World Development 161 (2023) 106086

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Disease and diversity in long-term economic development ☆

Javier A. Birchenall [a,*]

[a] Department of Economics, 2127 North Hall, University of California at Santa Barbara, CA 93106, United States

ARTICLE INFO

Article history:
Accepted 22 August 2022
Available online 16 September 2022

JEL classification:
J11
O11
P51

Keywords:
Communicable diseases
Economic development
Ethnic diversity

ABSTRACT

Ethnographic data and archeological censuses of cities suggest that sub-Saharan Africa lagged behind tropical America during pre-colonial times. Disease (i.e., environmentally determined pathogen stress) has a negative impact on pre-colonial economic conditions, as measured by the presence of large physical structures in ethnographic data. This negative relationship is seen primarily, but not exclusively in African societies. Using a simple coalitional game, I propose a causal path from disease to ethnic diversity. Ethnographic data suggests a positive effect of disease on ethnic diversity, and persistent effects on long-term economic development. Even today, pre-colonial factors influence income per capita and ethnolinguistic fractionalization.

1. Introduction

Africa's transition into modern economic growth has been a relatively slow and gradual process due to a multiplicity of factors. For the economic literature on African development, modern influences such as the legacies of European colonialism and the slave trade have been a focal point (Acemoglu et al., 2001; Acemoglu et al., 2002; Acemoglu et al., 2003; Austin, 2008; Bates et al., 2007; Glaeser et al., 2004; Nunn, 2008; Nunn and Wantchekon, 2011). There are also lines of continuity that emphasize pre-colonial conditions through political economy distortions, greater ethnic diversity, and the suboptimal provision of public goods (Bockstette et al., 2002; Cervellati et al., 2017; Cervellati et al., 2017; Depetris-Chauvin and Weil, 2018; Englebert, 2000; Fenske, 2014; Gennaioli and Rainer, 2007; Herbst, 2000; Michalopoulos, 2012; Michalopoulos and Papaioannou, 2013; Michalopoulos and Papaioannou, 2013). This paper is an empirical study of comparative development. It argues that sub-Saharan Africa lagged comparable areas before the European expansion, and that Africa's adverse disease environments have limited economic development since pre-colonial times.

The modern influences set off by the European expansion in the 1500s operated in regions with significant pre-existing differences. The absence of well-established data, however, is a critical limitation in studying the pre-colonial era. Pre-colonial data is difficult to find, and existing sources have a limited scope and quantitative focus. In this paper, I rely on Murdock and White's (1969) Standard Cross Cultural Sample (SCCS). I also consider alternative data sources, such as Chandler's, 1987 inventory of medium and large cities, and demographic data from McEvedy and Jones (1985). I focus on the SCCS because it consists of disaggregate data from 186 ethnographically well-described societies with different subsistence strategies. The SCCS was designed to be representative of all the pre-industrial societies in the world, and it was constructed to maximize independence in terms of cultural and historical origin. Societies are described from historic and ethnographic literature at the time coinciding with or just after contact with Western cultures when some reliable observer (i.e., traveler, missionary, trader, colonial agent, or anthropologist) first visited the society and wrote about it in sufficient detail (Pryor, 2005). The SCCS represents the earliest systematic "snapshot" of former European colonies in tropical Africa and the New World.

The SCCS measures various economic aspects of pre-modern societies. The SCCS records information about the presence and scope of *large or impressive structures*. A physical structure is easy to recognize by Western observers, hence it is an aspect of

☆ I would like to thank the Editor and two anonymous referees for many valuable suggestions. Philip Babcock, Sambit Bhattacharyya, Hoyt Bleakley, Bob Fogel, David Carr, Juan Carlos Cordoba, Dave Donaldson, Stanley Engerman, James Fenske, Nicola Gennaioli, Ana Maria Ibáñez, Hyeok Jeong, Michael Kevane, Gary Libecap, Ted Miguel, Kris Mitchener, Nathan Nunn, Pekka Hämäläinen, Phil Hoffman, Jeff Nugent, Bonnie Queen, James Robinson, Jean-Laurent Rosenthal, Raaj Sah, Romain Wacziarg, Warren Whatley, and seminar participants at USC, Universidad de los Andes (Bogotá), the All-UC Economic History group meeting in 2008, PACDEV in 2009, and the Econometric society for very useful comments. I would like to specially acknowledge the many suggestions by Peter Lindert, my discussant at the All-UC EH meeting.

* Corresponding author.

https://doi.org/10.1016/j.worlddev.2022.106086
0305-750X/

pre-colonial societies that can be measured with relative precision. Interpreting physical evidence is perhaps less arbitrary that interpreting, for example, measures of political centralization available in the SCCS or in their predecessors, i.e., Murdock's (1967) *Ethnographic Atlas*. (Agricultural societies in sub-Saharan Africa, for instance, are on average ranked as high as Eurasian societies in political centralization in the SCCS.[1]) Historically, impressive structures are a manifestation of a society's capacity to extract resources from peasants. Large structures are a proxy for tributary empires and economic conditions. As put by Barras (2016, p. 4–5), "buildings fulfil a complex of economic, social, cultural, and aesthetic functions [and] are the most durable of human artifacts."

In the SCCS, the presence and scope of large structures is significantly different between African societies and the rest of the world. Differences are evident even with respect to tropical America, a "control" area comparable to sub-Saharan Africa in many dimensions. Population densities were comparable (McEvedy and Jones (1985)), but Chandler's, 1987 inventory of cities suggests fewer and less densely populated cities in sub-Saharan Africa than in tropical America. While there were no large cities south of the Sahara around the time of the European expansion, Teothihuacán, currently Mexico city, was among the ten largest cities of the world in 400 AD (Chandler, 1987, p. 464). Westerners' direct accounts of large or impressive structures and archeological inventories of cities both tentatively suggest that societies in pre-colonial Africa were less developed than European and Asian societies.[2] Even in a more controlled comparative perspective, the previous accounts suggest lower economic sophistication relative to the tributary empires of tropical America.

As economic differences within the tropics potentially pre-date the European expansion, it is worth asking what factors may be able to explain differential comparative development in the past? Pathogen stress, a measure of the prevalence and severity of serious pathogenic diseases in the SCCS, is strongly and negatively related with the existence and scope of large or impressive structures. This relationship is driven primarily, but not exclusively, by African societies. For environmental reasons, sub-Saharan Africa has more adverse disease environments than societies in tropical areas elsewhere (Crosby, 2003; Curtin et al., 1995; Diamond and Panosian, 2006; McNeill, 1976; Wolfe et al., 2007). The association between disease and pre-colonial development is potentially causal, as tropical diseases are generally environmentally determined. Reverse causality is not an important concern because the control of endemic diseases was limited in the past. The pathogens considered here are transmitted through vectors environmentally determined, they are not by-products of changes in food production, and they have a regional distribution concentrated in tropical Africa. Even as a correlative relationship, the association between disease and pre-colonial development is robust to the inclusion of multiple control variables.

Many channels have been proposed to explain the influence of disease on economic outcomes. Malthusian views emphasize the impact of disease on population density, which is important for conflict-based theories in political science (Herbst, 2000) and for economic specialization (Allen, 2011). I control for differences in population density, but the considerable similarities in population density, technology, and geography between sub-Saharan Africa and tropical regions in the New World limit a Malthusian interpretation of economic differences within the tropics.[3] Disease has also been linked to economic conditions through its direct effect on labor productivity (Bhattacharyya, 2009; Bloom and Sachs, 1998; Kamarck, 1967). Disease, however, rarely influences individuals in isolation. Throughout history, endemic and epidemic diseases have typically acted on large groups of individuals.

The fundamental social mechanism proposed in this paper views reproductive, social, and cultural separation between human populations as a societal response to counter spillovers in high disease environments. The idea, formulated mathematically as the Nash equilibrium of a simple coalitional game with disease spillovers, is that the social structure is more fragmented in disease-prone areas. As Kurzban and Leary (Kurzban and Leary, 2001, p. 192) note, parasites put the "brakes" on sociality (Azevado, 1980; Cashdan, 2001; Fincher and Thornhill, 2008). I also test this causal path in the SCCS. African societies are more ethnically diverse than anywhere else in the world. In the SCCS, pathogen stress has a positive statistical influence on ethnic diversity, robust to many controls. Pathogen stress and ethnic diversity are also strongly associated with current income per capita and current indices of ethnolinguistic fractionalization. Overall, these findings suggest that ethnic fragmentation is partly a by-product of Africa's distinctive disease environment, which has negatively influenced the continent's pre-colonial and current economic conditions.

Some related literature. That tropical diseases represent major obstacles to economic development is a familiar theme to both historians and development economists. The importance of disease in long-term outcomes resounds with McNeill's (1976) historical thesis. McNeill (1976) argued that disease limited the consolidation of pre-modern states and empires contributing, for example, to the collapse of the Roman Empire and China's ancient Han Dynasty, and to the eventual European control of the New World (Price-Smith, 2008; Rosen, 2007; Zinsser, 1935). Regarding Africa, McNeill (1976, p. 43) asserts "[high exposure to disease], more than anything else, is why Africa remained backward in the development of civilization when compared to temperate lands (or tropical zones like those in the Americas), where prevailing ecosystems where less elaborated and correspondingly less inimical to simplification by human action." Africa's high disease burden has been particularly unfavorable since colonial times given the "ill-conceived efforts by European colonial administrators" to consolidate power and expand the agricultural frontier. These efforts had the unintended consequence of precipitating epidemic outbreaks of sleeping sickness and malaria during the nineteenth and twentieth centuries (McNeill, 1976, p. 42; Azevado, 1980).

The proposed link between disease and diversity complements a large literature that examines the causes behind Africa's high ethnolinguistic fractionalization. It is not possible to provide a detailed account of this literature, but a few notable examples are important to mention. Michalopoulos (2012) identified the influence of *physical barriers* (i.e., barriers geographically determined) on current ethnolinguistic fractionalization and Michalopoulos and Papaioannou (2013) studied the effect *political barriers* (i.e., artificial country boundaries) on current economic outcomes and civil conflict in Africa. This paper is consistent with

[1] Gennaioli and Rainer (2007), Heldring and Robinson (2012), Michalopoulos and Papaioannou (2013), and Osafo-Kwaako and Robinson (2013) are among the many studies that examine political centralization in the previous ethnographic surveys. Their goal is to study pre-colonial conditions *within* Africa and not comparative aspects *between* former European colonies in the tropics.

[2] This finding aligns with existing discussions by African economic historians, who have described pre-colonial state structures in Africa as weak (Connah, 2001; Curtin et al., 1995; Englebert, 2000; Fortes and Evans-Pritchard, 1940; Herbst, 2000; Hopkins, 1973). Some of the existing assessments see pre-colonial conditions in Africa through a European lens. My focus is on economic conditions in tropical areas, which are perhaps more comparable in some dimensions discussed below.

[3] Economists have almost exclusively associated pre-modern differences with technological disparities, often using a Malthusian logic (Ashraf and Galor, 2011; Comin et al., 2010). The Malthusian logic works well to account for economic and demographic differences between agricultural areas (i.e., the Near East and China) and non-agricultural ones (i.e., North America and Australia), but not for the comparative development of tropical areas with similar demography, technology, and geography.

the notion that disease acts as a *social barrier*. Social barriers that induce ethnic fragmentation are endogenous, and not easily amenable to causal identification using experimental or quasi-experimental methods. The enactment of distance-based rules, however, was a common social response to counter contagion in the past (Birchenall, 2014).

An advantage of focusing on pre-colonial times is that some plausible mechanisms linking disease and economic development can be discarded. An empirical association between pre-colonial conditions and modern outcomes could arise because of factors that pre-date the European arrival or because of factors that emerged in response to European influences.[4] An indirect influence of disease on economic conditions runs through the effect of settler mortality on colonial institutions and human capital (Acemoglu et al., 2001; Glaeser et al., 2004). It is not possible to decompress the influence of disease on economic development, but the importance of pathogen stress during pre-colonial times suggests that disease does not operate exclusively through European settlements, as argued by the *colonial origin hypothesis* in Acemoglu et al. (2001). In fact, pathogen stress is a better predictor of current institutions than settler mortality.

Outline. The rest of the paper proceed as follows. Section 2 introduces the main data sources and baseline estimates. Section 3 proposes a potential mechanism that links disease and ethnic diversity in a coalitional game where social integration carries the risk of disease contagion. I also consider the empirical validity of the proposed mechanism in the SCCS. Section 4 brings past influences to bear on current economic conditions and ethnic fragmentation, including a brief analysis of the relationship between pathogen stress, settler mortality, and political institutions. Section 5 presents some brief concluding remarks.

