUÍN, D. (1952): *Vocabulario de la Lengua General de todo el Perú llamada Lengua Qquichua, o del Inca*, Lima: Universidad Nacional Mayor de San Marcos, [1608].

- GOOTENBERG, P. (1991): “Population and Ethnicity in Early Republican Peru: Some Revisions,” *Latin American Research Review*, 26, 109–157.
- GORDON, M. M. (1964): *Assimilation in American Life: The Role of Race, Religion, and National Origins*, New York: Oxford University Press.
- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2016): “Long-Term Persistence,” *Journal of the European Economic Association*, 14, 1401–1436.
- HATTON, T. J. (2020): “Asylum Migration to the Developed World: Persecution, Incentives, and Policy,” *Journal of Economic Perspectives*, 34, 75–93.
- HENDERSON, J. V., A. STOREYGARD, AND D. N. WEIL (2012): “Measuring Economic Growth from Outer Space,” *American Economic Review*, 102, 994–1028.
- HJORT, J. (2014): “Ethnic Divisions and Production in Firms,” *The Quarterly Journal of Economics*, 129, 1899–1946.
- HODLER, R. AND P. A. RASCHKY (2014): “Regional Favoritism,” *The Quarterly Journal of Economics*, 129, 995–1033.
- JHA, S. (2013): “Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia,” *American Political Science Review*, 107, 806–832.
- (2018): “Trading for Peace,” *Economic Policy*, 33, 485–526.
- KLING, J. R., J. B. LIEBMAN, L. F. KATZ, AND L. SANBONMATSU (2004): “Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment,” KSG Working Paper No. RWP04-035.
- LA PORTA, R., F. LOPEZ-DE SILANES, A. SHLEIFER, AND R. VISHNY (1999): “The Quality of Government,” *The Journal of Law, Economics, and Organization*, 15, 222–279.
- LASKER, G. W. (1980): “Surnames in the Study of Human Biology,” *American Anthropologist*, 82, 525–538.
- (1985): *Surnames and Genetic Structure*, Cambridge University Press.
- LAZEAR, E. P. (1999): “Globalisation and the Market for Team-Mates,” *The Economic Journal*, 109, 15–40.
- LEWONTIN, R. C. (1972): “The Apportionment of Human Diversity,” *Evolutionary Biology*, 6, 381–398.
- LISSÓN CHÁVEZ, E. (1943): *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.

- LOHMANN VILLENA, G. (1986): *Francisco de Toledo. Disposiciones Gubernativas para el Virreinato del Perú (1569-1574)*, vol. I, Sevilla: Publicaciones de la Escuela de Estudios Hispanoamericanos de Sevilla, Consejo Superior de Investigaciones Científicas.
- LYNCH, J. (1962): *Administración Colonial Española, 1782-1810: El Sistema de Intendencias en el Virreinato del Río de la Plata*, Buenos Aires: Editorial Universitaria de Buenos Aires.
- MAGURRAN, A. E. (2004): *Measuring Biological Diversity*, Oxford, U.K.: Blackwell Science Ltd.
- MATIENZO, J. D. (1910): *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, [1567].
- MAYSHAR, J., O. MOAV, AND L. PASCALI (2020): “The Origin of the State: Land Productivity or Appropriability?” Working Paper.
- MCCRAE, R. R. (1996): “Social Consequences of Experiential Openness,” *Psychological Bulletin*, 120, 323–337.
- MCCRAE, R. R. AND P. T. COSTA (1997): “Conceptions and Correlates of Openness to Experience,” *Handbook of Personality Psychology*, 825–847.
- MEDINA, A. M. (1974a): “Las Reducciones en el Perú (1532-1600),” *Historia y Cultura*, 8, 141–172, museo Nacional de Arqueología e Historia del Perú.
- (1974b): “Las Reducciones en el Perú durante el Gobierno del Virrey Francisco de Toledo,” *Anuario de Estudios Americanos*, 31, 819–842, escuela de Estudios Hispanoamericanos de Sevilla, Consejo Superior de Investigaciones Científicas.
- (1993): “Las Reducciones Toledanas en el Perú,” in *Pueblos de Indios: Otro Urbanismo en la Región Andina*, Quito: Ediciones ABYA-YALA, 263–316.
- MICHALOPOULOS, S. (2012): “The Origins of Ethnolinguistic Diversity,” *American Economic Review*, 102, 1508–39.
- MICHALOPOULOS, S. AND E. PAPAIOANNOU (2013): “Pre-Colonial Ethnic Institutions and Contemporary African Development,” *Econometrica*, 81, 113–152.
- (2014): “National Institutions and Subnational Development in Africa,” *The Quarterly Journal of Economics*, 129, 151–213.
- (2018): “Spatial Patterns of Development: A Meso Approach,” *Annual Review of Economics*, 10, 383–410.
- MIGUEL, E. AND M. K. GUGERTY (2005): “Ethnic Diversity, Social Sanctions, and Public Goods in Kenya,” *Journal of Public Economics*, 89, 2325–2368.

- MONTALVO, J. G. AND M. REYNAL-QUEROL (2005): “Ethnic Diversity and Economic Development,” *Journal of Development Economics*, 76, 293–323.
- (forthcoming): “Ethnic Diversity and Growth: Revisiting the Evidence,” *Review of Economics and Statistics*.
- MOSCONA, J., N. NUNN, AND J. A. ROBINSON (forthcoming): “Segmentary Lineage Organization and Conflict in Sub-Saharan Africa,” *Econometrica*.
- MURDOCK, G. P. (1967): *Ethnographic Atlas*, University of Pittsburgh Press.
- MURRA, J. V. (1975): *Formaciones Económicas y Políticas del Mundo Andino*, Lima: Instituto de Estudios Peruanos.
- (1995): “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, 57–72.
- (2002): “Las Sociedades Andinas antes de 1532,” in *América Latina en la Época Colonial*, Barcelona: Crítica, vol. 1, 56–82.
- NUNN, N. AND D. PUGA (2012): “Ruggedness: The Blessing of Bad Geography in Africa,” *Review of Economics and Statistics*, 94, 20–36.
- NUNN, N. AND L. WANTCHEKON (2011): “The Slave Trade and the Origins of Mistrust in Africa,” *American Economic Review*, 101, 3221–52.
- PAZ SOLDÁN, M. F. (1877): *Diccionario Geográfico Estadístico del Perú*, Lima: Imprenta del Estado.
- PEASE, F. (1989): *Del Tawantinsuyu a la Historia del Perú*, Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- (1992): *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.
- PLATT, L. D. (1996): *Hispanic Surnames and Family History*, Baltimore, MD: Genealogical Publishing Company.
- PULGAR VIDAL, J. (1941): “Las Ocho Regiones Naturales del Perú,” in *Boletín del Museo de Historia Natural “Javier Prado”*, 17, 145–160, imprenta D. Miranda, Lima.
- (2012): *Geografía del Perú. Las Ocho Regiones Naturales del Perú*, Lima: Instituto de Ciencias de la Naturaleza, Territorio y Energías Renovables de la Pontificia Universidad Católica del Perú (INTE-PUCP), 12 ed.

- RELIGIOSOS FRANCISCANOS MISIONEROS DE LOS COLEGIOS DE PROPAGANDA FIDE DEL PERÚ (1998): *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, [1905].
- RENIEC (2012): *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).
- ROWE, J. H. (1946): “Inca Culture at the Time of the Spanish Conquest,” in *Handbook of South American Indians*, vol. 2, 183–330.
- SAIGNES, T. (1991): “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, 91–137.
- SÁNCHEZ-ALBORNOZ, N. (1978): *Indios y Tributos en el Alto Perú*, Lima: Instituto de Estudios Peruanos.
- SHANNON, C. E. (1948): “A Mathematical Theory of Communication,” *Bell System Technical Journal*, 27, 379–423.
- SIBLEY, C. G. AND J. DUCKITT (2008): “Personality and Prejudice: A Meta-Analysis and Theoretical Review,” *Personality and Social Psychology Review*, 12, 248–279.
- STERN, S. J. (1995): “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, 73–100.
- TAPIA, M. (2013): “Diagnóstico de los Ecosistemas de Montañas en el Perú,” Tech. rep., FAO–MINAM (Ministerio del Ambiente, Perú).
- TELLO, J. C. (1939): “Origen y Desarrollo de las Civilizaciones Prehistóricas Andinas,” in *Actas y Trabajos. XXVII Congreso Internacional de Americanistas*, Lima, vol. 1, 589–720.
- THE ECONOMIST (2004): “Diversity and Development,” July 2004.
- UNITED NATIONS (2010): “Beyond Conflict: Ensuring Lasting Peace in Africa,” *Africa Renewal*, December 2010.
- VOIGTLÄNDER, N. AND H.-J. VOTH (2012): “Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany,” *The Quarterly Journal of Economics*, 127, 1339–1392.
- VOLLMER, G. (1967): *Bevölkerungspolitik und Bevölkerungsstruktur im Vizekönigreich Peru zu Ende der Kolonialzeit, 1741-1821*, Berlin.

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

——— (2002): “Los Indios y la Conquista Española,” in *América Latina en la Época Colonial*, Barcelona: Crítica, vol. 1, 153–185.

ZUIDEMA, R. T. AND D. POOLE (1982): “Los Límites de los Cuatro Suyus Incaicos en el Cuzco,” *Bulletin de l’Institut Français d’Études Andines*, 11, 83–89.

