The Effect of the TseTse Fly on African Development (Job Market Paper)

Marcella Alsan[*][†]

October 29, 2012

Abstract

The TseTse fly is unique to the African continent and transmits a parasite harmful to humans and lethal to livestock. This paper tests the hypothesis that the presence of the TseTse reduced the ability of Africans to generate an agricultural surplus historically by limiting the use of domesticated animals and inhibiting the adoption of animal-powered technologies. To identify the effects of the fly, a TseTse suitability index (TSI) is created using insect physiology to model insect population dynamics. African tribes inhabiting TseTse-suitable areas were less likely to use draft animals and the plow, more likely to practice shifting cultivation and indigenous slavery, and had a lower population density in 1700. As a placebo test, the TSI is constructed worldwide and does not have similar explanatory power outside of Africa, where the fly does not exist. Current economic performance is affected by the TseTse through its effect on precolonial institutions.

Keywords: disease environment, agricultural productivity, institutions

[*]I am grateful to David Cutler, Paul Farmer, Claudia Goldin, Michael Kremer and Nathan Nunn for encouragement and detailed feedback. For additional comments I thank Alberto Alesina, Robert Bates, Hoyt Bleakley, Melissa Dell, Stanley Engerman, Daniel Fetter, James Feigenbaum, Erica Field, Joshua Gottlieb, Edward Glaeser, Richard Hornbeck, Lawrence Katz, Orlando Patterson, James Robinson, Dana Rotz and participants at seminars at Harvard University, EconCon 2012 and Massachusetts General Hospital. For assistance with FAO data and GIS I thank Giuliano Cecchi, Rafaelle Mattioli, William Wint and Jeff Blossom. Luca Maini provided excellent research assistance. Funding provided by NIH T32 AI 007433, Harvard Global Health Institute *Fellowship for Neglected Tropical Diseases* and the Harvard Department of Economics *Lab for Economic Analysis and Policy.* All errors are my own.

[†]Correspondence: National Bureau of Economic Research 1050 Massachusetts Avenue, Suite 418, Cambridge, MA 02138, U.S.A. (e-mail: malsan@fas.harvard.edu; website: http://www.people.fas.harvard.edu/~malsan/).

I Introduction

Many scholars suggest Africa's historically low population density, or its relative land abundance, played a pivotal role in shaping its development (Hopkins, 1973; Iliffe, 1995; Fenske, 2009; Herbst, 2000). This view maintains abundant arable land weakened state development by hindering the ability to broadcast power over sparsely settled territories. Moreoever, since labor was scarce and shadow wages high, household and slave labor substituted for wage labor. But why was land in historical Africa relatively abundant? The anthropologist Jack Goody (1971) argued agricultural technologies, such as the plow, that were used to improve food production in much of the rest of the Old World were slow to diffuse in Africa. However, an important and unique feature of African ecology that could have affected its ability to adopt technologies, agricultural productivity, population density and institutional development that has yet to be examined empirically is the TseTse fly.

Economists, historians and biologists have debated the role of the TseTse on African development. By circumscribing the use of domesticated animals as a source of draft power, and precluding the adoption of technologies complementary to draft power, the TseTse has been hypothesized to have hindered the ability of Africans to generate an agricultural surplus and easily transport goods overland. The entomologist T.A.M. Nash (1969, p. 31) writes, "It seems reasonable to suppose that for hundreds of years tsetse dictated that the economy of the African should be based on the hoe and the head-load, a dictatorship which he is now being freed by the petrol engine and the railway locomotive." Others have expressed skepticism that the TseTse could explain why African technology lagged behind Eurasia (Chaves et al., 2010).

This paper is the first to investigate whether the TseTse fly affected Africa's precolonial agricultural technologies, population density and institutions. The TseTse (*Glossina* species) is only found in Africa.[1] The fly feeds strictly on vertebrate blood and transmits

[1] The TseTse is a prehistoric species that likely predates continental separation 100 million years ago (Krafsur, 2009). Climate changes and glaciations are believed to have isolated the TseTse in Africa a very long time ago (Lambrecht, 1964, p. 2).

Trypanosomiasis, a parasite causing sleeping sickness in humans and *nagana* in domesticated animals.[2] Livestock tend to be more affected than people since there are more types of trypanosomes that can infect them and the fly preferentially feeds on nonhuman animal hosts (Leak, 1999; Owaga, 1985; Vale et al., 1986; Vale, 1974).[3]

The effects of the TseTse are examined within a model of agricultural production. Capital in this setting refers to livestock. TseTse-transmitted diseases are modeled as productivity parameters impacting factor inputs differentially, affecting domesticated animals more than humans, and not affecting land. The model predicts that tribal agricultural production becomes less capital intensive and the land-labor ratio rises as TseTse increases. These conditions hold so long as factor inputs are substitutable to some degree. Consistent with a Malthusian equilibrium, living standards are assumed to be constant. The size of the maximum agricultural surplus, which is declining in the TseTse, thereby determines the size of the equilibrium urban population.

Crucial for identification of the impact of the TseTse is its specific, nonlinear temperature and humidity requirements for viability. These physiological relationships have been elucidated through controlled laboratory experimentation on the fly (Bursell, 1960; Mellanby, 1937; Rajagopal and Bursell, 1965; Terblanche et al., 2008). The exact functional forms relating TseTse birth and death rates to climate are derived from the experimental data. Using insect population growth modeling, gridded climate data and geospatial software, the *potential* steady state TseTse population can be calculated. The TseTse suitability index (TSI)

[2] The TseTse is unique to Africa and TseTse-transmitted *Trypanosomiasis* is the subject of this paper. *T. cruzi* is in South and Central America and causes Chagas disease in humans. Three forms of trypanosomes causing disease in domesticated animals (*T. equiperdum, T. evansi* and *T. vivax)* have spread beyond Africa. *T. equiperdum* is a sexually transmitted infection of horses and will not be discussed further. *T. evansi* and *T. vivax* are believed to have been spread during the process of European colonization (FAO, 1998, p. 137). Animal trypanosomiasis outside of Africa was not as virulent as within Africa since it lacked a specialized vector for transmission (e.g. the TseTse) and a large reservoir population of immune wild game. Further background on the biology is provided in Section II.

[3] European explorers were convinced that the TseTse did not harm humans: "During my hunting excursions along the Teoge, I encountered the most extraordinary of insects, the Tsetse. Among the several scourges to which the traveller is subjected in the South African wilderness, one of the greatest is this insect; not, it is true, as to the wayfarer's own person, for he himself escapes almost unscathed, but as regards the horses and cattle" (Anderson and Fragmont, 1857, p. 488-489). Livingstone (1857, p. 80-81) remarked, "A most remarkable feature in the bite of the Tsetse fly is its perfect harmlessness in man and wild animals."

is the standardized value (Z-score) of this steady state population. The TSI is then linked to precolonial anthropological observations on African agricultural practices, institutions and urbanization. A detailed description of the ethnographic data is provided below. The regressions compare highly TseTse suitable areas to less TseTse suitable areas within Africa controlling separately for the individual factors in the TSI and their first-order interaction (robustness tests also include higher order terms).[4]

The TseTse is estimated to have had substantial effects on precolonial Africa: a one standard deviation increase in the TSI is associated with a 22 percentage point decrease in the likelihood an African ethnic group had large domesticated animals, a nine percentage point decrease in intensive cultivation and a seven percentage point decrease in plow use. A one standard deviation increase in the TSI is correlated with a 46 percent reduction in population density in 1700. Motivated by the land abundance literature, two institutions are explored in this paper: political centralization and indigenous slavery. TseTse suitability is associated with a 12 percentage point increase in the likelihood an ethnic group used slaves and an eight percentage point decrease in the probability it was centralized.

One explanation for these findings is that places hospitable for the fly were inhospitable to intensive agriculture and dense human settlement. To examine this hypothesis, the TSI is also constructed for ethnic groups outside of Africa. The TSI predicts significant, detrimental effects on development only within Africa. Simulating African development under a lower burden of TseTse indicates that Africa could have supported a population density closer to that of Eurasia under the counterfactual disease environment.[5] This finding is consistent with archeological evidence of civilizations supported by advanced agricultural systems in places where the fly could not survive, such as Great Zimbabwe and the Ethiopian highlands.

[4] The reduced form relationship between the TSI and precolonial outcomes is emphasized since a reliable historical map of TseTse is unavailable. The observed TseTse map is from 1973. Focusing on the TSI as opposed to the observed fly distribution also addresses endogenity concerns related to more advanced, centralized tribes being better able to control the fly. Admittedly this is more of a concern for the modern era since insecticides and medical treatments to combat *Trypanosomiasis* were not available until the mid-20th century.

[5] Caution should be used in interpreting the simulation results as the endogenous response to reducing the burden of disease from the fly is not considered.

This paper concerns institutional origins and thus the fundamental determinants of prosperity. The TSI has a negative correlation with current economic performance. This relationship appears to be driven by the TseTse's effect on shaping historical institutions and is robust to including colonizer and legal origin fixed effects. This set of findings provides evidence in favor of the Engerman and Sokoloff (2000) point of view on how endowments, such as the disease environment, may shape institutions and thereby have persistent effects on development.

This paper is organized as follows. The next section provides an historical overview of the role of livestock in agriculture and the adverse effects of TseTse-transmitted *Trypanosomiasis* on domesticated animals. Section III describes the construction of the TSI and presents a model of agricultural production in the presence of the TseTse. Section IV describes the data and the empirical framework for the historical analysis. Section V reports these results as well as the results of the placebo test and counterfactual simulation. Section VI examines the effect of the TseTse fly on African development today and Section VII concludes.