2. Data and baseline measurement

Societal data. The Standard Cross-Cultural Sample (SCCS) was developed by Murdock and White (1969) during the 1960s. The SCCS contains 186 pre-industrial societies with various subsistence strategies, including hunter-gatherers, fishers, pastoralists, horticulturalists, and agriculturalists. The SCCS provides extensive coded data constructed from historical records and published field research by ethnographers. Information about these societies is obtained through narratives and descriptions of Westerners in their first encounters with these societies. There is no "pristine" society in the SCCS; many of the native societies in Africa were not reached by Europeans until late in the nineteenth century or early in the twentieth century; see Table 1 below. The key, as pointed out by Pryor (2005, pp. 24–25), is to "ask how much these preindustrial societies have changed over the millennia until the pinpointed date." I assume that contamination in the SCCS is not a serious problem and that potential biases are corrected by controls for the date of pinpointing of the society. Further, the SCCS is not a random sample of all pre-industrial societies and, by construction, the SCCS is prone to measurement error.[5]

The SCCS has wide geographic coverage and it is the earliest and most complete systematic assessment of pre-industrial societies in the world (Fig. 1). The geographic composition of the 186 societies is: 32 societies are in sub-Saharan Africa, 35 in South and Central America, 24 in West Eurasia, 34 in East Eurasia, 31 in the insular Pacific and 30 in North America.[6] In the sample, there are 122 agricultural societies distributed relatively equally among the major regions of the world. The remaining societies are proto- or pre-agricultural societies such as the early agricultural societies of North America and hunter-gatherers (i.e., societies in which the contribution of agriculture to the local food supply is less than 10 percent according to the coded variable §v3 in the SCCS). I will add a control for pre- and proto-agricultural societies because their economic and political organizations are very different from the organization of agricultural societies; excluding these societies leads to virtually identical results.

Table 1 summarizes the main variables used here. I first focus on the degree of complexity in the buildings and structures in a society. I use the coded variable for *large or impressive structures*, which is §v66 in the SCCS. (I present complementary evidence of alternative proxies of pre-modern sophistication later on, and in a subsequent section I examine current income per capita.) Large and impressive structures have been historically associated with tributary empires, urban life, and surplus extraction. As Curtin et al. (1995, p. 72) note, "only a state with its permanent officials and central direction could easily mobilize the wealth of society for special purposes such as temple building or the support of a priesthood, a nobility, or a royal lineage." Large or impressive structures can be measured retrospectively with some accuracy. Subjective assessments also appear less important than in other economic and political measures in the SCCS. Large or impressive structures is organized in an ordinal scale with higher values representing higher complexity, as typically done in the SCCS. For agricultural societies, this variable is coded as follows, §v66: 1 = None (47 cases), 2 = Residences of influential individuals (18), 3 = Secular or public buildings (26), 4 = Religious or ceremonial buildings (24), 5 = Military structures (3), 6 = Economic or industrial buildings (4 cases). The ordering in the SCCS might not represent higher sophistication so I will also examine a dichotomous variable for the presence of religious, military, or industrial structures only, or for cases in which §v66 $>$ 3.

The SCCS provides a measure of political centralization in the form of jurisdictional hierarchy beyond the local community. This measure is coded as follows, §v237: 1 = No political authority beyond community, 2 = Petty chiefdoms, 3 = Larger chiefdoms, 4 = States, and 5 = Large states. (Several empirical studies listed later on have recently made use of this measure.) I will also use population density (§v64, defined as number of people per square mile) and community size (§v63). Both variables are ordinal scales from 1 to 7 but they essentially represent the logarithm of density and size, respectively. Controls for geography include distance from the Equator and log-altitude (§v183), and a measure of the agricultural potential in the region where the societies are located. Potential for agriculture is an index based on land slope, soil quality, and climate (§v921). I also use an explicit measure of food surplus in the society (§v21). Technological sophistication is a measure associated with the existence of writing and records (§v149), the fixity of residence (§v150), technological specializa-

[4] Ignoring the many changes between the pre-colonial era and the present leads to what Austen (1987) called a 'compression of history.' Considering influences that pre-date the European arrival serves to discriminate between competing aspects behind the comparative development of former European colonies. The use of pre-colonial intermediaries and institutions to exercise colonial control under *indirect rule* requires pre-existing political control. The presence of *extractive institutions* motivated by differences in disease environments or endowments in an example of emerging influences. Both views are prominent in the literature (Acemoglu et al., 2001; Bockstette et al., 2002; Heldring and Robinson, 2012).

[5] The societies in the SCCS are a culturally independent sub-sample of the *Ethnographic Atlas*, also consolidated by Murdock and White (1969). The Atlas contains many more societies than the SCCS but there is less information for each society, i.e., there is no measure of physical structures in the Atlas, which is one of the central outcomes of interest in this paper.

[6] The geographical distribution of societies differs slightly from the one in the SCCS (§v200) because I treat South and Central America as a unit. These regions and the West Indies form a single Neotropical region. (North America is part of the Nearctic zone.) Communication between South and Central America was far more common than between Central and North America; see Diamond (1997). I also treat African societies in the Sahel as part of sub-Saharan Africa whereas, in the SCCS, some are part of Eurasia. Obviously, later on I perform sensitivity analyses to examine if changes in the grouping affect the findings.

Table 1
Descriptive statistics for agricultural societies in the SCCS.

			Region		
	Overall mean	Std. Dev.	Africa	America	Eurasia
	A. Large or impressive structures (§v66)				
Large or impressive structures	2.42	1.40	1.50$^{(a,e)}$	2.03	2.91
	B. Centralization and demography				
Political centralization	2.45	1.21	2.57$^{(a)}$	1.60	2.71
Population density	4.65	1.69	4.80$^{(a)}$	3.24	5.10
Community size	3.92	1.63	3.92	3.50	4.08
	C. Geography and technology				
Distance from equator	16.64	12.48	9.15$^{(a,e)}$	13.30	20.67
Log-altitude	4.28	2.67	5.74 $^{(a,e)}$	4.65	3.63
Agricultural potential	17.70	2.52	18.42$^{(e)}$	18.30	17.21
Food surplus and storage	1.90	0.73	1.69$^{(e)}$	1.76	2.02
Technological sophistication	22.66	6.46	21.88$^{(a,e)}$	18.34	24.55
	D. Disease environments and ethnic diversity				
Pathogen stress	14.00	3.02	17.46 $^{(a,e)}$	13.73	12.82
Ethnic groups within 250 miles	7.34	9.51	18.73$^{(a,e)}$	3.76	4.31
Ethnic groups within 500 miles	23.09	26.19	59.19$^{(a,e)}$	11.34	13.64
	E. Date of pinpointing of the society (§v838)				
Pinpointing date	1876	219.05	1915$^{(a,e)}$	1868	1866

Note.– The sample is based on 122 agricultural societies in the SCCS. Means for societies in (sub-Saharan) Africa, (South and Central) America, and Eurasia, respectively. Many of these variables are ordinal in scale. $^{(a)}$ and $^{(e)}$ denote significant one-sided difference (at 10 percent level) between Africa and America, and Africa and Eurasia, respectively.

Fig. 1. Geographic distribution of the SCCS societies in the world.

tion (i.e., presence of pottery, metal work, and loom weaving, §v153), land transport (§v154), monetary exchange (§v155), and social stratification (§v158). I add these measures to construct an overall index.[7]

To measure the extent of disease prevalence during the pre-colonial era, I rely on a general measure of pathogen stress (§v1260). Pathogen stress is based on medical and public health sources on the latitude and longitude of the sample societies. Pathogen measures use data as close as possible to the defined dates for the sample societies' SCCS data (Low, 1990).[8] Seven pathogens (leishmanias, trypanosomes, malaria, schistosomes, filariae, spirochetes, and leprosy) are recorded in §v1253-1259 and are rated on a 3-point scale for frequency and severity. The individual scores are added to yield a total pathogen stress score. A high score represents many types of pathogens and more severe exposure. I will later on consider just the presence (regardless of the severity) of the pathogen as a robustness check. In principle, an analysis of individual pathogens is possible, but there is no reason to predict special associations with particular pathogens. Pathogens and their severity

[7] This practice is standard. Comin et al. (2010), for example, added the presence of different technologies to construct an overall index of pre-modern technological sophistication. Africa seems to have experienced an independent origin of iron work often cited as being part of the advancements spread with the Bantu expansions. As Austen, 1987, p. 14 note, however, "iron apparently made no dramatic impact upon early African agriculture." Cattle domestication also seems to have had an independent origin; see Austen (1987, chapter 1). Important independent achievements in writing, mathematics, and science also took place in the New World (Mann, 2005, pp. 16–20 and pp. 63–65)).

[8] Low (1990) studied disease and marriage systems. Thus there is no a priori sense of bias in the disease measures, which is an advantage of the indicators of disease in the SCCS.

also are spatially concentrated. An analysis based on a single disease would fail to recognize these complementarities.

Measures of ethnic diversity were compiled by Cashdan (2001), who calculated the number of ethnic groups present within a given radius (100–500 miles in 50-mile increments) of each SCCS society out of a "universe" of ethnographic societies in the world. My baseline analysis considers measures based on a 500 mile radius (§v1872). I also present results with a 250 miles radius (§v1867).

Baseline measurement. The summary characteristics of the agricultural societies in the SCCS in Table 1 suggest the following significant differences:

(i) Large or impressive structures were less prevalent in Africa compared to Eurasia or tropical America. The difference in the SCCS is likely accurate. One cannot nowadays point to a large sub-Saharan city in pre-modern times or to evidence of large public works typical of tributary empires. Pre-colonial cities and empires did exist in Africa long before Muslim and European contact (Connah, 2001; Coquery-Vidrovitch, 2011; Davidson, 1959; Hull, 1976),[9] (below)). Cities and centralized states, however, appear to be smaller in scale in sub-Saharan Africa than in any other agricultural society, especially given the technological sophistication of societies in Africa and Africa's repeated contact with the rest of the Old World.

(ii) Measured political centralization in sub-Saharan Africa was on average as high as in Eurasia, and higher than in tropical America. Measures of political centralization from ethnographic data have been used to study pre-colonial conditions in sub-Saharan Africa (Fenske, 2014; Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Osafo-Kwaako and Robinson, 2013). Previous studies have shown that political centralization is correlated with public goods provision and economic features such as the use of money among societies within Africa. It is unclear, however, if political centralization are based on a *comparative* scale that includes non-African societies. Measures of centralization in the SCCS might be relative and hence may not capture absolute differences across regions. With large buildings and structures, a later subsection uses (independent) archeological data to validate the ethnographic data.

(iii) In terms of demography, technology, and geography, tropical America is closer to sub-Saharan Africa than Eurasia. There is no statistical difference within the tropics in terms of agricultural potential and food surpluses. There is also no difference in community sizes between regions. Population density and technological sophistication are higher in sub-Saharan Africa compared to tropical America (although equal to Eurasia's population density). The similarities in these aspects are expected because agriculture originated independently in these tropical areas at similar times, around 4,000 BC and 4,500 BC (Smith, 1995). In addition, both areas had similar numbers of potentially domesticable animals, and large-seeded grass species suitable for agriculture (Diamond, 1997, Tables 8.1 and 9.2). In both, axes run mostly from North to South (Diamond, 1997, p. 177), and their areas are relatively the same (McEvedy and Jones, 1985 and Table 5 (below)). Their climates are also similar. Africa and tropical America are the only continents that cross the equator.

(iv) Pathogen stress is higher in sub-Saharan Africa relative to agricultural societies in any other part of the world. Tropical parasitic and infectious diseases have been heavily concentrated in sub-Saharan Africa. As noted by Diamond and Panosian (2006, pp. 32–33), for example, "there is [also] an asymmetry for the major infectious diseases of the tropics: they too arose overwhelmingly in the Old World, rather than in the New World." In their discussion of the geographic origins of tropical diseases, they note that:

> "[E]ven the most significant diseases which originated in the New World tropics, Chagas' disease and leishmaniasis (the latter also arose in the Old World tropics), have much less human impact than any of the six leading diseases of the Old World tropics (yellow fever, falciparum malaria, vivax malaria, cholera, dengue fever, and East African sleeping sickness). From Columbus's voyage of AD 1492 until the 1640s, European explorers and settlers in the New World suffered few deaths from indigenous disease, compared to the sizeable mortality that Europeans suffered in the Old World tropics and Native Americans suffered from introduced Eurasian diseases. For example, after Pizarro arrived in the tropics of Peru with an army of 168 Spaniards, not more than one or two of them died of disease over the next two years.".

