Colonization, Early Settlers and Development: The Case of Latin America

José G. Montalvo
Marta Reynal-Querol

This versión: June 2021

(July 2020)

Barcelona GSE Working Paper Series

Working Paper nº 1189

Colonization, Early Settlers and Development: The Case of Latin America*

José G. Montalvo

Universitat Pompeu Fabra-IPEG and Barcelona GSE

Marta Reynal-Querol

Universitat Pompeu Fabra-ICREA-IPEG and Barcelona GSE

Abstract

In this paper, we document the long-run impact of geographical heterogeneity in skills among the first settlers to Latin America. To this end, we compile administrative data on the early settlers in the Americas between 1492 and 1540, including among other characteristics, name, city of origin, destination, and occupation. From a methodological perspective, a focus on the initial period of colonization in Latin America offers several advantages. First, differences in the geographical distribution of occupations among the first settlers are likely to be accidental. Second, a setup that analyzes an area with a single colonizer (Spain) allows us to hold constant formal institutions and legal origin. Our results show a relevant effect of the skills of initial colonizers on long-run levels of development of the areas located around the original settlements. We find evidence of persistence in the form of market orientation and entrepreneurial spirit.

Keywords: Skills, Early Settlers, Persistence, Development.

*We are grateful for the comments and suggestions of Antonio Ciccone, Claudio Ferraz, Ernesto Dal Bo, Ruben Durante, Federico Finnan, Stelios Michalopoulos, Elias Papaioannou, Hillel Rapoport, Joachim Voth, David Yanagizawa-Drott, and especially to Oded Galor. We also thank participants in the Lacea-Ridge Workshop,

that of the International Economic Association, the PEDD International Conference, Fundación Ramon Areces -VIII Workshop NCID, IEA international Conference, 1st PSE-Barcelona Workshop on the economics of immigration and diversity, and numerous university seminars which include, Bocconi, Brown, Gothenburg, Kings College, Paris School of Economics, Paris Science Po, Universidad de los Andes, and University of Zurich. We benefited from exceptional research assistance on the part of Miriam Artiles, Maria del Mar Garcia Roldan, Juan Carlos Muñoz-Mora, Andreu Reixach, and Felipe Valencia. Financial support from the European Research Council under the European Community ERC Grant Agreement n.647514, the Spanish National Science Foundation AEI/FEDER, UE ECO2017-82696-P, and the Government of Catalonia (ICREA) is gratefully acknowledged. We are also thankful for the financial support provided by the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2015-0563).

> "*It is a known fact that the first to settle in a colony determine the pattern of the future society, while the later settlers adapt to the prevailing circumstances, even if they come from different regions.*"

Bernard Slicher Van Bath (1979), Latin America Around 1600 (p 187).

1 Introduction

Much recent work on comparative development investigates the relevance of large historical events. There has been a particular interest in colonialism and its persistent impact on the long-run development of affected countries.[1]

However, as argued by Michalopoulos & Papaioannou (2020), the colonial experience better resembles a "bundle treatment" than a single treatment. With regard to Africa, for example, the identity of colonizers is not common to all countries. This makes it difficult to disentangle the role of differences in institutional setup or legal origin from the influence of culture, human capital, and skills.

This paper analyzes the Spanish colonization of Latin America and, in particular, the role of heterogeneity in skills among the first settlers. Our choice of colonization area offers several crucial advantages. First, since all the countries were originally settled by the same colonial power, the findings are less subject to the bundle treatment criticism described above. Additionally, since we analyze the first travelers from Spain to the Americas, it is reasonable to assume that the geographical distribution of their occupations was as good as randomly assigned given that the colonizers were settling in previously unexplored, unknown areas. We show that settlers' skills were, in fact, very heterogeneous across locations. Importantly, this initial shock to the distribution of skills persists over time. Our results show a lasting legacy of the initial distribution of occupations: areas where the first settlers had a higher level of skills are currently more developed than those that were settled by low-skilled settlers. To the best of our knowledge, this is the first time that information on the skills of the first colonizers in the Americas has been used for economic research.

Moreover, while previous related research has used various statistical techniques (i.e., regression discontinuity, difference-in-differences, and instrumental variables) to

[1] For a recent review of the literature on the historical roots of economic development, see Nunn (2020).

disentangle the sources of differential development, the characteristics of our data generate a natural experiment that requires a minimal set of conditions.

Specifically, we exploit several particular features of the settlement process from Spain to the New World. First, for many years after the discovery of the Americas, there existed just one legal point of departure (Seville), which simplified the task of recording passenger lists. Second, the Spanish Crown imposed strict restrictions on travel for some groups (e.g., Jews, Muslims,...). Therefore, those who wanted to travel to the Americas had to show they did not belong to the banned population. Documents for this purpose along with passenger lists are preserved in the General Archive of the Indies (Seville) and, taken together, provide a detailed picture of the approximately 20,000 settlers in the Americas during the first period of colonization (1492-1540). Gathering this information required extensive archival work and the digitalization of long lists of travelers and settlers[2], and many old maps. The data, which include information on settlers' name, city of origin, destination, and occupation, show that, contrary to common belief, most of the settlers were not soldiers. That being said, the fundamentals of colonization imply that conquest and settlement were part of the same process.

Our empirical strategy is based on the fact that the initial colonizers had no knowledge of the land they intended to conquer. In particular, differences in the occupational structure of first settlers across settlements are likely to be accidental. We will provide historical evidence that supports this hypothesis. This is the initial phase of discovery, colonization and settlement, from 1492 to 1540, that took place simultaneously in the case of Latin American Spanish colonies. The historical evidence comes from narratives of the time, numerous documents of "capitulaciones" (contracts that conquerors sign with the King), and from the work of historians. This context therefore offers a quasirandom situation: given that the discovered territory had not previously been explored, their settling across it was as good as random. Moreover, specifically, we show that the geographical distribution of high- and low-skilled coloniz-

[2]The lists of travelers include some information on the origin of the migrants, while the list of settlers provides some additional information on the location of settlement and, more important, the fact that they arrived at the destination. In those years, mortality in transatlantic travel was quite high. Since we use only migrants that arrived in the Americas, we primarily refer to them as settlers.

ers was very similar regarding the pre-existing characteristics of the settlement areas, supporting the exogeneity assumption relative to the location of the first settlers. In addition, since the entire area was controlled by the same colonial power, we are able to hold constant both formal institutions and legal origin. Importantly, this pseudo-experiment shows that locations originally characterized by a greater proportion of high-skilled settlers today have a higher level of development than locations colonized by a low-skilled majority. It is important to emphasize that we are working with the first settlers. This means that they could not be attracted by the type of Spanish people who were already there because no Spanish settlers were present before the individuals we use in our analysis. This is important to bear in mind to avoid any confusion of the context. What these early settlers found in the new land was highly diverse climate, geography and indigenous groups. Collecting information to capture all these conditions required intense effort.

We also study the mechanism that could explain the long-term effect of occupations on differential development. In particular, we find evidence of persistence in skill level in the form of market orientation, agricultural technological transformation, infrastructure, and entrepreneurial spirit.[3] We use data from a period well after the first colonization wave (1573-1620) and contemporaneous data to provide evidence of this mechanism. We collected information on the economic activity in these early places in 1574 and 1620 to analyze whether the high-skilled places were already more market oriented a few years after their initial settlement.

We also discuss some threats to inference. First, we show that the results are robust to geographic, climatic, or precolonial characteristics, as well as to controlling for different colonization periods, the size of the colonization cohort, colonizers' route order, or penetration line. Second, although the legal origin and institutions of the colonizer, Spain, were the same in all of the areas, we consider whether the original official jurisdictional division and the time until their formal promulgation had a relevant impact on the results. Our estimations indicate that these jurisdictional divisions have no effect on the results. We also check that our findings are robust to alternative measures of skills and sizes of the settlement catchment areas.

Our results contribute to several lines of research. First, there is a growing

[3] Dell (2010) similarly shows that the persistent negative effect of Peru's mining "mita" on current incomes is due to less-developed infrastructure and weaker production in the market economy in areas without historical hacienda formation.

literature on the persistent effect of skills and human capital on economic development[4]. However, very few papers explore this issue in Latin America. Valencia (2019) finds that the higher educational attainment of some Guarani people, due to the presence of Jesuit missions, persisted 250 years later in areas of former Jesuit presence. Higher noncognitive abilities are consistent with occupational persistence and intergenerational knowledge transmission.[5] Rocha et al. (2017) show that an initial shock in the skill composition of migrants to some municipalities in Brazil in the late 19th and early 20th centuries persisted over time, leading to higher long-run income per capita. We similarly study the long-run effect of the distribution of skilled settlers to Latin America but reach back to the initial period of colonization. Like Rocha et al. (2017), who examine a single state in Brazil, we hold constant institutions by assessing an area under the same colonizer[6]. In our case, the issue of de jure institutions is arguably not an important caveat. Nonetheless, we demonstrate the robustness of our results to informal institutional arrangements that may have been at work at such an early stage of colonization.

Second, our paper builds on studies that distinguish among different types of human capital or abilities. Squicciarini & Voigtländer (2015), for example, find that it is the upper tail of human capital and not the average that is critical in the transition from stagnation to growth. Moreover, Gennaioli et al. (2013) emphasize the importance of entrepreneurial human capital in regional development.

Third, our research also relates to work on the effect of long-term population history on economic development. In this regard, there is a growing body of work that studies the impact of changes in the composition of population and migration on long-run economic performance (e.g., Spolaore & Wacziarg (2013) or Droller (2018)). Putterman & Weil (2010) examine the historical legacy of populations, setting the year 1500 as a reference date to calculate the share of contemporaneous population in each nation descended from people of different countries of origin. They find that

[4]See Michalopoulos & Papaioannou (2020) and Michalopoulos & Papaioannou (2017) for a detailed discussion of this body of work.

[5]In the case of Africa, Nunn (2011) also finds a persistent effect of missionary schools on educational attainment in various countries.

[6]In fact, during the initial period, there were no local formal institutions. Wantchekon et al. (2015) find that even in the absence of prior institutions, human capital has a large impact on economic development in colonial Dahomey (Benin).

the history of the ancestor population matters more than the history of the location. These results indicate that populations who settled in the Americas bore certain traits that made economic development more or less likely.

In a recent work, Easterly & Levine (2016) emphasize the importance of directly measuring colonial European settlements in the early stages of colonization.

Fourth, our results intersect with a relatively new literature on the very long-run impact of culture, occupations, and skills. Several studies have found that intergenerational mobility actually works much more slowly than predicted by traditional models. Barone & Mocetti (2016) provide evidence of persistence in belonging to elite professions after nearly six centuries. Opper & Andersson (2019) find that the entrepreneurial culture of the Ming Dynasty (1368-1644) persisted into the private activities of postreform China (1992-2012).[7] In this paper, we document the very long-run impact of a high proportion of high skilled professionals through the persistence of pro-market activities and entrepreneurial preferences.

Finally, as mentioned above, much of the related literature focuses on colonial shocks in Africa. Our paper joins a recent body of work that focuses on historical colonial experiences in Latin America, such as Valencia (2019).

2 The mechanics of conquest and settlement

The original settlement of Spaniards in the Americas was an organized endeavor, where private enterprise and legal coverage of the Spanish Crown generated a sophisticated set of incentives. The Spanish conquest was carried out by *huestes*, or expeditions, commonly known as *hueste conquistadoras* (conquest expeditions) or *hueste indianas* (Indian expeditions). The *huestes* consisted of a leader (traditionally called a *conquistador* or conqueror) and other individuals who, voluntarily and without payment, joined the group. Under the leader's protection, they undertook missions of discovery, conquest, and settlement. A *hueste* was a private endeavor; from the outset,

[7] Considering a shorter period, Alesina et al. (2020) document how the Chinese Communist Revolution failed to stop intergenerational transmission. Indeed, the grandchildren of prerevolution elites differ in cultural values, being more pro-market and individualistic. Cultural traits thus seem to overcome even the strongest attempts to subdue them.

the Crown made clear that it would not finance these groups. The expeditions did, however, have to share the benefits with the Crown, as they operated under a license granted by the latter to the leader. The Alexandrine Bulls, in particular the first bull, "Inter Caetera," recognized the Castilian Crown's claim to any discovered lands not already held by a Christian prince. In practical terms, this meant that no one could enter the newly discovered land without authorization from the King of Spain.

In fact, formally, *huestes* hoped of obtaining crown favors (land, forced labor, bounties, tax breaks and even improvement in the social status or titles of nobility). They were organized and structured according to a militia criterion, which must be self-sufficient to survive in unknown territories.[8]

Normally a *hueste* started when an individual asked permission to create it to go on an expedition to the Americas. The Indian *hueste* was an essentially voluntary expedition organized by a caudillo who had royal authorization to carry out a task of conquest, discovery, settlement, exploitation and evangelization in the Americas. The antecedent of this type of expedition is in Spain at the time of the Reconquest, but it has as its most immediate origin those that were organized for the conquests of the Canary Islands, with the difference that being the Indian expeditions were naval and terrestrial. The *hueste de las Indias* was a private company that reported benefits to both contracting parties (the Crown of Castile and individuals), who invested to obtain some kind of *mercedes* in the new world (lands, indigenous people, booty, exemption from taxes and social advancement) and that which, taking advantage of the adventurous spirit of its subjects, was basically interested in the settlement of the territories of the Crown for which it granted the respective royal authorization. The Indian Host was generally formed by a contract entered into between the Crown and the respective leader called a "capitulation" in which the royal authorization was granted for the future expedition to Las Indias that this leader would organize at his "cost and ammunition." Obtaining the permission of yore, the caudillo raised the "Royal Banner" at his house or in Seville and was made to tour the city with a drummer who also announced loudly that an expedition was going to be carried out and that those interested should register at the caudillo's house. The registration indicated whether the person who was registered made a contribution to the expedition. This was important because the grants were distributed pro rata to the contributions, similar to a commercial company. The members of the host, despite not being military,

[8] Ballesteros (1985).

were subject to military jurisdiction for the duration of the expedition. When the caudillo received authorization from the Crown, he also received (within the mandate to correctly carry out the expedition) the mandate to be able to judge the expedition members according to the rules of military justice (including capital punishment). There are many anecdotes of the hosts that indicate the absence of knowledge about the territory and the randomness of the hosts' destinations. When the host arrived in the Indies and the expedition began, they always deployed people in front with machetes to make their way, cutting through the vegetation of the jungle. Sometimes they found paths that could have been made either by indigenous groups or animals. It was said that some group had followed the route of an animal and that they were no longer heard from. They also sometimes divided the host into two groups, those who went by river and those who went by land, to increase the chances of survival. Sometimes when they met indigenous groups and asked them for precious materials, the indigenous people responded by sending the *huestes* in the wrong direction to get rid of them. Additionally, during the expedition, the members of the *huestes* left clothes and objects taped to trees to indicate that a *hueste* had gone through this route and that they were alive. They did this because if many years passed and nothing was known about them, the king could send another *hueste* in the same direction, believing that they had died. Thus, conflicts in the zones were avoided.

