Cities on Pre-Columbian Paths *

Bruno Barsanetti†

June 2019

Abstract

To what extent does the geographic distribution of pre-Columbian societies determine the location of New World cities? This paper provides evidence that some modern cities in southern Brazil concentrate around a pre-colonial trail. Historical accounts suggest that the Peabiru, as the path was named, was an important factor in explaining the location of the first European settlements, so the concentration is suggestive of path dependence. To separate the causal effects of the path from the effects of any geographic fundamentals that could correlate with it, I construct a counterfactual by exploring a region where European (Spanish) settlements were abandoned after a 17th-century slave raid. I show that proximity to the Peabiru is associated with higher population density and urbanization only in a region that was never abandoned; there is no effect in the area the Spanish abandoned.

*This work benefited from suggestions by Çagri Akkoyun, Roberto Aragão, Mariana Carvalho, Alípio Ferreira, Haritz Garro, Sara Hernandez, Yutaro Izumi, Cynthia Kinnan, Rafael Pucci, Prateek Raj, and Roberto Rosado; and from a seminar presentation at Northwestern University. All errors are mine.

†Northwestern University. E-mail: barsanetti@u.northwestern.edu
1 Introduction

Why do some sites develop into cities, concentrating population and economic activity, while other sites do not? In this paper, I provide evidence on how the pre-colonial distribution of economic activity affected city location in the Americas. Although today the continent has a well-developed network of cities, most of them only appeared after the European arrival. The cities did not emerge in a vacuum, but a priori it is not obvious to what extent the geographic distribution of pre-Columbian societies affected city placement. As the first colonizers were vastly outnumbered by the indigenous population, who often provided the Europeans with protection, trade opportunities and labor, colonial settlements had an incentive to locate near sites that were economically relevant before the European arrival. On the other hand, the European conquest led to a drastic reorganization of economic activity across the Americas, due to large population losses, an influx of new technologies, and the emergence of trade links with Europe and Africa. All these changes were likely to have impacted the benefits of locating in alternative sites, in which case the modern spatial distribution of economic activity could look quite different than the pre-colonial one.

In this paper, I ask: is there a causal effect of the spatial distribution of pre-Columbian societies on modern city location? There are two challenges in estimating the causal effect. The first is a data limitation issue: we lack good information on the spatial distribution of economic activity before the colonial period, especially at the geographically fine level that is necessary to study city placement. The second is an identification issue: modern and pre-colonial agglomerations may be located at the same site due to unobservable geographic features, in which case persistence does not reflect a causal relationship.

To solve the data limitation issue, I use the course of a network of indigenous paths in southern Brazil. The Peabiru was a trail system that crossed the Brazilian states of São Paulo and Paraná and provided a land route from the Atlantic coast to the Andes. Its strategic importance was recognized as early as 1525, when the first European traversed it in order to attack the Inca Empire, attracting conquest and settlement efforts from both Spain and Portugal. Over time, some towns developed on this path, eventually becoming large cities, of which São Paulo is the most notable case. Early European travelers of the Peabiru mentioned several populous villages along it, so a concentration of modern cities around the trail is indicative of persistence from pre-colonial agglomerations.

To account for the identification issue, I explore the 17th-century abandonment of Spanish settlements as a natural experiment. At the same time the Portuguese started the settlement of Brazil through the Atlantic coast, the Spanish occupied the western part of the Brazilian state of Paraná, advancing from Paraguay. Historical evidence shows that the Peabiru and the native villages that surrounded it were important factors behind the location of both Portuguese and Spanish settlements. In 1628, however, increased demand for slaves in Brazil led the Portuguese to organize a slave raid into the Spanish area. Spanish towns were militarily attacked and abandoned. There
was also a large negative shock to the native population, the majority of which was either captured, died or fled to other regions. Most importantly for identification, the area was not subsequently occupied by the Portuguese, being settled by Euroamericans only when western Paraná became an important frontier in the 20th century (Kohlhepp 2014). Therefore, the modern distribution of population in the formerly Spanish region is an ideal counterfactual, as it has not been affected by pre-Columbian populations and should mostly follow geographic fundamentals.

I implement this exercise by estimating a difference-in-differences model where the treatment is proximity to the Peabiru in the region that was occupied only by the Portuguese, not by the Spanish. Hence, I identify the causal effects by the difference between the coefficient on proximity to the path in the Portuguese region relative to the coefficient in the Spanish one. Balance tests, that use observable geographic fundamentals as the dependent variable, support the assumptions behind the identification strategy.

The results on different measures of the modern distribution of economic activity, such as population density, urbanization, and satellite nighttime lights, confirm the hypothesis that the concentration of economic activity around the indigenous path reflects a causal relationship. As an example, consider urbanization: only 5.8% of the sample is urban, but being closer than 10 km to the Peabiru increases this probability by 5.1 p.p. conditional on being in the Portuguese area. In the Spanish region, however, there is no change in urbanization with distance to the Peabiru.

I also investigate the mechanisms behind the results. I reject the hypothesis that persistence was completely due to indigenous trails being later converted into modern roads. The effects of the Peabiru on modern roads are small, so the effects of modern roads on the spatial distribution of economic activity would have to be unreasonably large for modern roads to completely mediate the effects of the pre-colonial path. Instead, the effects were probably due to agglomeration economies and coordination around settlements that were established during the colonial period.

I contribute to the empirical literature on the emergence of cities. There is already an extensive literature in economics on the emergence of European and Middle Eastern cities, that includes Bosker et al. (2013), Bosker and Buringh (2017), Michaels and Rauch (2018), Blaydes and Payke (2018), and Cermeño and Enflo (2018). Much less is known about city appearance and growth in the Americas. An exception is the evolution of United States cities from the 19th century (Desmet and Rappaport 2017, Nagy 2017). Studies of the persistence of the spatial distribution of economic activity in the continent usually focus on a coarser geographical unit of analysis, such as countries or states (Acemoglu et al. 2002, Maloney and Valencia Caicedo 2016).

This paper is also related to a literature that studies ancient roads. Wahl (2017) and Dalgaard et al (2018) examine the long-run effects of Roman roads, showing that a higher Roman road density in Europe leads to higher road and population density today. A paper that considers pre-Columbian roads is Martincus et al (2017), who use the Inca road network to construct an instrument for modern road building. Burghardt (1969) is an older study on how the indigenous trails in Canada’s
Niagara Peninsula gave way to a modern road network. In that context, the persistence of indigenous paths was limited.

Finally, this paper contributes to a vast literature that uses shocks such as war bombings and epidemics to study the persistence of geographic concentration over time\(^1\). An alternative strategy to identify path dependence is to select a geographic feature that no longer has any value, such as portage sites (Bleakley and Lin 2012) or depreciated colonial railroads (Jedwab et al. 2015), and then see if there is concentration of economic activity around it. Plausibly this is the case for the Peabiru, as a pre-Columbian road that could only be traversed on foot or by mule (and that has been mostly erased by deforestation and agriculture) is unlikely to be valuable in a modern economy. But notice that I use instead a more conservative strategy based on the interaction between the feature (the Peabiru) and a large shock (the abandonment of Spanish settlements). A similar identification strategy is used by Michaels and Rauch (2018), who compare urbanization around the sites of Roman towns in modern France and Britain.

2 Historical Setting

After Columbus’ first voyage to the Americas in 1492, Europeans began exploring the continent. The settlement of the Río de La Plata basin started in the early 16\(^{th}\) century, as the Spanish and Portuguese empires disputed the area. In 1516 Aleixo Garcia and a couple of other Europeans, members of the João Dias de Solis expedition to modern-day Uruguay, shipwrecked in Southern Brazil near the Island of Santa Catarina\(^2\). Staying for a decade in the region, Garcia not only learned about a system of indigenous trails (the Peabiru) that would allow him to reach the Andes, but also heard about a powerful kingdom with plenty of silver. In 1525 he departed in a military expedition with local Guaraní tribes through this path, eventually reaching and sacking parts of the Inca Empire near the Bolivian city of Sucre. Aleixo Garcia was killed on his way back, but news of the presence of silver led to increased interest and dispute for the area from both the Portuguese and Spanish crowns\(^3\).

The pre-Columbian path explored by Aleixo Garcia was part of a wider network of trails, which the indigenous referred to as the Peabiru\(^4\), that connected the Atlantic coast to the interior,

---


\(^2\)In this paper, the names of historical individuals are written in their (likely) native languages. Therefore, I use the Portuguese Aleixo Garcia instead of the Spanish Alejo Garcia. Geographical places in Brazil are written in Portuguese, except when widely known by a different word in English. Hence, I use Iguazu river instead of Iguacu river.

\(^3\)A good source for Garcia’s adventures is chapter 4, “O Outro Peru”, in Buarque de Holanda (1969). A list of many of the early explorers of the Peabiru is in Chmyz and Sauner (1971).

\(^4\)The Jesuits named the road as Saint Thomas Road, as they believed it was built by the first century apostle
crossing the Paraná river into Paraguay and then into Bolivia. Figure 2 shows the extent of these roads according to Maack (1959). A main artery was the one that started near São Vicente, then reached what is now the city of São Paulo before turning south into the state of Paraná. It then took a western direction, crossing the Paraná river a few kilometers above where the Guaiúra Falls used to be. Antonio Ruiz de Montoya, a Jesuit missionary in the 17th century, wrote this first-hand description of the trail: "a path eight spans wide, on which the grass grows very short although on either side of the path it grows nearly half a yard high" (Montoya 1993, p.75). The Peabiru soon became an important route for Europeans in the region, such as the newly appointed governor of Paraguay, Alvar Nuñez Cabeza de Vaca, who crossed it on his way to Asunción in 1541, recording the names of many Guarani villages that were located along it (Cabeza de Vaca 1906).

In 1494 the Portuguese and Spanish crowns had signed the Tordesillas Treaty, delimiting their areas of influence in the New World. Figure 1 shows where the Tordesillas Treaty line crossed the Brazilian territory, splitting the two states (São Paulo and Paraná) through which the Peabiru passed. The treaty was relatively vague and the technology to measure longitude still unavailable, so the Tordesillas line was not fully respected by settlers from either country (Cintra 2009). The European occupation of the region started from two settlements. First, a Portuguese group led by Martim Afonso de Sousa founded in 1532 the town of São Vicente at a point close to where the Peabiru reached the Atlantic Ocean, not before sending a failed expedition to traverse the path in order to reach Peru. And in 1537 the Spanish founded Asunción, from where many expeditions were sent to conquer and occupy the interior of the continent.

