Within-Group Heterogeneity in a Multi-Ethnic Society*

Miriam Artiles†

February 11, 2024

Abstract

Is ethnic diversity good or bad for economic development? Most studies find corrosive effects. I document that ethnic diversity need not spell poor development outcomes—a history of within-group heterogeneity can turn ethnic diversity into an advantage for long-run development. I collect data from a natural experiment of Peru's colonial history: the forced resettlement of native populations in the 16th century. This intervention forced together various ethnic groups into new jurisdictions. Where colonial officials concentrated populations with a history of within-group heterogeneity, who settled in complementary altitudes before colonization, ethnic diversity results in systematically lower costs and may even become advantageous.

JEL Codes: J15, N36, N56, O12, Z10, Z13

Keywords: Ethnic diversity, within-group heterogeneity, internal exchange, cooperation, culture, ethnic minorities, comparative development

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

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

1 Introduction

The effect of ethnic diversity on economic growth and development is a question of long-standing interest in economics. Following the initial work by Easterly and Levine (1997) and Alesina and Glaeser (2004), a large body of literature has examined the costs and benefits of ethnic diversity.[1] Most empirical studies find corrosive effects. When individuals within ethnic groups are homogeneous and groups differ in their preferences for policies or public goods, conflicting preferences can lead to inefficiencies in public good provision or to policy choices that may not benefit the entire society (e.g., Alesina, Baqir, and Easterly 1999; Miguel and Gugerty 2005). Inter-group tensions can also result in civil conflicts or exacerbate mistrust and lack of cooperation (e.g., Alesina and La Ferrara 2000; Fearon and Laitin 2003). On the other hand, some studies find that when ethnic groups differ in specializations, complementarities can sustain coexistence, facilitate inter-group trade, and generate economic gains (Jha 2013; Becker and Pascali 2019; Jedwab, Johnson, and Koyama 2019; Grosfeld, Sakalli, and Zhuravskaya 2020). While there is a general understanding that diversity brings opportunities and challenges, there is scarce evidence on which factors determine its positive or negative consequences. When is ethnic diversity good for economic development, and when is it bad?

I study whether the long-run effect of ethnic diversity on comparative development depends on exposure to within-group heterogeneity. Underlying previous literature on the effects of ethnic diversity is the assumption that individuals within ethnic groups tend to be homogeneous. However, ethnic groups are not necessarily homogeneous entities. Individuals within groups may differ along many dimensions, including preferences, economic activities or skills, as well as cultural and genetic traits (see Horowitz 1998; Ashraf and Galor 2013; Desmet, Ortuño-Ortín, and Wacziarg 2017). Recent empirical research shows that a deeper understanding of within-group heterogeneity can help shed light on the features that shape comparative economic growth and development (Ashraf and Galor 2013). However, whether this dimension helps explain the consequences of ethnic diversity remains to be explored.

This paper provides systematic evidence that exposure to within-group heterogeneity in complementary traits matters for understanding the long-run economic consequences of ethnic diversity. I analyze new data from a natural experiment of Peru's colonial history—the

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

forced resettlement of native populations in the 16th century. Unintentionally on the part of the Spanish colonizers, the resettlement forced together various ethnic groups into small-scale jurisdictions (parishes).[2] This intervention disrupted the internal settlement pattern, a key aspect of pre-colonial society. Before colonization, coethnic individuals settled in complementary altitudes and engaged in crop exchange to maximize the economic base:[3]

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

The resettlement accidentally altered the spatial distribution of coethnics growing altitude-specific crops, resulting in a setting where exposure to within-group heterogeneity was arguably orthogonal to ethnic diversity.[4] This unique scenario provides an opportunity to study the consequences of ethnic diversity across parishes with varying levels of past exposure to within-group heterogeneity. Were individuals from more heterogeneous ethnic groups better able to function in multi-ethnic societies? The answer is not obvious. Small-scale jurisdictions did not exist before colonization. Recent research shows that the benefits of ethnic diversity tend to flourish at the local level (Montalvo and Reynal-Querol 2021). The literature has also emphasized the positive role of local interactions (Desmet, Gomes, and Ortuño-Ortín 2020) and complementarities (Jha 2013, 2018) between ethnic groups. If, after being resettled, individuals with a history of within-group heterogeneity were more willing to engage with other ethnicities, mutually beneficial exchange from local inter-ethnic interactions might have become more frequent. Although trust tends to be higher among coethnics, individuals from more heterogeneous ethnic groups were already used to operating in diverse settings; they may have been better able to integrate with other ethnic groups.

[2]Throughout the paper, I use the term "ethnic group" to refer to the societies that coexisted in the Andean highlands of Peru before the Spanish conquest. I refer to the issue of ethnic identity in Section 3.

[3]This subsistence strategy is consistent with Michalopoulos (2012)'s idea that variation in geographic characteristics (e.g., elevation) may lead to specialization through the formation of zone-specific skills, applying the idea to specialization *within* ethnic groups. The internal economic organization of ethnic groups has received little attention in economics. See Fenske (2014) for supporting evidence that internal exchange in more ecologically diverse environments favored political centralization among pre-colonial African groups.

[4]This is not necessarily the case in general settings (e.g., on a global scale, ethnic diversity is negatively correlated with pre-colonial exposure to heterogeneity in specialized activities within groups; Figure A.1).

The first result of the paper documents the direct effect of ethnic diversity. Guided by the historical narrative, I geolocated parishes that were accidentally created close to spatial boundaries between ethnic groups. This narrative assumes that colonial officials were not fully aware of the vertical distribution of coethnics across space, making it unlikely for them to systematically consider ethnic boundaries when determining parish locations (Murra 1975). Given the vertical settlement pattern, geographic proximity to ethnic boundaries created quasi-random variation in ethnic diversity (Pease 1989). I provide empirical support for the historical narrative and validate ethnic diversity using surnames from colonial baptism records available for a subset of parishes. The results show a robust pattern when comparing contemporary living standards between parishes whose initial populations were ethnically diverse and those with an ethnically homogeneous founding population. On average, parishes with ethnic diversity tend to exhibit lower living standards in the long run. The results hold beyond geographic proximity to ethnic boundaries, highlighting the persistent consequences of forced diversity at the local level.

I then explore whether the overall effect of ethnic diversity depends on pre-colonial exposure to within-group heterogeneity. To compute a proxy for within-group heterogeneity, I mapped the distribution of complementary elevation zones within the homelands of ethnic groups. I validate this proxy using data from paleodietary reconstructions for a subset of the groups—the measure correlates with the consumption of carbon-enriched crops, likely coming from exchange between coethnics settled in complementary zones. The estimates show a negative coefficient on ethnic diversity and a positive coefficient on its interaction with average exposure to within-group heterogeneity. Where colonial officials concentrated individuals with a history of within-group heterogeneity, ethnic diversity results in systematically lower costs in terms of living standards and may even become advantageous. In parishes in the bottom 10 percent of exposure, ethnic diversity is systematically linked to decreased living standards (about 0.27 standard deviations). Conversely, in parishes in the top 10 percent, ethnic diversity is linked to increased living standards (about 0.21 standard deviations).

The estimated interaction effect persists when controlling for initial prosperity and geography. It also remains significant when considering administrative and ecclesiastical provinces, as well as proximity to the colonial mining *mita* (Dell 2010). Permutation-based p-values from randomization inference in the presence of heterogeneous treatment effects align with baseline results. Using a matching procedure to construct a sample of parishes with varying levels of exposure to within-group heterogeneity while ensuring similarity in other ethnic characteristics, such as pre-colonial political complexity and population density, shows consistent results.

In a setting that remains primarily agricultural, both historical and modern data support a complementarity between cultural and economic channels. First, the results align with the idea that the historical experience of within-group heterogeneity contributed to more open attitudes toward out-group members, facilitating inter-group interactions after forced resettlement. Despite having ethnically diverse initial populations, the data show a more integrated society, with a tendency for engaging in associational activities, non-subsistence agriculture, and local trade, where these populations had a historical experience of within-group heterogeneity sustained by internal exchange. Second, contemporary populations tend to perform better where the historical ethnic minority had a comparative advantage over the majority group. In turn, the results support the hypothesis that these ethnic minorities specialized in intermediary services related to agriculture (e.g., retailing), reducing inter-group competition and contributing to long-run coexistence. Consistently, mid-term data show a shift toward tertiary-sector activities, mainly driven by local trade, in a predominantly agricultural setting.[5]

This paper contributes to a substantial body of studies in development and political economy on the consequences of ethnic diversity. Most studies have documented negative effects (Alesina and La Ferrara 2005). In this literature, however, the role of within-group heterogeneity has been less explored. Ashraf and Galor (2013) conducted the first empirical study exploring the effect of this dimension on comparative development. Focusing on population heterogeneity, as proxied by genetic diversity across coethnics, the study finds a hump-shaped effect on economic development in both pre-colonial and modern times.[6] I focus on a genetically homogeneous region[7] where coethnic individuals settled in complementary altitudes and engaged in crop exchange during pre-colonial times. The results provide micro-level evidence that historical exposure to within-group heterogeneity in complementary traits can help us

[5]These results are not explained by pre-colonial crop diversity indirectly fostering occupation in non-agricultural sectors through the potential increased availability of skills (Fiszbein 2022).

[6]Underlying the positive side of this effect is the idea that the greater the degree of population heterogeneity is, the larger the set of potentially different traits among coethnics. If this heterogeneity translates into multiple specializations, gains from those that are complementary become more likely. Depetris-Chauvin and Özak (2020) provide evidence on the link between population heterogeneity and specializations across pre-modern societies.

[7]See Kemp, Tung, and Summar (2009), Valverde et al. (2016), and Nakatsuka et al. (2020) for genetic studies.

understand the long-run effects of ethnic diversity. Exploring potential mechanisms supports the idea that cultural transmission increased the likelihood of inter-ethnic interaction after resettlement. Being more willing to interact, the results then highlight the contribution of economic complementarities to sustaining long-run coexistence (Jha 2013; Becker and Pascali 2019; Jedwab, Johnson, and Koyama 2019; Grosfeld, Sakalli, and Zhuravskaya 2020).

These results contribute to studies showing that strategies to cope with environmental risk and adverse geography can shape culture (Nunn and Wantchekon 2011; Nunn and Puga 2012; Buggle and Durante 2021) and add to the literature on the long-run effects of cultural traits (e.g., Voigtländer and Voth 2012; Alesina, Giuliano, and Nunn 2013; Guiso, Sapienza, and Zingales 2016). Following Gennaioli and Rainer (2007) and Michalopoulos and Papaioannou (2013), the characteristics of pre-colonial ethnic groups have received increasing attention. The results suggest that further studying the internal economic organization of the groups can help us understand comparative development.[8] Section 2 summarizes the literature on ethnic diversity, Section 3 describes the historical context, Section 4 presents the empirical strategy, Section 5 explains the main results, Section 6 discusses mechanisms, and Section 7 concludes.

2 Literature on Ethnic Diversity and Hypothesis

A substantial body of literature has analyzed the consequences of ethnic diversity. The initial studies tended to emphasize its costs at various levels of analysis. Across countries and US localities, ethnic diversity has been associated with lower levels of economic growth, public good provision, and quality of government, as well as with greater political instability and civil conflict.[9] Using micro-level data, Miguel and Gugerty (2005) show that ethnic diversity is associated with lower public good provision in Kenya. Hjort (2014) focuses on the private sector, providing causal evidence for the effect of ethnic diversity on team productivity at

[8] Dippel (2014) has explored the internal political organization of the group. The author shows that when coethnic individuals form autonomous subpolities (i.e., they do not have a history of shared or centralized governance even though they are ethnically homogeneous), their forced coexistence in a centralized institutional system can negatively impact long-run development. See Esteban and Ray (2011) for a theoretical model of ethnic conflict in which coethnic individuals differ in attitudes and income. Moscona, Nunn, and Robinson (2020) provide evidence on segmentary lineage organization and conflict.

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

a flower plant in Kenya. The results show that teams of ethnically diverse workers are, on average, less productive than homogeneous teams. Additional evidence points toward a taste for discrimination against coworkers of different ethnic origin.

More recent papers have emphasized the role of local interactions. Desmet, Gomes, and Ortuño-Ortín (2020) provide cross-country evidence that local inter-ethnic interactions can contribute to weakening the costs of ethnic diversity for public good provision. Montalvo and Reynal-Querol (2021) have focused on the size of the unit of analysis, finding a positive relationship between ethnic diversity and economic growth at low levels of geographic aggregation. The authors find that a potential explanation for this result in the context of Africa is the increase in trade close to spatial boundaries between ethnic groups, suggesting ethnic specialization into complementary activities and inter-ethnic contact. Analyzing a voluntary resettlement program in Indonesia, Bazzi et al. (2019) show that inter-ethnic contact can promote societal integration in the long run.

These papers relate to recent studies on the positive role of inter-ethnic complementarities. Conditional on willingness to interact, the theoretical framework developed by Jha (2013, 2018) establishes that peaceful inter-ethnic coexistence can be sustained through the specialization of ethnic groups in complementary activities that are both costly to replicate and expropriate. Jha (2013) provides consistent empirical evidence on tolerance toward Muslims in South Asian medieval towns. Other studies have focused on anti-Semitism, with Jedwab, Johnson, and Koyama (2019) presenting evidence from the Black Death in Western Europe and Becker and Pascali (2019) from the Protestant Reformation in Germany. Consistently, Grosfeld, Sakalli, and Zhuravskaya (2020) show that during times of political stability in the Russian Empire, anti-Jewish pogroms were not prevalent in places where Jews specialized in intermediary occupations related to agriculture, providing insurance against economic shocks to peasants.

The existing literature has focused less on the potential role of within-group heterogeneity. This paper examines the importance of historical exposure to within-group heterogeneity in understanding the long-run consequences of ethnic diversity. If having a history of within-group heterogeneity increased the likelihood of inter-ethnic interaction after resettlement, it may have contributed to a higher frequency of mutually beneficial exchanges. The history of within-group heterogeneity, sustained by internal complementarities to reach a common goal (i.e., maximizing the economic base), may have shaped more open attitudes toward out-group members, in turn facilitating inter-ethnic interactions after forced resettlement.

3 Historical Background

Pre-colonial ethnic groups. By the time Spanish conquerors arrived, the Andean civilization comprised several coexisting groups that had been incorporated over the previous century into the Inca empire (1438-1525), such as *Chocorvos*, *Lucanas*, *Soras*, and *Chankas* (Tello 1939; Rowe 1946; Dulanto 2008). The 47 groups in my study region coexisted approximately after the disintegration of the *Wari* culture (ca. 1000) and before the Spanish conquest (ca. 1532).

How ethnically distinct were these groups from one another? In the study region, the term "ethnic group" was introduced by Murra (1975). The prevailing view is that the society comprised various groups with diverse linguistic roots (e.g., Rowe 1946; Murra 1975) and differentiated material cultures (e.g., Stanish 1989) following the collapse of *Wari* (Isbell 2010). Many groups, such as the *Atavillos* and *Chocorvos*, had their own language; for some groups (e.g., *Lucanas*) there is also anecdotal evidence that their languages coexisted with Quechua, the language of the Incas, during a period of indirect rule (Rowe 1946).[10] Group identity seems to have been reinforced by the absence of inter-group marriage. Specifically, the social unit is generally described as an endogamous group of several extended families with descent traced through the male line (Rowe 1946). The group usually claimed descent from a mythical ancestor, such as an animal or element of nature. This mythical kin was worshipped and sometimes honored with rites and sacrifices; see Garcilaso de la Vega (1960)[1609].[11]

Vertical settlement pattern. In the human ecology literature, the mountain environment of the Andean highlands is described as a vertical resource system (Brush 1976). Differences in elevation give rise to various microclimates within short distances, and each microclimate is,

[10] The issue of ethnic identity has been vaguely discussed in the literature, possibly due to the European perception of the region as culturally homogeneous despite the existence of ethnolinguistic differences at the time of Spanish contact. Charney (1998) argues that the use of the Spanish term "Indio" to collectively refer to all native peoples in official documents played a role in obscuring ethnic distinctions in the eyes of Europeans—"Indio" was not merely a label but the imposed *new ethnicity* for native individuals in the colonial legal system. Stanish (2001) points toward the interest of Inca and Spanish powers in promoting cultural unity via state propaganda.

[11] For example, the *Chankas* believed that they were closely connected to the Andean lion (*puma*). During festivities, they usually dressed in *puma* skins and adopted *puma* imagery. Bauer, Kellett, and Silva (2010) provide anecdotal evidence of public support for this identity.

in turn, suited to a different assortment of natural resources and crops.

Following the pioneering ethnohistoric work of Murra (1975), studies across various disciplines documented a pre-colonial subsistence strategy characterized by individuals from the same ethnic group simultaneously controlling different altitudes.[12] Specifically, Murra's model is often referred to as a zonal complementarity model (e.g., Stanish 1989; Aldenderfer 1993; Isbell and Silverman 2002). The group tried to maximize its economic base by establishing permanent settlements in vertically arranged zones (Murra 1975, 1995, 2002a,b). Since certain crops can only be grown at specific altitudes, these zones are interpreted as complements. By exchanging crops between populations settled in different zones, the group increased access to resources, thus maximizing total output at the group level.

Crop exchange seems to have been centralized at the group level, while maintaining ties to the extended family and the rest of the group (Pease 1989; Stanish 1989; Murra 2002b; Nash 2009).[13] Furthermore, Stern (1995, p. 76) suggests that "Andean rules of reciprocity and redistribution served to govern the exchanges ... 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." Crop sharing is also emphasized in native folklore. For example, Berezkin (2015)'s folklore catalog mentions the avaricious man motif—*A man does not share food with his wife or kinsfolk. He or his food is transformed (turns into a bird, into worms, etc.) in punishment*—in the *Conchucos*'s homeland, the same region that early chronicles describe as "very fertile and abundant, with many crops and resources that everyone has and sows" (Cieza de León 1962 [1553], p. 221).