The "Manual of Indian Law" provided legal guidance for Castilian expansion in America. From the outset, conquest, discovery, and settlement were all conceived as part of the same process. In fact, the individuals who composed the *huestes* were not soldiers, as has sometimes been assumed, but rather settlers who were originally merchants, doctors, peasants, etc. Furthermore, it was in the interests of the Crown that the already discovered and conquered territories not be dismantled by the migration of their inhabitants to other sites. In the words of the time, the land needed to be "ennobled" that is, filled with houses and inhabitants and its natural agricultural, mining, and livestock resources exploited. Until that occurred, no authorization was granted to migrate to other lands. Indeed, this explains why the actions of discovery, conquest and settlement were considered to be elements of the same endeavor. In this regard, Lockhart et al. (1976) provide an exemplary description:

> "Since conquest and settlement were one single ongoing process in Spanish America, we are a little reluctant to emphasize the distinction

between them....yet only in this way we can illustrate to what an extent the conquerors were acting like immigrants, businessman and settlers."

The locations of the settlements had to comply, following the indications of the Spanish administration, with the rules of Saint Thomas Aquinas, who listed the "indispensable" characteristics that a territory should have: access to potable water, proximity to construction material (wood), nearness to forests and cultivable land, etc. These recommendations applied to all the settlements.

Often, the initial resources of these expeditions were recorded in contracts, or licenses, called *capitulaciones*.

Evidence shows that the contributions made by the *hueste* participants were then considered when distributing the booty, whereby members were rewarded as a function of their inputs.

As evangelization was a fundamental objective of Castilian expansion to the Indies, the *huestes* did not accept non-Catholics. As a result, Muslims, Jews, heretics targeted by the Inquisition and their descendants, were excluded from these expeditions. Similarly prohibited were gypsies, married slaves without his wife and children, single women without a license, or married women without their husbands. This aspect is very relevant, as these restrictions were the reason for the Spanish administration's endeavor to record the name, origin, accompanying party, etc. of all the initial travelers to the Americas.

To summarize, early Spanish colonization of the Americas was mostly a process of urban occupation, where settlement and conquest went hand in hand. The location of the original sites was decided based on a common set of simple rules. Since these settlers were the first, they could not be attracted by the profile of Spanish settlers who were already there because no Spanish settler had previously been established there. They were the first who simultaneously discovered, conquered and settled. Finally, settlers were not typically soldiers but people of diverse occupational backgrounds.

3 Data

In this section, we describe the data. Gathering information on the first settlers to the Americas is possible thanks both to the fact that there was just one legal port of

departure and to the existence of restrictions on travel to the New World. To verify whether potential travelers met the criteria allowing for embarkation, the authorities required various documents, now preserved in the General Archive of the Indies in Seville.

Until 1668, the sole legal port of departure for ships going to *Las Indias* (the "Indies") was Seville. Beginning in 1492, the *Casa de la Contratación de las Indias* (House of Trade of the Indies; byname House of the Indies) compiled information on all passengers who traveled from Seville to the Indies as part of the commercial regulation of the American colonies. These files are archived in the abovementioned General Archive of the Indies, where the original information appears in two types of documents. The first comprises the so-called *informaciones y licencias*, or information and licenses, necessary to prove to the officers of the House of the Indies that they were not prohibited from traveling. As described in the previous section, Jews, Muslims, and those judged by the Inquisition were not permitted; only those of proven Christian origin could embark. Evidence in this regard included baptismal and marriage records containing demographic information not only on the passenger but also their accompanying party (wife and children). The preservation of such documents provides a rich source of information on these travelers. The second type of document consists of passenger lists, or the *Libro de asientos de pasajeros*. These files include the names of the passengers (double checked upon boarding the ship), the name of the vessel, and its captain. By merging the names of the passengers who appear in both types of documents, it is possible to construct registries that include the travelers' complete name, status as head of the family, marital status, wife and children traveling with him/her, town/city/place of origin in Spain, and destination city[9] in the Americas.[10]

[9]This element was not declared at embarkation but is possible to identify by searching the documents of the *Catalogo de los pasajeros a las indias* (Catalog of Passengers to the Indies).

[10]There is a general consensus among scholars of the colonial history of Latin America that these official lists, during the initial decades of colonization, captured most of the passengers. In the 18th century, a liberalization policy began to allow trade between colonies and many other Spanish ports. It is hence possible that an individual might have embarked on a ship to England and from there to the Spanish colonies aboard an illegal vessel transporting merchandise; at the time, non-Spanish ships were

Luis Rubio y Moreno, Deputy Director of the General Archive of the Indies, began the classification of these documents in 1915 and supervised work on the entries between 1492 and 1559. This endeavor was subsequently taken over by Cristobal Bermudez, who compiled all the information into three volumes.[11] Then, Peter Boyd-Bowman, after 30 years of laborious archival work, complemented the original compilation with his five volumes. He systematically organized all the data contained in the *Catalogo de los pasajeros a las indias* (Catalog of Passengers to the Indies), as well as added information from many other sources available in the General Archive, in an effort to enrich the personal information on the passengers, particularly relative to destination and occupation. With regard to destination, Boyd-Bowman utilized thousands of documents coming from the Americas (foundation letters of each village, contracts, payrolls, commercial agreements, royalties, *capitulaciones*, etc.) preserved in the General Archive. Note that these documents were not written by each individual but by the conqueror or the "escribano", or other member in the hueste designated by the conqueror.[12] There are documents from all villages, since it was in the interest of the settlement to be considered a village; otherwise, the king could not activate the privileges promised in the *capitulaciones*. The documents act as proof that the objective of the "hueste" has been achieved and mentioned who were there to obtain noble titles. Moreover, Boyd-Bowman was interested in the destination of the travelers and, therefore, considered only those passengers who matched with documentation from the Americas, so the final list comprises settlers who arrived in the New World.[13] If the documents included a reason for not arriving at the destination (drowned, executed, etc.) Boyd-Bowman reports the circumstances. Specifically, the first volume, Boyd-Bowman (1956), contains information from 1493 to 1519, while the second, Boyd-Bowman (1964), covers settlers from 1520 to 1540. These first two volumes thus

still prohibited in Castilian colonies. Captains may also have been occasionally bribed to allow a passenger to board without being legally registered.

[11]See Bermudez (1940-46). These volumes cover the period from 1509 to 1559. In the early 1980s, specialists of the General Archive published volumes IV to VII, which cover the period from 1560 to 1599.

[12]Therefore, there was no need to be literate to have one's name reported in one of those documents.

[13]Thus, explaining why the book is called *Primeros pobladores* (First Settlers) as opposed to first migrants or passengers.

include the first wave of settlers.[14]

Boyd-Bowman's work provides a particularly rich source of information in that he was able to match the names on the passenger list with many other documents. Through the use of more than 500 sources in addition to the catalog, he was able to add valuable data to the original registries. Boyd-Bowman also cleaned the data of double entries and completed information on the origin and destination of each settler. His work undeniably provides the best source of information on the origin, destination, occupation, and other background characteristics of the first settlers of the Americas.

Our final dataset thus contains information, classified by year of departure, on the name and surname of all the settlers, village of origin in Spain, initial destination in the Americas, occupation, and year and place of death. An example of the wealth of information provided for each settler is illustrated by the following example, taken from Boyd-Bowman's first volume:

3564. Gómez, Hernán, resident in Tui (province of Pontevedra), wife Francisca de Rubiales and sons Lucas and Juanico, destination Indies in 1512 (volume I page 595[15]), occupation court clerk, arrived in Santo Domingo in 1512, resident of Yaquimo (Santo Domingo) in 1514.

These detailed descriptions allow us not only to geographically locate the first settlers of the Americas but also to analyze their occupational profiles and the potential role of the latter in shaping local economic development. We consider only the initial period of settlement (1492-1540) since, during this period, colonization of the Americas

[14]They also include references by the author to the other three volumes, covering the period from 1541 to 1599. Unfortunately, however, the latter were unpublished manuscripts that went missing after Peter Boyd-Bowman died in 2010. Although we contacted Boyd-Bowman's family (widow, son, and daughter), as well as administrators and colleagues, those manuscripts could not be found. In addition, we checked the library of his university and inquired about a potential donation of the manuscripts, but we were unable to locate the three missing volumes. Regardless, this does not represent an important shortcoming as, for methodological reasons, we should work only with the settlers in the very first colonization wave.

[15]The volume and page refer to the location of the traveler in the compilation of Bermudez (1940-46).

was very extensive, ranging from northern Mexico to Argentina. Our database includes a total of 19,831 settlers. In less than 50 years, the Spanish settled and founded cities in a very large territory, from Zacatecas in Mexico to Argentina. In this first phase, the settlements were isolated, and they grew up in circles, with initially no connections between them. It is during this initial phase that the assumption of random location is most likely to hold. After this first phase, the settlers start exploring the territory around them, and the foundation on the interior after 1540 that started being organized from the initial settlements is less likely to be random. We do not consider these later settlements. Our analysis is concentrated in this first period during which settlement followed the penetration lines of discovery and conquest and in which the "direction to go" to conquer was decided in Spain without knowledge of any characteristics of the land to be conquered. In fact, in some cases, the direction leads to the ocean and no land to conquer. In this period, the only authority was the conqueror, and for many years, there were no formal institutions. At the end of this period, there were only two audiencias. These initial characteristics that define the inhospitable context of the initial conquest and settlement are crucial to understand the context in which the process of discovery, conquest and early settlement occurred, and the ignorance of the territory and limits of the area, together with the isolation of the "world," were part of it. Each settlement had to become organized and start a small community alone with the resources they had, including their abilities, the land and the help of indigenous groups. A situation in which the decision to settle in a particular area was often simply made because the group or their resources were exhausted. We control for climate, geography and indigenous groups that existed prior to the arrival of the settlers. However, since we are dealing with the first wave of settlements, issues related to the role of future viceroyalties, the potential attraction role of the type of settlers, the type of business developed in the area and so forth are not truly relevant.

In the original books of Peter Boyd-Bowman, there are 27,013 names in total. Among them, 7182 do not have a destination, usually because they are reported as drowned, hanged, executed, dead on the trip, etc. A few of them do not have a year of arrival, which we also decide to drop from the sample since in those cases we are uncertain of the time of arrival and, therefore, whether they belong to the group of settlers we consider.

Therefore, the number of settlers for which Boyd-Bowman reports a destination in the period of study (until 1540) is 19,831. In this very early time, voyage mortality on

the Atlantic passage was certainly very high. In fact, in a much later period, the 18th century, the single most important concern for immigrants was voyage mortality[16]. Mancke & Shammas (2005) estimates that for the period 1719-1736, the mortality rate of slaves in the middle passage was approximately 15% per voyage compared to approximately 10% among servants[17]. Following Grubb (1987), the mortality of German immigrants to Pennsylvania in a later period, 1727-1805, was 5.5%. To this rate, we have to add the debarkation morbidity rate, which came from afflictions so serious as to have caused these people to be bedridden. Many of these people may have also died. We should also note that the trip during the first years of the 16th century took two and a half months, while in the 18th century, sailing ships needed only six weeks. The death of slaves was approximately 100 deaths per month per 1000 persons. The lengthier the trip is, the higher the mortality per trip.

After digitalizing and codifying all of this information and using additional sources available in the General Archive (in particular, old maps depicting the positions of the original settlements), we geolocalized the destination locations of the settlers. While the majority of names of the cities of origin easily match their contemporary counterparts, matching the names of some destination areas with their current locations represents more of a challenge. Nevertheless, in many cases, the corresponding name is relatively straightforward. For other locations, an investigation into town name changes was necessary. We investigated the historical names of locations and were able to match the rest of the places. Once we had the contemporaneous name of a historical place, we could look for the geographical location. We have a total of 123 destinations (villages).

Among all the settlers, there are 13,692 for whom we have a destination that corresponds to the name of a foundation of a city or village, so we have a precise location. Usually, the year they arrive at this place is earlier than the foundation year. This is because they petition the king for the status of a village or city after they have settled. For other settlers, 6,136, instead of the name of the city they founded, Boyd-Bowman provided a name of the region or area. When the destination refers to an area, we look for the settlements (villages) that were settled in that particular period in this area. We have only one settlement in the area. To do this, we use a book by Lopez de Velasco (1573), who was a cosmographer sent by the Spanish

[16] Grubb (1987).

[17] Pages 30-31.

authorities in 1571, to map Spanish settlement at that time.[18]. Although some of the settlements considered by Lopez de Velasco (1573) were created after our limiting date (1540), most of our settlements can be matched to the settlements of previous periods. Lopez de Velasco (1573) was sent to Latin America to describe the geography of Las Indias. The main objective was to know all the settlements that were established up to 1570. We search Lopez de Velasco (1573) for the settlements that existed in the year the settler arrived in the area of the final destination. Logically, there were very few settlements in the area during that particular period, generally only one. Therefore, we assign the settler to the village or city that was founded in the period (year) closest to the year of arrival. All these places are among the 123 locations that we consider.

FIGURE 1: Early colonial settlements by number of settlers in 1540. *Source: Own elaboration.*

These criteria allow us to locate the first settlements of Spanish colonizers in the

[18]Lopez de Velasco was also known for constructing instruments to measure longitude using lunar eclipses.

Americas. Figure 1 depicts the locations of all the settlements identified for the period 1492-1540, where the size of the bubbles reflects the number of settlers in the final year of the period.

Various researchers have worked with the places listed in Lopez de Velasco (1573). De Solano (1990) listed 227 cities, villages, and mines founded up to 1574. Among them, he could find information on 189. These numbers included cities that were founded up to 1540 and others that were founded between 1540 and 1570. Our 123 settlements were created before 1541. In fact, 97 of the 123 destinations of Peter Boyd-Bowman appear in De Solano (1990) and Lopez de Velasco (1573). However, we check whether there are settlements in De Solano (1990) and Lopez de Velasco (1573) founded before 1540 that do not appear in Peter-Boyd Bowman. To do so, we search for the year of foundation of all 227 original places except for the 97 that we already have. Among the 227 (or 189 for the important ones) there are many places founded clearly after 1540 and before 1570. We then have places for which the year of foundation is not clear because different sources provide different years. We thus included those that were considered cities and those for which Lopez de Velasco (1573) had some information on the inhabitants. Among them, we find only 7 that were considered cities in which the year of foundation appears in at least one source before 1540. This indicates that the settlements in the volumes of Boyd-Bowman cover the most important early settlements of that period.

