Within-Group Heterogeneity in a Multi-Ethnic Society

Artiles, Miriam

Pontificia Universidad Católica de Chile

7 May 2023

Within-Group Heterogeneity in a Multi-Ethnic Society*

Miriam Artiles†

May 7, 2023

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 development. I collect data from a natural experiment of Peru's 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 climates of the Andes before colonization, ethnic diversity results in lower costs and may even become advantageous.

JEL Codes: J15, N16, O10, O12, Z10.

*I am grateful to Gianmarco León-Ciliotta, Luigi Pascali, Marta Reynal-Querol, and Joachim Voth for their guidance and support throughout this project. I am also grateful to Miguel Ángel Carpio, 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, three 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 side, some studies find that if ethnic groups differ in their specializations or skills, then the presence of complementarities can sustain coexistence, facilitate inter-group trade, and generate economic gains (Jha 2013; Becker and Pascali 2019; Jedwab, Johnson, and Koyama 2019; Montalvo and Reynal-Querol 2021). While there is a general understanding that diversity brings opportunities and challenges, there is scarce evidence on which factors determine its positive or negative consequences. When is ethnic diversity good for economic development, and when is it bad?

In this paper, I study whether the long-run effect of ethnic diversity on comparative development depends on 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, ethnicities are not necessarily homogeneous entities. Individuals within ethnic 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; Depetris-Chauvin and Özak 2020). However, whether this dimension contributes to explain

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

the consequences of ethnic diversity remains to be explored.

The main contribution of this paper is to document that exposure to within-group heterogeneity matters to understand the long-run economic consequences of ethnic diversity. I analyze new data from a natural experiment regarding Peru's colonial history—the forced resettlement of native populations in the 16th century. Unintentionally on the part of the Spanish colonizers, this intervention forced together various ethnic groups into small-scale jurisdictions. Where colonial officials concentrated individuals with a history of within-group heterogeneity, who, prior to colonization, settled in complementary climate zones of the Andes to maximize the economic base, ethnic diversity results in systematically lower costs and may even become advantageous.

Several features of the study setting are key to examine whether the consequences of ethnic diversity depend on exposure to within-group heterogeneity. First, ethnohistoric accounts suggest that, at the time of the intervention, geographic proximity to ethnic boundaries created quasi-random variation in the ethnic composition of new jurisdictions.[2] This happened as a result of a mismatch between the pre-colonial settlement pattern and the colonial notion of jurisdiction (Wachtel 1976; Pease 1989). Second, the Spanish intervention resulted in a setting where exposure to within-group heterogeneity was arguably orthogonal to ethnic diversity. The key element of pre-colonial society that the intervention altered was the settlement pattern. Before the Spanish conquest, coethnic individuals occupied different altitudes in an attempt to maximize the economic base. The anthropologist John Murra wrote:

> "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)[3]

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

[3]Different disciplines have documented the subsistence strategy of Andean ethnic groups. See Brush (1976),

Differences in elevation created complementary microclimates within short distances, which incentivized a vertical settlement pattern and internal exchange. The Andean subsistence strategy is consistent with Michalopoulos (2012)'s idea that variation in geographic characteristics, such as land quality and elevation, may lead to specialization through the formation of zone-specific skills, applying the idea *within* ethnic groups. The internal economic organization of ethnic groups, however, has received little attention in economics.[4]

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

The first result of the paper documents the direct effect of ethnic diversity, which I benchmark against previous results in the literature. Guided by the historical narrative, I identify parishes (colonial jurisdictions) that were accidentally created close to spatial boundaries between ethnic groups. The analysis relies on the assumption that Spanish officials were not fully aware of the vertical distribution of coethnic individuals over space. Given the vertical settlement pattern, parishes created close to ethnic boundaries accidentally concentrated populations from different ethnic origins (Pease 1989; Wachtel 1976).[5] I use a subsample of parishes for

Pease (1989), Stanish (1989), and Aldenderfer (1993), among others, for perspectives from human ecology, history, anthropology, and archaeology, respectively.

[4]See Moscona, Nunn, and Robinson (2020) for evidence on segmentary lineage organization.

[5]I provide empirical support for the historical narrative that colonial officials did not systematically consider ethnic boundaries for the location of parishes. On average, they did not select locations differently (i.e., in a way that resulted in systematic differences in proximity to ethnic boundaries) depending on the characteristics of native populations or geography.

which surnames from colonial baptism records are available to validate ethnic diversity. When comparing contemporary living standards between parishes whose initial populations were ethnically diverse and those with an ethnically homogeneous founding population, the results show a robust pattern. On average, parishes with ethnic diversity tend to exhibit lower living standards in the long run. I use data on local economic activity and access to public facilities from a variety of sources to capture living standards, including satellite imagery and census data. The result holds beyond geographic proximity to ethnic boundaries, highlighting the persistent consequences of forced diversity at the local level.

I then explore whether the average effect of ethnic diversity on living standards differs depending on pre-colonial exposure to within-group heterogeneity. I consider average exposure among the ethnic groups concentrated in each parish. To measure within-group heterogeneity during pre-colonial times, I construct spatial data on the distribution of vertical resource-producing zones within the homelands of ethnic groups. Using data from paleodietary reconstructions available for a subset of the groups, I document that, beyond crop variety, the measure of within-group heterogeneity is associated with the consumption of carbon-enriched crops likely coming from exchange between coethnics settled in complementary elevation zones.

When I examine the interaction effect, the estimates show a negative coefficient on ethnic diversity and a positive coefficient on its interaction with average exposure to within-group heterogeneity.[6] Ethnically diverse parishes whose initial populations were exposed to higher within-group heterogeneity tend to perform better in the long run. On average, in parishes in the bottom 10 percent of past exposure, ethnic diversity is associated with a decrease in living standards of 0.27 standard deviations. In the top 10 percent parishes, ethnic diversity is associated with a 0.21 standard deviation increase in living standards. Contemporary survey data on household consumption and medium-term outcomes show the same pattern.

The estimated interaction effect persists when controlling for initial prosperity and geography at the parish level. It is also robust to considering the administrative province and ecclesiastical

[6]Before presenting the main result, I provide evidence that more heterogeneous ethnic groups did not manipulate the locations of parishes at the time of the policy—parishes without ethnic diversity did not systematically concentrate populations from more heterogeneous ethnic groups.

jurisdiction in charge of the parish. I collect new data from archaeological sources to explore the correlates of within-group heterogeneity before colonization. Controlling for a broad set of potential correlates interacted with ethnic diversity alleviates concerns that other pre-colonial characteristics of ethnic groups, such as their level of socio-economic and institutional development, could be confounding the interaction effect. A matching procedure used to construct a sample of parishes with different levels of exposure to within-group heterogeneity whose populations were statistically similar in all other pre-colonial characteristics shows consistent estimates. The result is also robust to considering alternative scenarios on the spatial distribution of pre-colonial populations.[7]

Additional results support cultural transmission as the main mechanism. In line with qualitative evidence from early chronicles and native folklore, the results are consistent with the idea that pre-colonial internal exchange aimed at reaching a common goal (i.e., maximizing the economic base) contributed to the formation of cooperative behavior and more open attitudes toward out-group members, resulting in a more integrated society where populations engage relatively more in associational activities and in local trade. Using colonial baptism records to detect parents who were potentially from different ethnic groups offers a consistent interpretation. 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). Furthermore, where the historical ethnic minority likely had comparative advantage over the majority group, contemporary populations perform relatively better, suggesting that economic complementarities may also have contributed to sustain inter-ethnic coexistence (Jha 2013, 2018).

This paper contributes to a large body of studies in the development and political economy literatures on the consequences of ethnic diversity. The initial literature tended to emphasize the costs of ethnic diversity without considering the potential heterogeneity of individuals within ethnic groups. The empirical research has been conducted at different levels of analysis. Across countries and US localities, ethnic diversity has been associated with lower levels of economic growth, public good provision, and quality of government, as well as with greater

[7]Section 5.2 and Appendix B present supplementary analyses and robustness checks.

political instability and civil conflict.[8] Using micro-level data, Miguel and Gugerty (2005) show that ethnic diversity is associated with lower public good provision in Kenya. More recently, Hjort (2014) has focused on the private sector. The author provides causal evidence for the effect of ethnic diversity on team productivity at a flower plant in Kenya. The results show that teams of ethnically diverse workers are, on average, less productive than homogeneous teams. Evidence on potential mechanisms points toward a taste for discrimination against coworkers of different ethnic origin.

More recent papers have emphasized the role of local interactions in understanding the effects of ethnic diversity. Desmet, Gomes, and Ortuño-Ortín (2020) provide cross-country evidence that local inter-ethnic interactions can contribute to weaken the costs of ethnic diversity for public good provision. Montalvo and Reynal-Querol (2021) focus on the size of the unit of analysis, finding that ethnic diversity has a positive effect on economic growth at low levels of geographic aggregation. The authors argue that a potential explanation in the context of Africa is the increase in trade close to spatial boundaries between ethnic groups, suggesting ethnic specialization into complementary activities. These papers relate to recent studies on the positive role of inter-ethnic complementarities at the local level (Jha 2013, 2018; Becker and Pascali 2019; Jedwab, Johnson, and Koyama 2019).

Following Gennaioli and Rainer (2007), Michalopoulos and Papaioannou (2013), and Alesina, Giuliano, and Nunn (2013), the characteristics of pre-colonial societies and ethnic groups have received increasing attention. However, the potential heterogeneity of individuals within these groups has been less explored. Ashraf and Galor (2013) presented the first empirical study to explore the role of this dimension in explaining comparative development. The authors focus on population heterogeneity, as proxied by the degree of genetic diversity across coethnics, and find a hump-shaped effect on economic development in both pre-modern and modern times.[9] Dippel (2014) has explored a different source of heterogeneity—the

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

[9]Underlying the positive side of this effect is the idea that the greater the degree of population heterogeneity, 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. See Depetris-Chauvin and Özak

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 into a centralized institutional system can negatively impact long-run development. He provides empirical evidence from the formation of Native American reservations in the United States, showing that internal political conflict was the main underlying mechanism.[10]

In this paper, I focus on a genetically homogeneous region—highland Peru—where pre-colonial ethnic groups sought to maximize the economic base through centralized crop exchange between coethnics settled in complementary elevation zones.[11] The 16th-century intervention unintentionally altered the spatial distribution of coethnics, allowing us to explore the consequences of ethnic diversity in a setting with variation in past exposure to within-group heterogeneity. I examine the natural experiment and provide robust evidence that exposure to within-group heterogeneity in complementary traits contributed to overcome the drawbacks of ethnic diversity in the long run. The results are consistent with the idea that past exposure to within-group heterogeneity facilitated inter-ethnic interactions after the intervention, contributing to a more integrated society and sustaining long-run development.

The evidence on cultural transmission adds to the literature on the long-run effects of cultural traits (e.g., Nunn and Wantchekon 2011; Voigtländer and Voth 2012; Alesina, Giuliano, and Nunn 2013; Guiso, Sapienza, and Zingales 2016). This paper also contributes to a recent literature on the long-run consequences of forced displacements (Becker 2022). The results help us understand the effects of the displacement of indigenous populations as a result of colonization, a research topic with scarce evidence so far; see Valencia Caicedo (2019). In contemporary societies where multiple ethnicities coexist (e.g., as a consequence of forced displacements or due to voluntary migrations in an increasingly globalized world), understanding whether the consequences of ethnic diversity depend on exposure to within-group

(2020) for empirical evidence on the link between population heterogeneity and specializations across pre-modern societies.

[10] See also Esteban and Ray (2011) for a theoretical model of ethnic conflict in which coethnic individuals differ in attitudes and income.

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

heterogeneity is also important for policy discussions.

The remainder of the paper is organized as follows: Section 2 summarizes the historical context, Section 3 describes the data, Section 4 presents the empirical strategy, Section 5 explains the main result, Section 6 discusses potential mechanisms, and Section 7 concludes.

2 Historical Background

2.1 Pre-Colonial Context

Ethnic groups. By the time Spanish conquerors arrived, the Andean civilization comprised several coexisting groups that had been incorporated over the previous century to the Inca empire (1438-1525), e.g. *Chocorvos*, *Lucanas*, *Soras*, *Chankas*, *Quichuas*, *Cavinas*, *Huancas*, and *Aymaraes*, among others (Tello 1939; Rowe 1946; Dulanto 2008). In particular, the 47 groups in my study region coexisted after the disintegration of the *Wari* culture (ca. 1000) and before being conquered by the Spanish (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 most common view is that the Andean society comprised several groups with diverse linguistic roots (e.g., Rowe 1946; Murra 1975) and differentiated material cultures in both domestic and non-domestic contexts (e.g., Stanish 1989). For example, many groups, such as the *Atavillos* and the *Chocorvos*, had their own language; for some groups, such as the *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). However, the issue of ethnic identity has been vaguely discussed in the historical literature, possibly because the Peruvian Andes was long perceived by Europeans as a culturally homogeneous region despite ethnolinguistic differences being present at the time of Spanish contact.[12]

[12] Charney (1998) argues that the use of the Spanish term "Indio" to refer to all native peoples in official documents contributed to masking ethnic distinctions in the eyes of Europeans–"Indio" was not just 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.

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). In turn, the group usually claimed descent from a mythical ancestor, such as an animal or element of nature. This mythical kin was worshipped and sometimes honoured with rites and sacrifices. For example, the *Chankas* believed that they were closely connected to the Andean lion (*puma*). During festivities, they would usually dress in *puma* skins and coalesce around *puma* imagery; see Garcilaso de la Vega (1960)[1609] and Bauer, Kellett, and Silva (2010). Many individuals continue to support this identity in public gatherings today (see Appendix A.1).

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

After the pioneering ethnohistoric work of Murra (1975), studies across different disciplines applied his ideas to understand subsistence patterns in pre-colonial times (e.g., Brush 1976; Pease 1989; Stanish 1989; Aldenderfer 1993; Nash 2009). In the study region, the settlement pattern of a given ethnic group is described as a vertical archipelago. Specifically, archaeological and ethnohistoric research documents that subsistence was based on the simultaneous control of different elevation zones. Murra's model is often described as a zonal complementarity model (e.g., Stanish 1989; Aldenderfer 1993; Isbell and Silverman 2002b). The group tried to maximize the economic base by establishing permanent settlements in vertically arranged resource zones (Murra 1975, 1995, 2002a,b). Since certain crops can only be grown at specific altitudes, the zones can be 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. Rather than being organized by independent subpopulations, crop exchange seems to have been centralized at the group level (Murra 2002b). Furthermore, ties to the extended family and the rest of the group seem to have been retained to benefit from complementarities (Pease 1989; Stanish 1989, 2005; Murra 2002a; Nash 2009).[13]

Continuity after the Inca expansion. According to Murra's and subsequent research, this settlement pattern was already in place 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 suggests that the Inca government was indirect in the sense that each region was governed by the ruler of the corresponding ethnic group (Murra 1975, 2002b). This is a crucial characteristic of Inca rule because it supports the notion that ethnic traits were preserved during this period. One example is the festivity of the *Chankas* in honour of their mythical connection to the *puma*, which, according to early chronicles, was also celebrated during the Inca period (Garcilaso de la Vega (1960)[1609]). At the same time, ethnic rulers were pushed to continue with the control of vertical zones in their respective regions to sustain the empire (Murra 1956, 1975). The archaeologist and anthropologist John H. Rowe (1946) mapped the approximate extent of the groups at the time of the Spanish conquest (see Figure 1). The map, based on archaeological evidence and early ethnohistoric accounts, was published in the second volume of the *Smithsonian Handbook of South American Indians*.

[Figure 1 about here]

2.2 The Spanish Intervention

The contemporary administrative division of Peru has its origin in the initial colonial period. When Viceroy *Francisco de Toledo* first disembarked in Peru (1569), native populations followed the Andean pattern, living scattered along mountain slopes. This settlement pattern was seen as an "obstacle" for the Spanish administration. In the words of Spanish official *Juan*

[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.

de Matienzo, “the *indios*, for being isolated in *huaycos* and ravines, do not live in right order, and this is the main obstacle to be indoctrinated” (in Medina 1974a, p. 155).

To facilitate tribute collection and religious indoctrination, *Toledo* ordered the forced re-organization of native populations into residential (*reducciones*) and religious (*doctrinas*) jurisdictions. Between 1570 and 1575, colonial officials arranged the division of populations from all discovered lands in the Viceroyalty of Peru into *reducciones*. In turn, several *reducciones* were under the jurisdiction of a single *doctrina* or *parroquia*, a parish served either by the regular or secular clergy.[14] Section 4.1.1 describes colonial recommendations on desirable locations.