Continuity after the Inca expansion. According to Murra's research, this settlement pattern already existed during pre-Inca times (Murra 1956, 1975). The Inca expansion (1438–1525) was achieved through the gradual conquest of pre-existing groups. The dominant view is that this led to a dynamic process of state formation whereby differentiated regions or provinces were sequentially created based on ethnic identity (Rowe 1946). Ethnohistoric research

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

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

suggests that Inca government was indirect, with each region governed by the leader of the corresponding ethnic group (Murra 1975, 2002b). This is a crucial characteristic of Inca rule that suggests the preservation of ethnic traits during this period.[14] Additionally, ethnic rulers were pushed to maintain control over their respective vertical zones in order to sustain the empire (Murra 1956, 1975). Rowe (1946) mapped the approximate extent of the groups at the time of the Spanish conquest based on archaeological evidence and early ethnohistoric accounts. The map was published in the second volume of the *Smithsonian Handbook of South American Indians* (1948)—a georeferenced version is presented in Figure 1.

The colonial resettlement. The contemporary administrative division of Peru has its origins in the colonial period. When Spanish Viceroy *Francisco Toledo* arrived in Peru in 1569, native populations still lived scattered along mountain slopes. This settlement pattern was perceived as an "obstacle" to tribute collection and religious indoctrination. In words of the Spanish official *J. Matienzo*, "the *indios*, for being isolated in *huaycos* and ravines, do not live in right order, and this is the main obstacle to be indoctrinated" (in Medina 1974a, p. 155). Between 1570 and 1575, *F. Toledo* mandated the forced reorganization of native populations into residential jurisdictions (*reducciones*). In turn, several jurisdictions were assigned to a single *doctrina*, a parish served by either the regular or secular clergy.

The new model limited population movement, pointing against the exchange of resources between different elevation zones and thus creating a new paradigm for native populations (Pease 1989). Historical studies note that, in practice, the limitation of movement was effective at the parish level (Saignes 1991; Medina 1974a,b, 1993). This system was maintained throughout the entire colonial period, and at the time of independence from Spain, parishes were renamed as districts, forming the basis for the current third-level administrative division of the country. As a result of the ressetlement, 336 parishes were created in the study region, representing approximately 24 percent of current districts.[15]

[14] One example is the festivity of the *Chankas* in honor of their mythical connection to the *puma*, which, according to early chronicles, was still celebrated during the Inca period (Garcilaso de la Vega (1960)[1609]).

[15] Throughout the text, I use "parish" and "district" interchangeably. The study focuses on the Peruvian territory conquered by the Inca empire that remained in the Viceroyalty of Peru for the entire colonial period (1532-1810). The census prepared from 1791 to 1795 under the administration of *Gil de Taboada y Lemos* lists all parishes in this territory. Following the literature on Murra's model, I focus on parishes located in highland Peru (i.e., more than 500 meters above sea level). I exclude two capital parishes, *Cuzco* and *Arequipa*, along with six parishes now within Chilean territory.

4 Empirical Strategy

4.1 Average Effect of Ethnic Diversity

According to the historical narrative, geographic proximity to ethnic boundaries created quasi-random variation in ethnic diversity across parishes. This narrative is based on the assumption that colonial officials were not fully aware of the vertical distribution of coethnic individuals across space, making it unlikely for them to systematically consider ethnic boundaries when determining parish locations (Murra 1975). There was tension between the pre-existing settlement pattern—a native response to geography—and the Spanish notion of jurisdiction. Adhering to a horizontal conception of the world, the colonial administration concentrated populations using small-scale, continuous, and delimited jurisdictions (Medina 1974a,b, 1993). Given the vertical settlement pattern, parishes created close to ethnic boundaries unintentionally concentrated populations from different ethnic origins (Wachtel 1976; Pease 1989, 1992).

I first explore whether the ethnic composition of colonial parishes influenced long-run comparative development using the following specification:

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

where y_p is a contemporary development outcome for parish p, $Ethnic\ div_p$ is a dummy variable indicating whether parish p was accidentally created close to an ethnic boundary, X_p is a vector of parish-level control variables measured at baseline, and υ_p is an error term.[16] I start by defining $Ethnic\ div_p$ as a dummy variable for the presence of an ethnic boundary within a buffer of 10-km radius (three *leguas*) from the parish capital.[17] For this exercise, I use Rowe (1946)'s mapping of pre-colonial ethnic boundaries. In panel (a) of Figure 1, parishes with an ethnic boundary within the 10-km buffer are highlighted in yellow (35 percent of the

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

[17] Colonial accounts of the territories visited by Spanish officials describe distances commonly ranging between two and three *leguas*, the colonial measure of distance (Jiménez de la Espada 1881). See Paz Soldán (1877) for the correspondence between *leguas* and kilometers during the 16th century. When the distance between the capitals of two parishes is less than 10 km, I use equidistant boundaries to ensure non-overlapping buffers. The resulting buffers have a mean and median area of 240.44 km^2 and 256.51 km^2, respectively. Furthermore, an ethnic group is considered as part of the buffer only if its homeland occupies at least one percent of the buffer's area, ensuring that the ethnic group has at least one grid cell of 1 × 1 km inside the buffer.

sample), while those located further inside ethnic homelands are displayed in blue. I consider a range of different radii and placebos for ethnic boundaries in supplementary analyses.

Validating ethnic diversity. Unfortunately, the ethnicity of relocated individuals was not systematically registered. Appendix B provides supporting evidence that $Ethnic\ div_p$ captures diversity in ethnicity at the parish level. I constructed a dictionary of native linguistic roots and compiled a dataset of 112,340 individuals with native paternal surname using colonial baptism records from digital genealogical sources. The dataset includes 65 parishes, of which 20 percent have an ethnic border within the 10-km buffer, covering the period 1605–1780. I explore whether there was significantly higher surname diversity in parishes located close to ethnic boundaries compared to those located in the interior of ethnic homelands. Two contextual features are worth noting: (i) pre-colonial endogamy traced through the male line, and (ii) the absence of a family name system, with only first names related to mythical ancestors, before the Spanish conquest. The Catholic Church introduced the Hispanic family name system for the purpose of religious indoctrination. Qualitative evidence suggests that early priests commonly selected a Spanish first name, imposing the mythical first names of the parents as surnames (see Carpio and Guerrero 2021). Table B.1 shows that parishes located close to ethnic boundaries tend to exhibit higher surname diversity (between 0.41 and 0.56 standard deviations, on average) compared to parishes located in the interior of ethnic homelands (panels A-C); no such pattern is observed among individuals with non-native surnames (panel D).

Balance tests. I present empirical evidence supporting the historical narrative that colonial officials did not systematically consider ethnic boundaries when determining parish locations. Specifically, I show that factors that could have influenced locations and affected post-resettlement development did not vary significantly with proximity to ethnic boundaries.

Spanish officials may have followed recommendations or avoided locations where they suspected it would be easier for native populations to escape (e.g., plains or lower elevations). The colonial regulation of 1569–1570 described three desirable characteristics (Jiménez de la Espada 1881): (i) land quality and abundance, as enough land was needed for native families to work following their own rules of crop rotation; (ii) proximity to surface water, which was a key advantage for irrigation and sustaining populations that depended on subsistence agriculture; and (*iii*) distance to *huacas*, sacred native shrines that honored nature, to facilitate religious indoctrination. The results in Table 1 show that, on average, parishes founded on ethnically homogeneous populations and those whose initial populations were ethnically diverse are statistically similar in these characteristics. I start by exploring the mean and standard deviation of elevation and land caloric suitability (Galor and Özak 2016). There are no significant differences in these characteristics, alleviating the concern that colonial officials selected locations differently (i.e., in a way that resulted in systematic differences in proximity to ethnic boundaries) at different elevations or in plains as opposed to more rugged terrain. Log distance to perennial rivers is also balanced. I collected data on pre-colonial shrines to explore the third recommendation. On average, there are no significant differences in log distance to native shrines between ethnically diverse and non-diverse parishes. The table also shows balance in log distance to *mita* mines (Dell 2010) and local prosperity at the time of the policy, as proxied by the value of expected tribute.[18]

4.2 Heterogeneous Effects of Ethnic Diversity

The main specification of interest explores whether the overall effect of ethnic diversity differs depending on past exposure to within-group heterogeneity ($\beta_3 \neq 0$):

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

where $\overline{H}_p$ is a measure of average exposure to within-group heterogeneity. In particular, I consider a weighted average of the level of exposure among the ethnic groups concentrated in each parish: $\overline{H}_p = \sum_e w_{pe} H_e$, where w_{pe} is the area share of ethnic group e within the buffer of parish p, and H_e is an ethnic-level measure of within-group heterogeneity. Most parishes

[18] The colonial requirement to send native populations to *mita* mines started in 1573. Expected tribute was based on official assessment of the number of native individuals at the time of the policy (Cook 1982). Table A.1 provides evidence on statistical balance for other pre-colonial characteristics. To proxy for the threat of native attack at the time of the policy, I georeferenced data on pre-colonial defensive sites (e.g., fortresses, walled sites, and *pukaras*). The table shows balance in log distance to these sites, as well as in log distance to pre-colonial socio-economic and institutional centers (urban sites, elite residences, and political sites). I also explore log distance to pre-colonial infrastructures (Inca roads, canals, and bridges), as colonial officials could have been interested in exploiting them. Finally, colonial officials could also have been interested in specific crops, such as maize or potatoes (Brush 1976). Land caloric suitability for these crops is also balanced.

with ethnic diversity (85 percent) concentrated two ethnic groups, while the remaining 15 percent of parishes concentrated either three or four groups. Figure A.2 characterizes ethnic weights (w_{pe}), which range from 0.013 to 0.986 among parishes with ethnic diversity, with similar mean (0.467) and median (0.452) values.

Measuring within-group heterogeneity. Pre-colonial subsistence relied on crop exchange between coethnics settled in complementary elevation zones (Murra 1975). I constructed spatial data on the distribution of the zones to compute a proxy for within-group heterogeneity (H_e). According to the research of Pulgar Vidal (1941), a Peruvian geographer that combined local geography and native folklore to provide a comprehensive account of the mountain environment, there are five complementary zones in the study region: Yunga (*warm valley*, 500–2,300 m), Quechua (*temperate land*, 2,300–3,500 m), Suni or Jalca (*high land*, 3,500–4,000 m), Puna (*cold land*, 4,000–4,800 m), and Janca (*white land*, 4,800–6,768 m).[19] Each zone is traditionally known for specific crops. For example, the natural limit of maize cultivation is the Quechua zone, while grains such as quinua and kañiwa, as well as lupins such as tarwi, are best grown in the Jalca zone. Various potato varieties, which can provide more carbohydrates per hectare than maize at high altitudes, grow exceptionally well in the Puna zone (Burger and Merwe 1990; Sandweiss and Richardson 2008). I map the spatial distribution of the zones using 1×1 km elevation data from FAO's Harmonized World Soil Database (version 1.2).[20] The resulting map is shown in Panel (b) of Figure 1.[21]

Using this map, I compute the reciprocal of the Simpson (or Herfindahl) index, a common measure of diversity in ecological studies (Magurran 2004): $H_e = 1/\sum_j s_{ej}^2$, where s_{ej} is the area share of zone j within the homeland of ethnic group e. The index is normalized to one for the group with the highest value, increasing as the composition of zones becomes more diverse. Figure A.3 shows the density of the H index at the ethnic group level. Approximately 23 percent of the groups have an index value below 0.5, while the index for the remaining 77 percent ranges from 0.5 to 1, with similar mean and median values (0.661 and 0.682, respectively). Appendix A.1 provides empirical evidence that H_e does indeed explain crop

[19] Figures in parentheses refer to elevation in meters above sea level.

[20] I assign each grid cell (1×1 km at the equator) to a particular zone based on median elevation.

[21] Spanish officials founded parishes at different elevations; the percentage of ethnically diverse parishes ranges from 23.53 to 44.23, depending on the zone (Table A.2).

variety in the data. On average, a one standard deviation increase in H_e is associated with a 0.5 standard deviation increase in log crop variety, considering only native crops.

Validating within-group heterogeneity: evidence from pre-colonial diets. I explore pre-colonial diets and provide suggestive evidence that H_e helps explain carbon-enriched diets where such diets were unlikely in the absence of internal crop exchange. Biochemical analyses of archaeological human remains can inform the role that carbon-enriched crops play in individuals' diets. In particular, stable isotope measures of carbon in bone and dentin collagen ($\delta^{13}C_{col}$) can provide information on the presence of certain plants in the protein component of the diet (Ambrose 1993; Ambrose and Norr 1993). Plants characterized by using the C_4 photosynthetic pathway for carbon fixation have particularly high $\delta^{13}C_{col}$ values compared to those using the C_3 pathway. Maize, sorghum, and millet are well-known C_4 plants, while most plants, including tubers, use the C_3 pathway.

Wilson et al. (2022)'s database represents the most extensive effort at compiling comparable stable isotope values in the region. The database includes 196 archaeological individuals distributed across eight ethnic groups.[22] Five groups are observed in the Quechua zone, the upper limit of maize cultivation, while the remaining three groups are observed in the Jalca zone. For example, stable isotope values from bone collagen found in the Jalca zone of the *Soras* ($H_e = 0.77$) suggest carbon-enriched diets. While these isotope values might not be representative of the *Soras* population, it is noteworthy that they were found in a zone where growing maize, the main C_4 staple crop in the study region, was challenging due to geoclimatic conditions.[23] Tung and Knudson (2018) find similar evidence in southern Peru. The authors document carbon-enriched diets likely resulting from maize consumption in a non-suitable location, suggesting crop exchange between elevation zones.

In Table 2, I regress individual-level $\delta^{13}C_{col}$ scores on H_e. The positive correlation in Column 1 suggests that individuals from more heterogeneous ethnic groups tended to have more carbon-enriched diets, at least in their protein component. Column 2 includes zone

[22]The database covers 23 percent of the land area in the study region (Figure A.4). In the absence of individual-level data on ethnicity, I assign each individual to an ethnic group using Rowe (1946)'s ethnic boundaries and the geographic coordinates of the archaeological site where the human remains were found.

[23]Kiwicha, which is also a C_4 plant native to Peru, can present a carbon isotopic signature similar to that of maize in the study region (Turner, Kingston, and Armelagos 2010). However, it is unclear whether its consumption became widespread in this region during pre-colonial times (e.g., Tung and Knudson 2018).

fixed effects, thus comparing individuals settled in the same zone but from ethnic groups with different values of H_e. The positive correlation is also significant after controlling for differences in crop variety (Column 3) and average caloric suitability (Column 4) across ethnic groups' homelands. Although the data do not allow direct testing of internal exchange, Column 5 provides evidence that the ethnic boundary matters in explaining carbon-enriched diets: I replicate the analysis using grid cells instead of ethnic groups.[24] In line with the idea that ethnic boundaries matter, the results from the falsification exercise suggest no correlation between the grid-level H index and individual $\delta^{13}C_{col}$ scores. While the H index captures crop variety at different levels of geographical aggregation (Appendix A.1), it is associated with carbon-enriched diets only at the ethnic group level. This evidence is consistent with the narrative that crop exchange was centralized at the level of ethnic group (Murra 2002b).

Identifying assumptions. A causal interpretation of β_3 requires two conditions. First, ethnic diversity should not be determined by average exposure to within-group heterogeneity ($\overline{H}_p$). The 16th-century resettlement created a unique setting where $Ethnic\ div_p$ was likely orthogonal to $\overline{H}_p$. However, there might be concerns that more heterogeneous ethnic groups negotiated locations in the interior of ethnic homelands, leading ethnically homogeneous parishes to systematically concentrate populations from more heterogeneous ethnic groups. Table A.4 provides supporting evidence that parish locations were not significantly influenced by these groups. There is no significant correlation between within-group heterogeneity (H_e) and the total number of parishes where an ethnic group was concentrated. The proportion of parishes located close to ethnic boundaries and the average distance from a parish to the closest ethnic boundary are also uncorrelated with H_e. Consistently, Figure 2 shows a similar distribution of $\overline{H}_p$ among ethnically homogeneous parishes (left boxplot) and those concentrating populations from various ethnic groups (right boxplot)—$Ethnic\ div_p$ is not significantly correlated with $\overline{H}_p$ in the data (Table A.5).

Second, β_3 should not capture any differential effects of ethnic diversity due to correlates of within-group heterogeneity that may have been relevant for post-resettlement economic development. Table 3 explores the correlates of within-group heterogeneity (H_e). Columns 1-3 show that mean elevation, land caloric suitability, and river density are not significantly

[24] I define grid-cell size (50 × 50 km) so that the number of cells matches the number of ethnic groups.

correlated with H_e. There is no significant correlation between H_e and ethnic group size, as measured by land area (Column 4) and approximate population (Column 5) before colonization. However, H_e does exhibit a positive correlation with pre-colonial population density (Column 6), which could reflect economic prosperity (Ashraf and Galor 2011, 2013). Columns 7-9 explore pre-colonial socio-economic and institutional characteristics of the groups. In the absence of systematized ethnographic data, I collect information from archaeological sources (Appendix C). In line with Column 6, the data suggest that H_e is positively correlated with urbanization, as measured by a dummy for the presence of towns and urban centers within ethnic homelands (Column 7). Column 8 shows evidence consistent with the idea that incentives for internal exchange may lead to political centralization (Fenske 2014). In particular, I create a dummy for any material indicator that could evince political complexity (i.e., administrative centers and monumental architecture—public buildings and communal spaces, including temples, palaces, and complex mound platforms, as defined in Stanish 2001) and find a positive correlation with H_e. Column 9 shows no correlation with the presence of elite residences, nonetheless.[25] In robustness checks, I conduct various empirical exercises to mitigate the concern that pre-colonial correlates of H_e might confound the interaction effect.