3.1 Data on Occupations

For our study, the occupations of these first settlers are a fundamental variable. Traditionally, many historians have assumed that most early settlers in the Americas were soldiers, but as argued in section 2, the data suggest this to be a misconception. In this regard, Lockhart (1968) shows that early settlers came from a variety of occupational backgrounds.[19] Our data confirm his findings: we find 274 different occupations among the first settlers. Some examples include banker (*banquero*), lawyer (*abogado*), pharmacist (*boticario*), barber (*barbero*), butcher (*carnicero*), dentist (*dentista*), doctor (*doctor médico*), farmer (*labrador*), merchant or trader (*mercader*), weaver (*tejedor*), and servants (*siervos*).

[19] Leon (2007) describes travelers from Spain as a "small representation of Castilian society at the end of the XV Century."

Current datasets on labor market surveys usually incorporate information on the education and occupation of the labor force. As we explained above, we have the occupations of the settlers who first arrived in the Americas. However, given the organization and low level of penetration of education in the Middle Ages, the calculation of a proxy for human capital is quite complex. Human capital in the 16th century was very different from how we define human capital at present. One possibility is to attempt to construct a dictionary of occupational titles (DOT) to match occupation with the corresponding level of education. This is a difficult task at present, but it is even more complex when dealing with occupations of many centuries ago. We researched the descriptions of the occupations of the first settlers and attempted to match them with different levels of instruction. For this purpose, we analyzed the content of the occupations included in the list of first settlers and matched it with the corresponding level of education. The measurement of the human capital of the first colonizers is complex, mostly because the education system at the time was organized in a very different way. We explored some definitions using education, with no success. The basic sources are descriptions of those occupations at that time. For example, many works have been done on the formation process of escribanos, for example, exams that prove whether they know how to use technical books. However, only in very few cases do we find an uncontroversial match with a particular level of education.

It is, however, possible to classify occupations using a set of skills derived from the precise description of jobs at the time. To match skills to occupations, we follow the criteria of Ladero Quesada (1980), who analyzes the structure of Spanish society circa 1500. We consider individuals with occupations requiring skills such as being able to read and write or necessitating entrepreneurial drive to have a relatively high level of skills. Therefore, high skilled settlers include those with occupations such as doctors, lawyers, pharmacists, etc. Low skilled settlers include those with occupations ranging from peasants (*jornaleros*) to servants (*criados*). There is some evidence of the relative contribution of high skilled versus low skilled settlers during this period. For instance, Lockhart (1968) argues that "the only Spanish class that contributed truly almost nothing, as a functioning group, to Spanish Peruvian society, were the peasants."

We start working with the names that appear in Peter Boyd-Bowan's books. Originally in the books, 842 different names appear for occupations. However, many of them

are spelling errors. For example, the word "interpreter" is written as "INTERPETE" and "INTERPRETE". In both cases, this means interpreter, since in the first case, there is a spelling mistake. Additionally, the same occupation is written in different ways. For example, there are four different ways of writing merchant (mercader in Spanish): 1) mercader; 2) mercader lancero, 3) mercader tratante, and 4) mercaderes. The four names are included in one category named MERCADER. After cleaning all the data following the two above criteria, we have the list of 274 different occupations mentioned above.

We classify the 274 individual occupations into different categories. The first group is called "local business". This group includes artisans that had a small shop and atelier (it refers to craftsmen or individuals who performed skilled work with their hands), including standard artisan occupations such as colonial weavers and shoemakers, as well as businesses such as printing, metal smelting and building construction. Here, we also include owners of farms, cattle herders, fishermen and individuals who owned local businesses such as small boat owners. All these jobs have in common that they had to run a small business. They represent close to 10 percent of all settlers. We name the second category "royal officials" (usually public servants). This category includes administrators, explorers, soldiers and military officials of the conquest, and officials related to the treasury and judicial systems. They represent 18.76 percent of all settlers. The third category includes settlers with a college degree (e.g., physicians, doctors, notaries)[20] representing 4.67 percent of the sample, and the fourth category is "other high education", which includes for example interpreters and represents 5.16 percent of settlers. The fifth category includes "clergymen" or individuals related to the ecclesiastical system, including priests. This category represents 2.63 percent of all settlers; other categories are merchants with 1.97 percent of the sample, bankers and lenders with 0.07 percent of the sample and finally artists, which includes musicians, painters, singers, and writers, among others, and represents 0.4 percent of the sample (see Table A.8 in the Online Appendix).

[20]Before 1520, there were 13 universities in Spain: Lleida, Huesca, Girona, Barcelona, Valencia, Valladolid, Salamanca, Siguenza, Santiago, Alcala, Avila, Sevilla and Toledo. At that time, these universities offered studies in law, medicine, theology, and canon law. For this reason, we considered settlers with occupations related to these disciplines to be college educated.

All the groups defined above are what we call the high-skill categories, which cover 45 percent of the sample. The low-skill category basically includes two groups, servants either from the cities (called criados, or slaves) or from the countryside, the peasants, usually called "jornaleros", did not own any land, and they worked land from different owners in different places. Following Lopez de Quesada, in the description of Spanish Society in 1500, this group is considered to have "no occupation", although they represent the majority of the people at that time.

Figure A.1 (see Online Appendix) depicts each of the destinations according to the proportion of settlers who were high skilled, based on the previous definition. The colors of the dots on the map show clear heterogeneity across locations in terms of the first settlers' skills levels.

4 Methodology and basic results

In this section, we discuss the identification strategy and present the basic results.

Our aim is to determine whether the proportion of high skilled settlers among the early Spanish colonizers can explain differences observed in present-day levels of development around the original locations. Our identification strategy relies on the exogenous nature of the initial Spanish settlements. The territorial expansion from 1492 to 1540 of the urban centers settled by Spanish colonizers extended between the 25th parallel north and the 35th parallel south, reaching from Zacatecas (northern Mexico) to Buenos Aires.

In less than fifty years, colonizers had settled and founded cities across a very large territory, with considerably diverse geography, climate and indigenous groups.

Our choice of colonial area and time frame is driven by the possibility of generating the conditions of a pseudoexperiment, which helps identify the impact of skills and the persistence of this effect.

Crucially, we consider an area under a single colonial power, allowing us to avoid a bundling of institutions and human capital.[21] In addition, we use the period of initial colonization (1492-1540), when differences in the skills of the colonizers across

[21] In the next section, we present the results of a robustness exercise that considers the potential role of local and informal institutional arrangements.

settlements are likely to be exogenous. The location is also likely to be random and, therefore, exogenous and uncorrelated with the characteristics of potential economic activity in the place of destination. This is the initial phase of discovery, colonization and settlement, from 1492 to 1540, that took place simultaneously in the case of Spanish Latin American colonies. Understanding this context is key to understanding the uncertainty these first settlers faced. Settlers did not know the characteristics of the territory or of the climate of the new land or even the territorial limits. They were simultaneously explorers, conquerors and settlers. The context of these early 50 years is what defines the quasinatural experiment.

In this context, these settlers were the first Spaniards who arrived there. This means that they could not be attracted to particular locations by the previous settlers since they were the first.

Indeed, the first settlers had no idea whether certain areas offered a comparative economic advantage, given that the geography of the New World was unknown. As described above, their only guidance consisted of the recommendations of Saint Thomas Aquinas on suitable locations for settlements, together with the king's prohibition on migrating from a new settlement before it was firmly established.

Additional evidence similarly indicates a lack of knowledge about the geography and potential advantages of different areas of the Americas. During the initial phase of colonization, even the *capitulaciones* contained very imprecise geographical allocations, to the extent that in many cases, the area granted in fact consisted of a piece of the Pacific Ocean, a nonexistent island, or a high mountain. Garcia-Martinez (1970) argues that conquerors could be lucky or unlucky in terms of the land allocated to them in the *capitulacion*, reflecting just how random the process of initial settlement was. It was precisely due to such uncertainty about the geography of the New World that in 1571, the Spanish authorities decided to send Lopez de Velasco to demarcate and describe the geography of the settlements Lopez de Velasco (1573).

Accordingly, it is also very likely that the geographical distribution of settlers by skills is exogenous. Our approach exploits this natural experiment of the original Spanish colonization period, thus avoiding the elusive search for instruments for current/recent past levels of human capital.

To analyze the contemporary level of development of each area, we construct buffers of 20 km around the original settlements[22] and consider the effect of the

[22]In the online Appendix, we demonstrate the robustness of the results to buffers of

skills of the original settlers conditional on many other potential factors that could explain today's levels of economic development. Indeed, the original locations do not generally correspond to current administrative divisions, making it necessary to proxy for income per capita in the buffer around the original settlement. While the main outcome variable is per capita income for the buffer area around the original settlement, good-quality data on income per capita is difficult to find at the buffer level. We thus follow Henderson et al. (2012) and use luminosity at night as a proxy for development. Satellite night-light data are available from the National Oceanic and Atmospheric Administration. Chen & Nordhaus (2011) find that luminosity does, in fact, have informational value for countries, regions, and areas with poor-quality or missing data. Several papers have also found a very high correlation between luminosity and GDP at the country level. For instance, this type of data has been used by Michalopoulo & Papaioannou (2013) and Alesina et al. (2016), among others. We use Oak Ridge National Laboratory's (ORNL) Landscan data for population information at each buffer.

The basic specification is:

$$\log gdp_{ji} = \alpha_i + \beta HC_{ji} + \sum \gamma_k z_{kji} + \epsilon_{ji}$$

where $\log gdp_{ji}$ is the log of per capita night light in 2010. The variable HC is a proxy that measures the skill levels of the first settlers. The basic variable is the proportion of highly skilled settlers in 1540. Alternatively, we consider a dummy variable that takes value 1 in locations with a very high proportion of skilled workers (more than 75%).

We also consider controls for geography, using distance to the sea, average roughness,[23] and average elevation and control for climate using average precipitation and average temperature (CRU). We include as control variables soil quality (% fertile soil), shortest distance to a river, and shortest distance to a lake. We calculate averages for each buffer. In addition, we also include other variables that various papers in the literature have included, such as latitude, suitability measures (including caloric suitability index), and disease environment (malaria indices). Moreover, in this paper, we will include other variables that could be important for our analysis, such as

different sizes.

[23] Nunn & Puga (2012).

characteristics of the indigenous groups (population density), hierarchy level, and indigenous construction. In addition, we also analyze the role of informal institutions. Therefore, we include variables related to this, among other factors, that we describe in the next sections. Finally, we include the first year of the settlement and area fixed effects to complete the list of controls. To calculate the standard deviations, we cluster at different levels of aggregation (8 by 8 degrees and 10 by 10 degrees).

Table 1 summarizes the descriptive statistics of most of these variables. The sample includes those settlers who have a precise destination.[24] The Online Appendix provides a detailed description of all the variables.

TABLE 1: Descriptive statistics

	mean	sd	Number of settlements
Dependent Variable			
Log Night Light 2010 per capita	-3.056	0.743	123
Settlers and Skills			
Total Settlers	111.317	505.599	123
Year of First Settler	1526	10.596	123
% with Skills	0.684	0.317	123
Majority with Skills (>75 %)	0.504	0.502	123
% with Skills (all)	0.674	0.297	123
Initial Conditions and Audiencias			
Log Distance to Port	6.027	1.768	123
Log Distance to Seville	9.020	0.121	123
Precolonial Pop Density	0.249	0.588	98
Hierarchy Dummy	0.553	0.499	123
Temples Dummy	0.276	0.449	123
Log Distance to an Audiencia before 1540	6.329	1.879	123
No Audiencia at Foundation	0.081	0.274	123
Years without Audiencia	0.675	2.824	123
Geography and Climate Variables			
Log Temperature	4.394	0.755	123
Log Precipitation	2.915	0.273	123
Log Ruggedness Index	20.608	0.323	123
% Fertile Soil	50.994	30.331	123
Log Distance to River	2.770	0.407	123
Log Distance to Lake	5.166	1.206	123
Log Distance to Coastline	4.191	1.120	123
Latitude	11.533	12.591	123
Caloric Suitability	4.057	1.436	123
Agricultural Suitability	0.757	0.252	118
Malaria (Average Endemicity)	0.008	0.021	123

To analyze the hypothesis that the geographical distribution of settlers by skill is exogenous, we check whether pre-existing conditions such as geography, climate, pre-Colombian population density, and other characteristics of indigenous groups have explanatory power for the original distribution by occupation of settlers among

[24]The results presented in the tables in the following sections are not affected when using either the full sample or that containing settlers with a precise location. In the robustness section on the Online Appendix, we show that the results are not altered by using the full sample.

populated urban centers. Table 2 examines the relationship among geography, climate and other characteristics and the proportion of settlers with high skills by location. In column 1, we use the percentage of settlers with skills at each location.

In column 2, we consider the explanation of a dummy variable that takes a value of 1 if more than 75% of the settlers are high skilled. The two specifications show that geography and climate cannot explain the location of high-skilled versus low-skilled settlers. It would therefore seem that the distribution of high-skilled versus low-skilled settlers was random with respect to weather, geography, and soil characteristics, malaria environment and characteristics of indigenous groups.

TABLE 2: Determinants of the distribution of skills of the first colonizers

	Dependent Variable:	
	% with Skills (1)	Majority with Skills (2)
Log Temperature	0.021	0.015
	[0.058]	[0.110]
Log Precipitation	0.093	0.308
	[0.169]	[0.335]
Log Distance to Coastline	-0.064	-0.054
	[0.041]	[0.077]
Log Ruggedness Index	-0.033	-0.075
	[0.172]	[0.316]
Log Distance to River	0.148*	0.100
	[0.086]	[0.131]
Log Distance to Lake	0.046	0.082
	[0.034]	[0.060]
% Fertile Soil	0.000	0.000
	[0.001]	[0.002]
Latitude	0.002	0.004
	[0.003]	[0.005]
Agricultural Suitability	0.030	0.088
	[0.204]	[0.319]
Caloric Suitability	-0.000	-0.000
	[0.000]	[0.000]
Malaria (Average Endemicity)	-3.501*	-0.001
	[2.006]	[2.630]
Pre-Colonial Pop Density	0.055	0.102
	[0.053]	[0.099]
Hierarchy Dummy	0.050	0.137
	[0.068]	[0.126]
Temples Dummy	-0.057	-0.161
	[0.074]	[0.133]
Observations	96	96
R-squared	0.171	0.146

Notes - Robust standard errors in brackets.* Significant at 10%, ** significant at 5%, and *** significant at 1%.