Since the second half of the 16th century, many European settlements were built along the Peabiru. Consider the example of São Paulo, founded in 1554 at a plateau 800m above the sea level, 60km inland from São Vicente. The Peabiru not only provided the inhabitants of São Paulo with the necessary land access to the coast, but also opened up the possibility of venturing inland, which fitted the missionary goals of the Jesuits who first occupied the town (Campos 2006). It is also important to notice that São Paulo was located in the vicinity of an indigenous village controlled by an allied Tupi chief. Actually, strong alliances with local native groups were important for the success of the Portuguese settlements in Brazil (Caldeira 2017).

At around the same time, the Spanish settled a region they named the Guairá, building in 1557 their first settlement to the east of the Paraná river (in what is now Brazil). Ciudad Real del Guairá was named after a friendly Guarani chief that controlled the region and was built at the point where the Peabiru intersected the Paraná River. This fact is precisely pointed out by Ruy Diaz Melgarejo, the leader of the expedition that established the town: it was built on "the crossing of the road from Brazil". Over time, the Spanish built other settlements in the region, the most important

---

Saint Thomas. For this interesting story, see Buarque de Holanda (1969). Montoya (1993) and Lozano (1873) are contemporary Jesuit sources. In the native language, Peabiru means "soft path", possibly due to the short grass that grew on it.

of which was Villarrica del Espíritu Santo. Its first location was on the Peabiru, as attested by Melgarejo; some years later, it was moved to a nearby location at the margins of an important river (the Ivaí). Montoya (1993) estimated that by the early 17th century there were 140 male Spaniards in Villarrica and 40 in Ciudad Real. Through the encomienda, a Spanish institution that assigned a group of indigenous inhabitants to a Spanish settler, the encomendero, they explored the labor of the much larger Guarani populations living in the surroundings. Also, starting from 1609 Spanish Jesuits established themselves in the region. The two largest and longest-lived Jesuit reductions, Lozano and San Ignacio Mini, were located at the Paranapanema river, on the northern limits of the Guairá. Figure A.1 contains a map of Spanish settlement and the two main Jesuit reductions in the Guairá.

By the early 17th century, the European presence in the region was characterized by Portuguese settlements advancing from the coast and Spanish towns that were connected to Paraguay on the West. The colonization process in both areas was similar, characterized by alliances with local tribes, intermarriage, and the exploration of the indigenous labor force. Beyond the similarities, there was also the establishment of some trade routes, facilitated by the political union between Portugal and Spain from 1580 to 1640 (Caldeira 2017).

The situation changed when the inhabitants of São Paulo organized a slave-raiding expedition into the Spanish region in 1628. At first they raided the Jesuit reductions; priests and converted natives fled to Paraguay. Later on the Portuguese invaders attacked and destroyed the Spanish towns. This invasion not only interrupted the Spanish settlement of lands to the east of the Paraná River, but also led to a shift in the ethnic composition of the indigenous population, as Guarani peoples lost preeminence to Gê peoples, who were usually more hostile to Europeans (Mota 1994, Nimuendajú 2017). Writing in the 18th century, Lozano (1873, p.76) claims the region was pretty

---

6 For a description of archaeological findings in Villarrica, see Parellada (1995). After its abandonment, the inhabitants were transferred to a site in Paraguay, where a city of the same name still exists.

7 Melgarejo writes that Villarrica was founded “to the east of Ciudad Real, 60 leagues of it, between the sources of the Piquiri and the Ivaí, on the road traversed by Alvar Nuñez (Cabeza de Vaca) and Hernando de Tejo” (Cardozo 1938, p. 49). Author’s translation from the original Spanish: “al este de Ciudad Real, a sessenta leguas de la misma, entre las nacientes del Pyquiry y del Huybay, y en el camino por donde pasaron Alvar Nuñez y Hernando de Tejo”.

8 The number of natives in the surroundings of Villarrica, and actually in any part of the Guayrá, is not clear. Lozano (1873) estimates that when Villarrica was abandoned, there was an exodus of 500 Spaniards and 4,000 Guaranies. However, the number of indigenous inhabitants was probably much higher, since all the villages surrounding Villarrica were attacked by the slave raids (Cardozo 1938).

9 See Montoya (1993) and Sarreal (2014) for a history of the Jesuit presence in the Guairá. After the Portuguese arrival, these reductions were also re-established in different areas of the Spanish empire.

10 Guarani peoples in the Spanish case, Tupi peoples in the Portuguese.

11 In 1604 the Paraguayan governor, Saavedra, sent representatives to São Paulo with the purpose of promoting trade. In 1607, in a letter to the Spanish king, Saavedra also mentions a large movement of people across the land route (Cardozo 1938). The importance of the land route connecting the Portuguese and Spanish areas can also be exemplified by the route of the newly appointed governor to Asunción, Luis de Céspedes Xeria, who stopped in São Paulo in 1828 when on his way to Paraguay (Cavenaghi 2011).

12 See Montoya (1993), Cardozo (1938) and Caldeira (2017) for references.
much depopulated, particularly when compared to his estimates of 100,000 inhabitants at the time of Cabeza de Vaca’s journey. Interestingly there is a random element in the depopulation of the Spanish area relative to the Portuguese one, as in 1628 the Spanish colonial council ordered an assault on the town of São Paulo (Caldeira 2017). The orders took some time to reach Buenos Aires, and the Portuguese raid on the Guairá made the plan unfeasible. Therefore, a different timing of how the events unfolded could have flipped the direction of the natural experiment analyzed in this paper.

The 17th century not only saw the abandonment of the Spanish towns but also the Portuguese independence from Spain in 1640, which resulted in the decline of trade between São Paulo and Paraguay. Caldeira (2017) argues that, over time, the silver trafficking route that connected Potosí through Asunción all the way to São Vicente lost relevance to a route through Tucumán and Buenos Aires. Therefore, both depopulation and the decline of trade would have reduced the use of the Peabiru in the areas of the Guairá.

It is also important to highlight that the western portion of the state of Paraná, which corresponds to the area formerly occupied by the Spanish, would be settled only in the 20th century, when it became an important frontier (Kohlhepp 2014). The bottom map in Figure A.1 shows the location of the municipality seats in 1872, which are likely to indicate the largest urban centers at the time, against the Peabiru. Notice there were only two municipality seats in the formerly Spanish region. There were many more outside of it, and they clearly clustered around the pre-colonial road network.

There are many possible reasons why western Paraná remained unsettled until the early 20th century, including the high transportation costs to the urban centers to the east and the belligerence of the Gê peoples who became the major indigenous group in the area since the Guarani exodus. In the 19th century, a series of military campaigns resulted in the concentration of the indigenous inhabitants to small reservations, effectively clearing the area for the expansion of the agricultural frontier. The state of Paraná thus became a major agricultural frontier, attracting coffee farmers from neighboring São Paulo and pioneers from the southern Brazilian state of Rio Grande do Sul (Kohlhepp 2014). In this way, the occupation of the region starting in the 20th century is unlikely to be directly related to the distribution of the pre-Columbian population.

---

13The Portuguese raid on the Guairá occurred at a time when both Portugal and Spain had the same king. To understand this situation, it is good to keep in mind that these events were motivated by local conditions at a remote part of the empire. Furthermore, the Portuguese administration was never incorporated into the Spanish administration, and the king ruled his many domains by a system of councils.

14Chmyz and Sauner (1971), during their archaeological explorations of the Peabiru in western Paraná, interviewed the local population and noticed they were unaware it consisted of an indigenous road, rather believing it was built by the army for the movement of troops in a 19th century war.

15See Mota (1994) for a history of the wars against indigenous peoples in Paraná.
3 Data

In this section, I describe the main data sources used in this paper. The Appendix contains a description of auxiliary data sources that were used in the construction of control variables.

3.1 Historical Sources

The location of the Peabiru is obtained by digitizing the map by geographer Reinhard Maack (1959). The map is based on a collection of sources, that include the first-hand description of itineraries of 16\textsuperscript{th}-century European travelers as well as colonial period maps. Maack’s map is considered the best source on the Peabiru course (Chmyz and Sauner 1971, Colavite and Barros 2017) and it has been validated by archaeological evidence (Chmyz and Sauner 1971).

In order to define the contours of the Guairá, the area settled by the Spanish in the 16\textsuperscript{th} and early 17\textsuperscript{th} centuries, I use a natural borders approach by identifying the major rivers that encapsulate Spanish settlements: the Paraná to the west, the Iguazu to the south, the Paranapanema to the north, and the Tibaji to the east. The locations of Spanish towns and major Jesuit reductions are from archaeological maps of the Instituto de Terras, Cartografia e Geociências (ITCG-PR). The natural borders approach is adopted since there is no official delimitation of the border between Portuguese and Spanish areas, whereas these rivers are all major geographical obstacles. As an alternative, I also use a wider definition of the Guairá limits from Cardozo (1938). Instead of considering the Tibaji river as the eastern border of the Guairá, Cardozo (1938) uses the meridian 49.3 west.

The sample area is defined as sites in the current Brazilian states of São Paulo and Paraná that are within a distance of 50 km to the Peabiru. For the baseline results, I exclude a small portion of the state of Paraná that lies to the south of the Iguazu river, since it is not part of the Guairá and it is also unconnected (except through the Guairá) to the area of Portuguese settlement in the east; results are robust to the inclusion of this region into the sample. The sample region covers an area of 257,052 km\textsuperscript{2} and, in 2010, it had a population of more than 36 million inhabitants.

A note about nomenclature: I often refer to the Guairá as the Spanish region. For convenience, the rest of the sample is referred to as the Portuguese region.

3.2 Economic Activity

Data on the spatial distribution of current day economic activity comes from two sources. The first is the 2010 Brazilian census conducted by the Instituto Brasileiro de Geografia e Estatística (IBGE), which contains information on the population by census areas (setor censitário) from which I calculate population densities. Census areas are the smallest spatial units for which IBGE releases the population; the census areas in the sample have an average of 327 inhabitants. They
are also classified as either urban or rural. In Brazil, instead of following uniform criteria, an area is urban if it is administratively classified as so by the municipality it is in\textsuperscript{16}. Census areas classified as urban tend to have much higher population densities than rural ones, indicating the classification might be a proxy for where cities are. Appendix Figure A.2 displays the cumulative distribution of population density in rural and urban census areas.

