Founders of Prosperity: First Settlers and the Origins of Regional Development*

José G. Montalvo† Marta Reynal-Querol‡

Abstract

This paper introduces and analyzes a dataset documenting the first Spanish settlers to Latin America (1492–1540), comprising 19,831 individual settlers across 123 settlement locations with information on origins, destinations, and occupations spanning 274 distinct categories. These records, preserved through Spain's administrative requirements for emigration documentation and centralized port control through Seville, were systematically organized by historical scholars and have now been digitized and geocoded for economic analysis. The dataset reveals substantial spatial heterogeneity in settlers' skills composition. Critically, this spatial variation was not systematically predicted by environmental characteristics. Using comprehensive matrices linking Spanish origins with American destinations, we test whether settlers located in regions with climatological and geographical similarities to their origins. Environmental similarity explains virtually none of the variation in settlement patterns, and settlers from each province exhibit high fractionalization across destinations . These findings show that settlement patterns reflected uncertain conditions of exploration rather than systematic preferences.

*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 PID-2023-150768NB-I00, 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 Programe for Centers of Excellence in R&D (SEV-2024). This paper is a revised version of the BSE Working Paper 1189 of July 2020.

†Universitat Pompeu Fabra-IPEG and Barcelona GSE.

‡Universitat Pompeu Fabra-ICREA-IPEG and Barcelona GSE. Address: Ramon Trias Fargas 25-27, 08005 Barcelona. E-mail: marta.reynal@upf.edu

Using this novel dataset, we investigate whether spatial heterogeneity in settler human capital influenced long-run regional development, using nighttime luminosity in 2010 as a proxy for per capita income. We find that locations originally characterized by higher proportions of skilled settlers today exhibit substantially higher economic development. This relationship remains robust after controlling for geography, climate, land quality, disease environment, pre-colonial conditions, and institutional arrangements. Historical evidence from 1573–1620 reveals that high-skilled settlements developed more market-oriented economies early on, while contemporary data show they exhibit superior agricultural productivity, higher technology adoption, better infrastructure, and stronger entrepreneurial attitudes. These mechanisms explain the remarkable persistence of initial settler composition effects across nearly five centuries. The paper shows that comprehensive first-settler data enables a novel approach to how founding population skills shaped American development trajectories.

JEL classification numbers: O10, O15

> *"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

Recent scholarship on comparative economic development has increasingly investigated the persistent impact of major historical events, with particular attention to colonialism and its enduring effects on the long-run prosperity of affected nations. A central question in this literature concerns how early colonial experiences generate development trajectories that persist for centuries: What features of the initial colonization process determine whether regions emerge as economic centers or remain economically marginalized? Among the various explanations advanced—institutional arrangements, resource endowments, geographical characteristics, indigenous populations, disease environments—relatively little scholarly attention has focused on a fundamental dimension: the skills composition of the earliest settlers. This gap in the literature exists not due to a lack of theoretical interest but rather due to practical constraints: comprehensive data on early settlers, their characteristics, and their settlement locations have been unavailable to researchers.

This paper addresses this significant lacuna by introducing and analyzing a new dataset documenting in detail the first wave of Spanish settlers to Latin America (1492–1540). The value of this dataset merit emphasis before discussing our substantive findings. We have compiled comprehensive administrative records from the General Archive of the Indies in Seville documenting 19,831 individual settlers across 123 settlement locations, including their names, Spanish cities of origin, American destinations, and occupations. This dataset represents decades of archival work by historical scholars—most notably Cristóbal Bermúdez (1940–1946) and Peter Boyd-Bowman (1956, 1964)—who systematically matched passenger lists (*Libros de Asientos de Pasajeros*) with verification documents from the Americas (*informaciones y licencias*, foundation letters, contracts, payrolls, administrative records, *capitulaciones*). The result is a uniquely detailed record of early American colonization that we have now digitized, geocoded, and made available for empirical analysis.

The creation and analysis of this dataset provide several critical advantages over previous historical investigations. First, the comprehensiveness of the settler records is extraordinary: for each of the approximately 20,000 settlers included, we possess not merely a name but also their Spanish city of origin, American destination, declared occupation, and often information about family members and subsequent activities in the Americas. This level of detail—spanning from approximately 1492 through 1540—has never before been systematically organized for economic research. The occupational diversity evident in the data is remarkable: settlers pursued 274 distinct occupations, ranging from college-educated professionals (physicians, notaries, lawyers) and merchants to artisans and skilled craftspeople, as well as peasants and domestic servants. This granular occupational information allows for the classification of human capital along multiple dimensions—formal education (university-trained physicians and lawyers versus those with apprenticeship training), entrepreneurial orientation (merchants and local business owners versus servants and wage laborers), and specialized skills (notaries and interpreters versus general laborers).

Second, the administrative system that generated these records was specifically designed to produce the comprehensive documentation we now possess. The Spanish Crown established strict restrictions on emigration to the Americas—prohibiting Muslims, Jews, heretics prosecuted by the Inquisition, converts from Islam or Judaism, Roma people, unaccompanied women, and married individuals separated from spouses. To enforce these restrictions, prospective emigrants were required to provide documentation to officials of the *Casa de la Contratación de las Indias* (House of Trade of the Indies) established in Seville in 1503. Required documentation included baptismal certificates, marriage licenses, and testimonials affirming Christian lineage (collectively termed *informaciones y licencias*). Significantly, all transatlantic travel departed through a single legal port (Seville until 1668), which simplified the administrative task of recording passenger movements. These circumstances—religious exclusions, documentation requirements, port monopoly, and Crown interest in monitoring colonization—generated an extraordinarily complete historical record of early American migration and settlement. The preservation of these documents in the General Archive of the Indies provides a uniquely rich source for economic historical research.

Third, the data allow for a systematic study of settlement location patterns for the first time in the literature. Previous historical research has analyzed the conquistadors, the major military campaigns, and broad patterns of territorial expansion.

However, prior scholarship has rarely attempted to map the spatial distribution of settlers, classify them by skill level at each location, and document how these spatial patterns of human capital distribution vary across the Americas. Our geolocation of historical settlements using contemporary administrative division data and historical maps—particularly those produced by cosmographer Juan López de Velasco (1573)—enables the construction of settlement-specific variables measuring the skill composition of founding populations. Figure 2 in Section 3 presents the first spatial visualization of settler skill heterogeneity across early colonial settlements, revealing substantial geographical variation: some settlements were founded predominantly by low-skilled laborers, while others had concentrations of physicians, merchants, artisans, and other skilled professionals. As far as we know this spatial heterogeneity in settler composition had never been systematically documented before.

The availability of comprehensive data on the first settlers presents not only opportunities but also theoretical challenges. Conventional migration literature typically predicts the location of new migrants using information about previous migrants from the same origin who have already settled in destination areas. The seminal work by Card (2001) and subsequent studies construct instrumental variables based on historical settlement patterns, exploiting the tendency of migrants to locate in areas where earlier migrants from their origin regions had established communities. This "chain migration" or "network effect" framework posits that migrants follow established pathways, with prior settlers creating information channels, providing assistance to newcomers, and reducing migration costs through ethnic enclaves and support networks.

However, this conventional theoretical framework cannot apply to our context because we are analyzing the very first settlers—those who arrived when no prior Spanish communities existed in the Americas. There were no established migration networks, no information channels about specific destinations, and no ethnic enclaves to facilitate settlement. The challenge thus becomes: How can we understand the location patterns of the first settlers when the primary mechanism emphasized in migration literature (following previous settlers) is necessarily absent?

To address this challenge, we turn to an alternative theoretical framework based on geographical and climatological similarity between origins and destinations. The hypothesis is that when migrants lack information about destinations through established networks, they may select locations whose environmental characteristics

resemble their regions of origin. This conjecture has intuitive appeal: settlers from temperate Spanish regions might preferentially settle in temperate American zones, those from arid regions might favor similar climates, and settlers from mountainous provinces might choose highland settlements. Such preferences could arise from several mechanisms: familiarity with agricultural practices suited to particular climates, adaptation to specific temperature and precipitation patterns, knowledge of construction techniques appropriate for certain environments, or simply psychological comfort with familiar environmental conditions.

To investigate this hypothesis empirically, we constructed comprehensive matrices linking Spanish origins with American destinations through geographical and climatological measures. Specifically, we created a matrix of Spanish provinces $\times$ Latin American sites (6,027 entries) and an even more granular matrix of Spanish villages $\times$ Latin American sites (145,684 entries). For each origin-destination pair, we calculated absolute differences in key environmental characteristics: percentage of fertile soil, terrain ruggedness, average temperature, and average precipitation. These measures capture multiple dimensions along which environments might be perceived as similar or dissimilar. If settlers exhibited preferences for environmentally similar destinations, we would expect systematic relationships between these distance measures and actual settlement patterns—with settlers from each Spanish origin concentrating in American destinations whose geographical and climatological characteristics closely matched their origin regions.