5 Results

5.1 Main Results

Outcome variables. I examine contemporary living standards using various measures of local economic activity and access to public facilities. I first use luminosity data from satellite images at night to proxy for local economic activity (Michalopoulos and Papaioannou 2013). As a second proxy, I use non-subsistence agriculture (Dell 2010). Specifically, I use data from the agricultural census, which asks whether farmers primarily consume their harvest or sell it in local markets. The population census provides data on access to public sanitation and water. These variables are measured across different years, ranging from 1990 to 2000, depending on

[25] In Table A.6, I explore the correlation of within-group heterogeneity with various pre-colonial infrastructures. As expected, H_e is positively correlated with terraces. However, I find no evidence of a correlation with Inca roads, water canals, bridges, or food storage structures. The latter may be due to certain crops, such as potatoes, being sometimes simply spread on the ground or placed underground to freeze, leaving no archaeological record.

data availability, and 2010 to 2020, based on subsequent waves of the same data sources.[26]

Graphical analysis. Figure 3 displays the mean of each outcome variable (1990–2000) as a function of the number of ethnic groups. Overall, contemporary living standards are negatively associated with the number of ethnic groups concentrated in the 16th century. Based on the first principal component of the four variables, living standards are about 0.32 standard deviations lower, on average, in parishes that concentrated populations from different ethnic groups (35 percent of parishes).[27] To learn about the distribution of living standards, the left panel of Figure 4 plots the quintiles of the first principal component within bins of average exposure to within-group heterogeneity ($\overline{H}_p$) and $Ethnic\ div_p$ status. Below the median level of within-group heterogeneity ($\overline{H}_{p50}$=0.675, dashed line), most parishes in the top quintile of living standards (in yellow) are parishes with an ethnically homogeneous founding population. Above the median, however, the density of parishes in the top quintile drops drastically among ethnically homogenous parishes and increases in those built on various ethnic groups.

The right panel displays the distribution of living standards among parishes that concentrated populations from specific ethnic groups. The graphs illustrate different scenarios depending on the level of past exposure to within-group heterogeneity. Among parishes with *Lucanas* populations (below the median of within-group heterogeneity), the highest living standards are found in parishes with ethnically homogeneous populations, i.e., those exclusively comprising *Lucanas*. However, in parishes with *Cavinas* populations (slightly above the median) or *Conchucos* populations (substantially above the median), the top-ranking parishes are the ones where the *Cavinas* or *Conchucos* were resettled with other ethnic groups.

Baseline regression analysis. To analyze the average effect of ethnic diversity on long-run development, I follow the methodology in Kling et al. (2004) and Clingingsmith, Khwaja, and Kremer (2009). Specifically, I report the standardized average effect size (AES) across the different outcome variables, thus accounting for the covariance across underlying individual effects, jointly with heteroskedasticity-robust standard errors.

[26] Table A.7 presents summary statistics; Appendix C reports data sources and definitions.

[27] The first principal component refers to the log of average light intensity per capita (2000–2003), a dummy variable equal to one if the share of farmers practicing non-subsistence agriculture is above the median (1994), the share of dwellings with access to public sanitation (1993), and the share of dwellings with access to the public water network (1993), accounting for 55 percent of the total variance.

I first compare living standards between parishes whose initial populations were ethnically diverse and those with an ethnically homogeneous founding population (Equation 1). Panel A of Table 4 presents the results for 1990–2000. On average, living standards are 0.2 standard deviations lower in parishes built on ethnically diverse populations (Column 1). This disparity persists after considering parish-level baseline characteristics (X_p) and ecclesiastical jurisdiction fixed effects, accounting for potential differences across five colonial bishoprics (columns 2 and 3).[28] Table A.8 shows that geographic proximity to ethnic boundaries does not fully account for this result. I divided the study region into 10×10 km grid cells and created a dummy variable indicating whether a grid cell is within the buffer of an ethnically diverse parish. On average, grid cells located within the buffer of an ethnically diverse parish exhibit lower nightlight, even after controlling for proximity to ethnic boundaries.

I then explore the role of pre-colonial exposure to within-group heterogeneity. Table A.9 shows a positive correlation between average exposure ($\overline{H}_p$) and contemporary living standards. Nonetheless, this correlation is more pronounced among parishes with ethnic diversity (about 0.834) than among parishes where only one ethnic group was concentrated (about 0.420). The remaining columns of Table 4 present the results from estimating the interaction effect of ethnic diversity and $\overline{H}_p$ for the entire sample (Equation 2). Column 4 presents the unconditional specification for reference, while Column 5 presents the specification with parish-level baseline controls and ecclesiastical jurisdiction fixed effects. The estimated coefficient on ethnic diversity remains negative ($\hat{\beta}_1 < 0$); however, the coefficient on the interaction with $\overline{H}_p$ is consistently positive ($\hat{\beta}_3 > 0$). This pattern persists when I include fixed effects accounting for the colonial administrative province (44 provinces) instead of the ecclesiastical jurisdiction, jointly with standard errors clustered at the province level (Column 6).

On average, parishes built on ethnically diverse populations exhibit lower living standards. However, those with initial populations exposed to higher within-group heterogeneity tend to perform better in the long run. Similar estimates for the period 2010–2020 support that the documented pattern is not sensitive to specific years (Panel B of Table 4). Column 8 shows no direct effect on contemporary population density and urban status. Although these variables may correlate with other dimensions of contemporary development, incorporating them in the baseline specification (Column 7) suggests that the main result holds beyond

[28]The vector of parish-level baseline characteristics includes all variables in Table 1, longitude and latitude.

potential agglomeration effects (see, e.g., Michalopoulos and Papaioannou 2014). Table A.10 reports OLS estimates for each outcome variable separately, standard errors adjusted for spatial autocorrelation (Colella et al. 2019), and the R-squared coefficient. The results show that much of the documented overall effect of ethnic diversity is driven by differences in economic activity rather than access to public facilities.[29]

Figure 5 plots the estimated average effect size of ethnic diversity as a function of past exposure to within-group heterogeneity for 2010–2020 outcomes. Since $\overline{H}_p$ ranges from 0.301 to 1, the estimates imply that the negative average effect size of ethnic diversity decreases from -0.881 to -0.023 standard deviations as $\overline{H}_p$ reaches the median ($\overline{H}_{p50}$=0.675). However, in parishes built on various ethnic groups, a change from 0.6 to 1 of average within-group heterogeneity is linked to a 0.36 standard deviation increase in overall living standards compared to parishes built on a single ethnicity. This corresponds to a 4.8 percentage point increase in log nightlight per capita (relative to a mean of 0.056 log nightlight per capita) and a 38.44 percent increase in the probability of practicing non-subsistence agriculture (relative to a mean of 65 percent). The positive coefficient on ethnic diversity is statistically significant for $\overline{H}_p$ above 0.830, which corresponds to approximately 16.7 percent of the parishes.

Pre-colonial characteristics of ethnic groups. In Table 5, I show that the main result persists when controlling for pre-colonial correlates of within-group heterogeneity. I consider the set of pre-colonial ethnic characteristics analyzed in Table 3. In particular, I compute the weighted average of each characteristic among the ethnic groups concentrated in the parish and augment Equation 2 to control for the resulting average ($\overline{G}_p = \sum_e w_{pe} G_e$) and its interaction with ethnic diversity ($Ethnic\ div_p \times \overline{G}_p$). The first column shows the baseline specification (Column 6 of Table 4 for 2010–2020 living standards), Columns 2 to 8 introduce one characteristic at a time (only the coefficient on the interaction with ethnic diversity is reported), and Column 9 includes all characteristics. The results alleviate the concern that relevant socio-economic and institutional characteristics of ethnic groups could be driving the entire result. Column 10 shows that the positive interaction effect persists when using lasso methods to *select* the set of ethnic characteristics to be included (Belloni, Chernozhukov, and Hansen 2014).[30]

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

[30]The lasso routine considers all ethnic characteristics included in Column 9 of Table 5, land caloric suitability for maize, and ethnic infrastructures (Table A.6)—a total of 13 characteristics, of which 7 are *selected* by lasso.

Coarsened exact matching. In Table 6, I show the results from using a matching procedure to construct a counterfactual for parishes with high exposure to within-group heterogeneity (defined as $\overline{H}_p$ above the median). Specifically, I use coarsened exact matching (Iacus, King, and Porro 2012) to create a sample of parishes that present varying levels of exposure to within-group heterogeneity but are statistically similar in other pre-colonial characteristics ($\overline{G}_p$). In columns 1-2, the pre-colonial ethnic characteristics to be balanced by the matching algorithm are the characteristics used in Column 9 of Table 5. In columns 3-4, the procedure uses the set of lasso-*selected* characteristics.[31] Despite the reduced sample sizes, the results from the matched samples (columns 1 and 3) are consistent with the documented pattern.

To alleviate the concern that certain ethnic groups could be driving the results, the remaining columns of Table 6 control for the weight (w_{pe}) of the majority group, defined as the group with the highest area share within the 10-km buffer. The estimated interaction effect remains positive and statistically significant in both the matched and full samples. Additionally, Panel (a) of Figure A.5 shows consistent results when excluding all parishes with presence of a certain ethnic group, for one ethnic group at a time.

Randomization inference. The empirical analysis is guided by the historical narrative that the colonial resettlement resulted in quasi-random variation in ethnic diversity across parishes. Table 7 shows the results from permuting $Ethnic\ div_p$ within the regression sample. I use the procedure proposed by Young (2023) for randomization inference in the presence of heterogeneous treatment effects. In each iteration, $Ethnic\ div_p$ is permuted, and the interaction term with average exposure to within-group heterogeneity is recalculated accordingly. The results from 1,000 iterations show stable permutation-based p-values.

5.2 Supplementary Analyses

Pre-colonial land occupation and transition zones. The estimates in Table 4 are likely affected by non-classical measurement errors. A potential source of error is the underlying assumption that pre-colonial individuals were uniformly distributed over space. In Table A.12, I consider alternative scenarios. In the absence of historical data on the spatial distribution of

[31] Table A.11 documents that, in the matched samples, parishes that concentrated ethnically diverse and non-diverse populations continue to be balanced along geographic and initial factors.

the population, I use archaeological site records as evidence of land occupation. Columns 1 and 2 show consistent estimates after restricting the analysis to 20 km and 10 km around pre-colonial archaeological sites, respectively.[32] The baseline estimates capture the effect of potential exposure (i.e., another assumption is that all coethnic individuals were equally exposed to within-group heterogeneity). In Columns 3 and 4, I restrict the analysis to 20 km and 10 km around the transitions from one elevation zone to another. Although the results in Section 4.2 suggest that ethnic boundaries are relevant in explaining crop exchange, potential gains may have been higher around transition zones (Bates 2001). The estimates are similar to baseline results, suggesting no significant differences in exposure among coethnics.

Placebos for pre-colonial ethnic boundaries. Following Alesina, Michalopoulos, and Papaioannou (2016), Column 5 of Table A.12 presents the results from using artificial ethnic boundaries (from Thiessen polygons) instead of historical ones.[33] Compared to baseline results, the estimates are small and not significant, suggesting that historical ethnic boundaries matter. This aligns with the evidence in Section 4.2, where I used grid cells instead of Thiessen polygons. In Column 6, I use the boundaries of the first administrative demarcations of the colonial period (*corregimientos*). Close correspondence between the spatial boundaries of pre-colonial ethnic groups and *corregimientos* would suggest that the Spanish administration based the latter on prior knowledge of the spatial distribution of the groups. The estimated coefficients are not statistically significant, suggesting that this was not the case.

Sensitivity analyses. Table A.13 reports the results from varying the size of the buffer used to measure ethnic diversity. In columns 1-2 of Table A.14, I use an index of ethnic fractionalization ($1 - \sum_e w_{pe}^2$), while in columns 3-4, I use an alternative index of within-group heterogeneity ($\widetilde{H}_e = 1 - \sum_j s_{ej}^2$). Columns 5-6 consider the within-group heterogeneity of the majority group, instead of the weighted average of H_e, while columns 7-8 focus on the within-group heterogeneity of the minority group. The results are consistent with the documented pattern and support that both the majority and minority groups matter. Panel (b) of Figure A.5 displays point estimates and confidence intervals after excluding one parish at a

[32]This corresponds to reductions of 7.35 and 31.33 percent in total land area, respectively.

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

time, alleviating concerns about influential observations.

Additional robustness checks. Appendix A.2 shows that the main results of the paper hold for household consumption and addresses potential selective migration. Finally, it is important to consider that any potential effect of ethnic diversity after resettlement depends on the survival of the groups. The decline in native populations after European contact has been extensively documented by historical studies (e.g., Cook 1982). To the extent that all groups were similarly affected by disease and abuse, the estimates should be interpreted as the effect of ethnic diversity among the descendants of survivors. Table A.16 shows that the positive coefficient on the interaction term is robust to controlling for pre-resettlement variables related to the spread of smallpox—an infectious disease caused by the variola virus that may have affected native populations during the conquest. The table also shows results after controlling for variables related to the Inca period.

6 Mechanisms

Cultural transmission. According to early chronicles and ethnohistoric studies, pre-colonial internal exchange likely played a role in shaping a culture of cooperation (Section 3). This culture may have facilitated local inter-group interactions after forced resettlement, contributing to more open attitudes toward out-group members and fostering a trend toward a more integrated society. I first study inter-group contact during the colonial period. Did exposure to within-group heterogeneity favor inter-group interactions?

Comparing the two surnames of each individual in the colonial baptism records (1605–1780) offers the opportunity to explore inter-group unions.[34] The sample includes 17,411 individuals with native surnames across 41 parishes, of which 10 are parishes built on ethnically diverse populations. Since ethnic identities were not systematically registered, I use a measure of dissimilarity between the two surnames to *detect* parents who were potentially *from different ethnic groups*.[35] Given the sample size, a graphical summary of the data is provided in Figure 6.

[34]Each individual inherits two surnames in the Hispanic system of family names. The first surname corresponds to the paternal surname of the father, while the second corresponds to the paternal surname of the mother. Colonial marriage records from digital genealogical sources are limited in quantity and geographic coverage.

[35]I use the minimum number of spelling changes required to transform one surname into another (Levenshtein),

In line with the hypothesized channel, the left graph of Panel (a) suggests a positive correlation between average dissimilarity and exposure to within-group heterogeneity among parishes with ethnic diversity. The right graph shows a similar pattern when focusing on the share of unions with dissimilarity above 50 percent. Panel (b) replicates this exercise for the subsample of ethnically homogeneous parishes, revealing no significant correlation, which is reassuring.

It is worth noting that the ethnic identities of the sample can only be *detected* in historical data, primarily because native languages have gradually been displaced by regional varieties of Quechua and Spanish, shaping today's linguistic landscape. The visual analysis of baptism records sheds light on the long-run process of assimilation. Nonetheless, small-N results should be interpreted cautiously. Inter-ethnic union is also commonly used as a proxy for national integration (e.g, Bazzi et al. 2019). Contemporary survey data on individuals' self-reported identities are consistent with a more nation-oriented society (Table 8). Individuals tend to exhibit stronger identification with the state where the resettlement forced together ethnically diverse populations previously exposed to within-group heterogeneity (columns 1-3).[36] Consistently, these individuals are more likely to vote in presidential elections, a result that holds beyond trust in the state (columns 4-5). Parish-level data on volunteers for military service exhibit a similar pattern (columns 6-7). These estimates further align with the presence of neighborhood associations (columns 8-9), which in turn points toward a culture of cooperation (Guiso, Sapienza, and Zingales 2016).[37]

Economic complementarities. Being more willing to interact, inter-ethnic economic complementarities may have then sustained beneficial coexistence (Jha 2013). Complementarities may have been more likely where the ethnic minority belonged to a highly heterogeneous group but, conversely, the majority was relatively homogeneous (i.e., the minority likely complemented a desperate majority). To test this hypothesis, I define the minority group as

normalized by the length of the longest surname to be interpreted as the percentage of dissimilarity. See Dickens (2022) for an application of this measure to compute distance between languages.

[36]The ENAHO Peruvian survey (2004-2017) includes a question on whether individuals identify more strongly with their state administrative unit, ethnic or racial group, religion group, or other groups. All regressions include individual-level controls, thus comparing individuals with similar socio-demographic characteristics, and survey-year fixed effects, with standard errors clustered at the parish level.

[37]The results are robust to controlling for log population and consistent with individual-level data from the ENAHO survey on participation in voluntary associations (Table A.17). Unfortunately, ENAHO does not include trust-related questions, and the Latinobarometer and LAPOP surveys do not cover most of the sample.

the group with the lowest area share in the 10-km buffer and run regressions of the following form for the subsample of parishes built on ethnically diverse populations (Table 9, Panel A):

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

where $High\,min_p$ is a dummy variable indicating whether the minority belonged to a highly heterogeneous ethnic group (H_e above the 75th percentile) and $High\,maj_p$ is an analogous dummy for the majority group—the correlation between the two variables is 0.290. Consistent with the proposed interpretation, where the historical ethnic minority had a comparative advantage, contemporary populations tend to perform better ($\hat{\delta}_1 > 0$, column 2).[38]

The 1876 population census—the earliest post-colonial census with information on occupations—reveals that most parishes remained primarily agricultural by the late 19th century, with an average of 70 percent of the population employed in the agricultural sector. In the remaining columns, I explore whether there is empirical support for the hypothesis that ethnic minorities with a comparative advantage over the majority group specialized in intermediary services related to agriculture (e.g., retailing), thus reducing inter-ethnic competition (Grosfeld, Sakalli, and Zhuravskaya 2020). Consistently, columns 3 and 4 show that contemporary local retail markets for selling agricultural products tend to be located where the historical ethnic minority had a comparative advantage. To my knowledge, no such information is available for earlier years. However, in line with the idea of reduced competence over local resources, colonial data for the period 1700–1800 show that a larger share of the population had access to agricultural land in these parishes (columns 5 and 6).[39] This effect is offset where the minority loses its advantage over the majority group ($\hat{\delta}_1 + \hat{\delta}_3 \approx 0$).

Consequences: tertiary-sector occupation. Subsistence farming has historically been prevalent in the Andean highlands (Mayer 2002). This paper shows that non-subsistence agriculture flourished where the 16th-century resettlement brought together ethnically diverse populations previously exposed to within-group heterogeneity (Table A.10). This pattern holds beyond local crop diversity (Table A.18). Furthermore, it aligns with the appearance of

[38]In these parishes, the minium advantage of the minority over the majority group in terms of within-group heterogeneity is 4.34 percent, with a maximum of 34.17 percent.