A second measure of the spatial distribution of economic activity is from the DMSP/OLS Nighttime Lights data set, which consists of nighttime satellite images of lights intensity at a resolution of 30 arc-seconds. Nightlights have been widely used in recent years as a proxy for economic activity (Henderson \textit{et al} 2012). For comparability with census information, I use observations for the year of 2010. An advantage of nighttime lights information is that, in contrast to the endogenous borders of census areas, it provides a finer and relatively uniform partition of the space\textsuperscript{17}. The nighttime lights image is also used to define cities through a cutoff rule as in Harari (2017). I choose the cutoff so the area occupied by cities in the sample equals the area of urban census areas as classified by IBGE\textsuperscript{18}.

### 3.3 Summary Statistics

Table 1 displays summary statistics for the sample region. Units of observation are IBGE census areas. On average, they are at a distance of 20 km to the pre-colonial path network; 27.9\% of the sample area is closer than 10 km. The Guairá, as delimited by the natural borders, occupies 32\% of the sample. Population density is 141 inhabitants per km\(^2\), mostly concentrated in the urban areas, that cover 5.8\% of the sample.

### 4 Empirical Analysis

#### 4.1 Identification Strategy and Estimation

To investigate the concentration of economic activity around the Peabiru, I use a difference-in-differences model. Treated units are sites in the Portuguese regions that are close to the Peabiru. I estimate the following model by ordinary least squares\textsuperscript{19}:

\begin{align*}
\text{dependent variable} &= \beta_0 + \beta_1 \text{urban status} + \beta_2 \text{city dummy} + \text{other controls} + \epsilon
\end{align*}

\textsuperscript{16}See IBGE (2017) for a discussion of the urban-rural classification in Brazil.

\textsuperscript{17}Average pixel area is 0.76 squared kilometer.

\textsuperscript{18}This method results in any pixel with luminosity above 32 (out of 63) being classified as belonging to a city. Harari (2017) used a cutoff of 35 for India.

\textsuperscript{19}Two independent variables, the urban status and the city dummy, are binary. I still prefer to use a linear model instead of a probit as the interpretation of interaction terms is not obvious for non-linear models (Ai and Norton, 2003). Furthermore, the results are robust to the use of a saturated specification with no control variables, for which the predictions from a linear probability model are always between zero and one.
\[ y_i = \beta_{Peabiru} \times Portu_{gi} + \beta^S_{Peabiru} \times Span_{i} + \alpha_{Span_{i}} + X'_i \theta + u_i \] (1)

where \( Peabiru_{i} \) is a dummy variable indicating proximity to the Peabiru; in the baseline specification, it equals one if the distance to the Peabiru is less than 10 km\(^{20}\). The variables \( Span_{i} \) and \( Portu_{gi} \) are two mutually exclusive dummy variables: \( Span_{i} = 1 \) indicates that the census area or nighttime light pixel is inside the Guairá, the region originally settled by the Spanish, while \( Portu_{gi} = 1 \) indicates the rest of the sample. In the baseline specification, \( X_i \) is a vector of control variables that include state dummies, 2nd-degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline. All observations are weighted by area, so the coefficients are representative of the effects on a squared kilometer. In order to account for spatial correlation, I cluster standard errors by the micro-region, a regional classification of municipalities by IBGE. Micro-regions are geographical clusters of interconnected municipalities, which makes them appealing as clusters for inference when studying the spatial distribution of economic activity. The sample is covered by 59 micro-regions.

Dependent variables are all measures of how concentrated economic activity is in the area. From IBGE census data, I use the population density, the inverse hyperbolic sine of population density\(^{21}\), and a binary variable for the rural-urban status. From nighttime lights data, I use both the levels and the inverse hyperbolic sine of light intensity, and the classification of the grid into cities based on the cutoff rule described in Section 3.

Note that \( \beta \) is the coefficient on proximity to the Peabiru in the sample area that was never occupied by the Spanish (the Portuguese region). If there is concentration of economic activity around the pre-colonial paths only through the fact that the paths are close to favorable geographic fundamentals, then a positive relationship will hold both in the Portuguese and Spanish regions: \( \beta = \beta^S \). If, on the other hand, the Peabiru and the indigenous populations who lived around it attracted early European settlements that persisted due to path dependence in the location of cities, then the persistence should have been broken in the former Spanish Guairá, so \( \beta > \beta^S \). Therefore, \( \beta - \beta^S \) identifies the causal effect of the Peabiru on the density of economic activity.

But should we be worried about a correlation of the Peabiru with geographic fundamentals? I investigate this question by performing a balance test of observable fundamentals; the results are shown in Table 2. In this exercise, the dependent variables are the following\(^{22}\): (1) the average elevation; (2) the average terrain ruggedness index; (3) the number of days a year that support malaria transmission; (4) the inverse hyperbolic sine of the potential yield of coffee, which was an important crop for the expansion of the agricultural frontier of the states of São Paulo and Paraná; (5) the logarithm of the maximum calories that can be produced per hectare with indigenous staple

\(^{20}\)All distances are measured from the centroid of the census area or nighttime light pixel.

\(^{21}\)There are some zero population census areas, so the logarithm is not well defined.

\(^{22}\)See the Data Appendix for a description of the sources of these variables.
crops (beans, cassava, maize, and sweet potatoes); (6) same as (5), but including two common crops brought by Europeans: rice and wheat. In Panel A, I simply regress the fundamental on the dummy of distance to the Peabiru; in Panel B, I include the dummy for the Guairá and the interaction with proximity to the Peabiru. I cannot reject the null hypotheses that the fundamentals correlate with proximity to the Peabiru. This is the case both inside and outside the Spanish region; it is not possible to reject the hypotheses that the coefficients are the same in both regions.

It is then a conservative strategy to account for geographic fundamentals by comparing the coefficients in the Portuguese and the Spanish regions. A reason to do so is that there are many more unobservable fundamentals than observable ones. Furthermore, note that the coefficient $\beta^S$ on proximity to the path in the Guairá should be zero if there is no relationship of the path with favorable geographic fundamentals. Indeed, this is what I find.

### 4.2 Main Results

Before discussing the estimates of the coefficients in Equation 1, in Figure 3 I show a map of the Peabiru against night lights in 2010\textsuperscript{23}. The area to the southwest of the dashed red line corresponds to the old Spanish department of the Guairá. Note that the largest urban concentrations of the Portuguese area are all on the Peabiru: São Paulo and Curitiba near the northeastern and the southeastern corners of the map. Note also that in the Portuguese region there is a concentration of nightlights around the pre-colonial paths even far from these two large cities. Inside the area originally settled by the Spanish, on the other hand, there is apparently no relationship between nightlights and the Peabiru.

Table 3 reports the estimated coefficients of equation 1. In Panel A, I do not include controls. Column (1) shows that proximity to the Peabiru increases average population density by 154 inhabitants per km\(^2\) in the Portuguese region, but only by 12 inhabitants per km\(^2\) in the control region. The same pattern can be seen in column (2), which shows the coefficients when the dependent variable is the inverse hyperbolic sine of population density: there is a positive coefficient of 0.396 in the Portuguese region, but a negative (although not statistically significant) coefficient of -0.202 in the Spanish area. In this case, coefficients are more precisely estimated, and we can reject with a confidence level of 1% that they are the same in both regions. We can also use census data to learn the effects on urbanization: the coefficients in column (3) indicate that, in the Portuguese region, proximity to the path increases the probability that a site is urban by 5 p.p., while the coefficient is close to zero in the Guairá.

The results are much the same when nightlights are used. Relative to census areas, the nightlights grid is composed of cells of similar size. Since distance is measured from the centroid of the

---

\textsuperscript{23}The projection of the map is equidistant conic with reference parallels 26S and 21S, so distances are not much distorted.
cell, uniform sizes imply less measurement error. On the other hand, nightlights are top-coded, so they capture less of the variation that comes from densely populated urban areas. In columns (4) and (5), the dependent variable is the level and inverse hyperbolic sine of nightlights luminosity. In column (5), I consider the city dummy. Again, the estimates show the same patterns: proximity to the Peabiru has a strong effect on density only in the Portuguese region. The coefficients in the control region are always statistically indistinguishable than zero. Notice that the coefficients in columns (5) and (6) are close to the ones in columns (2) and (3), so the effects on density and urbanization are similar regardless of measuring them in the census or nightlights data.

In Panel B of Table 3, I include the baseline controls. Despite an increase in the explanatory power of the model (as measured by the R-squared), the coefficients on proximity to the Peabiru barely change, and neither do the statistical tests. After conditioning on observables, we still observe the same patterns.

### 4.3 Robustness

I perform a set of exercises to show that the results in Table 3 are robust to a variety of concerns.

The first set of robustness exercises relate to the model specification; they are shown in Appendix Table A.1. In Panel A, I use Cardozo’s (1938) definitions of the Guairá limits, as opposed to the natural borders approach used in the baseline. The coefficients change only slightly, but they are less precisely estimated. For two of the dependent variables, the level of the population density and the nightlights city dummy, the difference between the coefficients becomes borderline insignificant. Another robustness exercise about model specification consists of adding more controls on top of the baseline ones. In Panel B, I include all the geographic fundamentals that are dependent variables in Table 2: malaria transmission days, coffee potential yields, and the calories that can be produced with indigenous and European staple crops. In Panel C, I control for unobservable fundamentals by including instead a third degree polynomial of latitude and longitude. Such flexible specification accounts for regional differences in the determinants of density, such as the fact that regions that are farther away from the coast are less dense. In either case, the effects of the treatment either remain the same or increase.

A hypothesis behind the identification strategy is that not only is the correlation between unobservable geographic fundamentals and the Peabiru equal in the Spanish and Portuguese regions, but the effect of these fundamentals on the density of economic activity is also the same in both regions. However, since the Spanish settlement area was abandoned and settled much later than the Portuguese one, it is possible that the effects of different geographic fundamentals on the spatial distribution of population differ across the two areas. To understand whether this is a concern, in

---

24 Since these variables are constructed based on the contemporary climate and soil, I prefer not to include them among the baseline controls. They are potentially endogenous.

25 Henderson et al (2017) show that, when comparing countries that industrialized earlier or later, agricultural suit-
Appendix Table A.2 I include observable geographical attributes interacted with the Spanish settlement dummies as controls. In Panel A, I include only the control variables used in the baseline specification, while in Panel B the variables analyzed in Table 2 (caloric potential, coffee potential yield, malaria days) are also included. Results are robust to the interaction of the controls with the Spanish settlement dummy.