The empirical analysis examines whether environmental similarity predicts settlement patterns by regressing the percentage of settlers from origin "g" traveling to destination "c" on the absolute distances in fertile soil, ruggedness, temperature, and precipitation between that origin-destination pair. Strikingly, we find virtually no relationship between environmental similarity and settlement patterns. For the village-level analysis (145,684 observations), the R^2 is merely 0.003, indicating that environmental similarity explains less than one-third of one percent of the variation in settlement destinations. Even at the more aggregated province level (6,027 observations), the explanatory power rises only to $R^2 = 0.011$. These results suggest that climatological and geographical similarity between Spanish origins and American destinations played negligible roles in determining where settlers established themselves.

This finding has important implications. First, it shows that the absence of prior

settler networks cannot be replaced by environmental similarity as a predictive framework for settlement location. Second, it suggests that settlement patterns during the initial colonization phase were not systematically determined by characteristics that settlers could have anticipated or evaluated prior to departure. Third, it reinforces the historical evidence that settlers confronted profound uncertainty about destination characteristics and that settlement locations reflected factors other than settler preferences for familiar environments—factors such as the geographical imprecision of *capitulaciones* (royal contracts)[1], the sequential discovery of territories through exploratory expeditions, the exhaustion of expedition resources, which forced settlement at reached points, and the opportunistic establishment of communities when circumstances permitted.

Moreover, we document high fractionalization in settlement patterns: settlers from each Spanish province dispersed across multiple American destinations rather than concentrating in a few locations. The median fractionalization index across Spanish provinces is 0.809, indicating that settlers from typical provinces spread themselves across numerous destinations. This dispersed pattern further supports the interpretation that settlement locations were not driven by coordinated decisions or systematic preferences but reflected the chaotic and uncertain conditions of initial exploration and conquest.

Given the availability of this novel dataset documenting the settlement locations and human capital composition of first settlers, and given that settlement patterns do not appear systematically related to predictable environmental factors, the following question arises: Did the skills of initial settlers influence the long-run economic development of their settlement locations? This question represents a straightforward and compelling application of the new data. If human capital matters for economic productivity and growth—a hypothesis supported by extensive literature in development economics and economic history—then examining whether the human capital of founders influenced subsequent regional development provides valuable evidence on the importance of this factor.

The temporal scope of this inquiry is particularly significant. The dataset documents settlers from 1492 to 1540, representing 48 years of the initial colonization phase. Contemporary development outcomes can be measured in 2010 using satellite nighttime luminosity data. This creates a temporal span of nearly five centuries separating

[1]The location of the granted land could turned to be in the middle of the Pacific

the initial settler shock from contemporary outcome measurement—an extraordinarily long duration that provides evidence of the persistence of early conditions. If initial settler composition still influences development outcomes after five centuries and across multiple institutional, technological, and economic transformations, this would constitute remarkable evidence of path-dependence and the lasting significance of early human capital endowments.

The spatial dimension of the analysis adds value. The dataset covers over one hundred distinct settlement locations spanning from present-day Mexico to Argentina—encompassing extraordinary environmental diversity (tropical rainforests, high mountain plateaus, temperate zones, semi-arid regions), different indigenous populations with varied levels of political organization and technological sophistication, and locations at vastly different distances from early centers of Spanish colonial authority. By examining how settler skill composition across these geographically diverse locations relates to contemporary development, we can investigate whether human capital effects operate consistently across heterogeneous environmental and social contexts or whether effects are context-dependent.

Furthermore, the exclusive focus on Spanish territories during a single colonization period offers an important advantage: the formal institutional framework was uniform (Spanish colonial law and administrative structures), and the temporal context was shared (same era, same technological possibilities for transportation and communication). This institutional uniformity allows examination of how settler characteristics influenced development without confounding variation in institutional arrangements across regions. While earlier settlers could not have anticipated institutions and economic activities that would develop after 1540, the fact that all settlements operated within Spanish colonial legal frameworks and that initial settlement occurred within a compressed historical period provides useful analytical tractability.

The data reveal that settlers during this initial colonization phase were not primarily soldiers, as popular historical narratives often suggest, but rather a diverse cross-section of Spanish society. Approximately 45% of settlers fall into high-skill categories: merchants, physicians, notaries, lawyers, pharmacists, artisans (weavers, carpenters, blacksmiths, tailors), bankers, clergy, local business owners, and skilled interpreters. The remaining 55% comprises low-skill workers: predominantly peasants and servants, including domestic servants and attendants to high-status migrants. This substantial proportion of skilled settlers reflects the functional requirements

of establishing settlements in unknown, isolated territories. Expeditions required self-sufficient communities capable of managing trade, providing medical care, maintaining legal and administrative records, constructing buildings and infrastructure, and developing economic activities beyond subsistence agriculture. The presence of physicians, notaries, merchants, and artisans among settlers reflects these practical necessities.

The question of whether these initial settler compositions influenced long-run regional development is therefore not merely of academic interest but addresses a fundamental dimension of colonial history: What role did the skills and capabilities of founding populations play in shaping the subsequent development trajectories of American settlements? Given the availability of data documenting settler locations and skills for the first time, investigating this question represents a natural and valuable contribution to understanding both colonial history and the long-run determinants of regional development.

This paper makes several contributions. First, we introduce and document an unprecedented dataset on the first Spanish settlers to Latin America, providing detailed information on settler locations and occupational composition for the initial colonization phase (1492–1540). This dataset could also enable future researchers to investigate numerous questions regarding early colonization, settler characteristics, and colonial development.

Second, we address a theoretical challenge inherent in analyzing first settlers. Conventional migration theories based on following previous settlers cannot be applied when no previous settlers exist. By constructing comprehensive matrices of environmental characteristics linking Spanish origins with American destinations and demonstrating that climatological and geographical similarity played negligible roles in settlement location, we establish that settlement patterns reflected factors other than systematic preferences or predictable environmental selection. This finding provides important context for interpreting the relationship between settler composition and development outcomes.

Third, we provide the first systematic analysis of how spatial heterogeneity in settler human capital across locations relates to long-run regional development in Latin America. By documenting that settlement locations originally characterized by higher proportions of skilled settlers today exhibit significantly higher levels of development, we demonstrate the enduring importance of initial human capital distributions. The

magnitude and persistence of these effects suggest that the skills and capabilities of founding populations shape regional development trajectories in ways that persist for centuries. Some papers have emphasized the long-run persistence of historical factors in shaping development outcomes. Studies by Squicciarini & Voigtländer (2015), Gennaioli et al. (2013), Nunn (2020), Valencia (2019), and Rocha et al. (2017) have documented that human capital shocks and historical conditions can generate economic effects persisting across decades or even centuries. Additionally, research on migration selectivity by Pérez-Artés (2024) has established that Spanish emigrants to the Americas during the sixteenth to eighteenth centuries exhibited positive selection in numeracy and human capital relative to Spaniards who remained in Spain. However, while Pérez-Artés (2024) establishes the positive selection of migrants relative to non-migrants, and while the literature has documented the persistence of historical factors in development, no prior study has systematically examined how the spatial distribution of settler human capital across settlement locations within the Americas has influenced long-run regional development.

Fourth, by analyzing the initial settler composition across geographically diverse settlements within a uniform institutional framework (Spanish colonial law), we provide evidence on the role of human capital in development that is not confounded by institutional variation. This evidence contributes to ongoing scholarly debates regarding the relative importance of institutions versus human capital in explaining economic divergence.

The article proceeds as follows. Section 2 provides historical background on the mechanics of conquest and settlement, explaining the organizational structures through which Spanish expansion proceeded. Section 3 presents the data, documenting the sources, the geolocation of settlements, the classification of settlers skills, and descriptive statistics on settler composition across locations. Section 4 examines the determinants of settler location patterns, investigating what factors influenced where different types of settlers established themselves. Sections 5 and 6 present applications of these data to the analysis of early human capital in the long rund development and potential mechanisms. Section 5 analyzes the relationship between initial settler skill composition and contemporary regional development. Section 6 investigates the mechanisms through which initial settler composition may have influenced long-run development, examining historical evidence on early economic organization (1573–1620) and contemporary patterns in agricultural productivity,

technological adoption, infrastructure development, and entrepreneurial attitudes. Section 7 concludes with implications of the findings for understanding colonial legacies and the role of human capital in long-run development.

2 The Conquest: Institutional Background

The Spanish conquest of the Americas was fundamentally an institutional process aimed at extending the sovereignty of the Castilian Crown over newly discovered territories and integrating their populations into the Christian order. Upon successful subjugation of a territory, a colonial administration was established under the leadership of the expedition's commander, who assumed formal authority with titles such as governor, *adelantado*, or marshal. Members of the expeditionary force—the *hueste*—were assigned subordinate administrative, military, or settlement responsibilities. Ballesteros (1985) argues that this hierarchical organization generated a well-defined system of incentives, aligning royal objectives of territorial expansion and evangelization with private pursuits of wealth accumulation, land acquisition, and social advancement.