Pathogen stress is plausibly *exogenous*. The pathogens studied in the SCCS are primarily determined by environmental conditions and not by human action. Pathogens in the SCCS are unrelated to economic conditions (i.e., sanitation, childhood immunization, etc.) and this implies that reverse causality is not an important concern. Further, the SCCS pathogens are typically transmitted through vectors environmentally determined and not through an oral-fecal route which is strongly affected by urbanization and public health interventions. They rarely provoke immune reactions and these diseases are not by-products of changes in food production, as many of the "crowd diseases" in temperate areas (Wolfe et al., 2007). The African origin of these pathogens, and the reasons for their geographic concentration, provide some support to the exogeneity of disease.[10]

Finally, (v) ethnic diversity is also higher in sub-Saharan Africa compared to agricultural societies in any other part of the world. The measures of ethnic diversity in the SCCS are based on physical distance between societies. This measurement procedure differs from indices of ethnolinguistic fractionalization, which measure the probability that two randomly drawn people within a country will be from different ethnic groups. As societies are the political unit of the SCCS, it is not possible, and perhaps not adequate, to

[9] As Coquery-Vidrovitch (2011, p. 98) notes, cities in the area of Muslim influence "were more or less Islamized throughout history, but they were certainly not historical 'Muslim cities,' except perhaps Timbuktu." Ancient Egypt apparently did not have a strong influence on sub-Saharan Africa although cities appeared in the middle Nile earlier than in other areas. Meroë is the best known case. In 430 B.C., Meroë had about 20,000 inhabitants (Chandler, 1987, p. 461). Gold, ivory, slaves, and other mineral, animal and vegetable products were traded with Eurasia through the Nubian corridor that connected tropical Africa with Egypt. Some have suggested that this corridor was a cultural *cul-de-sac*, but such description might be inaccurate; see Connah (2001, p. 19). There were also many ancient Swahili coastal trading states in the eastern coast of Africa such as Kilwa; see Coquery-Vidrovitch (2011, Chapter 2).

[10] Far more tropical diseases originated and are still found in Africa partly because humans and Old World monkeys and apes are genetically closer and this implies higher disease transfers; (Diamond and Panosian, 2006; Wolfe et al., 2007). As McNeill (1976, p. 15) notes, "the array of parasites that infest wild primate populations is known to be formidable." Monkeys and apes in the Old World also serve as *reservoirs* and origin for many of the human diseases in Africa. (HIV/AIDS being a particular illustration in modern times.) This is not the case for New World monkeys. The absence of large mammals in the New World also meant reduced opportunities for disease transfers and therefore healthier environments (Wolfe et al., 2007). The unequal distribution of pathogens between Africa and tropical America is evident in the unbalanced nature of disease exchange after the European expansion (Crosby, 2003; Diamond, 1997; Hoeppli, 1969). The origin of diseases such as malaria in tropical Africa must be so early in human existence that some populations in Africa have developed adaptations at the genetic level. It is also relevant to clarify a common misconception about tropical disease in the Americas. As Diamond and Panosian (2006, pp. 33–34) note, "[m]any people initially associate the New World tropics with heavy mortality from infectious diseases when recalling the role of infection in the failed efforts by the French to construct a Panama Canal in 1881–88, and the similar obstacle infectious diseases posed when the United States subsequently undertook the building of the Panama Canal in the early 1900s (McCullough (1977)). The key to American success at the Canal was the defeat of the disease problem. But the diseases that caused so much mortality in Panama were yellow fever and malaria introduced from Africa."

construct identical measures as those used for countries. As community sizes are relatively equal in the SCCS, measures based on physical distance are likely to align well with those based on political units. Ethnic diversity in the SCCS is strongly positively correlated with measures of ethnolinguistic fractionalization in Easterly and Levine (1997). Further, as discussed in Cashdan (2001), SCCS measures of ethnic diversity are highly correlated with alternative estimates of linguistic diversity.

Reduced-form effects of disease. If pathogen stress is environmentally determined, the basic empirical strategy is straightforward. I first estimate regression equations of the following form:

$$Y_i = \delta \cdot \text{Pathogen}_i + X_i \cdot \phi + \varepsilon_i, \quad (1)$$

where Y_i measures the complexity of large or impressive structures in society i, Pathogen_i is the measure of pathogen stress in society i, and X_i are a series of controls that capture aspects that are potentially important for the presence and scope of large structures. The coefficient δ measures the reduced-form effect of pathogen stress conditional on X_i.

Table 2 reports the reduced-form estimates. Column (1) includes pathogen stress, not including any covariates. The point estimate is −0.102 (standard error 0.026) so that large or impressive structures were less complex in agricultural societies affected by heavy pathogen stress. Subsequent columns include control variables. It is not possible to consider only controls not themselves caused by pathogen stress. Some controls represent omitted aspects that may explain the influenced by disease on Y_i or may be caused by it. For example, as only large societies can devote resources to large or impressive structures, column (2) controls for demographic influences through the size of the society. This variable is statistically significant in some specifications, but the size and significance of δ changes little with the addition of this control.

Column (3) controls for geographic characteristics such as latitude, altitude, and agricultural potential. These controls marginally lower the value of δ. There is, however, no statistical difference in the point estimates of pathogen stress between columns (1) and (3). In column (4), I control for the amount of food surplus in the society and its technological sophistication. These variables are important for large buildings and structures in all specifications, but the point estimate for pathogen stress becomes more negative. In column (4), the point estimate for pathogen stress is −0.102 (s.e. 0.031), which is not distinguishable from the point estimate in (1). Column (5), adds the date of pinpointing of the society to control for potential measurement biases, but this variable has no effect in the estimation. Overall, the influence of pathogen stress on the complexity of large or impressive structures in society is stable even to the introduction of endogenous controls. Pathogen stress is likely to be capturing more than omitted or unobserved factors (Angrist and Pischke, 2009).

Columns (6) and (7) have a separate purpose. In both, I have included an African indicator for societies in this region, as in analyses of the *African dummy* in modern growth regressions (Collier and Gunning, 1999). Column (6) shows the mean difference between Africa and Eurasia. This difference is negative and significant. In column (7), I have included all controls and the African indicator variable. The point estimate for pathogen stress declines to −0.074 but it is still significant. Thus, the negative relationship between pathogen stress and large or impressive structures is *not* only driven by African societies or by the controls in (7).

The size of the estimated impact of pathogen stress is large. The difference in pathogen stress between tropical Africa and tropical America is $3.73 = (17.46 - 13.73)$, and the difference between tropical Africa and Eurasia is 4.64, Table 1. Further, the difference in large or impressive structures for these same regions is −0.53 and −1.41, respectively, Table 1. A point estimate of −0.10 indicates that about 70 percent (0.37/0.53) of the difference between Africa and tropical America can be accounted for by differences in pathogen stress. Similarly, about 30 percent (0.46/1.41) of the difference between Eurasia and Africa can be accounted for due to differences in pathogen stress. Finally, notice that many controls in X_i are endogenous. The fact that the estimates for δ change little makes the assumption of exogeneity in pathogen stress more credible.

Sensitivity analysis. Table 3 provides some robustness checks. Instead of using the order assumed by the SCCS, in panel I.A, I consider a dichotomous variable for the presence of religious, military, or industrial structures, which are considerably more likely to

Table 2
Reduced form estimates for the impact of disease on pre-colonial societies.

	Dependent variable: Large or impressive structures						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pathogen stress	−0.102***	−0.106***	−0.088**	−0.102***	−0.101***		− 0.074**
	(0.026)	(0.025)	(0.031)	(0.031)	(0.030)		(0.032)
Sub-Saharan Africa						−0.961***	−0.506**
						(0.20)	(0.25)
Community size		0.22***	0.20***	0.09	0.07		0.07
		(0.05)	(0.05)	(0.06)	(0.06)		(0.06)
Distance from equator			0.00	−0.01	−0.01		−0.01
			(0.01)	(0.01)	(0.01)		(0.01)
Log-altitude			−0.05	−0.04	−0.04		−0.02
			(0.04)	(0.03)	(0.03)		(0.03)
Agricultural potential			−0.01	−0.01	−0.01		−0.01
			(0.03)	(0.02)	(0.02)		(0.02)
Food surplus				0.30**	0.31**		0.30**
				(0.13)	(0.13)		(0.13)
Technological sophistication				0.06***	0.06***		0.05***
			(0.02)	(0.01)		(0.01)	
Pinpointing date					−0.00		−0.00
					(0.001)		(0.001)
R^2	0.18	0.24	0.24	0.32	0.32	0.20	0.34
N. Obs.	186	185	181	181	181	186	181

Note.– All specifications include (non-reported) controls for pre- and proto-agricultural societies. In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 3
Robustness checks for reduced form estimates of disease on development.

I. Dependent variable: Religious, military, or industrial structures only						
	(1)	(2)	(3)	(4)	(5)	(7)
			A. OLS estimates			
Pathogen stress	−0.025***	−0.027***	−0.022*	−0.026***	−0.026***	−0.020**
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
			B. Probit estimates			
Pathogen stress	−0.115***	−0.131***	−0.108**	−0.138***	−0.139***	−0.111**
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
N. Obs.	186	185	181	181	181	181
II. Dependent variable: Large or impressive structures						
	(1)	(2)	(3)	(4)	(5)	(7)
			A. Using the presence of the pathogen only			
Pathogen presence	−0.174***	−0.191***	−0.158**	−0.171***	−0.167***	−0.122*
	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
N. Obs.	186	185	181	181	181	181
			B. Order probit estimates			
Pathogen stress	−0.099***	−0.102***	−0.081**	−0.107***	−0.106***	−0.074**
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
			C. Controlling for population density rather than community size			
Pathogen stress	−0.102***	−0.107***	−0.093***	−0.105***	−0.103***	−0.073**
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
N. Obs.	183	182	178	178	178	178
III. Dependent variable: Large or impressive structures. Sub-samples						
	(1)	(2)	(3)	(4)	(5)	(7)
			A. Sub-Saharan Africa and tropical America only			
Pathogen stress	−0.060*	−0.085**	−0.080*	−0.080*	−0.07*	−0.05
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
N. Obs.	67	67	65	65	65	65
			B. Excluding Incas and Aztecs			
Pathogen stress	−0.098***	−0.103***	−0.078**	−0.094***	−0.093***	−0.068**
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
N. Obs.	184	183	179	179	179	179
			C. Excluding societies in the Sahel			
Pathogen stress	−0.074***	−0.085***	−0.079**	−0.096***	−0.088***	−0.062*
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
N. Obs.	158	157	153	153	153	153

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. The specifications coincide with those in Table 2.

provide evidence of strong pre-colonial conditions. In panel I.B, I use a probit estimate and the previous dichotomous dependent variable and report the point estimates for δ. In both cases, however, there is a negative and robust association between pathogen stress and this alternative measure of pre-colonial conditions, even in column (7) that includes a sub-Saharan Africa indicator.

In panel II.A, I consider the existence of the pathogens in the local environment of the society regardless of the severity and endemicity of the disease. The presence of the disease may be considered "more exogenous" than a measure that combines presence and severity. In this specification, the relationship between large structures and disease is equally strong. Panel II.B uses an order probit regression instead of a linear regression. The results, once again, are virtually unchanged. In panel II.C, I substitute community size for population density as a demographic control; there is also a negative association between pathogen stress and pre-colonial outcomes.

Finally, in panel III, I examine alternative sub-samples. I restrict the sample to tropical areas only to create a more meaningful comparison. I also exclude the Incas and Aztecs from the sample to increase comparability, and even exclude sub-Saharan African societies in the Sahel. In all these cases, the pattern observed in Table 2 remains unaltered with only minimal changes in the point estimates. The only notable change is in specifications (5) and (7) in panel III. In this set of regressions, which compare Africa and tropical America, the estimate of δ becomes insignificant once an African indicator is included. This is expected because the sample sizes are smaller and pathogen stress features limited variation.

Cities. The SCCS contains alternative measures of economic conditions, but they are difficult to validate.[11] Anthropometric measures are also available but they are not informative about comparative development.[12] Archeological censuses of cities are available, and they can corroborate the previous patterns. Archeological censuses are not disaggregated at the level of societies but invento-

[11] The SCCS contains a very large number of variables (i.e., over two thousand variables) some which could also serve as proxies for pre-colonial conditions. There are two limitations in using additional variables in the SCCS to assess pre-colonial development. First, one must be able to examine the SCCS data against independent assessments to identify systematic errors and biases. Urbanization in this subsection serves this purpose. Beyond urbanization, there seems to be no independent data for many of the aspects coded in the SCCS including measures of political centralization. Second, a large number of variables in the SCCS report information only about societies that are still present today. This means that many of the variables have a large number of missing observations. Using these variables will yield an incomplete (and perhaps incorrect) assessment of pre-colonial conditions.