In Appendix Table A.3, I show robustness to different sample selection procedures. In Panel A, the sample consists of sites up to 30km of the path, instead of 50km as I used in the baseline. Results hold in this case. In Panel B, I exclude sites within the municipalities of São Paulo and Curitiba; this is to address concerns that the results might be simply capturing the fact that these two state capitals, which are also the largest cities in the sample, were built along the Peabiru. The coefficients are similar to the baseline ones, indicating the results are not driven by the two cities. Another concern is that the sample area covers two states, but Spanish settlement occurred only in a portion of Paraná. If the concentration of urban activity around the path is different in Paraná than in São Paulo due to reasons other than the abandonment of the Spanish settlements, such as state level institutions and policies, they could be generating the patterns estimated so far. To show it is not the case, I estimate equation 1 on a restricted sample that consists only of the state of Paraná. Panel C displays these estimates. Despite being noisier, the coefficient of interest is now larger; results are robust to this sample restriction. Finally, in Panel D I include the portion of the state of Paraná that is to the south of the Iguazu river; coefficients are hardly affected.

The results are also robust to the use of different measures of proximity to the Peabiru. So far, I showed results when closeness to the pre-Columbian road network is an indicator if the distance to the road is less than 10km. In Appendix Table A.4, I show results when two alternative buffers are used: 5km (Panel A) and 15km (Panel B). Reassuringly, the effects are larger when using a 5km cutoff. Furthermore, in Appendix Figure A.3 I estimate a non-parametric version of the main equation, in which a step function of distance to the Peabiru is used instead. The graphs indicate that the positive effects of proximity to the path decay after the first 10 or 15km.

Since this paper is based on historical sources, I also discuss robustness to measurement error. The routes of the Peabiru were mapped by Reinhardt Maack (1959) many years after it was no longer used, so the main issue is that distance to the pre-Columbian road is measured only with error. As long as the measurement error problem in the control region is not more severe than it is in the Portuguese one, the results will be robust to it. However, if measurement error is stronger in the Spanish region, it will bias $\beta^S$ towards zero, increasing $\beta - \beta^S$. A simple solution to this problem is to use a larger grid as the unit of observation. I do so by aggregating the data to a grid ability relative to market access matters relatively more in the former.

---

26Since this restriction reduces the number of clusters, I use a wild bootstrap procedure to make inference on $\beta^S$ and $\beta^P$. See Cameron and Miller (2015).

27For some dependent variables, the point estimates for $\beta^S$ are negative. Hence, it is possible that measurement error actually leads to underestimates of $\beta - \beta^S$. 

13
of 20 minutes of latitude by 20 minutes of longitude. Now $Spain_{i}$, $Portug_{i}$ and $Peabiru_{i}$ are no longer dummy variables, but the share of each grid cell that is, respectively, in the Spanish region, the Portuguese region, or within 10km to the Peabiru. The estimated coefficients are shown in Panel A of Table A.5. They are noisier due to a loss of power, but for most dependent variables the coefficients of proximity to the path in the Portuguese region are statistically different than the one in the Spanish region at a 15% significance level. Note the coefficients are also larger, which indicates that measurement error could be an issue, reducing the estimated coefficients of the baseline equation.

Finally, I account for concerns of spatial correlation of the errors by experimenting with different inference methods. In Panel A of Appendix Table A.6, I report Conley (1999) standard errors. The standard errors are estimated assuming a uniform kernel that allows arbitrary correlation between the errors of observations that are up to 200km apart. In Panel B, I cluster the standard errors according to a grid of 1 degree of latitude by 1 degree of longitude. Results are robust to these alternative inference methods.

5 Mechanisms

The results discussed so far show that a pre-Columbian transportation network, the Peabiru, has a positive causal effect on urbanization and population density. It is natural to imagine that the effects are mediated by the modern transportation network, as ancient indigenous paths could have been eventually turned into modern roads. Dalgaard et al (2018) show a positive relationship between ancient (Roman) and modern transportation networks in Europe. On the other hand, Burghardt (1969) shows that few indigenous paths of Canada’s Niagara Peninsula became modern roads. What is the case for the Peabiru? Did it become part of the modern road network? And, if so, is it what explain the results?

I examine these questions by using georeferenced data on the main roads of Brazil in 2010; see the Appendix for details on the data sources and definitions. I first show the effects of the Peabiru on modern roads by estimating equation 1 when the dependent variable is a dummy equals one if the census area or the grid cell is within 10km of a modern road. Results are shown in Panel A of Table 4. In column (1) the observation units are census areas, while in column (2) they are nightlights grid cells. Around three-quarters of the sample area lies within 10km of a modern road. The results in column (2) imply that proximity to the Peabiru increases the probability a site is within 10km of a modern road by 8.5 p.p. in the Portuguese region and decreases it by 3.9 p.p. in the former Guairá. However, the coefficients are noisily estimated, and their difference is marginally statistically insignificant. But it is important to keep in mind that, although I cannot reject the hypothesis that the Peabiru had no effect on modern roads, they could actually be quite large; the upper bound of the 95% confidence interval for the difference $\beta - \beta^S$ corresponds to an
effect of 27.6 p.p..

I consider two questions regarding the intermediation by modern roads of the effects of the Peabiru on urbanization and population density:

1. **Is there any mediation through modern roads?** This question can be answered with a Sobel test. The Sobel test is based on the observation that the mediated effect is captured by the product of two coefficients: the effect of the Peabiru on modern roads (captured by \( \beta - \beta^S \)), and the coefficient on proximity to modern roads in a “horse race” equation that contains both proximity to the Peabiru and to modern roads. The coefficients for the “horse race” equation and the Sobel test statistic are shown in Panel B of Table 4. The t-statistic is low for all dependent variables, so I cannot reject the null hypothesis that there is no mediation. The coefficients indicate that only 11% to 18% of the effects of the Peabiru are mediated by modern roads.

2. **Is there any effect that does not operate through modern roads?** I examine this question through a simple exercise with a logic that can be seen in the figure below. The blue arrow represents a direct effect of the Peabiru, which does not occur through modern roads. The red arrows, on the other hand, indicate the indirect effect: how the Peabiru leads to modern roads, and then how modern roads affect population density. There is also a third arrow, which represents the reverse effect, from modern population density to modern roads. As roads are usually built to connect areas with significant economic activity, the reverse effect is likely to be positive. Hence, OLS estimates of the coefficient of roads on population density are likely to be larger than the true causal effect of roads.

![Diagram](image)

The question I am interested in is whether there is a direct effect. I do so by showing that the causal effect of modern roads on density and urbanization are too large to explain all the effects of the Peabiru. Such exercise is performed by considering the situation in which the direct effect is zero, so all effects of the pre-colonial paths are mediated by the modern
transportation network. In this case, closeness to the Peabiru interacted with the Portuguese region is a valid instrument to identify the causal effect of modern roads.\footnote{Redding and Turner (2015) denote this identification strategy as the historical route instrumental variable approach. See Duranton and Turner (2012) and Martincus et al (2017) for examples of papers that use this approach.}

Unfortunately, the coefficients shown in Panel A of Table 4 indicate that such instrumental variable would be weak. It is thus not much informative to use it to derive IV estimates of the impact of modern roads on the density of economic activity. Nevertheless, Table 4 gives enough information for a back-of-the-envelope calculation of how large the effects of roads on the spatial distribution of economic activity would be for all the effects of the Peabiru to be mediated by modern roads. For a conservative calculation, I combine the 95% upper bound of 0.276 on $\beta - \beta^S$ that is shown in Panel A with the baseline results from Table 3 to obtain “implied IV” estimates of the effects of modern roads.\footnote{Proximity to the Peabiru still might capture unobservable fundamentals, so it has to be used as an included instrument.} They are shown in Panel C of Table 4. Note they are at least twice as large as the OLS coefficients for the effects of modern roads on the density of economic activity. Since the OLS coefficients are likely to be over-estimates of the causal effect, the results suggest the existence of a direct effect.

In this section, I asked how much of the effects of the Peabiru are mediated by its conversion into modern roads, and how much is direct. The results are too noisy for a strong conclusion on whether there was some mediation through modern roads; a Sobel test does not reject the null hypothesis that there was no mediation. On the other hand, it is clear that the conversion of pre-Columbian trails into modern roads does not tell us the whole picture. There must be a direct effect.

What is it? The historical evidence discussed in Section 2 contains many cases of European settlements that were built along the indigenous paths. The settlements proved persistent, including many of the important cities in the sample.\footnote{Such as Itu, Sorocaba, Botucatu and Curitiba} Over time, agglomeration economies could have locked economic activity in these sites (Fujita et al. 1999, Allen and Donaldson 2018). Alternatively, they could have become focal points that coordinated expectations in a later period when agglomeration economies became more important, such as during the surge in urbanization starting from the late 19th century.

\section{Conclusion}

This paper contributes to our understanding of the roots of city location in the New World. As the European conquest resulted in a drastic change in the technology, the demography, and the trade

\textsuperscript{28}Redding and Turner (2015) denote this identification strategy as the historical route instrumental variable approach. See Duranton and Turner (2012) and Martincus et al (2017) for examples of papers that use this approach.

\textsuperscript{29}Proximity to the Peabiru still might capture unobservable fundamentals, so it has to be used as an included instrument.

\textsuperscript{30}The “implied IV” estimates are simply the ratio between the effects ($\beta - \beta^S$) on density or urbanization over the effects on modern roads.

\textsuperscript{31}Such as Itu, Sorocaba, Botucatu and Curitiba
links of the continent, it is not clear whether modern cities would be located in sites that were economically relevant during the pre-Columbian period. Still, there is plenty of historical evidence that proximity to indigenous settlements was important for early colonizers (Caldeira 2017). Many of the cities in the continent, including the two largest (Mexico City and São Paulo), are located near the site of indigenous villages or cities.

However, such observation does not tell us if the effect is causal, as persistence could be due to characteristics of the sites that mattered both today and in the past. In this paper, I address the identification challenge by studying a unique historical experiment. The invasion, depopulation and subsequent abandonment of the Guairá allowed me to construct a counterfactual of how city placement would look like if the geography of pre-colonial societies did not matter. I then show that, while proximity to a network of indigenous paths is associated with significant increases in population density and urbanization, no such relationship exists in the formerly Spanish region. A variety of empirical exercises indicate that these patterns are robust and likely to capture a causal effect.