Crucially, the acquisition of the Indies did not result from a centrally coordinated imperial strategy. Instead, it emerged from a decentralized process driven by private initiative operating under royal authorization. Conquest was undertaken by privately organized expeditions—commonly referred to as *huestes conquistadoras* or *huestes indianas*—composed predominantly of volunteers rather than conscripts. These forces did not constitute permanent royal armies and were not financed by the Crown. Rather, they functioned as privately funded enterprises subject to legal authorization, with the Crown retaining ultimate sovereignty and a share of the resulting gains.[2].

This arrangement rested on a legal framework established by the Alexandrine Bulls—most notably *Inter Caetera*—which recognized the Castilian Crown's exclusive claim over lands not governed by a Christian prince. In practical terms, Ballesteros (1985) argues that this framework meant that access to newly discovered territories was conditional on explicit royal approval, thereby formalizing conquest as a regulated private activity rather than an uncoordinated expansion.

[2]Ballesteros (1985)

2.1 Chronology of the Conquest

The conquest unfolded in distinct chronological phases, each characterized by specific patterns of expansion and institutional development. Broadly, the process extended from 1493—marked by Columbus's second voyage and the onset of permanent Spanish settlement—to approximately 1570, when Philip II formally declared the exploratory campaigns concluded. Over these seventy-seven years, the military and settlement dynamics that secured Spanish control over the Indies were consolidated.

The first phase, commonly designated as the Antillean stage, involved the gradual exploration and conquest of the Caribbean archipelagos. Expansion proceeded until Spanish expeditions reached *Tierra Firme*, the continental landmass bordering the Caribbean, corresponding to present-day Panama, Colombia, and Venezuela.

The second phase was characterized by the continental expansion of Spanish control. It encompassed the conquests of Mexico, Peru, and the New Kingdom of Granada (modern-day Colombia). Although routes of penetration were identified locally, legal authorization for these enterprises continued to be issued by the Crown in Spain.

The third phase consisted of interior expansion and the consolidation of territories secured during the second phase. Within this phase, the advance into the major southern river systems represents a particularly distinctive development.

These historical and chronological features define the institutional environment in which the *hueste indiana* operated.[3] Our analysis focuses on the first two phases, which encompass the initial processes of discovery, conquest, and settlement, preceding the subsequent consolidation and internal expansion from the primary colonial centers.

Within less than fifty years, Spanish settlers founded cities across an extensive area extending from Zacatecas in present-day Mexico to what is now Argentina. During this initial stage, settlements were highly isolated and emerged as independent nodes with minimal interconnection. Only after this early period of colonization (covering Phases 1 and 2) did settlers begin to explore surrounding territories (Phase 3), and from the 1540s onward, the establishment of new interior cities was typically organized from these original settlements. Our analysis excludes this later Phase 3 of expansion and concentrates exclusively on the first two phases of settlement.

[3] See next subsection.

2.2 The Organization of the *Hueste*

Individuals seeking authorization to explore and conquer new territories were required to obtain royal approval through colonial institutions such as the Council of the Indies or the *Casa de la Contratación*. Access to these bodies was often mediated by patronage networks, elite intermediaries, or court connections. Applicants were required to submit detailed plans and provide personal and financial guaranties.

Upon approval, the Crown issued a *Capitulación*, a contractual agreement granting the expedition leader the right—at his own expense—to conduct exploration, conquest, settlement, and pacification. The *Capitulación* specified the number of men the leader was authorized to recruit and conferred upon him the title of captain. Successful completion of the campaign could result in additional rewards, including appointment as governor, *adelantado*, or commander of the first fortress established in the conquered territory.

Once authorization had been obtained, the captain undertook the task of assembling the *hueste*. This process involved recruiting volunteers, raising capital, and securing provisions, weapons, ships, and supplies. In the case of maritime expeditions, additional costs included the acquisition or chartering of vessels, sailors' wages, and navigation provisions. Because expeditions were privately financed, conquest effectively operated as a joint venture between the Crown and private individuals, with expected returns taking the form of *mercedes*—land grants, access to indigenous labor, booty, tax exemptions, and social advancement.[4].

Contributions made by participants were often recorded and subsequently used to allocate rewards proportionally, resembling the operations of a commercial partnership. Although members of the *hueste* were not professional soldiers, they were subject to military jurisdiction for the duration of the expedition, including capital punishment for desertion or insubordination.[5].

Recruitment was conducted publicly through ritualized displays: the raising of the royal banner, the sounding of trumpets, and public proclamations. As described by del Castillo (1632), once enrolled, participants were legally bound to the expedition and forbidden from withdrawing under severe penalties.

The *hueste* was fundamentally a volunteer organization. Recruitment was regulated:

[4] Ballesteros (1985) and Padrón (1963)

[5] del Castillo (1632)

Muslims, Jews, heretics condemned by the Inquisition, gypsies, unmarried women without licenses, and certain enslaved individuals were formally excluded Given the centrality of evangelization to Castilian expansion, only Catholics were officially permitted to participate. These restrictions led the Spanish administration to maintain detailed records on the identities, origins, and family status of early migrants.[6].

Contrary to depictions of conquest as a purely military undertaking, members of the *hueste* were primarily settlers rather than soldiers. They included peasants, artisans, merchants, and professionals drawn from diverse occupational backgrounds. Discovery, conquest, and settlement were conceived as a unified process, both legally and in practice. As emphasized by Lockhart et al. (1976):

> "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 can we illustrate to what extent the conquerors were acting like immigrants, businessmen, and settlers."

2.3 Implications for the analysis

Early settlements were established under conditions of extreme uncertainty, with limited information regarding geography, climate, or indigenous societies. Initial urban centers were isolated and largely self-governing, with authority concentrated in the person of the conqueror. Formal institutions were largely absent at this stage; only later did administrative bodies such as *audiencias* emerge.

In sum, early Spanish colonization was primarily a process of urban occupation in which discovery, conquest, and settlement were inseparable. The first settlers simultaneously explored, conquered, and founded cities according to a shared set of simple institutional rules, in the absence of prior Spanish presence and under severe informational and environmental constraints. This context of decentralized initiative, initially weak institutions, and private investment is central to understanding both the character of early colonial expansion and the subsequent long-run institutional trajectories.

[6]See the "Manual of Indian Laws", Padrón (1963) and Ballesteros (1985)

For these reasons, and since our focus is on the first wave of settlements, in the empirical analysis, we control for climate, geography, and the presence of pre-existing indigenous groups, but the later establishment of viceroyalties, selective migration driven by prior Spanish settlement, or the composition of subsequent economic activities are not directly relevant. By construction, these settlements were founded by first movers in locations that had not yet been shaped by Spanish institutions, populations, or patterns of economic specialization.

3 Data

This section describes the construction of the dataset on the first Spanish settlers in the Americas and the way in which these data are linked to contemporary locations. The analysis exploits the fact that, during the initial phase of colonization, the Spanish Crown enforced a single legal port of departure and stringent controls on transatlantic travel, which generated detailed documentation that is now preserved in the General Archive of the Indies in Seville.

3.1 Original sources

Until 1668, Seville was the only legal port of departure for ships bound for *Las Indias* (the Indies). Beginning in 1492, the *Casa de la Contratación de las Indias* (House of Trade of the Indies) recorded information on all passengers traveling from Seville to the Americas as part of the regulation of colonial trade. These records, held in the General Archive of the Indies, appear in two main types of documents.

The first set consists of *informaciones y licencias* (information and licenses), which were required to establish that prospective travelers were legally entitled to embark. As discussed above, Jews, Muslims, converts, and individuals judged by the Inquisition were excluded; only those of proven Christian origin could travel. Supporting evidence included baptismal and marriage records that provide demographic information on passengers and their immediate family (spouses and children), which yield rich microdata on household composition.

The second set comprises the *Libros de asientos de pasajeros* (passenger lists), which report passengers' names (checked at boarding), the name of the vessel, and its captain. Merging information across these two sources allows the construction

of individual-level records including full name, family-head status, marital status, accompanying spouse and children, town or city of origin in Spain, and destination in the Americas.[7] Much later, in the eighteenth century, trade liberalization allowed departures from additional Spanish ports, and some travelers may have embarked on foreign ships or sailed under false documentation.

The classification and publication of these sources proceeded in several stages. Beginning in 1915, Luis Rubio y Moreno, Deputy Director of the General Archive of the Indies, organized and supervised the cataloguing of passenger records up to 1559; Cristóbal Bermúdez later compiled these materials into three printed volumes. Subsequent volumes covering 1560–1599 were produced by archivists in the 1980s. Building on this work, Peter Boyd-Bowman undertook three decades of archival research, producing five volumes that systematically reorganize and extend the *Catálogo de los pasajeros a las Indias*.

Boyd-Bowman's contribution is twofold. First, he cross-checked names from the passenger lists against many additional sources preserved in the General Archive—including foundation charters, contracts, payrolls, commercial agreements, royal grants, and *capitulaciones*—to augment the information on destination and occupation. These documents were written by conquerors, notaries (*escribanos*), or other appointed members of the *hueste*, so literacy was not required for an individual to appear in the record.[8] Second, Boyd-Bowman restricted attention to individuals whose presence in the Americas can be confirmed in colonial documents; the resulting compilation therefore captures settlers who reached the New World rather than all passengers who attempted the crossing.