Within four decades of the conquest of the Inca empire, the Spanish administration had completed a massive re-organization of native populations. The intention of the resettlement was not to create sustainable jurisdictions but to concentrate dispersed populations in a way more consistent with the Spanish conception of the world (i.e., small-scale, continuous, and delimited jurisdictions; Medina 1974a,b, 1993; Bauer, Kellett, and Silva 2010). Notably, there was tension at the time of the policy between the pre-existing settlement pattern, which was a native response to the mountain environment, and the Spanish notion of jurisdiction, based on the idea of a more horizontal world. Ethnohistoric accounts suggest that colonial officials did not consider the vertical distribution of coethnics over space (Murra 1975). Given the vertical settlement pattern, the new jurisdictions did not always respect pre-existing ethnic divisions (Pease 1989, 1992; Wachtel 1976, 2002).

The new model also limited population movement, pointing against the exchange of resources between different elevation zones and, thus, creating a new paradigm for native populations (Pease 1989).[15] Historical studies note that, in practice, the limitation of movement was effective at the parish level (Saignes 1991; Medina 1974a,b, 1993). In fact, this system was maintained throughout the entire colonial period, and, at the time of independence from

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

[15]Appendix A provides anecdotal evidence and examples of other Spanish practices that were at odds with the native knowledge of the environment.

Spain, parishes were called districts, forming the basis for what is currently the third-level administrative division of the country.[16]

3 Data Construction

3.1 Sample

I match contemporary districts to colonial parishes, which I refer to throughout the text as "parishes." This study focuses on the Peruvian territory conquered by the Inca empire that remained in the Viceroyalty of Peru for the entire colonial period (1532-1810). The census prepared from 1791 to 1795 under the administration of Viceroy *Gil de Taboada y Lemos* lists all parishes created in this territory (see Appendix A).

I start by matching districts to colonial parishes using each district's name and year of creation. Districts created as a result of the 16th-century intervention represent about 24 percent of current districts.[17] I then assign coordinates to each parish capital using a map from the Peruvian *Ministerio del Ambiente* (MINAM) that provides the name and coordinates of all existing population centers within each district. In most cases, the old parish capital remains the district capital. For districts where this is not the case (i.e., where the capital was changed after independence from Spanish rule), I assign the coordinates of the colonial capital.[18]

Following the historical literature on Murra's model, I focus on parishes located in highland Peru (i.e., more than 500 meters above sea level). I exclude the two capital parishes of Cuzco and Arequipa,[19] as well as six parishes that now form part of Chile. The final sample consists of 336 parishes; see Panel (a) of Figure 1.

---

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

[17]Appendix A reports modern aerial views of Yanque, created as a result of the intervention, as an example.

[18]I check for priests in charge of religious indoctrination during the colonial period using historical sources; see Lissón Chávez (1943), de Armas Medina (1953), de Córdoba Salinas (1957)[1651], and García (1997).

[19]Ancient DNA data show that Cuzco is also an exception to the trend of genetic homogeneity during the Inca period, which evidences population mobility in the administrative center of the empire (Nakatsuka et al. 2020).

3.2 Measuring Ethnic Diversity

3.2.1 A Measure of Ethnic Diversity

A measure of ethnic diversity would ideally be based on administrative data detailing the ethnicity of relocated individuals. However, such information was not systematically registered.[20] To detect parishes where colonial officials concentrated ethnically diverse populations, I follow the historical narrative that, given the vertical settlement pattern, parishes created close to ethnic boundaries accidentally concentrated populations from different ethnic origins (Pease 1989; Wachtel 1976).

Colonial accounts of the territories visited by Spanish officials describe distances that commonly lie between 2 and 3 *leguas*, the colonial measure of distance (e.g., Jiménez de la Espada 1881). I start by defining $Ethnic\ div_p$ as a dummy variable indicating the presence of an ethnic boundary within a buffer of 10-km radius (3 *leguas*) from the capital of parish p. For this exercise, I rely on Rowe (1946)'s mapping of the approximate extent of pre-colonial ethnic groups.[21] Panel (a) of Figure 1 shows the spatial distribution of the variable: parishes with an ethnic boundary within the buffer of 10-km radius are displayed in yellow (35 percent of the sample), while those located further inside ethnic homelands are displayed in blue. In robustness checks, I consider a range of different radii.

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

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

[21] See Paz Soldán (1877) for the correspondence between *leguas* and kilometers during the 16th century. Figure B.1 illustrates the buffer exercise. When the distance between the capitals of two parishes is less than 10 km, I use equidistant boundaries to ensure that the buffers do not overlap. The resulting buffers have mean and median areas of 240.44 km^2 and 256.51 km^2, respectively. Furthermore, an ethnic group is counted as part of the buffer only if its homeland occupies at least 1 percent of the buffer's area. This ensures that the ethnic group has at least one grid cell of 1 km $\times$ 1 km inside the buffer.

3.2.2 Validating Ethnic Diversity

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

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

where p_k represents the proportion of individuals with surname k in the population and K is the total number of different surnames. The first index, $D \in [0, 1]$, is a standard measure 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).[22] As long as any two individuals with the same surname inherited the surname from a common ancestor, S can be interpreted as the average uncertainty in predicting ancestry: if each surname has the same relative frequency in the population (surnames are evenly distributed across individuals), the uncertainty in predicting the most probable ancestor of a randomly selected individual will be high; in contrast, a more uneven distribution in which a few surnames are shared by a large portion of the population (e.g., an isolated community characterized by endogamous marriages) implies less uncertainty in predicting ancestry.

Isonymy methods make a strong assumption (i.e., that surname commonality directly translates into common ancestry).[23] Are these methods appropriate for this application? Two contextual features are worth noting. The first is pre-colonial endogamy with ancestry traced through the male line (Rowe 1946); the second is that no system of family names existed prior to the Spanish conquest, but first names related to mythical ancestors did. The Catholic Church

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

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

introduced the Hispanic system of family names for the purpose of religious indoctrination. While the potential adoption of Spanish surnames over time represents a limitation, qualitative evidence suggests that the common practice during the early colonial period was for priests to choose a Spanish first name, with the mythical first names of the individual's parents adopted as surnames (see Appendix C and Carpio and Guerrero (2021) for further details).

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

Empirical results. Since it can be reasonably assumed that not all historical records have been preserved, the results should be interpreted with caution.[24] Table 1 presents OLS estimates from regressing surname diversity measures on $Ethnic\ div_p$. In each column, the dependent variable is either the S index or the D index, constructed using individuals with native paternal surnames.

Panel A shows the baseline results. For each surname diversity index, the first column shows the unconditional correlation; the second column controls for the log number of individuals found in the records of the parish and 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 land caloric suitability during the pre-colonial period (Galor and Özak 2016), longitude, latitude, and log distance to perennial rivers; the last column includes ecclesiastical jurisdiction fixed effects, accounting for potential differences in the administration of baptism across five colonial bishoprics (Lima, Arequipa, Huamanga,

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

Trujillo, and Cuzco). Panel B shows the results obtained after dropping individuals whose surnames occur only once in the dataset.[25] In Panel C, I show the results obtained from using groups of similar surnames (instead of raw surnames) to compute surname diversity indices. This approach takes into account potential changes in the writing of surnames over time.[26] The results for the subsample of parishes with information suggest that, on average, parishes located close to ethnic boundaries exhibit higher levels of surname diversity among individuals with native surnames (between 0.41 and 0.56 standard deviations) than do parishes located in the interior of ethnic homelands.

[Table 1 about here]

3.3 Measuring Within-Group Heterogeneity

3.3.1 A Measure of Within-Group Heterogeneity

The subsistence strategy of the groups was based on crop exchange between coethnic individuals settled in vertically distributed resource zones. However, comprehensive data on the number of individuals settled in each zone prior to the Spanish conquest do not exist. In this paper, I construct spatial data on the distribution of resource-producing zones to compute a proxy for within-group heterogeneity.

For this task, I rely on the research of Pulgar Vidal (1941), a well-known Peruvian geographer. His work integrates local geography and native folklore, providing a well-established and comprehensive account of the mountain environment in Peru. In particular, he has identified five distinct natural resource zones in my 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.), where figures in parentheses refer to elevation in meters above sea level. Notably, each zone has traditionally been known for specific crops. For example, the natural limit of maize cultivation

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

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

is the Quechua zone; grains like quinua and kañiwa, as well as lupins like tarwi, are best grown in the Jalca zone; and various varieties of potatoes, which can provide more carbohydrates per hectare than maize at high altitudes, grow exceptionally well in the Puna (Pulgar Vidal 1941; Burger and Merwe 1990; Sandweiss and Richardson 2008).

I map the spatial distribution of the zones using elevation data from version 1.2 of the Harmonized World Soil Database (FAO). Specifically, I assign each grid cell in the FAO data (approximately 1 km × 1 km at the equator) to a particular zone based on median elevation. The resulting map is shown in Panel (b) of Figure 1. Spanish officials established parishes at different elevations (Table B.1). Furthermore, approximately 23.53, 34.57, 44.23, and 28.57 percent of parishes established in the Yunga, Quechua, Suni, and Puna zones, respectively, have an ethnic border within the buffer of 10-km radius.[27]

Using this map, I compute a measure of within-group heterogeneity. I start by considering the reciprocal of the Simpson or Herfindahl index, a common measure of diversity in ecological studies (Magurran 2004):[28]

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

where s_{ej} is the area share of zone j within the homeland of ethnic group e. The index increases as the composition of resource zones becomes more diverse. I normalize the index to 1 for the group with the highest value to facilitate the interpretation. Figure B.2 shows the density of the normalized 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).

[27] I exclude the Janca zone from the analysis (3.43 percent of the total territory in the study region) because it cannot be permanently inhabited due to oxygen constraints (e.g., Sandweiss and Richardson 2008). Pasture for camelids was the primary resource of this zone.

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

3.3.2 Validating Resource-Producing Zones

This section explores the extent to which resource zones are meaningful. Specifically, I explore whether the measure of 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.[29] 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.[30]

I start by computing H at the grid-cell level for 25 km $\times$ 25 km grid cells covering the entire study region. In Column 1 of Table 2, I regress the log of the number of crops on this measure. The estimated beta coefficient is positive (0.523) and statistically significant. The coefficient remains stable in magnitude and statistical significance when I control for an alternative potential predictor of crop variety—variation in elevation (Column 3).[31] The same pattern arises when including fixed effects that account for differences across hydrographic basins (Column 4),[32] log area (Column 5), and mean elevation (Column 6). Column 7 shows the same pattern when using the number of crops without being logarithmically transformed as an outcome variable.

According to unconditional OLS estimates, a one standard deviation increase in H is associated with a 0.523 standard deviation increase in log crop variety (Column 1). In the case of variation in elevation, the associated standard deviation increase in log crop variety is 0.398 (Column 2). However, this coefficient becomes small (0.007) and statistically insignificant after I control for H (Column 3). In Column 8, I substitute the grid-level H index with dummy variables indicating the number of resource zones within the grid cell. The magnitudes of the

[29] I follow the classification of native crops in Tapia (2013), who identifies 41 main native crops in the region. The 2012 agricultural census covers 38 of these crops.

[30] One limitation of using these census data is that farmers reported the list of crops harvested at the time of the census. Practices of crop rotation and fallow would thus affect the list of reported crops.

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

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

estimated coefficients increase with the number of zones (relative to those for grid cells with only one resource zone). Similar results arise across 50 km × 50 km grid cells (Table B.2) and ethnic groups (Table B.3).

[Table 2 about here]

3.3.3 Correlates of Within-Group Heterogeneity

In Table 3, I explore the pre-colonial correlates of within-group heterogeneity at the ethnic group level (H_e). Columns 1-3 show that mean elevation, land caloric suitability, and river density are not significantly correlated with within-group heterogeneity. Under autarky, group size coincides with market size, which could create incentives for specialization and innovation (Smith 1776). However, I find no statistically significant correlation between within-group heterogeneity and size of the ethnic group, as measured by land area (Column 4) and approximate population (Column 5) before colonization.[33] I do find that within-group heterogeneity is positively correlated with approximate population density (Column 6), which could reflect economic prosperity during the pre-colonial period (Ashraf and Galor 2011, 2013; Maloney and Valencia Caicedo 2016).

Columns 7-9 explore pre-colonial socio-economic and institutional characteristics. In the absence of systematized ethnographic data, I collect information from archaeological sources.[34] In line with Column 6, the data suggest that within-group heterogeneity is positively correlated with urbanization, as measured by a dummy for the presence of towns and urban centers within the ethnic group's homeland (Column 7). Column 8 shows evidence consistent with the idea that incentives for internal exchange may lead to political centralization (Fenske 2014). In particular, I create a dummy for any material indicator that could evince political complexity (i.e., administrative centers and monumental architecture—public buildings and communal spaces, including temples, palaces, and complex mound platforms, as defined in Stanish 2001) and find a positive correlation with within-group heterogeneity. Column 9 shows

[33]Population is available for 46 (out of 47) groups; see Appendix D for data sources.

[34]See Appendix D. Site-level archaeological data are receiving increasing attention in economics (Matranga and Pascali 2021). Importantly, although this type of data is subject to geographic coverage, the Peruvian Andes have a long tradition of archaeological research (Isbell and Silverman 2002a, 2008).

no correlation with the presence of elite residences, nonetheless. The results highlight the importance of considering these correlates in robustness exercises.[35]

[Table 3 about here]

3.3.4 Within-Group Heterogeneity and Pre-Colonial Diets

Table 2 shows that the classification of resource zones matters to capture crop variety. This result is a direct consequence of the mountain environment. In this section, I provide correlational evidence on how within-group heterogeneity relates to past human diets and crop exchange using data from paleodietary reconstructions. Biochemical analyses of archaeological human remains can inform on the role that carbon-enriched crops played in the diets of individuals. 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. Most plants use the C_3 pathway, including tubers. However, maize, sorghum, and millets are well-known C_4 plants.

I use the database of pre-colonial diets compiled by Wilson et al. (2022), the most extensive effort at compiling comparable stable isotope values in Peru, to detect carbon-enriched diets in regions where such diets were unlikely in the absence of crop exchange. For example, among the archaeological individuals found in the Jalca zone of the territory of the *Soras*, a highly heterogeneous group according to the classification of resource zones ($H_e = 0.77$), stable isotope values from bone collagen are typical of more carbon-enriched diets (individual $\delta^{13}C_{col}(‰)$ values range between -14.7 and -10.8, with an average value of -12.47).[36] Even

[35]In Table B.4, I explore the correlation of within-group heterogeneity with different types of pre-colonial infrastructures. Unsurprisingly, within-group heterogeneity is positively correlated with the presence of terraces. However, I find no evidence of correlation with Inca roads, nor evidence of correlation with canals, which could have facilitated the water flow, bridges, or food storage structures. The latter may be due to the fact that some crops, such as potatoes, were sometimes simply spread on the ground or placed underground to freeze, leaving no archaeological record.

[36]The values of $\delta^{13}C_{col}(‰)$ in C_3 plants tend to range between -20 and -35, with an average of -26.5 (Benson and Calvin 1948).

when these individuals are not representative of the *Soras* population, it is surprising that they were found in a zone where maize, the main C_4 staple crop in the study region, was difficult to grow given geoclimatic conditions.[37] Tung and Knudson (2018) find similar evidence in southern Peru—the authors document carbon-enriched diets likely coming from maize consumption in a non-suitable location, suggesting crop exchange between resource zones.

Wilson et al. (2022)'s database contains information on $\delta^{13}C_{col}$ from bone collagen for 196 individuals in the region and period of interest (i.e., highland Peru after the disintegration of the *Wari* culture and before the Spanish conquest). 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 individual's remains were found. The individuals are distributed across eight ethnic groups from the north, center, and south of the country (23 percent of the land area in the study region; see Figure B.3). All the groups had access to various resource zones according to Rowe (1946)'s map, with an H_e value ranging from 0.30 to 0.88. However, the data do not include individuals from different zones within the same ethnic group (Table B.5)—five groups are observed in the Quechua zone, the upper limit of maize cultivation, whereas the remaining three groups are observed in the Jalca zone, where maize was generally more difficult to grow (Pulgar Vidal 1941).

In Table 4, I regress individual-level $\delta^{13}C_{col}$ values on within-group heterogeneity (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 its protein component. To better understand this correlation, Column 2 includes zone fixed effects, thus comparing individuals settled in the same zone but from ethnic groups with different degrees of within-group heterogeneity. The correlation between the H index and $\delta^{13}C_{col}$ scores is positive and statistically significant, even after controlling for differences in crop variety across ethnic groups (Column 3). Similar results arise when controlling for the average caloric suitability of the ethnic homeland (Column 4) and in a more saturated specification that includes pre-colonial institutional and socio-economic controls at the ethnic group level (Column 5). Panel B shows

[37] 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).

that the results are similar when excluding children from the sample (14 individuals). The same pattern appears when, instead of using the continuous carbon isotope values as outcome variable, I use a dummy for whether $\delta^{13}C_{col}$ values reveal C_4 consumption (Table B.6).