A limitation of a natural experiment is that it only refers to local causal effects: the geographic features of pre-colonial societies had an effect on city placement in the region analyzed, but the effects they had on the rest of the continent are not clear. Probably, the Americas have regions that resemble the Guairá: by being settled later, any persistence was broken and the contemporary spatial distribution of economic activity will look quite different than the pre-colonial one. This is likely to be the case in low population density colonies, where an “expansion to the West” occurred much later, after independence.32 On the other hand, many regions of the continent were settled as early as the region I analyze: of the 10 largest cities in the continent, 8 were founded in the 16th century or earlier.33 In these areas, city location was possibly affected by the geography of pre-colonial societies. Further research is needed to show where there is a causal relation.

Finally, it is important to highlight that the effects of the Peabiru on density are quite substantial. A good comparison is with the importance of proximity to the shoreline. In the Portuguese region of the sample, average population density increases from 146.6 to 302.4 inhabitants per km² when comparing sites that are farther or closer than 10km to the Peabiru. The increase is smaller, of 188.9 to 269.9 inhabitants per km², when I do the same exercise with distance to the shoreline. This is a surprising comparison, given that Brazil, as the United States (Rappaport and Sachs 2003), has its population concentrated near the coast. The effects of proximity to the indigenous paths are not only causal; their magnitude is important.

---

32 Examples include Argentina, Canada, most of central Brazil, and the west and midwest of the USA.
33 They are São Paulo, Mexico City, Lima, Bogotá, Rio de Janeiro, Santiago, Caracas, and Buenos Aires. The two exceptions are New York and Los Angeles.
References


A Figures

Figure 1: Tordesillas Treaty line

Note: Brazil map denoting Spanish *de jure* areas in red and Portuguese *de jure* areas in green. The vertical line is the Tordesillas Treaty line, according to Cintra (2009). The two states highlighted are São Paulo and Paraná.
Figure 2: *Peabiru* and study area

Note: Peabiru is represented by the solid lines in green. The dotted black line indicates the contours of the states of São Paulo and Paraná. The Spanish settlement region (Guairá) is the area within the dashed red line. The orange vertical line is the Tordesillas Treaty line, according to Cintra (2009). The light blue lines represent the main rivers in the sample area. See the paper and the Data Appendix for the data sources. Albers Equal Area projection, with reference parallels 26S and 21S.
Note: Peabiru is represented by the solid lines in green. The dotted black line indicates the contours of the states of São Paulo and Paraná, whereas the Spanish settlement region is the area to the southwest of the dashed red line. In shades of orange, nighttime luminosity from the DMSP/OLS Nighttime Lights dataset. Equidistant Conic projection, with reference parallels 26S and 21S.
# B Tables

Table 1: Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to Peabiru (km)</td>
<td>60262</td>
<td>20.5</td>
<td>13.7</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km (%)</td>
<td>60262</td>
<td>27.9</td>
<td>44.9</td>
</tr>
<tr>
<td>Guairá (%)</td>
<td>60262</td>
<td>32.0</td>
<td>46.6</td>
</tr>
<tr>
<td>Dist. to Tordesillas Line</td>
<td>60262</td>
<td>161.5</td>
<td>193.1</td>
</tr>
<tr>
<td>Population Density in 2010</td>
<td>60262</td>
<td>141.9</td>
<td>1504.1</td>
</tr>
<tr>
<td>Urban District (%)</td>
<td>60262</td>
<td>5.8</td>
<td>23.3</td>
</tr>
<tr>
<td>State of Paraná (%)</td>
<td>60262</td>
<td>46.3</td>
<td>49.9</td>
</tr>
<tr>
<td>Distance to Shoreline (km)</td>
<td>60262</td>
<td>286.9</td>
<td>169.6</td>
</tr>
<tr>
<td>Distance to Main River (km)</td>
<td>60262</td>
<td>24.4</td>
<td>18.0</td>
</tr>
<tr>
<td>Mean Altitude (m)</td>
<td>60262</td>
<td>563.4</td>
<td>223.0</td>
</tr>
<tr>
<td>Mean Terrain Ruggedness (m)</td>
<td>60262</td>
<td>10.9</td>
<td>6.6</td>
</tr>
<tr>
<td>Area (km2)</td>
<td>60262</td>
<td>82.7</td>
<td>83.5</td>
</tr>
<tr>
<td>Population, 2010</td>
<td>60262</td>
<td>327.0</td>
<td>263.5</td>
</tr>
</tbody>
</table>

Note: The units of observation are the IBGE census areas. All statistics, except area and population, are weighted by area.
### Table 2: Balance of geographic fundamentals

<table>
<thead>
<tr>
<th></th>
<th>Average Elevation (m)</th>
<th>Terrain Ruggedness Index (m)</th>
<th>Malaria Transmission Days</th>
<th>IHS Coffee Potential Yield</th>
<th>Log Pot. Calories (Indigenous)</th>
<th>Log Pot. Calories (European)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: average correlation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km</td>
<td>-1.428</td>
<td>0.274</td>
<td>-2.984</td>
<td>-0.020</td>
<td>-0.041</td>
<td>-0.051</td>
</tr>
<tr>
<td></td>
<td>(22.183)</td>
<td>(0.490)</td>
<td>(4.871)</td>
<td>(0.013)</td>
<td>(0.049)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.005</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Panel B: correlation by region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>-1.481</td>
<td>0.575</td>
<td>-3.065</td>
<td>-0.018</td>
<td>-0.049</td>
<td>-0.081</td>
</tr>
<tr>
<td></td>
<td>(27.244)</td>
<td>(0.521)</td>
<td>(5.860)</td>
<td>(0.014)</td>
<td>(0.063)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>-0.942</td>
<td>-0.478</td>
<td>-4.145</td>
<td>-0.034</td>
<td>-0.040</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(33.801)</td>
<td>(1.026)</td>
<td>(8.186)</td>
<td>(0.021)</td>
<td>(0.047)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>Spanish</td>
<td>2.824</td>
<td>0.104</td>
<td>-10.933</td>
<td>-0.085**</td>
<td>-0.159</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(52.593)</td>
<td>(1.459)</td>
<td>(14.083)</td>
<td>(0.034)</td>
<td>(0.112)</td>
<td>(0.112)</td>
</tr>
<tr>
<td>β − β&lt;sup&gt;S&lt;/sup&gt;</td>
<td>-0.538</td>
<td>1.053</td>
<td>1.081</td>
<td>0.016</td>
<td>-0.008</td>
<td>-0.111</td>
</tr>
<tr>
<td>p-value β = β&lt;sup&gt;S&lt;/sup&gt;</td>
<td>0.990</td>
<td>0.363</td>
<td>0.916</td>
<td>0.521</td>
<td>0.915</td>
<td>0.193</td>
</tr>
<tr>
<td>Dependent Variable Mean</td>
<td>564.176</td>
<td>10.867</td>
<td>309.995</td>
<td>0.200</td>
<td>1.268</td>
<td>1.402</td>
</tr>
<tr>
<td>Dependent Variable Std. Dev.</td>
<td>232.774</td>
<td>7.410</td>
<td>58.775</td>
<td>0.128</td>
<td>0.539</td>
<td>0.502</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.000</td>
<td>0.001</td>
<td>0.008</td>
<td>0.110</td>
<td>0.019</td>
<td>0.006</td>
</tr>
<tr>
<td>Observations</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
</tr>
</tbody>
</table>

Note: The dependent variable in column (1) is the average elevation in a 1’ × 1’ grid cell, while in column (2) it is the average terrain ruggedness index in the same grid cell. The dependent variable in column (3) is the average of the number of days a year that support *P. falciparum* and *P. vivax* transmission. In column (4), the dependent variable is the inverse hyperbolic sine of the potential yield of coffee. In column (5), it is the logarithm of the maximum calories that can be produced with crops in the indigenous diet (beans, cassava, maize, sweet potatoes), while in column (6) I also include rice and wheat. All observations weighted by area. In the parentheses, robust standard errors clustered by micro-region. Notice β is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and β<sup>S</sup> is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. Statistical significance denoted by: * 10%, ** 5%, *** 1%
Table 3: Main results

<table>
<thead>
<tr>
<th></th>
<th>Pop. Density (1)</th>
<th>IHS Pop. Density (2)</th>
<th>Urban Area (3)</th>
<th>Lights (4)</th>
<th>IHS Lights (5)</th>
<th>City (Lights ≥ 32) (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: no controls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>154.303*</td>
<td>0.396***</td>
<td>0.050***</td>
<td>3.424***</td>
<td>0.442***</td>
<td>0.045**</td>
</tr>
<tr>
<td></td>
<td>(77.385)</td>
<td>(0.108)</td>
<td>(0.017)</td>
<td>(1.179)</td>
<td>(0.127)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>12.222</td>
<td>-0.202</td>
<td>0.007</td>
<td>0.046</td>
<td>-0.051</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(22.929)</td>
<td>(0.155)</td>
<td>(0.010)</td>
<td>(0.855)</td>
<td>(0.129)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Spanish</td>
<td>-111.524</td>
<td>0.256</td>
<td>-0.045*</td>
<td>-3.310**</td>
<td>-0.436**</td>
<td>-0.045*</td>
</tr>
<tr>
<td></td>
<td>(89.525)</td>
<td>(0.197)</td>
<td>(0.024)</td>
<td>(1.536)</td>
<td>(0.191)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>$\beta - \beta^S$</td>
<td>142.081</td>
<td>0.598</td>
<td>0.043</td>
<td>3.378</td>
<td>0.494</td>
<td>0.040</td>
</tr>
<tr>
<td>p-value $\beta = \beta^S$</td>
<td>0.083</td>
<td>0.003</td>
<td>0.030</td>
<td>0.024</td>
<td>0.008</td>
<td>0.070</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.004</td>
<td>0.011</td>
<td>0.020</td>
<td>0.034</td>
<td>0.039</td>
<td>0.018</td>
</tr>
<tr>
<td><strong>Panel B: baseline controls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>163.096*</td>
<td>0.387***</td>
<td>0.051***</td>
<td>3.761***</td>
<td>0.493***</td>
<td>0.049**</td>
</tr>
<tr>
<td></td>
<td>(85.291)</td>
<td>(0.112)</td>
<td>(0.018)</td>
<td>(1.222)</td>
<td>(0.124)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>16.037</td>
<td>-0.164</td>
<td>0.009</td>
<td>0.011</td>
<td>-0.066</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(31.005)</td>
<td>(0.181)</td>
<td>(0.011)</td>
<td>(0.992)</td>
<td>(0.134)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Spanish</td>
<td>109.742</td>
<td>0.241</td>
<td>0.043</td>
<td>0.884</td>
<td>0.053</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(119.488)</td>
<td>(0.372)</td>
<td>(0.033)</td>
<td>(1.922)</td>
<td>(0.240)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>$\beta - \beta^S$</td>
<td>147.059</td>
<td>0.551</td>
<td>0.042</td>
<td>3.751</td>
<td>0.559</td>
<td>0.044</td>
</tr>
<tr>
<td>p-value $\beta = \beta^S$</td>
<td>0.088</td>
<td>0.013</td>
<td>0.047</td>
<td>0.021</td>
<td>0.003</td>
<td>0.075</td>
</tr>
<tr>
<td>Dependent Variable Mean</td>
<td>141.271</td>
<td>2.289</td>
<td>0.058</td>
<td>5.919</td>
<td>1.068</td>
<td>0.058</td>
</tr>
<tr>
<td>Dependent Variable Std. Dev.</td>
<td>1435.676</td>
<td>1.520</td>
<td>0.233</td>
<td>12.811</td>
<td>1.592</td>
<td>0.234</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.014</td>
<td>0.093</td>
<td>0.077</td>
<td>0.087</td>
<td>0.092</td>
<td>0.053</td>
</tr>
<tr>
<td>Observations</td>
<td>60262</td>
<td>60262</td>
<td>60262</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
</tr>
</tbody>
</table>