Table 3 presents the results on the relationship between the proportion of high-skilled early settlers and long-term local development. The explanatory variable is the percentage of highly skilled early settlers in 1540. Column 1 shows that the locations that received a higher percentage of high-skilled early settlers are more developed at present than those that received a low proportion of high-skilled settlers. The estimation is not only statistically significant but economically very relevant: areas that received only high-skilled settlers have a 60% higher level of GDP per capita than

do areas that received only low-skilled settlers. This result is basically unaffected if we add controls for geography and climate as in column 2. The results do not change qualitatively if we add controls for land quality, as measured by the percentage of fertile soil, and for proximity to potable water, as measured by distance to the closest river or lake (column 3). These results are consistent with those described above and with the findings in Table 2, which show that land characteristics, geography, and climate did not determine the location of high-skilled versus low-skilled settlers. In column 4, we add latitude, and the results remain unchanged. In columns 5 and 6, we add two variables that capture land suitability. First, we use a variable from Ramankutty, and in column 6 following Galor & Özak (2016), we use the caloric suitability index instead of the Ramankutty et al. (2002) measure. Our results are robust to the addition of these variables. In column 7, we include a measure to capture the disease environment. We follow Arbatlı et al. (2020) and construct the mean level of Plasmodium falciparum malaria endemicity using Gething et al. (2011) estimates (see the Online Appendix). The results are unaffected by including this variable. In column 8, we include the first year of the settlement.

TABLE 3: Development and skills of the first colonizers

Dependent Variable: Log. Night Light 2010 per capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
% with Skills	0.610***	0.588**	0.668***	0.666***	0.681***	0.627**	0.508**	0.449**	0.449***	0.403***	0.395***	0.349***
	[0.230]	[0.236]	[0.231]	[0.231]	[0.243]	[0.240]	[0.224]	[0.217]	[0.119]	[0.129]	[0.102]	[0.100]
Log Temperature		0.092	0.098	0.076	0.167	0.098	0.143	0.176	0.176	0.180	0.242*	0.327***
		[0.099]	[0.111]	[0.111]	[0.121]	[0.115]	[0.114]	[0.114]	[0.140]	[0.139]	[0.134]	[0.111]
Log Precipitation		0.172	0.153	0.056	0.039	0.134	0.277	0.151	0.151	0.165	-0.147	-0.035
		[0.229]	[0.235]	[0.256]	[0.262]	[0.278]	[0.289]	[0.288]	[0.346]	[0.332]	[0.396]	[0.365]
Log Distance to Coastline		-0.048	-0.046	-0.024	-0.022	-0.010	-0.012	-0.110	-0.110	-0.085	-0.261***	-0.228***
		[0.085]	[0.085]	[0.090]	[0.089]	[0.092]	[0.092]	[0.099]	[0.087]	[0.080]	[0.086]	[0.064]
Log Ruggedness Index		0.292	0.302	0.300	0.459	0.224	0.106	0.067	0.067	0.026	0.410	0.179
		[0.326]	[0.324]	[0.319]	[0.329]	[0.336]	[0.328]	[0.326]	[0.320]	[0.306]	[0.307]	[0.289]
Log Distance to River			-0.317**	-0.299**	-0.296**	-0.288*	-0.331**	-0.240*	-0.240*	-0.278*	-0.169	-0.164*
			[0.152]	[0.147]	[0.146]	[0.152]	[0.151]	[0.140]	[0.127]	[0.134]	[0.110]	[0.090]
Log Distance to Lake			-0.057	-0.033	-0.053	-0.052	-0.031	-0.043	-0.043	-0.045	-0.016	-0.029
			[0.054]	[0.055]	[0.057]	[0.057]	[0.057]	[0.057]	[0.054]	[0.056]	[0.056]	[0.046]
% Fertile Soil			0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.002	0.001
			[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Latitude				0.008	0.010*	0.010*	0.009	0.014**	0.014**	0.013*	-0.014	0.033*
				[0.006]	[0.005]	[0.006]	[0.006]	[0.006]	[0.007]	[0.006]	[0.025]	[0.019]
Agricultural Suitability					-0.475							
					[0.296]							
Caloric Suitability						-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
						[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Malaria (Average Endemicity)							-7.264**	-5.300*	-5.300**	-4.986**	-1.880	-0.994
							[3.262]	[2.923]	[2.374]	[2.301]	[3.498]	[2.711]
Year of settlement	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Colonization Area FE	No	No	No	No	No	No	No	No	No	Yes	No	No
Audiencia FE (16th century)	No	No	No	No	No	No	No	No	No	No	Yes	No
Country FE	No	No	No	No	No	No	No	No	No	No	No	Yes
Observations	123	123	123	123	118	123	123	123	123	123	123	123
R-squared	0.068	0.101	0.135	0.148	0.187	0.154	0.187	0.269	0.269	0.289	0.408	0.527

Notes - Robust standard errors in columns 1 to 8. Robust standard errors clustered at grid cells of 10 degrees × 10 degrees in columns 9 to 12. Geographic and climate controls include log distance to coastline, log ruggedness index, log average temperature from 1961-1980, log average precipitation from 1961-1980, log distance to river, log distance to lake, % fertile soil, latitude, either caloric suitability or agricultural suitability, and malaria (average endemicity). All variable sources are detailed in Online Appendix B. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

In column 9, we cluster using cells of 10 by 10 degrees. In the Online Appendix, we show that the results are robust to clustering at other geographical levels: using

cells of 8 by 8 degrees.[25] and clustering at the audiencia level or country level (see Table A.1 in the Online Appendix). In columns 10 to 12, we include alternative fixed effects for historical boundaries. In column 10, we use colonization area fixed effects, which correspond to the early colonization zones during the 50 years of early discovery and colonization. In column 11, we instead use the audiencias of 1600, which capture the borders of an intermediate period, while in column 12, we use a measure of contemporaneous borders (country fixed effects). The basic results are also robust to the different characterizations of the boundaries of the territory over time. In the following sections, we consider the earlier borders, which are the colonization areas, because this is the period we are studying.[26]

In the Online Appendix B, we analyze the robustness of our results. First, we demonstrate their robustness to the use of alternative measures of high skill, the use of all data on settlers, and a change in the size of the buffer. Second, we analyze the sensitivity of the results to the inclusion of different types of high-skill occupations and to the inclusion of diversity of occupations. The only group of settlers that seems to affect the significance of the main variable are the settlers whose skills are related to occupations whose main characteristic was to run a "local business". This will become important when we explain the main mechanism and the persistent effects. Third we also show the results are robust to the inclusion of diversity of the place of origin of settlers. Finally, we perform some analyses to reinforce our identification strategy, such as randomization inference and placebo simulations. All the details for the robustness exercises can be found in Online Appendix B.

5 The role of initial conditions

In this section, we analyze the role of initial conditions, including population density, the role of some characteristics of indigenous societies, and the role of informal/initial institutions.

[25] We thank Hillel Rapoport for suggesting this clustering scheme.

[26] All results are robust to the use of audiencia fixed effects or country fixed effects.

5.1 Settlers, population density and indigenous characteristics

First, we consider whether the effect of high-skilled settlers is simply a reflection of the size of the settlement. Areas with high settler density could have had more skilled settlers than areas with low densities.[27] To check that this is not driving the results, Table 4 includes the log of total colonizers in early years. The estimation is unaffected by the inclusion of the size of the population that arrived in each settlement (column 1).

Furthermore, we assessed whether the settlers of the first location along a penetration route had different characteristics than colonizers who settled further along the same penetration route. For example, it could be the case that old or ill members could not continue with the *hueste* and were, therefore, those who settled in the first locations. To explore whether this could affect our results, we include the order of the respective settlements along a penetration route in column 2. The results are not affected by the inclusion of this new variable. Additionally, we control for the distance to the main ports in Latin America, following the work of Ellingsen (2020), and distance to Seville in column 3. None of these variables affect the main results. From columns 4 to 8, we include variables that capture the importance of precolonial societies. The most relevant variable measuring the success of these societies is precolonial population density. In column 4, we control for this variable using the data in Maloney & Valencia (2016). The coefficient on the percentage of high-skilled settlers is positive, statistically significant, and of similar size relative to the basic regressions. Alternatively, we could also control for the level of state development of the local groups using the hierarchy variables from Murdock (1959, 1967), which has been widely used in the literature on Africa. This variable appears in the Ethnographic Atlas of Murdock (1967), but it is incomplete for ethnic groups in Latin America. The idea is to capture the level of state development of the indigenous tribe that was at the location of the settlements. First, we map the settlement onto the Murdock Map for Latin America. This requires georeferencing the spatial distribution of native groups using Murdock's work for Latin America (Murdock, 1951, 1960). Then, we need to collect information on the characteristics of these groups. First, we attempt to match the names of the ethnic groups that appear in the Murdock map with the names

[27] The basic specification considers the controls from column 10 of Table 3.

that appear in the Ethnographic Atlas. For the cases that we match, we take the information about the hierarchy level from the Ethnographic Atlas, where available. For all the other cases, those that matched but had missing information and for the cases in which there was no matching, we collected information on the level of state development using the original sources and following the same definition as in the Ethnographic Atlas. The Online Appendix describes in detail the matching procedure and how we collect new information on state development for the groups that were in the territory of the 123 settlements in Latin America. The results, including the updated version of the hierarchy variables for Latin America, are presented in column 5. Again, the results on the effect of the percentage of high-skilled early settlers are unaffected by this variable. Finally, we could also attempt to capture the level of state development of the indigenous groups using archaeological information on the existence of precolonization temples. We use the dataset collected by Mayshar et al. (2020) that report the number of temples in each area[28]. The main source from which Mayshar et al. (2020) collected the information to construct these datasets is the Archaeological Atlas. The Online Appendix provides details on the definition and source of the data. The result using this alternative measure is included in column 6. We believe that with all this information, we are able to capture, as well as we possible, the level of state development of the indigenous groups. In all cases, the inclusion of these proxies does not affect the main result. As expected, precolonial density seems to be the most relevant proxy for state development of the indigenous groups before colonization. Additionally, and since we have constructed the overlap of the Murdock tribes with our 123 early settlements to be able to capture any other potential characteristics of the indigenous groups, we include Murdock fixed effects in column 7. As before, the characteristics of the indigenous groups do not affect the main results. Column 8 includes all the controls, showing that the results are robust to this complete specification.

[28]We thank Luigi Pascali for sharing the data with us.

TABLE 4: Controlling for initial conditions

Dependent Variable: Log. Night Light 2010 per capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% with Skills	0.404***	0.408***	0.398**	0.332**	0.407***	0.402***	0.420**	0.553**
	[0.127]	[0.133]	[0.158]	[0.138]	[0.126]	[0.132]	[0.173]	[0.228]
Log Total Early Settlers	-0.047							0.066
	[0.088]							[0.141]
Log Distance to Port			-0.002					0.086
			[0.027]					[0.060]
Log Distance to Seville			0.060					-4.312
			[0.972]					[3.419]
Pre-Colonial Pop Density				-0.579***				-0.562***
				[0.037]				[0.101]
Hierarchy Dummy					-0.068			
					[0.217]			
Temples Dummy						-0.150		
						[0.156]		
All controls column 10 of Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonizer Route Order	No	Yes	No	No	No	No	No	Yes
Murdock FE	No	No	No	No	No	No	Yes	Yes
Observations	123	123	123	98	123	123	123	98
R-squared	0.291	0.312	0.289	0.447	0.291	0.296	0.639	0.779

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

5.2 Early institutions

However, another important issue to consider is whether the effect of early skilled settlers is confounded by the impact of initial colonial institutions. In our analysis, the entire area falls under the Spanish Kingdom, where Spanish colonizers brought with them the same institutions. Nevertheless, historical evidence seems to indicate that there were, in fact, very few formal institutions at this very early stage of colonization. The only formal institution consisted of the audiencia, or a tribunal of administration and justice. The audiencias were responsible for ensuring the implementation of law and order and, in particular, intervening in cases where issues arose involving conquerors. However, audiencias were often far from early settlements; in other cases, an audiencia had not yet been established at the moment of founding of a given settlement.

In addition to the audiencias, a second source of rules and regulations was the conqueror. In signing a *capitulacion*, the conqueror was recognized as an informal source of law and order and as such held most of the power at the colonization stage. Thus, although there existed some audiencias, conquerors in areas far from the audiencias may have had few constraints on their control. To capture this early institutional framework, we use as proxies several measures of audiencias and the penetration paths of the conquerors.

Figure A.2 (see Online Appendix) depicts the two audiencias established before 1540: Santo Domingo (established in 1511) and Mexico (established in 1527). The

number of audiencias then increased rapidly after 1540, as shown in Figure A.3 (see Online Appendix). This means that before 1540, the political centers of many settlements were located quite far away, implying that the conquerors were likely the foremost source of law and order.

The expedition paths followed by the conquerors are known as "penetration lines."[29] We geodigitalized 51 expedition paths for the period before 1540, which correspond to 31 conquerors.

For each line, we have information about the starting year, end year, and name of the conqueror. Figure 2 depicts these penetration lines, showing that, as expected, early settlements are very close to these paths. Figure A.4 (see Online Appendix) provides a closer view of the penetration lines in Central America before 1540. We associated the location of the various settlements with their corresponding penetration line and the conqueror leading each expedition.

[29]Information on the penetration lines comes from the *Atlas del Descubrimiento de América y Oceanía*, which uses Montana (1943b) as a basic source.

FIGURE 2: Penetration lines before 1540

In Table 5, we analyze the robustness of the basic results to the inclusion of different proxies that capture the early institutional framework. In column 1, we include the log of the distance from each settlement to the closest audiencia, as a control for the latter's ability to exercise power over each location. In column 2, we include a dummy that takes a value of 1 if in the year of first settlement, there was no official audiencia (basically, those settlements before 1512). In column 3 we include the years without an official audiencia as an alternative measure of a lack of formal institutions. Finally, in column 4, we include the penetration line of the conqueror as a proxy for the conqueror-penetration line effect. The results presented in Table 5 show that taking into account early institutions does not alter the basic finding: the initial settlements with a high proportion of high-skilled settlers perform better in the long run than do those with many low-skilled settlers. Finally, in column 5, we include

all these variables together, and in column 6, we include all these variables, plus all variables from Table 4, showing that the main result is robust to this specification.