II Historical Background

Communicable disease has often been explored as a cause of Africa's underdevelopment (Bloom and Sachs, 1998; Gallup and Sachs, 2001; Sachs and Malaney, 2002). Although the literature has investigated the role of human pathogens on economic performance, it is largely silent on the impact of veterinary disease.[6] This is peculiar given the role livestock played in agriculture and as a form of transport throughout history. Prior to mechanization, domesticated animals were an important input into the agricultural production process. Livestock improved yields by providing manure for fertilizer, made use of leguminous fodder and served as a source of draft power. The agricultural revolution in England relied upon domesticated animals (Allen, 1999; Overton, 1996). The process was summarized by a

[6] Livingstone (1857) mentions the TseTse 67 times in his work, *Missionary Travels and Researches in South Africa*; by contrast, malaria is mentioned six times.

farmer in 1795: "No dung–no turnips–no bullocks–no barley–no clover nor...wheat" [quoted in Overton, 1996].

The adoption of domesticated animals and associated technologies also affected culture.[7] Alesina, Giuliano, and Nunn (2011) show that historical plow use is predictive of present-day gender norms, suggesting the comparative advantage men have in upper body strength led to a reduction in the role women played in farming. Technical progress in agricultural techniques (plow use, harnessing) led to higher returns from animal power. Pierre Bonnassie (2009, p. 40) argues these technical changes aided the decline of slavery in Western Europe: "on the one hand, water power, and on the other, an increased return accrued from animal labour (the return was quintupled in the case of the horse) took the place of human energy (represented by slave labour) in the most laborious and common of tasks."[8] The plow has also been linked to the formation of civil society. According to Lynn White (1978), the heavy plow needed such immense power to be pulled that several peasants would pool together their oxen: thus laying the basis for medieval cooperative society.

Although livestock disease has beleaguered farmers worldwide, African Animal *Trypanosomiasis* was particularly detrimental. Jared Diamond (1997, p. 186) writes that "the spread southward of Fertile Crescent domestic animals through Africa was stopped or slowed by climate and disease, especially by trypanosome diseases carried by tsetse flies." *Nagana* infects all forms of ungulates, whereas most other pathogens (i.e., glanders, rinderpest, foot-root) have a predilection for a particular species (Brown and Gilfoyle, 2010).[9] Rapid antigenic

[7] Some have also viewed large domesticated animals, particularly the horse, as crucial for conquest (Maudlin, 2006) and power consolidation. For example, the southward expansion of ethnic groups using cavalry in Northern Nigeria (e.g., Nupe and Oyo) was believed to have been limited by the TseTse (Law, 1977). Law (1977, p. 198) writes, "Oyo operations against Dahomey were restricted to brief raids, as the cavalry could not operate during the rainy season (presumably because of the danger from trypanosomiasis) and were hampered by the problem of securing fodder for the horses. Consequently, although they could overrun the country and defeat any Dahomian army which stood and fought, they could not effect a complete and permanent conquest, so that in the end, Dahomey had to be left autonomous and tributary."

[8] The price of a horse imported into the Oyo empire was at least twice the price of a slave (Law, 1977, p. 185).

[9] Anthrax and Brucella are as broad as nagana in infectious scope, though Brucella is not fatal to adult animals and immunity to anthrax has been shown to occur naturally in livestock (Turnbull et al., 1992). Vaccines exist today for both Brucella and Anthrax but not for *nagana*.

variation, the switching of proteins on the surface of the trypanosome so that a host cannot recognize the infectious agent, thwarts the animal's immune response (Borst and Rudenko, 1994).

A parasite that quickly kills its host will itself become extinct. African Animal *Trypanosomiasis* was able to survive because wild game served as a reservoir population in which they circulated. In general, for a lethal parasitic disease to thrive it requires a host population that is immune (such as the big game of Africa), and an efficient vector (such as the TseTse) which has coevolved for its transmission (part of the life cycle for the parasite is completed in the hindgut, salivary gland or proboscis of the fly).[10]

Evidence of the hardship TseTse posed to the keeping of livestock comes from the colonial record. Commissioner H.H. Johnston (1894) described the TseTse as the "greatest curse" nature laid upon Africa and remarked the "value of the country would be centupled" in the absence of the fly. Early colonists often resorted to the less advanced technologies that characterized the region. When describing the peculiar location of roads in the Province of Ukamba, Sir A. Harding (1897, p. 51) reported:

[10]Several different features lead to immunity in wild game including the presence of a trypanosome lethal factor (Mulla and Rickman, 1988). After 8,000 years, certain breeds of cattle (so called "trypanotolerant" breeds such as the N'dama) have been noted to have reduced susceptibility to *Trypanosomiasis*, though will succumb with a high enough parasite load (Murray et al., 1990).

As mentioned in footnote 2, there are animal trypanosomes outside of Africa, specifically *T. vivax* and *T. evansi* (also known as surra). Surra is a disease primarily affecting camels and horses.

The impact of animal-trypanosomiasis (*T. vivax* and *T. evansi)* outside of Africa differs from that within Africa in that there are no specialized vectors for transmission and it is therefore much less efficient. Specifically, biting flies act as needles that mechanically transmit the parasite from host to host. However, in order for trasmission to be sucessful the interval between fly feeds has to be very short as the trypansome dies when blood dries. Such a short interval between feeds typically only occurs when flies are interrupted while taking a bloodmeal. The TseTse can not only mechanically transmit but can also cyclically transmit, which means the trypanosome parasite actually multiplies in the TseTse gut and infective forms are stored in the salivary gland. The flies are infected for the rest of their lives. Whenever the TseTse takes a bloodmeal, the fly emits salivary anticoagulant that helps it feed and the trypanosomes are injected into the host blood stream along with the anticoagulant. This translates into a much higher transmission rate (rate at which an initial infected animal/human can give rise to new infections) and more subspecies of trypanosomes that have developed to survive within the TseTse and can harm many different species of large domesticated animals within Africa (FAO, 1998, p. 6). For example, in Africa there is not only the trypanosomes mentioned above but also those that rely on the TseTse to complete part of their replication cycle: *T. congolense, T. simiae, T. godfreyi, T. brucei brucei, T. brucei rhodesiense* and *T. brucei gambiense.* Finally in Africa, wild game are immune to the trypanosome parasites and thus serve as a reservoir of the disease. This does not seem to be the case outside of Africa (Luckins and Dwinger, 2004; FAO, 1998, p.140), meaning that the disease cannot be as acutely lethal in domesticated animals or else it will not be propagated.

the presence of Tsetse-fly preclude the animal transport by carts, which in the interior is the great incentive for road-making. In Witu, for instance, where the Administration built a good carriage-road from the capital to the Port of Mkunumbi, the bullocks employed for the waggons on it all died, and the old wretched system of human porterage has still to be resorted to for transport

The problem was especially acute given that the TseTse's ecological niche was in the most fertile areas in Africa. As can be seen in Figure I, the suitability of land for agriculture and the TSI are positively correlated. It is remarkable that no country within Africa is both inhospitable for the fly and yet highly suitable for agriculture. European colonists noted excellent pasture without evidence of cattle and remarked that railways would be needed to penetrate the African interior.[11]

III Conceptual Framework

III.1 Constructing the TseTse Suitability Index (TSI)

III.1.1 Insect Physiology

There are two main reasons to develop the TSI. First, a reliable map of the precolonial TseTse distribution is not available. Using a modern map of fly distribution may lead to skewed results since climate change has altered the location of the fly.[12] The TSI can be constructed with historical climate data thereby mitigating this concern. Second, a measure of potential as opposed to observed TseTse, especially for modern outcomes, purges the estimates of bias arising from states with stronger institutions being able to better control the fly.

[11] Bryden (1900, p. 288-230) described the following: "Some few years since, when Rhodesia was first being opened up, an attempt was made to run a line of coaches from the Pungwe River, on the East Coast, towards Mashonaland. Splendid American coaches were imported and plenty of fine horses and mules got out. But the Tse-tse-fly beat this spirited enterprise. The horses and mules died, the attempt had to be abandoned and the coaches still lie rotting in the wilderness. In its turn, however, I am glad to say the deadly Tse-tse has been overcome by the railway."

[12] Moore and Messina (2010) model how climate change has affected the TseTse distribution within Kenya. Paleoclimatic data from tree cores and ice rings (Mann et al., 2008) demonstrate changes in Africa's temperature, especially in the latter half of the 20th century (Appendix Figure A.I) which might affect fly distribution.

Other researchers have developed suitability indices for the TseTse (Rogers and Randolph, 1986; Wint and Rogers, 2000); however, these are not adequate for the purposes of this paper since they use inputs that could be considered endogenous, such as the distribution of cattle. Important for the approach herein is that results from controlled laboratory experiments are used to identify abiotic climate factors that affect insect physiology and thus determine the reproductive success of the fly. In general, insects are particularly sensitive to changes in the conditions of their environment due to their large surface area to volume ratio, which affects thermoregulation and water balance (Schowalter, 2011, p. 31). The TseTse is distinguished from most other insects by its method of reproduction, known as adenotrophic viviparity, which results in a low number of offspring production per adult female. After ovulation, the egg develops in the uterus of the female until reaching the third instar larva stage (Jackson, 1949). Larva are deposited onto the soil where they quickly burrow and encapsulate to form pupa. The pupa have fat stores that last for the approximately month-long period needed to metamorphosize into adult flies (Glasgow, 1963).

Pupa metabolism is highly temperature dependent. Extremes of temperature lead the pupa to metabolize lipid reserves too quickly or slowly, thereby exhausting energy stores before metamorphosis is complete. Figure II panel (A), adapted from Bursell (1960) and Rajagopal and Bursell (1965), demonstrates pupa survival as a function of temperature. The dots represent data from lab experiments and a line representing the best quadratic fit is shown (the exact equation corresponding to this line and all equations used to develop the TSI are contained in Appendix Table B.1). At temperatures below 22 degrees Celsius, adult flies lapse into a chill coma (Terblanche et al., 2008). TseTse flies, especially at the young, tenereal stage, are susceptible to dessication at low humidities (Mellanby, 1937). The saturation deficit is a combination of humidity and temperature–it captures the difficulty organisms have in transpiration in hot and humid weather. Figure II panel (B) shows that at lower humidities (higher saturation deficits) fly mortality increases.[13] These physiological

[13] Teneral comes from *tener,* the Latin word for tender, and describes the softness of the fly body and immaturity of its thoracic musculature prior to the first bloodmeal. Insects are particularly vulnerable to

relationships are combined in a closed population growth model of the fly.