TABLE 1: Ethnic Diversity and Measures of Surname Heterogeneity (1605-1780)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Variable:							
	S Index	S Index	S Index	S Index	D Index	D Index	D Index	D Index
Panel A: Baseline								
Ethnic diversity (dummy)	0.536** [0.208]	0.456** [0.180]	0.543** [0.213]	0.528** [0.205]	0.507*** [0.167]	0.466** [0.178]	0.552** [0.217]	0.484** [0.223]
Panel B: Drop surnames of frequency 1								
Ethnic diversity (dummy)	0.487** [0.205]	0.418** [0.157]	0.503** [0.191]	0.487** [0.188]	0.481*** [0.170]	0.446** [0.175]	0.521** [0.210]	0.455** [0.215]
Panel C: Grouped surnames								
Ethnic diversity (dummy)	0.507** [0.211]	0.426** [0.182]	0.507** [0.212]	0.501** [0.203]	0.488*** [0.169]	0.445** [0.177]	0.527** [0.214]	0.468** [0.219]
Observations	66	66	66	66	66	66	66	66
Ln total pop. (1605-1780)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
% Non-native surnames (1605-1780)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Geography	No	No	Yes	Yes	No	No	Yes	Yes
Ln dist to colonial mine	No	No	Yes	Yes	No	No	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	Yes	No	No	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. The dependent variable is one measure of surname heterogeneity, either the Shannon index (S index) or one minus the Herfindahl index (D index), constructed using individuals with native first surname. Ethnic diversity is a dummy variable that takes value 1 if there is more than one tribe within a 10km buffer from the parish capital, and 0 otherwise. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Columns (2) and (6) add the log total number of individuals found in baptism records and the share of individuals with non-native first surname (1605-1780). All variables except ethnic diversity are standardized to have mean 0 and standard deviation 1.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 2: Balance Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Variable:							
	Mean Elevation	Variation in Elevation	Mean Caloric Suit.	Variation in Caloric Suit.	Ln Dist. to River	Ln Pop. (~ 1575)	Ln Dist. to Mine	Av. Within- Group H
Ethnic div (dummy)	0.110 [0.104] (0.125)	0.174 [0.116] (0.135)	-0.028 [0.113] (0.136)	0.032 [0.118] (0.131)	-0.066 [0.114] (0.127)	-0.045 [0.114] (0.131)	-0.048 [0.115] (0.137)	0.141 [0.112] (0.133)
Observations	336	336	336	336	336	336	336	336

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets; Conley standard errors corrected for spatial dependence with a distance cutoff of 50km in parentheses. Each column indicates a dependent variable. The table reports the coefficient on the ethnic diversity dummy variable, taking value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of the within-group heterogeneity index, computed using the area share of each ethnic group within the 10km buffer as weight. All variables except ethnic diversity are standardized to have mean 0 and standard deviation 1.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 3: Correlates of Within-Group Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Variable:							
	Mean Elevation	Variation in Elevation	Mean Caloric Suit.	Variation in Caloric Suit.	Ln River Density	Group Area	Ln Group Area	Ln Pop. Density (~ 1532)
Within-Group H	-0.109 [0.103]	0.658*** [0.106]	0.139 [0.107]	0.319*** [0.080]	0.154 [0.113]	-0.152 [0.122]	-0.105 [0.145]	0.188 [0.149]
Observations	49	49	49	49	49	49	49	48

Notes. OLS estimates. The unit of observation is the ethnic group. Robust standard errors in brackets. Each column indicates a dependent variable. All variables are standardized to have mean 0 and standard deviation 1. Regressions include a quadratic polynomial in longitude-latitude of the ethnic group's centroid.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 4: Overall Effect of Ethnic Diversity

	(1)	(2)	(3)	(4)	(5)	(6)
Contemporary Living Standards (AES)						
Ethnic diversity (dummy)	-0.205*** [0.076]	-0.222*** [0.065]	-0.221*** [0.065]	-0.212*** [0.065]	-0.159*** [0.060]	-0.168** [0.066]
Observations	336	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	No
Ln tributary pop. (~ 1575)	No	No	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contemporary Living Standards (AES)								
Ethnic diversity (dummy)	-1.064*** [0.328]	-1.131*** [0.281]	-1.168*** [0.284]	-1.145*** [0.280]	-0.967*** [0.263]	-0.931*** [0.259]	-0.809*** [0.259]	-0.713** [0.287]
Ethnic div \times Av. Within-Group H	1.258*** [0.459]	1.335*** [0.390]	1.392*** [0.393]	1.370*** [0.385]	1.186*** [0.375]	1.182*** [0.370]	1.129*** [0.365]	0.974** [0.410]
Observations	336	336	336	336	336	336	336	336
p -value for joint significance	0.001	0.000	0.000	0.000	0.000	0.001	0.008	0.045
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Ln tributary pop. (\sim 1575)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No	No
Religious Order FE	No	No	No	No	No	Yes	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of the within-group heterogeneity index, computed using the area share of each ethnic group within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces. The p -value refers to the joint significance of ethnic diversity terms.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 6: Robustness I - Interaction Effect of Av. Variation in Elevation and in Pre-Colonial Caloric Suitability

	(1)	(2)	(3)	(4)	(5)	(6)
Contemporary Living Standards (AES)						
Ethnic diversity (dummy)	-0.747*** [0.269]	-0.725*** [0.265]	-0.876*** [0.256]	-0.779*** [0.248]	-0.701*** [0.268]	-0.705*** [0.261]
Ethnic div \times Av. Within-Group H	1.229*** [0.388]	1.117*** [0.374]	1.020** [0.398]	1.024*** [0.360]	1.068*** [0.393]	1.046*** [0.373]
Ethnic div \times Av. Variation in Elevation	-0.355 [0.332]	-0.143 [0.323]			-0.348 [0.332]	-0.156 [0.321]
Ethnic div \times Av. Variation in Caloric Suit.			0.169 [0.233]	0.122 [0.192]	0.182 [0.203]	0.114 [0.193]
Observations	336	336	336	336	336	336
Ecclesiastical Jurisd. FE	Yes	No	Yes	No	Yes	No
Colonial Province FE	No	Yes	No	Yes	No	Yes
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. Density	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	Yes	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of the within-group heterogeneity index, computed using the area share of each ethnic group within the 10km buffer as weight. Similarly, average variation in elevation and in pre-1500CE caloric suitability refer to the weighted average of group-level characteristics, normalized as to take value 1 for the group with the highest value. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). The following control variables vary at the parish level. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 7: Robustness II - Interaction Effect of Av. Land Area, Population Density and Maize Suitability

	(1)	(2)	(3)	(4)	(5)	(6)
Contemporary Living Standards (AES)						
Ethnic diversity (dummy)	-0.741** [0.315]	-0.702** [0.321]	-0.680** [0.273]	-0.685** [0.270]	-0.837*** [0.260]	-0.752*** [0.260]
Ethnic div × Av. Within-Group H	0.953** [0.409]	0.991** [0.420]	0.932** [0.375]	1.029*** [0.371]	1.138*** [0.355]	1.075*** [0.346]
Ethnic div × Av. Land Area	-0.001 [0.003]	-0.001 [0.003]				
Ethnic div × Ln Av. Population Density (~ 1532)			0.163** [0.079]	0.101 [0.080]		
Ethnic div × Av. Suit. for Maize					-0.185 [0.198]	-0.078 [0.201]
Observations	336	336	336	336	336	336
Ecclesiastical Jurisd. FE	Yes	No	Yes	No	Yes	No
Colonial Province FE	No	Yes	No	Yes	No	Yes
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. Density	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	Yes	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of the within-group heterogeneity index, computed using the area share of each ethnic group within the 10km buffer as weight. Similarly, average land area ($km^2/1000$), the log of average population density by 1532 and average suitability for maize refer to the weighted average of group-level characteristics. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). The following control variables vary at the parish level. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 8: Robustness III - Av. Group Characteristics, Majority Group FE and Excluding Coastal Groups

	(1)	(2)	(3)	(4)	(5)	(6)
Contemporary Living Standards (AES)						
Ethnic diversity (dummy)	-0.777*** [0.246]	-0.744*** [0.248]	-0.913*** [0.257]	-0.731*** [0.260]	-0.843*** [0.312]	-0.743** [0.314]
Ethnic div \times Av. Within-Group H	0.972*** [0.350]	1.043*** [0.350]	1.201*** [0.361]	1.088*** [0.372]	1.136*** [0.430]	1.060** [0.436]
Observations	336	336	336	336	301	301
Av. Group Characteristics	Yes	Yes	No	No	No	No
Majority Group FE	No	No	Yes	Yes	No	No
Excluding Coastal Groups	No	No	No	No	Yes	Yes
Ecclesiastical Jurisd. FE	Yes	No	Yes	No	Yes	No
Colonial Province FE	No	Yes	No	Yes	No	Yes
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. Density	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	Yes	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of the within-group heterogeneity index, computed using the area share of each ethnic group within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). The vector of average group characteristics includes the weighted average of group area, mean elevation, variation in elevation, mean pre-1500CE caloric suitability, variation in pre-1500CE caloric suitability, log river density, and log population density by 1532, computed using the area share of each ethnic group within the 10km buffer as weight. The majority group refers to the ethnic group with the highest area share within the 10km buffer. Columns (5)-(6) exclude parishes in which the ethnic group with the highest area share within the 10km buffer is a coastal group. The following control variables vary at the parish level. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 9: Mid-Term Outcomes: Structural Change and Literacy Rate (1876 Population Census)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Variable:							
	Share of Emp. in Primary Sector		Share of Emp. in Secondary Sector		Share of Emp. in Tertiary Sector		Literacy Rate	
Ethnic diversity (dummy)	0.406*** [0.150]	0.395*** [0.134]	-0.219* [0.124]	-0.231** [0.116]	-0.187*** [0.060]	-0.164*** [0.054]	-0.245*** [0.084]	-0.130** [0.051]
Ethnic div \times Av. Within-Group H	-0.531** [0.219]	-0.529*** [0.187]	0.290* [0.165]	0.310** [0.148]	0.240*** [0.091]	0.219*** [0.081]	0.270** [0.115]	0.143** [0.070]
Observations	282	282	282	282	282	282	282	282
Mean Dep. Var.	0.791	0.791	0.137	0.137	0.072	0.072	0.151	0.151
Geography	No	Yes	No	Yes	No	Yes	No	Yes
Ln tributary pop. (\sim 1575)	No	Yes	No	Yes	No	Yes	No	Yes
Ln distance to colonial mine	No	Yes	No	Yes	No	Yes	No	Yes
Demographic controls 1792	No	Yes	No	Yes	No	Yes	No	Yes
Ecclesiastical Jurisd. FE	No	Yes	No	Yes	No	Yes	No	Yes
Religious Order FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. Outcomes in columns (1)-(6) refer to the share of male population in employment. In columns (7) and (8), the outcome is the literacy rate of the male population (those who can read and/or write). Regressions are weighted by total male population in 1876. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 10: Cultural Transmission Mechanism: Contemporary Associations