TABLE 5: Audiencias and colonizer effect

Dependent Variable: Log. Night Light 2010 per capita

	(1)	(2)	(3)	(4)	(5)	(6)
% with Skills	0.407***	0.384**	0.337**	0.672***	0.627**	0.808***
	[0.131]	[0.160]	[0.146]	[0.125]	[0.229]	[0.220]
Log Dist to Audiencia	0.068				0.140	0.014
	[0.062]				[0.112]	[0.118]
No Audiencia at Foundation		0.170			-1.142***	
		[0.579]			[0.234]	
Years without Audiencia			0.060		0.128	0.122
			[0.046]		[0.094]	[0.183]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes	Yes	Yes
Colonizer Route FE	No	No	No	Yes	Yes	Yes
All controls of Table 4	No	No	No	No	No	Yes
Observations	123	123	123	123	123	98
R-squared	0.304	0.290	0.302	0.652	0.710	0.939

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. The vector of controls from Table 3 includes log total settlers (first year), log distance to port, log distance to Seville, precolonial population density, Murdock FEs, and colonizer route order. In column 6, the coefficient on No Audiencia at Foundation is dropped because of multicollinearity. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

6 Analyzing the persistence mechanism

The results of the previous sections show that the areas in Latin America that received the highest proportions of high-skilled settlers have greater income per capita today than places that received a high proportion of low-skilled settlers. In this section, we assess several possible drivers behind the channels of persistence. The economic literature has documented notable persistence of important historical events on comparative development. Much work centers around early institutions, cultural traits, and human capital. In our case, the early institutions of the Americas are relatively uniform and linked to a single colonial power. We also control for the possibility of informal early institutions, namely, powerful conquerors far from the audiencias. The source of the initial shock is the different distribution of skills of the first settlers.[30]

The mechanism for persistence is related to the degree of entrepreneurship needed to develop high-skill occupations. As argued above, occupations considered high skilled share a common characteristic: they imply a certain degree of entrepreneur-

[30]Occupations have frequently been employed in studies that analyze very long-run processes. For instance, the intergenerational mobility literature often uses occupational change as a metric of mobility when conducting comparisons with earlier periods. See Long & Ferrie (2013) or Barone & Mocetti (2016).

ship. Murphy et al. (1991) and Baumol (1990) likewise consider the central role of entrepreneurship in the process of economic growth. In line with Baumol (1990), we define entrepreneurs broadly as "persons who are ingenious and creative in finding ways that add to their own wealth, power, and prestige."[31] Artisans, doctors, etc. must necessarily invest and accept some risk to develop their activities. By contrast, the low-skilled group includes laborers, such as peasants and servants, who work for others.

Our analysis therefore relies on differences in the supply of local entrepreneurial skills and assesses their persistence over time. Prior research shows evidence of the long-term stability of certain attitudes. For instance, Necker & Voskort (2014) and Dohmen et al. (2012) provide empirical support for the persistence of attitudes such as the willingness to take risks. There is also evidence of persistence in occupations over generations.[32] Opper & Andersson (2019) analyze the long-term stability of entrepreneurial culture and talent pools using occupational choice in preindustrial China. They find strong persistence in the regional distribution of entrepreneurial skills, which correlates with the performance of regions in postreform China.

In light of such studies, which find persistence to be driven by the intergenerational transmission of cultural values and attitudes, it seems reasonable to analyze this factor as the first mechanism. A high level of entrepreneurship can arguably accelerate the structural transformation of the economy when the conditions for such change are met.[33]

There are many potential channels that could explain the persistence of this shock. It is important to note that in the case of Latin America, agriculture remains a very important economic activity. We focus on three factors suggested to be important by the recent literature: the adoption of new technologies in agriculture, the market orientation of agricultural production, and attitudes towards entrepreneurship. To

[31]This definition is similarly adopted in Opper & Andersson (2019).

[32]See, for example, Long & Ferrie (2013). The focus here is less on intergenerational occupational persistence than on the transmission of the entrepreneurial content of the occupations. In this regard, recent research by Barone & Mocetti (2016) shows evidence of persistence of elite professions across many generations.

[33]In the same spirit, Opper & Andersson (2019) finds that the regional distribution of a measure of entrepreneurship during the Ming and Qing Dynasties explains the differential economic growth of regions in China in the postreform period (1992-2012).

examine this channel, we use different measures. To check the relevance of these mechanisms, we construct proxies to capture the adoption of new technologies, the efficiency in the use of land, and market-oriented agriculture. Moreover, we also use historical data on a market-oriented economy for 1573-1620 and a proxy for entrepreneurial attitudes using Latinobarometer. We also consider other mechanisms that are not supported by the data.

6.1 Adoption of new technologies in agriculture and the efficient use of land

Recent research has investigated the adoption of new technology in agriculture as a potential source of long-term persistence. Valencia (2019) studied the adoption of GE soy seeds in Brazil in the spirit of Griliches (1957) and Bustos et al. (2016). Valencia (2019) tests whether areas with higher human capital—closer to Jesuit missions—adopted this new agriculture technology faster than areas with lower levels of human capital. Bustos et al. (2016) obtain an exogenous measure of technological change in agriculture by using estimates of potential soy and maize yields across geographical areas of Brazil. Gollin et al. (2018) examines the economic impact of high-yielding varieties in developing countries.

Following this literature, we consider cereals, rather than soy, which is the most relevant crop in the area we study. To capture the adoption of new technologies, we use a measure from GAEZ-FAO. Following the literature, the adoption of new technologies could be captured by closely examining how close to the frontier research they are. We want to show that areas with high-skilled settlers use land with the most innovative techniques for irrigation, mechanization and seeds. FAO-GAEZ calculates the ratio between the actual versus potential yield of each small area. We use variables for the ratio of actual versus potential yield. Potential yield captures what the land could produce if the best techniques, the best seeds, etc. were used based on the geographic characteristics and weather characteristics of the geographical area. The literature has used similar variables for soy in Brazil. We find that places that had a higher proportion of skilled settlers among its early colonizers are closer to the frontier research (the actual production is closer to the potential production) than areas that had low-skilled settlers among the early colonizers (Table 6, column 1). Therefore, areas that had a high proportion of high skilled settlers among their early colonizers

are more prone to adopt new agricultural technologies characterized by high yield, irrigation and mechanization than areas with a low proportion of high-skilled settlers. Therefore, these areas lead to the structural transformation of agriculture.

As a result of this structural change in the agricultural sector, we also obverse greater efficiency in the use of land in areas that were originally settled by high-skill settlers. Thus, these areas have a higher proportion of cultivated land over total land. To calculate this ratio, we first compute the proportion of land within a 20 km buffer from the colonial settlement that is used as cropland (i.e., territory employed for the cultivation of food) in the year 2000. This data is based on satellite information compiled by Ramankutty et al. (2008), available from Columbia University's Center for International Earth Science Information Network (CIESIN).[34]

Second, we also compute the proportion of land within a 20 km buffer from the colonial settlement that is suitable for agriculture. We use the index of land suitability for cultivation by Ramankutty et al. (2002), which is based on temperature and soil conditions and distributed by the Center for Sustainability and the Global Environment (SAGE), at the University of Wisconsin-Madison.[35]

Using this information, we compute the proportion of land used as cropland over the land that is suitable for cultivation (*proportion of cropland/proportion of land suitable for cultivation*). Note that the data are only available for 118 out of 123 buffers.

The results in column 2 of Table 6 suggest a positive relationship between the above ratio and the level of high-skill early settlers among the first colonizers.

[34]Ramankutty, N., A.T. Evan, C. Monfreda, and J.A. Foley. 2010. Global Agricultural Lands: Croplands, 2000. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4C8276G. Ramankutty, N., A.T. Evan, C. Monfreda, and J.A. Foley. 2008. Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000. Global Biogeochemical Cycles 22 (1): GB1003. http://dx.doi.org/10.1029/2007GB002952

[35]Ramankutty, N., J.A. Foley, J. Norman, and K. McSweeney. 2002. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. Global Ecology and Biogeography 11 (5), 377-392. https://doi.org/10.1046/j.1466-822x.2002.00294.x

6.2 Market-oriented economy

Another persistence mechanism that has been discussed in the literature in the context of Latin America is the market orientation of the products.[36] This mechanism is also associated with entrepreneurship. We can find additional evidence for the modernization of agriculture and improved productivity if we consider the market orientation of agriculture. For this purpose, we use two more variables. First, we take the percentage of market agricultural workers from IPUMS. In line with previous results, we find that places that received a larger proportion of high-skilled settlers have a larger percentage of market-oriented agricultural workers (column 3, Table 6).

We also assess market orientation by examining the existing infrastructures, in particular roads, which facilitate the transportation of products to other places.[37] Infrastructure often requires time to build and follows historical patterns. We examine whether places where the original settlers had a comparatively high skill level also had more roads, allowing them to connect with other areas. Data on contemporary infrastructure are obtained from the Seamless Digital Chart of the World (SDCW) v10.[38] The SDCW is based on the Digital Chart of the World (DCW), the most comprehensive global geodatabase currently available. The main variable of interest is the total length (in km) of any type of road within a 20 km buffer from the colonial settlement. Any type of roads refers to all operational roads, including primary roads (with and without a median) and secondary roads (without a median by definition). We define the following variable:

$$Roads\ per\ capita = \log\left(0.1 + \frac{Road\ Length\ (km)}{Total\ Population}\right)$$

where *Total Population* within a 20 km buffer from the colonial settlement is computed using the LandScan raster data for 2010. In columns 4 and 5 of Table 6, we find a positive relationship between *roads per capita* and the high-skilled proportion of first

[36] For instance, Dell (2010) shows that the long-run persistence of mita can be observed in agricultural market participation or the proportion of households selling agricultural products in the market.

[37] Dell (2010) shows that mita areas currently have a lower density of regional road networks and paved/gravel regional roads than non-mita areas.

[38] See http://www.worldgeodatasets.com/basemaps/

colonizers, measured as the percentage of colonizers with high skills.

The results indicate that locations characterized by a high proportion of high-skilled settlers experience a faster agricultural transformation and higher market orientation than locations with a low proportion of high-skilled settlers.

TABLE 6: Mechanisms – Agricultural transformation

	Dependent Variable:				
	Tech Frontier Cereals (1)	Eff Cropped Land (2)	Mkt Agr Workers (3)	Roads Per Capita (4)	Roads Per Capita (5)
% with Skills	0.399**	0.067*	0.113**	0.007*	0.007**
	[0.166]	[0.039]	[0.042]	[0.004]	[0.003]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes	Yes
Observations	123	118	95	123	123
R-squared	0.482	0.240	0.193	0.133	0.133

Notes - Robust standard errors are clustered at grid cells of 10 degrees × 10 degrees in columns 1 to 4 and at the country level in column 5. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. The dependent variables are as follows: (1)ratio of actual over potential yield for cereal crops within a 20 km buffer from the colonial settlement; (2) ratio of cropped land over land that is suitable for cultivation within a 20 km buffer from the colonial settlement; (3) fraction of the male adult population that works as market-oriented skilled agricultural workers (IPUMS); and (4-5) log roads length (km) per capita within a 20 km buffer from the colonial settlement. In column (5) we clustered at country level. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

Having shown that areas where settlers had a high proportion of high skilled occupations show a higher density of roads, we also investigate whether there is some historical evidence on market orientation during an earlier time. We have codified information coming from two sources that refers to the situation of Spanish settlements in 1574 and 1620. The main sources are Lopez de Velasco (1573) and Vazquez de Espinosa (1620). Both provide a detailed description of the vegetation, fauna, and the products each place produced, the number of inhabitants at that time, and the basic location of the place. All this information is based on the observation of the geographers who traveled to Latin America during the period 1570-73 and in 1620 to attempt to obtain information on the locations of the settlements and what the settlers were doing. First, we have codified information on the economic activities of each of the settlements. The original sources indicate whether a particular area produces agricultural products, vegetables, livestock, or manufacturing, mining and craft products (agricultura, hortalizas, ganaderia, manufactura, mineria y artesania). We found information on 97 of the 123 original settlements.[39] During these early periods, as explained in the previous section, the settlements were isolated, and therefore, having a diversified economy could be a sign of prosperity. We constructed a variable that captures the diversity of the products, which is the sum of the number of different products they produce in each place. The variable ranges from 0 to 5. Column 1 of Table 7 shows that areas with more skilled settlers have a more diversified

[39]We collected information on the products that each settlement produced.

economy.

We can also examine commercial roads circa 1600. Slicher Van Bath (2010) provides a map of the terrestrial trade routes between 1574 and 1628. This map is constructed based on the information available in Lopez de Velasco (1573) and Vazquez de Espinosa (1620). The distances are calculated as straight lines connecting nodes, and they can be used to calculate the distance of each settlement to the closest trade route. We digitized the map and overlapped it with our settlements. We fist use the log of the distance. The result in column 2, although going in the right direction, is not significant. In column 3, we use our dummy variables for high-skilled settlers, and the results indicate that places with a large majority of high-skilled settlers are closer to trade routes. In addition to this, in column 4, we construct a dummy for whether the route is within the 20 km buffer, and we use it as the dependent variable. The results indicate that places with a large majority of high-skilled settlers have a higher probability of being close to a trade route than a settlement populated with low-skill settlers.

TABLE 7: Mechanisms – Trade in 1620

	Dependent Variable:			
	Economic Diversification in 1620 (1)	Log Dist to Trade Route in 1620 (2)	Log Dist to Trade Route in 1620 (3)	Dummy Trade Route in 1620 (4)
% with Skills	0.907**	-1.343		
	[0.394]	[1.056]		
Majority with Skills (>75)			-1.253**	0.223***
			[0.523]	[0.075]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes
Observations	97	97	97	97
R-squared	0.446	0.449	0.471	0.464

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. The dependent variables are as follows: (1) colonial economic diversification; (2-3) log distance to the closest colonial trade route; and (4) dummy taking value 1 if distance to the closest trade route is less than 40km. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

6.3 Attitudes towards entrepreneurship

The previous two mechanisms offer an indirect explanation for the effect of entrepreneurship on the persistence of the impact of the original proportion of high-skill settlers. We can also attempt to directly measure this effect using contemporaneous surveys. To the extent that risk attitudes and, in particular, entrepreneurship persist over time, areas with a high level of entrepreneurs among the original settlers should display higher economic growth in the present. To provide evidence on this second mechanism, we employ information from the Latinobarometer, using responses to the following question as a proxy for attitudes towards entrepreneurship: *"What is the most impor-*

tant thing in order for a country to successfully develop? " We regard as a positive attitude towards entrepreneurship the answer: "Having an active entrepreneurial class." Each individual in the Latinobarometer is associated with a specific location, indicating their city of residence. We use this information to match the individuals with our original locations. (See Online Appendix for a description of the matching procedure) After concluding the matching, we calculated the percentage of individuals who replied "Having an active entrepreneurial class" to the question *"What is the most important thing in order for a country to develop successfully?*" The results in Table 8 show that areas that had a higher proportion of high-skilled first settlers have more positive attitudes towards the role of entrepreneurship in development than do areas that had lower levels of high-skilled early colonizers, supporting the hypothesis of the transmission of entrepreneurial attitudes over generations. We show the results when excluding (column 1) or including (column 2) individual characteristics. Moreover, they are robust to including country fixed effects rather than area fixed effects, as column 3 and column 4 show.

TABLE 8: Mechanism – Having an active entrepreneurial class

Dependent Variable: Having an active entrepreneurial class

	(1)	(2)	(3)	(4)
% with Skills	0.097*	0.096*	0.092*	0.122*
	[0.048]	[0.055]	[0.049]	[0.057]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes
Individual Characteristics	No	Yes	No	Yes
Country FE	No	No	Yes	Yes
Observations	54	54	54	54
R-squared	0.448	0.507	0.733	0.789

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. The vector of individual variables includes respondent characteristics: average age, proportion of females, and proportion of individuals by social class categorical variables. The dependent variable is the share of respondents that agreed with "What is the most important thing in order for a country to develop successfully? Having an active entrepreneurial class". * Significant at 10%, ** significant at 5%, and *** significant at 1%.