[39]This information is available for 44 percent of the parishes with ethnic diversity.

intermediary services related to agriculture, primarily supported by ethnic minorities (Table 9, Panel A). I use the 1876 population census to explore the share of the population employed in retail and services (Table 9, Panel B). In line with the documented pattern, the estimates show a shift toward tertiary-sector activities in a predominantly agricultural setting. As past exposure to within-group heterogeneity increases, employment tends to become more oriented toward the tertiary sector where the resettlement brought together ethnically diverse populations.[40] Consistently, decomposing tertiary-sector employment shows that local trade drives the overall effect. Census data on 21st-century employment show a similar pattern.[41]

7 Conclusion

A large body of literature has examined the implications of ethnic diversity for human living standards and comparative development. However, the role of within-group heterogeneity has been less explored. This paper shows that exposure to within-group heterogeneity in complementary traits matters for understanding the long-run effects of ethnic diversity. I collect new data from a natural experiment of Peru's colonial history and find consistent evidence that, where colonial officials concentrated populations with a history of within-group heterogeneity, who settled in complementary altitudes of the Andes before colonization, ethnic diversity results in systematically lower costs and may even become advantageous.

The subsistence strategy of pre-colonial populations, characterized by a single ethnic group having simultaneous control over altitude-specific resources, contributed to shaping the long-run effects of ethnic diversity in the study setting. Similar subsistence strategies are known to have existed in other cultures, such as those of Bali and Polynesia. The results suggest that further studying the internal economic organization of ethnic groups can help us understand comparative development. This paper also sheds light on the long-run effects of forced displacements (Becker 2022). The consequences of indigenous displacement due to colonization is a research topic with limited evidence (see Valencia Caicedo 2019). In

[40] I find no significant evidence that pre-colonial crop diversity may have indirectly fostered secondary-sector (Fiszbein 2022) or tertiary-sector occupation through the potential increased availability of skills (Table A.19).

[41] Contemporary data on technology adoption in agriculture show no statistically significant differences. However, chemical fertilizers and similar products have traditionally been imported in the study setting–their availability may not have been uniform across the territory (results available upon request).

contemporary societies where multiple ethnicities coexist (e.g., due to forced displacements or voluntary migrations in an increasingly globalized world), understanding whether the consequences of ethnic diversity depend on exposure to within-group heterogeneity can also inform policy discussions.

References

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

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

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

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

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

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

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

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

Ambrose, Stanley H. 1993. "Isotopic Analysis of Paleodiets: Methodological and Interpretive Considerations." In *Investigations of Ancient Human Tissue: Chemical Analysis in Anthropology*. Philadelphia, USA: Gordon and Breach Science Publishers, 59–130.

Ambrose, Stanley H and Lynette Norr. 1993. "Experimental Evidence for the Relationship of the Carbon Isotope Ratios of Whole Diet and Dietary Protein to Those of Bone Collagen and Carbonate." In *Prehistoric Human Bone: Archaeology at the Molecular Level*. Berlin: Springer, 1–37.

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

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

Bates, Robert H. 2001. *Prosperity and Violence: The Political Economy of Development*. W. W. Norton & Company.

Bauer, Brian S, LC Kellett, and Aráoz Silva. 2010. *The Chanka: Archaeological Research in Andahuaylas (Apurimac)*. Cotsen Institute of Archaeology, UCLA. With contributions by Sabine Hyland and and Carlo Socualaya Dávila.

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

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

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

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects After Selection Among High-Dimensional controls." *The Review of Economic Studies* 81 (2):608–650.

Berezkin, Yu E. 2015. "Folklore and Mythology Catalogue: Its Lay-out and Potential for Research." *The Retrospective Methods Network* 1 (S10):58–70.

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

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

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

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

Charney, Paul. 1998. "A Sense of Belonging: Colonial Indian Cofradias and Ethnicity in the Valley of Lima, Peru." *The Americas* 54 (3):379–407.

Cieza de León, Pedro de. 1962. *La Crónica del Perú*. Madrid: Espasa–Calpe. [1553].

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

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

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

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

Dell, Melissa. 2010. “The Persistent Effects of Peru’s Mining Mita.” *Econometrica* 78 (6):1863–1903.

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

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

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

Desmet, Klaus, Shlomo Weber, and Ignacio Ortuño-Ortín. 2009. “Linguistic Diversity and Redistribution.” *Journal of the European Economic Association* 7 (6):1291–1318.

Dickens, Andrew. 2022. “Understanding Ethnolinguistic Differences: The Roles of Geography and Trade.” *The Economic Journal* 132 (643):953–980.

Dippel, Christian. 2014. “Forced Coexistence and Economic Development: Evidence from Native American Reservations.” *Econometrica* 82 (6):2131–2165.

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

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

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

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

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

Fiszbein, Martin. 2022. “Agricultural Diversity, Structural Change, and Long-Run Development: Evidence from the United States.” *American Economic Journal: Macroeconomics* 14 (2):1–43.

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

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

Grosfeld, Irena, Seyhun Orcan Sakalli, and Ekaterina Zhuravskaya. 2020. "Middleman Minorities and Ethnic Violence: Anti-Jewish Pogroms in the Russian Empire." *The Review of Economic Studies* 87 (1):289–342.

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

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

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

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

Iacus, Stefano M, Gary King, and Giuseppe Porro. 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching." *Political Analysis* 20 (1):1–24.

Isbell, William H. 2010. "La Arqueología Wari y la Dispersión del Quechua." *Boletín de Arqueología PUCP* (14):199–220.

Isbell, William H and Helaine Silverman. 2002. "Theorizing Variations in Andean Sociopolitical Organization." In *Andean Archaeology I*. Springer, 3–11.

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

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

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

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

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

Kling, Jeffrey R, Jeffrey B Liebman, Lawrence F Katz, and Lisa 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, Rafael, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert Vishny. 1999. "The Quality of Government." *The Journal of Law, Economics, and Organization* 15 (1):222–279.

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

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

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

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

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

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

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

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

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

Murra, John V. 1956. "The Economic Organization of the Inca State." Tech. rep., The University of Chicago.

———. 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.

———. 2002a. *El Mundo Andino: Población, Medio Ambiente y Economía*, vol. 24. Lima: Instituto de Estudios Peruanos.

———. 2002b. “Las Sociedades Andinas antes de 1532.” In *América Latina en la Época Colonial*, vol. 1. Barcelona: Crítica, 56–82.

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

Nash, Donna J. 2009. “Household Archaeology in the Andes.” *Journal of Archaeological Research* 17 (3):205–261.

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

Nunn, Nathan and Leonard Wantchekon. 2011. “The Slave Trade and the Origins of Mistrust in Africa.” *American Economic Review* 101 (7):3221–52.

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

Pease, Franklin. 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.

Pulgar Vidal, Javier. 1941. “Las Ocho Regiones Naturales del Perú.” In *Boletín del Museo de Historia Natural “Javier Prado”*, 17. Lima: Imprenta D. Miranda, 145–160.

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

Rowe, John Howland. 1946. “Inca Culture at the Time of the Spanish Conquest.” In *Handbook of South American Indians*, vol. 2. US Government Printing Office, 183–330.

Saignes, Thierry. 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.

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

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

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

Stern, Steve 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.

Tello, Julio César. 1939. "Origen y Desarrollo de las Civilizaciones Prehistóricas Andinas." In *Actas y Trabajos. XXVII Congreso Internacional de Americanistas*, vol. 1. Lima: Librería e Imprenta Gil, 589–720.

Tung, Tiffiny A and Kelly J Knudson. 2018. "Stable Isotope Analysis of a Pre-Hispanic Andean Community: Reconstructing Pre-Wari and Wari era Diets in the Hinterland of the Wari empire, Peru." *American Journal of Physical Anthropology* 165 (1):149–172.

Turner, Bethany L, John D Kingston, and George J Armelagos. 2010. "Variation in Dietary Histories Among the Immigrants of Machu Picchu: Carbon and Nitrogen Isotope Evidence." *Chungara, Revista de Antropología Chilena* 42 (2):515–534.

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

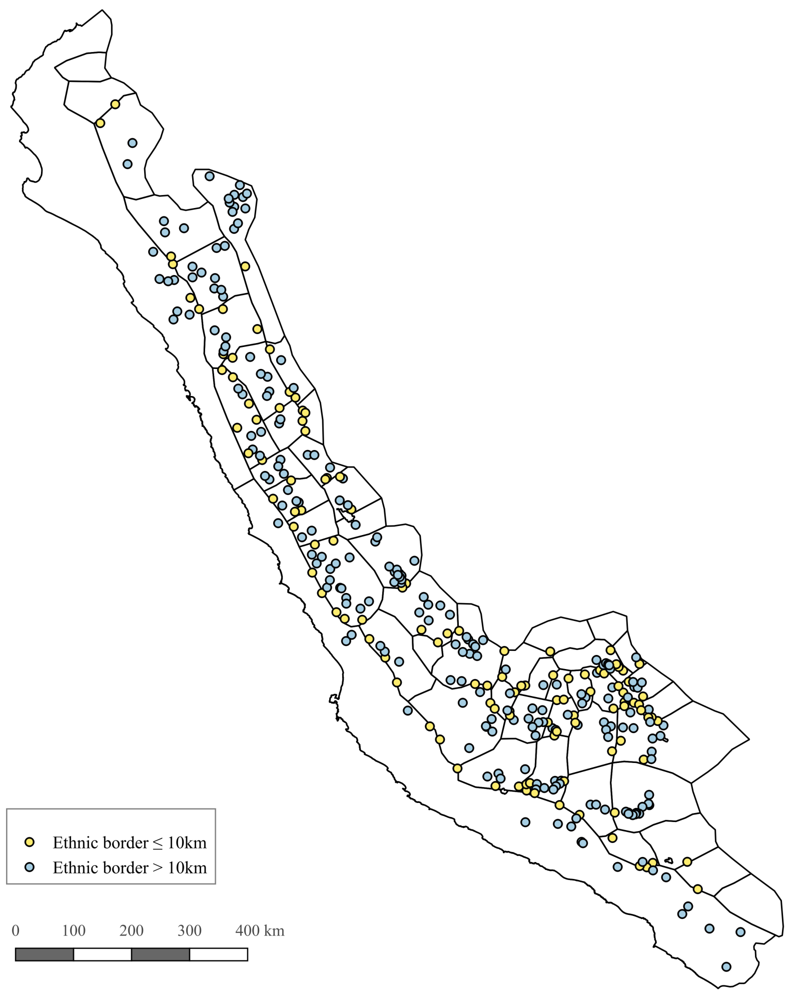
Valverde, Guido, María Inés Barreto Romero, Isabel Flores Espinoza, Alan Cooper, Lars Fehren-Schmitz, Bastien Llamas, and Wolfgang Haak. 2016. "Ancient DNA Analysis Suggests Negligible Impact of the Wari Empire Expansion in Peru's Central Coast during the Middle Horizon." *PLoS One* 11 (6):e0155508.

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

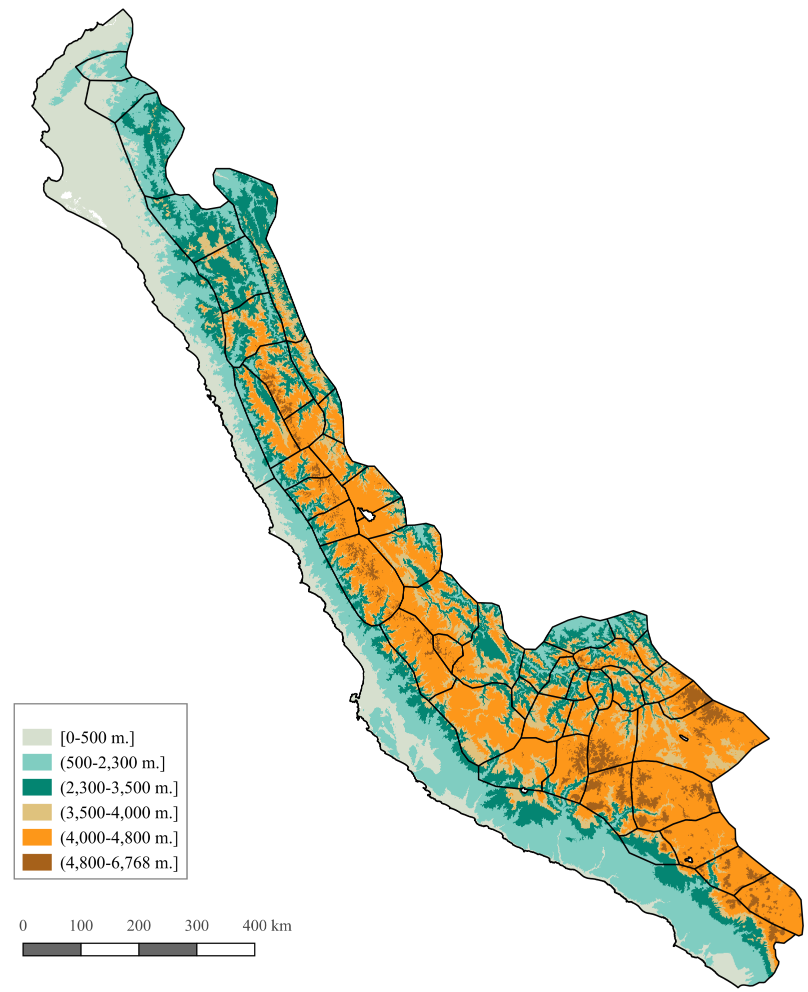
Wachtel, Nathan. 1976. *Los Vencidos. Los Indios del Perú frente a la Conquista Española (1530-1570)*. Madrid: Alianza Editorial.

Wilson, Kurt M, Weston C McCool, Simon C Brewer, Nicole Zamora-Wilson, Percy J Schryver, Roxanne Lois F Lamson, Ashlyn M Huggard, Joan Brenner Coltrain, Daniel A Contreras, and Brian F Codding. 2022. "Climate and Demography Drive 7000 Years of Dietary Change in the Central Andes." *Scientific Reports* 12 (1):1–16.

Young, Alwyn. 2023. "Asymptotically Robust Permutation-based Randomization Confidence Intervals for Parametric OLS Regression." *European Economic Review* :104644.



(a) Colonial Parishes



(b) Complementary Elevation Zones

FIGURE 1: Colonial Parishes and Elevation Zones

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

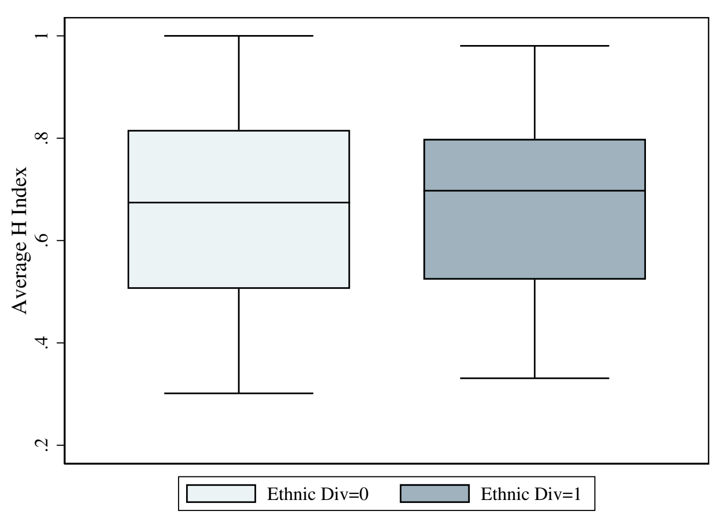


Figure 2: Ethnic Diversity and Average Within-Group Heterogeneity

Notes. Boxplots of average within-group heterogeneity ($\overline{H}_p$) by ethnic diversity status across parishes. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise.

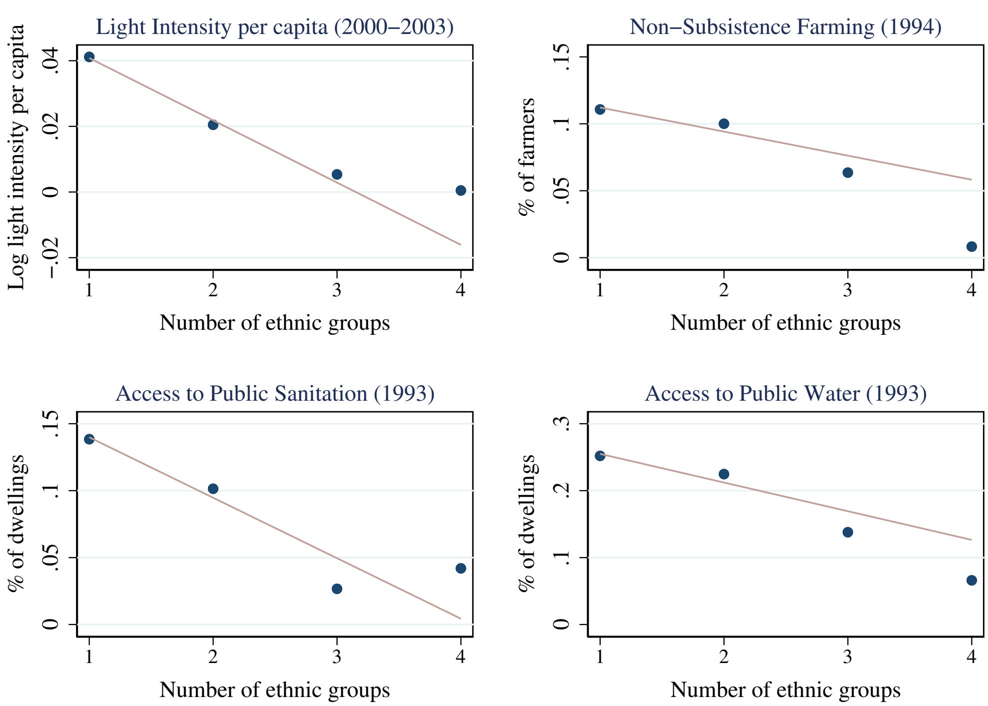


Figure 3: Number of Ethnic Groups and Contemporary Development

Notes. The graphs display the mean of each outcome variable as a function of the number of ethnic groups using parish-level data. Specifically, the x-axis refers to the number of ethnic groups within a buffer of 10-km radius from the parish capital. Among parishes with ethnic diversity, 85 percent of parishes concentrated two ethnic groups only.The outcome variables refer to the log average light intensity per capita (2000–2003), the share of farmers practicing non-subsistence agriculture (1994), the share of dwellings with access to public sanitation (1993), and the share of dwellings with access to the public water network (1993).