III.1.2 Population Growth Model of the TseTse Fly

Let B represent the birth rate, which is temperature dependent, and M represent the mortality of adult flies from dessication.[14] In the absence of intraspecies competition, the TseTse has a constant, net nonnegative growth rate, λ, defined as:

$$\lambda = B - M. \tag{1}$$

The equations for B and M are found by fitting curves to the data points from laboratory experiments, as in Figure II panels (A) and (B). Substituting in climate data from Africa, B typically exceeds M and there is no steady state. Therefore, a second form of mortality that is not due to climate, but rather attributable to competition among flies, is introduced. This is known as density dependent mortality, Δ, and can be expressed as:[15]

$$\Delta = \phi \left(N\right)^{\psi}, \tag{2}$$

It is now straightforward to derive the steady state equilibrium population (the value, N^*, such that $\lambda - \Delta = 0$):

$$N^* = \left(\frac{\lambda}{\phi}\right)^{\frac{1}{\psi}}. \tag{3}$$

Intuitively, the steady state fly population will be bigger the larger the difference between the birth and death rates. Implementing this model requires a choice of parameter values. Fortunately, May et al. (1974) have studied the stability conditions of a similar model,

dessication during molts (Schowalter, 2011).

[14] The birth rate is proxied for by the pupa cohort survival rate. Each adult female fly has an average of one pupa every ten days. The adult mortality rate is also for a ten day period. Predation could affect the mortality rate, but tends to be most relevant at intermediate population densities, when density is sufficient for a high rate of encounter but prior to predator satiation (Schowalter, 2011, p. 143). It is not known how relevant predation is in controlling the TseTse population (FAO, 1982).

[15] TseTse flies have been observed to exhibit density-dependent mortality due to competition for bloodmeals and oviposition sites (Rogers and Randolph, 1984).

which rest upon ψ, and have determined the steady state oscillates about its equilibrium for parameter values $2 > \psi > 0$. For the TseTse Suitability Index, $\phi = 0.025$ and $\psi = 1.25$, though a sensitivity analysis (Appendix Table A.II) shows results are robust over the entire range.[16]

The historical analysis uses climate data from the 20th century reanalysis version 2 (20CRv2), a retrospective analysis produced by the National Oceanic and Atmospheric Administration's (NOAA) Earth Science Research Laboratory Physical Sciences Division in collaboration with the University of Colorado CIRES Climate Diagnostics Center (Compo et al., 2011). 20CRv2 is the earliest climate data set available that covers Africa and includes the indicators necessary for constructing the TSI.

The 20CRv2 data set provides the first estimates of daily global tropospheric measures from the late 19th century to the present at a 2° spatial resolution. Earlier reanalyses had difficulty recreating weather for historical periods since upper-air observations typically increased over time and data assimilation methods could not adequately adjust for variation in observation networks. The 20CRv2 uses advanced assimilation methods (the Kalman Ensemble technique) to develop a more accurate representation of late 19th century weather. For the historical analysis, the average of daily mean temperature and relative humidity for the first year of available data (1871) are used.[17]

The steady state TseTse population is computed for each ethnic group using equation (3). A 3-dimensional figure of the relationship between climate factors and the steady state population is shown in Figure II panel (C). The TSI joined with a continental map of African ethnic groups is shown in Figure III panel (A). The observed TseTse distribution in 1973 is shown in Figure III panel (B) (Ford and Katondo, 1977). The correlation between the historical TSI and the fraction of a group's area where the TseTse is observed is 0.62. A scatter plot of the relationship between the fraction of area TseTse-infested and the historical

[16] This particular value was chosen based on the results of experimentation with *Aedes aegypti* ($\psi = 0.922 \pm 0.47$) (Legros et al., 2009, p.14).

[17] Results (available on request) demonstrate that using a longer period of time or different years of climate data deliver similar results.

TSI is shown in Appendix Figure A.II.

III.2 A Model of Precolonial Agricultural Production

The analysis focuses on the effect of trypanosome parasites transmitted by the TseTse on historical African development. Understanding how the TseTse affected the health of humans and domesticated animals is relatively straight-forward. To illustrate how the disease might have affected the ability of African ethnic groups to generate an agricultural surplus and support a large, nonagricultural population, a simple partial equilibrium model of agricultural production is introduced. TseTse-transmitted diseases are represented as productivity parameters that differentially affect the three inputs: capital (draft animals and complementary technologies such as the plow), labor and land.

Consider a chief of an ethnic group seeking to maximize agricultural surplus (π), denominated in calories, by choosing the optimal number of draft animals (D) agricultural laborers (L_A) and hectares of land (T). Tsetse is modeled as an endowment reducing the productivity or survival of agricultural labor to (δ) and of animals to ($\delta\theta$), where $0 < \theta, \delta < 1$. This set of assumptions highlights the diminished productivity of livestock and agricultural labor in the presence of the TseTse and the enhanced susceptibility of draft animals to disease. Production of calories (y) is represented by a decreasing-returns-to-scale CES technology:

$$y = \left[(\delta\theta D)^\rho + (\delta L_A)^\rho + T^\rho\right]^{\frac{\gamma}{\rho}}. \quad (4)$$

Let w be the wage in calories needed for an agricultural worker to perform his tasks. Let r represent the calories a draft animal needs to sustain it and z represent the part of y (i.e., grain) that must be sown into the ground. The objective function of the chief can therefore be written as:

$$\max_{L_A, D, T} \pi = y - rD - wL_A - zT. \quad (5)$$

Define $\frac{D}{L_A}$ as capital intensity and $\frac{T}{L_A}$ as the land-labor ratio. The first order conditions from surplus optimization yield:

$$\frac{D^*}{L_A^*} = \theta^{\frac{\rho}{1-\rho}} \left(\frac{w}{r}\right)^{\frac{1}{1-\rho}} \tag{6}$$

$$\frac{T^*}{L_A^*} = \delta^{\frac{\rho}{\rho-1}} \left(\frac{w}{z}\right)^{\frac{1}{1-\rho}}. \tag{7}$$

Differentiating the above expressions with respect to the TseTse productivity parameters demonstrates that agricultural production becomes less capital intensive and has a higher land-labor ratio as the burden of TseTse-transmitted diseases increase.[18] These conditions hold if and only if the elasticity of substitution, $\sigma = 1/(1-\rho)$, is greater than one, implying the inputs are substitutable in the production process.[19]

The vector of inputs that optimizes surplus can be substituted into the production function to determine the effect of TseTse-related diseases on the maximal size of the surplus (π^*). This profit function identifies the surplus in terms of the exogenous parameters of the model $(w, r, z, \theta, \delta)$:

$$\pi^* = \left[\frac{1}{\gamma}^{\frac{\gamma}{\gamma-1}} - \frac{1}{\gamma}^{\frac{1}{\gamma-1}}\right] \left[\left(\frac{r}{\theta\delta}\right)^{1-\sigma} + \left(\frac{w}{\delta}\right)^{1-\sigma} + z^{1-\sigma}\right]^{\frac{\gamma}{(\gamma-1)(1-\sigma)}}. \tag{8}$$

Intuitively, the surplus is decreasing in the calories required by domesticated animals (r) and the agricultural workforce (w). The surplus is also declining in the burden of TseTse-

[18] The precolonial ethnographic data does not include measures of capital stock or land per agricultural worker, but does have binary variables for the presence or absence of large domesticated animals and intensive agriculture at the ethnic group level. The TSI does not distinguish between human and livestock disease since it is based upon physiological characteristics of the fly. The diseases (*nagana* or sleeping sickness) are determined by the parasite the fly transmits.

[19] Capital and labor seem obvious substitutes. The notion that land was a substitute for capital or labor is consistent with the historical record. Intensive agriculture (a short-fallow system) required the plow to remove grass roots and domesticated animals for their draft power and manure to replete the soil (Boserup, 1966, p.16). In the absence of these inputs, societies continued to rely on the older method of slash-and-burn farming (a long-fallow system). Intensively farming the land also required more labor than shifting farming because of the additional soil preparation tasks that it entailed (Boserup, 1966, p.44). It is possible to relax the assumption of a common elasticity of substitution among all three parameters by using a nested CES. This delivers similar comparative statics (Equations (6) and (7)) but does not permit a closed form solution to the value function (in Equation (8)).

transmitted diseases. Inspection of the expression for π^* shows why this is so–diminishing survival in the presence of the TseTse acts to inflate the cost of the animate inputs, which are inversely related to the size of the surplus.

Galor and Ashraf (2011) provide empirical support for the hypothesis that the preindustrial world was well characterized by Malthusian dynamics. In a Malthusian equilibrium, the size of the population adjusts (for example, through famine or the passion between the sexes) so that living standards are maintained at a constant, near subsistence level. Using this definition of equilibrium it is straightforward to derive the size of the nonagricultural population. The nonagricultural population include the artisans, religious and military not devoted entirely to food production and are referred to simply as the urban population, L_U. Define $\bar{c}$ as the constant number of calories that each nonagricultural individual consumes in the Malthusian equilibrium. Then the number of nonagricultural people the ethnic group can support will be the maximal number of calories it can produce divided by the number of calories per nonagricultural person: $L_U^* = \frac{\pi^*}{\bar{c}}$.[20] By reducing the size of the agricultural surplus, the TseTse is predicted to reduce the size of its equilibrium urban population.