6.4 Evidence on Alternative mechanisms

Are there alternative mechanisms that could explain the persistence of skills over time in Latin America? While the literature on the persistence of large historical events often analyzes the education channel, our data are difficult to match with the level of instruction, leading us to use occupations and their associated skill levels. We do, however, check whether the variance in skills across the original locations could have led to persistent differences in education. Indeed, the latter may indicate that the unobservable differences in education, proxied by occupations, are the main factor explaining the persistence. Online Appendix C shows that using data from Census, Ipums, and Historical data from Jesuit missions there is no evidence of this

mechanism. Another potential mechanism of persistence is an agglomeration effect[40] induced by high-skilled settlements attracting more people and thus becoming more densely populated than locations with low-skilled colonizers. Employment density is positively related to productivity at the regional level; the same effect could be at work at the local level. Using data on population density from three different sources, the Population Density Grid from the Gridded Population of World V. 3, Census and LANDSCAN, we find no evidence of this mechanism. Also using historical data on population during 1573, there is no evidence of this mechanism. (See Online Appendix)

7 Concluding remarks

In this paper, we use administrative data on the first colonizers of the Americas to analyze the long-run effect of skill differences in the original settlements.

To this end, we construct a dataset containing the first 20,000 settlers in the Americas, which includes their complete name, place of origin, destination, and occupation. To the best of our knowledge, this is the first time that information on the skills of the first settlers in the Americas has been used for economic research.

The early years of Latin American colonization (1492-1540) provide a unique natural experiment for evaluating the effect of skills on long-run development. First, differences in the geographical distribution of skills are exogenous. Second, the area was colonized by a single colonial power, allowing us to hold constant formal institutions and legal origin. Previous analyses of the legacy of colonial powers have addressed the bundle effect of institutions and human capital using various instruments (which are sometimes difficult to justify) or current/recent past levels of human capital. Our approach provides a novel and credible alternative to such methodologies.

The results indicate that locations where a greater number of early settlers had high skilled occupations currently have a higher level of development than those where early settlers were less skilled. Our findings are robust to many alternative specifications. Specifically, we find evidence of persistence in the form of market orientation and entrepreneurial spirit.

[40] See Rocha et al. (2017).

References

Acemoglu, D., & Dell, M. (2010). Productivity differences between and within countries. *American Economic Journal: Macroeconomics*, *2*(1), 169–88.

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

Alesina, A., Michalopoulo, S., & Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, *124*(2), 428-488.

Alesina, A., Seror, M., Yang, D., You, Y., & Zeng, W. (2020). Persistence through revolutions. *NBER WP 1060*.

Alsan, M. (2015). The effect of the tsetse fly on african development. *American Economic Review*, *105*(1), 382–410.

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

Ballesteros, M. (1985). La conquista indiana. *Cuadernos Historia 16*, *16*(172).

Barone, G., & Mocetti, S. (2016). Intergenerational mobility in the very long run: Florence 1427-2011. *Banca de Italia WP 1060*.

Baumol, W. (1990). Entrepreneurship: productive, unproductive and destructive. *Journal of Political Economy*, *98*(5), 893-921.

Bermudez, C. (1940-46). *Catálogo de pasajeros a indias durante los siglos xvi, xvii y xviii.* (Vols. vol, I, 1509-1534; vol. II, 1535-1538, y vol. III, 1539-1559). CSIC.

Boyd-Bowman, P. (1956). *Indice geobiográfico de cuarenta mil pobladores españoles de américa en el siglo xvi* (Vol. I).

Boyd-Bowman, P. (1964). *Indice geobiográfico de cuarenta mil pobladores españoles de américa en el siglo xvi* (Vol. II).

Bustos, P., Caprettini, B., & Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, *106*(6), 1320–65.

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, *108*(21), 8589–8594.

Dell, M. (2010). The persisten effect of peru's mining mita. *Econometrica*, *78*(6), 1863-1903.

De Solano, F. (1990). *Ciudades hispanoamericanas y pueblos de indios* (Vol. 2). Editorial CSIC-CSIC Press.

Dohmen, T., Falk, A., & Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, *79*(2), 645-677.

Droller, F. (2018). Migration, population composition and long run economic development: Evidence from settlements in the pampas. *The Economic Journal*, *128*(614), 2321–2352.

Easterly, W., & Levine, R. (2016). The european origins of economic development. *Journal of Economic Growth*, *21*(3), 225–257.

Ellingsen, S. (2020). *Free and protected: Trade and breaks in long-term persistence.* (Working Paper)

Galor, O., & Özak, Ö. (2016). The agricultural origins of time preference. *American Economic Review*, *106*(10), 3064–3103.

Garcia-Martinez, B. (1970). Ojeada a las capitulaciones para la conquista de america. *Revista de Historia de America*, *69*, 1-40.

Gennaioli, N., La Porta, R., Lopez-de Silanes, F., & Shleifer, A. (2013). Human capital and regional development. *The Quarterly journal of economics*, *128*(1), 105–164.

Gething, P. W., Patil, A. P., Smith, D. L., Guerra, C. A., Elyazar, I. R., Johnston, G. L., ... Hay, S. I. (2011). A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria journal*, *10*(1), 1–16.

Gollin, D., Hansen, C. W., & Wingender, A. (2018). *Two blades of grass: The impact of the green revolution* (Tech. Rep.). National Bureau of Economic Research.

Grubb, F. (1987). Morbidity and mortality on the north atlantic passage: Eighteen-century german immigration. *Journal of Interdisciplinary History*, *17*(3), 565-585.

Hamy, A. (1892). *Documents pour server a l'histoire des domiciles de la compagnie de jesus dans le monde entier de 1540 a 1773.*

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, *102*(2), 994–1028.

Ladero Quesada, M. A. (1980). *Historia de america latina. hechos, documentos, polemicas. españa en 1492* (Vol. I). Editorial Hernando.

Leon, M. (2007). Pasajeros del segundo viajes de cristobal colon. *Revista de Estudios Colombinos*(3), 20-60.

Lockhart, J. (1968). *Spanish peru, 1532-1560: A colonial society.* University of Wisconsin Press Madison.

Lockhart, J., Otte, E., Lockhart, J. M., & Knight, A. (1976). *Letters and people of the spanish indies: sixteenth century* (Vol. 22). Cambridge University Press.

Long, J., & Ferrie, J. (2013). Intergenerational occupational mobility in great britain and the united states since 1850. *American Economic Review*, *103*(4), 1109-1137.

López De Velasco, J. (1894). *Geografía y descripción universal de las indias* (Vol. reprinted from the original version from 1573). Establ. tip. de Fortanet.

Maloney, W., & Valencia, F. (2016). The persistence of (subnational) fortune. *Economic Journal*, *126*(598), 2363-2401.

Mancke, E., & Shammas, C. (2005). *The creation of the british atlantic world*.

Mayshar, J., Moav, O., & Pascali, L. (2020). *The origin of the state: Land productivity or appropriability?* (Working paper)

Michalopoulo, S., & Papaioannou, E. (2013). Pre-colonial institutions and contemporary african development. *Econometrica*, *81*(1), 113-152.

Michalopoulos, S., & Papaioannou, E. (2017). *The long economic and political shadow of history* (Vol. I). Washington DC: Center for Economic Policy Research.

Michalopoulos, S., & Papaioannou, E. (2020). Historical development and african development. *Journal of Economic Literature*, *58*(1), 53-128.

Montana, A. (1943a). *Atlas del descubrimiento de américa y oceanía*. Barcelona: Ed. Miguel Salvatella.

Montana, A. (1943b). *Descubrimiento, exploraciones y conquistas de españoles y portugueses en américa y oceanía*. Barcelona: Ed. Miguel Salvatella.

Morales Padrón, F. (1988). *Atlas histórico cultural de américa*. Comisión de Canarias para la conmemoración del V centenario del descubrimiento de América.

Murdock, G. P. (1951). *Outline of south american cultures* (Vol. 2). New Haven: Human Relations Area Files.

Murdock, G. P. (1959). *Africa: Its peoples and their culture history*. New York: McGraw-Hill Book Company.

Murdock, G. P. (1960). *Ethnographic bibliography of north america*. New Haven: Human Relations Area Files.

Murdock, G. P. (1967). *Ethnographic atlas*. Pittsburgh: University of Pittsburgh Press.

Murphy, K., Shleifer, A., & Vishny, R. (1991). The allocation of talent: implications for growth. *Quarterly Journal of Economics*, *106*(2), 503-530.

Necker, S., & Voskort, A. (2014). Intergenerational transmission of risk attitudes - a reveled preference appoach. *European Economic Review*, *65*, 66-89.

Nunn, N. (2014). Gender and missionary influence in colonial africa. In E. Akyeampong, R. Bates, N. Nunn, & J. Robinson (Eds.), *African development in historical perspective* (p. 489-512). Cambridge: Cambridge University Press.

Nunn, N. (2020). The historical roots of economic development. *Science*, *367*(6485).

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, *94*(1), 20–36.

Opper, S., & Andersson, F. (2019). Are entrepreneurial cultures stable over time? historical evidence from china. *Asian Pacific Journal of Management*, *36*, 1165-1192.

Putterman, L., & Weil, D. N. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *The Quarterly journal of economics*, *125*(4), 1627–1682.

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. geographic distribution of global agricultural lands in the year 2000. *Global biogeochemical cycles*, *22*(1).

Ramankutty, N., Foley, J. A., Norman, J., & McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography*, *11*(5), 377–392.

Rocha, R., Ferraz, C., & Soares, R. R. (2017). Human capital persistence and development. *American Economic Journal: Applied Economics*, *9*(4), 105–36.

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

Squicciarini, M. P., & Voigtländer, N. (2015). Human capital and industrialization: Evidence from the age of enlightenment. *The Quarterly Journal of Economics*, *130*(4), 1825–1883.

Steward, J. H. (1946). *Handbook of south american indians* (Vol. 143).

Trigger, B. G., Washburn, W. E., Adams, R. E., MacLeod, M. J., Salomon, F., & Schwartz, S. B. (1996). *The cambridge history of the native peoples of the americas: Mesoamerica* (Vol. 2). Cambridge University Press.

Valencia, F. (2019). The mission: Human capital transmission, economic persistence, and culture in south america. *The Quarterly Journal of Economics*, *134*(1), 507–556.

Van Bath, B. S. (2010). *Hispanoamérica en torno a 1600*. Universidad de Alicante.

Vázquez de Espinosa, A. (1948). *Compendio y descripción de las indias occidentales* (Vol. reprinted from the original version from 1620).

Wantchekon, L., Klašnja, M., & Novta, N. (2015). Education and human capital externalities: evidence from colonial benin. *The Quarterly Journal of Economics*, *130*(2), 703–757.

Online Appendix of the paper:

Colonization, Early Settlers and Development: The Case of Latin America

José G. Montalvo
Universitat Pompeu Fabra-IPEG and Barcelona GSE
Marta Reynal-Querol
Universitat Pompeu Fabra-ICREA-IPEG and Barcelona GSE

A Online Appendix – Tables

TABLE A.1: Standard errors clustered at different levels

Dependent Variable: Log. Night Light 2010 per capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
% with Skills	0.403***	0.403**	0.403***	0.395***	0.395**	0.395***	0.349**	0.349*	0.349**
	[0.127]	[0.152]	[0.116]	[0.116]	[0.160]	[0.121]	[0.142]	[0.169]	[0.133]
All controls column 10 of Table 3 (except Colonization Area FE)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonization Area FE	Yes	Yes	Yes	No	No	No	No	No	No
Audiencia FE	No	No	No	Yes	Yes	Yes	No	No	No
Country FE	No	No	No	No	No	No	Yes	Yes	Yes
Cluster	8×8d cells	Audiencia	Country	8×8d cells	Audiencia	Country	8×8d cells	Audiencia	Country
Observations	123	123	123	123	123	123	123	123	123
R-squared	0.289	0.289	0.289	0.408	0.408	0.408	0.527	0.527	0.527

Notes - Robust standard errors clustered (a) at grid cells of 8 degrees × 8 degrees in columns 1, 4, and 7; (b) at the audiencia level in columns 2, 5, and 8; (c) at the country level in columns 3, 6, and 9; and (*d*) at the level of colonization period in columns 5 and 9. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

B Online Appendix – Robust Analysis

In this section, we analyze the robustness of the results. First, we demonstrate their robustness to the use of alternative measures of high skill, the use of all data on settlers, and a change in the size of the buffer. Second, we analyze the sensitivity of the results to the inclusion of different types of high-skill occupations. Finally, we perform some analysis to reinforce our identification strategy, such as randomization inference and placebo simulations.

In Table A.2, we perform the first robust analysis. In column 1, we analyze the sensitivity of the results to the use of an alternative measure of high skills. We construct a dummy variable to capture locations with a very high proportion of high-skilled settlers. Specifically, we include a dummy that takes a value of 1 if the percentage of high-skilled settlers is larger than 75% (which is the median of the sample of the 123 locations). The results are in the same direction as when using the percentage of high-skilled settlers and indicate that locations that received a high percentage of high-skilled settlers perform better in the present.

In column 2, we include the settlers for who we know the region, and we infer a precise settlement. In the previous sections, we considered those settlers for whom we know their exact destination. However, for other settlers, rather than their exact destination, we have a reference to the region. In the data section, we discussed how we approached that situation to incorporate this information. Column 2 of Table A.2 presents the results using all the settlers, showing that the estimation is consistent with the basic findings.[41]

In the last three columns of this table, we consider the effect of changing the size of the buffer that defines the area of interest. Specifically, we consider 10 km, 20 km, and 30 km buffers. Moreover, we also address another potential issue. Situations of overlap can arise across different buffers. In such cases, we drop the area that was founded later. This approach will ensure our exogeneity assumption.

The results in Table A.2 confirm the basic results: the proportion of high-skilled settlers has a positive effect on long-run development. The estimated coefficient for the percentage of high-skilled settlers is very similar to that obtained in previous tables.

[41]The distribution of skills is almost identical in both samples: the settlers with exact destination and all the settlers.