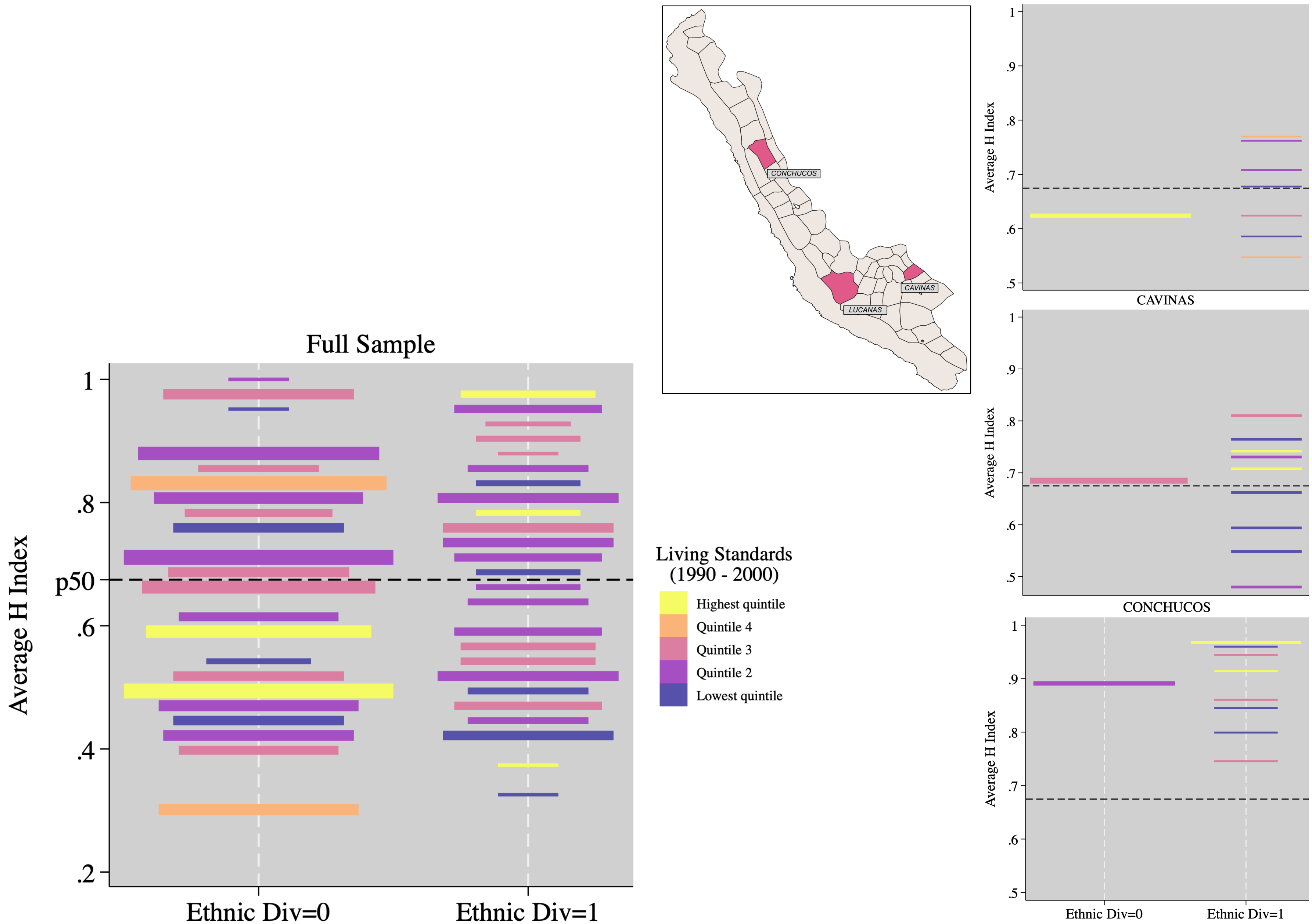


Figure 4: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development

Notes. The color legend indicates the quintiles of the first principal component of the log of average light intensity per capita (2000–2003), a dummy variable equal to one if the share of farmers practicing non-subsistence agriculture is above the median (1994), the share of dwellings with access to public sanitation (1993), and the share of dwellings with access to the public water network (1993), within bins of average exposure to within-group heterogeneity ($\overline{H}_p$) and ethnic diversity status. The size is scaled according to the relative frequency of each combination of ethnic diversity and binned within-group heterogeneity. The black dashed line indicates the median of average exposure to within-group heterogeneity. The left panel includes all parishes; the right panel includes parishes with populations from the *Lucanas*, *Cavinas*, and *Conchucos* ethnic groups, separately.

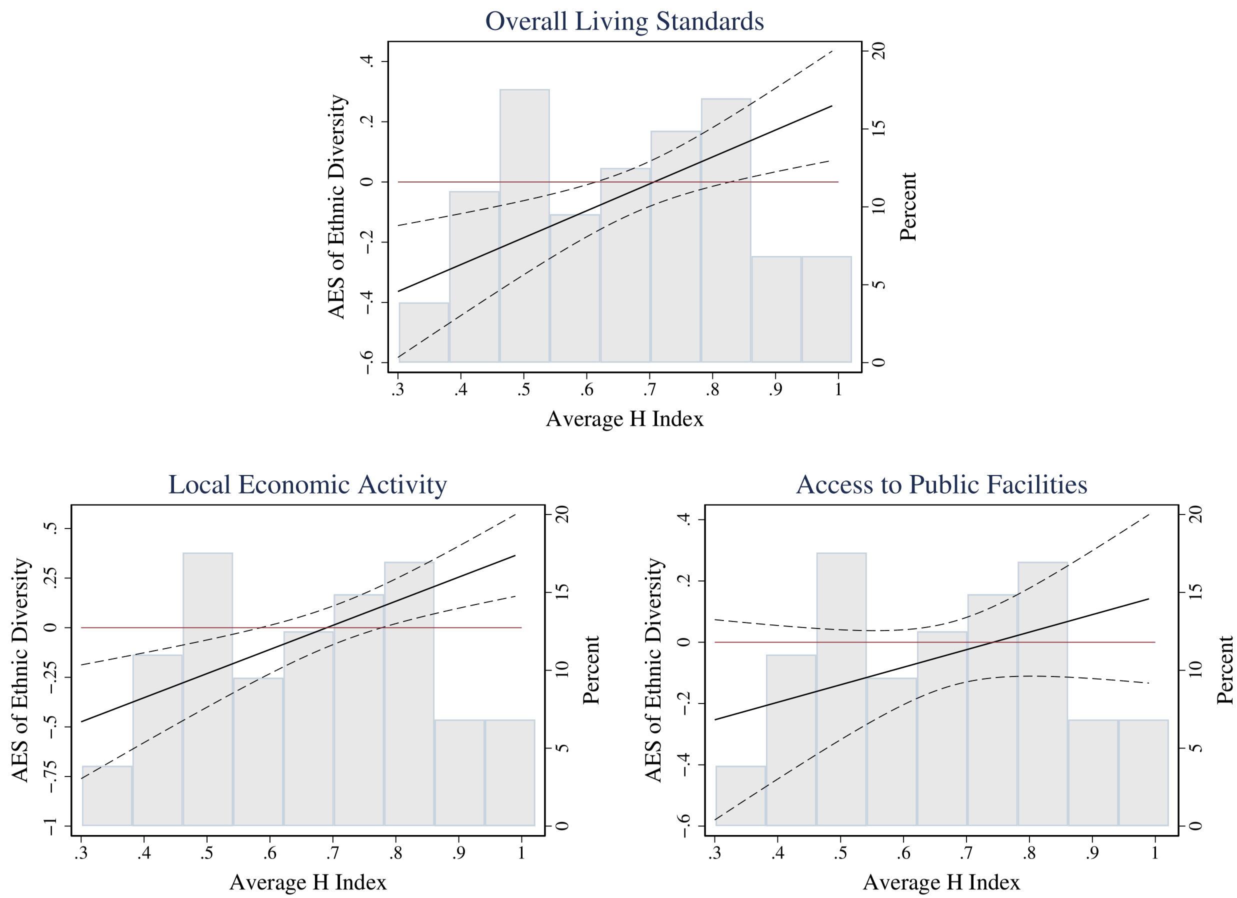
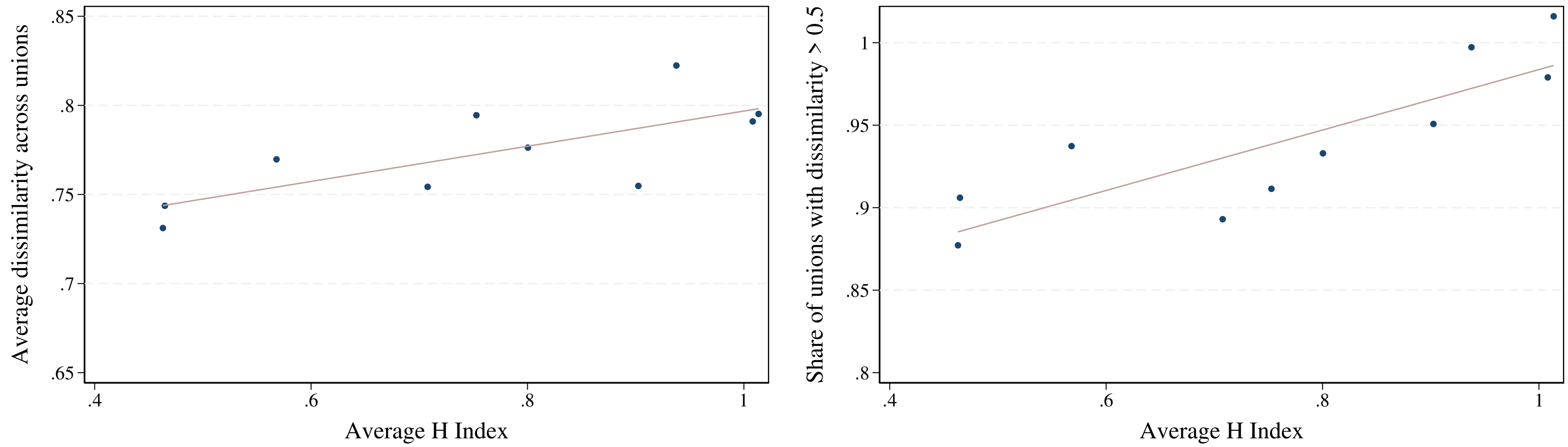
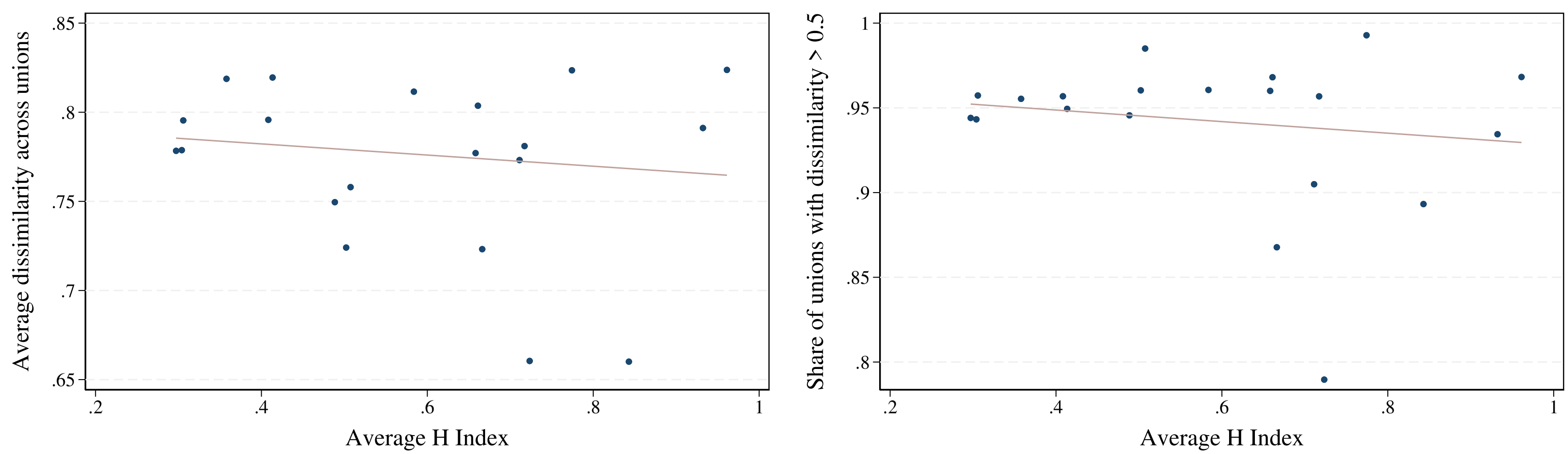


FIGURE 5: The Effect of Ethnic Diversity on Contemporary Development

Notes. The solid line represents the standardized average effect size (AES) of ethnic diversity after control variables and colonial province fixed effects. Dashed lines represent 95 percent confidence intervals. The x-axis and the histogram in the background refer to the average level of within-group heterogeneity among the ethnic groups concentrated in each parish (the right y-axis indicates the percentage of parishes). The AES for local economic activity refers to the log of average light intensity per capita (2010–2013) and non-subsistence agriculture—a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median (2012). The AES for access to public facilities refers to the share of dwellings with access to public sanitation (2017) and the share of dwellings with access to the public water network (2017). The AES for overall living standards refers to the four variables.



(a) Parishes with ethnic diversity (10 parishes)



(b) Parishes without ethnic diversity (31 parishes)

FIGURE 6: Within-Group Heterogeneity and Inter-Group Unions (1605–1870)

Notes. The figure presents binned scatterplots. In the graphs on the left-hand side, the y-axis variable is average dissimilarity, defined as the mean normalized Levenshtein distance across all unions of the parish. In the graphs on the right-hand side, the y-axis variable refers to the share of unions with dissimilarity above 0.5. All graphs control for the log number of individuals found in the parish records, mitigating concerns about parish size influencing the results. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise.

Table 1: Balance Tests for Ethnic Diversity

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

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

Table 2: Within-Group Heterogeneity and Pre-Colonial Diets

	Dependent Variable: Carbon Isotope Score ($\delta^{13}C_{col}$)				
	(1)	(2)	(3)	(4)	(5)
H index (ethnic group level)	0.501	0.279	0.851	0.261	
	[0.153]**	[0.089]**	[0.145]***	[0.071]***	
	(0.145)***	(0.094)***	(0.107)***	(0.069)***	
H index (grid cell level)					-0.028
					[0.334]
					(0.194)
Number of individuals	196	196	196	196	196
Zone FE	No	Yes	Yes	Yes	Yes
Ln (# native crops)	No	No	Yes	No	No
Land caloric suitability	No	No	No	Yes	Yes
Number of ethnic groups	8	8	8	8	–
Number of grid cells (50 × 50 km)	–	–	–	–	8

Notes. The unit of observation is the individual. The table reports OLS estimates. Standard errors clustered at the ethnic group (columns 1-4) or grid cell (column 5) level in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). *Ln (# native crops)* refers to the number of different native crops found within the ethnic homeland (see Appendix A.1). *Land caloric suitability* refers to the average caloric suitability of the ethnic homeland or grid cell (Galor and Özak 2016). All regressions control for the longitude and latitude of the archaeological site where the individual's remains were found. All variables are standardized to have zero mean and standard deviation equal to one. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Pre-Colonial Correlates of Within-Group Heterogeneity

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

Notes. The unit of observation is the ethnic group. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). All variables except for dummies are standardized to have zero mean and standard deviation equal to one. Population is available for 46 (out of 47) groups; see Appendix C. The dummy variables for urbanization, political complexity, and elite residences take value 1 for 12.77, 21.28, and 21.28 percent of the groups, respectively.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 4: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development

	Overall Living Standards (AES)							Ln Pop. Den. and Urban Status (AES)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 1990 – 2000								
Ethnic diversity	-0.200***	-0.167**	-0.117**	-0.826***	-0.554**	-0.598***	-0.721***	0.270
	[0.070]	[0.066]	[0.060]	[0.261]	[0.222]	[0.208]	[0.178]	[0.247]
Ethnic div. × Av. H index				0.938***	0.655**	0.780**	0.978***	-0.442
				[0.364]	[0.329]	[0.306]	[0.252]	[0.371]
Joint significance (p-value)				0.001	0.013	0.010	0.000	0.482
Panel B: 2010 – 2020								
Ethnic diversity	-0.143**	-0.125*	-0.104*	-0.842***	-0.555**	-0.616**	-0.632***	0.034
	[0.067]	[0.065]	[0.062]	[0.281]	[0.259]	[0.249]	[0.191]	[0.287]
Ethnic div. × Av. H index				1.036***	0.677*	0.879**	0.894***	-0.027
				[0.394]	[0.369]	[0.386]	[0.275]	[0.465]
Joint significance (p-value)				0.005	0.055	0.036	0.004	0.968
Baseline controls	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Ecclesiastical jurisd. FE	No	No	Yes	No	Yes	Yes	Yes	Yes
Colonial province FE	No	No	No	No	No	Yes	Yes	Yes
Ln pop. den. and urban status	No	No	No	No	No	No	Yes	No
Number of parishes	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Columns 1-7 report the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2000–2003 in Panel A and 2010–2013 in Panel B), non-subsistence agriculture—a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median (1994 in Panel A and 2012 in Panel B), the share of dwellings with access to public sanitation (1993 in Panel A and 2017 in Panel B), and the share of dwellings with access to the public water network (1993 in Panel A and 2017 in Panel B). Column 8 reports the AES for the log of population density and a dummy taking value 1 if the share of urban population is above the median (1993 in Panel A and 2017 in Panel B). Robust standard errors (columns 1-5), clustered at the province level (columns 6-8), in brackets. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. The p-value refers to the joint significance of ethnic diversity terms.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development — Pre-Colonial Characteristics of Ethnic Groups