The first volume of Boyd-Bowman (1493–1519) and the second (1520–1540) cover the initial wave of settlement.[9]

Boyd-Bowman's reconstruction yields exceptionally detailed microdata because he matched the passenger lists to numerous independent documents, removed obvious

[7]Destinations are not always explicitly reported at embarkation but can often be recovered using the *Catálogo de los pasajeros a las Indias*.

[8]This feature limits selection on literacy into the documentary record.

[9]The remaining three volumes, which extended coverage to 1599, remained unpublished and were lost after Boyd-Bowman's death in 2010; extensive searches in family papers and institutional libraries have not recovered them.

duplicate entries, and completed missing information on the origin and destination whenever possible. For the early period under study, his volumes constitute the most comprehensive source on the origin, destination, occupation, and other background characteristics of Spanish settlers in the Americas.

3.2 Data on settlers

The resulting dataset contains, for each settler and year of departure, the individual's name and surname, occupation, village of origin in Spain, and destination in the Americas. The level of detail can be illustrated by an entry 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), occupation court clerk, arrived in Santo Domingo in 1512, resident of Yaquimo (Santo Domingo) in 1514.

These biographies allow for the geographic placement of early settlers and the reconstruction of their occupational structure, which is central to the analysis of the long-run impact of initial human capital. The empirical work focuses on the first colonization wave, 1492–1540, during which Spanish settlement extended from northern Mexico to Argentina.

Boyd-Bowman's books contain 27,013 individual names in total. Of these, 7,182 have no recorded destination because the individual died en route (drowned, executed, or otherwise reported as having perished) or because the documentation does not specify arrival. In a few additional cases, the year of arrival is missing; these observations are dropped since their inclusion would introduce uncertainty about whether the individual belongs to the initial wave. Consequently, the final dataset of settlers in the period up to 1540 consists of 19,831 individuals.

Voyage mortality in this early period was high by historical standards, consistent with evidence from later episodes of Atlantic migration. For the eighteenth century, Mancke & Shammas (2005) reports mortality rates of approximately 15 percent per voyage for slaves and 10 percent for servants on the Middle Passage, while Grubb (1987) estimates a mortality rate of 5.5 percent for German immigrants to Pennsylvania during 1727–1805, to which morbidity at debarkation must be added. Travel conditions

were worse in the early sixteenth century, when the crossing took roughly two and a half months compared with about six weeks in the eighteenth century, and slave voyages recorded on the order of 100 deaths per 1,000 persons per month. The longer duration and harsher conditions of early Atlantic crossings imply higher mortality per voyage.

After digitizing and coding the individual-level records, and using additional archival sources (notably early maps of settlements), the destinations were georeferenced. Matching the historical names of cities of origin to their contemporary counterparts in Spain is straightforward in most cases, but destination names in the Americas often require additional historical work because many places were renamed or merged. When necessary, information on historical toponyms was collected to identify the modern location, and the corresponding geographic coordinates were then assigned. The final dataset identifies 123 distinct destinations for the period 1492–1540.

Among all settlers, 13,692 have a destination that coincides with the name of a city or village foundation, providing an exact location. The recorded arrival year generally precedes the formal foundation date because colonial authorities requested city or village status from the Crown only after settlement had occurred. For the remaining 6,136 settlers, Boyd-Bowman reports a region or area rather than a specific town. In these cases, the strategy is to identify which settlement(s) in that area existed at the time of arrival.

To this end, the analysis uses the information from López De Velasco (1894), a cosmographic survey commissioned in 1571 to map Spanish settlements in the Indies. Most of the locations in the sample can be matched to places that already existed by that date. For each settler whose destination is recorded as an area, López De Velasco (1894) is used to identify settlements in that area that existed at the time of arrival; in practice, there are typically very few such settlements, most often only one. The settler is then assigned to the village or city whose foundation year is closest to the arrival year. All such matched locations are among the 123 destinations used in the empirical analysis.

These procedures allow the reconstruction of the first settlement network of Spanish colonizers in the Americas. Figure 1 maps all identified settlements for 1492–1540; bubble size reflects the number of settlers present in 1540.

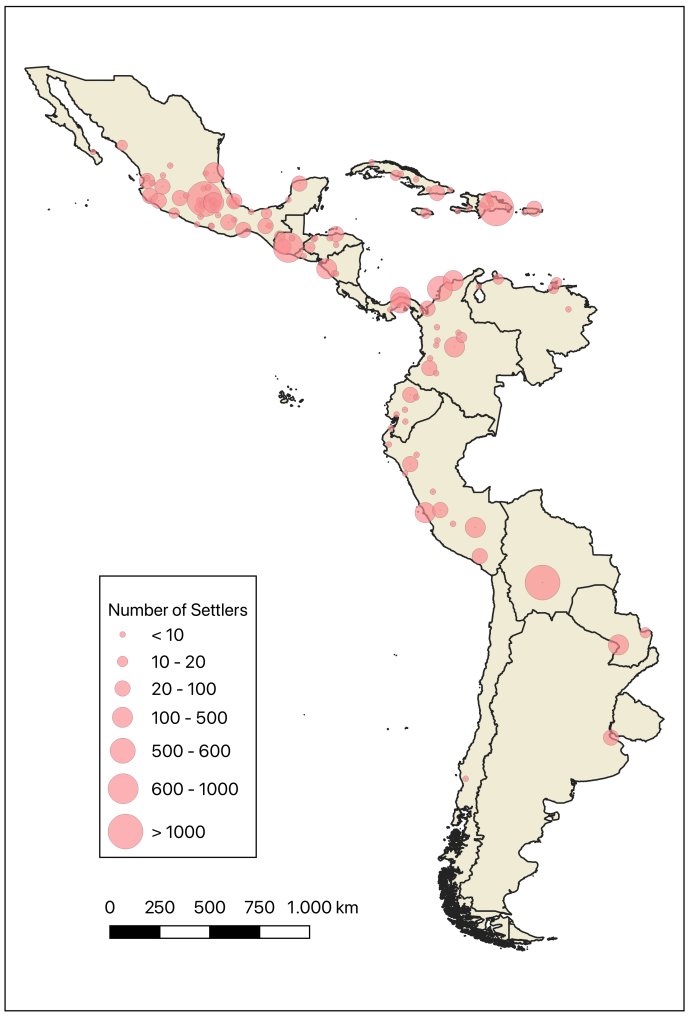

FIGURE 1: Early colonial settlements by number of settlers in 1540. *Source: Authors' calculations based on Boyd-Bowman and archival sources.*

Previous scholars have also exploited López De Velasco (1894) to enumerate early settlements. Solano (1990), for example, lists 227 cities, villages, and mines founded up to 1574, with detailed information on 189 of them, including both pre-1540 foundations and those established between 1540 and 1570. The 123 destinations in the present data were all created before 1541; 97 of them appear both in Solano (1990) and in Lopez de Velasco (1573). To assess coverage, the year of foundation was collected for the remaining places in Solano's list (excluding the 97 already matched). Many of these are clearly founded after 1540, and others have ambiguous dates; among the latter, only seven are cities for which at least one source reports a foundation date before 1540 and for which Lopez de Velasco (1573) provides information on inhabitants. This

comparison suggests that the settlements in Boyd-Bowman's volumes cover virtually all of the major early foundations in the period under study.

3.3 Data on settlers' occupations

Occupational information for the first settlers is a key ingredient of the analysis, as it underpins the construction of measures of initial human capital at the settlement level. Standard labor-force surveys provide direct information on education and occupation, but in the early sixteenth century, the education system was organized very differently and schooling was far from universal, which makes it difficult to map historical occupations into modern educational categories.

Historians have long emphasized the military component of early colonization, often assuming that soldiers dominated the first waves of migration. However, qualitative work documents substantial occupational heterogeneity among settlers; Lockhart (1968), for instance, shows that migrants to Spanish Peru came from a wide array of backgrounds. The data corroborate this view: after cleaning and aggregation (described below), 274 distinct occupations are observed among early settlers, including bankers (*banqueros*), lawyers (*abogados*), pharmacists (*boticarios*), barbers (*barberos*), butchers (*carniceros*), dentists (*dentistas*), physicians (*doctores médicos*), farmers (*labradores*), merchants or traders (*mercaderes*), and weavers (*tejedores*).

Constructing a proxy for human capital from these occupations is challenging. Attempts to infer formal education directly—for example, by matching occupations to institutional degrees—are complicated by the limited and uneven diffusion of schooling and by the fact that many occupations do not map neatly onto clearly defined educational requirements. Historical work on specific professions, such as *escribanos* (notaries), often documents examination requirements and the use of technical manuals, but such evidence is not systematically available for all occupations. Only a small subset of occupations can be unambiguously matched to specific levels of formal education.