[Table 4 about here]

Although the available data do not allow direct testing of internal exchange, Column 5 provides evidence that the ethnic boundary matters in explaining the consumption of carbon-enriched crops. I replicate the analysis using grid cells instead of ethnic groups.[38] 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 (Section 3.3.2), it is only associated with carbon-enriched diets at the ethnic group level. This evidence is consistent with the narrative that crop exchange was centralized at the group level (Murra 2002b).

4 Empirical Strategy

4.1 Average Effect of Ethnic Diversity

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

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

where y_p is a contemporary development outcome for parish p, $Ethnic\ div_p$ is a dummy variable indicating whether the parish was created close to an ethnic boundary, X_p is a vector of parish-level control variables measured at baseline, and υ_p is an error term.[39] The historical

[38] I define the grid-cell size (50 km × 50 km) so that the number of cells matches the number of ethnic groups included in columns 1-5 of Table 4.

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

narrative suggests that geographic proximity to ethnic boundaries created quasi-random variation in ethnic diversity across parishes. The analysis relies on the assumption that colonial officials were not fully aware of the vertical distribution of coethnic individuals over space (Murra 1975). Thus, they were unlikely to have systematically considered ethnic boundaries when deciding the location of new parishes. Given the vertical settlement pattern, parishes created close to ethnic boundaries accidentally concentrated populations from different ethnic origins (Pease 1989; Wachtel 1976).

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

4.1.1 Balance Tests for Ethnic Diversity

The colonial regulation of 1569–1570 described three desirable characteristics for location (Jiménez de la Espada 1881). The first was land quality and abundance. Enough land was needed to be worked by native families following their own rules of crop rotation. Land plots were thought to be the primary means for paying tribute.[40] The second characteristic was access to water. Proximity to surface water, which in this context meant access to the system of Andean river basins, was a key advantage for the irrigation of land and the possibility of sustaining populations that mainly depended on subsistence agriculture. Finally, to facilitate religious indoctrination, the locations would ideally be far from *huacas*, sacred native shrines that generally honored nature. Local officials were also tasked with destroying the houses where native families used to live before the resettlement. Shortly after the creation of new jurisdictions, families refusing to relocate were to be punished and forced to move.

The extent to which Spanish officials applied the recommendations is unclear (Pease 1989). Nonetheless, the results in Table 5 show that, on average, parishes built on ethnically

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

homogeneous populations and those whose initial populations were ethnically diverse are statistically similar in the highlighted factors. I start by exploring the mean and standard deviation of terrain characteristics, such as elevation and pre-colonial land caloric suitability. There are no statistically significant differences in these characteristics, which alleviates concerns regarding the possibility that colonial officials selected locations differently (i.e., in a way that resulted in systematic differences in proximity to ethnic boundaries) at different elevations or in plains as opposed to more rugged terrain. Log distance to perennial rivers is also balanced. Finally, I collected data on the location of pre-colonial shrines to explore the third recommendation. On average, ethnically diverse and non-diverse parishes do not differ significantly in log distance to native shrines. 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.[41]

[Table 5 about here]

Table B.7 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 geo-referenced 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.

[41] The legal requirement to send native populations to mines subjected to the *mita* started in 1573. The expected tribute was based on colonial officials' assessment of the number of individuals present at the time of the policy (Cook 1982; Puente Brunke 1991).

4.2 Heterogeneous Effects of Ethnic Diversity

I am interested in exploring whether the average effect of ethnic diversity on economic development 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 the ethnic-level measure of pre-colonial within-group heterogeneity.

A causal interpretation of β_3 requires two conditions. First, to introduce the interaction term in equation 2, ethnic diversity should not be determined by $\overline{H}_p$. Considering the quasi-random nature of the variation in ethnic diversity, $Ethnic\ div_p$ should be orthogonal to $\overline{H}_p$. However, one may be concerned about the possibility that, for example, more heterogeneous ethnic groups negotiated locations in the interior of ethnic homelands, in which case parishes with an ethnically homogeneous founding population would systematically concentrate populations from more heterogeneous ethnic groups. Before presenting the main result, I document that $Ethnic\ div_p$ is uncorrelated with $\overline{H}_p$ in the data. I also provide supporting evidence that more heterogeneous ethnic groups did not manipulate locations during the resettlement.

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. In robustness checks, I control for the pre-colonial correlates of within-group heterogeneity and their interactions with ethnic diversity. I also conduct additional empirical exercises to alleviate the concern that omitted variables related to pre-colonial characteristics of ethnic groups could be confounding the interaction effect.

4.2.1 Within-Group Heterogeneity and Colonial Locations

In Table B.8, I examine whether there is evidence that more heterogeneous ethnic groups tried to manipulate colonial locations, thus inducing changes in the ethnic composition of parishes. I start by aggregating the number of parishes at the ethnic group level. Columns 1 and 2 show that within-group heterogeneity (H_e) is not significantly correlated with the total number of parishes in which the ethnic group was concentrated, nor is it correlated with the proportion of parishes located close to spatial boundaries between ethnic groups, on average. Column 3 shows that the average distance from a parish to the closest ethnic boundary is also uncorrelated with within-group heterogeneity. This supports the idea that ethnic groups with more heterogeneous subpopulations did not have the chance to influence average distance to ethnic boundaries at the time of the policy. As a result, parishes built on ethnically homogeneous populations should not systematically concentrate populations from more heterogeneous ethnic groups. Figure 2 shows that the distribution of average exposure to within-group heterogeneity ($\overline{H}_p$) is indeed similar among ethnically homogeneous parishes (left boxplot) and those that concentrated populations from various ethnic groups (right boxplot). Consistently, Table B.9 documents that ethnic diversity is not significantly correlated with the average level of exposure to within-group heterogeneity across parishes.

5 Results

5.1 Main Results

I start by exploring contemporary living standards across parishes. In particular, I use different measures of local economic activity and access to public facilities previously employed in the literature. First, I follow the empirical literature in using luminosity data from satellite images at night to proxy for local economic activity (Michalopoulos and Papaioannou 2013, 2018). The second proxy I consider is non-subsistence agriculture (Dell 2010). Subsistence farming has traditionally been a widespread practice in the Andean highlands (Mayer 2002). Specifically, the Peruvian agricultural census asks farmers whether they dedicate most of their harvest to self-consumption or, instead, to sale or trade in local markets. Finally, population censuses provide data on access to public sanitation and to the public water network. The variables are measured in different years of 1990–2000, depending on data availability (Table B.10 presents summary statistics; Appendix D reports data sources and definitions).

Figure 3 displays the mean of the different outcome variables as a function of the number of ethnic groups. Contemporary living standards are negatively associated with the number of ethnic groups concentrated in the 16th century. For example, based on the first principal component of the four variables, living standards are 0.32 standard deviations lower, on average, in parishes that concentrated populations from different ethnic groups (35 percent of parishes).[42] To learn about the distribution of living standards, the first panel of Figure 4 plots the mean of the first principal component within bins of average exposure to within-group heterogeneity ($\overline{H}_p$) and ethnic diversity. Below the median level of within-group heterogeneity ($\overline{H}_{p50}$=0.675), most parishes in the top quintile of living standards (in yellow) are parishes with an ethnically homogeneous founding population. Above the median, however, the percentage of parishes in the top quintile drops drastically among ethnically homogenous parishes and increases in those built on various ethnic groups. As an example, the second panel plots the distribution of living standards for parishes that concentrated populations from the *Lucanas*, *Cavinas*, and *Conchucos* ethnic groups, separately. The graphs show different scenarios depending on the average level of past exposure to within-group heterogeneity. Among parishes below the median (e.g., *Lucanas*), the top parishes are those that concentrated ethnically homogeneous populations. Above the median (e.g., *Cavinas*, *Conchucos*), the top parishes are those with populations from different ethnic groups.

Table 6 presents the first statistical examination of the data. To analyze the overall 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

[42]The first principal component of the four variables–log 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)–accounts for 55 percent of the total variance in the data.

across underlying individual effects, jointly with heteroskedasticity-robust standard errors. I first compare contemporary living standards between parishes whose initial populations were ethnically diverse and those with an ethnically homogeneous founding population (equation 1). On average, living standards are 0.2 standard deviations lower in parishes built on ethnically diverse populations. Neither parish-level baseline characteristics nor ecclesiastical jurisdiction fixed effects fully explain this result (Columns 2 and 3).[43] Table B.11 shows that geographic proximity to ethnic boundaries is not responsible for this result either. In particular, I divide the study region into 10 km × 10 km grid cells and create a dummy variable indicating whether a grid cell is part of the buffer of an ethnically diverse parish. Having been part of an ethnically diverse parish is negatively correlated with log nightlight per capita even after accounting for proximity to ethnic boundaries.

I then explore the role of within-group heterogeneity. Column 4 shows that pre-colonial exposure to within-group heterogeneity ($\overline{H}_p$) is positively correlated with contemporary living standards. This correlation is stronger among parishes with ethnic diversity (Column 5) than among parishes where only one ethnic group was concentrated (Column 6). The regressions include parish-level baseline characteristics and ecclesiastical jurisdiction fixed effects. In Column 7, I present the results from estimating the interaction effect of ethnic diversity and average exposure to within-group heterogeneity for the whole sample (equation 2). The estimated coefficient on ethnic diversity is negative, but, consistently, its interaction with $\overline{H}_p$ is positive. This pattern persists when I include fixed effects that account for the colonial administrative province rather than the ecclesiastical jurisdiction (Column 8).[44]

[Table 6 about here]

Table 7 explores this interaction effect in more detail. Panel A shows the results for 1990–2000, while Panel B explores living standards for 2010–2020 based on different waves of the same data sources. All columns except the first, presented for reference, include parish-level baseline characteristics and colonial province fixed effects. I report robust standard errors

[43] The vector of baseline controls includes all variables in Table 5, longitude, and latitude.

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

clustered at the province level in brackets.[45] The estimates are similar for the two decades, showing that the documented pattern is not sensitive to specific years. On average, ethnic diversity is robustly associated with lower living standards. However, parishes whose initial populations were exposed to higher within-group heterogeneity tend to perform relatively better in the long run. Separating the effects into local economic activity and access to public facilities shows that much of the overall effect is driven by differences in economic activity (Columns 2, 4, and 6).[46] Columns 3, 5, and 7 show similar results when controlling for log population density and a rural dummy variable in each decade (Michalopoulos and Papaioannou 2013, 2014).

[Table 7 about here]

Figure 5 plots the estimated average effect size of ethnic diversity as a function of past exposure to within-group heterogeneity (Panel B of Table 7). Dashed lines represent 95 percent confidence intervals. 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, going from 0.6 to 1 of average within-group heterogeneity is associated with 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 parishes in the sample. The next section presents supplementary exercises and robustness checks.

[Figure 5 about here]

[45] Table B.12 reports OLS estimates for each outcome variable, separately, and standard errors adjusted for spatial autocorrelation (Colella et al. 2019), as well as the R-squared.

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

5.2 Supplementary Analyses

5.2.1 Pre-Colonial Characteristics of Ethnic Groups

Individuals from more heterogeneous ethnic groups were likely exposed to stronger economic and institutional development before the resettlement (Section 3.3.3). In Table 8, I show that the main result persists when controlling for the pre-colonial correlates of within-group heterogeneity. I consider the set of pre-colonial ethnic-level characteristics analyzed in Table 3. In particular, I compute the weighted average of each characteristic among the ethnic groups concentrated in the parish and extend 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 for reference (Column 2 of Table 7 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 pre-colonial ethnic characteristics (Belloni, Chernozhukov, and Hansen 2014).[47]

[Table 8 about here]

In Table 9, 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 with different levels of exposure to within-group heterogeneity that are statistically similar in all other pre-colonial characteristics ($\overline{G}_p$). In Panel A, the pre-colonial characteristics to be balanced by the matching algorithm are the baseline characteristics used in Column 9 of Table 8. In Panel B, the procedure uses the set of lasso-*selected* pre-colonial characteristics. Table B.13 documents that, in the matched samples, parishes that concentrated

[47] Apart from those in Column 9 of Table 8, the set of characteristics to be considered by the lasso routine includes all variables related to pre-colonial infrastructures (Table B.4) and land caloric suitability for maize–a total of 13 characteristics, of which 7 are *selected* by lasso.

ethnically diverse and non-diverse populations continue to be balanced along geographic and initial factors. Despite the reduced sample size, the estimated AES of ethnic diversity in the matched samples is consistent with previous results (Column 1). Estimates using the first principal component of the four development outcomes as dependent variable and matching weights show the same pattern (columns 2 and 3).[48] Finally, Figure B.4 shows that the coefficients of interest are qualitatively and quantitatively similar after excluding all parishes in which a particular ethnic group was present (for one ethnic group at a time), which alleviates concerns that certain ethnic groups could be driving the result.

[Table 9 about here]

5.2.2 Pre-Colonial Land Occupation, Transition Zones, and Placebos

The estimates in Tables 6 and 7 are likely affected by non-classical measurement error. A potential source of error is the underlying assumption that individuals were uniformly distributed over space during the pre-colonial period. In Table 10, I consider alternative scenarios. In the absence of historical data on the spatial distribution of the population within ethnic groups, I follow recent archaeological studies in using pre-colonial site records as evidence of land occupation (Morrison et al. 2021). Columns 1 and 2 show the estimates of equation 2 after restricting the total land area in the region of analysis to a distance of 20 km and 10 km around pre-colonial archaeological sites, respectively, in order to compute H_e.[49] The point estimates are slightly higher, although the estimated effect of ethnic diversity echoes the baseline result ($-0.616 + 0.879 \times \overline{H}_p$). A second underlying assumption is that all coethnic individuals were equally exposed to within-group heterogeneity. In Columns 3 and 4, I compute H_e after restricting the total land area to 20 km and 10 km around the transitions from one elevation zone to another, respectively. Although the available data suggest that

[48] Weights account for imbalance in the number of parishes with high and low within-group heterogeneity (i.e., $\overline{H}_p$ above or below the median, respectively) for each combination of the categories of coarsened characteristics. The AES routine does not accept weights.

[49] This corresponds to a reduction in the total land area of 7.35 and 31.33 percent, respectively. For this exercise, I combine a public inventory of pre-colonial archaeological sites (*Catastro de Monumentos Arqueológicos Prehispánicos*, Ministerio de Cultura, Perú) with my own survey of published archaeological studies, catalogues, and handbooks (see Ravines Sánchez 1985, Ramos Giraldo 2001, and Isbell and Silverman 2002a, 2008).

ethnic boundaries are relevant in order to explain crop exchange (Section 3.3.4), gains could have been higher around zones of transition (Bates 2001). The estimates are similar to baseline ones, suggesting that exposure among coethnics did not differ significantly with distance to transition zones.

The empirical analysis also relies on the mapping of historical ethnic boundaries. Given the historical nature of the map (Figure 1), it is important to assess how precise the delineated boundaries are. In Column 5 of Table 10, I follow the approach in Alesina, Michalopoulos, and Papaioannou (2016). In particular, I perform the analysis using artificial ethnic boundaries (from Thiessen polygons) rather than historical ones.[50] Compared to baseline results, the estimates are small and not statistically significant, which suggests that historical ethnic boundaries matter. This is in line with the evidence in Section 3.3.4, where I use grid cells instead of Thiessen polygons. In Column 6, I use the approximate boundaries of the first administrative demarcations created after the Spanish conquest (*corregimientos*) instead. Close correspondence between the spatial boundaries of pre-colonial ethnic groups and *corregimientos* would suggest that the Spanish administration created the latter based on prior knowledge of the spatial distribution of the groups. The fact that the point estimates are not statistically different from zero suggests that this was not the case.

[Table 10 about here]

5.2.3 Sensitivity Analyses and Additional Robustness Checks

The appendix reports estimates from using alternative definitions to measure the key explanatory variables. Table B.14 reports the results from varying the size of the buffer used to detect the ethnic origins of the populations. Table B.15 shows the results from using an index of ethnic fractionalization ($Ethnic\ frac_p = 1 - \sum_e w_{pe}^2$) to measure of ethnic diversity. However, since w_{pe} represents the area share of ethnic group e within the buffer of parish p rather than the exact population share, this measure may arguably be more affected by measurement error than $Ethnic\ div_p$. Table B.16 employs a Herfindahl index ($\widetilde{H}_e = 1 - \sum_j s_{ej}^2$) to measure

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

within-group heterogeneity. Reassuringly, the previous estimates are consistent with baseline results. Figure B.5 displays point estimates and confidence intervals after excluding one parish at a time, alleviating concerns about influential observations. Appendix B.2.1 shows that the main result of the paper holds for household consumption and addresses selective migration.