III.3 Labor Coercion in a General Equilibrium Framework

The model above predicted land-labor ratios were increasing in the TseTse. Nieboer (1900) and Domar (1970) both observed that high land-labor ratios were positively correlated with labor coercion in the historical record. Acemoglu and Wolitzky (2011) use a principal agent framework embedded in a general equilibrium model to define the conditions under which this might occur. They identify two "equilibrium interactions." The first is called the labor demand effect: demand for labor rises as the output price increases making coercive tactics more attractive. The second interaction is the outside option effect: labor scarcity discourages coercion through this channel because it increases the marginal product of labor in other (unmodeled) sectors. Specifically, the laborer has an outside option of $\bar{u} = u(\tilde{L}) - \frac{\xi}{1-\xi}G$ where

[20] With free mobility between agricultural and nonagricultural sectors and in the absence of labor coercion, it would be natural to equate w and $\bar{c}$.

$u(\tilde{L})$ is the wage in the noncoercive sector, G is the average aggregate coercion in the economy and $\frac{\xi}{1-\xi}$ are the odds of being placed back into another coercive relationship. Whether labor scarcity encourages coercion is deduced from the difference between the increase in the price of output (labor demand effect) versus the increase in reservation utility (outside option effect). Since TseTse-transmitted diseases are predicted to decrease the ability of precolonial African ethnic groups to generate an agricultural surplus and support urban centers, the "outside option" of low-skill work in cities would likely have been negligible. Furthermore, the demand for agricultural labor (at least relative to capital) would have been high.[21] Taken together, it seems plausible that high land-labor ratios encouraged by the TseTse would be associated with labor coercion.

IV Data and Empirical Framework: Historical Analysis

Motivated by the conceptual framework presented above, the empirical analysis focuses on how the TseTse affected agricultural practices (plow and draft animal use, intensive cultivation), urbanization (population density and urban centers) and institutions (the practice of indigenous slavery and political centralization). Figure IV panel (A) provides a visual representation of the reduced form relationship between large domesticated animals and the TSI. Figure IV panel (B) is constructed by creating twenty equal sized (5 percentile point) bins of the TSI and plotting the mean probability of large domestic animal usage within each bin. The map and plot demonstrate that ethnic groups inhabiting parts of Africa that were less suitable for the fly heavily relied upon domesticated animals, whereas those located in places more suitable for the TseTse did not. The main estimating equation is presented below and further explores this within-Africa heterogeneity:

$$y_j = \alpha_0 + \delta TSI_j + \mathbf{X}_j'\Omega + \varepsilon_j, \tag{9}$$

[21] It can be shown that $\frac{\partial L^*}{\partial \delta} < 0$ if inputs are fairly substitutable ($\gamma < \rho$).

where y_j represents one of the precolonial outcomes associated with ethnic group j. The vector $\mathbf{X}'_j$ denotes the set of climate variables comprising the TSI (i.e., temperature and humidity) and their first order interaction. The identification strategy in this section can be seen graphically as the difference between Figure II panel (C), the nonlinear interaction of temperature and humidity that comprises the TSI, and panel (D), the first order interaction between temperature and humidity. It should be emphasized that similar results are obtained when including second order terms in climate (i.e., quadratic forms in temperature, humidity and temperature^2*humidity^2).

Most of the outcome variables are taken from the *Ethnographic Atlas*, a world-wide database that includes the historical features of 1263 ethnic groups, of which 533 are in Africa (Murdock, 1967).[22] The data are cross-sectional and are meant to capture the characteristics of ethnic groups prior to European settlement. The observations are coded by Murdock and summarize field work performed by anthropologists primarily during the 19th and early 20th century.

Historical gridded population data for the year 1700 are available through the History Database of the Global Environment (HYDE) version 3.0 (Goldewijk, 2005). This data set was constructed by collecting historical population estimates at the sub-national level. Hindcasting to the base year was performed by using regional growth rates if no other data sources were available. The authors cross-checked their results with McEvedy and Jones (1978) and arrive at similar estimates.[23] The location of cities with over 20,000 inhabitants in the year 1800 is included as an alternative measure of urbanization (Chandler, 1987).

The outcome data are spatially combined with Murdock's *Tribal Map of African Ethnicities* (Murdock, 1959b), which includes the location of 843 ethnic group areas. Ethnic groups from the *Ethnographic Atlas* are joined to the *Map* using the procedure described by

[22] Although there are 1265 ethnic groups originally, Chilcotin and Tokleau are entered twice with slightly different outcomes in the database and those groups are therefore dropped from the analysis.

[23] In personal communication with the authors of the data set, waterways were also used to interpolate population, though their code was not made available. Whether an ethnicity was landlocked or had access to a river enter as controls in the analysis.

Fenske (2009).[24] There is no map of boundaries for ethnic groups outside of Africa, a point returned to below.

$\mathbf{X}_j'$ includes other plausibly exogenous controls. Absolute latitude controls for different agro-ecological zones. Irrigation, trade and fishing, as well as an alternative form of transportation, would have been influenced by waterways, and access to such is also included in the analysis. Longitude captures differences in the Eastern and Western parts of the continent. Soil fertility and other environmental conditions might also affect aggregate agricultural productivity, and thus a summary measure of agricultural suitability, developed by the Food and Agricultural Organization (FAO), is included as a covariate.[25] *Plasmodium Falciparum,* has been singled out as an obstacle to growth in Africa due to its affect on human health (Bloom and Sachs, 1998; Gallup and Sachs, 2001). Others have argued that genetic mutations confer a tolerance to severe malaria and thus its effect may be overstated (Weil, 2010). The malaria ecology index by Kiszweski et al. (2004) is used to approximate the prevalence of different forms of malaria. Altitude reflects the privileged position of the African highlands–relatively free from insect vectors, easier to defend and with ample precipitation. A correlation matrix between the TSI and these geographic and climate features is shown in Appendix Figure A.III.

It is unlikely that each ethnic group can be thought of as an independent observation, given that many share a common cultural ancestry. This will not be entirely captured by standard errors that control for spatial correlation, since migratory patterns (such as the Bantu expansion) may place groups far apart despite a common lineage. The most popular way to deal with spatial correlation in the data is to use Conley's covariance matrix, a

[24] Since there is not a one-to-one match for all observations in the *Atlas* and those in the *Map*, Fenske developed an algorithm to join unmatched ethnic groups based on an alternative name, supergroup or location. 523 ethnic groups are matched this way–one outlier observation is dropped bringing the sample to 522. The "Matches" excel file is available on his website: `http://www.jamesfenske.com/`.

[25] The FAO methodology characterizes the climate, soil and terrain conditions relevant to agricultural production and compares these requirements with observed conditions to develop a global dataset of maximum potential crop yields under varying input levels. This index has been described in detail by Nunn and Qian (2011, p. 609-610). Similar to their analysis, suitability for rainfed crops is used in this empirical exercise to approximate historical agricultural conditions.

weighted average of spatial autocovariances, with the weights declining linearly to zero until a pre-specified cut point is reached (Conley, 1999). This would be inadequate in a setting where spatial and genealogical correlation are both at work. Errors are therefore clustered at the level of cultural provinces, which are groupings from Murdock's book, *Africa: Its People and Their Cultural History.* Reconstruction of the attributes of ethnic groups by Murdock is based on written and archeological records, linguistic evidence, common cultigens and the conservatism of certain features in societal organization (Murdock, 1959a, p. 42). These provinces capture both spatial and genealogical correlation. The sample of ethnic groups included in the analysis is shown in Appendix Figure A.IV, with shading to represent the 44 different clusters.[26]

V Historical Results

V.1 Agriculture

The TseTse prefer to feed on nonhuman animals and, unlike wild game, livestock are not immune. The presence of large domesticated animals at the ethnic group level is coded as a binary variable equal to one if the ethnic group uses bovines, camelids, or equines. Each cell in Table I reports the coefficient on the TSI from a separate regression using Equation (9). Table I column (1) controls only for the climate variables in the TSI. A one standard deviation increase in the TSI decreases the probability of an ethnic group possessing large domesticated animals by 21.6 percentage points. Moving across the columns, geographic and malaria controls are added. The point estimate remains stable, reducing concern for selection on unobservables (Altonji, Elder, and Taber, 2005). The preferred specification is reported in Table I column (3) and includes geographic, climate and malaria controls. A one standard deviation increase in the TSI is associated with a statistically significant, 21.9

[26]Note that 478 of 522 ethnicities are represented in the map and 44 ethnicities are joined to one of the represented 478 using the algorithm developed by Fenske (2009). Conley standard errors are used in a robustness check.

percentage point decrease in the probability an ethnic group possesses large domesticated animals, which is approximately equal to the simple difference between ethnic groups with and without access to a river.[27]

African agricultural technology may not have advanced in many places because of the TseTse. First, without draft animals, a plow is hard to use. Second, shifting agriculture is a labor saving technique, since the number of hours necessary to burn a forest and remove the stumps and rocks is far fewer then what would be needed to continuously farm the same plot (Boserup, 1966). Third, intensive agriculture requires fertilizer, since the soil is rapidly depleted of nutrients by repeated cultivation. Without animal dung, farmers would need the long fallow of shifting cultivation to allow time for the soil to replenish. The correlation between the TseTse and shifting agriculture is positive: a one standard deviation in the TSI decreases intensive agriculture by 8.7 percentage points. A one standard deviation increase in the TSI is also associated with a 6.7 percentage point decrease in aboriginal usage of the plow.

Agricultural practices and cultural norms are often intertwined. For example, Alesina, Giuliano, and Nunn (2011) find evidence in support of the Boserupian hypothesis that historical plow use led to a gendered division of labor in agriculture. Table I row (4) reports the results of a (0,1) variable indicating whether females perform the majority of agricultural tasks. The distribution of the fly is positively correlated with the participation of women in agriculture: a one standard deviation increase in the TSI is associated with a statistically significant 19.5 percentage point increase in the probability that females are participating heavily in agriculture.