Table A.2: Robust analysis

Dependent Variable: Log. Night Light 2010 per capita

	Majority with Skills (1)	All settlers (2)	Non-Overlap 10km (3)	Non-Overlap 20km (4)	Non-Overlap 30km (5)
Majority with Skills (>75)	0.218**				
	[0.092]				
% with Skills		0.361**	0.454***	0.455**	0.316**
		[0.164]	[0.116]	[0.158]	[0.147]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes	Yes
Observations	123	123	117	112	108
R-squared	0.284	0.283	0.254	0.294	0.255

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

Another objection that could be raised is that what matters is not the percentage of highly skilled but the diversity of occupations. It could be the case that in places in which there is a larger percentage of high-skilled settlers, there are also more different types of occupations, and therefore what matters would be diversity rather than the percentage of high-skilled settlers. To address this issue, we construct a fractionalization index (1-Herfindhal index) of occupations using the percentage of settlers that have each occupation for each settlement. We first construct the fractionalization measure using all the occupations including the settlers with low skills and then only among the different types of high-skilled settlers, listed in the previous table in section 3. The results are in Table A.3, in which we always include all controls from Table 3. In column 1, we use the fractionalization of high-skilled settlers together with our main variable of interest, which is the percentage of high-skilled settlers. Our results are maintained, and fractionalization has no effect. In column 2, we drop our variable of interest to investigate whether both variables were capturing the same phenomenon. However, fractionalization remains insignificant. In columns 3 and 4, we do the same but use all types of settlers when constructing the fractionalization index (all types of high- plus low-skill settlers). In all cases, the result is the same. This indicates that the effect of high-skilled settlers is not driven by the possibility that those places also have greater diversity of occupations.

Table A.3: Fractionalization with occupations

Dependent Variable: Log. Night Light 2010 per capita

	(1)	(2)	(3)	(4)
% with Skills	0.407***		0.424***	
	[0.132]		[0.118]	
Frac with Occupations (Settlers with Skills)	-0.265	-0.260		
	[0.203]	[0.216]		
Frac with Occupations (All Settlers)			-0.165	-0.113
			[0.144]	[0.162]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes
Observations	123	123	123	123
R-squared	0.301	0.276	0.293	0.267

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

Results are also not affected if we include the fractionalization of place of origin of the settlers (see Table A.4).

Table A.4: Fractionalization with place of origin

Dependent Variable: Log. Night Light 2010 per capita

	(1)	(2)	(3)	(4)
% with Skills	0.389***		0.387***	
	[0.131]		[0.132]	
Frac with Place of Origin (Region)	-0.135	-0.172		
	[0.110]	[0.151]		
Frac with Place of Origin (Province)			-0.142	-0.174
			[0.116]	[0.144]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes
Observations	123	123	123	123
R-squared	0.292	0.270	0.293	0.271

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FE. The places of origin are grouped into 18 regions and 49 provinces. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

Ladero Quesada (1980), who analyzes the structure of Spanish society circa 1500, also explains the importance of different occupations at that time. Therefore, another question that could be raised is whether all highly skilled occupations played the same role in the development of these early settlements. To analyze this question, we constructed the main variable of interest, the percentage of highly skilled occupations, considering each time one of the different types of high-skilled occupations as being low skilled to determine whether there is any important occupation that drives the results. The highly skilled occupations that we suppress each time are listed in section 3. Table A.5 presents the results. In column 1, we subtract the category of "local Business" in columns 2 and 3 "college" and "other high educ". In columns 4 and 5, we exclude "merchants" and "bankers and lenders". In column 6, we remove "clergymen", and in columns 7 and 8, we exclude "artists" and "royal officials". The results in nearly all columns from 2 to 8 are robust. The only group of settlers that seems to affect the significance of the main variable are the settlers whose skills are related to

occupations whose main characteristic was to run a "local business". This will become very important when we explain the main mechanism and the persistent effects.

TABLE A.5: Subtracting occupation categories

Dependent Variable: Log. Night Light 2010 per capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% with Skills (subtracting: entrepreneurs)	0.144							
	[0.121]							
% with Skills (subtracting: college)		0.396***						
		[0.115]						
% with Skills (subtracting: other high educ)			0.475***					
			[0.116]					
% with Skills (subtracting: merchants)				0.404***				
				[0.124]				
% with Skills (subtracting: bankers and lenders)					0.405***			
					[0.129]			
% with Skills (subtracting: clergymen)						0.326**		
						[0.122]		
% with Skills (subtracting: artists)							0.388**	
							[0.135]	
% with Skills (subtracting: royal officials)								0.247*
								[0.141]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	123	123	123	123	123	123	123	123
R-squared	0.269	0.289	0.301	0.290	0.289	0.280	0.287	0.276

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. The category of royal officials includes administrators, nobility and military. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

Finally, we perform some analysis to provide more evidence that supports our identification strategy. Although we believe that there exists much historical evidence on the randomness of the distribution of settlers across settlements, we want to reinforce this idea by conducting a randomization inference. Specifically, we estimate the baseline regression (column 10 of Table 3) using randomization inference to assess whether the effect of the variable of interest (percentage of settlers with high skills) is unlikely to be observed by chance. We perform permutation tests on the basis of Monte Carlo simulations and compute the corresponding p-value after randomly permuting 500 times the percentage of settlers with skills. The coefficient remains statistically significant at the 5 percent level, both when permuting the percentage of settlers with high skills without imposing any strata and when permuting it within exogenous cells of 10 degrees x 10 degrees.

Moreover, we also consider placebo distribution of effects. In particular, we perform simulations that randomly assign the location of all settlers in our data to the 123 settlements. We then compare the findings with our results. In the first simulation exercise, we obtain the individuals shuffled randomly across the 123 cities while maintaining the total number of individuals in each city as in the original historical data. We run this simulation 500 times, so we can compute the percentage of settlers with skills at the city level from each of the 500 simulations. Using the baseline

specification (column 10 of Table 3), the coefficient on the simulated percentage of settlers with skills is positive and statistically significant in only 7.8 percent of regressions (39 out of 500 regressions). In the second simulation exercise, we obtain the individuals shuffled randomly across the 123 cities while allowing the total number of individuals in each city to vary. In this case, we also find that the coefficient on the simulated percentage of settlers with skills is positive and statistically significant in a similar percentage of regressions.[42]

[42]If we conduct the same analysis but include all the settlers, we obtain similar results.

C Online Appendix – Evidence on Alternative Mechanisms

Are there alternative mechanisms that could explain the persistence of skills over time in Latin America? While the literature on the persistence of large historical events often analyzes the education channel, our data are difficult to match with the level of instruction, leading us to use occupations and their associated skill levels. We do, however, check whether the variance in skills across the original locations could have led to persistent differences in education. Indeed, the latter may indicate that the unobservable differences in education, proxied by occupations, are the main factor explaining the persistence.

As a preliminary exercise, we explore the relationship between Jesuit missions and early settler skills. The recent literature suggests a connection between the location of Jesuit establishments and current levels of education.[43] We assess whether places with a greater proportion of skilled settlers are more closely correlated with Jesuit missions than those with fewer skilled settlers.[44] The main source of information on Jesuit presence in our area of analysis is Hamy (1892), which includes, among others, a catalog of all Jesuit establishments located around the world between 1540 and 1773. While the original edition is preserved in the Roman Jesuit Archives, Boston College has an accessible digitalized version. For each address, the catalog provides the name (in Latin and French) and kind of the establishment and the Jesuit province to which it belongs. The establishments are classified into five categories: *Collège* (school), *Rèsidence* or *Maison de probation* (housing), *Sèminaire* (higher education) and *Mission*.

In most cases, the name is sufficient for identifying the location of the Jesuit establishment, as it commonly coincides with that of the colonial settlement. For those that match our original settlements, we use the coordinates already collected. Otherwise, we employ the coordinates of contemporary city locations provided by Google Maps.

We define two different variables: a dummy that takes a value of 1 if there is a

[43] See Valencia (2019) for the case of Latin America and Nunn (2014) for nineteenth-century Christian missions in Africa.

[44] Several seminar participants wondered whether Jesuit missions might have been located close to high-skilled populations.

Jesuit presence within a 20 km buffer from the colonial settlement and the log of the distance, in km, to the closest Jesuit location.

Columns 1 and 2 of Table A.6 analyze the relationship between skills and Jesuit missions. We find no relationship between the proportion of high-skilled settlers and Jesuit presence. The same is true if we consider the distance to the missions in the buffer area.

It is also possible to directly study the relationship between the proportion of skills among the original settlers and contemporary education. To this end, we employ two variables from IPUMS, linking each colonial settlement to a unique second-level subnational unit and using the name and coordinates of contemporary municipalities.

The variable *Years of schooling* indicates the average number of years of schooling to reach the highest grade/level of education completed among males aged 18 to 55.[45] The variable *education* records the fraction of the population at least 15 years of age that has not completed basic education (elementary, primary and secondary school) according to modern population censuses.

Columns 3 (Years of schooling) and 4 (Education) in Table A.6 show that the proportion of highly skilled settlers among the original colonizers is not a statistically significant determinant of current levels of education in these locations.

TABLE A.6: Mechanisms – Jesuit missions and education

	Dependent Variable:			
	Jesuit Mission Dummy (17th Century) (1)	Log Dist Jesuit Mission (17th Century) (2)	Years of Schooling (IPUMS) (3)	Education (Modern Census) (4)
% with Skills	0.036	1.914	-0.703	-0.031
	[0.078]	[1.581]	[0.454]	[0.052]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes
Observations	123	123	103	101
R-squared	0.238	0.252	0.240	0.435

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. The dependent variables are as follows: the Jesuit mission dummy, which takes a value of 1 if there is Jesuit presence within a 20 km buffer from the colonial settlement, 0 otherwise; log of the distance in km to the closest Jesuit location; mean years of schooling among the male adult population (IPUMS); and percentage of the population at least 15 years of age that has not completed basic education (elementary, primary and secondary school) from official population censuses. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

Another potential mechanism of persistence is an agglomeration effect[46] induced by high-skilled settlements attracting more people and thus becoming more densely populated than locations with low-skilled colonizers. Employment density is positively related to productivity at the regional level; the same effect could be at work at

[45] We top coded this variable at 18 years of schooling.

[46] See Rocha et al. (2017)

the local level. Using data on population density from three different sources, the Population Density Grid from the Gridded Population of World V. 3, Census and LANDSCAN, we find no evidence of this mechanism.

To analyze this mechanism, we check the effect of population density in 2000 using the Population Density Grid from the Gridded Population of the World (GPW, v.3), which provides information on population density at 2.5 arc-minute resolution for 1990, 1995, and 2000.[47] The population density corresponds to the 20 km buffer around the original colonial settlement. We find no evidence for this mechanism.

The same result is obtained when the dependent variable is the total population in 2010 from LandScan or when we use the total population from the census. The results are shown in columns 1, 2 and 3 of Table A.7.

We also searched for information on the historical population of the original settlements. Traditional historical sources are not useful for this exercise because those sources provide information for large urban centers. Fortunately, Lopez de Velasco (1573) provides information on inhabitants of these early settlements. We have collected this information. We were able to match 78 places. The results are in line with the results using contemporaneous data.

TABLE A.7: Mechanisms – Agglomeration

	Dependent Variable:			
	Log Pop Density (GPW) (1)	Log Pop (LandScan) (2)	Log Pop (Census) (3)	Log Pop in 1570 (4)
% with Skills	-0.631**	-0.455	-0.662	-0.413
	[0.243]	[0.321]	[0.401]	[0.375]
All controls column 10 of Table 3	Yes	Yes	Yes	Yes
Observations	123	123	123	78
R-squared	0.191	0.183	0.143	0.415

Notes - Robust standard errors clustered at grid cells of 10 degrees × 10 degrees. The vector of controls from column 10 of Table 3 includes geographic (log distance to coastline and log ruggedness index) and climate (log average temperature from 1961-1980 and log average precipitation from 1961-1980) variables, log distance to river, log distance to lake, fertile soil, latitude, caloric suitability, malaria (average endemicity), year of first settler, and colonization area FEs. The dependent variables are as follows: (1) log contemporary population density within a 20 km buffer from the colonial settlement (year 2000 data from the Gridded Population of the World, GPW); (2) log contemporary population within a 20 km buffer from the colonial settlement (year 2010 data from LandScan); (3) log contemporary population from official population censuses; and (4) log number of *vecinos* in 1570 from Lopez de Velasco. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

[47] Center for International Earth Science Information Network (CIESIN), Columbia University and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World Version 3 (GPWv3): Population Density Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw (downloaded in February 2018)

D Online Appendix – Data sources and definitions

D.1 Main Dependent Variable

$$Log\ (Night\ Light\ per\ capita) = \log\left(\frac{0.1 + Total\ Night\ Light}{0.1 + Total\ Population}\right)$$

Total Night Light. Sum of light intensity values within a 20 km buffer around the colonial settlement for 2010. Source: authors' computation using cloud-free satellite night light coverage (30 arc-second raster data from satellite F18) of the DMSP-OLS Nighttime Lights Time Series (NOAA's National Geophysical Data Center).

Total Population. Sum of population counts within a 20 km buffer around the colonial settlement for 2010. Source: authors' computation using the LandScan 2010 High Resolution Global Population Data Set (30 arc-second raster data), copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory.

D.2 Settlers and Skills

Source: authors' computation after digitizing the information in the first and second volumes of Boyd-Bowman (1956, 1964).

Percentage with High Skills. Proportion of highly skilled settlers at the end of the period.

TABLE A.8: Descriptive statistics: occupations

	Number of cities with at least one settler	Percentage over all settlers	Percentage over settlers with skill
Local business	52	9.45%	20.70%
Royal officials	94	18.76%	41.07%
College	51	4.67%	10.23%
Other high educ	54	5.16%	11.29%
Clergymen	39	2.63%	5.75%
Merchants	26	1.97%	4.30%
Bankers and lenders	4	0.07%	0.14%
Artists	17	0.40%	0.87%

FIGURE A.1: Share of settlers with skills. *Source: Own elaboration.*

Majority with high Skills. Dummy variable taking value 1 if more than 75% of settlers are high skilled, 0 otherwise.
Log Total Early Settlers. Natural log of the total number of early settlers.

D.3 Initial Conditions and Audiencias

D.3.1 Distance to Port and to Seville

Log Distance to Port. Natural log of the distance (great circle distance, km) from the colonial settlement to the closest port during the colonial period (Acapulco, Panama, Veracruz, Cartagena, or Callao). Source: authors' computation.
Log Distance to Seville. Natural log of the distance (great circle distance, km) from the colonial settlement to the port of Seville. Source: authors' computation.

D.3.2 Audiencias

Source: authors' computation using information on audiencias from Morales Padrón (1988).

Log Distance to Audiencia. Natural log of the distance (great circle distance, km) from the colonial settlement to the closest audiencia. We only consider audiencias established before 1540. The first audiencia was Santo Domingo (1511),– and its jurisdiction included all territories discovered until the foundation of subsequent audiencias. The second and third audiencias were Mexico (1527) and Panama (1538), respectively. The audiencia of Panama is a special case because, in practice, it was established in 1540, but it was abolished in 1542 and reestablished in 1563. We focus on the audiencias of Santo Domingo and Mexico.

No Audiencia at Foundation. Dummy taking value 1 if the first settler arrived before 1511, 0 otherwise.

Years without Audiencia. Number of years without an official audiencia.

FIGURE A.2: Audiencias established before 1540

Audiencia FE (16th century). The audiencias at the end of the 16th century correspond to (1) Santo Domingo, (2) Mexico, (3) Panama, (4) Lima, (5) Guatemala, (6) Guadalajara, (7) Santa Fe de Bogota, (8) Charcas, (9) Quito, and (10) Chile.