	(1)	(2)	(3)	(4)
	Dep. Variable:			
	Presence of Neigh. Associations (2002)		Share of Land Managed by Agr. Associations (1994)	
Ethnic diversity (dummy)	-0.327** [0.151]	-0.313** [0.154]	-0.365** [0.151]	-0.393** [0.155]
Ethnic div \times Av. Within-Group H	0.476** [0.237]	0.470** [0.235]	0.431** [0.211]	0.459** [0.218]
Observations	333	333	334	334
Mean Dep. Var.	0.129	0.129	0.295	0.295
Geography	No	Yes	No	Yes
Ln tributary pop. (~ 1575)	No	Yes	No	Yes
Ln distance to colonial mine	No	Yes	No	Yes
Demographic controls 1792	No	Yes	No	Yes
Ecclesiastical Jurisd. FE	No	Yes	No	Yes
Religious Order FE	No	Yes	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. In columns (1)-(2), the dependent variable is a dummy indicating the presence of neighborhood associations (2002 national registry of municipalities). In columns (3)-(4), the dependent variable is the share of land managed by agricultural associations (1994 agricultural census). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 11: Cultural Transmission or Economic Complementarities?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ln(number of occupations) in 1876						Contemporary	
	Primary Sector		Secondary Sector		Tertiary Sector		Living Standards (AES)	
High minority	0.043 [0.138]	0.131 [0.141]	0.546*** [0.175]	0.524*** [0.175]	0.709*** [0.205]	0.581*** [0.182]	0.506** [0.200]	0.447** [0.196]
High majority	0.032 [0.107]	-0.011 [0.115]	0.440* [0.251]	0.220 [0.279]	0.307* [0.173]	0.122 [0.198]	0.170 [0.144]	0.329** [0.149]
High min. × High maj.	-0.032 [0.210]	-0.245 [0.233]	-0.381 [0.344]	-0.762** [0.358]	-0.734** [0.349]	-0.871** [0.402]	0.145 [0.307]	0.134 [0.294]
Observations	97	97	97	97	97	97	117	117
Mean number of occupations	2.990	2.990	12.268	12.268	7.433	7.433		
Geography	No	Yes	No	Yes	No	Yes	No	Yes
Ln tributary pop. (~ 1575)	No	Yes	No	Yes	No	Yes	No	Yes
Ln distance to colonial mine	No	Yes	No	Yes	No	Yes	No	Yes
Demographic controls 1792	No	Yes	No	Yes	No	Yes	No	Yes
Ecclesiastical Jurisd. FE	No	Yes	No	Yes	No	Yes	No	Yes
Religious Order FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Regressions for the subsample of parishes with ethnic diversity. High majority is a dummy variable indicating whether the within-group heterogeneity value of the ethnic majority (ethnicity with the highest area share within the 10km buffer) is above the 75th percentile, and 0 otherwise. High minority refers to the analogous dummy variable for the ethnic minority (ethnicity with the lowest area share within the 10km buffer). Outcomes in columns (1)-(6) refer to $\log(1+\text{number of occupations})$ from the 1876 population census; available for 97 parishes. Columns (7) and (8) report the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

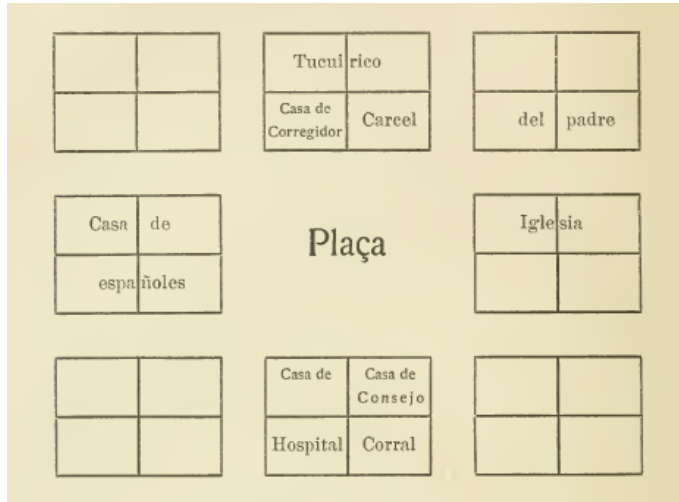
A Appendix - Historical

Native groups. The following table displays the names of native groups used in the analysis (49 groups). Alternative forms of native names in parentheses are from Rowe (1946) and Tello (1939). Columns present the area share of each production zone within the ethnicity's homeland, following the classification of Pulgar Vidal (1941). Groups from Rowe (1946) that were not under the Viceroyalty of Peru, and thus were not covered by the census of Taboada y Lemos (1791-95), are not part of the analysis. Most of these groups were under the jurisdiction of the Audiencia of Charcas, in the Viceroyalty of Río de la Plata, at the time of the census. In particular, the following groups are not covered: Aymaran groups of *Pacasa* or *Pacaje*, *Caranga* or *Caranca*, *Charca*, *Quillaca* or *Quillagua*, *Omasuyo*, and *Collahuaya* (all in Bolivia today); non-Aymaran groups of *Cochapampa*, *Yampará*, *Chicha*, *Lipe*, and *Uru* (all in Bolivia today); *Tarapacá* (Chile today); and the *Lupaca* and *Colla* Aymaran groups (both in the Department of Puno, Peru, today). The *Moyopampa* group, in the Amazonian region of Pulgar Vidal (1941), also lies outside the area of interest. Finally, the groups of *Tarata* and *Calva* do not intersect with any parish buffer.

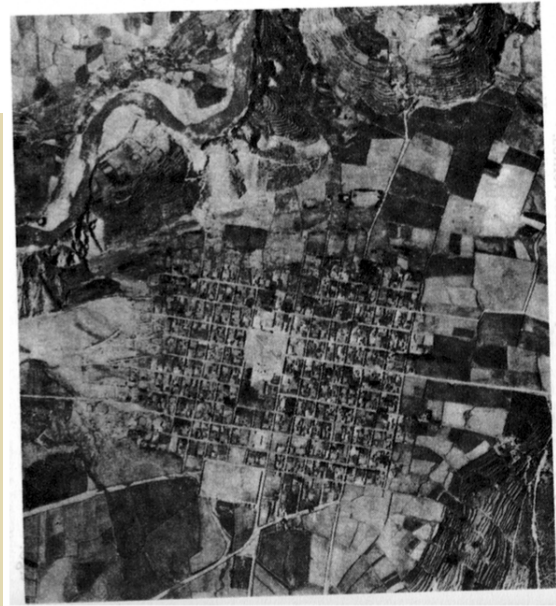
Tribes and Provinces of the Inca Empire (*circa 1530*)

Name	Yunga	Quechua	Suni	Puna	Janca
<i>Angará (Ankara)</i>	0.0294	0.2057	0.2768	0.4856	0.0025
<i>Arequipa (Ariquepa, Ariquipay)</i>	0.1554	0.4725	0.2087	0.1358	0.0277
<i>Atavillo (Atavillo, Atabillo)</i>	0.0695	0.1459	0.0642	0.6682	0.0522
<i>Ayavaca (Ayabaca, Ayauaca, Ayawaka)</i>	0.9422	0.0578	-	-	-
<i>Aymar� (Aymaraes, Aymarays)</i>	0.0037	0.1868	0.3063	0.5024	0.0008
<i>Cajamarca (Caxamarca, Cazamalca, Cassamarca, Kaxamarka)</i>	0.2368	0.5581	0.1867	0.0184	-
<i>Cajatampo (Caxatambo)</i>	0.0748	0.2009	0.1326	0.5231	0.0685
<i>Cana (Kana)</i>	-	0.0092	0.2188	0.6915	0.0805
<i>Canchi (Kanchi)</i>	-	0.0176	0.1233	0.4664	0.3926
<i>Caruma</i>	0.0100	0.0712	0.0602	0.6960	0.1627
<i>Cavana (Cabana, Cauana)</i>	0.0293	0.1058	0.1169	0.5095	0.2385
<i>Cavina (Cavina, Cabina, Cavi�a, Cavi�a, Cabi�a, Kawina)</i>	-	0.1391	0.3580	0.4789	0.0241
<i>Chachapoya (Chacha)</i>	0.2324	0.5294	0.2064	0.0318	-
<i>Chanca (Changa, Chanka)</i>	0.2334	0.4006	0.2247	0.1385	0.0028
<i>Chilque (Chillque, Chilqui, Chillke)</i>	0.0013	0.3197	0.3814	0.2966	0.0010
<i>Chimu</i>	0.7966	0.1614	0.0394	0.0026	-
<i>Chinchaycocha (Chinchaykocha)</i>	0.0036	0.0931	0.1738	0.7279	0.0017
<i>Choclococha</i>	-	0.0071	0.0726	0.8145	0.1057
<i>Chocorvo (Chocoruo, Chocorbo, Chucurpu, Chukurpu)</i>	0.0099	0.0851	0.1288	0.7253	0.0509
<i>Chumpivilca (Chumbivilca, Chumbivilca, Chumpi-willka)</i>	-	0.0255	0.1694	0.6054	0.1997
<i>Collagua (Kollawa)</i>	0.0002	0.0111	0.0331	0.7617	0.1939
<i>Conchucos (Conchucos, Konchucos)</i>	0.0650	0.2635	0.2262	0.3522	0.0932
<i>Contisuyo (Condesuyo, Cuntisuyo, Condes, Kontisuyo)</i>	0.0240	0.1137	0.0724	0.6159	0.1740
<i>Cotapampa (Cotabamba, Kotapampa)</i>	0.0424	0.3976	0.2558	0.3038	0.0004
<i>Cusco (Cuzco, Cozco, Inca, Inga)</i>	0.0221	0.2260	0.2859	0.4180	0.0480
<i>Huacrachuco (Huacrachucu, Wakrachuko)</i>	0.0349	0.3595	0.3836	0.2221	-
<i>Huamachuco (Guamachuco, Huamachucu, Wamachuko)</i>	0.1567	0.3846	0.2479	0.2103	0.0005
<i>Huamali (Guamali)</i>	0.0343	0.2594	0.3247	0.3817	-
<i>Huambo (Guambo, Wambo)</i>	0.4111	0.4643	0.1243	0.0003	-
<i>Huanca (Guanca, Wanka)</i>	0.0594	0.2058	0.2943	0.4349	0.0057
<i>Huancapampa (Huancabamba, Guancabamba, Wankapampa)</i>	0.6222	0.3661	0.0117	-	-
<i>Huayla (Guayla, Huaylla, Wayla)</i>	0.1539	0.3266	0.1751	0.3359	0.0085
<i>Hu�nuco (Guanuco, Huanucu, Wanuku)</i>	0.0009	0.1393	0.3557	0.5025	0.0017
<i>Lare (Lari)</i>	0.3341	0.4772	0.1688	0.0199	-
<i>Ocro (Okro)</i>	0.1521	0.2159	0.1121	0.4537	0.0662
<i>Omasayo (Omasuyo, Vmasuyo)</i>	-	0.1115	0.2177	0.6690	0.0018
<i>South Paracas</i>	0.6913	0.2006	0.0533	0.0450	0.0098
<i>Parinacocha (Parihuacocha)</i>	0.0568	0.2364	0.2295	0.4752	0.0021
<i>Paucartampo (Paucartambo, Paucartampu)</i>	0.0210	0.3169	0.3643	0.2977	-
<i>Pinco (Pinko)</i>	-	0.1000	0.1794	0.6470	0.0737
<i>Quechua (Quichua, Quichiua, Kichiwa)</i>	0.1560	0.3479	0.2223	0.2666	0.0072
<i>Rucana (Lucana, Rukana)</i>	0.0052	0.1348	0.2798	0.5799	0.0002
<i>Sora</i>	0.0400	0.4635	0.3254	0.1711	-
<i>Tarma (Tarama)</i>	0.0017	0.1034	0.2220	0.6728	-
<i>Ubina</i>	-	0.0027	0.0457	0.9043	0.0473
<i>Vilcapampa (Vilcabamba)</i>	0.4852	0.4012	0.1013	0.0122	-
<i>Vilcas (Vilcas, Bilcas, Vilcshuaman, Vilcasguaman)</i>	0.0497	0.3615	0.2920	0.2907	0.0060
<i>Yanahuara (Yanaguara)</i>	0.0925	0.4487	0.2012	0.2520	0.0056
<i>Yauyo</i>	0.0612	0.1677	0.1021	0.5802	0.0888

Notes. The classification of elevation zones follows Pulgar Vidal (1941): 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]. Alternative forms of tribal names are from Rowe (1946) and Tello (1939). Following Rowe (1946), *Cusco* refers to Inca tribal lines either by blood or by privilege. The *Chimu kingdom* refers to the individual valleys of: *Tumbez (Tumbes, Tumpiz, Tumpiz)*; *Chira; Piura; Olmos; Lambayeque; Pacasmayo or Jequetepeque (Xequetepeque)*; *Chicama; Chim  (Chimo)*; *Vir ; Chao; Chimbote or Santa (Sancta)*; *Nepe a or Guambacho; Casma; Huarmey (Guarmey)*; and *Parmunca (Paramonga)*. *South Paracas* refers to the individual valleys of: *Huaura (Guaura)*; *Chanca; Lima (Rima, Rimac)*; *Lurin; Chilca (Chilca)*; *Mala; Huarco (Guarco)*; *Chincha; Pisco; Ica; Nazca; Acari; Yauca; Atico; Caraveli; Oco a; Camana; Quilca (Quilca)*; *Tampo (Tambo)*; *Moquehua (Moquegua)*; *Locumba; Sama*; and *Arica*.