Finally, it is important to note that any potential effect of ethnic diversity after the resettlement should be conditional on the survival of ethnic groups. The decline in native populations after European contact has been well documented by historical studies (Cook 1982; Denevan 1992). To the extent that all groups were similarly affected by disease and abuse, the estimates should be interpreted as the effect of ethnic diversity among the descendants of survivors. Table B.19 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 and known to have affected native populations before the resettlement—and to controlling for variables related to the Inca period.[51] I also consider variables related to the structure of the population by the late colonial period (Table B.20).[52]

6 Mechanisms

6.1 Cultural Transmission

This section studies cultural transmission as an underlying mechanism. Strategies to cope with environmental risk and adverse geography can help sustain cooperation and more trusting

[51] 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), log density of Inca roads, which connected Cuzco with the rest of the territories, and log population density. 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 (*suyus*) and after excluding parishes that concentrated groups potentially affected by Inca resettlements (Bongers et al. 2020), separately.

[52] The first two columns of Table B.20 show that the results are robust to controlling for the log of the "indigenous" population and the percentage of the "mestizo" (indigenous-Spanish mixed) population in the parish, defined according to the 1791–1795 census. The remaining columns consider variables related to the administration of religion, a major goal of the resettlement. Specifically, the results are robust to controlling for the log number of priests per capita and to the inclusion of religious order fixed effects. Only one parish in the sample was administered by the Jesuit order, which has been shown to positively influence long-run human capital and economic development (Valencia Caicedo 2019).

attitudes over time (Nunn and Wantchekon 2011; Nunn and Puga 2012; Buggle and Durante 2021).[53] A potential explanation to the documented pattern of development is that pre-colonial internal exchange aimed at reaching a common goal (i.e., maximizing the economic base) contributed to the formation of a culture of cooperation. After resettlement, the transmission of cooperative behavior and more open attitudes toward out-group members may have facilitated local interactions between ethnic groups.[54]

Ethnohistoric studies and early chronicles offer a consistent interpretation. For example, 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." Anecdotally, the importance of crop sharing can also be found in native folklore. For example, Berezkin (2015)'s folklore catalogue 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). Furthermore, during Inca times, the chronicles suggest that "if it was necessary for someone to do something else in an emergency, like war or some other urgent matter, the other Indians of the community worked the fields of the absent man without asking or receiving any compensation beyond their food, and, this done, each cultivated his own fields. This assistance which the community rendered to its absent members caused each man to return home willingly when he had finished his job, for he might find on his return after long absence that a harvest which he had neither sown nor reaped was gathered into his house" (Cobo 1890 [1653], in Rowe 1946, p. 266).

[53] Nunn and Puga (2012) document the indirect effect of ruggedness on the development of African countries by allowing protection from slave traders. Separately, a culture of mistrust has been shown to persist among the descendants of individuals affected by the slave trade (Nunn and Wantchekon 2011). Buggle and Durante (2021) find that European regions exposed to higher environmental risk during the pre-modern era exhibit higher levels of inter-personal trust today. The study argues that, in face of variability in temperature and precipitation, farmers developed cooperative strategies that contributed to the emergence of more trusting attitudes.

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

I explore in the data whether the formation of a culture of cooperation and more favorable attitudes toward out-group members is a plausible channel. I first study inter-group contact during the colonial period. Did exposure to within-group heterogeneity favor inter-group interactions? Comparing the first and second surnames of each individual in the sample of colonial baptism records offers the opportunity of exploring inter-group unions.[55] The sample includes 17,411 individuals with native first and second surnames distributed across 41 parishes, of which 10 are parishes built on ethnically diverse populations. Since I do not observe ethnicity but only surnames in the data, I use a measure of dissimilarity (Levenshtein distance, L) between the two surnames of each individual to detect parents who were potentially *from different ethnic groups*. The measure, defined as the minimum number of spelling changes required to transform one surname into another, is divided by the length of the longest surname to be interpreted as the percentage of dissimilarity between surnames.[56]

Figure 6 presents a graphical summary of the data. I compute the percentage of dissimilarity for each union and then obtain the average dissimilarity at the parish level. The left graph of Panel (a) suggests a positive correlation with exposure to within-group heterogeneity in the subsample of parishes with ethnic diversity. The right graph shows a similar pattern when using instead the share of unions with dissimilarity levels above 50 percent. The two graphs account for the log of the total number of individuals found in the records of the parish. Panel (b) replicates the exercise for the subsample of parishes without ethnic diversity. Reassuringly, the results suggest that no correlation exists in this subsample.[57]

[Figure 6 about here]

The data suggest that inter-ethnic unions may have been more likely where individuals had a history of within-group heterogeneity, resulting in a more integrated society.[58] Small-N

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

[56]See Dickens (2022) for an application of this measure to compute distance between languages.

[57]Despite the small sample size, the results from OLS regressions are consistent with this interpretation (Table B.21, Columns 1-7). The same pattern holds at the union level (Columns 8-11).

[58]Marriage between different ethnic groups is often used as an indicator of societal integration (Gordon 1964; Bazzi et al. 2019).

results, however, should be interpreted with caution. Contemporary survey data on individuals' self-reported identity show a consistent interpretation. When asked "Which group do you identify with the most?," where the resettlement forced together ethnically diverse populations, individuals tend to identify more strongly with the state the higher the level of past exposure to within-group heterogeneity, compared to where the resettlement concentrated ethnically homogeneous populations (Table 11, Columns 1-3).[59] All regressions include individual-level controls (gender, age, age squared, years of schooling, civil status, and mother tongue), thus comparing individuals with similar socio-demographic characteristics, and survey-year fixed effects, with standard errors clustered at the parish level. These individuals are also more likely to vote in presidential elections (Column 4), a result that holds beyond trust in the state (Column 5). Administrative data on the number of volunteers for military service show the same pattern (Columns 6-7), which is consistent with a more integrated society.

[Table 11 about here]

In Table 12, I then analyze participation in voluntary organizations as a proxy for cooperative behavior (Guiso, Sapienza, and Zingales 2016). Columns 1-4 use the individual-level survey data to explore participation in different types of associational groups such as neighborhood, labor, and professional associations. Conditional on individual-level socio-demographic controls, year fixed effects, and log associations per capita, where the resettlement resulted in ethnically diverse jurisdictions, individuals are more likely to associate with others the higher the level of past exposure to within-group heterogeneity.[60] The estimates are consistent with census data on the share of farmers organized in the form of communal associations or committees (Column 5) and with administrative data on the presence of neighbourhood associations (Column 6) for the earliest years available.

[Table 12 about here]

[59]The ENAHO Peruvian household survey includes a question on whether individuals identify more strongly with their state administrative unit, ethnicity or race, religion group, native community, or other.

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

6.2 Structural Transformation

This section explores the extent to which the documented pattern of development was accompanied by a shift in the structure of economic activity from agriculture toward non-agricultural sectors. On average, non-subsistence agriculture predominates where the 16th-century intervention resulted in ethnically diverse populations with a high level of past exposure to within-group heterogeneity. This pattern holds beyond the local availability and diversity of crops (Table B.22), which suggests that a history of within-group heterogeneity sustained by internal exchange may have contributed to more favorable attitudes toward local trade.[61] In turn, complementary skills and technologies may have flourish to support a market-oriented society.

In Table 13, I explore data from the 1876 population census, the earliest post-independence census with detailed information on occupations. In particular, I classify the different occupations in the sample by sector of economic activity and then compute the share of population employed in each sector. Most of the sample continued to be predominantly agricultural by the late 19th century, with 70 percent of the population employed in the primary sector, on average. However, as past exposure to within-group heterogeneity increases, employment tends to be relatively more oriented toward tertiary-sector activities where the intervention resulted in ethnically diverse populations, compared to where the resettlement resulted in populations from a single ethnic group (Column 1). The change in the structure of economic activity happened to the detriment of agriculture (Column 3). Early literacy could have been an advantage in the transition out of the agricultural sector (Porzio, Rossi, and Santangelo 2022, Column 7).[62] Decomposing tertiary-sector employment into local

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

[62] Consistently, 20th-century data show that, in these places, farmers are more open to receive technical support in order to improve agricultural productivity (Table B.23, Column 1). Data on technology adoption in agriculture show no statistically significant differences (Columns 2-5). However, since chemical fertilizers and similar products have traditionally been imported in the study setting (see https://agraria.pe/noticias/peru-importa-1-2-millones-de-toneladas-de-fertilizantes-sint-26839 or https://www.technoserve.org/blog/building-

trade and other services shows that trade is driving this effect (Columns 9 and 11), which is consistent with the result that more farmers sell their products in local markets. Census data on 21th-century employment show the same pattern (Columns 10 and 12).

[Table 13 about here]

I find no statistically significant evidence that these places were more involved in manufacturing by the late 19th century, which suggests they were no more industrialized (Columns 5 and 6), nor evidence that past exposure to crop heterogeneity might have indirectly fostered industrialization (Fiszbein 2022) or tertiary-sector occupation through the potential increased availability of skills (Table B.24). Rather, in a setting that continues to be predominantly agricultural, the empirical evidence points toward a relatively more integrated society with a tendency to engage in local trade where ethnically diverse populations had a historical experience of within-group heterogeneity sustained by internal exchange.

6.3 Economic Complementarities

Where the ethnic minority had comparative advantage over the majority group, economic complementarities may also have sustained beneficial coexistence (Jha 2013, 2018; Becker and Pascali 2019).[63] In the study setting, 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 the desperate majority). To test this hypothesis, I define the minority group as the one with the lowest share of area in the 10-km buffer, resulting in 44 different ethnic minorities (out of 47 ethnic groups in the sample), and run regressions of the following form for the subsample of parishes built on ethnically

farmers-resilience-to-the-fertilizer-crisis-in-peru/), they may not have been equally available across the territory. To my knowledge, no such data is available for the previous century.

[63] The theoretical framework developed in Jha (2013, 2018) establishes that peaceful inter-group coexistence can be sustained through the specialization of ethnic groups into complementary activities that are costly to replicate and expropriate. Jha (2013) provides consistent empirical evidence on tolerance toward Muslims in Hindu societies. Other studies have focused on anti-Semitism, finding consistent empirical results; see Becker and Pascali (2019) for evidence in the context of the Protestant Reformation in Germany and Jedwab, Johnson, and Koyama (2019) for evidence from the Black Death in Europe.

diverse populations (Table 14):

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

where y_p refers to contemporary living standards, $High\, min_p$ is a dummy variable indicating whether the minority belonged to a relatively heterogeneous ethnic group (H_e above the 50th percentile in Column 1 and above the 75th percentile in Column 2), and $High\, maj_p$ is an analogous dummy variable for the majority group. The results do not rule out economic complementarities as a potential channel—where the historical ethnic minority likely had a comparative advantage, contemporary populations perform relatively better ($\hat{\delta}_1 > 0$). In Column 3, I explore whether, in line with this interpretation, marketplaces for the exchange of local goods tend to be located in these places. In particular, I use a dummy for the presence of local retail markets as outcome variable. Consistently, the estimated coefficient on $High\, min_p$ is positive and statistically significant. This result is not explained by total population (Column 4) or agglomeration (Column 5).

[Table 14 about here]

7 Conclusion

A large body of empirical literature has examined the implications of ethnic diversity for economic growth and development. However, we know relatively little about the role of within-group heterogeneity. This paper shows that exposure to within-group heterogeneity in complementary traits matters for understanding the long-run effect of ethnic diversity on comparative development. I collect new data from a natural experiment of Peru's colonial history and find robust evidence that, where colonial officials concentrated populations with a history of within-group heterogeneity, who, prior to colonization, settled in complementary climates of the Andes to maximize the economic base, ethnic diversity results in systematically lower costs and may even become advantageous.

In a setting that continues to be predominantly agricultural, additional evidence shows that, despite having built society based on ethnically diverse populations, there is a relatively more integrated society with a tendency to engage in associational activities and in local trade where these populations had a historical experience of within-group heterogeneity sustained by internal exchange. Furthermore, where the historical ethnic minority likely had comparative advantage over the majority group, contemporary populations perform relatively better, suggesting that economic complementarities may also have contributed to sustain inter-ethnic coexistence.

The subsistence strategy of pre-colonial Andean groups, characterized by a single ethnic group having simultaneous control over altitude-specific resources, contributed to shape 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 studying the internal economic organization of ethnic groups will likely help understand comparative development.

References

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

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

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. URL http://www.wwnorton.com/college/titles/polisci/bates/welcome.htm.

Bauer, Brian S, LC Kellett, and Aráoz Silva. 2010. *The Chanka: Archaeological Research in Andahuaylas (Apurimac)*. University of California, Los Angeles: Cotsen Institute of Archaeology Press. 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.

Benson, AA and M Calvin. 1948. "The Path of Carbon in Photosynthesis." *Science* 107 (2784):476–480.

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

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

Bongers, Jacob L, Nathan Nakatsuka, Colleen O'Shea, Thomas K Harper, Henry Tantaleán, Charles Stanish, and Lars Fehren-Schmitz. 2020. "Integration of Ancient DNA with Transdisciplinary Dataset Finds Strong Support for Inca Resettlement in the South Peruvian Coast." *Proceedings of the National Academy of Sciences* .

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.

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

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.

Cobo, Bernabé. 1890. *Historia del Nuevo Mundo (Publicado por Primera Vez con Notas y Otras Ilustraciones de D. Marcos Jiménez de la Espada)*. Sevilla: Sociedad de Bibliófilos Andaluces. [1653].

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.

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

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

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

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).

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

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

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.

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

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.

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

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

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

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

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].

Gootenberg, Paul. 1991. "Population and Ethnicity in Early Republican Peru: Some Revisions." *Latin American Research Review* 26 (3):109–157. URL http://www.jstor.org/stable/2503666.

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

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 and Helaine Silverman. 2002a. *Andean Archaeology I*. Springer.

———. 2002b. "Theorizing Variations in Andean Sociopolitical Organization." In *Andean Archaeology I*. Springer, 3–11.

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

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.

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

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

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

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

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.

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

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

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

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

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

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

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

McCrae, Robert R and Paul T Costa. 1997. "Conceptions and Correlates of Openness to Experience." *Handbook of Personality Psychology* :825–847.

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.

———. 2018. “Spatial Patterns of Development: A Meso Approach.” *Annual Review of Economics* 10:383–410.

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* .

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

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

Murra, J. 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-Fondo Editorial de la Pontificia Universidad Católica del Perú.

———. 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, D. 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.

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

Porzio, Tommaso, Federico Rossi, and Gabriella Santangelo. 2022. "The Human Side of Structural Transformation." *American Economic Review* 112 (8):2774–2814.

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

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.

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

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.

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.

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

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.

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

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

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

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

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: printed for W. Strahan; and T. Cadell.

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.

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

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

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

Tello, 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.