Instead, the classification is based on the skill content of occupations, defined in terms of literacy, specialized knowledge, and entrepreneurial responsibility. Following Ladero (2004), who analyzes the structure of Spanish society around 1500, occupations that required reading and writing, specialized training, or the management of a business are treated as relatively high skill. This group includes doctors, lawyers,

pharmacists, notaries, and similar professions. At the other end of the spectrum, low-skill occupations include peasants (*jornaleros*) and servants (*criados*), who typically lacked land ownership and formal training. Lockhart (1968) notes that peasants contributed relatively little as an organized class to colonial society, which is consistent with this broad division.

The construction of occupational categories proceeds as follows. Boyd-Bowman's original books report 842 distinct occupation labels. Many differences arise from spelling errors (e.g., "INTERPETE" versus "INTERPRETE," both meaning interpreter) or stylistic variants of the same underlying occupation. For example, there are four variants for merchant (*mercader*): *mercader*, *mercader lancero*, *mercader tratante*, and *mercaderes*, which are all consolidated into a single category, MERCADER. After correcting obvious spelling mistakes and collapsing synonymous labels, the number of distinct occupations falls to 274.

These 274 occupations are grouped into several broad categories. The first category, "local business," includes artisans and small proprietors who operated workshops or small enterprises (such as weavers, shoemakers, printers, metal smelters, builders), as well as farmers, cattle herders, fishermen, and small boat owners; these occupations share the common feature of running a small business and account for roughly 10 percent of settlers. The second category, "royal officials," includes administrators, explorers, soldiers, military officers, treasury officials, and judicial personnel, representing 18.76 percent of settlers.

The third category consists of occupations requiring a university degree (e.g., physicians, lawyers, notaries), which represent 4.67 percent of the sample. Before 1520, Spain had 13 universities (Lleida, Huesca, Girona, Barcelona, Valencia, Valladolid, Salamanca, Sigüenza, Santiago, Alcalá, Ávila, Sevilla, and Toledo), offering mainly law, medicine, theology, and canon law; settlers in these professions are classified as university-educated.[10] The fourth category, "other high education," includes interpreters and other occupations that require advanced linguistic or technical skills and accounts for 5.16 percent of settlers. The fifth category comprises "clergymen," including priests and other ecclesiastical personnel (2.63 percent). Finally, merchants (1.97 percent), bankers and lenders (0.07 percent), and artists (musicians, painters, singers, writers, and related professions, 0.4 percent) form additional distinct categories

[10]This definition reflects the structure of higher education in early sixteenth-century Spain.

(see Table A.1 in the Online Appendix A).

All of these groups together define the high-skill set, which covers 45 percent of settlers. The complementary low-skill category is composed primarily of servants and landless peasants (*jornaleros*), who worked the land of different owners and often changed location. Following Boyd-Bowman's classification, the servant group includes domestic servants, individuals with no clearly defined occupation, and attendants of university graduates, among others. Contemporary descriptions of Spanish society around 1500 emphasize that servants and peasants constituted the majority of the population (Lopez de Quesada, 1500), and Boyd-Bowman reports that 58 percent of settlers belonged to these groups. In the final dataset, the corresponding figure is 55 percent.

Figure 2 maps the 123 destinations according to the proportion of settlers classified as high skill. The variation in color intensity across locations indicates substantial heterogeneity in the occupational and skill composition of the first settlers.

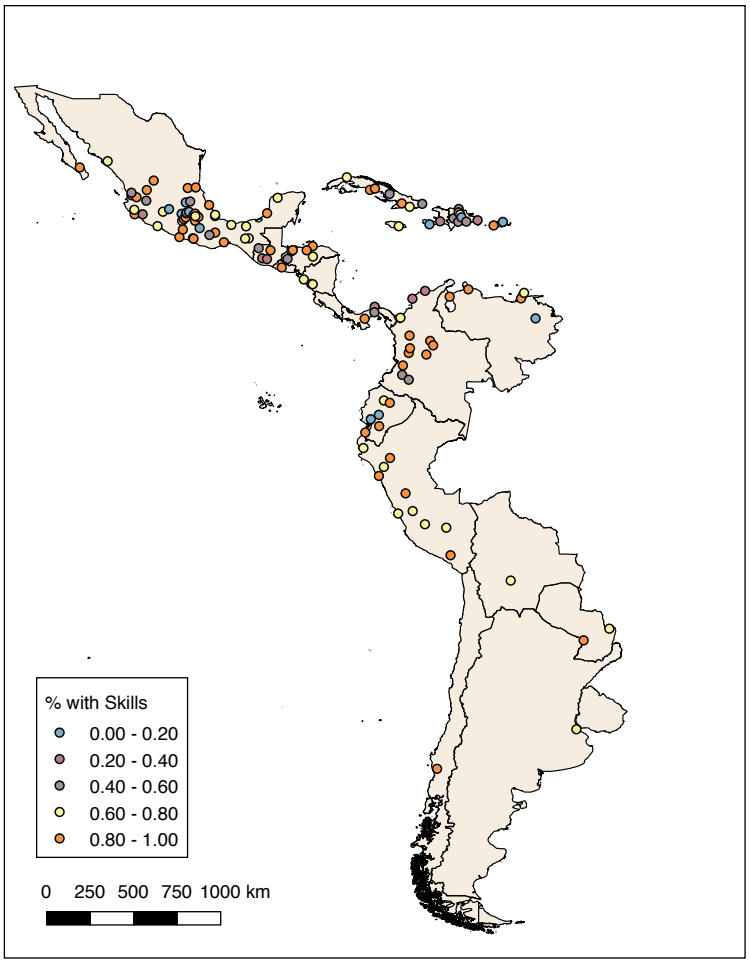

FIGURE 2: Proportion of settlers with skills.

Table 1 reports descriptive statistics for all the variables aggregated at the location level. A detailed description of all variables is provided in the Appendix A.

TABLE 1: Descriptive statistics

	mean	sd	Number of settlements
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

4 Understanding the location of first settlers

The empirical strategy exploits variation in the skills of the very earliest settlers, focusing on the initial phase of Spanish colonization rather than on mid-colonial settlement patterns, which have been the emphasis of most of the existing literature. This focus offers the advantage that first settlers could not have been influenced by previous Spanish migrants, but it also raises the challenge of understanding what determined the locations of these initial settlements.

Between 1492 and 1540, Spanish colonizers established urban centers across a vast territory stretching roughly from the 25th parallel north to the 35th parallel south, from Zacatecas (northern Mexico) to Buenos Aires. In less than fifty years, they founded cities in regions with highly diverse geography, climate, and indigenous populations. A key feature of the setting is that the entire area was governed by a single colonial power, Spain, which allows formal institutions and legal origin to be held constant and avoids the confounding bundle of institutions and human capital present in multi-colonizer comparisons.[11] Moreover, by restricting attention to the

[11]Section 5.3 presents robustness checks that consider local and informal institutional

initial colonization period (1492–1540), differences in the skill composition of settlers across locations cannot be driven by the pull effect of preexisting Spanish communities, since there were no earlier Spanish settlers. This period corresponds to the phase of discovery, conquest, and settlement that unfolded nearly simultaneously across Spanish America, and it is characterized by high uncertainty about geography, climate, and territorial limits. First settlers acted at once as explorers, conquerors, and residents.

Contemporaries had little information about the comparative economic advantages of different locations, given that the geography of the New World was largely unknown. As discussed above, simple rules—such as scholastic recommendations inspired by Saint Thomas Aquinas on how to choose settlement sites, together with the Crown's prohibition on leaving a new settlement before it was firmly established—provided the main guidance.

Additional historical evidence points to substantial ignorance about the physical characteristics of the territory and the economic potential of specific areas. During the initial phase of colonization, *capitulaciones* often allocated extremely imprecise geographic areas, sometimes granting a strip of the Pacific Ocean, a nonexistent island, or a mountain range as the basis for conquest. García-Baquero (1992) argues that conquerors could be "lucky" or "unlucky" in the land assigned to them, underscoring the quasi-random nature of early allocations. Precisely because of this uncertainty, the Crown later dispatched López de Velasco to describe and demarcate the geography of the Indies in 1571.López De Velasco (1894)

At the same time, it is necessary to consider potential endogeneity in the choice of settlement locations, even in this very early context. The analysis proceeds in two steps. First, it documents that geography, climate, and indicators of indigenous institutional development are similar across locations with a high versus low proportion of high-skill settlers (Section 4.1). Second, it investigates potential push factors, drawing on migration theory and, in particular, the shift-share approach of Card (2001). In standard applications, a key push mechanism is the presence of previous migrants from a given origin in the destination; in this setting, that mechanism is mechanically absent, as the focus is on the first settlers. Accordingly, the analysis explores whether early migrants may instead have preferred regions whose climate and geography resembled those of their origin and evaluates whether such a "climatological" push factor can predict settlement patterns (Section 4.2).

---

arrangements.

4.1 Determinants of the distribution of skills of the first settlers

This subsection examines whether locations with a high versus low proportion of high-skill settlers differ systematically in pre-existing characteristics. Specifically, it assesses whether geography, climate, pre-Columbian population density, and attributes of indigenous societies have explanatory power for the initial occupational composition of settlers across urban centers.