(a) Model of village in 1567



(b) Modern aerial view of Yanque (Collaguas)



(c) Google Earth view of Yanque (Collaguas)

Spanish Planning

Notes. Subfigure (a) shows the model of village designed in 1567 by the Spanish official Juan de Matienzo, Matienzo (1910)[1567]. Subfigure (b) shows a modern aerial view of Yanque, Collaguas, created as a result of the resettlement policy of the 16th century (Servicio Aerofotográfico Nacional del Perú, in Medina (1993)). Subfigure (c) shows the contemporary Google Earth view of Yanque.

B Appendix - Surnames

Identification of native surnames. Hispanic and foreign surnames are excluded 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.⁵⁶ I complement Basque surnames using a list of surnames provided by the Real Academia de la Lengua Vasca.

In order to identify native surnames, I constructed a dictionary of linguistic roots from the Quechuan and Aymaran language families. There is not a 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.

⁵⁶Check suggestions by the Biblioteca Nacional de España (BNE) here. The list of surnames in Carraffa and Carraffa (1920–1963) can also be accessed through The Library of Congress.

Baptism Record from FamilySearch.org

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

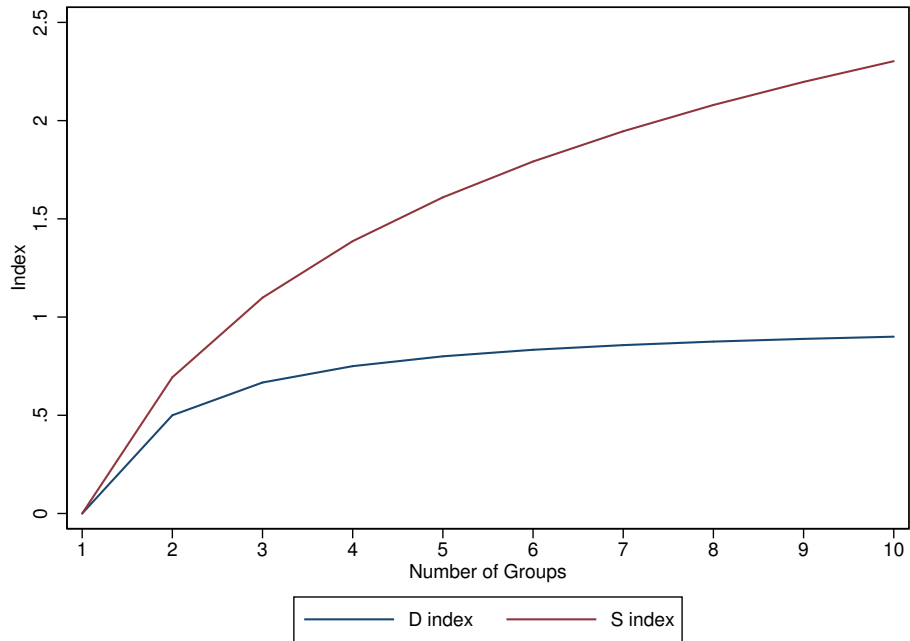
Cita

"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.

Descriptive Statistics - Dataset of Baptisms

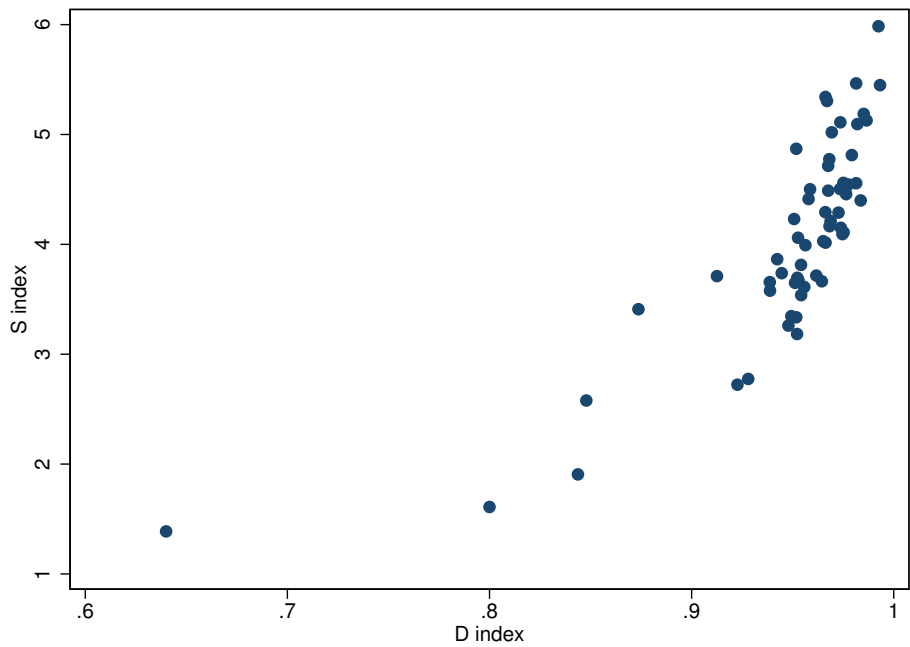
	# Individuals	# Parishes	Mean # Individuals	Median # Individuals
<i>By period</i>				
[1605, 1625]	848	8	106	16.5
(1625, 1650]	5,039	19	265.211	145
(1650, 1675]	8,033	30	267.767	125.5
(1675, 1700]	19,195	40	479.875	209
(1700, 1725]	17,947	49	366.265	197
(1725, 1750]	21,172	46	460.261	205
(1750, 1780]	40,106	63	636.603	184
<i>Full period</i>				
	112,340	66	1,702.121	561.5

Notes. Statistics refer to individuals with native first surname. The first panel reports statistics by time period: number of individuals, number of parishes, and number of individuals in the mean and median parish. The second panel reports analogous statistics for the full period.



S and D Indices

Notes. The figure plots each index for different numbers of groups of equal size. For example, when the x-axis is 5, each index is computed for $K = 5$ groups, each of them with a proportion of 0.2 in the population.



S and D Indices in the Data

Notes. The figure plots the relationship between the D and S indices in the data (66 parishes), computed using the sample of individuals with native first surname.

C Appendix - Data

Geographic characteristics

Mean elevation. Average elevation across all grid cells with centroid within the unit of analysis. 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 (in meters) based on information from the NASA Shuttle Radar Topographic Mission.

Variation in elevation. Standard deviation of elevation across all grid cells with centroid within the unit of analysis. 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 (in meters) based on information from the NASA Shuttle Radar Topographic Mission.

Mean pre-1500CE caloric suitability. Average caloric suitability in the pre-1500 period across all grid cells with centroid within the unit of analysis. Source: author's computation using the Caloric Suitability Index (Galor and Özak 2016). They provide 5 arc-minute raster data with average potential crop yield given the set of available crops in the pre-1500 period.

Variation in pre-1500CE caloric suitability. Standard deviation of caloric suitability in the pre-1500 period across all grid cells with centroid within the unit of analysis. Source: author's computation using the Caloric Suitability Index (Galor and Özak 2016). They provide 5 arc-minute raster data with average potential crop yield given the set of available crops in the pre-1500 period.

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

Characteristics of Ethnic groups

Group area. Total land area of the ethnic group (km^2). Source: author's computation after georeferencing the approximate extent of ethnic groups at the time of the Spanish conquest (Rowe 1946).

Ln river density. Natural log of total river length (km , only permanent rivers) within the ethnic homeland divided by total land area (km^2). Source: author's computation using

watercourse and inland water area features from version 10.0 of the Seamless Digital Chart of the World.

Ln population density of ethnic group by 1532. Natural log of approximate population divided by total land area (km^2). Source: author's computation using population figures in Cook (1982, 2010). I add the first estimate of tributary population between 1532 and 1575 for all population centers within the ethnic homeland. The resulting population estimates cover 48 out of the 49 ethnic groups used in the analysis.

Suitability for maize. Average potential caloric yield of maize across all grid cells with centroid within the ethnic homeland. Source: author's computation using 5 arc-minute raster data with the potential caloric yield of maize under rain-fed low-input agriculture provided by Galor and Özak (2016).

Contemporary outcomes

Ln light intensity per capita. Natural log of 0.01 plus average light intensity per capita for the period 2000-2003. The average sum of light intensity values across all grid cells with centroid within the 10km buffer (yearly average for 2000-2003) is divided by total population within the same buffer (year 2000). By adding a small constant before computing the logarithm, parishes for which light intensity is reported to be zero (86 parishes) are not dropped from the analysis. The minimum estimate of population for parishes with zero nightlight is of 440 individuals, approximately. Sources: average cloud free coverages of the DMSP-OLS Nighttime Lights Time Series (30 arc-second raster data from satellite F15) produced by the NOAA's National Geophysical Data Center; 30 arc-second raster data with population counts for the year 2000 come from version 4.10 of the Gridded Population of the World (Center for International Earth Science Information Network – CIESIN).

Indicator for 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 rather than to own consumption is above the median value in the sample ($P_{50} = 0.03$), and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI).

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

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

Presence of neighborhood associations. Dummy variable indicating the presence of neighborhood associations in 2002. Source: *Registro Nacional de Municipalidades*, provided by the National Institute of Statistics (INEI).

Share of land managed by agr. associations. Share of land managed by farmers in agricultural associations. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI).

Mid-term outcomes

Share of employment by economic sector. Share of male employment in the primary, secondary and tertiary sectors. Source: author's computation using data from the 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878).

Ln number of occupations by economic sector. Natural log of 1 plus number of occupations among the male population, separately for the primary, secondary and tertiary sectors. Source: author's computation using data from the 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878).

Literacy rate. Literacy rate of the male 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).