[12] By this I mean *biological standards of living*based on anthropometric measures such as height. In particular, individuals in some African societies (i.e., the Tutsi) are exceptionally tall because height is a genetic adaptation that helps diffuse heat. Individuals in other societies (i.e., the Pygmies), are exceptionally short because they lack, for genetic reasons, the growth spur during adolescence. Any cross-sectional inference based on height would be uninformative because of the importance of these genetic adaptations. In a separate line of inquiry, Ehret (1998) used *linguistic archaeology* to trace out the diffusion and origin of pre-colonial technologies, but only for sub-Saharan Africa.

Table 4
Cities in Africa and the New World.

Year	North Africa	Sub-Saharan Africa				North America	South and Central America
		Muslims	Middle Nile and Ethiopia	Rest	Total		
			A. Number of cities with populations over 20,000 inhabitants				
800	10	0	2	3	5	0	10
1000	13	0	1	4	5	0	9
1200	18	6	2	4	12	0	10
1300	18	8	2	5	15	0	11
1400	18	8	2	9	19	0	18
1500	19	13	3	8	24	1	16
			B. Number of cities with populations over 40,000 inhabitants				
800	4	0	0	1	1	0	2
1500	7	4	0	2	6	0	6

Note.– Data from Chandler (1987, pp. 39–57). The indigenous cities in sub-Saharan Africa cover mostly Ghana, Zimbabwe and the Bantus. The middle Nile corresponds to Dongola (modern Sudan) and Kaffa. North Africa includes cities in the Mediterranean (i.e., Arabian, Egypt, Spanish Africa, and Aloa) and the Maghreb. Muslim cities are mostly in the Sahel.

Table 5
Pre-colonial population size in Africa and the Americas.

Region	Area	Biraben (1979)				McEvedy and Jones (1985)		
		400 BC	AD	1000	1500	AD	1000	1500
Africa								
North	2	10	14	9	9	8	11	8
Sub-Saharan	25	7	12	30	78	8	22	38
The Americas								
North	20	1	2	2	3	0.4	0.7	1.3
South and Central	20	7	10	16	39	4	8	13
World population		153	252	253	461	170	265	425

Note.– Population in millions. Area (mill. km ^{2}) from McEvedy and Jones (1985). North Africa includes the Maghreb, Libya and Egypt. The area in North Africa does not include the Sahara. North America includes the US, Canada, and the Caribbean.

ries of cities and their sizes vary over time and so they complement the "snapshot" in the SCCS.

Table 4 reports the number of cities with populations over 20 and 40 thousand inhabitants in sub-Saharan Africa and tropical America, which are comparable regions. (There is no systematic information about tropical cities and towns with less than 20 thousand inhabitants.) The inventory of cities comes from Chandler (1987) which, according to Connah (2001), provides accurate patterns of city formation in Africa. The table reports different time periods and divides sub-Saharan Africa into three sub-regions. The cities in regions with high Arab influence are coded as Muslims while the Middle Nile and Ethiopia are regions likely influenced by ancient Egypt. The cities in the rest of sub-Saharan Africa could be considered as "indigenous" though such classification is not necessarily relevant for the analyses presented here.[9]

Table 4 shows that during pre-Muslim times (i.e., 800), the number of medium- and large-sized cities in sub-Saharan Africa, 5 and 1, was half of the number of cities in tropical America, 10 and 2. After 1200, city formation changed rapidly in Africa due to the Arab-Muslim influence. In 1500, for example, Africa and tropical America had an equal number of large cities.

The civilizations in South and Central America originated in tropical rainforests similar to the African rainforest. None of their major cities were coastal and, in terms of latitude, their tropical cities were close to the cities in sub-Saharan Africa. For example, the latitude of the main cities of the Inca empire was about -13° and most Aztec cities were at latitudes near 20° to 22°. In Africa, the ruins of the Great Zimbabwe are at a latitude of -20°, while the cities of the Mali empire and predecessors (i.e., Djenné-Djenno, Gao, Kumbi Saleh), at a latitude of 15°. The latitudes of Meroë and Aksum (in East Africa) were 16° and 14°. None of these cities were coastal. (An exception are the Swahili coastal trading cities in the eastern coast; see footNote 9 for detailed references on city formation in Africa.).

The inventory of cities also serves to estimate *urbanization rates*, which are perceived as meaningful measures of pre-modern conditions (Acemoglu et al., 2002). To measure urbanization rates, Table 5 reports demographic data from Biraben (1979) and McEvedy and Jones (1985), which are independent sources. I use the size of the urban population in Table 4 and the total population sizes in Table 5.[13] For the year 800, these measures yield a range given by

$$\left(\frac{\text{African urbanization rate}}{\text{American urbanization rate}}\right)_{800} \in [0.18, 0.27],$$

which implies that the urbanization rate in sub-Saharan Africa was between one-fifth to one-third of the urbanization rate of tropical America. In 1500, Africa's urbanization rate was

$$\left(\frac{\text{African urbanization rate}}{\text{American urbanization rate}}\right)_{1500} \in [0.47, 0.68],$$

which is still lower than in tropical America.

Urbanization rates are consistent with the differences in the number and scale of cities in Chandler (1987), and with the reports about large buildings and structures across the large geographic regions in the SCCS. If large buildings and structures in the SCCS provide a "snapshot" of pre-colonial economic conditions, the upward trend in urbanization rates in Africa suggests that the SCCS

[13] To estimate the size of the urban population, I multiply the number of cities at each size by the cut-off size. For tropical America this number in 1500 is $10 \times 20,000 + 6 \times 40,000 = 440,000$. The same estimate for sub-Saharan Africa is 600,000. (If Muslim cities are excluded, the size of the urban population in Africa would only be 260,000.) Using population size from Biraben (1979) gives urbanization rates for tropical America of 1.12 percent and for Africa it gives 0.76 percent. Using population size from McEvedy and Jones (1985), the estimates are 3.38 and 1.57 percent. For 800, I use the urban estimates from 800 and the population sizes of 1000, and report the ratio between these values.

would likely overestimate African economic conditions in the past, for example, during the pre-Muslim era.

Differences in urbanization rates can be mapped into differences in *income per capita*. Acemoglu et al. (2002), Table 2) estimated an elasticity of GDP per capita to the urbanization rate of 0.038. Based on this elasticity, income per capita in sub-Saharan Africa in 800 was between 5 to 10 percent lower than income per capita in tropical America. In 1500, the income difference was between 1.5 to 7 percent, which is of the same order of magnitude as the estimate of GDP per capita provided by Maddison (2001, Table4.1).

Differences in urbanization rates during pre-colonial times suggest a caveat to the *Reversal of Fortune*argument in Acemoglu et al. (2002). Among former European colonies, Acemoglu et al. (2002) found a strong negative relationship between urbanization rates in 1500 and current income per capita. There has been no reversal in the ranking of economic conditions in the tropical world, i.e., between sub-Saharan Africa and tropical America. Acemoglu et al. (2002), Table 3 (1)) provides a point estimate of urbanization rates in 1500 on current GDP per capita for former European colonies of -0.078 (standard error 0.026). This estimate implies that current GDP per capita in Africa should be as much as 25 percent higher than in tropical America. Current income per capita differences are of the order of 4:1 in the opposite direction.

In Acemoglu et al.'s (2002) language, sub-Saharan Africa experienced a *Persistence of Misfortunes* relative to tropical America. This does not contradict Acemoglu et al. (2002). (Their empirical analysis of urbanization excluded sub-Saharan Africa arguing for measurement difficulties, although they used data from Chandler (1987) for other world regions.) Acemoglu et al. (2002), explain why North America and Australia are richer than tropical America and Africa. Their analysis, however, is not informative about differences within the tropics.

As land areas are roughly the same, Table 5 suggests that sub-Saharan Africa was more densely populated than tropical America; almost as densely populated as North Africa and Eurasia. This pattern seems to contradict the difference in urbanization rates. Population density, however, was *not* associated with urbanization in Africa.[14] In all other world regions, urbanization rates and population densities are strongly positively associated. Africa's distinct pattern has been documented in multiple independent analyses of political scientists (Fortes and Evans-Pritchard, 1940; Vengroff, 1976). In Appendix A, I verify that large and impressive structures are strongly positively associated to population density in all regions of the world, but sub-Saharan Africa.

Overview. Pathogen stress is negatively associated with the presence of large buildings and structures during the pre-colonial era. This association is robust and consistent with independently assessed measures of urbanization. Pathogen stress is primarily determined by environmental conditions, so the previous association is possibly a causal process. Obviously, such an association might reflect omitted variables, misspecification errors, and/or measurement errors. Given the data limitations, it is not possible to rule out any of these explanations. The previous associations might also be the reflection of biases in the data generating process or plain limited data quality. An advantage of focusing in comparisons within tropical areas is that they may potentially reduced Eurocentric biases. When seen through the eyes of European observers, a contrast between European societies and any tropical area in the past was likely to give a disproportionate attention to European achievements and views of the world. By focusing in the tropical world, the Eurocentric biases may be differenced out.

3. Disease and diversity

Several mechanisms could explain the influence of disease on past economic conditions. Pathogens may have a direct influence on pre-colonial outcomes through, for instance, a biological channel associated with reduced work energy and labor productivity (Bhattacharyya, 2009; Bloom and Sachs, 1998; Fogel, 1994; Kamarck, 1967). Given that diseases often acted on large groups of individuals, pathogens may also have an indirect influence on the social, economic, and political organization of society. This section proposes and tests an indirect channel of causation between pathogen stress and past economic conditions. The proposed channel is based on the hypothesis that disease limits social integration and increases ethnic diversity and fragmentation.

A causal process. The proposed causal process between disease and diversity argues that pathogen avoidance encourages social fragmentation and diversity as a way to counter negative spillovers in agricultural societies. To gain some perspective, Michalopoulos (2012) has shown that *physical barriers* to social integration (i.e., land quality and elevation), increased ethnolinguistic fractionalization. The thesis in this subsection is that *social barriers* act in a similar way.[15]

Consider a simple N-player symmetric game, where each of the players (i.e., ethnic groups) decides on the formation of a coalition of size $n \leqslant N$. The population size of group i is given by ℓ_i and their disease environment is associated with a *disease risk* given by $0 < d_i < 1$. Disease is risky in part because diseased individuals are unable to work. In the absence of coalitions, for example, production for group i would be given by the expected output from having ℓ_i workers producing under diminishing returns and facing a disease risk at a rate d_i. Expected output would be $y_i = (1-d_i)\ell_i^{\alpha} + d_i 0 = (1-d_i)\ell_i^{\alpha}$.

In a social world, disease is also risky because it carries negative spillovers, i.e., *contagion*. For group i, the disease risk of interacting with "outside" groups is multiplicative and given by $(1-d_1)\dots(1-d_{i-1})(1-d_{i+1})\dots(1-d_{n_i})$. This is the probability that *at least one*of the other $n_i - 1$ groups carries a communicable disease. In simple terms, for group i, all that matters to be negatively impacted by disease in a social setting is that at least one of the outside groups could transmit diseases. When the disease environments of the outside groups are identical, the risk of contagion is given by $(1-d_{-i})^{n_i-1}$. The risk of contagion is increasing in the number of outside groups in the coalition.

Social interactions with outside groups also carry benefits. By bringing outsiders into the coalition, available resources are augmented by $\ell_1 + \dots \ell_{i-1} + \ell_{i+1} \dots + \ell_{n_i} = (n_i - 1)\ell_{-i}$, which, again, assumes that the given population sizes of the outside groups are identical. The main decision for members of group i is the number of outside groups to invite to their coalition, i.e., $0 \leqslant n_i \leqslant N$. This decision balances the negative disease spillovers and the positive labor productivity gains. Under continuous choices for n_i, the unre-

[14] Data on population density contradict Herbst (2000), who argued that the lack of state consolidation in pre-colonial Africa was a consequence of a relatively low population density. Herbst (2000) argued that low population density made state control more difficult and competition for space less attractive. Implicit in Herbst (2000) is a comparison between state formation in Africa and Europe. The comparison between pre-colonial Africa and the Americas suggests a weaker link between state formation and population densities. The Americas had smaller population densities than Africa, but their independent development of states resembles the patterns seen in the earliest Eurasian states.