Table 2 reports regressions of the proportion of high-skill settlers at each location on geography, climate, and precolonial and indigenous characteristics. In column (1), the dependent variable is the percentage of settlers classified as high skill; in column (2), the dependent variable is a dummy that equals one when more than 75 percent of settlers at a location are high skill. Across both specifications, the coefficients on geography and climate are small and statistically insignificant, and there is no systematic association with soil quality, malaria environment, or measures of indigenous development.

Taken together, these results suggest that the distribution of high- versus low-skill settlers is not systematically related to observable geographic or climatic characteristics, nor to pre-existing indigenous institutional features. In this sense, the initial allocation of high-skill settlers across locations appears conditionally balanced, providing support for the identifying assumption that early variation in skills is plausibly exogenous to local fundamentals.

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%.

4.2 Predicting first settlers location using migrants' home-region climate and geography profiles

Section 4.1 shows that observable geographic and climatic characteristics are not correlated with the share of high-skill settlers across locations. This subsection nonetheless shows, from a different perspective, the nature of the "push" factors shaping the location choices of the first settlers, taking inspiration from the migration literature and, in particular, from the shift-share strategy introduced by Card (2001).

In standard applications, the Card-type instrument for immigrant inflows interacts national inflows from each origin group with the historical distribution of that group across destinations, thereby separating push factors (changes in national-origin inflows) from pull factors (local conditions). Adapting this logic to the early-settler setting, an analog for the proportion of high-skill settlers would take the form

$$SP_{jc} = \sum_{g} \tau_{gj} \lambda_{gc} M_g,$$

where M_g is the total number of settlers from the Spanish province g arriving in Latin America, τ_{gj} is the share of high-skill settlers ($j = 1$) from province g, and λ_{gc} is the fraction of settlers from province g residing in the destination city c in the previous period.

A central novelty of this context is that the analysis focuses precisely on the first generation of settlers who founded new villages. The stock of settlers in any Latin American city prior to their arrival is therefore zero, so λ_{gc} cannot be constructed using lagged settlement patterns in the usual way. Recent historical work, such as Buchardi, Chaney, and Hassan (2018), uses nineteenth-century migration networks to instrument for current ancestry, but such an approach is infeasible here because there is no earlier stock of Spanish settlers in the Americas. This raises a fundamental question: in the absence of prior settlement, what determined the destinations of the first migrants?

One plausible hypothesis is that emigrants preferred regions whose geography and climate resembled those of their origin. If so, the probability that settlers from province g chose city c should be higher when the environmental characteristics of g and c are similar, and such similarity measures could be used to approximate λ_{gc}. The empirical exercise tests this prediction.

The first step is to link each individual settler to both a Spanish province and a village of origin; in total, the data include 49 provinces and around 1,300 villages. Figure 3 shows the spatial distribution of villages of origin across Spain.

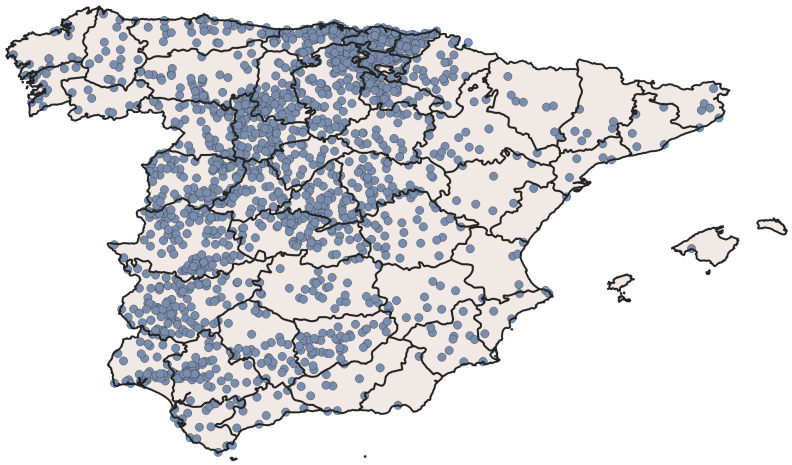

FIGURE 3: Spanish villages of origin

Next, a province–city matrix is constructed, with one observation for each Spanish province–Latin American city pair, yielding 6,027 entries (49 provinces by 123 cities). For each pair (g, c), the variable Per_{gc} measures the percentage of settlers from province g who migrated to city c. For example, out of 120 settlers from the province of Álava, 2 settled in Cajamarca (Peru), so Per_{gc} equals 1.67 percent for the Álava–Cajamarca pair. The baseline specification regresses Per_{gc} on a vector X_{gc} of absolute differences in fertile soil, ruggedness, temperature, and precipitation between province g and city c:

$$Per_{gc} = \alpha + X'_{gc}\gamma + \varepsilon_{gc}.$$

Tables 3 and 4 report the results using Spanish provinces and villages as origins, respectively. In each table, column (1) presents the baseline specification, column (2) adds fixed effects for Spanish origins (province or village), and column (3) includes destination-city fixed effects. Column (4) replaces city fixed effects with "colonizer route" fixed effects, comparing cities within the same conquest route, which is relevant because, in practice, individual choices were likely constrained to the set of locations included in a given *hueste's* penetration line.

TABLE 3: Testing the climatological theory (with Spanish provinces)

Dependent Variable: Per_{gc}

	(1)	(2)	(3)	(4)
Absolute Dist. in Precipitation	-0.273**	-0.253*	-0.175	0.149
	[0.115]	[0.138]	[0.148]	[0.141]
Absolute Dist. in Temperature	-1.313***	-1.833***	0.174	-1.381***
	[0.234]	[0.256]	[0.204]	[0.228]
Absolute Dist. in Ruggedness	1.257***	1.408***	4.493	0.228
	[0.205]	[0.197]	[4.658]	[0.220]
Absolute Dist. in % Fertile Soil	-0.001	-0.001	-0.000	-0.005***
	[0.002]	[0.002]	[0.001]	[0.002]
Spanish Province FE	No	Yes	Yes	Yes
City of destination FE	No	No	Yes	No
Colonizer Route FE	No	No	No	Yes
Observations	6027	6027	6027	6027
R-squared	0.009	0.011	0.634	0.107

Notes - The unit of observation is the Spanish province–Latin American city pair. Robust standard errors in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

TABLE 4: Testing the climatological theory (with Spanish villages)

Dependent Variable: Per_{gc}

	(1)	(2)	(3)	(4)
Absolute Dist. in Precipitation	-0.309***	-0.368***	-0.159**	0.015
	[0.034]	[0.037]	[0.080]	[0.036]
Absolute Dist. in Temperature	-0.930***	-1.526***	-0.013	-1.446***
	[0.083]	[0.098]	[0.118]	[0.090]
Absolute Dist. in Ruggedness	0.790***	1.160***	-0.278*	-0.041
	[0.060]	[0.073]	[0.153]	[0.093]
Absolute Dist. in % Fertile Soil	-0.000	-0.001	0.001	-0.003***
	[0.001]	[0.001]	[0.001]	[0.001]
Spanish Village FE	No	Yes	Yes	Yes
City of destination FE	No	No	Yes	No
Colonizer Route FE	No	No	No	Yes
Observations	145,684	145,684	145,684	145,684
R-squared	0.002	0.003	0.242	0.044

Notes - The unit of observation is the Spanish village–Latin American city pair. Robust standard errors in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

The idea is to use the predicted values from these regressions as an analogue for λ_{gc}. In practice, however, the predicted λ_{gc} is only weakly correlated with the actual shares Per_{gc}. Using village fixed effects (column (2) of Table 4), the correlation between predicted and observed Per_{gc} is 0.0549, while the corresponding correlation in the province-level analysis is 0.1049 (column (2) of Table 3). These weak correlations indicate that similarity in climate and geography between origin and destination is a poor predictor of the distribution of settlers across cities during early colonization, and hence that a Card-style instrument based on predicted λ_{gc} is not feasible in this context.

Several robustness checks confirm this conclusion. First, the analysis redefines τ_{gj} as a dummy for whether the share of high-skill settlers from province g exceeds thresholds such as 50 percent or 75 percent. Second, it imputes missing values in the province–city matrix when Per_{gc} falls below certain cutoffs (e.g., 1 percent, 10 percent). Third, it separately predicts the numerator and denominator of λ_{gc}. Parallel exercises using villages instead of provinces yield similar results. In all cases, the predictive power of climatic and geographic similarity for actual settlement patterns remains very limited.

If climate and geography were important determinants of location choice, settlers from a given province or village would concentrate in a small subset of Latin American destinations. Instead, the data show substantial dispersion. Using Per_{gc}, a fractionalization index of destinations, $DestFrac_g$, is constructed for each province to measure the diversity of its settlers' destinations. Figure 4 displays the distribution of this index across provinces. Most provinces exhibit high fractionalization, with the 25th percentile at 0.761, the median at 0.809, and the 75th percentile at 0.836, indicating

that settlers from a given province spread across many different cities.