Control variables

Demographic characteristics 1792. Separate variables for the shares of indigenous, mestizo, slave and Spanish population. Source: census of Viceroy Gil de Taboada y Lemos (1791-95), published in Vollmer (1967). It provides information at the parish level on the number of individuals by caste category (indigenous, mestizo, free, slave and Spanish), separately by gender, as well as on the number of individuals related with the ecclesiastical system, including priests.

Ln tributary population by 1575. Natural log of approximate tributary population by 1575. Source: Cook (1982, 2010). The data exist for 128 out of 336 parishes used in the analysis; for the remaining parishes the data is imputed using the mean tributary population of the colonial province to which the parish belonged.

Ln distance to colonial mine. Natural log of the geodesic distance (*km*) from the parish capital to the closest mining center during the colonial period (i.e., either Huancavelica or Potosí mines). Source: author's computation using coordinates of mining centers from Dell (2010).

Ln population density. Natural log of total population per square kilometer. Source: 1993 national population and housing census, conducted by the National Institute of Statistics (INEI).

Ecclesiastical jurisdiction. Colonial bishopric to which the parish belonged (Lima, Arequipa, Huamanga, Trujillo and Cuzco). Source: Unanue, J. H, (1797): *Guía Política, Eclesiástica y Militar del Virreynato del Perú para el Año de 1797*.

Religious order. One of the regular orders (Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús), more than one religious order, or secular clergy if no specific order was in charge of the parish during most of the colonial period. Sources: Lissón Chávez (1943), de Armas Medina (1953), Córdoba y Salinas (1957)[1651], and García (1997).

Colonial province. Administrative province (*partido*) to which the parish belonged by the end of the 18th century (44 provinces). Source: census of Viceroy Gil de Taboada y Lemos (1791-95), published in Vollmer (1967).

D Appendix - Figures

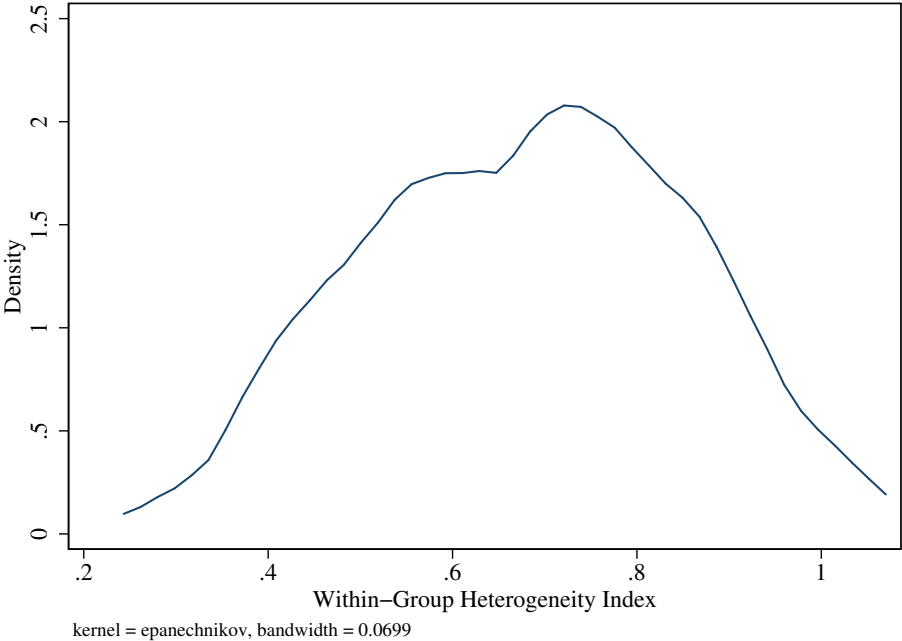


FIGURE A.1: Density of Within-Group Heterogeneity

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

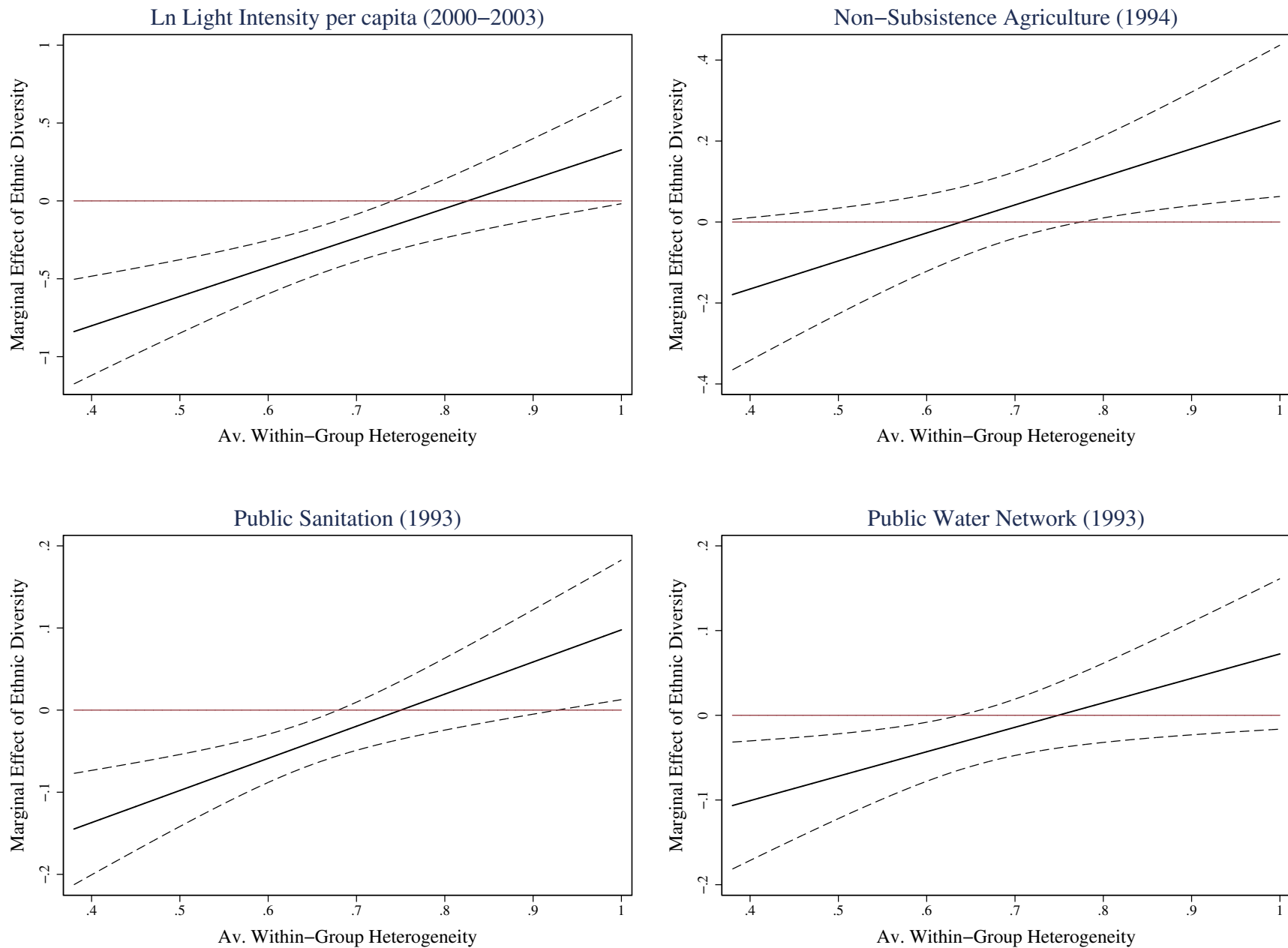
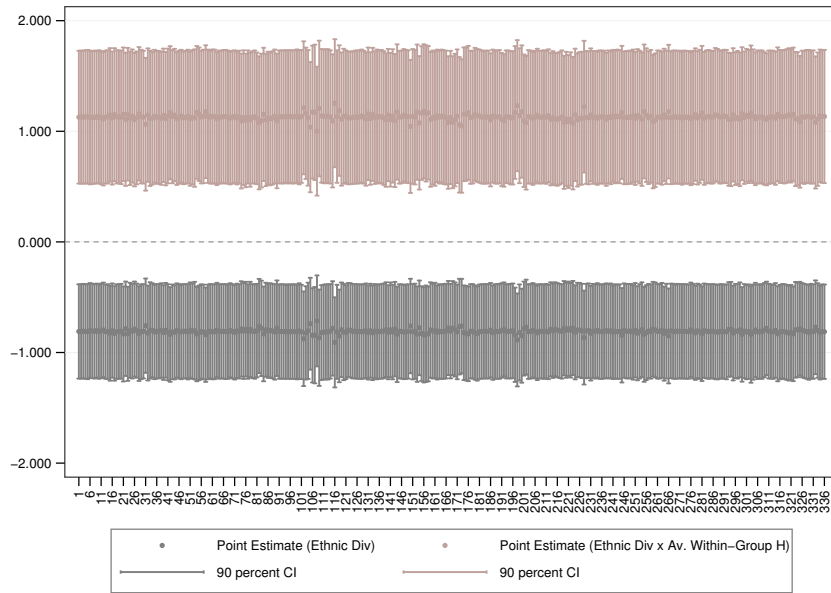
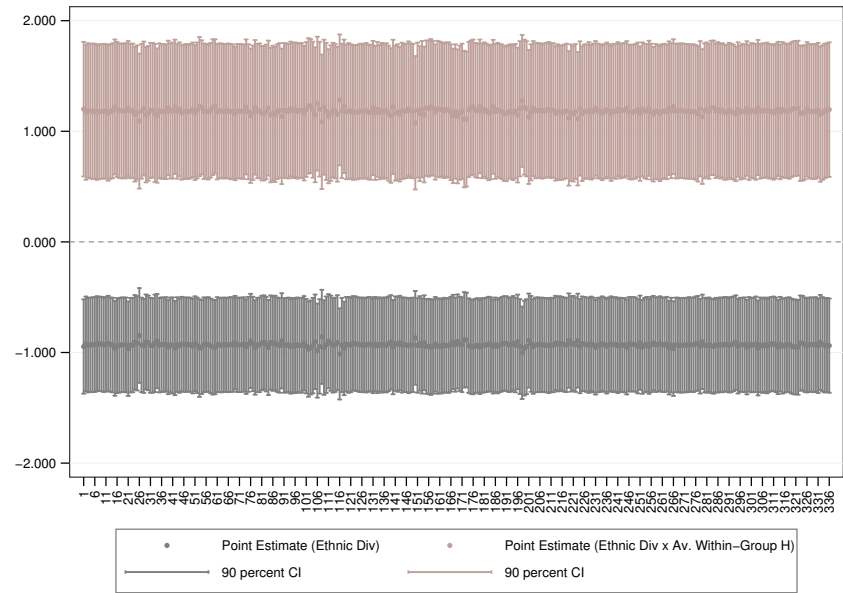


FIGURE A.2: Marginal Effect of Ethnic Diversity - Contemporary Outcomes

Notes. Estimated marginal effects of ethnic diversity on contemporary outcomes (ln average light intensity per capita, indicator for non-subsistence agriculture, share of dwellings with access to public sanitation, and share of dwellings with access to the public water network) after baseline controls, religious order fixed effects, and ecclesiastical jurisdiction fixed effects. Dashed lines represent 90% confidence intervals.