FIGURE A.3: Audiencias at the end of the 16th century

D.3.3 Colonization Area and Route

Colonization Area FE. The colonization area fixed effect accounts for the three main areas of colonization during the conquest. The first area corresponds roughly to La Española, Cuba, Puerto Rico, Jamaica, and central Mexico. The second area covers the rest of central Mexico, as well as some regions of Central America and the north coast of South America, and the third area corresponds roughly to South America and northern Mexico. Source: Morales Padrón (1988).

Colonizer Route FE. colonizer route or penetration line (expedition paths followed by the conquerors at the time of discovery). Source: authors' computation after georeferencing all discovery routes from the *Atlas del Descubrimiento de América y Oceanía* (Montana, 1943a,b).

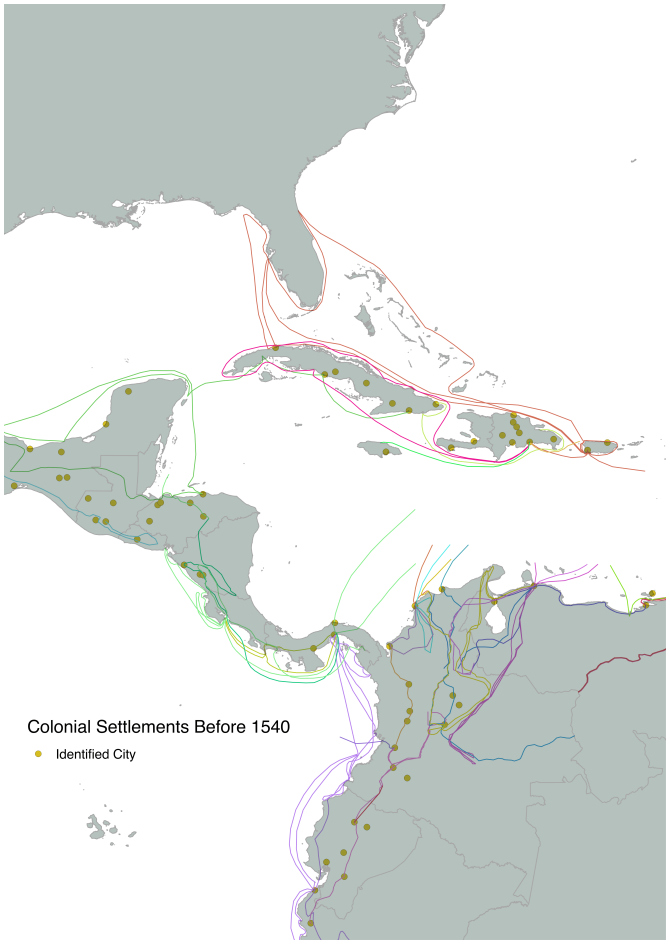

FIGURE A.4: Details of penetration lines in Central America

Colonizer Route Order. Chronological order of settlements along the same colonizer route. Source: authors' computation using Montana (1943a,b).

D.3.4 Precolonial variables

Precolonial Pop Density. Precolonial pop density over 100. Source: Maloney & Valencia (2016) provide data on precolonial population density for first-level administrative units (NUTS1). We assign an administrative unit to each settlement. The data are available for 98 settlements located in 66 administrative units (49 with a unique settlement and 17 with more than one settlement).

Temples Dummy. Dummy taking value 1 if there is at least one precolonial temple within a 20 km buffer around the colonial settlement, 0 otherwise. Source: authors' computation using data collected by Mayshar et al. (2020) from the *Archaeological*

Atlas of the World.

Hierarchy Dummy. Dummy taking value 1 if the precolonial tribe in the territory of the colonial settlement had two, three or four levels of jurisdictional hierarchy beyond the local community level, 0 otherwise. See Online Appendix E for details on the matching procedure and how we collect new information on state development for precolonial tribes.

Murdock FE. Fixed effect accounting for the tribe in the territory of the colonial settlement prior to the Spanish conquest. See Online Appendix E for details on the matching procedure.

D.4 Geography and Climate Variables

Log Temperature. Natural log of average temperature (1961-1990) within a 20 km buffer around the colonial settlement. Source: authors' computation using the CRU TS3.10 Dataset (30 arc-minute).

Log Precipitation. Natural log of average precipitation (1961-1990) within a 20 km buffer around the colonial settlement. Source: authors' computation using the CRU TS3.10 Dataset (30 arc-minute).

Log Ruggedness Index. Natural log of the average ruggedness index within a 20 km buffer around the colonial settlement. Source: authors' computation using grid-cell-level data (30 arc-seconds) on terrain ruggedness created by Nunn & Puga (2012).

% Fertile Soil. Average percentage of fertile soil within a 20 km buffer around the colonial settlement. Source: authors' computation using grid-cell-level data (30 arc-second) from the FAO Digital Soil Map of the World (DSMW).

Log Distance to River. Natural log of the distance (great circle distance, km) from the colonial settlement to the closest river. Source: authors' computation using data from version 10 of the Seamless Digital Chart of the World (SDCW).

Log Distance to Lake. Natural log of the distance (great circle distance, km) from the colonial settlement to the closest lake. Source: authors' computation using data from version 10 of the Seamless Digital Chart of the World (SDCW).

Log Distance to Coastline. Natural log of the distance (great circle distance, km) from the colonial settlement to the closest coastline point. Source: authors' computation using coastlines from version 3.7 of the Gridded Population of the World

(GPW) project.

Caloric Suitability. Average caloric suitability index (potential caloric yield attainable based on the crops that were suitable for cultivation in the pre-1500 period) within a 20 km buffer around the colonial settlement. Source: authors' computation using grid-cell-level data (5 arc-minute) from Galor & Özak (2016).

Agricultural Suitability. Average agricultural suitability index (fraction of land that is suitable to be used for agriculture) within a 20 km buffer around the colonial settlement. Source: authors' computation using grid-cell-level data (30 arc-minute) from Ramankutty et al. (2002).

Malaria (Average Endemicity). Average malaria endemicity within a 20 km buffer around the colonial settlement. Source: authors' computation using grid-cell-level data (30 arc-seconds) from Gething et al. (2011).

D.5 Mechanisms

Efficient Cropped Land. Average proportion of land area used as cropland (land used for the cultivation of food) over the land that is suitable for cultivation within a 20 km buffer around the colonial settlement. Source: authors' computation using data from Ramankutty et al. (2002) and Ramankutty et al. (2008).

Technological Frontier − Cereals. Average ratio of actual over potential yield for cereal crops within a 20 km buffer around the colonial settlement. Source: authors' computation using data from the GAEZ Global Agro-Ecological Zones project.

Market-Oriented Agricultural Workers. Fraction of the male adult population (18-55 years old) working as market-oriented skilled agricultural workers. Source: IPUMS. Variable OCCISCO provides the person's primary occupation, coded according to the major categories in the International Standard Classification of Occupations (ISCO-88). We use information from one modern census for each country following the selection of years in Acemoglu & Dell (2010).

Roads Per Capita. Natural log of total road length (km) divided by population within a 20 km buffer around the colonial settlement. Source: authors' computation using data on contemporary roads (all primary and secondary roads) from version 10 of the Seamless Digital Chart of the World (SDCW) and data on population from the LandScan 2010 High Resolution Global Population Data Set.

Economic Diversification in 1620. Sum of the number of different economic

activities (agriculture, vegetables, livestock, manufacturing, mining, craft products) in the settlement by 1620. Source: we codified information on the economic activities of each of the settlements from Lopez de Velasco (1573) and Vazquez de Espinosa (1620). The original sources that we found do not provide information on quantities but only whether they produce agriculture, vegetables, livestock, manufacturing, mining and crafts products (agricultura, hortalizas, ganaderia, manufactura, mineria y artesania). We found 97 of our 123 settlements in these books. We read the description of the place and collected information on the products each settlement produced.

Log Distance to Trade Route in 1620. Natural log of the distance (great circle distance, km) from the colonial settlement to the closest trade route by 1620. Source: authors' computation after georeferencing all routes in Van Bath (2010)'s *Hiaspanoamérica en torno a 1600*. Van Bath (2010) provides a map of the land trade routes between 1574 and 1628. This map was constructed based on the information from Lopez de Velasco (1573) and Vazquez de Espinosa (1620).

Dummy for Trade Routes in 1620. Dummy taking value 1 if there is a trade route within a 20 km buffer around the colonial settlement. Source: authors' computation using the map of colonial trade routes in Van Bath (2010).

Having an active entrepreneurial class. Share of respondents who agreed with "What is the most important thing in order for a country to develop successfully? Having an active entrepreneurial class". Source: Latinobarometer. Each individual in the Latinobarometer is associated with a specific location, indicating their city of residence. We use this information to match the individuals with our original locations. For the maching procedure we georeferenced all the locations found in the Latinobarometer.[48] Once we extracted the latitude and longitude of each location, we matched these locations with the original settlements using two methods. We matched two places if their coordinates coincide exactly, and we can identify them by name as being the same. We regard as a positive match those places where the Latinobarometer location falls within the 20 km buffer from the original settlement city. Upon completion of the automatic match, we also conducted a manual check. Specifically, we checked that the coordinates of all matched cities are correct to discard false positives. In a second step, we assessed whether those cities that were not matched with any location in the Latinobarometer could be matched manually. This

[48]We employed Google maps API, using the name of the city, region, and country as search parameters.

may occur when the name in the Latinobarometer is very different from that in Google maps. In total, we were able to match 75.6 % of the locations using the buffer method and 67.5% using the exact method.

Jesuit Mission Dummy (17th Century). Dummy taking value 1 if there is a Jesuit mission within a 20 km buffer around the colonial settlement, 0 otherwise. Source: authors' computation after georeferencing information on Jesuit presence in Hamy (1892).

Log Distance to Jesuit Mission (17th Century). Natural log of the distance (great circle distance, km) from the colonial settlement to the closest Jesuit mission. Source: authors' computation after georeferencing information on Jesuit presence in Hamy (1892).

Years of Schooling (IPUMS). Mean years of schooling among the male adult population (18-55 years old). Source: IPUMS. We use information from one contemporary census for each country following the selection of years in Acemoglu & Dell (2010).

Education (Census). Percentage of the population at least 15 years of age who has not finished basic education (elementary, primary and secondary school). Source: We use information from one modern census for each country following the selection of years in Acemoglu & Dell (2010).

Log Population Density (GPW). Natural log of average contemporary population density within a 20 km buffer around the colonial settlement. Source: authors' computation using grid-cell-level data (2.5 arc-minute) from version 3 of the Gridded Population of the World (GPW) project.

Log Population (LandScan). Natural log of contemporary population within a 20 km buffer around the colonial settlement. Source: authors' computation using the LandScan 2010 High Resolution Global Population Data Set (30 arc-second raster data).

Log Population (Census). Natural log of contemporary population. Source: population census. We use information from one modern census for each country following the selection of years in Acemoglu & Dell (2010).

Log Population in 1570. Natural log of number of inhabitants (*vecinos*) in 1573. Source: We collected this information from the book of Lopez de Velasco (1573). Matching could be performed for 97 colonial settlements.

E Online Appendix – Precolonial Tribes and State Development

Information on the level of state development of precolonial tribes predominantly comes from G.P. Murdock (1967)'s *Ethnographic Atlas*, which contains ethnographic information for 1,265 tribes or ethnic groups. First, to match these data to the colonial settlements in our dataset, we first used Murdock (1951, 1960)'s map of native peoples in the Americas and assigned to each settlement the name of the native group whose area in the map contains the colonial settlement. Following this stage 111 of our 123 settlements have the name of an indigenous group mapped to them, with there being 48 unique names in total. We were unable to assign a name to the remaining 12 settlements due to gaps in Murdock's map. Following this we attempted to match the group names in Murdocks's map to the Ethnographic Atlas. We were able to find a direct match for 17 ethnic names. These 17 ethnic tribes, includes 63 settlements. For the remaining tribes names, we used The Handbook of South American Indians (Steward, 1946), supplemented with The Cambridge History of the Native Peoples of the Americas (Trigger et al., 1996) to find alternate names for the same groups so that a match could be made with the Ethnographic Atlas. Following this step, 13 more names, (which includes 22 additional settlements) were matched to a group in the Ethnographic Atlas. It is worth noting that due to some groups in Murdock's map being subgroups of larger groups detailed in the Ethnographic Atlas these initial 13 names were remapped to only 9 unique names during this stage of the process.

To find a match for these final 38 unmatched settlements (26 settlements corresponding to 18 tribe unique names and 12 with no associated name) we made use of the centroids for ethnic groups detailed in the Ethnographic Atlas and matched the settlements to the group whose centroid was closest. As a result all 38 settlements were matched with Ethnographic Atlas entries, there being 22 unique ethnic groups, 7 of which other settlements had been matched to in previous steps and 15 of which were new. This is equivalent to the method used in Alsan (2015) of constructing Thiessen polygons to assign areas to groups based on their centroid. The average distance from a settlement to the closest centroid was 190 km, which is in keeping with the distances used for the slightly more common approach of drawing buffer zones around centroids (for example, Alesina et al. 2013 uses a 200 km radius). After this process all 123 settlements were matched to a group in the ethnographic atlas, with there being 41

unique groups.

We follow the literature in using the number of hierarchies or jurisdictional levels beyond the local community as a measure of state development (variable v33 in the Ethnographic Atlas). However, this variable was missing for 10 of the native groups (covering 18 settlements) in the Ethnographic Atlas. To fill this gap, we compared the descriptions of these groups in several sources to the descriptions of groups with correct Ethnographic Atlas data and to the description of the hierarchy variable given in the Ethnographic Atlas. As with the name matching procedure, the primary text relied upon was The Handbook of South American Indians. In addition, The Cambridge History of The Native Peoples of the Americas was used for the Lacandon, Mixtec, Otomi, Tepehuan, and Zapotec. The Handbook of South American Indians is a collection of 6 volumes (and a 7th index volume) of academic works in ethnography concerning the native peoples of South America. The Cambridge History of The Native Peoples of the Americas is a similar work in that it provides a comprehensive overview of most indigenous groups within a given region (in this case, Mesoamerica). While it tends to focus on historical details, as opposed to ethnography, it still provides adequate detail on the social organization of the ethnic groups, and we used it to reasonably infer a value for our hierarchy variable.

We ultimately have 41 different ethnic groups in our territory of interest, all matched to a colonial settlement. We use this information to include fixed effects accounting for the precolonial tribe. Of the 41 tribes, 18 have no levels of hierarchy beyond the local community (matched to 32 colonial settlements), 11 have one level (matched to 23 settlements), 6 have two levels (matched to 50 settlements), 5 have three levels (matched to 11 settlements), and the 1 remaining tribe has four levels (matched to 7 settlements).