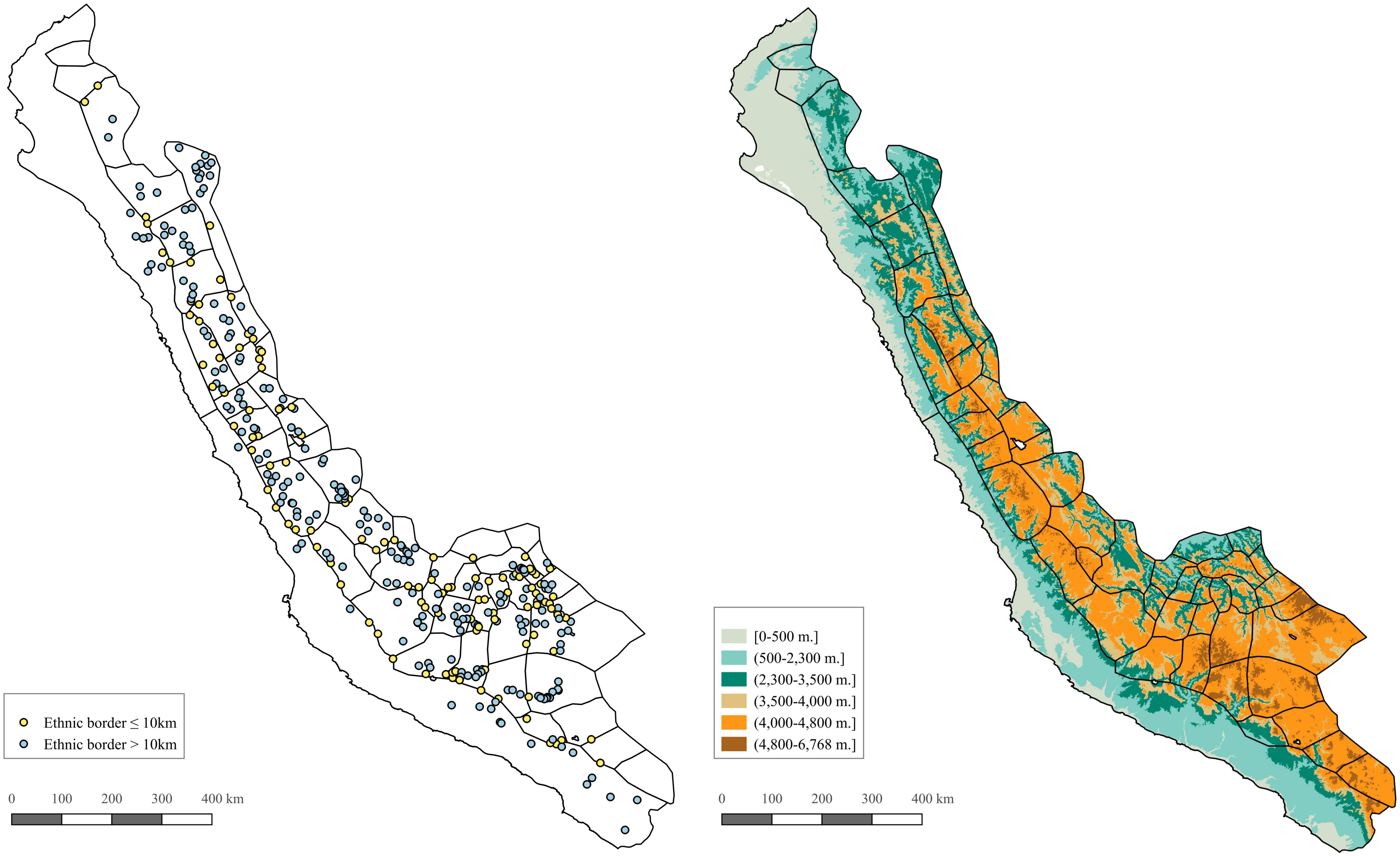
Vollmer, Günter. 1967. *Bevölkerungspolitik und Bevölkerungsstruktur im Vizekönigreich Peru zu Ende der Kolonialzeit, 1741-1821*. Berlin: Gehlen.

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

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.

Zuidema, Reiner Tom and Deborah Poole. 1982. "Los Límites de los Cuatro Suyus Incaicos en el Cuzco." *Bulletin de l'Institut Français d'Études Andines* 11 (1-2):83–89.



(a) Colonial Parishes (b) Resource-Producing Zones

FIGURE 1: Colonial Parishes and Resource-Producing Zones

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

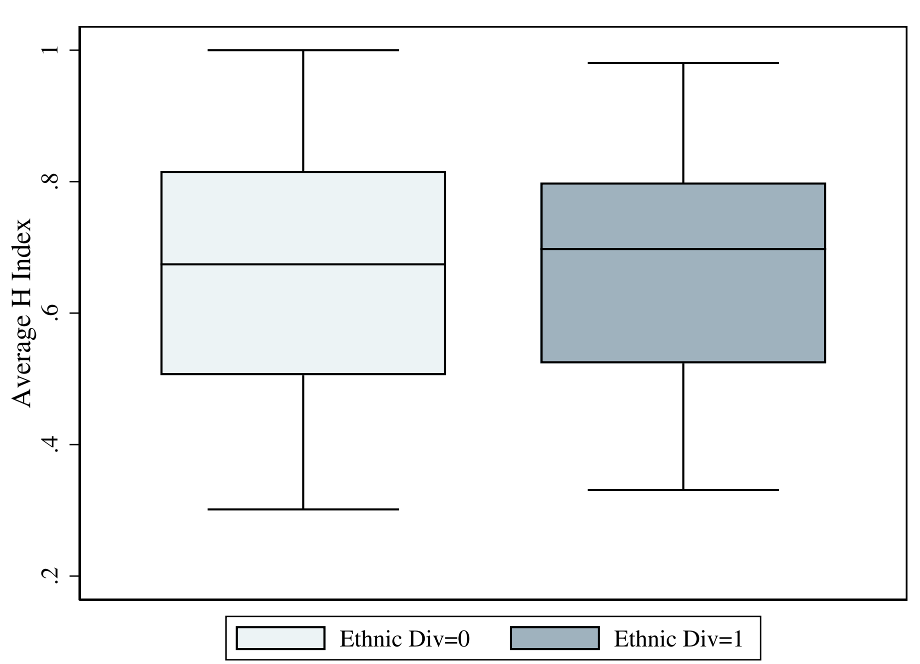


FIGURE 2: Boxplots: Ethnic Diversity and Average Within-Group Heterogeneity

Notes. Boxplots of average within-group heterogeneity ($\overline{H}_p$) 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.

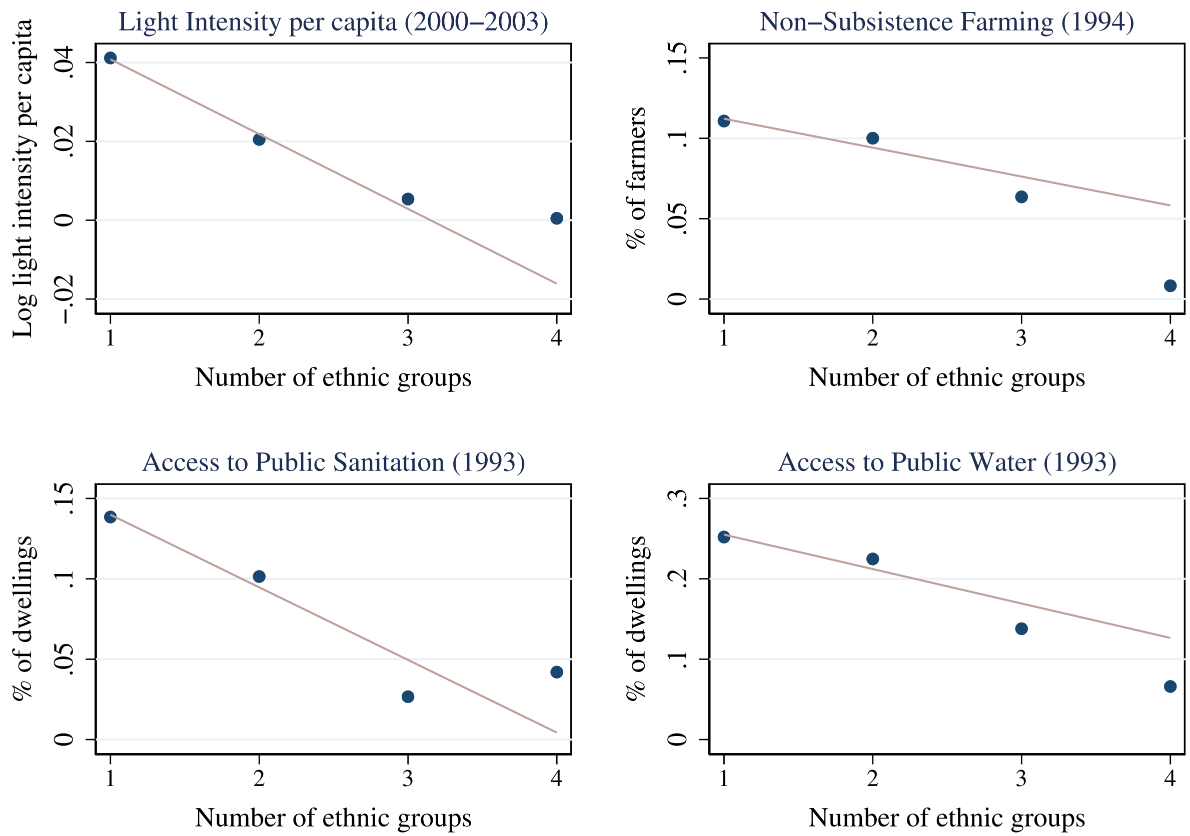


FIGURE 3: Number of Ethnic Groups and Contemporary Development

Notes. Mean of the different outcome variables–log average light intensity per capita (2000–2003), share of farmers practicing non-subsistence agriculture (1994), share of dwellings with access to public sanitation (1993), and share of dwellings with access to the public water network (1993)–as a function of the number of ethnic groups. The figure uses parish-level data. The x-axis refers to the number of ethnic groups within a buffer of 10-km radius from the parish capital. Most of the parishes with ethnic diversity (85 percent) concentrated two ethnic groups. The remaining 15 percent of parishes concentrated either three or four groups.

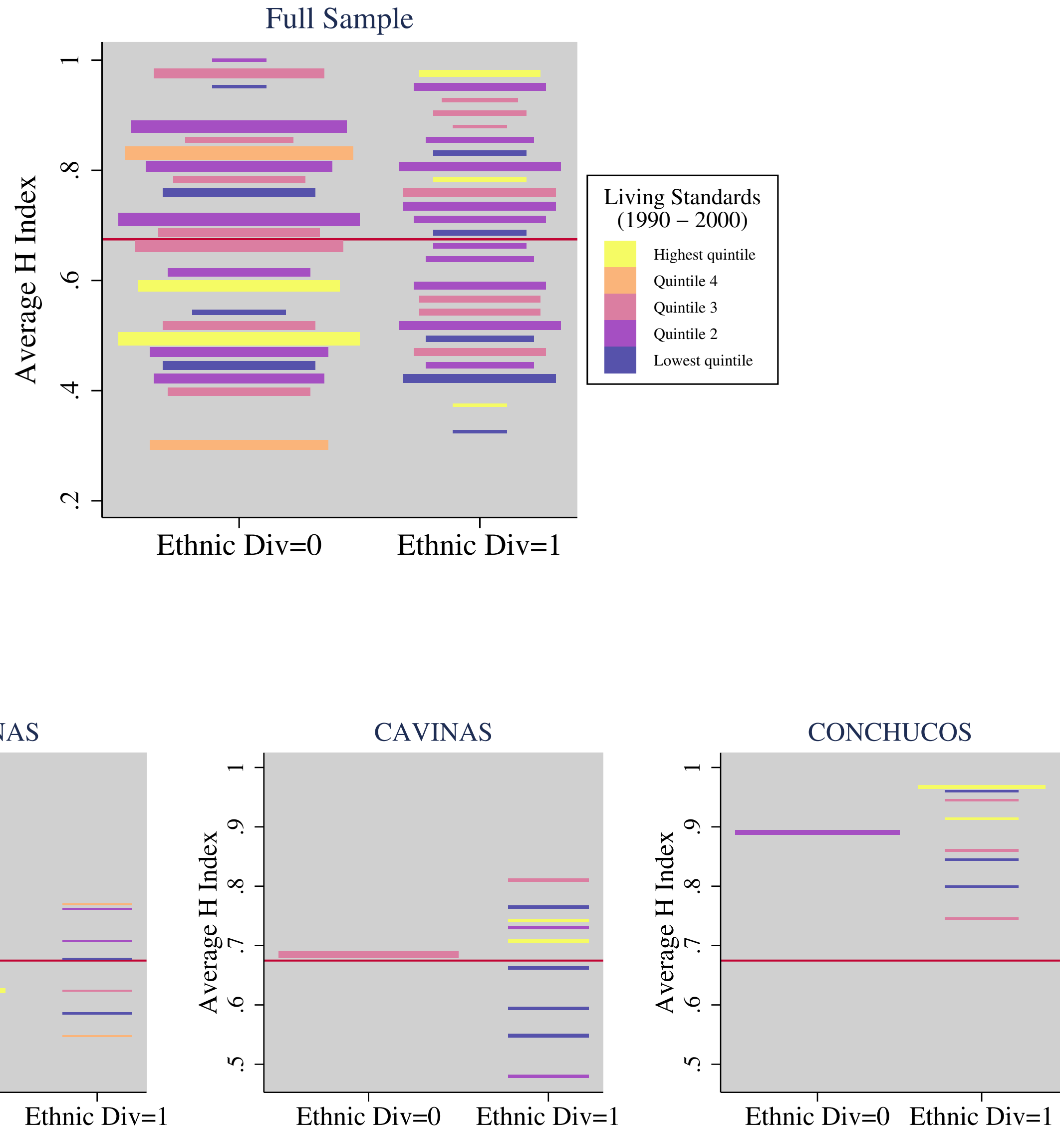


FIGURE 4: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development

Notes. The color legend indicates the mean of the first principal component of four variables–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. 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 size is scaled according to the relative frequency of each combination of ethnic diversity and binned within-group heterogeneity. The red line indicates the median of average exposure to within-group heterogeneity. The first panel includes all parishes, the second panel includes parishes that concentrated populations from the *Lucanas*, *Cavinas*, and *Conchucos* ethnic groups in separate graphs.

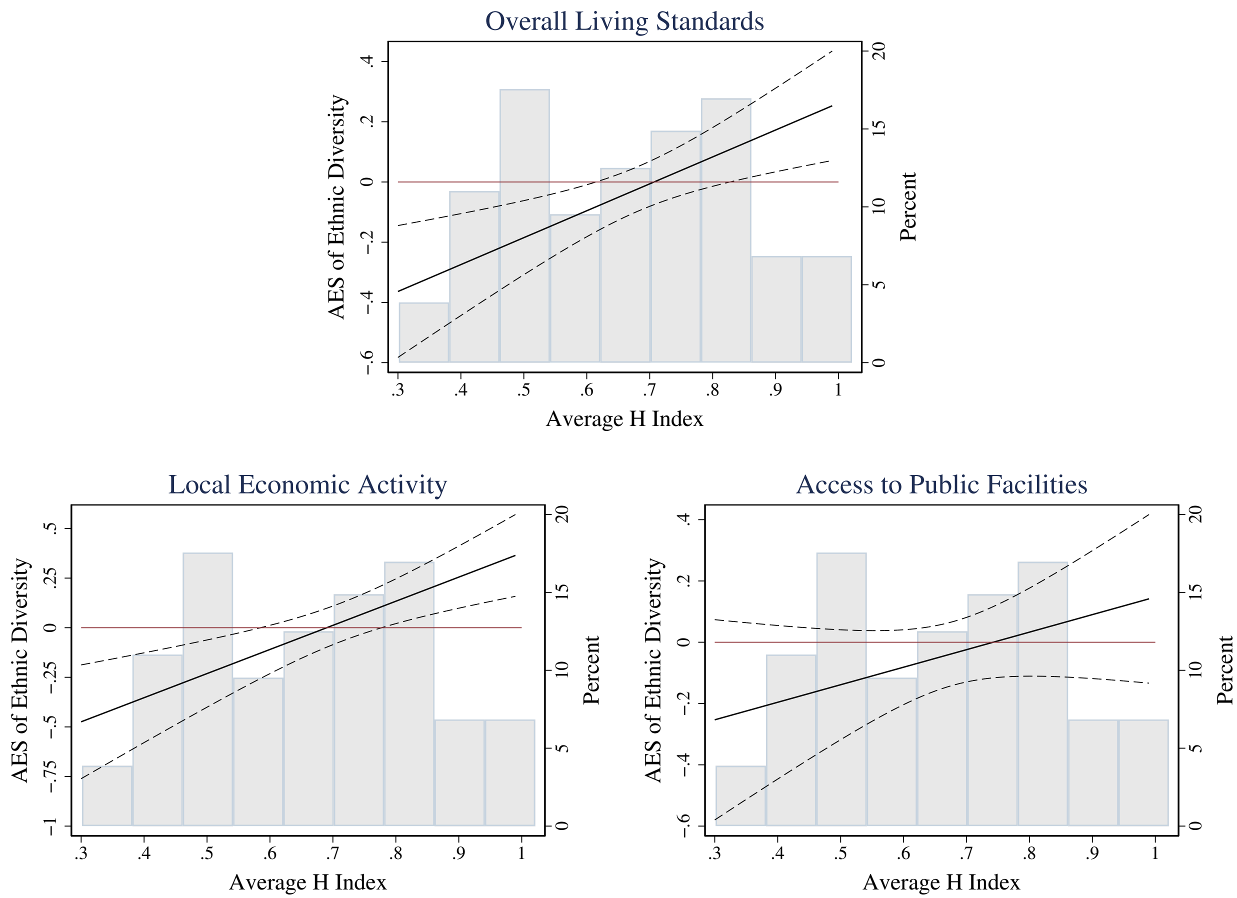
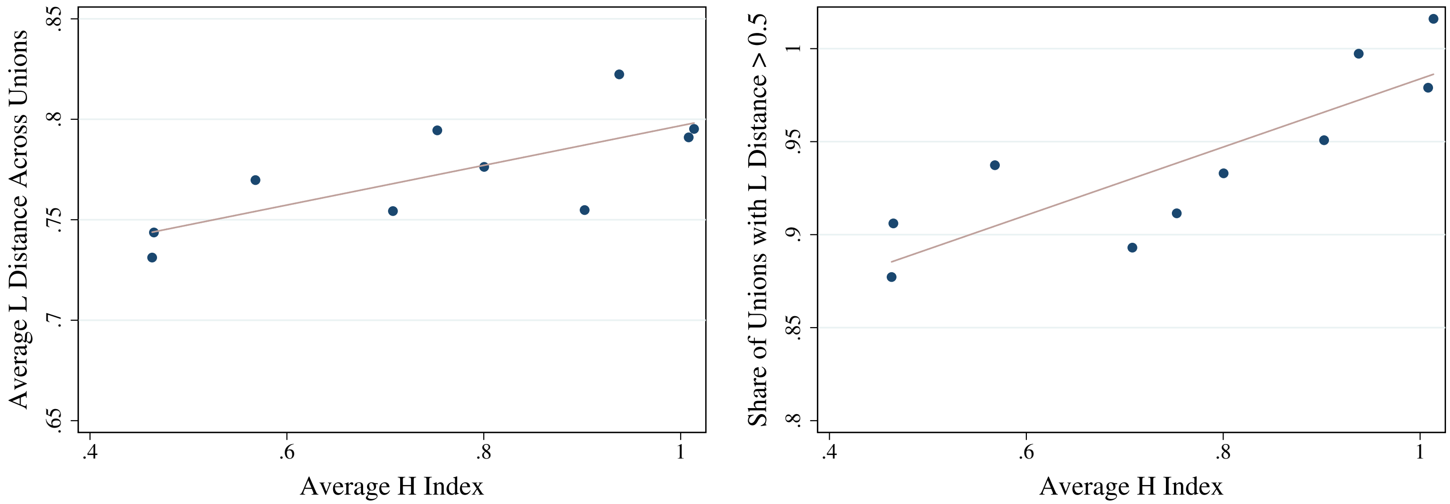
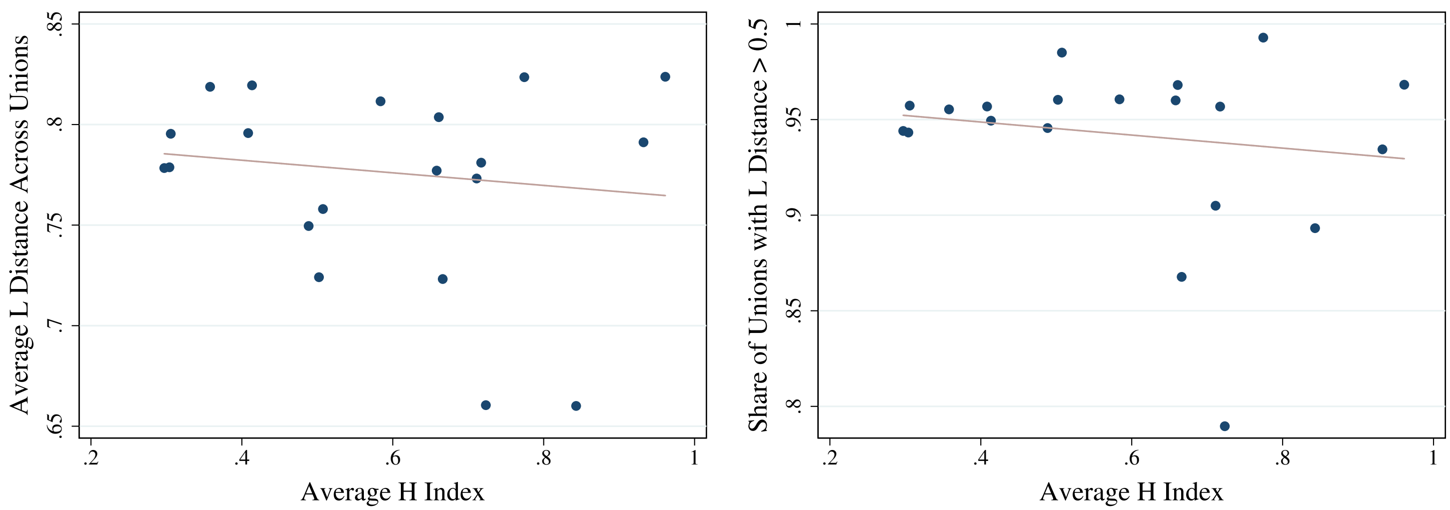