A back of the envelope calculation demonstrates how the effects of the TseTse might have

[27] These results are only capturing the effect of the TseTse on the extensive margin. Even where draft animals were present, livestock might not have been able to graze on the best pasturage, or otherwise used as productively in the presence of the TseTse. For instance, employment of large livestock as pack animals would not have been feasible in the presence of intervening fly belts. Interaction of the TSI with a dummy for river has a negative and significant coefficient (-0.09 s.e. 0.04), indicating further substitution away from using large domesticated animals for transport in the presence of the fly. Further, milking livestock is negatively associated with TseTse suitability, suggesting an effect on the intensive margin (Appendix Table A.IV).

translated into relative land abundance. Shifting agriculture can produce up to 0.2 quintals of grain per hectare sustaining a maximum of 10 inhabitants per square kilometer (Mazoyer et al., 2006, p. 116) whereas intensively farming the land using animal powered technologies and fertilizer can support a maximum of 55 inhabitants per square kilometer (Mazoyer et al., 2006, p. 282). In the sample, the mean population density in 1700 in Africa is 2 inhabitants per square kilometer (Goldewijk, 2005, p. 354), suggesting that most ethnic groups were performing shifting (if any) agriculture. Imagine a small ethnic group with approximately 30,000 square kilometers of land (the average size in the sample) to allocate to either intensive or shifting agriculture. At baseline, assume none of the territory is allocated to intensive farming. A one standard deviation decrease in the TSI would increase intensive farming by 28 percent, increasing the average population density of the ethnic group by approximately 41 percent.

V.2 Urbanization

Both an agricultural surplus and transportation networks are important for urbanization (Bairoch, 1988). Two outcome variables from different data sources are used in the analysis to measure urbanization: the logarithm of population density (inhabitants per square kilometer) in 1700 from HYDE and the presence of an urban center in 1800 from Chandler (1987). A one standard deviation increase in the TSI is associated with a statistically significant 46 percent reduction in population density. Urbanization in precolonial Africa was rare, with only four percent of ethnic groups containing at least one city. A one standard deviation increase in TseTse suitability is associated with a significant three percentage point decrease in the probability of the ethnic group having an urban center.

V.3 Institutions

The measure of political centralization is constructed following Gennaioli and Rainer (2007) who use the jurisdictional hierarchy variable in the *Ethnographic Atlas*. This variable equals zero for groups lacking any form of centralized state, one for petty chiefdoms, two for large paramount chiefdoms/petty states and three or four for large states. An ethnic group is considered centralized if it has a value greater than one for the jurisdictional hierarchy variable. Consistent with the above results, there is a negative correlation between the TseTse and centralized states. A one standard deviation increase in the TSI decreases the probability of an ethnic group being classified as centralized by 8.2 percentage points.[28]

The entomologist J.P. Glasgow (1963, p. 3) conjectured that the practice of indigenous slavery and the presence of the TseTse were related: "Nearer the equator the use of draught or pack animals was impossible, and such trade as occurred depended on transport by human carriers. This circumstance, we may suppose, encouraged the growth of slavery, and the existence of slavery as a flourishing indigenous institution must have facilitated the export of African slaves." Using the ethnographic data there is empirical support for a positive correlation between the TseTse and the practice of indigenous slavery at the ethnic group level. A one standard deviation increase in the TSI is associated with a statistically significant 11.7 percentage point increase in the probability an African ethnic group used slave labor.[29]

V.4 Two-Stage Least-Squares Estimates

The most reliable map of observed TseTse was created in 1973 and displays only whether the fly is present or absent without regard for the intensity of infestation (Ford and Katondo, 1977). The TseTse distribution is believed to have changed over time due to elimination campaigns as well as climate change. The 1973 map likely suffers from measurement error– it is not the result of one large entomological survey but a compilation of many surveys at different points in time performed by various researchers using nonstandardized methodologies.

[28] Estimation with an ordered logit produces similar, statistically significant, results.

[29] Appendix Table II reports summary measures of the effect of the TseTse fly on agriculture, urbanization and institutions. The average effect size (AES) coefficients calculate the mean (standardized) effect of the TSI across various outcomes and is calculated following Kling et al. (2004). Seemingly unrelated regression is used to estimate the sample variance of the AES estimator. The results are reported for both the full sample and the Tropics subsample. In general, the findings reinforce those obtained when the outcomes are examined individually and are not sensitive to the threshold TSI used to define the control group.

Finally, there are multiple ways the TseTse could affect a complex outcome like population density. First it affects the ability to generate an agricultural surplus as mentioned above. But there are also direct effects of sleeping sickness on the survival of humans. Therefore, the exclusion restriction, in the very narrowly defined sense of what channel the TseTse operates (as opposed to whether the TseTse is affecting an outcome), will not hold. For all these reasons, the reduced form has been emphasized throughout the paper. However, there may still be interest in the first-stage, as a measure of the goodness of fit of the TSI with the observed TseTse distribution. Furthermore, classical measurement error in the map should be corrected by using instrumental variables.[30]

Two-stage least-squares is used to instrument for the fraction of an ethnic group's land that is TseTse-infested with the historical TSI. The results are reported in Table II. The IV estimates are larger than the corresponding OLS estimates consistent with measurement error. The first-stage F-statistic is approximately ten or above (it varies due to changes in the sample size by different outcomes).[31] Since the TSI is derived from the steady state fly population and the fraction is derived from the land area that is TseTse infested, the magnitudes in Table I and Table II are not directly comparable, though the results are consistent. For example, according to the two-stage least square estimates, a 0.1 increase in the fraction of land that is TseTse-infested leads to a 22 percentage point reduction in the probability the ethnic group used large domesticated animals, similar to Table I column (3).

V.5 Robustness

A series of robustness tests are performed. To reassure that the correlations reported above are not attributable to a tropical climate, the sample of ethnic groups is limited to those that are wholly contained between the Tropics of Capricorn and Cancer (see Table I, column (4)). All of the results remain statistically significant and economically relevant despite this

[30] Aigner (1973) showed that, even with a binary variable, measurement error would lead to attenuation bias.

[31] Estimates with a balanced sample produce similar results and can be found in Appendix Table A.V.

restriction.

The second major concern is that the TSI is picking up climate factors that are inhospitable to the keeping of livestock, use of the plow, or human settlement. Figure I suggests that this is not the case–agricultural suitability and TSI are *positively* correlated. This concern is also addressed below by showing that the TSI does not have the same predictive power outside of Africa. Another straightforward approach is to include more climate variables as controls. Since both the relationship between temperature and pupa survival as well as humidity and adult mortality are quadratic, second order terms in these climate variables are entered into the regression. Table III column (1) includes these first order and second order terms, as well as their higher order interactions. The point estimates (with the exception of plow use) remain almost identical to the baseline case. It appears unlikely that climate alone is driving the results.

A third concern is that the TseTse overlaps considerably with another vector-borne infectious disease with potentially deleterious effects on African development: malaria. The most efficient *Anopheles* mosquito in tropical Africa prefers human hosts over cattle (WHO, 2001). Thus, the biological mechanisms whereby malaria could influence outcomes such as the keeping of livestock and use of the plow is much weaker than for TseTse-transmitted *Trypanosomiasis*. Table I used the Kiszewski et al., (2004) malaria ecology index. Table III column (2) uses a different measure, the malaria distribution in 1900 (Hay et al., 2004). This measure is coarser than the malaria ecology index (with 5 as opposed to 35 categories) yet delivers similar results.

A fourth concern is that the TSI is constructed by cherry-picking parameter values. Two approaches allay these concerns. A different method to predict the fly distribution is employed. The climatic conditions for fly survival are taken from field research by Rogers and Randolph (1986) and are converted into a binary indicator of "optimal fly survival."[32] With

[32] Rogers and Randolph (1986) define the optimum as the joint condition that the temperature lie between 22 and 27 degrees Celsius and the saturation deficit lie between 6 and 14 mm Hg. This is not the preferred method since they rely on field observations (which may be influenced by human activity) instead of laboratory experiments.

the exception of the centralization outcome, the alternative TSI produces almost identical results to the model developed above (see Table III, column (3)). Second, a sensitivity analysis of parameter values is undertaken and results reported in Appendix Table A.III.

A fifth set of issues involves statistical inference and the choice of regression model. Conley standard errors are reported in Table IV column (4) with cutoff values of 5 degrees latitude and longitude. All results remain significant at conventional levels. Probit and negative binomial models, where appropriate, are employed instead of linear probability models in Table III column (5). Again, the results are broadly similar to Table I.

A sixth concern is with the *Ethnographic Atlas*. The Atlas is meant to capture, to the fullest extent possible, the attributes of African ethnic groups prior to European settlement. Clearly, trade between Africa and Eurasia was a source of influence before colonization. For the current study, the main threat is that the export slave trade may have encouraged indigenous slavery, discouraged centralization and reduced population. Thus, findings attributed to the TseTse could instead be picking up the effects of export slavery.[33] Fortunately, using data from Nunn and Wantchekon (2011), it is possible to control for the number of slaves exported by ethnic group when examining these particular outcomes (Appendix Table A.VI). This approach may be criticized for including as a control a variable that could be considered an outcome thereby leading to biased estimates (Angrist and Pischke, 2009, p. 64).[34] Empirically, including slave exports as a control hardly changes the point estimates

[33] Most historians agree that indigenous slavery antedated export slavery, but the conditions of African slaves within Africa were less harsh than the conditions of African slaves outside of Africa. A debate surrounds the extent to which slave conditions within Africa worsened as a result of the external slave trade (Lovejoy, 1983).

[34] The bias can be signed. Specifically (ignoring control variables for simplification), denote slave exports by z, then:

$$y_j = \mu_o + \alpha TSI_j + \pi z_j + \varepsilon_j \quad \text{(a)}$$
$$z_j = \mu_1 + \phi TSI_j + \eta_j \quad \text{(b)}$$

Assume $\phi > 0$. Substituting (b) into (a) demonstrates:

$$y_j = \mu_o + \pi\mu_1 + (\alpha + \pi\phi)\, TSI_j + \pi\eta_j + \varepsilon_j \quad \text{(c)}$$

The estimated effect of TSI on the outcome of interest when z is not included (c) will be greater than the effect when it is included, $|\alpha + \pi\phi| > |\alpha|$ if α and π both have positive or negative covariance with the compared to Table I column (3).