(a) Colonial Province FE



(b) Ecclesiastical Jurisdiction FE

FIGURE A.3: Robustness - Excluding Parishes One by One

Notes. The figure displays point estimates and 90% confidence intervals for the coefficients on ethnic diversity and on ethnic diversity \times average within-group heterogeneity after excluding parishes one by one (i.e., each regression in the x-axis excludes one parish). Specifically, the figure reports the standardized average effect size (AES) across contemporary outcomes (ln average light intensity per capita, indicator for non-subsistence agriculture, share of dwellings with access to public sanitation, and share of dwellings with access to the public water network) after baseline controls, religious order fixed effects, and either colonial province (panel a) or ecclesiastical jurisdiction (panel b) fixed effects.

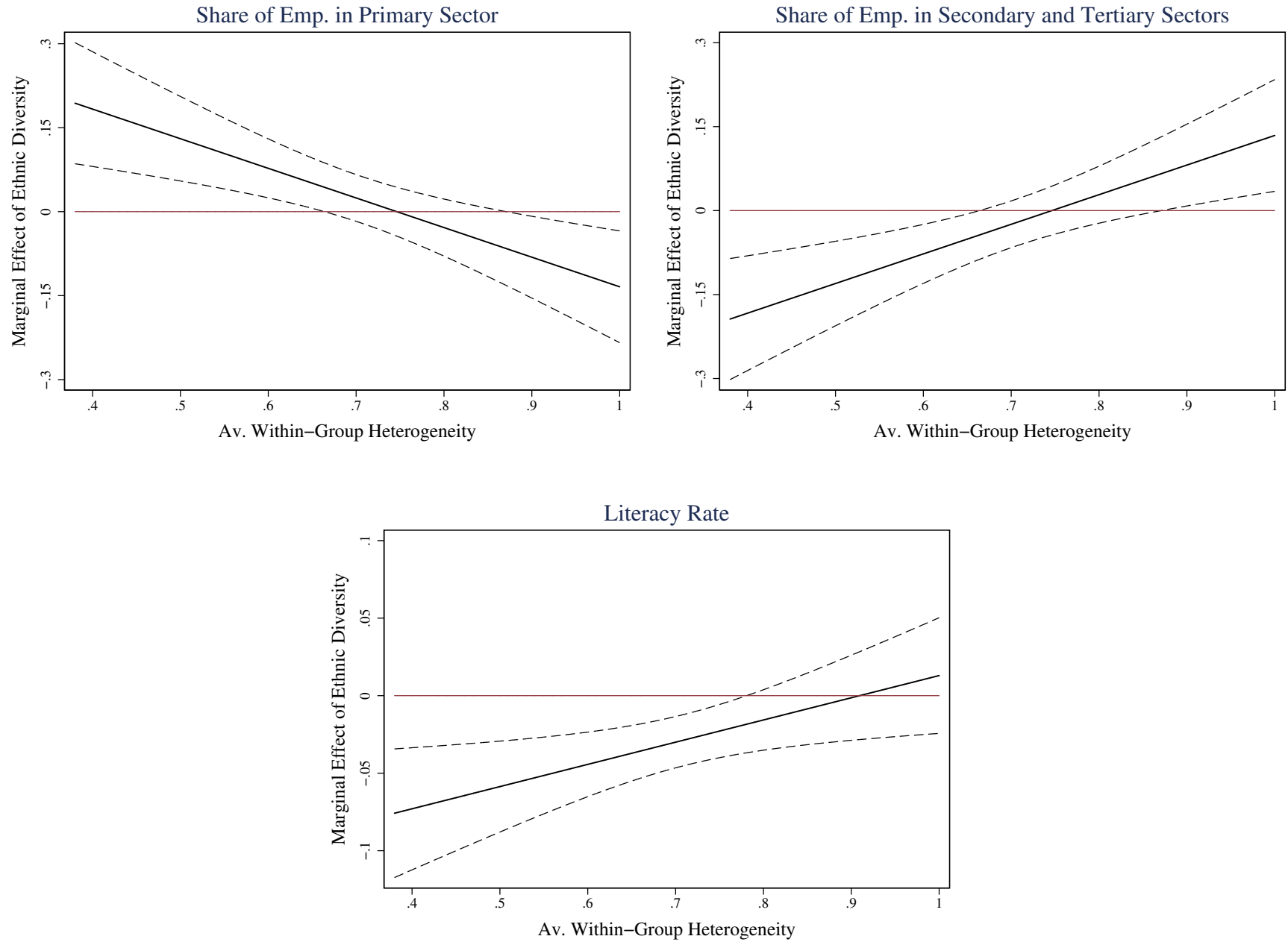


FIGURE A.4: Marginal Effect of Ethnic Diversity - Midterm Outcomes (1876)

Notes. Estimated marginal effects of ethnic diversity on mid-term outcomes (share of male employment in the primary sector, share of male employment in secondary and tertiary sectors, and literacy rate of the male population by 1876) after baseline controls, religious order fixed effects, and ecclesiastical jurisdiction fixed effects. Dashed lines represent 90% confidence intervals.

E Appendix - Tables

- Table A.1: summary statistics.
- Table A.2: correlation at the ethnic group level between within-group heterogeneity and crop diversity.
- Tables A.3 and A.4: overall effect of ethnic diversity for individual outcomes (log light intensity per capita, indicator for non-subsistence agriculture, access to public sanitation and access to the public water network).
- Table A.5: AES of ethnic fractionalization for the whole sample.
- Tables A.6 and A.7: main result for individual outcomes (log light intensity per capita, indicator for non-subsistence agriculture, access to public sanitation and access to the public water network).
- Table A.8: AES of average within-group heterogeneity for the whole sample and separately for parishes with and without ethnic diversity.
- Table A.9: main result using ethnic fractionalization instead of the ethnic diversity dummy variable.
- Table A.10: main result measuring within-group heterogeneity as 1-Herfindahl index.
- Table A.11: main result using a robust version of the ethnic diversity dummy variable.
- Table A.12: main result using a living standards index (standardized score of the first principal component for the four outcomes) as dependent variable.
- Table A.13: Conley standard errors.
- Table A.14: additional robustness checks related with the Inca period.
- Table A.15: main result excluding parishes for which the 10km buffer intersects an ethnic border and a *corregimiento* border at the same time.
- Table A.16: placebo analysis using *corregimiento* borders instead of ethnic borders.
- Table A.17: evidence from household-level survey data on participation in neighborhood associations and unions.
- Table A.18: evidence on potential inter-ethnic unions during the colonial period.
- Table A.19: evidence from household-level survey data on identity formation.
- Table A.20: Table 11 controlling for log population in 1876 and using a fractionalization index to measure diversity of occupations.

TABLE A.1: Descriptive Statistics

	Whole sample			Ethnic div = 1			Ethnic div = 0		
	obs.	mean	sd	obs.	mean	sd	obs.	mean	sd
Contemporary variables									
Total light intensity per capita (2000-2003)	336	0.035	0.063	117	0.019	0.026	219	0.044	0.074
Non-subsistence agriculture (share of agr. units, 1994)	336	0.105	0.158	117	0.094	0.138	219	0.111	0.168
Public sanitation (share of dwellings, 1993)	336	0.122	0.169	117	0.090	0.135	219	0.138	0.182
Public water network (share of dwellings, 1993)	336	0.238	0.212	117	0.211	0.192	219	0.252	0.221
Total number of dwellings (1993)	336	1755.542	2716.989	117	1658.880	2244.647	219	1807.183	2941.980
Population density (1993)	336	40.054	93.976	117	32.918	55.442	219	43.865	109.058
Geography variables									
Mean elevation	336	3432.450	670.576	117	3480.341	529.292	219	3406.865	734.882
Variation in elevation	336	459.078	182.124	117	479.734	188.314	219	448.043	178.185
Mean caloric suitability	336	696.912	938.347	117	679.946	911.994	219	705.976	954.068
Variation in caloric suitability	336	614.852	656.938	117	628.710	700.915	219	607.448	633.730
Distance to river (km)	336	3.676	5.494	117	3.615	6.000	219	3.709	5.218
Latitude	336	-12.270	2.892	117	-12.493	2.602	219	-12.151	3.034
Longitude	336	-74.690	2.467	117	-74.482	2.419	219	-74.801	2.490

Notes. The unit of observation is the parish. Total light intensity per capita (from 30 arc-second raster data) refers to the sum of light intensity across all grid cells with centroid within a 10km buffer from the parish capital divided by total population within the same buffer. Elevation (from 30 arc-second raster data) and pre-1500CE caloric suitability (from 5 arc-minute raster data) measures refer to the mean or variation (sd) across all grid cells with centroid within the 10km buffer. Longitude and latitude correspond to the parish capital. Non-subsistence agriculture, access to public infrastructure, the total number of dwellings and population density in 1993 refer to the corresponding contemporary district. Distance to river is the geodesic distance from the parish capital to the closest permanent river. Data sources and definitions for all variables are reported in Appendix C.

Parishes by Elevation Zone of the Parish Capital

	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

TABLE A.2: Within-Group Heterogeneity and Crop Diversity

	(1)	(2)	(3)	(4)
	Dep. Variable: Crop Diversity			
Within-Group H	0.207*** [0.076]	0.288*** [0.067]	0.494*** [0.167]	0.492*** [0.179]
Observations	49	49	49	49
Adjusted R-squared	0.094	0.272	0.581	0.558
Ethnic group area	No	Yes	Yes	Yes
Geography	No	No	Yes	Yes
Zone profile FE	No	No	No	Yes

Notes. OLS estimates. The unit of observation is the ethnic group. Robust standard errors in brackets. The index of crop diversity is defined as 1/Herfindahl index, normalized as to take value 1 for the group with the highest crop diversity. Data on native crops come from the 2012 national agricultural census. Geography controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log river density (total river length/group area), and a quadratic polynomial in longitude-latitude of the ethnic group's centroid. The zone profile FE accounts for ethnic groups with presence of the same elevation zones within the group's area.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Crop data. The 2012 national agricultural census, conducted by the National Institute of Statistics (INEI), provides information on the area planted with each crop at the moment of the census for the whole national territory. In order to apply the census form, the territory within each district was divided into several units (*Sector de Enumeración Agropecuario*, SEA), each one comprising on average 100 agricultural producers. Using the map of territorial units provided by INEI, I assign agricultural producers to the homelands of ethnic groups. I focus on native crops (see Tapia 2013) and compute a measure of crop diversity at the ethnic group level using the share of area planted with each crop group (roots and tubers, cereals, fruit trees, vegetables, and pulses) within the ethnic homeland.