FIGURE 5: The Effect of Ethnic Diversity on Contemporary Development

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

	Dependent Variable: Surname Diversity (1605 – 1780)							
	S Index	S Index	S Index	S Index	D Index	D Index	D Index	D Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Baseline							
Ethnic diversity	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							
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							
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)**
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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: Validating Resource-Producing Zones

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

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

Table 3: 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)
Observations	47	47	47	47	46	46	47	47	47

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

Table 4: Within-Group Heterogeneity and Pre-Colonial Diets

	Dependent Variable: Carbon Isotope Score ($\delta^{13}C_{col}$)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: All Individuals					
H index (ethnic group level)	0.501	0.279	0.851	0.261	0.728	
	[0.153]**	[0.089]**	[0.145]***	[0.071]***	[0.062]***	
	(0.145)***	(0.094)***	(0.107)***	(0.069)***	(0.055)***	
H index (grid cell level)						-0.028
						[0.334]
						(0.194)
Number of individuals	196	196	196	196	196	196
	Panel B: Excluding Children					
H index (ethnic group level)	0.554	0.303	0.874	0.269	0.750	
	[0.152]***	[0.106]**	[0.148]***	[0.085]**	[0.065]***	
	(0.144)***	(0.112)***	(0.107)***	(0.081)***	(0.058)***	
H index (grid cell level)						-0.047
						[0.343]
						(0.194)
Number of individuals	182	182	182	182	182	182
Zone FE	No	Yes	Yes	Yes	Yes	Yes
Ln (# Crops)	No	No	Yes	No	No	No
Land caloric suitability	No	No	No	Yes	Yes	No
Institutional and socioecon. controls	No	No	No	No	Yes	No
Number of ethnic groups	8	8	8	8	8	–
Number of grid cells	–	–	–	–	–	8

Notes. The unit of observation is the individual. The table reports OLS estimates. Standard errors clustered at the ethnic group 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 (# Crops)* refers to the number of different native crops within the ethnic homeland. *Land caloric suitability* refers to the average caloric suitability of the ethnic homeland. The vector of institutional and socio-economic controls includes ethic-level dummies for pre-colonial political complexity, urbanization, and elite residences. All regressions control for the longitude and latitude of the archaeological site where the individual's remains were found. All variables except dummies are standardized to have zero mean and standard deviation equal to one. The grid-cell level H index refers to grid cells of 50km × 50km. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: 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 6: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development I

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

Notes. The unit of observation is the parish. Robust standard errors in brackets. 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 (2000–2003), an indicator for non-subsistence agriculture (1994, 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 (1993), and the share of dwellings with access to the public water network (1993). The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to 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 7: Ethnic Diversity, Within-Group Heterogeneity, and Contemporary Development II

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

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

Table 8: Robustness: Pre-Colonial Characteristics of Ethnic Groups I

	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), an indicator for 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 9: Robustness: Pre-Colonial Characteristics of Ethnic Groups II

	Overall Living Standards (2010 – 2020)		
	Matched Sample (AES)	Matched Sample (PCA)	Matched Sample and Weights (PCA)
	(1)	(2)	(3)
	Panel A: Baseline Pre-Colonial Characteristics		
Ethnic diversity	-0.490	-0.789	-0.561
	[0.165]***	[0.426]*	[0.398]
		(0.316)**	
Ethnic div. × High Av. H	0.737	1.084	1.189
	[0.287]**	[0.698]	[0.683]*
		(0.335)***	
Number of parishes	103	103	103
	Panel B: Lasso-*Selected* Pre-Colonial Characteristics		
Ethnic diversity	-0.475	-0.825	-0.862
	[0.152]***	[0.291]***	[0.295]***
		(0.268)***	
Ethnic div. × High Av. H	0.889	1.359	1.415
	[0.225]***	[0.469]***	[0.449]***
		(0.434)***	
Number of parishes	78	78	78
Baseline controls	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). 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 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 8) in Panel A and the lasso-*selected* characteristics (Column 10 of Table 8) in Panel B. Column (1) 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), an indicator for 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). Columns (2) and (3) report corresponding estimates when using the first principal component of the four previous variables as the dependent variable (standardized to have zero mean and standard deviation equal to one). 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 10: 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), an indicator for 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 11: Cultural Transmission Mechanism – Evidence From Societal Integration

	Dependent Variable:						
	Which Group Do You Identify With The Most? (2004-2017)			Voting in Presidential Elections (2007-2011)		Military Service (2008-2014)	
	State	Ethnicity or Race	Religion	Dummy (Yes, I Voted in the 2006 Election)		Ln (1 + Av. Volunteers)	Dummy (Volunteers ≥ 0)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ethnic diversity	-0.155	0.020	0.001	-0.052	-0.052	-0.128	-0.228
	[0.050]***	[0.013]	[0.027]	[0.035]	[0.035]	[0.048]**	[0.095]**
	(0.056)***	(0.018)	(0.030)	(0.043)	(0.044)	(0.037)***	(0.073)***
Ethnic div × Av. Within-Group H	0.208	-0.026	0.006	0.104	0.104	0.240	0.361
	[0.071]***	[0.018]	[0.038]	[0.047]**	[0.047]**	[0.082]***	[0.130]***
	(0.088)**	(0.024)	(0.049)	(0.054)*	(0.054)*	(0.076)***	(0.127)***
Trust in the state					0.011		
					[0.011]		
					(0.012)		
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	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
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

Notes. The unit of observation is the individual in Columns 1-5 (individual-level data from yearly waves of the ENAHO Peruvian household survey) and the parish in Columns 6-7. In brackets, robust standard errors clustered at the parish (Columns 1-5) or province (Columns 6-7) level. 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. The question 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 mother tongue. Columns 6-7 control for the log 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 12: Cultural Transmission Mechanism – Evidence From Participation in Voluntary Associations

	Dependent Variable:					
	Participation Dummy (2004-2017)				Dummy Neigh.	Dummy (% Farmers in Communal Assoc.
	Neigh. Assoc.	Professional Assoc.	Labor Assoc.	Some Assoc.	Assoc. (2002-2003)	≥ Median) (1994)
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic diversity	-0.059	-0.030	-0.058	-0.140	-0.421	-0.343
	[0.036]	[0.014]**	[0.031]*	[0.052]***	[0.159]**	[0.145]**
	(0.048)	(0.017)*	(0.033)*	(0.056)**	(0.127)***	(0.155)**
Ethnic div. × Av. H index	0.096	0.047	0.074	0.206	0.672	0.514
	[0.046]**	[0.021]**	[0.039]*	[0.068]***	[0.238]***	[0.220]**
	(0.066)	(0.030)	(0.039)*	(0.072)***	(0.198)***	(0.240)**
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	–	–
Year FE	Yes	Yes	Yes	Yes	–	–
Number of parishes	280	280	280	280	336	334
Number of individuals	54,676	54,676	54,676	54,676	–	–
Mean Dep. Var.	0.076	0.036	0.063	0.157	0.190	0.500

Notes. The unit of observation is the individual in Columns 1-4 (individual-level data from yearly waves of the ENAHO Peruvian household survey) and the parish in Columns 5-6. In brackets, robust standard errors clustered at the parish (Columns 1-4) or province (Columns 5-6) level. 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. Column 5 controls for the log of the total population and Column 6 for the log of the total number of farmers. In Columns 1-4, the vector of individual-level controls includes gender, age, age squared, years of schooling, civil status, and mother tongue. The vector of baseline controls includes parish-level mean and standard deviation of elevation, mean and standard deviation of land caloric suitability, longitude, latitude, log distance to perennial rivers, log distance to 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 13: Mechanisms: 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	-0.111***	-0.250**	0.204*	0.340**	-0.092	-0.090
	[0.041]	[0.104]	[0.110]	[0.141]	[0.137]	[0.056]
Ethnic div. × Av. H index	0.151**	0.340*	-0.293*	-0.459**	0.142	0.119
	[0.067]	[0.170]	[0.155]	[0.216]	[0.208]	[0.073]
Mean Dep. Var.	0.073	0.336	0.687	0.517	0.240	0.147
	Can Read and/or Write		Tertiary Sector: Local Trade		Tertiary Sector: Other Services	
	(1876)	(2007-2017)	(1876)	(2007-2017)	(1876)	(2007-2017)
	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic diversity	-0.078**	-0.021	-0.110***	-0.108**	-0.001	-0.077*
	[0.034]	[0.035]	[0.039]	[0.044]	[0.003]	[0.038]
Ethnic div. × Av. H index	0.091**	0.020	0.152**	0.147**	-0.001	0.106
	[0.039]	[0.057]	[0.065]	[0.067]	[0.004]	[0.066]
Mean Dep. Var.	0.102	0.823	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
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. 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 outcomes refer to the share of the population employed in the tertiary (1-2), primary (3-4), and secondary (5-6) sectors, the share of the population who can read and/or write (7-8), the share of the population employed in local trade (9-10), and the share of the population employed in other services of the tertiary sector (11-12). The outcomes for the period 2007-2017 refer to the average from the 2007 and 2017 population censuses. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 14: Mechanisms: Cultural Transmission or Economic Complementarities?

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

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

Appendix

Within-Group Heterogeneity in a Multi-Ethnic Society

Miriam Artiles

PUC-Chile

May 7, 2023

Contents

A Appendix - Historical Setting

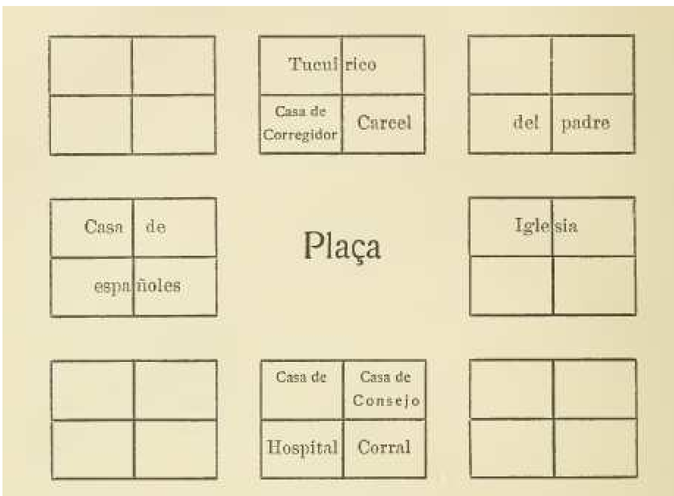
A.1 Contemporary images of *Chanka* identity. The left figure shows a *Chanka* message from the 2002 election claiming "Somos *Chankas*" (translated as "We are *Chankas*"). The figure on the right shows a *Chanka* festivity in 2003 honouring their connection to the puma. The two images are taken from Bauer, Kellett, and Silva (2010).

A.2 Anecdotal evidence: the Spanish administration in a mountain environment. The Andean environment was different from what Spanish observers of the 16th century had seen before (p. 55, Murra 2002b). Interestingly, as noted by Pulgar Vidal (1941, 2012), Spanish chronicles and colonial documents employed morphological terms used in 16th-century Spain to describe the geography of the Peruvian territory. In particular, they distinguished between three major regions. The word *coast* was used for flat territories with direct access to the ocean, *sierra* for the mountainous territory of the Andes, and *jungle* for the Amazon rainforest. These terms have persisted over time to describe the Peruvian territory in a broad way. However, Pulgar Vidal's studies, which began by analyzing indigenous knowledge of geography embedded in native folklore, support that neither these terms were used by native populations to describe the territory nor a concept of territory divided into three main regions existed before the Spanish colonization. Pulgar Vidal (1941, 2012)'s work distinguishes five natural resource zones in my study region (see Section 3.3.1). Historical studies also suggest that colonial officials were not fully aware of native practices at the time of the intervention. For example, it was common for native populations to use different combinations of crop rotation and fallowing in order to increase soil productivity. A common source of conflict with native populations appeared when Spanish officials found uncultivated lands−they generally thought the lands were abandoned when, in fact, they were in a fallow period (Pease 1989).

A.3 The census of 1791-95. Parishes are displayed by administrative region (*intendencia*) and province (*partido*). After the Bourbon reforms of 1784-1785, the viceroyalty was divided into *intendencias*, and *intendencias* were, in turn, divided into *partidos*. The census covers the territory under the Viceroyalty of Peru, thus excluding parishes in the *intendencia* of Puno. Puno was under the jurisdiction of the Audiencia of Charcas (modern Bolivia), in the Viceroyalty of Río de la Plata, until 1795 (*Real Cédula* of February 1, 1796); see Lynch (1962, p. 67-68) for more details. A summary of the census was published as an appendix to Manuel Fuentes' *Memorias de los virreyes que han gobernado el Peru* (1859, vol. 6, p. 6-9). The document was signed by José Ignacio de Lequanda and dated January 10, 1796. The whole census with figures at the parish level was later published in Vollmer (1967), where it is referred to as "Census of 1792." The census is considered a baseline for the study of population before independence from Spain (Gootenberg 1991).