[15] Particular case studies about the role of geography on the Nigerian highlands and the Liangshan mountains between China and Tibet have been referenced by Cashdan (2001, p. 977). Disease also limited trade in sub-Saharan Africa through a separate mechanism briefly discussed by Hull (1976, p.9): "the *TseTse*fly, which attacked animals and rendered them unreliable as conveyors of goods. Consequently, nearly everything had to be carried atop human heads. Transport was therefore expensive and enormously inefficient." Alsan (2015) provides a case study of the *TseTse* fly in African development.

stricted value $(1-d_i)\arg\max\{(1-d_{-i})^{n_i-1}[\ell_i+(n_i-1)\ell_{-i}]^{\alpha}\}$ serves to derive group i's best response:

$$n_i^* = \min\left\{\frac{\ell_{-i}-\ell_i}{\ell_{-1}} - \frac{\alpha}{\ln(1-d_{-i})}, N\right\}. \quad (2)$$

This expression says that group i would weakly prefer a larger coalition if the outside groups are more populous (i.e., higher $\ell_{-i}-\ell_i$) and healthier (i.e., lower d_{-i}).

In a symmetric Nash equilibrium, the advantage of a more populous outside labor force disappears, so the equilibrium size of the coalition is $n^*(d)=\min\{-\alpha/\ln(1-d), N\}$. Comparative statics imply that the number of ethnic groups in a coalition is weakly decreasing in disease risk

$$\frac{\partial n^*(d)}{\partial d} \leq 0 \quad \text{with} \quad \lim_{d\downarrow 0} n^*(d) = N \quad \text{and} \quad \lim_{d\uparrow 1} n^*(d) = 0. \quad (3)$$

In other words, societies facing more adverse disease environments are less likely to interact with outside groups and hence less likely to form socially cohesive units. Disease, simply put, limits interactions by enacting a *social barrier* between unranked ethnic groups.

How fragmented is society? A conventional measure of (ethnic) fractionalization (e.g., Easterly and Levine (1997)) uses Herfindahl's concentration, as in $e = 1-\sum_i^N N^{-1}\left[n_i^*/N\right]^2$, or

$$e^*(d) = 1-[n^*(d)/N]^2 \simeq 1-[\min\{\alpha/d, N\}/N]^2,$$

where $\ln(1-d)\simeq -d$ is valid for small values of d. As $e^*(d)$ shows, social fragmentation is weakly increasing in disease risk, as higher values of d hinder social integration; see (3). How economically costly is disease and social fragmentation? In equilibrium, when $n^*(d)<N$, aggregate output is given by $y^*(d)=(1-d)^{n^*(d)}(n^*(d)\ell)^{\alpha}$, or

$$y^*(d) \simeq (1-\alpha)[\alpha/d]^{\alpha}\ell^{\alpha},$$

where $(1-d)^{\alpha/d}\simeq 1-\alpha$. The effects of disease on aggregate output show up in total factor productivity as disease influences how production is organized in society. Therefore, not only does disease lead to less cohesive social units, but it also lowers aggregate efficiency and output.

As both $y^*(d)$ and $e^*(d)$ vary with d, the previous model stresses that economic and social conditions are endogenous, which is important when thinking about causal identification. In addition, while the model is highly simplified, it showcases the social tension between disease and economic integration. This tension is not unique to ethnic fragmentation. In Birchenall (2014), I have expanded on the theoretical analysis of disease and social outcomes, and catalogued a number of historical instances in which past societies dealt with pathogens and communicable diseases through social barriers. I have not listed these instances here to avoid repetition. Segregation due to disease is consistent with systematic studies of biologists and anthropologists. For example, Cashdan (2001) and Fincher and Thornhill (2008) studied the geographical patterns of ethnic group distributions and language in the world. Without focusing on the mechanisms connecting disease and diversity, they showed that environmental factors such as unpredictable climate, pathogen stress, and habitat diversity shaped ethnic diversity. Cashdan (2001, pp. 975–976) concluded that "one of the strongest environmental predictors of high ethnic diversity is pathogen stress" and that "this pattern of relationships suggests that pathogens may be an important force in limiting the size of chiefdom/states." Similarly, Fincher and Thornhill (2008, p. 1289) conclude that "the worldwide distribution of indigenous human language diversity is strongly positively related to human parasite diversity."

Disease and diversity in the past. This subsection documents a strong positive association between pathogen stress and precolonial ethnic diversity. This association can be interpreted as causal under the assumption that disease is exogenous. I also show a strong negative association between ethnic diversity and large buildings and structures. The nature of this relationship is complicated because ethnic diversity and large buildings are both endogenous variables (i.e., both $y^*(d)$ and $e^*(d)$ are functions of d).

Fig. 2 plots the number of ethnic groups within 500 miles of each agricultural society in the SCCS against pathogen stress, across the main geographic regions studied here. The figure shows a strong positive association between disease and ethnic diversity; more ethnic groups are present in more disease-prone areas. The figure also shows that the previous relationship is primarily, but not only, driven by African societies.

Table 6 presents OLS estimates of Eq. (1) using the number of ethnic groups within 500 miles as dependent variable, and similar specifications as Table 2. Overall, the relationship between pathogen stress and ethnic diversity is positive. Column (1) includes no controls and shows that higher pathogen stress is associated with higher ethnic diversity. The magnitude and significance of this effect is robust to the addition of demographic controls in the form of community size, column (2). Community size is an important control because measures of ethnic diversity in the SCCS are based on physical distance. Column (3) controls for geographic factors such as latitude, altitude, and agricultural potential. Latitude and altitude often play a significant role. For instance, ethnic diversity is concentrated in tropical and mountainous areas. The point estimate for disease declines in column (3) from 4.6 to 3.3, but the statistical significance of the point estimates survives the inclusion of these geographic controls. This suggests that the association between pathogen stress and diversity is not masking channels of geographic variability previously studied by the literature; see, e.g., Michalopoulos (2012). Controlling for food surplus and technological sophistication in the society, column (4), leaves the previous results virtually unaffected. Finally, column (5) adds the date of the pinpointing of the society. This control is significant, but it has a small quantitative impact on the point estimate for pathogen stress.

As before, columns (6) and (7) add a regional control for sub-Saharan Africa. Column (6) shows the mean difference in ethnic diversity between Africa and Eurasia. This difference is large, pos-

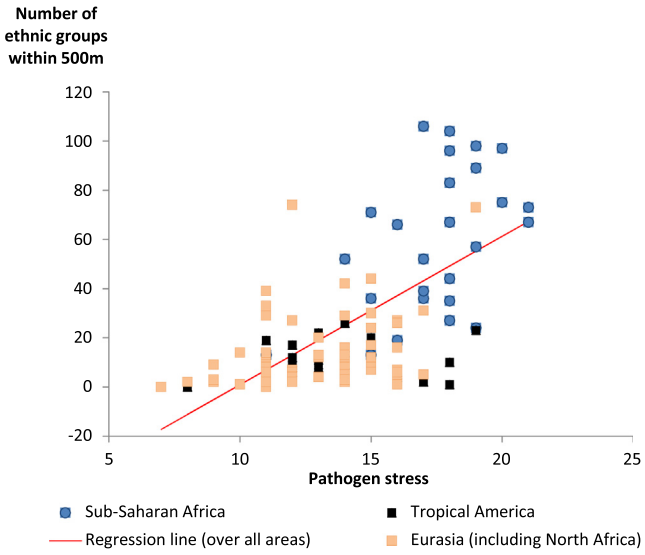

Fig. 2. Pathogen stress and number of ethnic groups (withing 500 miles) in agricultural societies in the SCCS.

Table 6
Disease and diversity in the pre-colonial era.

	Dependent variable: Ethnic groups within a 500 miles radius						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pathogen stress	4.64***	4.59***	3.29***	3.30***	3.26***		1.42**
	(0.60)	(0.60)	(0.68)	(0.70)	(0.69)		(0.59)
Sub-Saharan Africa						45.33***	35.20***
						(5.34)	(5.49)
Community size		1.06	1.11	1.08	1.59		1.33
		(1.19)	(1.21)	(1.40)	(1.45)		(1.22)
Distance from equator			−0.39***	−0.39***	−0.37**		−0.37***
			(0.12)	(0.15)	(0.15)		(0.12)
Log-altitude			1.65***	1.69***	1.60***		0.73
			(0.55)	(0.56)	(0.56)		(0.48)
Agricultural potential			−0.41	−0.40	−0.31		−0.37
			(0.54)	(0.55)	(0.54)		(0.50)
Food surplus				0.91	0.89		2.00
				(2.37)	(2.33)		(2.21)
Technological sophistication				−0.01	−0.07		0.01
			(0.33)	(0.33)		(0.28)	
Pinpoint date of society					0.03**		0.03**
					(0.01)		(0.01)
R^2	0.31	0.31	0.35	0.36	0.37	0.45	0.54
N. Obs.	183	182	178	178	178	183	178

Note.– All specifications include (non-reported) controls for pre-and proto-agricultural societies. In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

itive, and statistically significant; see also Table 1. Including all controls and the African indicator in column (7) reduces the point estimate for pathogen stress to 1.42 (s.e. 0.59). This point estimate, however, is still significant. Thus, as in Table 3, the relationship between pathogen stress and ethnic diversity is not driven only by African societies or by the controls included in column (7). This observation is important because Tables 1 and 6 show that the regional variation in ethnic diversity is large. For example, the R_2 in Table 6, column (6), suggests that the sub-Saharan African control accounts for a considerable fraction of the total variation in ethnic diversity in the world. Columns (6) and (7) also show that the set of controls in Table 6 account for about 20 percent of Africa's ethnic diversity, as measured by the decline in the African indicator in column (7) relative to column (6); the African indicator declines by about 20 percent, from 45 to 35.

To assess the quantitative significance of disease, one can use the point estimates of Table 6 and the difference in pathogen stress between Africa and tropical America, 3.73, or the difference between Africa and Eurasia, 4.64. The point estimate for disease in column (7), 1.42, the lowest point estimate in Table 6, implies that one should see about 5.3 and 6.5 more ethnic groups in Africa relative to the previous regions. Pathogen stress thus accounts for about 11 and 14 percent of the baseline differences seen in Table 1. A difficulty with Table 6, and therefore with the previous quantitative assessment, is that the estimates of disease on diversity are not as "stable" as in Tables 2 and 3. The decline in these point estimates as the number of controls increases suggests an omitted variable correlated with disease and diversity.

Sensitivity analysis. Given the large number of variables assembled in the SCCS (see footnote 11), there is no limit to the controls one can add to the previous specifications. I consider some of them in the sensitivity analyses that follow. This subsection pays special attention to alternative sub-samples and additional controls. I also consider some more standard checks, such as those in Table 3. Panel I.A in Table 7, for example, considers just the presence of the pathogens in the local environment. The point estimate for disease follows the same pattern as in Table 6, but disease becomes insignificant once the African indicator is included. (The p-value, however, is only 0.11 in column (7).) As in Table 6, column (7), the effects of disease weaken because most of the variation in ethnic diversity is between African and non-African regions, and the African indicator absorbs most of this variation. Panels I.B and I.C show that replacing community size for population density as a control has no effect on the point estimates but restores significance in column (7), and that using the number of ethnic groups in a 250 mile radius yields similar results.

Panel II in Table 7 studies several sub-samples. One of the most interesting comparisons is between tropical America and sub-Saharan Africa. The New World developed independently from the Old World, and as I previously noted, tropical America and sub-Saharan Africa shared many similarities in terms of technology and geography, but an important asymmetry in their disease environments. Despite the large reduction in sample sizes, the point estimates of disease in this comparison (panel II.A) are larger and more precisely estimated than in the benchmark cases of Table 6.

Sub-Saharan Africa had repeated contact with Eurasia through the Nile River, the Sahara (by the Arab trade that started during the seventh century AD), and the Indian ocean (by the East African trade in medieval times). Given their repeated interactions before the European expansion, panel II.B considers a comparison between Eurasia and sub-Saharan Africa. The effects of disease are also stronger than in Table 6.[16]

The only sub-samples in which disease has no systematic effect on ethnic diversity are those that include North and sub-Saharan Africa, and those that exclude tropical Africa entirely, panels C and D. It is not accurate to state that disease does not matter in these specifications. Disease has a positive effect on diversity even when demographic controls are in place, columns (1) and (2). This positive effect, however, is accounted for by the geographic, technological, and environmental factors included in columns (3) and (4). These specifications show that geographic, technological, and environmental factors help explain ethnic diversity in non-African societies but do not help account for Africa's ethnic diver-

[16] Contact was enough to expose Africa to the epidemiological conditions of Eurasia. McNeill (1976, p. 130) noted that "when African slaves began to come to the new world after 1500, they suffered no spectacular die-off from contact with European diseases, which is sufficient demonstration that in their African habitat some exposure to the standard childhood diseases of civilization must have occurred." As Curtin et al. (1995, p. 242) note, even in the more distant regions, "the Bantu-speaking farmers of southern Africa had some immunities to help protect them from the European diseases that exterminated the indigenous people of Tasmania and decimated those of Patagonia."