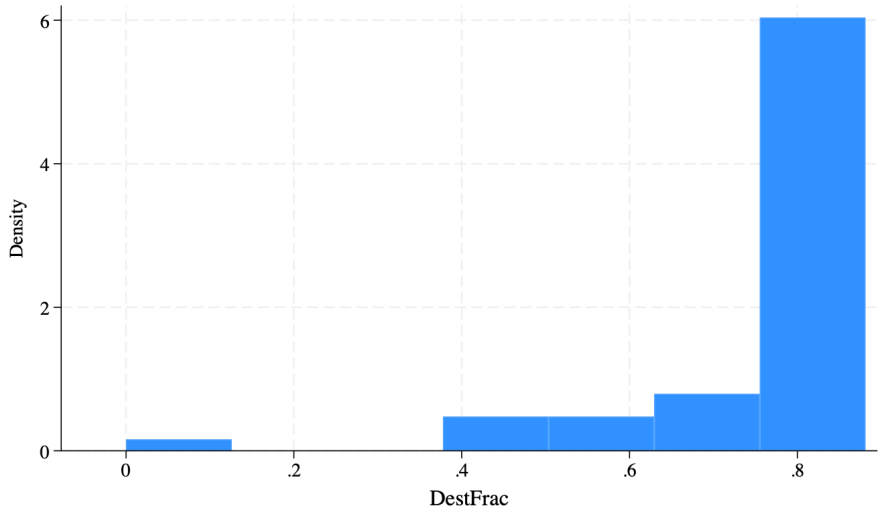

FIGURE 4: Fractionalization of destinations by province

Had the predicted values from the climatological regressions closely tracked the observed shares Per_{gc}, they could have been used to construct a shift-share style instrument in the spirit of Card (2001), using the predicted λ_{gc} instead of the historical distribution. The weak predictive performance of climate and geography implies that this strategy is not viable. At the same time, the results reinforce the view that the location choices of the first settlers did not systematically follow patterns of geographic or climatic similarity with their places of origin. While caution is warranted, the evidence is consistent with the interpretation that the distribution of first settlers across locations contains a substantial random component, which mitigates concerns about the endogeneity of the location of the settlers in the context of the initial wave of colonization.

5 Early human capital and long run development: an application

As we argued before during the initial phase of colonization, the land granted to the settlers was extremely imprecise in geographical terms, since they did not know the correct demarcation of the land in the New World. The Capitulaciones sometimes

granted the migrants a strip of the Pacific Ocean, a nonexistent island, or a mountain range. Historians have shown that conquerors could be "lucky" or "unlucky" in the land assigned to them, underscoring the quasi-random nature of early allocations. In the previous section, we have shown that the settlement patterns do not appear systematically related to predictable environmental factors. Given all this evidence on the exogenous nature of the location of the first settlers, we can try to answer the question of how the skills of the initial settlers influence the long-run economic development of their settlement locations. This question represents a compelling application of the new dataset. If human capital matters for economic productivity and growth — a hypothesis supported by extensive literature in development economics and economic history — then examining whether the human capital of founders influenced subsequent regional development provides valuable evidence on the importance of this factor.

The temporal scope of this inquiry is particularly significant. The dataset documents settlers from 1492 to 1540, representing approximately 48 years of the initial colonization phase. Since locations do not correspond with administrative units that have an estimation of current production, contemporary development outcomes have to be calculated using an alternative measure. One possibility is to use satellite nighttime luminosity data in the specific area of the original settlement.[12] If initial settler composition still influences development outcomes after five centuries and across multiple institutional, technological, and economic transformations, this would constitute remarkable evidence of path-dependence and the lasting significance of early human capital endowments.

5.1 Estimating the effect of early skills' composition on current development

To relate the skills of early settlers to contemporary development, the analysis constructs circular buffers with a radius of 20 km around each original settlement.[13] Because historical settlements rarely coincide with modern administrative units, out-

[12]We are aware that nighttime luminosity is an imprecise measure of development, but in this particular exercise, it is difficult to find a better alternative to approximate output.

[13]Section 5.3.2 shows that the results are robust to alternative buffer sizes.

comes are measured at the buffer level rather than at the level of current municipalities or counties.

The main outcome variable is per capita income in the area surrounding each original settlement. Direct income data at this spatial resolution are unavailable, so the analysis follows Henderson, Storeygard, and Weil (2012) and uses satellite luminosity at night as a proxy for economic activity. Night-light data are obtained from the National Oceanic and Atmospheric Administration, and the population in each buffer is measured using Oak Ridge National Laboratory's LandScan grid. Chen and Nordhaus (2011) show that luminosity contains valuable information on economic activity, especially in settings with poor or missing income data, and several studies find a high correlation between night lights and GDP (e.g., Michalopoulos and Papaioannou 2013; Alesina, Harnoss, and Rapoport 2016). The basic estimating equation is

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

where $\log gdp_{ji}$ denotes log night light per capita in 2010 for buffer j in area i, HC_{ji} is a measure of the skill composition of the first settlers, and z_{kji} is a vector of control variables. The baseline measure of HC_{ji} is the proportion of high-skill settlers in 1540; an alternative specification uses a dummy that equals one for locations where more than 75 percent of settlers are high skill.

The controls include geography and climate: distance to the sea, average terrain roughness (Nunn and Puga 2012), average elevation, average precipitation, and average temperature (CRU data). Additional geographic controls are the share of fertile soil, distance to the nearest river, and distance to the nearest lake, all averaged at the buffer level. The specification also includes latitude, measures of agricultural suitability and caloric suitability, and indices of disease environment (malaria endemicity).

Beyond physical geography, the analysis controls for precolonial and institutional conditions. These variables include precolonial population density, measures of the hierarchical organization of indigenous societies, and indicators of indigenous construction, as well as measures of informal institutions, which are described in detail in subsequent sections. Institutional variables directly related to colonial administration include the distance to the nearest *Audiencia* before 1540, an indicator for the absence of an *Audiencia* at foundation, and the number of years the settlement remained without an *Audiencia*. All specifications include the year of the first settler and area

fixed effects. Standard errors are clustered at different spatial aggregation levels (8-by-8-degree and 10-by-10-degree grids). Figure 5 and Figure 6 provide descriptive evidence consistent with the main empirical results. Figure 5 plots kernel densities of log night light per capita in 2010 for settlements with a high proportion of high-skill settlers (more than 75 percent) and for the remaining settlements. The distribution for high-skill locations is shifted to the right, indicating higher contemporary development in places initially settled by more skilled colonizers.

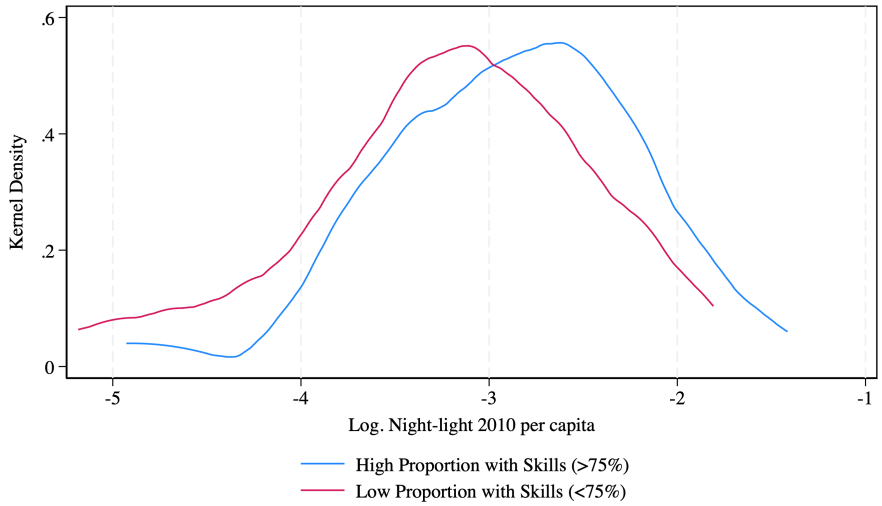

FIGURE 5: Proportion of settlers with skills (kernel density).

Figure 6 presents a binned scatterplot of log night light per capita in 2010 against the percentage of high-skill settlers, residualized with respect to geography and climate controls (log distance to coastline, log ruggedness, log temperature, and log precipitation). The positive slope of the fitted line illustrates the strong association between the initial skill composition of settlers and current development.