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

A.4 Contemporary aerial views of Yanque. Subfigure (a) shows the model of *reducción* designed in 1567 by Matienzo (1910)[1567]. Subfigures (b) and (c) show contemporary aerial views of Yanque, created as a result of the resettlement policy in the 16th century (*Servicio Aerofotográfico Nacional del Perú*, in Medina (1993), and Google Earth imagery, respectively).



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

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

(c) Aaerial view of Yanque in 2019 (Google Earth)

B Appendix - Figures and Tables

B.1 Summary Statistics and Exploratory Analyses

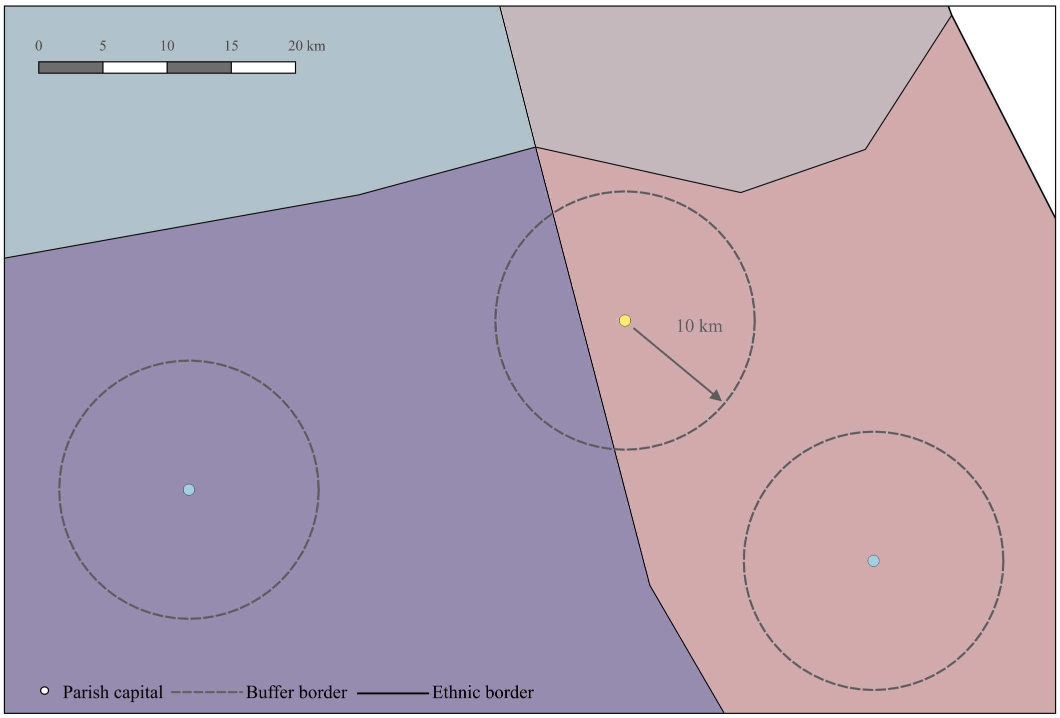


FIGURE B.1: Buffer Exercise

Notes. Construction of 10-km buffer around each parish capital. The map is displayed using a World Geodetic System projection (WGS 1984).

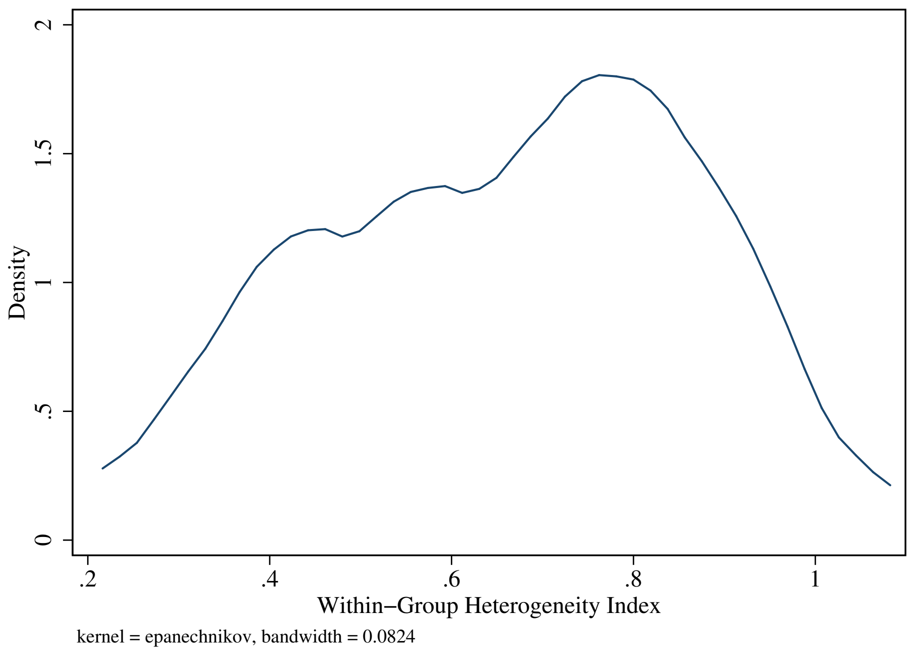


FIGURE B.2: Density of Within-Group Heterogeneity

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

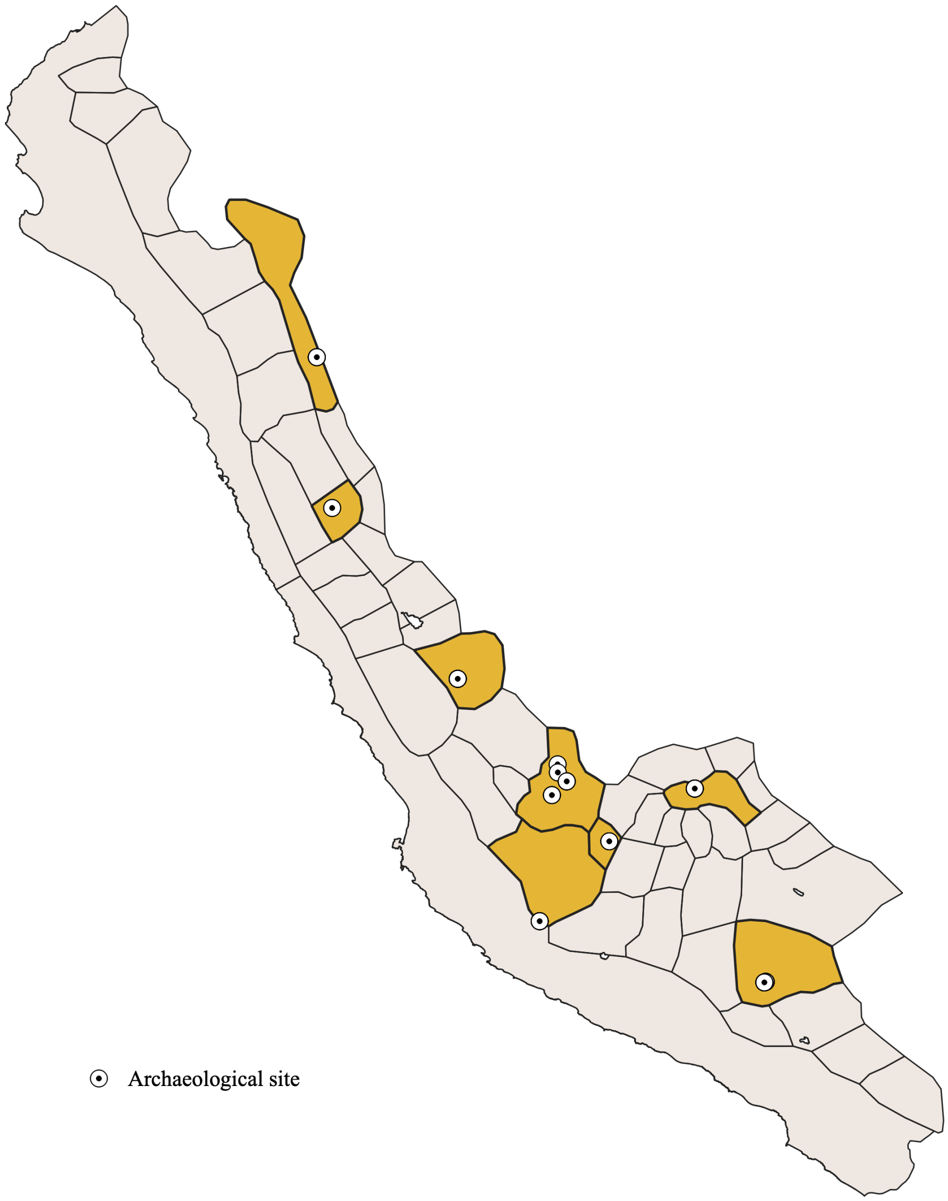


Figure B.3: Archaeological Sites with Information on Pre-Colonial Diets

Notes. The map shows the location of archaeological sites with information on pre-colonial individual diets. Geographic coordinates are from Wilson et al. (2022). Lines in black represent the approximate extent of ethnic groups according to Rowe (1946)'s map. The map is displayed using a World Geodetic System projection (WGS 1984).

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

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

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

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

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

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

Table B.3: Validating Resource-Producing Zones – Ethnic Groups

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

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

Table B.4: 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)
Observations	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 B.5: Coverage of Wilson et al. (2022)'s Data on Pre-Colonial Diets

Ethnic Group	# Individuals	Zone
Chachapoya	10	Quechua
Collagua	45	Suni or Jalca
Cusco	59	Quechua
Huanca	40	Suni or Jalca
Pinco	3	Quechua
Rucana	16	Quechua
Sora	4	Suni or Jalca
Vilcas	19	Quechua
Total	196	

Notes. Individuals are assigned to ethnic groups using the geographic coordinates of the archaeological site where the individual's remains were found and Rowe (1946)'s ethnic boundaries. Geographic coordinates are from Wilson et al. (2022)'s data.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.6: Within-Group Heterogeneity and Pre-Colonial Diets – Dummy Dependent Variable

	Dependent Variable: Dummy for Consumption of C_4 Plants ($\delta^{13}C_{col} \geq -14$)					
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: All Individuals			
H index (ethnic group level)	0.234	0.139	0.343	0.129	0.205	
	[0.062]***	[0.030]***	[0.056]***	[0.024]***	[0.091]*	
	(0.058)***	(0.033)***	(0.031)***	(0.024)***	(0.083)**	
H index (grid cell level)						0.013
						[0.138]
						(0.081)
Number of individuals	196	196	196	196	196	196
			Panel B: Excluding Children			
H index (ethnic group level)	0.241	0.135	0.338	0.112	0.180	
	[0.059]***	[0.033]***	[0.055]***	[0.026]***	[0.093]*	
	(0.055)***	(0.037)***	(0.026)***	(0.025)***	(0.085)**	
H index (grid cell level)						0.001
						[0.134]
						(0.074)
Number of individuals	182	182	182	182	182	182
Zone FE	No	Yes	Yes	Yes	Yes	Yes
Ln (# Crops)	No	No	Yes	No	No	No
Land caloric suitability	No	No	No	Yes	Yes	No
Institutional and socioecon. controls	No	No	No	No	Yes	No
Number of ethnic groups	8	8	8	8	8	–
Number of grid cells	–	–	–	–	–	8

Notes. The unit of observation is the individual. The table reports OLS estimates. Standard errors clustered at the ethnic group 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 (# Crops)* refers to the number of different native crops within the ethnic homeland. *Land caloric suitability* refers to the average caloric suitability of the ethnic homeland. The vector of institutional and socioeconomic controls includes ethic-level dummies for pre-colonial political complexity, urbanization, and elite residences. All regressions control for the longitude and latitude of the archaeological site where the individual's remains were found. All variables except dummies are standardized to have zero mean and standard deviation equal to one. The grid-cell level H index refers to grid cells of 50km × 50km. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.7: 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 B.8: H Index and Ethnic Diversity

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

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

Table B.9: 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 B.10: 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 D.

B.2 Robustness and Sensitivity Analyses

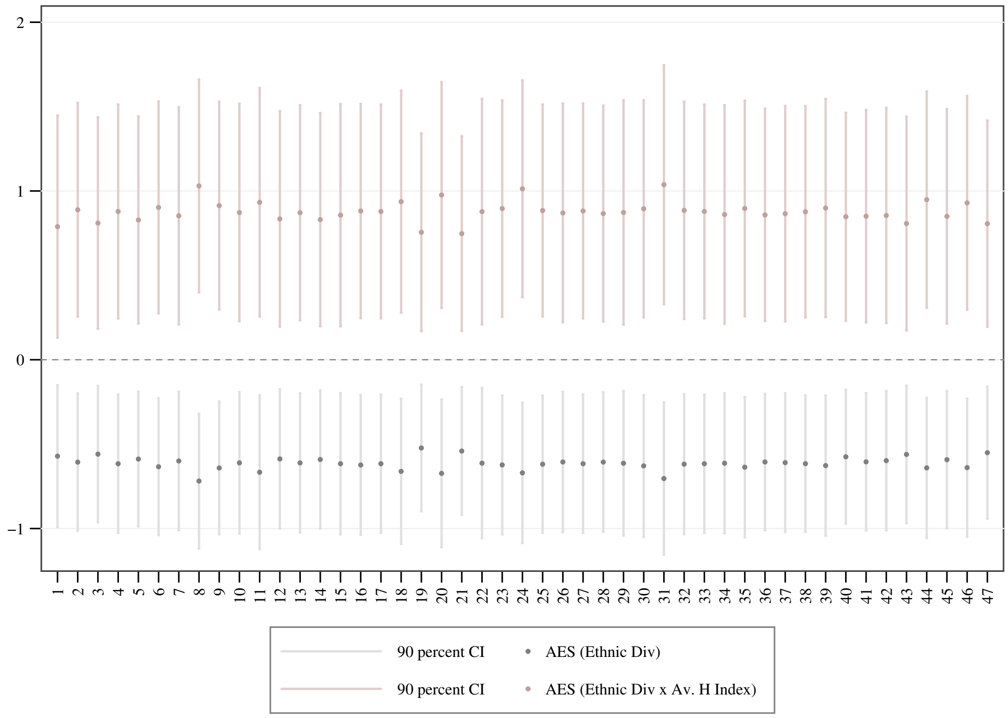


FIGURE B.4: Influential Observations – Excluding One Ethnic Group at a Time

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), an indicator for 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). Each regression excludes all parishes in which a given ethnic group was present (for one ethnic group at a time; the ethnic group is indicated on the x-axis).

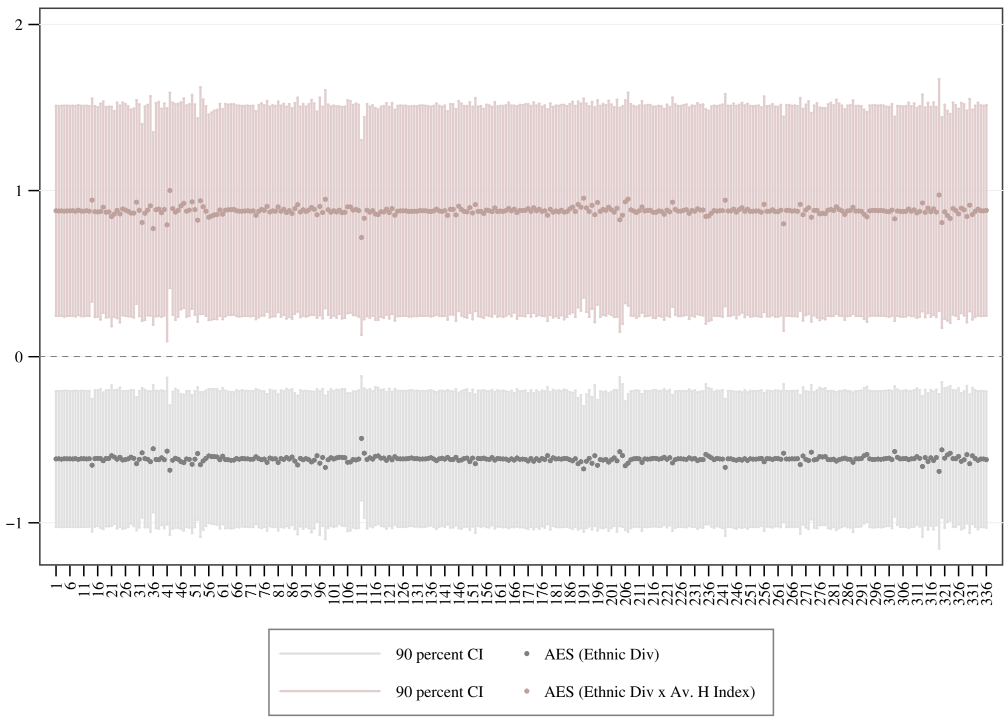


Figure B.5: Influential Observations – Excluding One Parish at A Time

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), an indicator for 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). Each regression excludes one parish (indicated on the x-axis) at a time.