TABLE A.3: Overall Effect of Ethnic Diversity on Local Economic Activity

	(1)	(2)	(3)	(4)	(5)
Dep. Variable: Ln Light Intensity per capita (2000-2003)					
Ethnic diversity (dummy)	-0.383*** [0.097]	-0.364*** [0.089]	-0.366*** [0.090]	-0.365*** [0.088]	-0.327*** [0.088]
Mean Dep. Var.	-3.613	-3.613	-3.613	-3.613	-3.613
Dep. Variable: Non-Subsistence Agriculture (1994)					
Ethnic diversity (dummy)	-0.017 [0.017]	-0.038** [0.015]	-0.038** [0.015]	-0.036** [0.015]	-0.027* [0.014]
Mean Dep. Var.	0.500	0.500	0.500	0.500	0.500
Observations	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	No	No	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. The dependent variables are log average light intensity per capita (2000-2003) in the first panel and the share of agricultural producers devoting most of the harvest to sale or trade in local markets (1994 agricultural census) in the second panel. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.4: Overall Effect of Ethnic Diversity on Access to Public Infrastructure

	(1)	(2)	(3)	(4)	(5)
Dep. Variable: Sanitation (1993)					
Ethnic diversity (dummy)	-0.048*** [0.018]	-0.050*** [0.016]	-0.049*** [0.016]	-0.046*** [0.016]	-0.037** [0.015]
Mean Dep. Var.	0.122	0.122	0.122	0.122	0.122
Dep. Variable: Water Network (1993)					
Ethnic diversity (dummy)	-0.041* [0.023]	-0.035* [0.021]	-0.036* [0.021]	-0.034 [0.021]	-0.022 [0.020]
Mean Dep. Var.	0.238	0.238	0.238	0.238	0.238
Observations	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes
Ln pop. Density	No	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	No	No	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. The dependent variables are the share of occupied dwellings with access to public sanitation in the first panel and the share of occupied dwellings with access to the public network of water supply in the second panel, both from the 1993 population and housing census. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.5: Overall Effect of Ethnic Fractionalization

	(1)	(2)	(3)	(4)	(5)
Contemporary Living Standards (AES)					
Ethnic fractionalization	-0.565*** [0.186]	-0.522*** [0.160]	-0.519*** [0.160]	-0.486*** [0.161]	-0.340** [0.146]
Observations	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	No	No	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic fractionalization is a Herfindahl index computed using the area share of each ethnic group within a 10km buffer from the parish capital. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.6: Ethnic Diversity, Within-Group Heterogeneity and Local Economic Activity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Variable: Ln Light Intensity per capita (2000-2003)							
Ethnic diversity (dummy)	-1.856*** [0.470]	-1.855*** [0.416]	-1.852*** [0.415]	-1.875*** [0.414]	-1.706*** [0.417]	-1.641*** [0.421]	-1.323*** [0.467]
Ethnic div \times Av. Within-Group H	2.192*** [0.657]	2.191*** [0.592]	2.185*** [0.593]	2.215*** [0.586]	2.016*** [0.592]	1.973*** [0.601]	1.638** [0.667]
Mean Dep. Var.	-3.613	-3.613	-3.613	-3.613	-3.613	-3.613	-3.613
Dep. Variable: Non-Subsistence Agriculture (1994)							
Ethnic diversity (dummy)	-0.849*** [0.320]	-0.702*** [0.267]	-0.631*** [0.243]	-0.629*** [0.238]	-0.458** [0.233]	-0.443* [0.232]	-0.401* [0.227]
Ethnic div \times Av. Within-Group H	1.110*** [0.424]	1.008*** [0.353]	0.935*** [0.326]	0.919*** [0.321]	0.686** [0.327]	0.693** [0.328]	0.648** [0.325]
Mean Dep. Var.	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Observations	336	336	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	No	No	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No
Religious Order FE	No	No	No	No	No	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The dependent variables are log average light intensity per capita (2000-2003) in the first panel and a 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 value (1994 agricultural census) in the second panel. In the second panel, regressions are weighted by the total number of dwellings. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.7: Ethnic Diversity, Within-Group Heterogeneity and Access to Public Infrastructure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Variable: Sanitation (1993)							
Ethnic diversity (dummy)	-0.462** [0.207]	-0.359*** [0.106]	-0.365*** [0.104]	-0.348*** [0.096]	-0.301*** [0.096]	-0.294*** [0.092]	-0.136* [0.076]
Ethnic div \times Av. Within-Group H	0.553* [0.299]	0.475*** [0.168]	0.490*** [0.159]	0.475*** [0.148]	0.399*** [0.144]	0.391*** [0.140]	0.199* [0.109]
Mean Dep. Var.	0.122	0.122	0.122	0.122	0.122	0.122	0.122
Dep. Variable: Water Network (1993)							
Ethnic diversity (dummy)	-0.382* [0.202]	-0.290*** [0.106]	-0.266** [0.109]	-0.262** [0.104]	-0.216** [0.102]	-0.216** [0.099]	-0.080 [0.091]
Ethnic div \times Av. Within-Group H	0.444 [0.294]	0.384** [0.163]	0.359** [0.162]	0.361** [0.154]	0.286* [0.149]	0.289* [0.148]	0.111 [0.127]
Mean Dep. Var.	0.238	0.238	0.238	0.238	0.238	0.238	0.238
Observations	336	336	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. Density	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	No	No	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No
Religious Order FE	No	No	No	No	No	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The dependent variables are the share of occupied dwellings with access to public sanitation in the first panel and the share of occupied dwellings with access to the public network of water supply in the second panel, both from the 1993 population and housing census. Regressions are weighted by the total number of dwellings. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.8: Within-Group Heterogeneity and Contemporary Development

	(1)	(2)	(3)	(4)	(5)	(6)
Contemporary Living Standards (AES)						
	Whole Sample		Ethnic div = 1		Ethnic div = 0	
Dummy (Av. Within-Group H > 75 percentile)	0.242*** [0.086]	0.204** [0.098]	0.517*** [0.160]	0.719*** [0.227]	0.164 [0.105]	0.053 [0.135]
Ln pop. Density	No	No	No	No	No	No
Dummy (Av. Within-Group H > 75 percentile)	0.131 [0.081]	0.120 [0.089]	0.369*** [0.140]	0.431** [0.195]	0.051 [0.100]	-0.003 [0.134]
Ln pop. Density	Yes	Yes	Yes	Yes	Yes	Yes
Observations	336	336	117	117	219	219
Ecclesiastical Jurisd. FE	Yes	No	Yes	No	Yes	No
Colonial Province FE	No	Yes	No	Yes	No	Yes
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	Yes	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.9: Ethnic Fractionalization, Within-Group Heterogeneity and Contemporary Development

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Contemporary Living Standards (AES)						
Ethnic fractionalization	-1.944** [0.797]	-2.460*** [0.658]	-2.523*** [0.664]	-2.399*** [0.662]	-1.935*** [0.649]	-1.835*** [0.633]	-1.565** [0.648]
Ethnic frac \times Av. Within-Group H	2.047* [1.116]	2.895*** [0.942]	3.000*** [0.955]	2.859*** [0.949]	2.388** [0.940]	2.338** [0.913]	2.081** [0.932]
Observations	336	336	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	No	No	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No
Religious Order FE	No	No	No	No	No	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic fractionalization is a Herfindahl index computed using the area share of each ethnic group within a 10km buffer from the parish capital. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.10: Ethnic Diversity, Within-Group Heterogeneity (1-Herf.) and Contemporary Development

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Contemporary Living Standards (AES)							
Ethnic diversity (dummy)	-1.307*** [0.419]	-1.368*** [0.375]	-1.428*** [0.382]	-1.402*** [0.376]	-1.176*** [0.355]	-1.088*** [0.347]	-1.022*** [0.357]
Ethnic div \times Av. Within-Group H (1-Herf.)	1.845*** [0.686]	1.923*** [0.611]	2.028*** [0.623]	1.998*** [0.612]	1.708*** [0.590]	1.615*** [0.576]	1.645*** [0.584]
Observations	336	336	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	No	No	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No
Religious Order FE	No	No	No	No	No	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity (1-Herfindahl index), computed using the area share of each ethnic group within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.11: Ethnic Diversity (Robust Dummy), Within-Group Heterogeneity and Contemporary Development

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Contemporary Living Standards (AES)						
Ethnic diversity (robust dummy)	-0.795** [0.357]	-1.113*** [0.319]	-1.149*** [0.321]	-1.089*** [0.317]	-0.901*** [0.295]	-0.846*** [0.291]	-0.664** [0.299]
Ethnic div \times Av. Within-Group H	0.836* [0.498]	1.336*** [0.444]	1.394*** [0.448]	1.327*** [0.442]	1.126*** [0.421]	1.074*** [0.416]	0.883** [0.423]
Observations	336	336	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	No	No	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No
Religious Order FE	No	No	No	No	No	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital (robust version: area share of ethnic group within buffer ≥ 10 percent), and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.12: Living Standards Index (PCA)

	(1)	(2)	(3)	(4)
Dep. Variable: Living Standards Index				
Ethnic diversity (dummy)	-1.277*** [0.384]	-1.021** [0.414]	-1.741*** [0.490]	-1.039** [0.438]
Ethnic div \times Av. Within-Group H	1.592*** [0.547]	1.391** [0.586]	2.291*** [0.743]	1.420** [0.618]
Observations	336	336	336	336
Population Weights	No	No	Yes	Yes
Ecclesiastical Jurisd. FE	Yes	No	Yes	No
Colonial Province FE	No	Yes	No	Yes
Geography	Yes	Yes	Yes	Yes
Ln pop. Density	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The dependent variable is contemporary living standards, as proxied by the standardized score of the first principal component for the following outcomes: ln average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Columns (3) and (4) are weighted by the total number of dwellings in 1993. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.13: Ethnic Diversity, Within-Group Heterogeneity and Contemporary Development
(Conley Standard Errors)

	(1)	(2)	(3)	(4)
Dep. Variable: Living Standards Index				
Distance Cutoff:				
	30km		50km	
Ethnic diversity	-1.277*** [0.370]	-1.021*** [0.372]	-1.277*** [0.354]	-1.021*** [0.354]
Ethnic div \times Av. Within-Group H	1.592*** [0.521]	1.391*** [0.524]	1.592*** [0.495]	1.391*** [0.500]
Observations	336	336	336	336
Ecclesiastical Jurisd. FE	Yes	No	Yes	No
Colonial Province FE	No	Yes	No	Yes
Geography	Yes	Yes	Yes	Yes
Ln pop. Density	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes

Notes. OLS estimates. The unit of observation is the parish. Conley standard errors corrected for spatial dependence with a distance cutoff of 30km (columns 1-2) or 50km (columns 3-4) in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The dependent variable is contemporary living standards, as proxied by the standardized score of the first principal component for the following outcomes: ln average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.14: Inca-Period Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)
Contemporary Living Standards (AES)						
Ethnic diversity (dummy)	-0.869*** [0.253]	-0.815*** [0.259]	-0.915*** [0.257]	-0.801*** [0.256]	-0.878*** [0.308]	-0.669** [0.314]
Ethnic div \times Av. Within-Group H	1.067*** [0.359]	1.133*** [0.364]	1.141*** [0.369]	1.098*** [0.362]	1.212*** [0.427]	0.938** [0.435]
Observations	336	336	336	336	275	275
Inca region (<i>suyu</i>) FE	Yes	Yes	No	No	No	No
Ln distance to Inca site	No	No	Yes	Yes	No	No
Excluding potential Inca-affected groups	No	No	No	No	Yes	Yes
Ecclesiastical Jurisd. FE	Yes	No	Yes	No	Yes	No
Colonial Province FE	No	Yes	No	Yes	No	Yes
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. Density	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	Yes	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of the within-group heterogeneity index, computed using the area share of each ethnic group within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). The following control variables vary at the parish level. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces. Columns 1-2 include fixed effects accounting for the four major Inca regions (*suyus*) into which the empire was divided (Zuidema and Poole 1982). Columns 3-4 control for log distance to the closest pre-Hispanic site, including Inca administrative centers, connected by the Inca road network (Qhapaq Ñan); see *Guía de Identificación y Registro del Qhapaq Ñan* (Ministerio de Cultura). The fact that the study is focused on parishes located in the highland region alleviates concerns regarding potential Inca resettlements from the north to the south coast of Peru (Bongers et al. 2020). Columns 5-6 exclude parishes with presence of other groups potentially affected by Inca resettlements according to the historical literature (Rowe 1946; de La Espada 1881). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.15: Ethnic Diversity, Within-Group Heterogeneity and Contemporary Development

(*Corregimiento* Robustness)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Contemporary Living Standards (AES)						
Ethnic diversity (dummy)	-1.255*** [0.386]	-1.297*** [0.327]	-1.347*** [0.326]	-1.333*** [0.323]	-1.194*** [0.293]	-1.120*** [0.294]	-0.974*** [0.279]
Ethnic div \times Av. Within-Group H	1.597*** [0.541]	1.627*** [0.446]	1.701*** [0.444]	1.684*** [0.437]	1.544*** [0.418]	1.458*** [0.421]	1.346*** [0.396]
Observations	293	293	293	293	293	293	293
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (\sim 1575)	No	No	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No
Religious Order FE	No	No	No	No	No	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Regressions exclude parishes for which the buffer intersects an ethnic border and a *corregimiento* border at the same time. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.16: Ethnic Diversity, Within-Group Heterogeneity and Contemporary Development

(*Corregimiento* Placebo)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Contemporary Living Standards (AES)							
<i>Corregimiento</i> Dummy	0.781*	-0.178	-0.178	-0.059	0.051	-0.002	-0.095
	[0.409]	[0.348]	[0.349]	[0.349]	[0.321]	[0.321]	[0.287]
<i>Corregimiento</i> Dummy \times Av. Within- <i>Corregimiento</i> H	-1.187**	0.150	0.143	0.009	-0.057	0.069	0.183
	[0.537]	[0.456]	[0.456]	[0.456]	[0.426]	[0.429]	[0.383]
Observations	336	336	336	336	336	336	336
Geography	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln pop. density	No	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	No	No	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	No	No	No	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	No	Yes	Yes	No
Religious Order FE	No	No	No	No	No	Yes	Yes
Colonial Province FE	No	No	No	No	No	No	Yes

Notes. The unit of observation is the parish. Robust standard errors in brackets. The *Corregimiento* dummy takes value 1 if there is more than one *corregimiento* within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-*Corregimiento* H refers to the weighted average of within-*corregimiento* heterogeneity, computed using the area share of each *corregimiento* within the 10km buffer as weight. The table presents the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Colonial province fixed effects account for 44 administrative provinces.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.17: Ethnic Diversity, Within-Group Heterogeneity and Participation

(2004-2017 Household-Level Survey Data)

	(1)	(2)	(3)	(4)
Dep. Variable: Participation in				
	Neigh. Associations		Unions	
Ethnic diversity (dummy)	-0.111*	-0.084**	-0.068*	-0.072**
	[0.062]	[0.039]	[0.036]	[0.031]
Ethnic div \times Av. Within-Group H	0.133	0.112**	0.103**	0.108**
	[0.085]	[0.056]	[0.040]	[0.043]
Observations	52,494	52,494	52,494	52,494
Number of districts	280	280	280	280
Mean Dep. Var.	0.075	0.075	0.062	0.062
Year FE	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
Geography	No	Yes	No	Yes
Ln tributary pop. (\sim 1575)	No	Yes	No	Yes
Ln distance to colonial mine	No	Yes	No	Yes
Demographic controls 1792	No	Yes	No	Yes
Ecclesiastical Jurisd. FE	No	Yes	No	Yes
Religious Order FE	No	Yes	No	Yes

Notes. OLS estimates. Standard errors in brackets are clustered at the parish level. The dependent variable takes value 1 if some household member participates in a certain type of association, and 0 otherwise (ENAHO household survey). Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. All regressions include year fixed effects, personal characteristics of the household head (gender, age, age squared, years of schooling, civil status and mother tongue) and log total associations per capita in the district. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.18: Within-Group Heterogeneity and Potential Inter-Group Unions (1605-1870)

	(1)	(2)	(3)	(4)	(5)
Dep. Variable:					
Share of Unions between Linguistically-Distant Individuals					
Ethnic diversity (dummy)	-0.463*** [0.151]	-0.436*** [0.155]	-0.341* [0.176]	-0.352 [0.216]	-0.536** [0.218]
Ethnic div \times Av. Within-Group H	0.653** [0.273]	0.622** [0.282]	0.554* [0.309]	0.618* [0.342]	0.679* [0.350]
Observations	61	61	61	61	61
Mean Dep. Var.	0.297	0.297	0.297	0.297	0.297
Ln total pop. (1605-1780)	No	Yes	Yes	Yes	Yes
Potential partners (1605-1780)	No	Yes	Yes	Yes	Yes
Geography	No	No	Yes	Yes	Yes
Ln distance to colonial mine	No	No	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	No	No	No	Yes	Yes
Religious Order FE	No	No	No	Yes	Yes
Population Weights	No	No	No	No	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. The dependent variable is the share of unions with Levenshtein distance \geq 75th percentile ($L_{75} = 7$). Column (2) controls for the log total number of individuals (1605-1780) and for the mean share of potential partners, defined as people in the parish with whom the individual has L distance $\geq L_{75}$. Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy. Column (5) is weighted by total population in 1605-1780.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.19: Ethnic Diversity, Within-Group Heterogeneity and Identity Formation

(2004-2017 Household-Level Survey Data)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable: <i>Which group do you identify most with?</i>						
	Administrative Region		Ethnicity or Race		Native Community	
Ethnic diversity (dummy)	-0.190** [0.086]	-0.190*** [0.060]	0.017 [0.021]	0.038** [0.017]	0.255*** [0.094]	0.167*** [0.060]
Ethnic div \times Av. Within-Group H	0.228** [0.116]	0.240*** [0.082]	-0.023 [0.028]	-0.046* [0.024]	-0.307** [0.130]	-0.232*** [0.083]
Observations	52,494	52,494	52,494	52,494	52,494	52,494
Number of districts	280	280	280	280	280	280
Mean Dep. Var.	0.504	0.504	0.039	0.039	0.214	0.214
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Geography	No	Yes	No	Yes	No	Yes
Ln tributary pop. (~ 1575)	No	Yes	No	Yes	No	Yes
Ln distance to colonial mine	No	Yes	No	Yes	No	Yes
Demographic controls 1792	No	Yes	No	Yes	No	Yes
Ecclesiastical Jurisd. FE	No	Yes	No	Yes	No	Yes
Religious Order FE	No	Yes	No	Yes	No	Yes

Notes. OLS estimates. Standard errors in brackets are clustered at the parish level. The dependent variable takes value 1 if the individual answers that she/he is most identified with a certain group (administrative region/ethnicity or race/native community), and 0 otherwise (ENAH0 household survey). Ethnic diversity is a dummy variable that takes value 1 if there is more than one group within a 10km buffer from the parish capital, and 0 otherwise. Av. Within-Group H refers to the weighted average of within-group heterogeneity, computed using the area share of each ethnic group within the 10km buffer as weight. All regressions include year fixed effects and personal characteristics (gender, age, age squared, years of schooling, civil status and mother tongue). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.20: Cultural Transmission or Economic Complementarities?

	(1)	(2)	(3)	(4)	(5)	(6)
	Primary Sector		Secondary Sector		Tertiary Sector	
	Ln(occupations)	Frac occupations	Ln(occupations)	Frac occupations	Ln(occupations)	Frac occupations
High minority	0.130 [0.138]	-0.011 [0.069]	0.487*** [0.182]	0.098** [0.044]	0.551*** [0.192]	0.046 [0.071]
High majority	-0.011 [0.115]	-0.035 [0.077]	0.213 [0.279]	0.053 [0.050]	0.116 [0.199]	0.023 [0.056]
High min. × High maj.	-0.245 [0.233]	-0.006 [0.145]	-0.747** [0.360]	-0.127** [0.062]	-0.860** [0.409]	-0.176 [0.116]
Observations	97	97	97	97	97	97
Mean number of occupations	2.990	2.990	12.268	12.268	7.433	7.433
Ln male pop. 1876	Yes	No	Yes	No	Yes	No
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Ln tributary pop. (~ 1575)	Yes	Yes	Yes	Yes	Yes	Yes
Ln distance to colonial mine	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls 1792	Yes	Yes	Yes	Yes	Yes	Yes
Ecclesiastical Jurisd. FE	Yes	Yes	Yes	Yes	Yes	Yes
Religious Order FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. OLS estimates. The unit of observation is the parish. Robust standard errors in brackets. Regressions for the subsample of parishes with ethnic diversity. High majority is a dummy variable indicating whether the within-group heterogeneity value of the ethnic majority (ethnicity with the highest area share within the 10km buffer) is above the 75th percentile, and 0 otherwise. High minority refers to the analogous dummy variable for the ethnic minority (ethnicity with the lowest area share within the 10km buffer). Outcomes in columns (1)-(6) refer to $\log(1+\text{number of occupations})$ from the 1876 population census; available for 97 parishes. Columns (7) and (8) report the standardized average effect size (AES) for four outcomes: log average light intensity per capita (2000-2003), indicator for non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993). Geographic controls include mean elevation, standard deviation of elevation, mean pre-1500CE caloric suitability, standard deviation of pre-1500CE caloric suitability, log distance to river, and a quadratic polynomial in longitude-latitude. Demographic controls include separate variables for the shares of indigenous, mestizo, slave, and Spanish populations by 1792. The ecclesiastical jurisdiction is the colonial bishopric in charge of the parish (Lima, Arequipa, Huamanga, Trujillo, or Cuzco). Religious order fixed effects refer to Santo Domingo, La Merced, San Francisco, San Agustín, Compañía de Jesús, more than one order, or secular clergy.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.