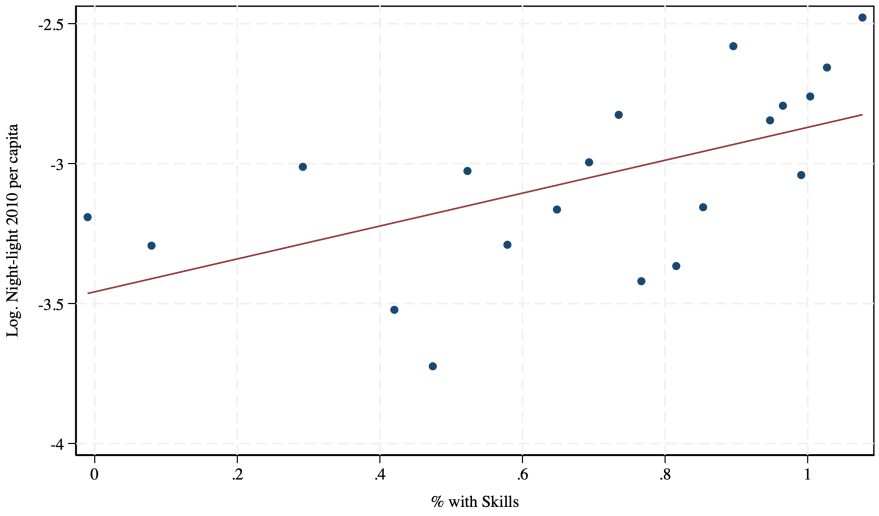

FIGURE 6: Log night light per capita and percentage of settlers with skills. *Notes: Binscatter between log night light per capita in 2010 and the percentage of settlers classified as high skill, after controlling for log distance to coastline, log ruggedness, log temperature, and log precipitation.*

Table 5 presents the OLS 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 lower proportion of high-skilled settlers. The estimation is not only statistically significant but also economically very relevant: areas that received only high-skilled settlers have a 60% higher level of GDP per capita than 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 Appendix A). The results are unaffected by the inclusion of this variable. In Column 8, we include the first year of the settlement. In Column 9, we cluster using cells of 10 by 10 degrees. In the Appendix A.2, we show that the results are robust to clustering at other geographical levels: using cells of 8 by 8 degrees.[14] and clustering at the audiencia level or country level (see table in the Appendix A.2). 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.[15]

TABLE 5: 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 Appendix A. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

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

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

5.2 Randomization Inference and Montecarlo Simulation

We perform some analyzes to provide more evidence that supports our results, such as randomization inference and placebo simulations. In the previous section (Section 4), we provide some evidence that indicates location endogeneity does not seem to be an issue in our context. We reinforce this idea by conducting randomization inference. Specifically, we estimate the baseline regression (Column 10 of Table 4) 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 based on Monte Carlo simulations and compute the corresponding p-value after randomly permuting the percentage of settlers with skills 500 times. 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 a 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 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 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.[16]

5.3 The role of initial conditions

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

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

5.3.1 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.[17] To check that this is not driving the results, Table 6 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 older 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 that appear in the Ethnographic Atlas. For the cases that we match,

[17] The basic specification considers the controls from Column 10 of 5

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 Appendix C 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 reports the number of temples in each area[18]. The main source from which Mayshar et al. (2020) collected the information to construct these datasets is the Archaeological Atlas. The Appendix A 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 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.

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

TABLE 6: 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 4 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%.

5.3.2 Early institutions

Another important issue to consider is whether the effect of early high 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 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 the 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.1 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.2. 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.[19] 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 the name of the conqueror. Figure 7 depicts these penetration lines, showing that, as expected, early settlements are very close to these paths. Figure A.3 provides a closer look at 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.

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

FIGURE 7: Penetration lines before 1540

In Table 7, 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 7 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 6, showing that the main result is robust to this specification.

TABLE 7: *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 4 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 4 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%.

In Online Appendix D, 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 and to the inclusion of diversity of occupations.

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 formal 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.[20]

A potential mechanism for persistence could be 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 entrepreneurship. 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."[21] . 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.[22]. 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

[20]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).

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

[22]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 the persistence of elite professions across many generations.

met.[23]

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 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 for which we do not find empirical support.

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 these results, 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. 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

[23]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).

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 (8, 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 observe 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).[24]

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.[25]

[24]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

[25]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/

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 8 suggest a positive relationship between the above ratio and the level of high-skill early settlers among the first colonizers.

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.[26] 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 obtain 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 10).

We also assess market orientation by examining the existing infrastructures, in particular roads, which facilitate the transportation of products to other places.[27] 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.[28] 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

10.1046/j.1466-822x.2002.00294.x

[26] 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.

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

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

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 8, we find a positive relationship between *roads per capita* and the high-skilled proportion of first 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 8: 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 4 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.[29] 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 9 shows that areas with more skilled settlers have a more diversified 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 it shows the right sign, 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 9: 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 4 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%.

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

6.3 Attitudes towards entrepreneurship

The previous two mechanisms offer an indirect explanation for the link between entrepreneurship and the original proportion of high-skill settlers. We can also attempt to directly measure this link 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 mechanism, we employ information from the Latinobarometer, using responses to the following question as a proxy for attitudes towards entrepreneurship: *"What is the most important 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 C 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 10 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 10: 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 4 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 E shows that using data from the Census, Ipums, and Historical data from Jesuit missions, there is no evidence of this mechanism. Another potential mechanism of persistence is an agglomeration effect[30] 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 E)

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 settlers in the Americas, which includes their complete names, 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 setting for evaluating the effect of skills on long-run development. First, we show that differences in the geographical distribution of skills do not depend on geography, climate, or characteristics of the indigenous groups. Moreover, they cannot be predicted

[30] See Rocha et al. (2017)

by similarity in geography or climate of the place of origin, which reduces our concerns about endogeneity. Second, the area was colonized by a single colonial power, allowing us to hold formal institutions and legal origin constant. Previous analyzes 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 alternative to such methodologies.

The results indicate that locations where a greater number of early settlers had highly 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.

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.

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.

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.

del Castillo, B. D. (1632). *Historia verdadera de la conquista de la nueva españa*. Madrid: Imprenta del Reino.

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

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

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.

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*.

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

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)

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.

Padrón, F. M. (1963). *Historia del descubrimiento y conquista de américa*. Madrid: Editora Nacional.

Pérez-Artés, M. d. C. (2024). Numeracy selectivity of Spanish migrants in colonial America (sixteenth–eighteenth centuries). *The Economic History Review*, *77*(2), 503–522. doi: 10.1111/ehr.13279

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.

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).

A Online Appendix – Data sources and definitions

A.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.

A.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.1: 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%

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.

A.3 Initial Conditions and *audiencias*

A.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.

A.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.1: 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.2: Audiencias at the end of the 16th century

A.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).

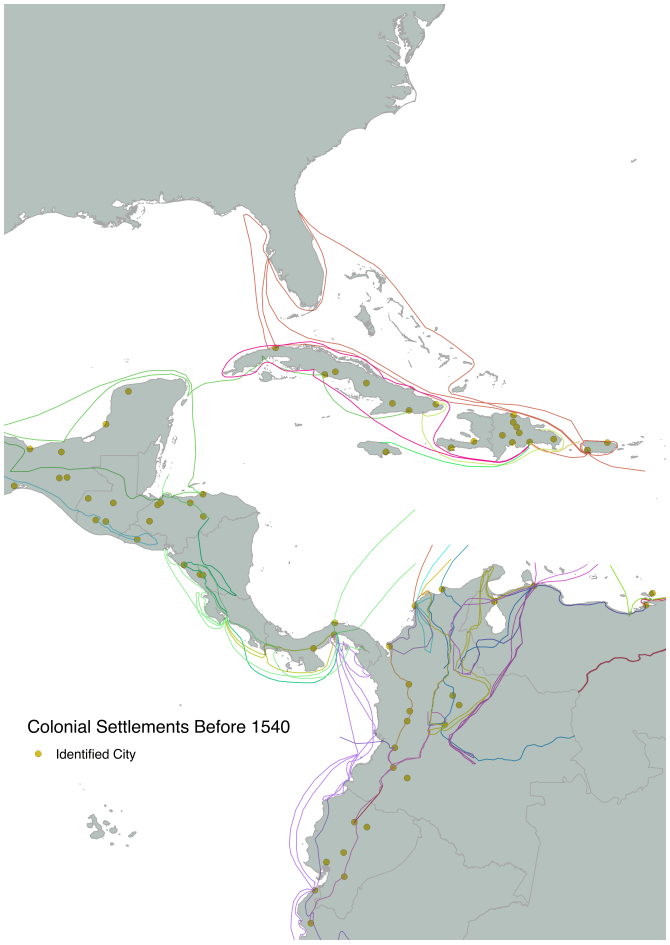

FIGURE A.3: Details of penetration lines in Central America

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

A.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 Appendix C 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 Appendix C for details on the matching procedure.

A.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).

A.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.[31] 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

[31]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.

B Online Appendix - Clustered standard errors at different levels

TABLE A.2: 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 4 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%.

C 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 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).

D Online Appendix – Robust Analisis

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.3, 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.3 presents the results using all the settlers, showing that the estimation is consistent with the basic findings.[32]

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.3 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.

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

Table A.3: 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 4 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.5, 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.Results are 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 4 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%.

Table A.5: 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 4 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%.

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.6 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.6: 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 4 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%.

E 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.[33] We assess whether places with a greater proportion of skilled settlers are more closely correlated with Jesuit missions than those with fewer skilled settlers.[34] The main source of information on

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

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

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 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.7 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.[35] 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.7 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.

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

TABLE A.7: 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 4 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[36] 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.

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.[37] 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.8.

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

[36] See Rocha et al. (2017)

[37] 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)

have collected this information. We were able to match 78 places. The results are in line with the results using contemporaneous data.

TABLE A.8: 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 4 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%.