Table B.11: Light Intensity per capita and Geographic Proximity to Ethnic Boundaries

	Dependent Variable: Light Intensity per capita (2000 – 2003)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic diversity	-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). Ethnic diversity takes value 1 if the grid cell is part of 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 B.12: 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 rural dummy	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Number of parishes	336	336	336	336	336	336	336	336	336	336	336	336

Notes. The unit of observation is the parish. 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) an indicator for 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 rural dummy variable takes vale 1 if the share of rural population is above the median, and 0 otherwise (1993 in Panel A and 2017 in Panel B). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.13: 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]
Observations	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]
Observations	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 8) in Panel A and the lasso-*selected* characteristics (Column 10 of Table 8) 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 B.14: 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), an indicator for 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 B.15: Robustness: Using Fractionalization to Measure Ethnic Diversity (Ethnic $\text{frac}_p = 1 - \sum_e w_{pe}^2$)

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

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

Table B.16: Robustness: Using a Herfindahl Index to Measure Within-Group Heterogeneity ($\widetilde{H}_e = 1 - \sum_j s_{ej}^2$)

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

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

B.2.1 Household Consumption and Selective Migration

I explore the extent to which post-resettlement selective migration might help explain the results. Although population movement was restricted after the resettlement, individuals may have managed to escape. Selective migration would be a potential channel of persistence if the most *capable* individuals from poorer parishes managed to escape to better-off ones, and if the characteristics they brought with them were (*i*) heritable and (*ii*) relevant for economic development (Dell 2010; Lowes and Montero 2021).

Two scenarios are worth analyzing. First, the most *capable* individuals 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, would help explain the negative coefficient on ethnic diversity.[64] However, it would not help explain the positive coefficient on the interaction term. Another possibility is that ethnically diverse parishes with individuals from more heterogeneous ethnic groups (type A parishes) received the most *capable* individuals from more disadvantaged parishes (those with ethnic diversity but low within-group heterogeneity; type B parishes). This second scenario would help explain the positive sign. Note, however, that the average distance between type A and B parishes is 51 km (defining parishes with high within-group heterogeneity as those with $\overline{H}_p$ above the 75th percentile), which may have made migration difficult.

In Tables B.17 and B.18, I examine whether the different types of selective migration can fully explain the results. For this exercise, I use survey data on household consumption from the Peruvian Institute of Statistics (INEI). The outcome variable is the log of real household consumption per capita. I use the spatial deflators provided by INEI 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.

[64] The 1993 population census provides information on the share of individuals who were not born in the district. In my sample, these individuals represent only 0.008 percent of the district's population, on average. Furthermore, I find no statistically significant differences between ethnically diverse and nondiverse parishes in the share of nonlocal population. However, in the 2007 census, this share was 2.2 percent higher in parishes with an ethnically homogeneous founding population. For reference, Dell (2010) finds that in-migration was 4.8 percent higher in non*mita* districts (using 1993 census data).

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 located in parishes without ethnic diversity (Table B.17) or in parishes with ethnic diversity and high $\overline{H}_p$ (Table B.18) for different bandwidths (x). The surveys cover 219 parishes, of which 82 are parishes with ethnic diversity. Columns 1 and 2 show that the main result of the paper holds for household consumption. The results are qualitatively similar when either type of selective migration is assumed. For the statistical significance of the interaction term to decline, a high rate of the second type of selective migration (e.g., 8 or 10 percent) has to be assumed.

Table B.17: Ethnic Diversity, Within-Group Heterogeneity, and Household Consumption (I)

	Dependent Variable: Log Real Household Consumption Per Capita						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ethnic diversity	-1.264	-0.515	-0.457	-0.416	-0.407	-0.390	-0.372
	[0.434]***	[0.238]**	[0.224]**	[0.219]*	[0.218]*	[0.217]*	[0.215]*
	(0.454)***	(0.195)***	(0.204)**	(0.211)**	(0.214)*	(0.218)*	(0.217)*
Ethnic div. × Av. H index	1.604	0.748	0.701	0.668	0.677	0.671	0.663
	[0.609]***	[0.351]**	[0.336]**	[0.331]**	[0.330]**	[0.329]**	[0.328]**
	(0.616)***	(0.339)**	(0.358)**	(0.364)*	(0.370)*	(0.374)*	(0.372)*
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	10,792	10,632	10,473	10,314	10,155
% Trimmed	–	–	2.2 %	4.2 %	6.2 %	8.2 %	10.2 %

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 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 mother tongue 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 B.18: Ethnic Diversity, Within-Group Heterogeneity, and Household Consumption (II)

	Dependent Variable: Log Real Household Consumption Per Capita						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ethnic diversity	-1.264	-0.515	-0.484	-0.457	-0.434	-0.414	-0.391
	[0.434]***	[0.238]**	[0.237]**	[0.237]*	[0.237]*	[0.238]*	[0.237]
	(0.454)***	(0.195)***	(0.193)**	(0.193)**	(0.193)**	(0.194)**	(0.199)**
Ethnic div. × Av. H index	1.604	0.748	0.688	0.638	0.595	0.556	0.516
	[0.609]***	[0.351]**	[0.348]**	[0.347]*	[0.348]*	[0.349]	[0.347]
	(0.616)***	(0.339)**	(0.332)**	(0.331)*	(0.329)*	(0.330)*	(0.335)
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	10,944	10,923	10,902	10,881	10,860
% Trimmed	–	–	2.2 %	4.2 %	6.2 %	8.2 %	10.2 %

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 located in parishes with ethnic diversity and average within-group heterogeneity above the 75th percentile (see % trimmed in the last row). The vector of individual-level controls includes gender, age, age squared, years of schooling, civil status, and mother tongue 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 B.19: 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), an indicator for 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 6 includes fixed effects for the four major Inca regions (*suyus*) according to the mapping of Zuidema and Poole (1982). Column 7 excludes parishes that concentrated groups potentially affected by Inca resettlements according to Rowe (1946). Nonetheless, the fact that the study focuses on parishes located in the highland region alleviates concerns regarding potential Inca resettlements from the north to the south coast of Peru (Bongers et al. 2020). 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 B.20: Robustness: Post-Resettlement Variables Related to the Structure of the Population and Religion

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

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. 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), an indicator for 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.3 Mechanisms: Supporting Evidence

Table B.21: Within-Group Heterogeneity and Inter-Group Unions (1605−1870)

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

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

Table B.22: Non-Subsistence Agriculture Beyond 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 (# Crops)		0.047			0.001	
		[0.060]			[0.085]	
		(0.053)			(0.072)	
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 an indicator for 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). *Ln (# Crops)* and *Frac. Crops* refer to native crops only. 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 B.23: Mechanisms: Technology Adoption in Agriculture and Attitudes Toward Technical Support

	Dependent Variable:				
	Dummy "Technical Support is Necessary" (1994)	Technology Adoption in Agriculture			
		Dummy Insecticides (1994)	Dummy Improved Seeds (1994)	Dummy Bio. Control (1994)	Dummy Chemical Fertilizer (1994)
	(1)	(2)	(3)	(4)	(5)
Ethnic diversity	-0.338	-0.258	-0.368	-0.118	-0.021
	[0.185]*	[0.221]	[0.201]*	[0.205]	[0.231]
	(0.078)***	(0.175)	(0.185)**	(0.195)	(0.225)
Ethnic div. × Av. H index	0.479	0.339	0.458	0.203	-0.031
	[0.286]	[0.342]	[0.315]	[0.315]	[0.376]
	(0.124)***	(0.229)	(0.273)*	(0.357)	(0.346)
Baseline controls	Yes	Yes	Yes	Yes	Yes
Colonial province FE	Yes	Yes	Yes	Yes	Yes
Number of parishes	334	334	334	334	334
Mean Dep. Var.	0.500	0.500	0.500	0.500	0.500

Notes. The unit of observation is the parish. Robust standard errors clustered at the level of the colonial province in brackets. In parentheses, standard errors are corrected for spatial dependence with a distance cutoff of approximately one degree at the equator (Colella et al. 2019). Data from the 1994 agricultural census on whether farmers consider that technical support is necessary in order to improve agricultural productivity, use of insecticides, improved seeds, fertilizers, and knowledge of biological control. The outcome variables are dummies for whether the share of farmers is above the median. The median values for each variable are: technical support in necessary ($P_{50} = 0.828$), insecticides ($P_{50} = 0.330$), improved seeds ($P_{50} = 0.086$), biological control ($P_{50} = 0.034$), and chemical fertilizer ($P_{50} = 0.261$). All regressions control for the log of the total number of farmers. 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 B.24: Mechanisms: Structural Transformation – Evidence from Pre-Colonial Crop Fractionalization

	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
	Can Read and/or Write		Tertiary Sector: Local Trade		Tertiary Sector: Other Services	
	(1876)	(2007-2017)	(1876)	(2007-2017)	(1876)	(2007-2017)
	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic diversity	0.090	-0.071	-0.033	-0.100	0.010	-0.041
	[0.062]	[0.057]	[0.087]	[0.079]	[0.007]	[0.111]
Ethnic div. × Av. Crop Frac.	-0.202	0.124	0.061	0.185	-0.024*	0.073
	[0.129]	[0.108]	[0.186]	[0.144]	[0.014]	[0.211]
Mean Dep. Var.	0.102	0.823	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
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, the share of the population who can read and/or write (7-8), the share of the population employed in local trade (9-10), and the share of the population employed in other services of the tertiary sector (11-12). The outcomes for the period 2007-2017 refer to the average from the 2007 and 2017 population censuses. 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. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C Appendix - Surnames

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

Colonial baptism records. The website FamilySearch.org provides access to baptism records from colonial Peru. The organization, which seeks to help trace users' ancestry, seeks volunteers from around the world to make indexed genealogical records freely available. The results in Table 1 use information from the collection "Perú, bautismos, 1556-1930."[65] Each baptism record includes information on the full name and gender of the individual, name of the parish, and date of baptism. The original handwritten record has also been uploaded in some cases and can be easily accessed.

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

URL

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

Información indexada

Ocultar

Imagen no disponible

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

Cita

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

Identification of native surnames. I first excluded Hispanic and foreign surnames from the analysis. The main source for the identification of Hispanic surnames is Platt (1996), which includes an index of Hispanic surnames developed in Latin America and the United States. The author writes "the word Hispanic refers to individuals born in Latin America or the United States, whose parents speak Spanish and whose principal cultural background was Spanish." This source includes the list of surnames in Carraffa and Carraffa (1920–1963), the traditional reference for Hispanic surnames.[66] I complement Basque surnames using a list provided by the Real Academia de la Lengua Vasca.

In order to identify native surnames, I then constructed a dictionary of linguistic roots from the Quechuan and Aymaran language families. There is no unique source for the identification of surnames from these families. The transformation of native surnames over time (*castellanización*), as well as the presence of many regional varieties of Quechua and Aymara, make necessary the combination of different (temporal and regional) sources. For Quechua, the main sources are the classic dictionary by González Holguín (1952)[1608] and a recent dictionary compiled by the Academia Mayor de la Lengua Quechua (2005). I

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

also include the list of names provided by the Peruvian *Registro Nacional de Identificación y Estado Civil* (RENIEC 2012). For Aymara, the main sources are the classic dictionary by Bertonio (2011)[1612], the list of surnames provided by De Lucca (1983), and a recent dictionary compiled by CONADI (2011). I complement the analysis using two additional sources: (1) *Vocabulario Políglota Incaico*, originally compiled by Franciscan missionaries in Peru, which provides an extensive list of words in four dialects of Quechua (varieties of Cuzco, Ayacucho, Junín and Ancash) and Aymara, see Fide (1998)[1905]; and (2) the *An Crúbadán-Corpus Building for Minority Languages* project, which provides downloadable text datasets for different dialects of Quechua and Aymara based on online text resources, including translations of the Bible and the Universal Declaration of Human Rights.

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

Descriptive Statistics - Dataset of Baptisms

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

D 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). It provides 30 arc-second raster data with median elevation based on information from the NASA Shuttle Radar Topographic Mission (SRTM).

Variation in elevation. Standard deviation of elevation across all grid cells with centroid within a 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), which provides information on potential crop yield given the set of available crops before 1500CE (5 arc-minute raster data).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Ln light intensity per capita. Natural log of 1 plus average light intensity per capita. The average sum of light intensity values across all grid cells with centroid within the 10-km buffer is divided by total population within the same buffer. Source: average cloud free coverages of the DMSP-OLS Nighttime Lights Time Series, produced by the NOAA's National Geophysical Data Center, which provide 30 arc-second yearly raster data. Data from satellites F15 and

F18 for the periods 2000-2003 and 2010-2013, respectively (yearly averages from the same satellite). Version 4.10 of the Gridded Population of the World (Center for International Earth Science Information Network−CIESIN) provides 30 arc-second raster data with population counts for the years 2000 and 2010. Population counts are developed through the uniform areal-weighting method using census data adjusted to match the United Nation's population counts at the country level. Period: contemporary.

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

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

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

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

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

Share of pop. by sector of economic activity. Share of population employed in the primary, secondary, and tertiary sectors. Source: author's coding using data on occupations from the 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878). 21th-century data come from the 2007 and 2017 population and housing censuses, conducted by the National Institute of Statistics (INEI). Period: post-independence.

Literacy rate. Share of the population who can read and/or write. Source: 1876 population census (*Censo General de la República del Perú formado en 1876*, published: Lima, 1878). 21th-century data come from the 2007 and 2017 population and housing censuses, conducted by the National Institute of Statistics (INEI). Period: post-independence.

Dummy neighborhood association. Dummy variable taking value 1 for the presence of neighborhood associations, and 0 otherwise. Source: 2002-2003 *Registro Nacional de Municipalidades*, provided by the National Institute of Statistics (INEI). Period: contemporary.

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

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

Ln volunteers for military service. Natural log of 1 plus the average number of volunteers for military service between 2008 and 2014. Source: administrative data provided by the Peruvian *Ministerio de Defensa*. Period: contemporary.

Dummy volunteers for military service. Dummy variable taking value 1 if there is at least one volunteer between 2008 and 2014, and zero otherwise. Source: administrative data provided by the Peruvian *Ministerio de Defensa*. Period: contemporary.

Dummy "technical support is necessary". Dummy variable taking value 1 if the share of agricultural producers reporting that "technical support is necessary in order to improve agricultural productivity" is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

Dummies technology adoption in agriculture. Separate dummy variables for whether the share of agricultural producers reporting to use insecticides, improved seeds, or chemical fertilizers, or to know about biological control, is above the median, and 0 otherwise. Source: 1994 national agricultural census, conducted by the National Institute of Statistics (INEI). Period: contemporary.

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

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, conducted by the National Institute of Statistics (INEI). The question was not repeated for other presidential elections. Period: contemporary.

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 identification with a state administrative unit, ethnicity or race, and religion). Source: 2004-2017 ENAHO surveys, conducted by the National Institute of Statistics (INEI). Period: contemporary.

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 data on the first records (1532–1575) of tributary population from Cook (1982, 2010). I consider all population centers within the ethnic homeland.

Ln population density [ethnic level]. Natural log of 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, and 0 otherwise. Source: author's computation using Rowe (1946)'s ethnic boundaries and information on pre-colonial archaeological sites in Ravines Sánchez (1985), Ramos Giraldo (2001), Isbell and Silverman (2002a, 2008), and the inventory of pre-colonial archaeological sites (*Catastro de Monumentos Arqueológicos Prehispánicos*) developed by SIGDA (Sistema de Información Geográfica de Arqueología, 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, as defined in Stanish 2001. Source: author's computation using Rowe (1946)'s ethnic boundaries and archaeological records; for the main sources of information on pre-colonial archaeological sites see *Dummy urbanization*.

Dummy elite residences [ethnic level]. Dummy variable taking value 1 for the presence of elite residences, and 0 otherwise. Source: author's computation using Rowe (1946)'s ethnic boundaries and archaeological records; for the main sources of information on pre-colonial archaeological sites see *Dummy urbanization*.

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, and otherwise. Source: author's computation using Rowe (1946)'s ethnic boundaries and archaeological records; for the main sources of information on pre-colonial archaeological sites see *Dummy urbanization*.

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 map of the Inca road network (Qhapaq Ñan) produced by SIGDA (Sistema de Información Geográfica de Arqueología, Ministerio de Cultura, Perú), accessed in March 2021.