Table 7
Robustness checks for reduced form estimates of disease and diversity.

I. Dependent variable (for A and B): Ethnic groups within a 500 miles radius						
	(1)	(2)	(3)	(4)	(5)	(7)
			A. Using the presence of the pathogen only			
Pathogen presence	7.58***	7.49***	4.78***	4.81***	4.72***	1.65
	(1.15)	(1.14)	(1.25)	(1.29)	(1.26)	(1.05)
			B. Controlling for population density rather than community size			
Pathogen stress	4.64***	4.64***	3.29***	3.38***	3.35***	1.55***
	(0.60)	(0.59)	(0.67)	(0.68)	(0.68)	(0.59)
			C. Using ethnic groups within a 250 miles radius as dependent var.			
Pathogen stress	1.49***	1.47***	1.04***	1.05***	1.04***	0.50**
	(0.22)	(0.22)	(0.25)	(0.26)	(0.26)	(0.22)
II. Dependent variable: Ethnic groups within a 500 miles radius. Sub-samples						
	(1)	(2)	(3)	(4)	(5)	(7)
			A. Sub-Saharan Africa and tropical America only			
Pathogen stress	5.48***	5.26***	3.64***	3.42***	3.34***	1.69*
	(0.86)	(0.87)	(1.09)	(1.12)	(1.05)	(0.93)
N. Obs.	67	67	65	65	65	65
			B. Societies in the Old World			
Pathogen stress	5.87***	5.80***	4.88***	5.13***	5.13***	2.82**
	(0.62)	(0.61)	(0.74)	(0.73)	(0.73)	(0.78)
N. Obs.	118	117	113	113	113	113
			C. Sub-Saharan and North Africa only			
Pathogen stress	6.75***	6.49***	1.75	2.27*	2.30*	1.41
	(1.20)	(1.19)	(1.26)	(1.14)	(1.17)	(1.29)
N. Obs.	43	43	41	41	41	41
			D. Non-African societies			
Pathogen stress	1.40***	1.40***	0.37	0.33	0.35	0.35
	(0.43)	(0.44)	(0.57)	(0.55)	(0.54)	(0.54)
N. Obs.	151	150	148	148	148	148
III. Dependent variable: Ethnic groups within a 500 miles radius. Added controls						
	(1)	(2)	(3)	(4)	(5)	(7)
			A. Controlling for slavery			
Pathogen stress	4.38***	4.33***	2.93***	2.95***	2.79***	1.34**
	(0.58)	(0.58)	(0.64)	(0.66)	(0.65)	(0.56)
			B. Controlling for political centralization			
Pathogen stress	4.69***	4.64***	3.37***	3.36***	3.30***	1.50**
	(0.62)	(0.62)	(0.69)	(0.69)	(0.69)	(0.60)

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. The specifications coincide with those in Tables 2 and 6.

sity. The point of this paper is that pathogen stress, heavily concentrated in sub-Saharan Africa due to environmental factors, serves in part as the "missing link" to account for Africa's ethnic diversity.

Omitted variables might still be unaccounted for. I next consider additional specifications that add controls that have been relevant for explaining ethnic diversity. Panel III.A includes a control for slavery in the society.[17] Slavery was practiced during pre-colonial times in Africa and it is a likely determinant of ethnic diversity. As Nunn (2008, p. 164) noted, "slave trades tended to weaken the links between villages, thus discouraging the formation of larger communities and broader ethnic identities." At the same time, slavery may have been rendered possible because of ethnic diversity, so it is clear that slavery is an endogenous control. Still, while the effect of slavery (not reported) is often statistically significant, its inclusion does not weaken the baseline estimates.

Linguistic, cultural, and ethnic fragmentation tend to disappear after the consolidation of states (Weber, 1976).[18] Food surplus and technological sophistication are important determinants of state consolidation. They are statistically significant determinants of the presence and scope of large buildings and structures; see, columns (4), (5), and (7) of Table 2. These variables, however, do not weaken the effect of disease on diversity in the previous specifications. Panel III, specification B, add controls for political centralization. These controls leave the baseline estimates of disease on diversity unaffected.

Finally, the effect of disease on ethnic diversity is primarily driven by sub-Saharan Africa. It is thus possible that an "African factor" might be responsible for the previous findings. A biased (i.e., Eurocentric) view of Africa or Africa's long history relative to other world regions are some examples of such "African factor." (For instance, as Ashraf and Galor (2013) and Spolaore and Wacziarg (2009) argue, Africa's long history is likely responsible for Africa's high genetic diversity and possibly its ethnic diversity.) To examine this possibility, I consider the effect of disease on diversity in pre- and proto-agricultural societies (i.e., hunter-gatherers). These societies are smaller in scale, less politically sophisticated, and geographically mobile so one should expect no effect of pathogenic stress on diversity, i.e., pathogen avoidance is more easily achieved in small and mobile societies. If there is a built-in bias toward Africa or an "African factor," one should also find a significant African indicator for this hunter-gatherers.

Table 8 presents three specifications. In (1), I control for disease but no other covariate, in (2) I include an African indicator and no other control, and in (3) I control for all the influences considered in the richest specification of Table 6. Sample sizes are small, but

[17] A detailed analysis of slavery and forced labor during the pre-colonial era is beyond the scope of this paper but it has been partly carried out by Patterson (1982) for some societies in the SCCS. Patterson (1982) coded the presence and approximate origin of slave populations in the world. I use §v917 in the SCCS, as a control.

[18] Weber (1976, p. 7), for example, noted in his analysis of state consolidation in France that "[d]iversity had not bothered earlier centuries very much. It seem part of the nature of things, whether from place to place or one social group to another. But the Revolution had brought with it the concept of national unity as an integral and integrating ideal at all levels, and the ideal of oneness stirred concern about shortcomings. Diversity became imperfection, injustice, failure, something to be noted and to be remedied."

Table 8
Disease and diversity in non-agricultural societies.

	Dep. var.: Ethnic groups in 500 m.		
	(1)	(2)	(3)
Pathogen stress	0.15		−2.30
	(1.02)		(1.66)
Sub-Saharan Africa		27.16**	22.83
		(13.39)	(17.10)
Full set of controls	No	No	Yes
R^2	0.00	0.08	0.25
N. Obs.	64	64	64

Note.– The sample is based on pre- and proto-agricultural societies. Full set of controls corresponds to specification (7) in Table 6. In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

column (1) finds no relationship between ethnic diversity and disease for non-agricultural societies. There is, however, a significant difference in ethnic diversity between African and non-African regions, column (2). The point estimate of the African indicator is smaller than in Table 6 and its explanatory power and statistical significance are smaller. Once the full set of controls is included, the African indicator and pathogen stress *both* become insignificant. In conclusion, there appears to be no systematic relationship between disease and diversity for non-agricultural societies and no systematic difference in ethnic diversity between non-agricultural societies in Africa and non-African regions.[19]

4. The persistence of pre-modern influences

Past economic and social outcomes are interesting on their own, but history also matters as it aids in the understanding of current economic and social developments. This section examines the persistence of pre-modern influences. The discussion is brief because a complete treatment of the persistence of historical factors requires a separate paper. This section reports four findings: (i) pathogen stress is negatively associated with current income per capita, (ii) pathogen stress is positively associated with current measures of ethnolinguistic fractionalization (ELF), (iii) pre-colonial ethnic diversity is negatively associated with current economic conditions and positively associated with current ELF, and (iv) pathogen stress was a central contributor to the high mortality of European settlers during the colonization of tropical areas.

Modern data. To link the SCCS to current outcomes, I match each of the 186 societies to the countries where these societies located. Appendix D lists the countries on which each society resided according to the latitude and longitude provided by the SCCS. (To ensure a correct imputation, I used the Human Relations Area Files, which separately lists the location of different cultures around the world.) Appendix D also provides a map with the location of each of the societies. The matching of the SCCS to countries is fairly straightforward for agricultural societies because there is limited geographic mobility in these societies. A concern is that, sometimes, many societies locate in one country, and sometimes one society locates in many countries. I allow for both possibilities and average SCCS variables across countries. This concern is important because divided societies are more likely to experience civil wars, as noted by Michalopoulos and Papaioannou (2013). As a robustness check, I have also used the country where the society is primarily located, but the results are virtually identical.

Some societies, notably China, have been politically independent and unified for centuries. For these societies, there is only a subtle distinction between pre-modern societies and modern states. Other societies, notably the empires in the New World, were disintegrated by European influences. Using modern income per capita thus requires special care due to colonization. I control for colonial influences in three main ways. First, I focus on former colonies only. This essentially restricts the sample to countries in Africa and the New World, which has been an important comparison throughout the paper. Second, I control for population densities in 1500 as they were arguably an important determinant of colonial policies. Less densely populated colonies provided fewer incentives for *extractive institutions* (Acemoglu et al., 2002). Finally, I control for whether the country was a British, French, German, Spanish, Italian, Dutch, or Portuguese colony. I also add numerous controls associated with religion, natural endowments, and geography and climate.[20]

Pathogens, income, and fractionalization. Table 9 presents two sets of results. Panel A reports the reduced-form association between the log of GDP per capita in 1995 and pathogen stress. This reduced-form regression is the modern version of Tables 2 and 3. Panel A shows that pathogen stress has a negative effect on income per capita. Controls for colonial influences, columns (2) and (3), weaken the point estimate of pathogen stress, but disease and current income per capita are still negatively associated in these specifications. There is a negative effect of pathogen stress on income per capita even after the many additional controls for natural endowments and geography are in place. To gain some perspective, there is a total of 28 controls in specification (6).

Panel B uses the average of five different indices of *ethnolinguistic fractionalization* from Easterly and Levine (1997), as used in Acemoglu et al. (2001). These regressions are the modern version of Tables 6 and 7. Panel B shows three things: first, the influence of pathogen stress on ELF is positive and statistically significant, with the exception of specification (6) in which pathogen stress is not significantly different from zero for reasons that I discuss below. Second, in all other specifications, the point estimate for pathogen stress is stable. Table 9, panel B, thus suggests that pathogen stress is important for understanding *current* ethnic fragmentation. In conjunction with the findings of the previous section, it also suggests that the effects of disease were not drastically perturbed by colonial influences. Third, an important component of ethnolingusitic fragmentation is environmentally determined. Because specification (6) has numerous environmental controls, the inclusion of temperature, humidity, latitude and similar variables weakens the significance of pathogen stress. Essentially, there is a strong correlation between environmental factors, the severity of pathogen stress, and ethnic fragmentation behind the estimate of disease in panel B, specification (6).

[19] These estimates also discount systematic differences between Africa and non-African regions due to sample selection and prejudice about Africa. For example, since most of the population movements in the long pre-agricultural period have been out of Africa, the sample of societies in Africa may be a negatively selected sample of hunter-gatherers whereas the hunter-gatherers that migrated into the Americas had positive attributes that later conduced to faster agricultural developments. Further, because most of the societies in Africa were described at the peak of colonization, there may be prejudice against African societies. The fact that there are no predictable differences between the non-agricultural societies in Africa and the rest of the world suggests that prejudice and selection may not be crucial concerns.

[20] Unless otherwise noted, I will rely on the variables coded by Acemoglu et al. (2002, Appendix 2) since I want to explicitly minimize any interference introduced by including multiple variables from the SCCS. I obtained all the data from their online Appendix. *Religion* includes controls for catholic, protestant, muslim, and other. *Endowments* include mineral reserves associated with coal production, and measures of natural reserves as percent of the world for gold, iron, silver, zinc, and oil. This set of controls also includes whether a country is landlocked or an island. *Geography*controls for average temperature, minimum monthly high, maximum monthly high, minimum monthly low, and maximum monthly low. It also includes humidity variables associated with morning minimum, morning maximum, afternoon minimum, and afternoon maximum as well as soil quality in the form of steppe (low latitude), steppe (middle latitude), desert (middle latitude), dry steppe wasteland, desert dry winter, and highland. The richest specification has 28 controls.

Table 9
Income per capita, disease, and ethnolinguistic fragmentation (ELF).