Note: The dependent variable in column (1) is the population density of the census area, winsorizing the top and bottom 1%. In column (2), the dependent variable is the inverse hyperbolic sine of population density. In column (3), it is a binary variable equals to one if the census area is classified as urban by IBGE. In column (4) and (5), the dependent variables are the level and the inverse hyperbolic sine of the luminosity measured for each pixel. In column (6), it is a dummy variable if luminosity is greater or equal than 32; it is a proxy for where cities are located, see main text. In Panel B, control variables include state fixed effects, 2nd degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline. All observations weighted by area. In the parentheses, robust standard errors clustered by micro-region. Notice $\beta$ is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and $\beta^S$ is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. Statistical significance denoted by: * 10%, ** 5%, *** 1%
Table 4: Mediation by modern roads

Panel A: effects of the Peabiru on modern roads

<table>
<thead>
<tr>
<th></th>
<th>Census areas</th>
<th>Nightlights grid cells</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>0.054 (0.037)</td>
<td>0.085** (0.034)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>-0.057 (0.076)</td>
<td>-0.039 (0.071)</td>
</tr>
<tr>
<td>β − βS</td>
<td>0.111</td>
<td>0.123</td>
</tr>
<tr>
<td>p-value β = βS</td>
<td>0.190</td>
<td>0.119</td>
</tr>
<tr>
<td>Upper bound (95% CI) on β − βS</td>
<td>0.276</td>
<td>0.276</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Pop. Density (1)</th>
<th>IHS Pop. Density (2)</th>
<th>Urban Density (3)</th>
<th>Lights (4)</th>
<th>IHS Lights (5)</th>
<th>City (Lights ≥ 32) (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dist. to Peabiru ≤ 10km</td>
<td>25.382 (30.797)</td>
<td>-0.118 (0.148)</td>
<td>0.013 (0.010)</td>
<td>0.256 (0.812)</td>
<td>-0.027 (0.102)</td>
<td>0.007 (0.011)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>128.847 (78.009)</td>
<td>0.462** (0.186)</td>
<td>0.034* (0.020)</td>
<td>2.967** (1.467)</td>
<td>0.434*** (0.162)</td>
<td>0.035 (0.023)</td>
</tr>
<tr>
<td>Distance to road ≤ 10km</td>
<td>163.774* (86.566)</td>
<td>0.805*** (0.130)</td>
<td>0.069*** (0.021)</td>
<td>6.359*** (1.154)</td>
<td>1.013*** (0.116)</td>
<td>0.071*** (0.019)</td>
</tr>
<tr>
<td>Sobel test t-statistic</td>
<td>1.083 (1.036)</td>
<td>1.292 (1.170)</td>
<td>1.226 (1.119)</td>
<td>1.516 (1.187)</td>
<td>1.552 (1.187)</td>
<td>1.452 (1.187)</td>
</tr>
<tr>
<td>Share mediated</td>
<td>11.00% (11.00%)</td>
<td>13.95% (13.95%)</td>
<td>15.42% (15.42%)</td>
<td>17.25% (17.25%)</td>
<td>18.23% (18.23%)</td>
<td>16.56% (16.56%)</td>
</tr>
</tbody>
</table>

Panel C: OLS results and implied IVs

|                        | 167.364* (87.610) | 0.818*** (0.131)      | 0.070*** (0.021)   | 6.451*** (1.145) | 1.026*** (0.114) | 0.073*** (0.019)       |
|                        | 532.822 (60262)   | 1.996 (60262)         | 0.152 (60262)      | 13.591 (327795)  | 2.025 (327795)   | 0.159 (327795)         |
| Observations           | 60262 (60262)     | 60262 (60262)         | 60262 (327795)     | 327795 (327795)  | 327795 (327795)  | 327795 (327795)        |

Note: The dependent variable is binary variable equals to one if the census area (in column 1) or the grid cell (column 2) is within 10km of a modern road. Control variables include state fixed effects, 4th degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline. All observations weighted by area. In the parentheses, robust standard errors clustered by micro-region. Notice β is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and βS is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. The upper and lower bounds refer to a 95% confidence interval for β − βS. Statistical significance denoted by: * 10%, ** 5%, *** 1%
C Appendix Figures and Tables

Figure A.1: Historical maps

(a) Spanish settlements
(b) 1872 municipality seats

Note: Peabiru is represented by the solid lines in green. Left map: the three Spanish towns in red; in blue, the two largest Jesuit misiones, Loreto and San Ignacio Mini. Coordinates from the Instituto de Terras, Cartografia e Geociências (ITCG-PR). Right map: Municipality seats in the 1872 census in red. Coordinates from IBGE. Equidistant Conic projection, with reference parallels 26S and 21S.
Figure A.2: Population density in rural and urban areas

Note: Data from the 2010 census. The horizontal axis is the inverse hyperbolic sine of population density. The cumulative distribution is conditional on urban or rural classification. Census areas are weighted by area.
Figure A.3: Non-parametric effects of proximity

(a) Panel A: population density in levels (left) and IHS (right)

(b) Panel B: nightlights in levels (left) and IHS (right)

(c) Panel C: urban area (left) and city dummy (right)

Note: The solid line represents the estimates for the difference between the coefficient for the step function in the Portuguese and Spanish regions. The step size is 2km, and the omitted group are observations at a distance to the Peabiru that is within the interval (48, 50]. Control variables include state fixed effects, 2nd degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline. All observations weighted by area. Robust standard errors clustered by microregion. The dotted line indicates the 90% confidence interval.
Table A.1: Robustness: alternative specifications

<table>
<thead>
<tr>
<th></th>
<th>Pop. Density (1)</th>
<th>IHS Pop. Density (2)</th>
<th>Urban Area (3)</th>
<th>Lights (4)</th>
<th>IHS Lights (5)</th>
<th>City (Lights ≥ 32) (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Cardozo’s definition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10 km × Portuguese</td>
<td>170.989*</td>
<td>0.377***</td>
<td>0.052***</td>
<td>3.471***</td>
<td>0.445***</td>
<td>0.045**</td>
</tr>
<tr>
<td></td>
<td>(94.749)</td>
<td>(0.109)</td>
<td>(0.018)</td>
<td>(1.163)</td>
<td>(0.116)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10 km × Spanish</td>
<td>19.263</td>
<td>-0.086</td>
<td>0.013</td>
<td>0.879</td>
<td>0.077</td>
<td>0.015</td>
</tr>
<tr>
<td>Spanish</td>
<td>(24.456)</td>
<td>(0.156)</td>
<td>(0.010)</td>
<td>(0.937)</td>
<td>(0.141)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>β − β²</td>
<td>-36.253</td>
<td>-0.310</td>
<td>0.008</td>
<td>-4.114*</td>
<td>-0.626**</td>
<td>-0.048</td>
</tr>
<tr>
<td>p-value β = β²</td>
<td>(92.728)</td>
<td>(0.365)</td>
<td>(0.043)</td>
<td>(2.129)</td>
<td>(0.253)</td>
<td>(0.032)</td>
</tr>
<tr>
<td></td>
<td>151.725</td>
<td>0.463</td>
<td>0.039</td>
<td>2.592</td>
<td>0.367</td>
<td>0.030</td>
</tr>
<tr>
<td><strong>Panel B: controls include fundamentals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10 km × Portuguese</td>
<td>162.347*</td>
<td>0.387***</td>
<td>0.050***</td>
<td>3.978***</td>
<td>0.528***</td>
<td>0.051**</td>
</tr>
<tr>
<td></td>
<td>(85.373)</td>
<td>(0.101)</td>
<td>(0.016)</td>
<td>(1.177)</td>
<td>(0.118)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10 km × Spanish</td>
<td>-14.049</td>
<td>-0.278</td>
<td>0.000</td>
<td>-0.264</td>
<td>-0.107</td>
<td>0.001</td>
</tr>
<tr>
<td>Spanish</td>
<td>(31.695)</td>
<td>(0.196)</td>
<td>(0.011)</td>
<td>(0.919)</td>
<td>(0.123)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>β − β²</td>
<td>148.298</td>
<td>0.109</td>
<td>0.050</td>
<td>4.243</td>
<td>0.635</td>
<td>0.050</td>
</tr>
<tr>
<td>p-value β = β²</td>
<td>0.062</td>
<td>0.004</td>
<td>0.010</td>
<td>0.006</td>
<td>0.001</td>
<td>0.033</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.032</td>
<td>0.159</td>
<td>0.126</td>
<td>0.124</td>
<td>0.124</td>
<td>0.076</td>
</tr>
<tr>
<td><strong>Panel C: 3rd degree polynomial of coordinates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10 km × Portuguese</td>
<td>197.392***</td>
<td>0.421***</td>
<td>0.059***</td>
<td>4.494***</td>
<td>0.569***</td>
<td>0.060***</td>
</tr>
<tr>
<td></td>
<td>(98.256)</td>
<td>(0.095)</td>
<td>(0.014)</td>
<td>(1.000)</td>
<td>(0.092)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10 km × Spanish</td>
<td>-60.760</td>
<td>-0.286</td>
<td>-0.007</td>
<td>-0.224</td>
<td>-0.047</td>
<td>-0.002</td>
</tr>
<tr>
<td>Spanish</td>
<td>(36.795)</td>
<td>(0.179)</td>
<td>(0.009)</td>
<td>(0.855)</td>
<td>(0.111)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>β − β²</td>
<td>258.151</td>
<td>0.707</td>
<td>0.066</td>
<td>4.719</td>
<td>0.616</td>
<td>0.062</td>
</tr>
<tr>
<td>p-value β = β²</td>
<td>0.039</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.090</td>
<td>0.275</td>
<td>0.252</td>
<td>0.350</td>
<td>0.293</td>
<td>0.233</td>
</tr>
<tr>
<td>Observations</td>
<td>60262</td>
<td>60262</td>
<td>60262</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
</tr>
</tbody>
</table>