Lastly, in the empirical work, African ethnic group boundaries are taken as given. However, the TseTse may have increased the size of the ethnic group's land area by encouraging shifting agriculture. Theory predicts that the land/labor ratio should increase. Indeed, there is a positive correlation between the TSI and the land area of a ethnic group. To understand whether the main results reported in Table I are driven by the effect of the TseTse on land area size, the logarithm of land area (in kilometers^2) is included as a control in Table III column (6).[35] The results are almost identical to Table I column (3).[36]

V.6 The Differential Effect of the TseTse Suitability Index in Africa

To exclude the possibility that the TSI is identifying generic patterns between climate and agriculture, population etc., it is necessary to check for correlations between the TSI and the outcomes of interest outside of Africa, where the fly does not exist. It is straightforward to extend the TseTse population growth model to cover the globe. The sample now includes all ethnic groups in the *Ethnographic Atlas*. The estimating equation is as follows:

$$y_j = \beta\ TSI_j + \delta\ TSI_j \cdot I_j^{Africa} + \mathbf{X}_j'\Omega + I_j^{Africa}\mathbf{X}_j'\Gamma + \sum_k \gamma_k I_j^k + \eta_j, \tag{10}$$

outcome, y and $|\alpha + \pi\phi| < |\alpha|$ if their covariance with y differs. Comparing Table I with Appendix Table A.VI, the difference between $\alpha + \pi\phi$ and α is very small.

[35] In another specification, the tribal boundaries are ignored and equally sized "virtual countries" are created within Africa, following the work of Michalopoulos (2011). The average of tribal practices within each virtual country is used as the dependent variable. These results are reported in Appendix Table A.VII.

[36] Communicable diseases are often associated with externalities and, as highlighted by Kremer and Miguel (2004), empirical results can be misleading if they are ignored. One spillover effect that could drastically change the interpretation of the results was if economic activity from highly TseTse suitable areas was simply displaced to ethnic groups in less TseTse suitable areas. On the other hand, diffusion of technologies might have been hindered by having neighbors that were affected by the TseTse. To investigate violations of the Stable Unit Treatment Value Assumption (SUTVA), the average TSI of the ethnic groups within a 10 degree radius of a given ethnic group's centroid is included in the analysis. The results are presented in Appendix Table A.VIII. The coefficients on "nearby TSI" reinforce the general pattern of the main effect and the absolute value of the sum of own TSI and neighbor TSI is often greater than own TSI alone, suggesting that TseTse had negative spillover effects for nearby ethnic groups.

where I_j^k indicates that ethnic group j is in continent k and γ_k is a continent fixed effect. This approach is similar to the main estimating equation except the TSI enters as the main effect and TseTse is identified as the specific interaction between a dummy variable for Africa and the TSI. All other geographic and climate covariates from Table I column (3) are included as main effects and as interactions with the Africa dummy. Thus, Africa is allowed to differ from the rest of the world in many ways, not just the TseTse.[37]

A challenge to this analysis is that only the center of ethnic groups are mapped outside of Africa, not the entire boundaries. The standard approach to overcome this problem is to draw a circular "buffer zone" around the centroid. This approach is shown in panel (C) of Figure V. If the buffer zones are chosen too large, they overlap, making it difficult to allocate territory to mutually exclusive ethnic groups. If the buffer zones are too small, they will poorly approximate the actual boundaries. The approach followed in this paper is to construct Thiessen polygons, which more nearly approximate boundaries (compare Figure V panels (A) and (B)).[38] The boundaries for 1244 of the 1263 observations in the *Ethnographic Atlas* are able to be mapped in this way. The majority of ethnic groups dropped in this process are small island communities.[39] A comparison between the δ coefficients from Equation (9) for the *Map* versus Thiessen polygons sample is shown in Table IV. The coefficients are not significantly different.

Results from Equation (10) are shown in Table V. In many of the cases, the sign on the

[37] Standard errors are clustered by language families which are broader categories than cultural provinces though still capture spatial and cultural relatedness. This change was necessary because provinces are not available outside of Africa. Number of cities in 1800 was not included as an outcome because the spatial coverage of Chandler (1987) does not overlap with all parts of the *Atlas*.

[38] For a set of points S in Euclidean space, a Thiessen polygon is one such that every point in the constructed polygon is closer to one such point p than to any other point in S. Within Africa, Thiessen polygons have a higher correlation with the Murdock map boundaries than the buffer zone technique. The one drawback is that, for observations with identical centroids, the Thiessen polygons will be identical (though this would also occur with the use of buffer zones).

[39] The excluded ethnic groups, with locations in parentheses, include: Abelam (Papua New Guinea), Aymara (coast of Peru), Buem (coast of Ghana), Ifaluk (Micronesia), Kapingamarangi (Micronesia), Kunda (Zambia), Lamotrek (Micronesia), Manihikia (Cook Islands), Nomoians (Micronesia), Onotoa (Gilbert Islands), Ontong-Ja (Solomon Islands), Raroians (French Polynesia), Rotumans (Fiji), Trukese (Micronesia), Tukulor (coastline of West Africa), Ulithians (Micronesia), Upolu (Samoa), Yami (Orchid Island) and Yapese (Micronesia). Eight ancient civilizations (before 1500) are also excluded (these include Ancient Egypt, Aryans, Babylonia, Romans, Icelander, Uzbeg, Khmer and Hebrews).

$(TSI \cdot I^{Africa})$ interaction is different from the main effect. Furthermore, this interaction is statistically significant–indicating a differential effect of the TSI within Africa. The coefficient on the main TSI effect is usually insignificant, except for female participation in agriculture which has the opposite sign and is significant. This result appears to be driven exclusively by North America–places inhospitable for the fly tend to rely on hunting, which is predominantly done by men. One concern may be that these results are being driven by the difference between highly advanced Europe and Africa. Appendix Table A.IX panel (A) repeats the placebo test, but this time excluding European ethnic groups. The results are very similar. Appendix Table A.IX panel (B) estimates Equation (10) comparing Africa to other parts of the Tropical world. The sample size is now much smaller with less variation. Despite this limitation, the results are remarkably similar.

V.6.1 Africa without the TseTse: Archeological Evidence

The results presented so far are supportive of the hypothesis that the TseTse was a constraint on agricultural production and urbanization in historical Africa. One natural question to pose is how Africa would have developed in the absence of the fly. Historians, geographers and archeologists provide qualitative evidence that Africa without the TseTse would have been far more advanced.

The civilization of Great Zimbabwe was located on a plateau between the Zambezi and Limpopo rivers and has been described by archeologist Tim Connah (1987, p. 228) as a "peninsula in a sea of tsetse." Garlake (1978) noted that the boundaries of the Great Zimbabwe complex corresponded to the climatic boundaries of the TseTse described by Rogers and Randolph (Figure VI). The Great Zimbabwe plateau also falls within the bottom tenth percentile of the historical TSI distribution. Garlake (1978, p. 491) writes "there are no stone enclosures at any distance into the tsetse-infested lowlands. Indeed, only three exist more than ten miles beyond the suggested maximum extent of the tsetse infestation in 1896." Such stark boundaries may have been because the people of Great Zimbabwe greatly relied on cattle, as deduced from skeletal remains of livestock around the site. Their economy was complex, integrating cereal agriculture, pastoralism and trade. The elliptical building at the center of the city was by far the largest single precolonial structure in sub-Saharan Africa. Connah (1987, p. 224) states the archeological evidence from the Zimbabwe Plateau suggests "both urbanization and state formation during the first half of the second millennium AD" and compares the city with London or Paris during the same period.[40]

V.6.2 Simulation

The qualitative evidence cited above can be supplemented with a quantitative analysis. Using the specification presented in Equation (10), predicted values of the precolonial outcomes are generated for the Africa sample. The average values of these predicted outcomes are shown in Table VI, column (1). To represent a reduction in the burden of TseTse-related disease, every element in the $\left(TSI \cdot I^{Africa}\right)$ vector is reduced by one standard deviation within the Africa sample. This change brings the mean of the TSI within Africa to be approximately equivalent to the mean of the TSI outside of Africa. The average values of the predicted outcomes using the new $\left(TSI \cdot I^{Africa}\right)$ vector are shown in Table VI column (2). Africa would have nearly doubled in population and plow usage, it would have increased large domesticated animal usage by approximately a third, and slavery and female agriculture would have been substantially reduced. The outcomes for Africa are now much closer to those of Eurasia, which are shown in Table VI column (3). For example, the average logarithm of population density in 1700 is 1.63 in simulated Africa and 1.75 in Eurasia.

This exercise comes with the caveat that it does not take into account the endogenous response to an historical reduction in the burden of disease from the TseTse. Perhaps Africa would have been conquered and colonized earlier without the TseTse barrier.[41] Today, many ecologists view the TseTse as a keystone species, (a species critical for supporting an

[40] Theories for the decline of Great Zimbabwe include ecological disaster from overgrazing and climate change leading to an encroachment of fly belts onto the plateau.

[41] The late colonization of Africa is thought to be TseTse-related. Horseback-riding Muslim armies presumably had difficulty penetrating further South than the Sahel (Fukuyama, 2011, p. 91).

ecosystem) because it serves as a check on agricultural expansion and overgrazing, thereby protecting African biodiversity.

VI The TseTse and Current Development

Disease could affect development through its historical effect on shaping institutions and/or through contemporaneous impacts on health. There are various approaches to disentangle these two channels. One is to study eradication campaigns. Places where the fly no longer poses a threat to health would permit isolation of the historical channel. Unfortunately no large scale TseTse eradication campaigns have been successful. Another approach would be to find locations where the climate has changed sufficiently so that places historically suitable for the fly are no longer suitable or vice-versa. Since temperature changes would likely affect the Northern and Southern limits of where the fly can potentially exist, this approach may be suitable for a fine geospatial analysis. The country-level TSI does not change significantly enough over time for such an identification strategy to yield meaningful results.[42] OLS estimates of current economic performance on the fraction of a country's land that is TseTse-infested will be biased since wealthier states with better institutions may be better able to control the TseTse. The approach herein is to instrument for observed TseTse using the TSI constructed with modern climate data. The modern climate data are from the East Anglia Climate Research Unit and span the years 1961-1990. The data are made available as monthly means at a 10 minute resolution. If a correlation between the contemporaneous TseTse distribution and income per capita is noted, the test will be to see if it remains significant after controlling for the historical institutions shown to be correlated with the TSI.