	Overall Living Standards (AES, 2010 – 2020)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ethnic diversity	-0.616**	-0.891***	-0.690***	-0.719***	-0.618**	-0.611**	-0.625**	-0.614***	-0.958**	-0.360
	[0.249]	[0.335]	[0.257]	[0.262]	[0.274]	[0.259]	[0.245]	[0.223]	[0.446]	[0.579]
Ethnic div. × Av. H index	0.879**	0.851**	1.064***	0.709*	0.903**	0.898**	0.996***	0.900**	1.318***	1.174***
	[0.386]	[0.358]	[0.401]	[0.363]	[0.397]	[0.390]	[0.385]	[0.367]	[0.448]	[0.393]
Ethnic div. × Av. elevation		0.378							-0.248	-0.569
		[0.319]							[0.538]	[0.554]
Ethnic div. × Av. caloric suitability			-0.733						-1.423	-1.424*
			[0.526]						[0.989]	[0.858]
Ethnic div. × Av. ln river density				0.210					0.347	0.298
				[0.143]					[0.231]	[0.244]
Ethnic div. × Av. ln population density					0.019				-0.006	-0.059
					[0.113]				[0.135]	[0.149]
Ethnic div. × Av. urbanization						-0.070			0.085	
						[0.194]			[0.335]	
Ethnic div. × Av. political complexity							-0.133		-0.277	-0.134
							[0.144]		[0.242]	[0.198]
Ethnic div. × Av. elite residences								0.008	0.266	
								[0.113]	[0.216]	
Ethnic div. × Av. water canals										-0.094
										[0.170]
Ethnic div. × Av. ln road density										-0.058
										[0.068]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lasso characteristics	No	No	No	No	No	No	No	No	No	Yes
Number of parishes	336	336	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2010–2013), non-subsistence agriculture (2012, a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). Column 10 shows the coefficients of lasso-selected characteristics (Belloni, Chernozhukov, and Hansen 2014). The lasso algorithm uses the first principal component of the four previous outcomes as the dependent variable. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 6: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development — Coarsened Exact Matching

	Overall Living Standards (AES, 2010 – 2020)					
	CEM Matched Sample (Baseline Characteristics)		CEM Matched Sample (Lasso Characteristics)		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic diversity	-0.490***	-0.447	-0.475***	-0.841**	-0.226***	-0.201*
	[0.165]	[0.401]	[0.152]	[0.351]	[0.072]	[0.106]
Ethnic div. × High Av. H (dummy)	0.737**	0.737**	0.889***	0.945***	0.378***	0.378***
	[0.287]	[0.288]	[0.225]	[0.255]	[0.129]	[0.129]
Number of parishes	103	103	78	78	336	336
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes
Majority group weight	No	Yes	No	Yes	No	Yes

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The matched sample refers to the sample of parishes selected by the coarsened exact matching (CEM) algorithm (Iacus, King, and Porro 2012) as the counterfactual group for parishes with average within-group heterogeneity above the median (*High Av. H* dummy). The pre-colonial characteristics to be balanced by the algorithm are the baseline characteristics (Column 9 of Table 5) in Panel A and the lasso-*selected* characteristics (Column 10 of Table 5) in Panel B. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2010–2013), non-subsistence agriculture (2012, a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. Columns 2, 4 and 6 control for the area share of the ethnic group with the highest share of area within the 10-km buffer. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 7: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development — Randomization Inference

		p-value				
	Coeff.	Baseline	Minimum	Maximum	Randomized	# Parishes
	Dependent Variable: Non-Subsistence Agriculture (2012)					
Ethnic diversity	-0.292	0.093	0.089	0.090	0.089	336
Ethnic div. × Av. H index	0.596	0.027	0.031	0.032	0.031	336
	Dependent Variable: Light Intensity per capita (2010–2013)					
Ethnic diversity	-0.108	0.004	0.008	0.009	0.008	336
Ethnic div. × Av. H index	0.121	0.010	0.019	0.020	0.019	336

Notes. The unit of observation is the parish. The coefficient and baseline p-values refer to OLS regressions with robust standard errors clustered at the level of the colonial province. The outcomes are the log of average light intensity per capita and non-subsistence agriculture—a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median. The regressions include baseline controls (see Table 6) and colonial province fixed effects. The minimum, maximum and randomized p-values result from permuting *Ethnic div.* (1,000 iterations) and recalculating *Av. H index* accordingly; see Young (2023).

TABLE 8: Mechanisms: Cultural Transmission

	Dependent Variable:								
	Which Group Do You Identify With The Most? (2004-2017)			Voting in Presidential Elections (2007-2011)		Military Service (2008-2014)		Dummy Neighborhood Associations	
	State	Ethnicity or Race	Religion	Dummy (Yes, I Voted in the 2006 Election)		Ln (1 + Av. Volunteers)	Dummy Volunteers	2002-2003	2012-2014
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ethnic diversity	-0.155	0.020	0.001	-0.052	-0.052	-0.129	-0.230	-0.436	-0.529
	[0.050]***	[0.013]	[0.027]	[0.035]	[0.035]	[0.067]*	[0.106]**	[0.206]**	[0.162]***
	(0.056)***	(0.018)	(0.030)	(0.043)	(0.044)	(0.056)**	(0.076)***	(0.224)*	(0.230)**
Ethnic div × Av. H index	0.208	-0.026	0.006	0.104	0.104	0.243	0.365	0.693	0.711
	[0.071]***	[0.018]	[0.038]	[0.047]**	[0.047]**	[0.114]**	[0.148]**	[0.313]**	[0.234]***
	(0.088)**	(0.024)	(0.049)	(0.054)*	(0.054)*	(0.102)**	(0.127)***	(0.341)**	(0.329)**
Trust in the state					0.011				
					[0.011]				
					(0.012)				
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	–	–	–	–
Year FE	Yes	Yes	Yes	Yes	Yes	–	–	–	–
Number of parishes	280	280	280	186	186	336	336	336	336
Number of individuals	52,875	52,875	52,875	17,422	17,422	–	–	–	–
Mean Dep. Var.	0.511	0.037	0.170	0.831	0.831	0.064	0.134	0.372	0.354

Notes. The unit of observation is the individual in columns 1-5 (data from yearly waves of the ENAHO Peruvian household survey) and the parish in columns 6-9. In brackets, robust standard errors clustered at the parish (columns 1-5) or province (columns 6-9) levels. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The question on having voted in the 2006 presidential election (columns 4-5) was only asked in the 2007-2011 waves of the survey, and it was not repeated for other presidential elections. In columns 1-5, the vector of individual-level controls includes gender, age, age squared, years of schooling, civil status, and language spoken at home. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.

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

TABLE 9: Mechanisms: Economic Complementarities and Tertiary-Sector Occupation

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Economic Complementarities					
	Overall Living Standards (AES, 1990 – 2000)		Dummy Agricultural Retail Market (1993)		Access to Agricultural Land (Mean Share of Pop, 1700 – 1800)	
High min	0.684***	0.957***	0.343**	0.292*	0.368***	0.360***
	[0.230]	[0.346]	[0.165]	[0.152]	[0.103]	[0.125]
High maj	0.244	0.393	-0.058	-0.100	-0.060	-0.018
	[0.280]	[0.263]	[0.073]	[0.083]	[0.058]	[0.062]
High min × High maj		-0.494	-0.191	-0.195	-0.284*	-0.337*
		[0.647]	[0.154]	[0.147]	[0.153]	[0.179]
Ln population (1993)				0.078*		
				[0.039]		
Ln mean population (1700 – 1800)						-0.059*
						[0.029]
Number of parishes	117	117	117	117	51	51
Mean Dep. Var.	–	–	0.077	0.077	0.828	0.828
	Panel B: Tertiary-Sector Occupation (Share of Pop)					
	Tertiary Sector		Tertiary Sector: Local Trade		Tertiary Sector: Other Services	
	(1876)	(2007 – 2017)	(1876)	(2007 – 2017)	(1876)	(2007 – 2017)
Ethnic diversity (dummy)	-0.111***	-0.250**	-0.110***	-0.108**	-0.001	-0.077*
	[0.041]	[0.104]	[0.039]	[0.044]	[0.003]	[0.038]
Ethnic div. × Av. H index	0.151**	0.340*	0.152**	0.147**	-0.001	0.106
	[0.067]	[0.170]	[0.065]	[0.067]	[0.004]	[0.066]
Number of parishes	282	336	282	336	282	336
Mean Dep. Var.	0.073	0.336	0.064	0.124	0.009	0.104
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. *High min* is a dummy variable indicating that the minority group's H index is above the 75th percentile. Variable *High maj* is an analogous dummy for the majority group. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. In Panel B, regressions are weighted by the square root of the total population. Outcome variables are indicated at the top of each column. The outcome variables for the period 2007-2017 refer to the average from the 2007 and 2017 population censuses. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A Online Appendix - Figures and Tables

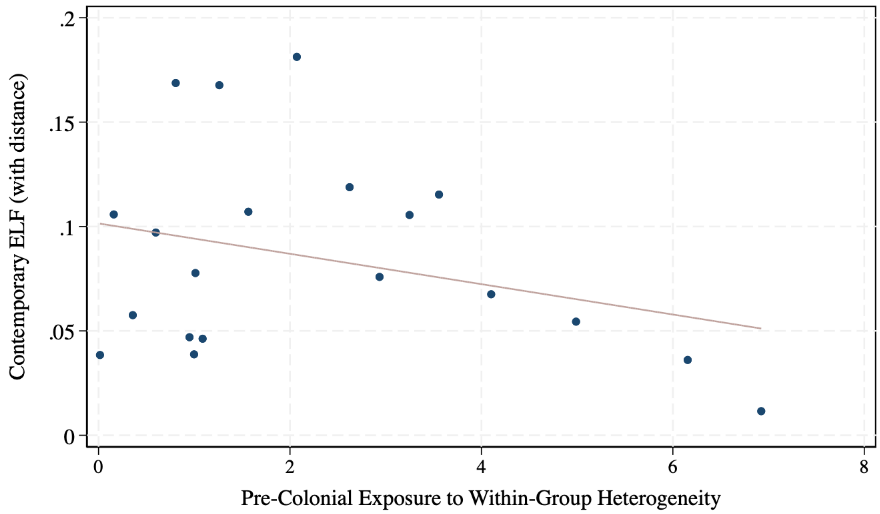


FIGURE A.1: Country-Level Correlation

Notes. Binscatter between country-level ethnolinguistic fractionalization and pre-colonial exposure to within-group heterogeneity in specialized activities. Contemporary ethnolinguistic fractionalization (with linguistic distance) is computed using ethnolinguistic groups from the *Ethnologue* database. The x-axis refers to a weighted average of the number of specialized activities within pre-colonial groups, considering the corresponding ancestors of *Ethnologue* groups in the *Ethnographic Atlas*. The number of specialized activities refers to variables 55 to 65 from the *Ethnographic Atlas*, for specialization in metal working, weaving, leather working, pottery making, boat building, house construction, gathering, hunting, fishing, animal husbandry, and agriculture. Weights refer to the population share of each *Ethnologue* group in a country, computed using population data from *LandScan*.

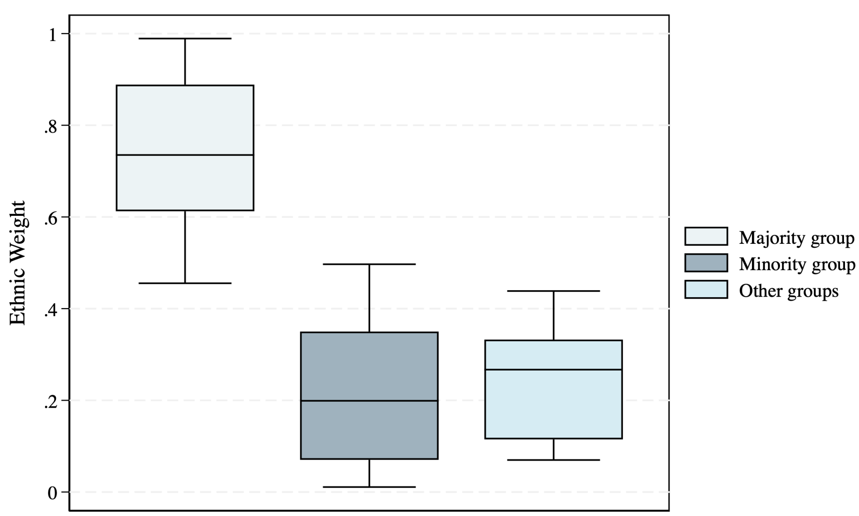


FIGURE A.2: Boxplots of Ethnic Weights

Notes. The ethnic weight (w_{pe}) is defined as the area share of ethnic group e within a 10-km buffer from the capital of parish p. The boxplots refer to the ethnic weights of the majority group (i.e., the group with the highest area share within the 10-km buffer), the minority group (i.e., the group the lowest area share), and other groups, considering parishes with ethnic diversity. The category "Other groups" applies only for parishes that concentrated either three or four groups (15 percent of parishes with ethnic diversity).

Figure A.3: Density of Within-Group Heterogeneity

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

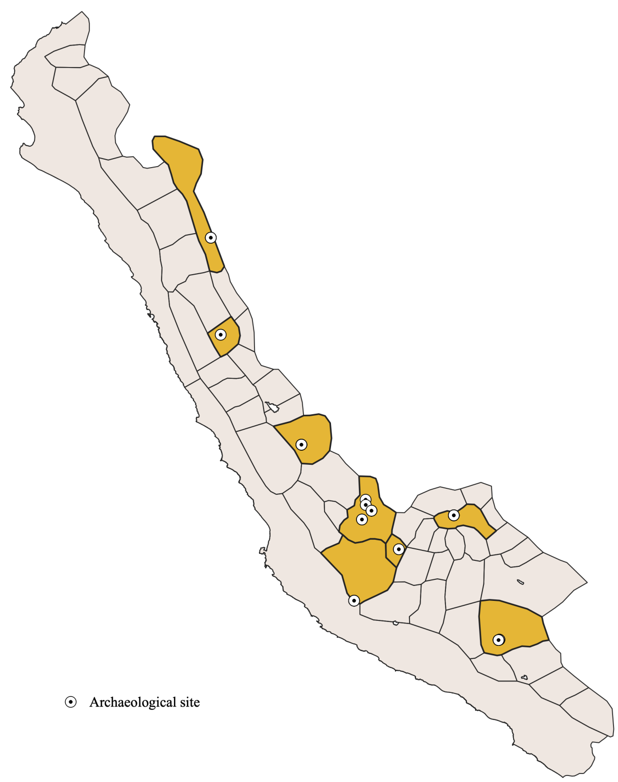
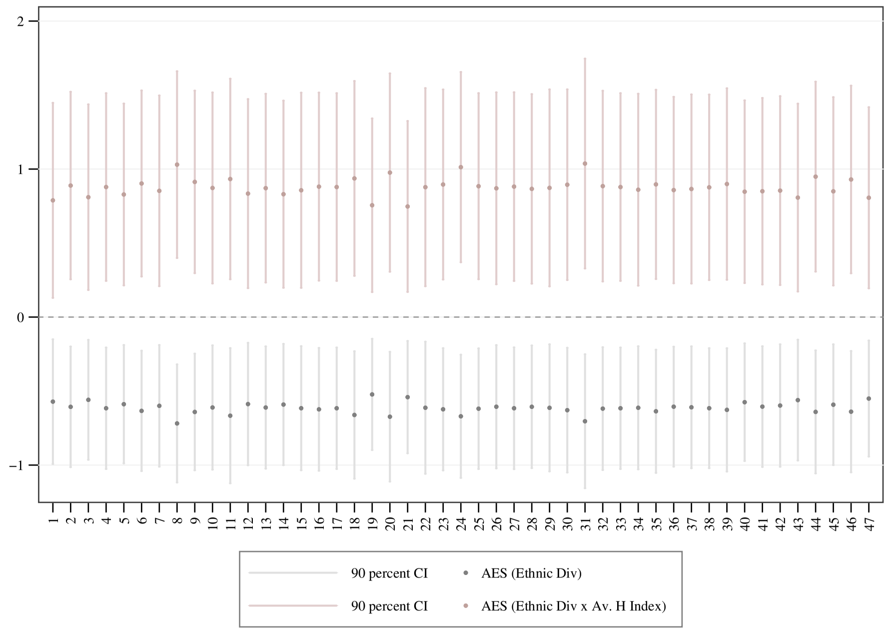
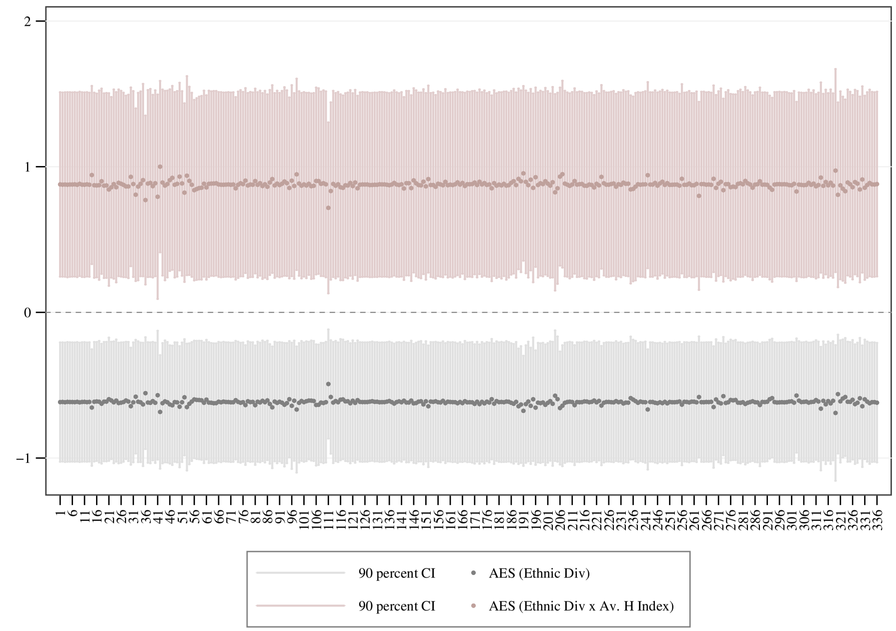


Figure A.4: Archaeological Sites with Information on Pre-Colonial Diets

Notes. Archaeological sites with information on pre-colonial individual diets. Geographic coordinates are from Wilson et al. (2022). Lines in black represent the extent of pre-colonial groups according to Rowe (1946)'s map.