Note: The dependent variable in column (1) is the population density of the census area, winsorizing the top and bottom 1%. In column (2), the dependent variable is the inverse hyperbolic sine of population density. In column (3), it is a binary variable equals to one if the census area is classified as urban by IBGE. In column (4) and (5), the dependent variables are the level and the inverse hyperbolic sine of the luminosity measured for each pixel. In column (6), it is a dummy variable if luminosity is greater or equal than 32; it is a proxy for where cities are located, see main text. Control variables vary by Panel; see main text. Notice β is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and β² is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. Statistical significance denoted by: * 10%, ** 5%, *** 1%.
Table A.2: Robustness: different coefficients for control variables

<table>
<thead>
<tr>
<th></th>
<th>Pop. Density</th>
<th>IHS Pop. Density</th>
<th>Urban Area</th>
<th>Lights</th>
<th>IHS Lights</th>
<th>City (Lights ≥ 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: baseline controls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>161.494*</td>
<td>0.381***</td>
<td>0.050***</td>
<td>3.763***</td>
<td>0.491***</td>
<td>0.049**</td>
</tr>
<tr>
<td></td>
<td>(86.246)</td>
<td>(0.113)</td>
<td>(0.018)</td>
<td>(1.252)</td>
<td>(0.128)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>2.398</td>
<td>-0.210</td>
<td>0.003</td>
<td>-0.246</td>
<td>-0.108</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(18.543)</td>
<td>(0.176)</td>
<td>(0.008)</td>
<td>(0.722)</td>
<td>(0.100)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>(\beta - \beta^S)</td>
<td>159.096</td>
<td>0.591</td>
<td>0.047</td>
<td>4.008</td>
<td>0.599</td>
<td>0.047</td>
</tr>
<tr>
<td>p-value (\beta = \beta^S)</td>
<td>0.076</td>
<td>0.007</td>
<td>0.015</td>
<td>0.007</td>
<td>0.000</td>
<td>0.042</td>
</tr>
<tr>
<td><strong>Panel B: extended set of controls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>160.691*</td>
<td>0.374***</td>
<td>0.049***</td>
<td>3.950***</td>
<td>0.527***</td>
<td>0.051**</td>
</tr>
<tr>
<td></td>
<td>(84.929)</td>
<td>(0.103)</td>
<td>(0.017)</td>
<td>(1.198)</td>
<td>(0.119)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>-8.199</td>
<td>-0.273</td>
<td>-0.001</td>
<td>-0.343</td>
<td>-0.110</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(14.526)</td>
<td>(0.167)</td>
<td>(0.007)</td>
<td>(0.640)</td>
<td>(0.091)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>(\beta - \beta^S)</td>
<td>168.890</td>
<td>0.647</td>
<td>0.050</td>
<td>4.293</td>
<td>0.637</td>
<td>0.051</td>
</tr>
<tr>
<td>p-value (\beta = \beta^S)</td>
<td>0.053</td>
<td>0.002</td>
<td>0.006</td>
<td>0.003</td>
<td>0.000</td>
<td>0.021</td>
</tr>
<tr>
<td>Observations</td>
<td>60262</td>
<td>60262</td>
<td>60262</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
</tr>
</tbody>
</table>

Note: The dependent variable in column (1) is the population density of the census area, winsorizing the top and bottom 1%. In column (2), the dependent variable is the inverse hyperbolic sine of population density. In column (3), it is a binary variable equals to one if the census area is classified as urban by IBGE. In column (4) and (5), the dependent variables are the level and the inverse hyperbolic sine of the luminosity measured for each pixel. In column (6), it is a dummy variable if luminosity is greater or equal than 32; it is a proxy for where cities are located, see main text. Control variables include state fixed effects, 2nd degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline, and the interactions with Spanish settlement. In Panel B, control variables also include days of malaria transmission, the inverse hyperbolic sine of the potential yield of coffee, and the logarithms of the potential calories, as well as the interactions with Spanish settlement. All observations weighted by area. In the parentheses, robust standard errors clustered by micro-region. Notice \(\beta\) is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and \(\beta^S\) is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. Statistical significance denoted by: * 10%, ** 5%, *** 1%
Table A.3: Robustness: alternative samples

<table>
<thead>
<tr>
<th>Panel A: sample is 30km buffer around the Peabiru</th>
<th>Pop. Density</th>
<th>IHS Pop. Density</th>
<th>Urban Area</th>
<th>Lights</th>
<th>IHS Lights</th>
<th>City Lights (Lights ≥ 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>153.975***</td>
<td>0.394***</td>
<td>0.054***</td>
<td>3.818***</td>
<td>0.479***</td>
<td>0.050***</td>
</tr>
<tr>
<td></td>
<td>(70.352)</td>
<td>(0.102)</td>
<td>(0.016)</td>
<td>(1.042)</td>
<td>(0.102)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>27.339</td>
<td>-0.089</td>
<td>0.013</td>
<td>0.375</td>
<td>-0.005</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(29.603)</td>
<td>(0.186)</td>
<td>(0.010)</td>
<td>(0.850)</td>
<td>(0.111)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>β − βs</td>
<td>126.636</td>
<td>0.483</td>
<td>0.041</td>
<td>3.443</td>
<td>0.484</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>(29.603)</td>
<td>(0.186)</td>
<td>(0.010)</td>
<td>(0.850)</td>
<td>(0.111)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>p-value β = βs = 0</td>
<td>0.054</td>
<td>0.028</td>
<td>0.029</td>
<td>0.013</td>
<td>0.002</td>
<td>0.056</td>
</tr>
<tr>
<td>Observations</td>
<td>49757</td>
<td>49757</td>
<td>49757</td>
<td>238993</td>
<td>238993</td>
<td>238993</td>
</tr>
</tbody>
</table>

Panel B: excludes state capitals

| Dist. to Peabiru ≤ 10km × Portuguese | 105.879*** | 0.327*** | 0.041*** | 3.264*** | 0.465*** | 0.040*** |
| | (52.475) | (0.092) | (0.015) | (1.027) | (0.117) | (0.016) |
| Dist. to Peabiru ≤ 10km × Spanish | 12.815 | -0.166 | 0.009 | 0.003 | -0.067 | 0.005 |
| | (24.467) | (0.181) | (0.011) | (0.958) | (0.131) | (0.013) |
| βs = β − λ | 93.064 | 0.493 | 0.032 | 3.261 | 0.532 | 0.036 |
| p-value βs = 0 | 0.096 | 0.020 | 0.088 | 0.024 | 0.004 | 0.092 |
| Observations | 38914 | 38914 | 38914 | 325315 | 325315 | 325315 |

Panel C: only the state of Paraná

| Dist. to Peabiru ≤ 10km × Portuguese | 239.671*** | 0.811 | 0.099*** | 7.413*** | 0.875*** | 0.107*** |
| | (90.472) | (0.112) | (0.024) | (1.776) | (0.127) | (0.031) |
| Dist. to Peabiru ≤ 10km × Spanish | -0.089 | -0.216 | 0.003 | -0.218 | -0.095 | 0.001 |
| | (21.523) | (0.169) | (0.008) | (0.738) | (0.102) | (0.010) |
| β − βs | 239.760 | 1.027 | 0.097 | 7.631 | 0.969 | 0.105 |
| p-value β = βs | 0.064 | 0.029 | 0.017 | 0.001 | 0.004 | 0.017 |
| Observations | 11913 | 11913 | 11913 | 152109 | 152109 | 152109 |

Panel D: includes regions south of Iguazu river

| Dist. to Peabiru ≤ 10km × Portuguese | 156.811* | 0.382*** | 0.048*** | 3.544*** | 0.465*** | 0.046** |
| | (82.318) | (0.105) | (0.017) | (1.183) | (0.121) | (0.019) |
| Dist. to Peabiru ≤ 10km × Spanish | 15.723 | -0.166 | 0.009 | 0.009 | -0.066 | 0.005 |
| | (30.621) | (0.184) | (0.011) | (0.988) | (0.133) | (0.013) |
| β − βs | 141.087 | 0.548 | 0.039 | 3.535 | 0.531 | 0.041 |
| p-value β = βs | 0.090 | 0.013 | 0.053 | 0.026 | 0.005 | 0.089 |
| Observations | 61135 | 61135 | 61135 | 338819 | 338819 | 338819 |

Note: The dependent variable in column (1) is the population density of the census area, winsorizing the top and bottom 1%. In column (2), the dependent variable is the inverse hyperbolic sine of population density. In column (3), it is a binary variable equals to one if the census area is classified as urban by IBGE. In column (4) and (5), the dependent variables are the level and the inverse hyperbolic sine of the luminosity measured for each pixel. In column (6), it is a dummy variable if luminosity is greater or equal than 32; it is a proxy for where cities are located, see main text. Control variables include state fixed effects, 2nd degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline. All observations weighted by area. In the parentheses, robust standard errors clustered by micro-region. Notice β is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and βS is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. Statistical significance denoted by: * 10%, ** 5%, *** 1%
Table A.4: Robustness: alternative cutoffs