Extractive economic institutions, such as slavery, could have persistent negative effects on

[42] The country-level correlation between historical and the modern TSI is 0.89 when using the same climate data set (20CRv2) to calculate the two measures. Panel data that would permit a finer geospatial analysis exploiting climate change since the mid to late 20th century as the exogenous variation in TseTse prevalence has not been found.

development for several reasons. First, slave systems tended to be technologically backwards– this is because the producers (the slaves) had no incentive to innovate as all rewards would be expropriated by their masters (Acemoglu and Robinson, 2012, p. 172). Second, slavery hindered human capital formation. Borjas (1992) has modeled how ethnic skill differentials act as an externality persisting across generations. Given the importance of human capital for modern growth (Goldin and Katz, 2009), such skill differentials may help to explain national income differences today. Third, as with the export slave trade, the internal slave trade might have encouraged ethnic fractionalization (Nunn, 2008) and mistrust (Nunn and Wantchekon, 2011), thereby contributing to poor governance. Pluralistic political institutions that are not centralized may devolve into chaos (Acemoglu and Robinson, 2012, p. 81), whereas centralized ones can enforce rules, deliver public goods and encourage economic growth (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2011, 2012).

To investigate whether the correlation between income per capita and observed TseTse are mediated through the channel of extractive economic or political institutions, a population weighted average of institution i in country c across ethnic groups j is constructed similar to the plow measure of Alesina, Giuliano, and Nunn (2011). Define $\mathbf{I}_j^i$ as a variable equal to one if ethnic group j was historically characterized by institution i and zero otherwise, where $i \in \{Indigenous\ Slavery, Centralization\}$. Let $L_{j,c}$ denote the number of individuals in ethnic group j living in country c and L_c is the total population of the country. The country-level measure of the prevalence of historical institution i in the population of country c is therefore:

$$\begin{aligned} Institution_c^i &= \frac{\sum_j L_{j,c} \cdot \mathbf{I}_j^i}{L_c} \\ i &\in \{Indigenous\ Slavery, Centralization\} \end{aligned} \tag{11}$$

This index captures the proportion of a country's current inhabitants whose ancestors lived in a centralized society or practiced slavery. The population data are from LandScan 2000 at a resolution of 1 kilometer. When constructing $L_{j,c}$ it is assumed that individuals identified within the boundaries of an ethnic group belong to it. Although there will be error in this index due to migration, it is remarkable how many individuals continue to inhabit their ancestral homeland.[43] Two-stage least-squares is used to estimate:

$$y_c = \alpha_1 + \delta TseTse_c + \sum_i \mu_i Institution_c^i + \mathbf{X}_c' \Omega + \varepsilon_c, \tag{12}$$

where y_c is the logarithm of income per capita in 2005 using income data from Maddison.[44] African countries with a population exceeding 250,000 are included. This restriction mainly excludes small island nations that have no TseTse yet high incomes.

The cross country relationship between the fraction of a country infested with the TseTse and income today is negative (see Table VII, panel (A), column (1)). In column (2), a limited set of climate and geographic controls are added (specifically, mean temperature, the logarithm of suitable land for agriculture and whether a country is landlocked) without a large change in the TseTse coefficient. Institutions set up by the colonizer or jurisdictional heritage might have persistent effects on development (La Porta et al., 1997; North, 1990). Colonizer and legal origins fixed effects are added next to investigate this hypothesis. In column (3), these controls appear to have a limited effect on the TseTse coefficient, reducing it by approximately one-fourth from the baseline specification in column (1). In column (4), precolonial institutions found to be correlated with the TSI are included. This reduces the point estimate coefficient on TseTse by approximately half and it loses statistical significance.[45] Similar to the findings of Michalopoulos and Papaioannou (2012), precolonial centralization has a positive and significant effect on current economic performance. On the other hand, indigenous slavery has a negative and significant effect on per capita income.

[43] Using the 2005 Afrobarometer survey for Nigeria, 2,292 respondents who list an indigenous ethnicity, 2,037 are living in the area of the ancestors (89 percent).

[44] There may be concern for "division bias" in these estimates as the weighted average of historical institutions is a per capita measure. Using a non-population based estimate of development, Government Effectivness in 2005, from Kaufmann et al. (2010) produces similar results. See Appendix Table A.X.

[45] The standard errors are too large to reject the possibility that the coefficient is actually the same as in column (1). The reduced form of equation (12) is shown in Appendix Table A.XI.

Once controlling for historical institutions, the TseTse does not have an independent, statistically significant effect on income per capita, though the sample of countries is small and the estimates are imprecise. Historically, there was a positive correlation between societies who used slaves and those who traded slaves to the Europeans. In column (5) the log of slaves exported is included as an additional control–the TseTse coefficient drops slightly further and the coefficients on both centralization and indigenous slavery remain significant at conventional levels.[46] Combining the estimates of Table VII column (4) with the simulation results in Table VI, the average African country would be approximately 30 percent wealthier had the fly not have influenced its historical institutions.

VII Concluding Remarks

This study has investigated the effect of the TseTse on African development. Using insect population growth models and laboratory experiments of TseTse physiology, a suitability index for TseTse was constructed. This index was then joined with ethnographic data on precolonial African agricultural practices, institutions and urbanization. Historical TseTse suitability was associated with less advanced agricultural practices, the use of slaves and a lower population density within but not outside of Africa.

Simulating African development with a lower burden of historical TseTse-transmitted disease demonstrated that population density in precolonial Africa would have doubled and indigenous slavery halved. These results should be interpreted with caution given that they do not allow for an endogenous response to the fly's removal. However, the predictions are broadly consistent with the archaeological record–civilizations such as Great Zimbabwe–with levels of technology and urbanization rivaling contemporaries in the rest of the Old World, flourished in areas of Africa inhospitable to the fly.

[46] Malaria is an oft-cited reason for African underdevelopment. Including the malaria ecology index in this set of regressions results in a coefficient with the "wrong" sign that is not significant and reduces the first-stage F-statistic. The coefficient on TseTse remains significant at convential levels but is estimated much less precisely.

The findings suggest TseTse-associated disease continues to influence development mainly through its effect on precolonial institutions. The cross-country exercise suffers from a small sample size and thus imprecise estimates. Microeconomic evidence from TseTse elimination campaigns would improve the estimates of the effect of current TseTse-transmitted *Trypanosomiasis* on the livelihoods of African peasants and represents an important area for future research.

References

Acemoglu, D. and J. A. . Robinson (2012). *Why Nations Fail: The Origins of Power, Prosperity and Poverty.* New York: Crown Publishers.

Aigner, D. (1973). Regression with a Binary Independent Variable Subject to Errors of Observation. *Journal of Econometrics 1*, 49–60.

Alesina, A., P. Giuliano, and N. Nunn (2011). On the Origins of Gender Roles: Women and the Plough. *National Bureau of Economic Research Working Paper Series No. 17098.*

Allen, R. C. (1999). Tracking the Agricultural Revolution in England. *The Economic History Review 52*, 209–235.

Altonji, J., T. Elder, and C. Taber (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy 113*, 151–184.

Anderson, C. and J. Fragmont (1857). *Lake Ngami: or, Explorations and discoveries during four years' wanderings in the wilds of southwestern Africa.* New York: Dix, Edwards & Co.

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

Bairoch, P. (1988). *Cities and Economic Development : From the Dawn of History to the Present.* Chicago: University of Chicago Press.

Bloom, D. E. and J. D. Sachs (1998). Geography, Demography, and Economic Growth in Africa. *Brookings Papers on Economic Activity 1998*, 207–295.

Bonnassie, P. (2009). *From Slavery to Feudalism in South-Western Europe.* Cambridge, U.K.: Cambridge University Press.

Borjas, G. (1992). Ethnic Capital and Intergenerational Mobility. *Quarterly Journal of Economics 107*, 123–150.

Borst, P. and G. Rudenko (1994). Antigenic Variation in African Trypanosomes. *Science (New York, N.Y.) 264*, 1872–3.

Boserup, E. (1966). *The Conditions of Agricultural Growth: The Economics of Agrarian Change Under Population Pressure.* Chicago: Aldine Pub. Co.

Brown, K. and D. Gilfoyle (2010). *Healing the Herds: Disease, Livestock Economies, and the Globalization of Veterinary Medicine.* Athens, Ohio: Ohio University Press.

Bryden, H. (1900). *Animals of Africa.* London: Sands and Company.

Bursell, E. (1960). Loss of water by Excretion and Defecation in the Tsetse Fly. *Journal of Experimental Biology 37*, 689–697.

Chandler, T. (1987). *Four Thousand Years of Urban Growth : An Historical Census*. Lewiston, New York: St. David's University Press.

Chaves, I., S. L. Engerman, and J. Robinson (2010). Reinventing the Wheel: The Economic Benefits of Wheeled Transportation in Early Colonial British West Africa. Working Paper.

Compo, G., J. Whitaker, and P. e. a. Sardeshmukh (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meterological Society 137*, 1–28.

Conley, T. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics 92*, 1–45.

Connah, G. (1987). *African Civilizations : Precolonial Cities and States in Tropical Africa : An Archaeological Perspective*. Cambridge, U.K.: Cambridge University Press.

Diamond, J. M. (1997). *Guns, Germs, and Steel : The Fates of Human Societies* (1st ed.). New York: W.W. Norton.

Domar, E. D. (1970). The Causes of Slavery or Serfdom: A Hypothesis. *The Journal of Economic History 30*, 18–32.