(a) Influential Ethnic Groups



(b) Influential Parishes

FIGURE A.5: Influential Observations

Notes. Standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) and 90 percent confidence intervals after control variables and colonial province fixed effects. The standardized AES refers to the following outcomes: the log of average light intensity per capita (2010–2013), non-subsistence agriculture (2012, a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). In panel (a), each regression excludes all parishes with presence of a given ethnic group (indicated on the x-axis) at a time. In panel (b), each regression excludes one parish (indicated on the x-axis) at a time.

Table A.1: Balance Tests−Additional Pre-Colonial Characteristics

	Ethnic Diversity = 1		Ethnic Diversity = 0				
	mean	sd	mean	sd	Diff.	p-value[a]	p-value[b]
(1) Ln dist. to defensive site	4.186	0.705	4.161	0.927	-0.025	[0.783]	[0.856]
(2) Ln dist. to urban site	4.209	0.741	4.125	0.982	-0.084	[0.382]	[0.559]
(3) Ln dist. to political site	4.086	0.672	3.931	0.984	-0.155	[0.090]*	[0.234]
(4) Ln dist. to elite residence	4.041	1.018	3.926	1.259	-0.115	[0.366]	[0.204]
(5) Ln dist. to road	1.409	2.597	0.983	2.458	-0.427	[0.145]	[0.187]
(6) Ln dist. to canal	3.900	0.810	3.913	0.908	0.012	[0.900]	[0.908]
(7) Ln dist. to bridge	4.249	0.760	4.250	0.783	0.001	[0.991]	[0.992]
(8) Caloric suitability for maize	357.869	1010.757	304.422	899.243	-53.447	[0.632]	[0.691]
(9) Caloric suitability for potato	595.073	755.261	623.936	800.354	28.863	[0.744]	[0.851]
Number of parishes	117	117	219	219	336	336	336

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

Table A.2: Frequency of Parishes by Ethnic Diversity and Elevation Zone

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

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

A.1 Validating Complementary Elevation Zones

This section explores whether the measure of within-group heterogeneity does indeed explain crop variety in the data. For this exercise, one would ideally use land suitability data for all crops available before 1500. In the absence of these data, I use information on native crops from modern sources. I rely on the 2012 agricultural census, which provides geo-referenced data for an extensive set of native crops. I explore the determinants of crop variety across grid cells of different sizes, as well as across ethnic groups, based on the number of native crops reported by farmers at the time of the census.[42]

I start by computing H for 25 km × 25 km grid cells covering the entire study region. In Column 1 of Table A.3, I regress the log of the number of crops on this measure. The estimated beta coefficient is positive (0.523) and statistically significant. The coefficient remains stable in magnitude and statistical significance when I control for another potential predictor of crop variety—variation in elevation (Column 3). According to unconditional OLS estimates, a one standard deviation increase in H is associated with a 0.523 standard deviation increase in log crop variety (Column 1). In the case of variation in elevation, the associated standard deviation increase in log crop variety is 0.398 (Column 2). However, this coefficient becomes small (0.007) and statistically insignificant after I control for H (Column 3).

The same pattern arises when including fixed effects for the 62 hydrographic basins in the study region (Column 4), log area (Column 5), and mean elevation (Column 6). In Column 7, I substitute the grid-level H index with dummy variables indicating the number of zones within the grid cell. The magnitudes of the estimated coefficients increase with the number of zones (relative to those for grid cells with only one zone). Columns 8 and 9 report similar results across 50 km × 50 km grid cells and ethnic groups, respectively.

[42] I follow the classification of native crops in Tapia (2013), who identifies 41 main native crops. The 2012 agricultural census covers 38 of these crops. Farmers were only required to report the list of crops harvested at the time of the census—crop rotation and fallow practices can affect the list of reported crops.

Table A.3: Validating Elevation Zones

	Dependent Variable: Ln (# Native Crops)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
H index	0.523		0.518	0.505	0.483	0.498		0.531	0.527
	[0.040]***		[0.042]***	[0.050]***	[0.049]***	[0.050]***		[0.064]***	[0.106]***
	(0.047)***		(0.064)***	(0.048)***	(0.048)***	(0.047)***		(0.084)***	(0.086)***
SD of elevation		0.398	0.007	0.064	0.072	-0.025	0.001	-0.041	0.135
		[0.043]***	[0.043]	[0.073]	[0.074]	[0.067]	[0.072]	[0.068]	[0.141]
		(0.061)***	(0.074)	(0.077)	(0.077)	(0.061)	(0.079)	(0.034)	(0.157)
Ln land area					0.124	0.157	0.172	0.413	0.387
					[0.019]***	[0.020]***	[0.027]***	[0.048]***	[0.126]***
					(0.021)***	(0.021)***	(0.031)***	(0.047)***	(0.124)***
Mean elevation						-0.193	-0.279	-0.022	0.134
						[0.059]***	[0.081]***	[0.058]	[0.130]
						(0.062)***	(0.090)***	(0.060)	(0.106)
Dummy (number of zones=2)							1.067		
							[0.166]***		
							(0.225)***		
Dummy (number of zones=3)							1.838		
							[0.167]***		
							(0.215)***		
Dummy (number of zones=4)							2.027		
							[0.205]***		
							(0.246)***		
Hydrographic basin FE	No	No	No	Yes	Yes	Yes	Yes	No	No
Observations	526	526	526	526	526	526	526	148	47

Notes. The unit of observation is the 25 km × 25 km grid cell in Columns 1-7, the 50 km × 50 km grid cell in Column 8, and the ethnic group in Column 9. The table reports OLS estimates. Robust standard errors in brackets; clustered at the basin level in Columns 4-7. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). All variables except dummies are standardized to have zero mean and standard deviation equal to one. All regressions control for longitude and latitude.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: H Index and Ethnic Diversity

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

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

Table A.5: Average H Index and Ethnic Diversity

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

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

Table A.6: Pre-Colonial Correlates of Within-Group Heterogeneity: Infrastructure

	Dependent Variable:				
	Dummy Storage	Dummy Terraces	Ln Road Density	Dummy Canals	Dummy Bridges
	(1)	(2)	(3)	(4)	(5)
H index	0.032	0.058	0.125	-0.008	-0.038
	[0.052]	[0.032]*	[0.131]	[0.062]	[0.056]
	(0.054)	(0.029)**	(0.143)	(0.046)	(0.045)
Number of ethnic groups	47	47	47	47	47

Notes. The unit of observation is the ethnic group. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). All variables except dummies are standardized to have zero mean and standard deviation equal to one. The dummy variables for food storage structures, terraces, canals, and bridges take value 1 for 14.89, 10.64, 27.66, and 23.40 percent of the groups, respectively. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.7: Summary Statistics for Contemporary Outcomes

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

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

Table A.8: Light Intensity and Geographic Proximity to Ethnic Boundaries

	Dependent Variable: Light Intensity per capita (2000 – 2003)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnically diverse parish (dummy)	-0.067	-0.089			-0.059	-0.082	-0.052	-0.075
	[0.028]**	[0.028]***			[0.024]**	[0.025]***	[0.024]**	[0.024]***
	(0.040)*	(0.037)**			(0.031)*	(0.031)***	(0.030)*	(0.030)**
	(0.041)	(0.039)**			(0.035)*	(0.038)**	(0.034)	(0.038)**
Ethnic border < 10km (dummy)			-0.034	-0.038	-0.018	-0.016		
			[0.034]	[0.032]	[0.033]	[0.032]		
			(0.049)	(0.043)	(0.048)	(0.042)		
			(0.050)	(0.038)	(0.050)	(0.039)		
Ln dist. to ethnic border							0.015	0.015
							[0.015]	[0.014]
							(0.026)	(0.020)
							(0.027)	(0.017)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2730	2730	2730	2730	2730	2730	2730	2730

Notes. The unit of observation is the 10 km × 10 km grid cell. The table reports OLS estimates. Robust standard errors in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately 0.5 and 1 degree at the equator (Colella et al. 2019). The dependent variable is the log of average light intensity per capita (2000-2003). *Ethnically diverse parish* takes value 1 if the grid cell is within the buffer of an ethnically diverse parish (considering the buffer of 10-km radius from the parish capital), and 0 otherwise. The vector of control variables includes mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, and log river density. All variables except dummies are standardized to have zero mean and standard deviation equal to one.

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

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

	Overall Living Standards (AES)					
	Panel A: 1990 – 2000			**Panel B:** 2010 – 2020		
	Full Sample	Ethnic Div = 1	Ethnic Div = 0	Full Sample	Ethnic Div = 1	Ethnic Div = 0
	(1)	(2)	(3)	(4)	(5)	(6)
Average H index	0.481**	0.834*	0.420*	0.545**	0.986**	0.432
	[0.192]	[0.495]	[0.223]	[0.218]	[0.439]	[0.267]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Ecclesiastical jurisd. FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	336	117	219	336	117	219

Notes. The unit of observation is the parish. Robust standard errors in brackets. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2000–2003 in Panel A and 2010–2013 in Panel B), non-subsistence agriculture—a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median (1994 in Panel A and 2012 in Panel B), the share of dwellings with access to public sanitation (1993 in Panel A and 2017 in Panel B), and the share of dwellings with access to the public water network (1993 in Panel A and 2017 in Panel B). Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.

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

TABLE A.10: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development−Individual Effects

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

Notes. The unit of observation is the parish. The table reports OLS estimates. Robust standard errors clustered at the level of the colonial province in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The dependent variables are: (1) the log of average light intensity per capita (2000-2003 in Panel A and 2010-2013 in Panel B), (2) non-subsistence agriculture—a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median (1994 in Panel A and 2012 in Panel B), (3) the share of dwellings with access to public sanitation (1993 in Panel A and 2017 in Panel B), and (4) the share of dwellings with access to the public water network (1993 in Panel A and 2017 in Panel B). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. The urban dummy variable takes vale 1 if the share of urban population is above the median, and 0 otherwise (1993 in Panel A and 2017 in Panel B).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.11: Balance Tests for Ethnic Diversity – Matched Samples

	Dependent Variable:							
	Mean Elevation	SD of Elevation	Mean Caloric Suit.	SD of Caloric Suit.	Ln Dist Perennial River	Ln Dist Native Shrine	Ln Expected Tribute (16th c.)	Ln Dist *Mita* Mine
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Matched Sample Based on CEM for Baseline Pre-Colonial Characteristics								
Ethnic diversity	59.402	27.378	-4.773	-10.989	-0.281	-0.012	-0.135	0.405**
	[0.377]	[0.574]	[0.635]	[0.476]	[0.207]	[0.948]	[0.539]	[0.046]
Number of parishes	103	103	103	103	103	103	103	103
Panel B: Matched Sample Based on CEM for Lasso-*Selected* Pre-Colonial Characteristics								
Ethnic diversity	-78.014	17.340	14.227	13.645	-0.583	-0.099	-0.241	0.232
	[0.461]	[0.763]	[0.125]	[0.249]	[0.146]	[0.762]	[0.244]	[0.413]
Number of parishes	78	78	78	78	78	78	78	78

Notes. The unit of observation is the parish. Robust standard errors in brackets. The matched sample refers to the sample of parishes selected by the coarsened exact matching (CEM) algorithm (Iacus, King, and Porro 2012) as the counterfactual group for parishes with average within group heterogeneity above the median. The pre-colonial characteristics to be balanced by the algorithm are the baseline characteristics (Column 9 of Table 5) in Panel A and the lasso-*selected* characteristics (Column 10 of Table 5) in Panel B. The table reports estimates from using the corresponding matching weights. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.12: Robustness: Pre-Colonial Land Occupation, Transition Zones, and Placebos

	Overall Living Standards (AES, 2010 – 2020)					
	Pre-Colonial Land Occupation		Transition Zones		Placebos	
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic diversity	-0.648**	-0.779**	-0.656**	-0.687**		
	[0.260]	[0.310]	[0.277]	[0.309]		
Ethnic div. × Av. H index (20km correction)	0.922**					
	[0.401]					
Ethnic div. × Av. H index (10km correction)		1.057**				
		[0.452]				
Ethnic div. × Av. H index (20km transition-zone buffer)			0.918**			
			[0.418]			
Ethnic div. × Av. H index (10km transition-zone buffer)				0.946**		
				[0.454]		
Dummy (Artificial ethnic border within parish buffer)					-0.267	
					[0.495]	
Dummy × Av. H index (Artificial)					0.388	
					[0.617]	
Dummy (*Corregimiento* border within parish buffer)						0.149
						[0.314]
Dummy × Av. H index (*Corregimiento*)						-0.203
						[0.421]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2010–2013), non-subsistence agriculture (2012, a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.13: Robustness: Varying the Buffer Size

	Overall Living Standards (AES, 2010 – 2020)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	7km	8km	9km	10km	11km	12km	13km	14km	15km
Ethnic diversity	-0.439*	-0.444**	-0.511***	**-0.616****	-0.422**	-0.460**	-0.517**	-0.551**	-0.514**
	[0.247]	[0.221]	[0.195]	**[0.249]**	[0.202]	[0.206]	[0.201]	[0.236]	[0.243]
Ethnic div. × Av. H index	0.686*	0.683**	0.777**	**0.879****	0.578*	0.690**	0.816**	0.867**	0.773*
	[0.380]	[0.336]	[0.312]	**[0.386]**	[0.343]	[0.344]	[0.347]	[0.385]	[0.409]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	336	336	336	336	336	336	336	336	336
% parishes intersecting ethnic border	27.38	30.36	32.14	34.82	37.20	40.18	42.26	45.24	46.43

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of r-km radius from the parish capital, for $r \in \{7, 8, 9, 10, 11, 12, 13, 14, 15\}$ (indicated at the top of each column), and 0 otherwise. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2010–2013), non-subsistence agriculture (2012, a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.14: Robustness: Alternative Diversity Indices

	Overall Living Standards (AES, 2010 – 2020)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic frac	-1.583**	-1.167*						
	[0.634]	[0.639]						
Ethnic frac × Av. H index	1.964**	1.788*						
	[0.881]	[0.929]						
Ethnic diversity (dummy)			-0.955***	-0.668**	-0.646**	-0.515**	-0.898***	-0.651***
			[0.321]	[0.290]	[0.255]	[0.225]	[0.245]	[0.204]
Ethnic div × Av. $\widetilde{H}$ index			1.440***	1.140**				
			[0.542]	[0.529]				
Ethnic div × Av. H index (majority group)					0.738**	0.716**		
					[0.375]	[0.342]		
Ethnic div × Av. H index (minority group)							1.188***	0.986***
							[0.336]	[0.303]
Baseline controls	No	Yes	No	Yes	No	Yes	No	Yes
Colonial province FE	No	Yes	No	Yes	No	Yes	No	Yes
Number of parishes	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. Columns 1-2 use $\textit{ethnic frac}_p = 1 - \sum_e w_{pe}^2$ as a measure of ethnic diversity, while columns 3-8 use *ethnic div*, which takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. As a measure of within-group heterogeneity, columns 1-2 use the weighted average defined in the main text, $H_p = \sum_e w_{pe} H_e$, where $H_e = 1/\sum_j s_{ej}^2$; columns 3-4 use a Herfindahl index to measure H_e instead, specifically $\widetilde{H}_e = 1 - \sum_j s_{ej}^2$; columns 5-6 use within-group heterogeneity (H_e) of the majority group in the parish, defined as the group with the highest area share within the 10-km buffer, instead of a weighted average; while columns 7-8 use within-group heterogeneity (H_e) of the minority group. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2010–2013), non-subsistence agriculture (2012, a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.2 Household Consumption and Selective Migration

Although population movement was restricted after the resettlement, individuals may have managed to escape. For example, the most *capable* individuals (i.e., those from more heterogeneous ethnic groups) from ethnically diverse parishes may have tried to join their coethnics located in parishes with an ethnically homogeneous founding population. This type of selective migration, which represents a return to the pre-resettlement configuration, may help explain the overall negative impacts of ethnic diversity if the characteristics these individuals brought with them were (*i*) heritable and (*ii*) relevant for economic development (Dell 2010; Lowes and Montero 2021).

In Table A.15, I use survey data on household consumption to explore this possibility. The outcome variable is the log of real household consumption per capita. I use the spatial deflators provided by the Peruvian Institute of Statistics and follow Dell (2010) in subtracting public transfers received by the household. Real household consumption without transfers is divided by the number of household members to obtain a per capita measure. Although the surveys are not representative at the local level, they allow me to trim the sample by dropping the top x percent of households from more heterogeneous ethnic groups (i.e., among those with average within-group heterogeneity above the median) located in parishes with an ethnically homogeneous founding population, for different bandwidths (x). The annual surveys from 2004 to 2007 cover 219 parishes, of which 137 are parishes built on ethnically homogeneous populations. Columns 1 and 2 show that the main result of the paper holds for household consumption. The results for the trimmed samples in the remaining columns are consistent with the main result of the paper. Selective migration after independence from Spanish rule might also help explain the results. However, data from the 1993 population census suggest low migration rates—non-native individuals represent only 0.008 percent of a parish's population in the sample, on average.