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Dependent variable: Log-GDP per capita					
Pathogen stress	−0.219***	−0.161***	−0.130***	−0.135***	−0.108***	−0.084**
	(0.026)	(0.027)	(0.033)	(0.033)	(0.040)	(0.044)
R^2	0.43	0.54	0.63	0.65	0.73	0.80
N.Obs/N.controls	67/1	66/2	66/10	65/12	65/20	65/28
	B. Dependent variable: Ethnolinguistic fractionalization (ELF)					
Pathogen stress	0.035***	0.037***	0.030***	0.032***	0.034**	−0.008
	(0.009)	(0.011)	(0.011)	(0.011)	(0.014)	(0.016)
R^2	0.13	0.14	0.34	0.41	0.57	0.68
N.Obs/N.controls	71/1	68/2	68/9	68/12	68/20	68/28
Controls included						
Pop. density 1500	No	Yes	Yes	Yes	Yes	Yes
Colonizer id.	No	No	Yes	Yes	Yes	Yes
Religion	No	No	No	Yes	Yes	Yes
Endowments	No	No	No	No	Yes	No
Geography	No	No	No	No	No	Yes

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. For variable definitions please see the main text, specially footnote20.

Recall from Table 1 that the difference in pathogen stress in sub-Saharan Africa and tropical America is $3.73 = 17.46 - 13.73$. This difference changes little once the societies have been matched to actual countries. (For instance, the difference at the country level is $3.30 = 16.88 - 13.58$.) The log-income per capita gap between both regions in the sample is $-1.05 = \ln(\$4,478) - \ln(\$1,564)$. Using the OLS point estimate of panel A, column (6), the lowest point estimate in the table, gives a predicted difference of $-0.084 \times 3.73 = -0.31$. This difference is about 30 percent of the actual difference between both regions, which is a large value especially as I have already considered some competing explanations. For example, as I have controlled for colonizer identities and population density in 1500, the estimates already take into account potential colonial influences associated with differential colonial policies and the *Reversal of Fortune* studied by Acemoglu et al. (2002). I have also controlled for religion, natural endowments, and geography and climate. In fact, the R_2 in the richest specification is about 80 percent.[21]

Appendix C considers several alternative specifications. I use presence of pathogens, discounting their severity and endemicity. I also add controls for food surplus and technological sophistication based on the SCCS data. The findings of Table 9 remain unchanged in these specifications. I also expand the sample to include non-colonies, and find that pathogen stress plays a larger role on income per capita and ethnic fragmentation in this specification. In other checks, I excluded sub-Saharan Africa from the sample. None of the results are significant, which is consistent with the discussion of the previous section. Sub-Saharan Africa is the most important source of variation in pathogen stress and ethnic diversity or ethnolinguistic fractionalization in the data. Thus excluding this region renders most estimates insignificant. Even a contrast between North Africa and sub-Saharan, however, often yields strong and significant results. Finally, I include countries in tropical America and Africa only. In this case, the point estimates are also often negative and significant, as in Table 9.

Ethnic diversity, income, and fractionalization. Table 10 presents three sets of results: panel IA reports OLS correlations between GDP per capita and pre-colonial ethnic diversity. Panel IB reports the same correlations but with current ELF instead of the pre-colonial measure. The correlations between income and ELF are well-known (Easterly and Levine, 1997), but the general change in the point estimates in Panel B provides a useful benchmark for Panel A. Finally, panel II reports OLS correlations between pre-colonial ethnic diversity and current ELF.

Ethnic diversity and income per capita are endogenous variables. As the simple analytical model suggests, both are outcomes that respond to disease environments making a causal assessment of their relationship difficult (Appendix B). Table 10, panels A and B, shows that past levels of ethnic diversity and current levels of ethnolinguistic fractionalization are negatively correlated with income per capita. The correlation between pre-colonial ethnic diversity and income is less robust than the correlation between current ELF and income, possibly due to differences in sample sizes. The general pattern in panels A and B, however, is similar in the sense that the addition of controls weakens these correlations. For example, the point estimates in panel A decline as the endowments and geographic controls are added; specifications (5) and (6) are not statistically significant. One difference between these panels is that panel A has considerably fewer observations compared to panel B. For instance, as Appendix C shows (Table C4, panel C), if non-colonies are added to the sample in Table 10, pre-colonial ethnic diversity remains significant, even in the richest specifications.[22]

Panel II provides a more compelling view of the role of pre-colonial diversity on current outcomes. This panel shows that pre-colonial and current measures of ethnic diversity are strongly positively related and that their relationship is stable across specifications. This means that the current high level of ethnic fragmentation in Africa is not an outcome due exclusively to colonial influences. In fact, Table 11 suggests that colonization did not drastically altered the role of pre-colonial influences on current ethnic diversity. For instance, if the current high levels of ethnic fragmentation in Africa were solely due to colonial factors, one should see a marked decline in the point estimates in panel II, columns (2) and

[21] I cannot directly control for urbanization which is a more precise measure of pre-colonial conditions in Acemoglu et al. (2002). The measures of urbanization previously developed for sub-Saharan Africa apply to the entire area and do not vary within Africa. One of the robustness checks in Appendix C adds large or impressive structures as a control. This leaves the main results of Table 9 intact.

[22] The significance of pre-colonial diversity on current income per capita once non-colonies are added to the sample could be due to larger sample sizes or to the fact that past ethnic diversity is more important for non-colonies than for European colonies. For instance, a homogenous population during pre-modern times might be an important determinant of modern economic growth; see, e.g., Ashraf and Galor (2013) and Spolaore and Wacziarg (2009) for discussions centered on the role of genetic diversity on current economic outcomes. Appendix C also uses, for robustness, measures of ethnic diversity based on the number of ethnic groups within a 250 miles radius and several other checks. The point estimates change in the same way as in the main tables, but all specifications yield negative and often significant results.

Table 10
Income per capita, ethnolinguistic fractionalization (ELF), and ethnic diversity.

I. Dependent variable: Current log-GDP per capita	(1)	(2)	(3)	(4)	(5)	(6)
			A. Pre-colonial ethnic diversity			
Ethnic groups within 500 m.	−0.019***	−0.014***	−0.009**	−0.010**	−0.004	−0.003
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
R^2	0.23	0.47	0.57	0.59	0.70	0.79
N.Obs/N.controls	67/1	66/2	66/10	65/12	65/20	65/28
			B. Ethnolinguistic fractionalization			
Ethnolinguistic fractionalization (ELF)	−1.382***	−1.264***	−1.083***	−1.119***	−0.838***	−0.783**
	(0.31)	(0.26)	(0.28)	(0.30)	(0.28)	(0.33)
R^2	0.18	0.50	0.58	0.58	0.70	0.75
N.Obs/N.controls	87/1	85/2	85/9	85/12	84/20	84/28
II. Dependent variable: Ethnolinguistic fractionalization (ELF)	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic groups within 500 m. (×100)	0.66***	0.77***	0.65***	0.66***	0.60***	0.45***
	(0.07)	(0.08)	(0.09)	(0.09)	(0.12)	(0.10)
R^2	0.32	0.37	0.51	0.56	0.64	0.71
N.Obs/N.controls	71/1	68/2	68/9	68/12	68/20	68/28
Controls included						
Pop. density 1500	No	Yes	Yes	Yes	Yes	Yes
Colonizer id.	No	No	Yes	Yes	Yes	Yes
Religion	No	No	No	Yes	Yes	Yes
Endowments	No	No	No	No	Yes	No
Geography	No	No	No	No	No	Yes

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. For variable definitions please see the main text, specially footnote 20.

(3), as colonizer controls are added. The point estimate for ethnic diversity does not decline as a result of religion, endowments, or geographic controls. Thus, Table 10 shows that pre-colonial diversity is related to current outcomes, though its effects are more likely to be channeled through ELF rather than as a direct influence on income per capita.

Pathogen stress and settler mortality. By examining the pre-colonial past, it is possible to rule out alternative causal links between disease and economic development that emphasize colonial influences. The *colonial origins*hypothesis proposed by Acemoglu et al. (2001) considered a causal path through which disease lowers current income per capita. Briefly stated, this hypothesis argues that settler mortality rates influenced current outcomes through their effect on European settlements and political institutions. I next show that pathogen stress was an important component behind the mortality faced by European settlers, as settler mortality becomes insignificant as a determinant of institutions once pathogen stress is controlled for.

Table 11 examines the influence of disease on the protection against expropriation risk, which is a key measure of institutional quality. Column (1) includes settler mortality on the baseline sample of Acemoglu et al. (2001). Column (2) presents the same estimate for the 58 countries with data on settler mortality also linked to the SCCS. The point estimates in both columns are nearly identical and consistent with Acemoglu et al. (2001), Tables 3–6). The estimates for pathogen stress in columns (3) and (4) show that the sample in which pathogen stress and settler mortality are both available is representative. Column (5) shows that when both pathogen stress and settler mortality are included as predictors of protection against expropriation risk, settler mortality becomes insignificant. This suggests that pathogen stress is likely the ultimate determinant of the adverse mortality conditions faced by Europeans in sub-Saharan Africa and the New World during colonial times. Column (6) shows that restricting the sample to former colonies has no effect on the previous estimates. In column (6), settler mortality is also statistically insignificant while pathogen stress remains significant.

What conclusions can be made based on Table 11? It is likely that the (exclusion) restrictions required to identify the causal effect of political institutions on long-term economic development through settler mortality may be invalid. The reason is that pathogen stress, a better predictor of current institutions than settler mortality, influenced ethnic diversity and past economic conditions. The problem is not that disease environments are not important for political institutions (i.e., a weak instrument). On the contrary, the issue is that there are many different pathways for disease environments to influence current and past social and economic structures. While disease likely influenced European settlements and political institutions through settler mortality, disease also likely influenced the social and economic organization of production on areas under European influence, but long before the European expansion.[23]

5. Some concluding remarks

This paper studied the influence of disease on economic development, particularly the relationship between pathogen stress and pre-colonial economic and social conditions in sub-Saharan Africa and the tropical regions of the New World. I used the complexity of large or impressive structures in ethnographic accounts of pre-industrial societies at the time of first contact with Westerners to measure pre-colonial economic conditions. Large and impressive structures are not pure measures of economic conditions. Yet, they indicate a society's capacity to generate, extract, and mobilize resources. In the SCCS, the pace at which small-scale societies developed into large-scale tributary empires and states was adversely affected by pathogen stress.

The path connecting disease and economic development in pre-colonial times relied on the idea that parasitic diseases limit social

[23] A modified view, proposed by Glaeser et al. (2004), argues that European settlements also influenced human capital, not only political institutions. Both views, however, place the influence of disease during colonization at the center stage of current economic disparities.

Table 11
Protection against expropriation risk, settler mortality, and pathogen stress.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Protection against expropriation risk					
Pathogen stress			−0.361***	−0.327***	−0.257***	−0.242***
			(0.046)	(0.059)	(0.068)	(0.069)
Settler mortality	−0.646***	−0.654***			−0.238	−0.234
	(0.125)	(0.152)			(0.179)	(0.197)
Linked to SCCS	No	Yes	Yes	Yes	Yes	Yes
Country in AJR	Yes	Yes	No	Yes	Yes	Yes
Only colonies	No	No	No	No	No	Yes
R^2	0.30	0.29	0.42	0.40	0.42	0.38
N. Obs.	74	58	81	58	58	54

Note.– In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level. Linked to SCCS means that at least one society in the SCCS resided in the country. Country in AJR means that the country is part of the sample studied by Acemoglu et al. (2001). Only colonies means that the country is a former European colony.

integration and productive efficiency through *social barriers*. This idea was expressed in formal terms as a Nash equilibrium of a simple coalitional game where social groups are potentially at risk due to disease spilliovers. In the SCCS, high disease prevalence is associated with high levels of ethnic diversity. This association is robust across specifications and controls, although the negative effects of disease on the social organization of pre-colonial societies tend to be concentrated in sub-Saharan Africa. The effects of disease on economic development and ethnic diversity are also persistent. They remain relevant even once many controls for colonial influences are taken into account. As infectious and parasitic diseases have been highly prevalent in tropical Africa due to environmental reasons, the findings tentatively suggest that Africa's slow and complex transition into modern economic growth may have some deep roots in its pre-colonial past.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Population density and urbanization

During the pre-colonial era, existing demographic accounts suggest that sub-Saharan Africa was more densely populated than tropical America. Yet, urbanization rates were lower and large and impressive structures were less prevalent in Africa than in any other region. This Appendix argues that African development followed a divergent pattern for urbanization. It shows that in all other world regions in the SCCS, large and impressive structures are strongly positively associated with population density. In sub-Saharan Africa, these variables are independent. This Appendix also provides additional remarks about this pattern, and the popular use of population density to measure economic conditions during the pre-colonial era.