<table>
<thead>
<tr>
<th></th>
<th>Pop. Density (1)</th>
<th>IHS Pop. Density (2)</th>
<th>Urban Area (3)</th>
<th>Lights (4)</th>
<th>IHS Lights (5)</th>
<th>City (Lights ≥ 32) (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A:</strong> ≤ 5km</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 5km × Portuguese</td>
<td>172.228*</td>
<td>0.483***</td>
<td>0.054***</td>
<td>3.940***</td>
<td>0.483***</td>
<td>0.054**</td>
</tr>
<tr>
<td></td>
<td>(89.162)</td>
<td>(0.112)</td>
<td>(0.018)</td>
<td>(1.253)</td>
<td>(0.121)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 5km × Spanish</td>
<td>18.570</td>
<td>-0.325</td>
<td>0.012</td>
<td>0.357</td>
<td>-0.013</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(30.023)</td>
<td>(0.213)</td>
<td>(0.012)</td>
<td>(1.017)</td>
<td>(0.132)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>β − βS</td>
<td>153.657</td>
<td>0.808</td>
<td>0.042</td>
<td>3.583</td>
<td>0.496</td>
<td>0.047</td>
</tr>
<tr>
<td>p-value β = βS</td>
<td>0.080</td>
<td>0.001</td>
<td>0.051</td>
<td>0.031</td>
<td>0.007</td>
<td>0.066</td>
</tr>
<tr>
<td><strong>Panel B:</strong> ≤ 15km</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 15km × Portuguese</td>
<td>141.825*</td>
<td>0.373***</td>
<td>0.041**</td>
<td>3.166**</td>
<td>0.442***</td>
<td>0.039**</td>
</tr>
<tr>
<td></td>
<td>(81.831)</td>
<td>(0.107)</td>
<td>(0.016)</td>
<td>(1.194)</td>
<td>(0.126)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 15km × Spanish</td>
<td>3.785</td>
<td>-0.227</td>
<td>0.005</td>
<td>-0.166</td>
<td>-0.074</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(27.917)</td>
<td>(0.149)</td>
<td>(0.010)</td>
<td>(0.903)</td>
<td>(0.127)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>β − βS</td>
<td>138.040</td>
<td>0.601</td>
<td>0.036</td>
<td>3.332</td>
<td>0.516</td>
<td>0.037</td>
</tr>
<tr>
<td>p-value β = βS</td>
<td>0.109</td>
<td>0.002</td>
<td>0.055</td>
<td>0.031</td>
<td>0.006</td>
<td>0.106</td>
</tr>
<tr>
<td>Observations</td>
<td>60262</td>
<td>60262</td>
<td>60262</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
</tr>
</tbody>
</table>

Note: The dependent variable in column (1) is the population density of the census area, winsorizing the top and bottom 1%. In column (2), the dependent variable is the inverse hyperbolic sine of population density. In column (3), it is a binary variable equals to one if the census area is classified as urban by IBGE. In column (4) and (5), the dependent variables are the level and the inverse hyperbolic sine of the luminosity measured for each pixel. In column (6), it is a dummy variable if luminosity is greater or equal than 32; it is a proxy for where cities are located, see main text. Control variables include state fixed effects, 2nd degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline. All observations weighted by area. In the parentheses, robust standard errors clustered by micro-region. Notice β is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and βS is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. Statistical significance denoted by: * 10%, ** 5%, *** 1%
Table A.5: Robustness: measurement error

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Peabiru × Portuguese</td>
<td>146.845</td>
<td>1.410***</td>
<td>0.082***</td>
<td>6.181***</td>
<td>0.898***</td>
<td>0.081**</td>
</tr>
<tr>
<td></td>
<td>(90.138)</td>
<td>(0.292)</td>
<td>(0.029)</td>
<td>(2.260)</td>
<td>(0.239)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>Peabiru × Spain</td>
<td>-19.214</td>
<td>-0.527</td>
<td>-0.002</td>
<td>-1.163</td>
<td>-0.418</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(123.909)</td>
<td>(0.591)</td>
<td>(0.038)</td>
<td>(2.525)</td>
<td>(0.466)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>β − βS</td>
<td>166.059</td>
<td>1.937</td>
<td>0.083</td>
<td>7.344</td>
<td>1.317</td>
<td>0.087</td>
</tr>
<tr>
<td>p-value β = βS</td>
<td>0.314</td>
<td>0.009</td>
<td>0.105</td>
<td>0.050</td>
<td>0.020</td>
<td>0.115</td>
</tr>
<tr>
<td>Observations</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
<td>255</td>
</tr>
</tbody>
</table>

Note: The unit of observation is a 20 minutes × 20 minutes grid cell. Grid cells only contain areas that are within 50km of the Peabiru, so their shapes and areas will not be the same. The dependent variable in column (1) is the population density of the grid cell. In column (2), the dependent variable is the logarithm of population density. In column (3), it is the share of the grid cell that is classified as urban by IBGE. In column (4) and (5), the dependent variable are the level and the logarithm of the average luminosity measured for each pixel. In column (6), it is the share of the grid cell with luminosity greater or equal than 32. Control variables include state fixed effects, 2nd degree polynomials for altitude and ruggedness, and the share of the grid cell that is closer than 10 km to a major river or to the shoreline. All observations weighted by area. In the parentheses, robust standard errors clustered according to a 1 degree × 1 degree grid. Notice β is the coefficient on the share within 10km to the Peabiru interacted with the share in the Portuguese region, and βS is the coefficient on the share within 10km to the Peabiru interacted with the share in Spanish region. Statistical significance denoted by: * 10%, ** 5%, *** 1%
Table A.6: Robustness: inference

<table>
<thead>
<tr>
<th></th>
<th>Pop. Density</th>
<th>IHS Pop. Density</th>
<th>Urban Area</th>
<th>Lights</th>
<th>IHS Lights</th>
<th>City (Lights ≥ 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Conley standard errors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>163.096*</td>
<td>0.387***</td>
<td>0.051**</td>
<td>3.761***</td>
<td>0.493***</td>
<td>0.049**</td>
</tr>
<tr>
<td></td>
<td>(84.148)</td>
<td>(0.133)</td>
<td>(0.021)</td>
<td>(0.717)</td>
<td>(0.073)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>16.037</td>
<td>-0.164</td>
<td>0.009*</td>
<td>0.011*</td>
<td>-0.066</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(16.314)</td>
<td>(0.096)</td>
<td>(0.005)</td>
<td>(0.427)</td>
<td>(0.069)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>(\beta - \beta^S)</td>
<td>147.059</td>
<td>0.551</td>
<td>0.042</td>
<td>3.751</td>
<td>0.559</td>
<td>0.044</td>
</tr>
<tr>
<td>p-value (\beta = \beta^S)</td>
<td>0.080</td>
<td>0.003</td>
<td>0.056</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>Panel B: 1×1 degree clusters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Portuguese</td>
<td>163.096*</td>
<td>0.387***</td>
<td>0.051***</td>
<td>3.761***</td>
<td>0.493***</td>
<td>0.049***</td>
</tr>
<tr>
<td></td>
<td>(90.857)</td>
<td>(0.103)</td>
<td>(0.017)</td>
<td>(1.075)</td>
<td>(0.124)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Dist. to Peabiru ≤ 10km × Spanish</td>
<td>16.037</td>
<td>-0.164</td>
<td>0.009</td>
<td>0.011</td>
<td>-0.066</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(26.294)</td>
<td>(0.148)</td>
<td>(0.008)</td>
<td>(0.706)</td>
<td>(0.097)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>(\beta - \beta^S)</td>
<td>147.059</td>
<td>0.551</td>
<td>0.042</td>
<td>3.751</td>
<td>0.559</td>
<td>0.044</td>
</tr>
<tr>
<td>p-value (\beta = \beta^S)</td>
<td>0.105</td>
<td>0.004</td>
<td>0.029</td>
<td>0.006</td>
<td>0.001</td>
<td>0.032</td>
</tr>
<tr>
<td>Observations</td>
<td>60262</td>
<td>60262</td>
<td>60262</td>
<td>327795</td>
<td>327795</td>
<td>327795</td>
</tr>
</tbody>
</table>

Note: The dependent variable in column (1) is the population density of the census area, winsorizing the top and bottom 1%. In column (2), the dependent variable is the inverse hyperbolic sine of population density. In column (3), it is a binary variable equals to one if the census area is classified as urban by IBGE. In column (4) and (5), the dependent variable are the level and the inverse hyperbolic sine of the luminosity measured for each pixel. In column (6), it is a dummy variable if luminosity is greater or equal than 32; it is a proxy for where cities are located, see main text. Control variables include state fixed effects, 2nd degree polynomials for altitude and ruggedness, and binary variables equal to one if closer than 10 km to a major river or to the shoreline. All observations weighted by area. In the parentheses, standard errors. In Panel A, standard errors as in Conley (1999), with a cutoff of 200km and an uniform kernel. In Panel B, robust standard errors with observations clustered in 1x1 degree grid cells. Notice \(\beta\) is the coefficient on the indicator of proximity to the Peabiru interacted with the Portuguese region, and \(\beta^S\) is the coefficient on proximity to the Peabiru interacted with the Spanish settlement dummy. Statistical significance denoted by: * 10\%, ** 5\%, *** 1\%
D Data Appendix

In this Appendix, I describe auxiliary data sources.

**Digital elevation model.** The altitude and ruggedness variable is created using as source the ALOS Global Digital Surface Model, a 1 arcsec × 1 arcsec raster released by the Japan Aerospace Exploration Agency (JAXA). For each census area or grid cell, altitude is calculated as the mean in that area. The terrain ruggedness index (TRI) is the square root of the average squared differences in altitude between each pixel and its eight surrounding pixels. The mean of this index is taken over the census area or grid cell.

**Shoreline.** The shoreline information is taken from the World Vector Shorelines, available from the Global Self-consistent Hierarchical High-resolution Geography database (GSHHG).

**Main rivers.** The location of the main rivers is from the Brazilian Agência Nacional das Águas (ANA). I use version 1.3 of the Base Hidrográfica Ottocodificada. To identify the main rivers, I use the ANA map of the Paraná river basin and select the main rivers as all those highlighted in the map. In the sample area, these rivers are the Paraná, Iguazu, Negro, Ivaí, Paranapanema, Itararé, Tibaji, Itapetininga, Pardo, do Peixe, Tietê and Sorocaba.

**Malaria.** Malaria transmission suitability is from the Malaria Atlas. I use the mean of the days per year that could support Plasmodium falciparum transmission and the days per year that could support P. vivax transmission. Each of these is a 1 arcsec × 1 arcsec raster.

**Agricultural suitability.** Coffee potential yields and the caloric potential are constructed from the Global Agro-ecological Zones (GAEZ) dataset from FAO. In all cases, I use the potential yield under rain-fed, low-input agriculture, so it reflects agricultural suitability with the traditional technology. The caloric potential of each crop follows Galor and Özak (2016).

**Main roads.** The data is from the Base Cartográfica do Brasil ao Milionésimo, version 3, released in 2010 by IBGE. Main roads were the paved and built federal or state roads.