Table A.15: Ethnic Diversity, Within-Group Heterogeneity, and Household Consumption

	Dependent Variable: Log Real Household Consumption Per Capita						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ethnic diversity	-1.264	-0.515	-0.473	-0.446	-0.443	-0.438	-0.429
	[0.434]***	[0.238]**	[0.224]**	[0.221]**	[0.219]**	[0.217]**	[0.215]**
	(0.454)***	(0.195)***	(0.205)**	(0.211)**	(0.215)**	(0.215)**	(0.217)**
Ethnic div. × Av. H index	1.604	0.748	0.728	0.717	0.734	0.750	0.756
	[0.609]***	[0.351]**	[0.336]**	[0.333]**	[0.331]**	[0.329]**	[0.326]**
	(0.616)***	(0.339)**	(0.360)**	(0.367)*	(0.371)**	(0.372)**	(0.375)**
Baseline controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	219	219	219	219	219	219	219
Number of households	10,967	10,967	10,772	10,609	10,430	10,258	10,092
% Trimmed	–	–	2.2 %	4.2 %	6.2 %	8.2 %	10.2 %
R-Squared	0.205	0.370	0.356	0.346	0.337	0.331	0.324

Notes. The unit of observation is the household. The table reports OLS estimates. Robust standard errors clustered at the parish level in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The dependent variable is the log of real household consumption per capita (annual survey data for the period 2004–2007). In columns 3–7, the sample is trimmed by dropping the top x percent of households from more heterogeneous ethnic groups (i.e., among those with average within-group heterogeneity above the median) located in parishes without ethnic diversity; see % trimmed in the last row. The vector of individual-level controls includes gender, age, age squared, years of schooling, civil status, and language spoken at home of the household head. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A.16: Robustness: Inca-Period Variables and Pre-Resettlement Spread of Smallpox

	Overall Living Standards (AES, 2010 – 2020)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ethnic diversity	-0.616**	-0.218	-0.538**	-0.618**	-1.708	-0.613**	-0.610**
	[0.249]	[0.815]	[0.258]	[0.274]	[1.414]	[0.252]	[0.292]
Ethnic div. × Av. H index	0.879**	0.916**	0.959**	0.903**	1.071***	0.871**	0.887**
	[0.386]	[0.376]	[0.377]	[0.397]	[0.387]	[0.388]	[0.425]
Ethnic div. × Av. Ln dist. smallpox outbreak		-0.034			0.109		
		[0.061]			[0.130]		
Ethnic div. × Av. Ln road density			-0.046		-0.143		
			[0.061]		[0.121]		
Ethnic div. × Av. Ln population density				0.019	-0.070		
				[0.113]	[0.137]		
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inca region (*suyu*) FE	No	No	No	No	No	Yes	No
Excluding groups potentially affected by Inca resettlements	No	No	No	No	No	No	Yes
Number of parishes	336	336	336	336	336	336	241

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The table reports the standardized AES (Kling et al. 2004; Clingingsmith, Khwaja, and Kremer 2009) across four outcomes: the log of average light intensity per capita (2010–2013), non-subsistence agriculture (2012, a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median), the share of dwellings with access to public sanitation (2017), and the share of dwellings with access to the public water network (2017). I control for different variables that may have increased transmission risk during the first epidemic wave of smallpox after the Spanish conquest (1524–1526): log distance to the closest outbreak (Tomebamba or Cuzco, Column 2), log density of Inca roads, which connected Cuzco with the remaining territories (Column 3), and log population density (Column 4). I control for the weighted average of each of these ethnic-level variables among the groups concentrated in the parish and their interactions with ethnic diversity. The last two columns of the table present estimates after adding fixed effects for the four major Inca regions or *suyus* (Column 6) and excluding parishes that concentrated groups potentially affected by Inca resettlements (Column 7), according to Rowe (1946). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.17: Mechanisms: Cultural Transmission

	Dependent Variable:			
Panel A:	Military Service (2008-2014)		Dummy Neighborhood Associations	
	Ln (1 + Av. Volunteers)	Dummy Volunteers	2002-2003	2012-2014
	(1)	(2)	(3)	(4)
Ethnic diversity	-0.128	-0.228	-0.436	-0.531
	[0.048]**	[0.095]**	[0.205]**	[0.157]***
	(0.037)***	(0.073)***	(0.222)**	(0.220)**
Ethnic div × Av. H index	0.240	0.361	0.694	0.714
	[0.082]***	[0.130]***	[0.315]**	[0.227]***
	(0.076)***	(0.127)***	(0.344)**	(0.318)**
Ln population	Yes	Yes	Yes	Yes
Number of parishes	336	336	336	336
Panel B:	Participation Dummy (2004-2017)			
	Neighborhood Associations	Professional Associations	Labor Associations	Some Association
	(1)	(2)	(3)	(4)
Ethnic diversity	-0.059	-0.030	-0.058	-0.140
	[0.036]	[0.014]**	[0.031]*	[0.052]***
	(0.048)	(0.017)*	(0.033)*	(0.056)**
Ethnic div. × Av. H index	0.096	0.047	0.074	0.206
	[0.046]**	[0.021]**	[0.039]*	[0.068]***
	(0.066)	(0.030)	(0.039)*	(0.072)***
Individual controls and Year FE	Yes	Yes	Yes	Yes
Number of parishes	280	280	280	280
Number of individuals	54,676	54,676	54,676	54,676
Baseline controls	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes

Notes. The unit of observation is the parish in Panel A and the individual in Panel B (Individual-level data from yearly waves of the ENAHO Peruvian household survey.). In brackets, robust standard errors clustered at the province level (Panel A) or parish level (Panel B). In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. In panels A and B, the vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines. In Panel B, the vector of individual-level controls includes gender, age, age squared, years of schooling, civil status, and language spoken at home. Regressions in Panel B control for the log of the total number of associations per capita.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.18: Non-Subsistence Agriculture and Local Crop Availability

	Dependent Variable:					
	Non-Subsistence Agriculture (1994)			Non-Subsistence Agriculture (2012)		
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic diversity	-0.289	-0.297	-0.291	-0.292	-0.292	-0.289
	[0.162]*	[0.168]*	[0.161]*	[0.158]*	[0.159]*	[0.156]*
	(0.132)**	(0.134)**	(0.132)**	(0.155)*	(0.155)*	(0.151)*
Ethnic div. × Av. H index	0.511	0.519	0.514	0.596	0.596	0.592
	[0.265]*	[0.273]*	[0.264]*	[0.242]**	[0.244]**	[0.238]**
	(0.220)**	(0.222)**	(0.219)**	(0.214)***	(0.214)***	(0.207)***
Ln (# Native Crops)		0.047			0.001	
		[0.060]			[0.085]	
		(0.053)			(0.072)	
Native Crop Frac.			-0.033			0.042
			[0.169]			[0.186]
			(0.174)			(0.177)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	336	336	336	336	336	336

Notes. The unit of observation is the parish. The table reports OLS estimates. Robust standard errors clustered at the level of the colonial province in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. The dependent variables is non-subsistence agriculture—a dummy variable for whether the share of farmers practicing non-subsistence agriculture is above the median (1994 in Columns 1-3 and 2012 in Columns 4-6). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.

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

TABLE A.19: Pre-Colonial Crop Diversity and Structural Transformation

	Dependent Variable: Share of Population					
	Tertiary Sector		Primary Sector		Secondary Sector	
	(1876)	(2007-2017)	(1876)	(2007-2017)	(1876)	(2007-2017)
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic diversity	-0.023	-0.215	0.067	0.309	-0.045	-0.094
	[0.087]	[0.243]	[0.268]	[0.308]	[0.270]	[0.134]
Ethnic div. × Av. Crop Frac.	0.038	0.390	-0.134	-0.561	0.097	0.171
	[0.187]	[0.457]	[0.502]	[0.568]	[0.509]	[0.251]
Mean Dep. Var.	0.073	0.336	0.687	0.517	0.240	0.147
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of parishes	282	336	282	336	282	336

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. The outcomes refer to the share of the population employed in the tertiary (1-2), primary (3-4), and secondary (5-6) sectors. Ethnic diversity takes value 1 if there is an ethnic border within a buffer of 10-km radius from the parish capital, and 0 otherwise. Crop fractionalization refers to native crops only. Regressions are weighted by the square root of the total population. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to pre-colonial native shrines, log 16th-century expected tribute, and log distance to *mita* mines.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Online Appendix - Surnames

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

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

where p_k represents the proportion of individuals with surname k in the population and K is the total number of different surnames. The first index, $D \in [0, 1]$, is a standard measure of diversity based on the Simpson or Herfindahl index. The second index, $S \in [0, \ln(K)]$, takes its theoretical basis from information theory (Shannon 1948). As long as any two individuals with the same surname inherited the surname from a common ancestor, S can be interpreted as the average uncertainty in predicting ancestry: if each surname has the same relative frequency in the population (surnames are evenly distributed across individuals), the uncertainty in predicting the most likely ancestor of a randomly selected individual will be high; in contrast, a more uneven distribution in which a few surnames are shared by a large portion of the population (e.g., an isolated community characterized by endogamous marriages) implies less uncertainty in predicting ancestry.

I created a dataset of colonial individuals with native paternal surnames using baptism records from digital genealogical sources (FamilySearch.org). Each record reports the full name of the individual, parish, and date of baptism (1605–1780). The dataset provides information for 112,340 individuals.[43] Table B.1 presents OLS estimates from regressing surname diversity measures (either the S index or the D index) on $Ethnic\ div_p$. Since it can be reasonably assumed that not all historical records have been preserved, the results should be interpreted with caution. Panel A shows the baseline results. For each surname

[43] I use information from the collection "Perú, bautismos, 1556-1930," accessed in December 2018. The number of parishes with information varies by year. The mean parish comprises 1,726 individuals with native paternal surnames, of whom 857 are men, relative to a sample mean of 1,627 individuals per parish according to the census of 1791–1795 (of which 769 are men).

diversity index, the first column shows the unconditional correlation; the second column controls for the log number of individuals found in the records of the parish and for the share of individuals with non-native surnames; the third column accounts for potential differences in the mean and standard deviation of elevation, mean and standard deviation of pre-colonial land caloric suitability, longitude, latitude, and log distance to perennial rivers; the last column includes ecclesiastical jurisdiction fixed effects, accounting for potential differences in the administration of baptism across five colonial bishoprics. Panel B shows the estimates obtained after dropping individuals whose surnames occur only once in the dataset, which results in a sample size of 106,124 individuals. In Panel C, I show the estimates obtained from using groups of similar surnames (instead of raw surnames) to compute surname diversity indices.[44] This approach accounts for potential changes in the writing of surnames over time. Panel D restricts the analysis to individuals with non-native paternal surnames.

Identification of native surnames. In order to identify native surnames, I constructed a dictionary of linguistic roots from the Quechuan and Aymaran language families. There is no unique source for the identification of surnames from these families. 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.

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

Table B.1: Validating Ethnic Diversity

	Dependent Variable: Surname Diversity (1605 – 1780)							
	S Index	S Index	S Index	S Index	D Index	D Index	D Index	D Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Baseline (Native Surnames)							
Ethnic diversity	0.512	0.447	0.533	0.558	0.481	0.448	0.477	0.503
	[0.212]**	[0.182]**	[0.198]***	[0.202]***	[0.170]***	[0.182]**	[0.183]**	[0.203]**
	(0.213)**	(0.207)**	(0.220)**	(0.214)***	(0.217)**	(0.201)**	(0.240)**	(0.250)**
	Panel B: Non-Unique Surnames (Native Surnames)							
Ethnic diversity	0.462	0.408	0.476	0.502	0.454	0.428	0.436	0.456
	[0.209]**	[0.159]**	[0.175]***	[0.185]***	[0.173]**	[0.178]**	[0.170]**	[0.188]**
	(0.208)**	(0.180)**	(0.210)**	(0.212)**	(0.218)**	(0.196)**	(0.230)*	(0.246)*
	Panel C: Grouped Surnames (Native Surnames)							
Ethnic diversity	0.482	0.416	0.504	0.525	0.461	0.427	0.454	0.478
	[0.215]**	[0.184]**	[0.199]**	[0.199]**	[0.172]***	[0.181]**	[0.180]**	[0.199]**
	(0.213)**	(0.205)**	(0.214)**	(0.198)***	(0.216)**	(0.199)**	(0.238)*	(0.240)**
	Panel D: Non-Native Surnames							
Ethnic diversity (dummy)	0.284	0.116	0.105	0.079	0.333	0.185	0.199	0.214
	[0.257]	[0.101]	[0.114]	[0.115]	[0.209]	[0.116]	[0.144]	[0.167]
	(0.223)	(0.097)	(0.110)	(0.098)	(0.195)*	(0.126)	(0.133)	(0.134)
Number of parishes	65	65	65	65	65	65	65	65
Ln total individuals (1605–1780)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
% Non-native surnames (1605–1780)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Parish-level controls	No	No	Yes	Yes	No	No	Yes	Yes
Ecclesiastical jurisd. FE	No	No	No	Yes	No	No	No	Yes

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

C Online Appendix - Data Sources and Definitions

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

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

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

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

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

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

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

Ln distance to mita mine. Natural log of the geodesic distance (km) from the parish capital to the closest *mita* mine. Source: author's computation using data from Dell (2010).

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

Administrative province. Categorical variable indicating the colonial administrative province (*partido*). Source: census of 1791-95; see Appendix A.3.

***Ln distance to native shrine, defensive site, canal, bridge*.** Natural log of the geodesic distance (km) from the parish capital to the closest pre-colonial shrine/defensive site/canal/bridge. Source: author's computation using pre-colonial archaeological data; see *Dummy urbanization [ethnic level]*.

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

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

***Non-subsistence agriculture*.** Dummy variable taking value 1 if the share of agricultural producers devoting most of the harvest to sell in local markets is above the median. Source: 1994 and 2012 national agricultural censuses; 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 and 2017 national population and housing censuses; National Institute of Statistics (INEI).

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

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

Urban status. Dummy variable taking value 1 if the share of urban population is above the median. Source: 1993 and 2017 national population and housing censuses; National Institute of Statistics (INEI).

Identification with the state, ethnicity or race, and religion. Dummy variables taking value 1 if the individual reports to identify more strongly with a certain group (separate variables for state administrative units, ethnicity or race, and religion). Source: 2004-2017 ENAHO surveys; National Institute of Statistics (INEI).

Dummy voted in the 2006 presidential election. Dummy variable taking value 1 if the individual reports to have voted in the 2006 presidential election. Source: 2007-2011 ENAHO surveys; National Institute of Statistics (INEI).

Participation in voluntary associations. Dummy variables taking value 1 if the individual reports to participate in a voluntary association (separate variables for neighborhood, professional, and labor associations). Source: 2004-2017 ENAHO surveys; National Institute of Statistics (INEI).

Ln volunteers for military service. Natural log of 1 plus the average number of volunteers for military service between 2008 and 2014. Source: *Ministerio de Defensa*, Perú.

Dummy neighborhood association. Dummy variables taking value 1 for the presence of neighborhood associations during the periods 2002-2004 and 2012-2014. Source: yearly data from the *Registro Nacional de Municipalidades*; National Institute of Statistics (INEI).

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

Access to agricultural land. Mean share of the population with access to agricultural land between 1706 and 1800. Source: author's computation using data from *Tierra y población en el Perú (ss. xviii–xix)* (compiled by Pablo Macera, published: Lima, 1972).

Tertiary-sector occupation. Share of population employed in the tertiary sector. Source: author's computation using data from the 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878), and the 2007 and 2017 population and housing censuses, conducted by the National Institute of Statistics (INEI).

Mean elevation [ethnic level]. Average elevation across all grid cells with centroid within the ethnic homeland. Source: author's computation using Rowe (1946)'s ethnic boundaries; see

Mean elevation.

Mean caloric suitability [ethnic level]. Average pre-1500 land caloric suitability across all grid cells with centroid within the ethnic homeland. Source: author's computation using Rowe (1946)'s ethnic boundaries; see *Mean caloric suitability*.

Ln river density [ethnic level]. Natural log of total river length (km, only perennial rivers) divided by total land area (km^2). Source: author's computation using Rowe (1946)'s ethnic boundaries; see *Ln distance to perennial river*.

Ln land area [ethnic level]. Natural log of total land area (km^2) within the ethnic homeland. Source: author's computation using Rowe (1946)'s ethnic boundaries.

Ln population [ethnic level]. Natural log of approximate population by the time of the Spanish conquest. Source: author's computation using Rowe (1946)'s ethnic boundaries and tributary population data (1532–1575) from Cook (1982). All population centers with centroid within the ethnic homeland are considered.

Ln population density [ethnic level]. Natural log of total population divided by land area. Source: see *Ln population [ethnic level]*.

Dummy urbanization [ethnic level]. Dummy variable taking value 1 for the presence of pre-colonial towns or urban centers. Source: author's computation using Rowe (1946)'s ethnic boundaries and information on pre-colonial archaeological sites in Ravines (1985), Giraldo (2001), Isbell and Silverman (2002, 2008), and the inventory of pre-colonial archaeological sites (*Catastro de Monumentos Arqueológicos Prehispánicos*) developed by SIGDA (Ministerio de Cultura, Perú). The inventory was accessed in March 2021.

Dummy political complexity [ethnic level]. Dummy variable taking value 1 for the presence of pre-colonial administrative centers and monumental architecture—public buildings and communal spaces, including temples, palaces, and complex mound platforms. Source: see *Dummy urbanization [ethnic level]*.

Dummy elite residences [ethnic level]. Dummy variable taking value 1 for the presence of elite residences. Source: see *Dummy urbanization [ethnic level]*.

Dummies for different types of infrastructure [ethnic level]. Separate dummy variables taking value 1 for the presence of terraces, food storage structures, canal, or bridges. Source: see *Dummy urbanization [ethnic level]*.

Ln road density [ethnic level]. Natural log of total road length (km) divided by total land area

(km^2) within the ethnic homeland. Source: author's computation using Rowe (1946)'s ethnic boundaries and the Inca road map (Qhapaq Ñan) produced by SIGDA (Ministerio de Cultura, Perú), accessed in March 2021.

Appendix - References

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

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

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

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

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

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

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

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

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

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

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

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

FIDE. 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] Editor: Cerrón-Palomino, R.

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

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

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

Isbell, William H and Helaine Silverman. 2008. *The Handbook of South American Archaeology*. Springer.

Kling, Jeffrey R, Jeffrey B Liebman, Lawrence F Katz, and Lisa 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.

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

———. 1985. *Surnames and Genetic Structure*. Cambridge University Press.

Lowes, Sara and Eduardo Montero. 2021. "Concessions, Violence, and Indirect Rule: Evidence from the Congo Free State." *The Quarterly Journal of Economics* 136 (4):2047–2091.

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

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

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

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

RENIEC. 2012. *Introducción a un Tesoro de Nombres Quechuas en Apurímac*. Lima, Perú: Registro Nacional de Identificación y Estado Civil, jointly with Terra Nuova and Apurimac ONLUS.

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

Tapia, Mario E. 2013. "Diagnóstico de los Ecosistemas de Montañas en el Perú." Tech. rep., FAO–MINAM (Ministerio del Ambiente, Perú).