Table A.1, panel I, shows that there is no systematic association between population density and the presence of large buildings and impressive structures *in* Africa. Panel II shows the same results for the presence of religious, military, or industrial structures only. These results contrast with a positive association in all other agricultural societies in the world. Specifically, column (1) of Table A.1 reports the OLS correlation between the presence of large buildings and population density in all societies. Column (2) examines the same correlation within Africa. This correlation is actually negative although not statistically significant. In non-African regions, this correlation is strong and positive, column (3). This is also true for societies in Eurasia, column (4). In tropical America, the correlation is actually stronger than in non-African societies, column (5), and not solely driven by the presence of large empires such as the Incas and the Aztecs, column (6).

The pattern documented in Table A.1 has also been evident to African political scientists. Fortes and Evans-Pritchard (1940, pp. 7–8), for example, in discussing the influence of demography on African political systems, noted that

> "[I]t would be incorrect to suppose that governmental institutions are found in those societies with greatest [population] density. The opposite seems to be equally likely, judging by our material. The density of the Zulu is 3.5, of the Ngwato 2.5, of the Bemba 3.75 per square mile, while that of the Nuer is higher and of the Tallensi and Logoli very much higher. It might be supposed that the dense permanent settlements of the Tallensi would necessarily lead to the development of a centralized form of government, whereas the wide dispersion of shifting villages among the Bemba would be incompatible with centralized rule. The reverse is actually the case. In addition to the material collected in this book, evidence from other African societies could be cited to prove that a large population in a political unit and a high degree of political centralization do not necessarily go together with great density."

Population density provides a poor indication of political centralization in Africa. Population density might also be a poor proxy for economic conditions in the pre-colonial era. Acemoglu et al. (2002) provide several reservations regarding the use of population density as a proxy for pre-modern conditions. For example, the empirical relationship between population density and income per capita is weaker than that between urbanization rates and economic conditions. For instance, there is no cross-sectional association between population density and income per capita today. This means that differences in population density during the pre-colonial era cannot be translated into differences in income per capita. In contrast, the association between urbanization and income per capita is strong and robust even today (Acemoglu et al., 2002, Table 2). The use of population density as an indicator of pre-colonial conditions is based on a Malthusian logic where population density reflects differences in technologies (Ashraf and Galor, 2011). However, as Acemoglu et al. (2002, 1243) noted, while the Malthusian model suggests that higher densities may reflect higher income per capita, "the main thrust of Malthus' work was how a higher than equilibrium level of population increases death rates and reduces birth rates to correct itself. A high population density could therefore be reflecting an 'excess' of population, causing low income per capita." Overall, while there is no single proxy for economic conditions in the past, population density

Table A.1
OLS correlation between large or impressive structures and population density.

	All societies (1)	Sub-Saharan Africa (2)	Non-African societies (3)	Eurasia (4)	Tropical America (5)	Tropical America (no empires) (6)
I. Dependent variable: Large or impressive structures						
Population density	0.22***	−0.29	0.28***	0.23***	0.37***	0.30*
	(0.05)	(0.23)	(0.05)	(0.07)	(0.12)	(0.15)
R^2	0.20	0.14	0.30	0.27	0.35	0.29
II. Dependent variable: Religious, military, or industrial structures only						
Population density	0.06***	−0.05	0.08***	0.07***	0.13***	0.10**
	(0.01)	(0.05)	(0.01)	(0.02)	(0.03)	(0.04)
R^2	0.12	0.10	0.16	0.13	0.35	0.26
N. Obs.	184	32	152	118	34	32

Note.– No empires means that the Incas and Aztecs have been excluded. All specifications include a (non-reported) control for pre- and proto-agricultural societies. In parentheses are robust standard errors. ***, ** and * denote statistical significance at the 1, 5, and 10 percent level.

might not provide a reliable indicator of past economic development in economies with similar technological and geographic conditions.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.worlddev.2022.106086.

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review, 91*, 1369–1401.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distributions. *Quarterly Journal of Economics, 107*, 1231–1294.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2003). An African Success Story: Botswana. In D. Rodrik (Ed.), *Search of Prosperity: Analytic Narratives on Economic Growth* (pp. 80–122). Princeton University Press.

Allen, R. (2011). *Global Economic History: A Very Short Introduction*. Oxford University Press.

Alsan, M. (2015). The Effect of the TseTse Fly on African Development. *American Economic Review, 105*, 382–410.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ashraf, Q., & Galor, O. (2011). Dynamics and Stagnation in the Malthusian Epoch. *American Economic Review, 101*, 2003–2041.

Ashraf, Q., & Galor, O. (2013). The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *American Economic Review, 103*, 1–46.

Austen, R. A. (1987). *African Economic History: Internal Development and External Dependency*. James Currey.

Austin, G. (2008). The 'Rversal of Fortune' Thesis and the Compression of History: Perspectives from African and Comparative Economic History. *Journal of International Development, 20*, 996–1027.

Azevado, M. J. (1980) Epidemic Disease among the Sara of Southern Chad, 1890-1940, in Hartwig, G.W. and Patterson, K.D., Disease in African History, Duke University Press, 119-152.

Barras, R. (2016). *A Wealth of Buildings: Marking the Rhythm of English History*. Palgrave Macmillan.

Bates, R. H., Coatsworth, J. H., & Williamson, J. G. (2007). Lost Decades: Postindependence Performance in Latin America and Africa. *Journal of Economic History, 67*, 917–943.

Bhattacharyya, S. (2009). Root Causes of African Underdevelopment. *Journal of African Economies, 18*, 745–780.

Biraben, J.-N. (1979). Essai sur L'évolution du Nombre des Hommes. *Population, 34*, 13–25.

Birchenall, J. A. (2014) "A Theory of Disease and Development, UCSB Mimeo.

Bockstette, V., Chanda, A., & Putterman, L. (2002). States and Markets: The Advantage of an Early Start. *Journal of Economic Growth, 7*, 347–369.

Bloom, D. E., & Sachs, J. (1998). Geography, Demography, and Economic Growth in Africa, Brookings Papers on Economic Activity. *Economic Studies Program, 29*, 207–296.

Cashdan, E. (2001). Ethnic Diversity and its Environmental Determinants: Effects of Climate. *Pathogens, and Habitat Diversity, American Anthropologist, 103*, 968–991.

Cervellati, M., Esposito, E., & Sunde, U. (2017). Long-Term Exposure to Malaria and Development: Disaggregate Evidence for Contemporaneous Africa. *Journal of Demographic Economics, 83*, 129–148.

Cervellati, M., Sunde, U., & Valmori, S. (2017). Pathogens, Weather Shocks, and Civil Conflicts. *Economic Journal, 127*, 2581–2616.

Chandler, T. (1987). *Four Thousand Years of Urban Growth: An Historical Census*. St. Davis University Press.

Collier, P., & Gunning, J. W. (1999). Explaining African Economic Performance. *Journal of Economic Literature, 37*, 64–111.

Comin, D. A., Easterly, B., & Gong, E. (2010). Was the Wealth of Nations Determined in 1000 B.C.? *American Economic Journal: Macroeconomics, 2*, 65–97.

Connah, G. (2001). *African Civilizations: An Archaeological Perspective*. Cambridge University Press.

Coquery-Vidrovitch, C. (2011). *The History of African Cities South of the Sahara: From the Origins to Colonization*. Wiener: Princeton.

Crosby, A. W. (2003). *The Columbian Exchange: Biological and Cultural Consequences of 1492*. Greenwood Publishing Group.

Curtin, P., Feierman, S., Thompson, L., & Vansina, J. (1995). *African History: From Earliest Times to Independence*. New York: Longman.

Davidson, B. (1959). *The Lost Cities of Africa*. Little, Brown and Company.

Depetris-Chauvin, E., & Weil, D. N. (2018). Malaria and Early African Development: Evidence from the Sickle Cell Trait. *Economic Journal, 128*, 1207–1234.

Diamond, J. (1997). *Guns, Germs and Steel*. Vintage: The Fates of Human Societies.

Diamond, J., & Panosian, C. (2006). Continental Asymmetry of the Origins of Human Infectious Diseases. In P. Hämäläinen (Ed.), *When Disease Makes History: Epidemics and Great Historical Turning Points* (pp. 17–44). Helsinki Univ. Press.

Easterly, W., & Levine, R. (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics, 112*, 1203–1250.

Ehret, C. (1998) An African Classical Age: Eastern and Southern Africa in World History, 1,000 B.C. to A.D. 400, University of Virginia Press.

Englebert, P. (2000). *State Legitimacy and Development in Africa*. Lynne Rienner Publishers.

Fenske, J. (2014). Ecology, Trade, And States In Pre-Colonial Africa. *Journal of the European Economic Association, 12*(3), 612–640.

Fincher, C. L., & Thornhill, R. (2008). A Parasite-driven Wedge: Infectious Diseases may Explain Language and other Biodiversity. *Oikos, 117*(1289), 1297.

Fogel, R. (1994). Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy. *American Economic Review, 84*, 369–395.

Fortes, M., & Evans-Pritchard, E. E. (1940). *African Political Systems*. Oxford University Press.

Gennaioli, N., & Rainer, I. (2007). The Modern Impact of Precolonial Centralization in Africa. *Journal of Economic Growth, 12*, 185–234.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do Institutions Cause Growth? *Journal of Economic Growth, 9*, 271–303.

Heldring, L. and Robinson, J. A. (2012) Colonialism and Economic Development in Africa, NBER Working Paper No. 18566.

Herbst, J. (2000). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton University Press.

Hoeppli, R. (1969) "Parasitic Diseases in Africa and the Western Hemisphere: Early Documentation and Transmission by the Slave Trade, Acta Tropica, Supplementum No. 10.

Hopkins, A. G. (1973). *An Economic History of West Africa*. Columbia University Press.

Kamarck, A. (1967). *The Economics of African Development*. Praeger Publishers.

Kurzban, R., & Leary, M. R. (2001). Evolutionary Origins of Stigmatization: the Functions of Social Exclusion. *Psychological Bulletin, 127*, 187–208.

Hull, R. W. (1976). *African Cities and Towns before the European Conquest*. W.W. Norton.

Low, B. S. (1990). Marriage Systems and Pathogen Stress in Human Societies. *American Zoologist, 30*, 325–339.

Maddison, A. (2001). *The World Economy: A Millennial Perspective*. OECD Development Centre.

Mann, C. C. (2005). *1491: New Revelations of the Americas Before Columbus*. Alfred A. Knopf.

McEvedy, C. and Jones, R. (1985) Atlas of World Population History, Penguin Books.

McNeill, W. H. (1976). *Plagues and People*. Anchor Books.

Michalopoulos, S. (2012). The Origins of Ethnolinguistic Diversity. *American Economic Review, 102*, 1508–1539.

Michalopoulos, S., & Papaioannou, E. (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica, 81*, 113–152.

Michalopoulos, S. and Papaioannou, E. (2013). The Long-Run Effects of the Scramble for Africa, NBER Working Paper No. 17620.

Murdock, G. P. (1967). *Ethnographic Atlas*. University of Pittsburgh Press.
Murdock, G. P., & White, D. R. (1969). Standard Cross-Cultural Sample. *Ethnology, 9*, 329–369. Updates are available online at http://repositories.cdlib.org/imbs/socdyn/wp/Standard_Cross-Cultural_Sample.
Nunn, N. (2008). The Long-Term Effects of Africa's Slave Trades. *Quarterly Journal of Economics, 123*, 139–176.
Nunn, N., & Wantchekon, L. (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review, 101*, 3221–3252.
Osafo-Kwaako, P., & Robinson, J. (2013). Political Centralization in pre-Colonial Africa. *Journal of Comparative Economics, 41*, 6–21.
Patterson, O. (1982). *Slavery and Social Death: A Comparative Study*. Harvard University Press.
Price-Smith, A. T. (2008). *Contagion and Chaos: Disease, Ecology, and National Security in the Era of Globalization*. MIT Press.
Pryor, F. L. (2005). *Economic Systems of Foraging, Agricultural, and Industrial Societies*. Cambridge University Press.
Rosen, W. (2007). *Justinian's Flea: Plague, Empire, and the Birth of Europe*. Penguin.
Smith, B. D. (1995). *The Emergence of Agriculture*. Scientific American Library.
Spolaore, E., & Wacziarg, R. (2009). The Diffusion of Development. *Quarterly Journal of Economics, 124*, 469–529.
Vengroff, R. (1976). Population Density and State Formation in Africa. *African Studies Review, 19*, 67–74.
Weber, E. (1976). *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914*. Stanford University Press.
Wolfe, N. D., Dunavan, C. P., & Diamond, J. (2007). Origins of Major Human Infectious Diseases. *Nature, 447*, 279–283.
Zinsser, H. (1935). *Rats, Lice and History*. Little, Brown and Company.