Engerman, S. L. and K. L. Sokoloff (2000). History Lessons: Institutions, Factors Endowments, and Paths of Development in the New World. *The Journal of Economic Perspectives 14*, 217–232.

FAO (1982). Training Manual for Tsetse Control Personnel. *2*, 1–79.

FAO (1998). A field guide for the diagnosis, treatment and prevention of african animal trypanosomiasis.

Fenske, J. (2009). Does Land Abundance Explain African Institutions? Technical report, Economic Growth Center, Yale University.

Ford, J. and K. Katondo (1977). Maps of Tsetse Flies (Glossina) Distribution in Africa, 1973, according to Sub-Generic Groups on a Scale of 1: 5 000 0000. *Bulletin of Animal Health and Production in Africa 25*, 187–193.

Fukuyama, F. (2011). *The Origins of Political Order : From Prehuman Times to the French Revolution* (1st ed.). New York: Farrar, Straus and Giroux.

Gallup, J. and J. Sachs (2001). The Economic Burden of Malaria. *The American Journal of Tropical Medicine and Hygiene 64*, 85–96.

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

Garlake, P. S. (1978). Pastoralism and Zimbabwe. *The Journal of African History 19*, 479–493.

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

Glasgow, J. P. (1963). *The Distribution and Abundance of Tsetse*. International Series of Monographs on Pure and Applied Biology. New York: Macmillan.

Goldewijk, K. K. (2005). Three Centuries of Global Population Growth: A Spatial Referenced Population (Density) Database for 1700-2000. *Population & Environment 26*, 343–367.

Goldin, C. and L. Katz (2009). *The Race Between Education and Technology*. Cambridge, Mass: Belknap Press of the Harvard University Press.

Goody, J. (1971). *Technology, Tradition and the State in Africa*. London: Oxford University Press.

Grigg, D. B. (1980). *Population Growth and Agrarian Change : An Historical Perspective*. Cambridge Geographical Studies ; 13. Cambridge [Eng.]: Cambridge University Press.

Harding, A. (1897). *Report by Sir A. Harding on the Condition and Progress of the East Africa Protectorate from its Establishment to the 20 July, 1897*.

Hay, S. I., C. A. Guerra, A. J. Tatem, A. M. Noor, and R. W. Snow (2004). The Global Distribution and Population at Risk of Malaria: Past, Present, and Future. *The Lancet Infectious Diseases 4*, 327–336.

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

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

Iliffe, J. (1995). *Africans: The History of a Continent*. Cambridge, U.K.: Cambridge University Press.

Jackson, C. H. N. (1949). The Biology of Tsetse Flies. *Biological Reviews 24*, 174–199.

Johnston, H. (1894). *Report by Commissioner Johnston of the First Three Years' Administration of the Eastern Portion of British Central Africa*. Her Majesty's Stationery Office.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2010). The Worldwide Governance Indicators: Methodology and Analytical Issues. *World Bank Policy Research Paper*.

Kiszweski, A., A. Mellinger, A. Spielman, P. Malaney, S. Sachs, and J. Sachs (2004). A Global Index Representing the Stability of Malaria Transmission. *The American Journal of Tropical Medicine and Hygiene 70*, 486–498.

Kling, J., J. Liebman, L. Katz, and L. , Sanbonmatsu (2004). Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment. *Harvard mimeo*.

Krafsur, E. S. (2009). Tsetse flies: Genetics, Evolution, and Role as Vectors. *Infection Genetics and Evolution 9*, 124–141.

Kremer, M. and E. Miguel (2004). "worms: Identifying Impacts on Eduaction and Health in the Presence of Treatment Externalities". *Econometrica 72*, 159–217.

La Porta, R., F. Lopez-De-Silanes, A. Shleifer, and R. W. Vishny (1997). Legal Determinants of External Finance. *The Journal of Finance 52*(3), 1131–1150.

Lambrecht, F. L. (1964). Aspects of evolution and ecology of Tsetse flies and Trypanosomiasis in prehistoric African environment. *The Journal of African History 5*, 1–24.

Law, R. (1977). *The Oyo Empire c.1600-c.1836: A West African Imperialism in the Era of the Atlantic Slave Trade*. Oxford: Clarendon Press.

Leak, S. (1999). *Tsetse Biology and Ecology: Their Role in the Epidemiology and Control of Trypanosomosis*. CABI Publishing in association with International Livestock Research Institute.

Legros, M., A. Lloyd, Y. Huang, and F. Gould (2009). Density-Dependent Intraspecific Competition in the Larval Stage of Aedes aegypti (Diptera: Culicidae):Revisiting the Current Paradigm. *Journal of Medical Entomology 46*, 409–419.

Livingstone, D. (1857). *Missionary Travels and Researches in South Africa*. London: Harper and Brothers.

Lovejoy, P. E. (1983). *Transformations in Slavery : A History of Slavery in Africa*. Cambridge, U.K.: Cambridge University Press.

Luckins, A. G. and R. H. Dwinger (2004). *The Trypanosomiases*. CAB International.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni (2008). Proxy-based Reconstructions of Hemispheric and Global Surface Temperature Variations over the Past Two Millennia. *Proceedings of the National Academy of Sciences 105*, 13252–13257.

Maudlin, I. (2006). African Trypanosomiasis. *Annals of Tropical Medicine and Parasitology*, 679–708.

May, R., G. Conway, M. Hassell, and T. Southwood (1974). Time-Delays, Density-Dependence and Single-Species Oscillations. *Journal of Animal Ecology 43*, 747–770.

Mazoyer, M., L. Roudart, and J. Membrez (2006). *A History of World Agriculture: From the Neolithic Age to the Current Crisis*. Earthscan.

McEvedy, C. and R. Jones (1978). *Atlas of World Population History*. Hammondsworth: Penguin.

Mellanby, K. (1937). Experimental work with the Tsetse-Fly, Glossina Palpalis in Uganda. *Bulletin of Entomological Research 28*, 611–632.

Michalopoulos, S. (2011). The Origins of Ethnolinguistic Diversity. *American Economic Review 102*, 1508–1539.

Michalopoulos, S. and E. Papaioannou (2011). Divide and Rule or the Rule of the Divided? Evidence from Africa. *National Bureau of Economic Research Working Paper Series No. 17184*.

Michalopoulos, S. and E. Papaioannou (2012). Pre-colonial Ethnic Institutions and Contemporary African Development. National Bureau of Economic Research Working Paper No. 18224.

Moore, N. and J. Messina (2010). A Landscape and Climate Data Logistic Model of Tsetse Distribution in Kenya. *PLOS One 5*, 1–10.

Mulla, A. F. and L. R. Rickman (1988). How do African Game Animals Control Trypanosome Infections? *Parasitology Today 4*, 352–354.

Murdock, G. P. (1959a). *Africa, its Peoples and their Culture History.* New York: McGraw-Hill.

Murdock, G. P. (1959b). Tribal Map of Africa.

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

Murray, M., J. C. Trail, and G. D. D'Ieteren (1990). Trypanotolerance in Cattle and Prospects for the Control of Trypanosomiasis by Selective Breeding. *Revue Scientifique et Technique*, 369–386.

Nash, T. (1969). *Africa's Bane: The Tsetse Fly.* London: Collins.

Nieboer, H. (1900). *Slavery as an Industrial System: Ethnological Researches.* The Hague.

North, D. (1990). *Institutions, Institutional Change, and Economic Performance.* Cambridge University Press.

Nunn, N. (2008). The Long-term Effects of Africa's Slave Trades. *The Quarterly Journal of Economics 123*, 139–176.

Nunn, N. and N. Qian (2011). The Potato's Contribution to Population and Urbanization: Evidence from A Historical Experiment. *The Quarterly Journal of Economics 126*, 593–650.

Nunn, N. and L. Wantchekon (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review 101*, 3221–3252.

Overton, M. (1996). *Agricultural Revolution in England: The Transformation of the Agrarian Economy, 1500-1850.* Cambridge, U.K.: Cambridge University Press.

Owaga, M. L. A. (1985). Observations on the Efficacy of Buffalo Urine as a Potent Olfactory Attractant for Glossina Pallidipes Austen. *International Journal of Tropical Insect Science 6*, 561–566.

Rajagopal, P. K. and E. Bursell (1965). The Effect of Temperature on the Oxygen Consumption of Tsetse Pupae. *Bulletin of Entomological Research 56*, 219–225.

Rogers, D. J. and S. E. Randolph (1984). A Review of Density-Dependent Processes in Tsetse Populations. *International Journal of Tropical Insect Science 5*, 397–402.

Rogers, D. J. and S. E. Randolph (1986). The Distribution and Abundance of Tsetse Flies. *Journal of Animal Ecology 55*, 1007–1025.

Sachs, J. and P. Malaney (2002). The Economic and Social Burden of Malaria. *Nature 415*, 680–685.

Schowalter, T. (2011). *Insect Ecology: An Ecosystem Approach*. Elsevier Science.

Terblanche, J. S., S. Clusella-Trullas, J. A. Deere, and S. L. Chown (2008). Thermal tolerance in a South-East African Population of the Tsetse Fly Glossina Pallidipes (Diptera, Glossinidae): Implications for Forecasting Climate Change Impacts. *Journal of Insect Physiology 54*, 114–127.

Turnbull, P. C., M. Doganay, P. M. Lindeque, B. Aygen, and J. McLaughlin (1992). Serology and Anthrax in Humans, Livestock and Etosha National park Wildlife. *Epidemiol Infect 108*, 299–313.

Vale, G. A. (1974). The Responses of Tsetse Flies (Diptera, Glossinidae) to Mobile and Stationary Baits. *Bulletin of Entomological Research 64*, 545–588.

Vale, G. A., S. Flint, and D. R. Hall (1986). The Field Responses of Tsetse Flies, Glossina spp. (Diptera: Glossinidae), to Odours of Host Residues. *Bulletin of Entomological Research 76*, 685–693.

Weil, D. N. (2010). The Impact of Malaria on African Development over the Longue Duree. Working Paper.

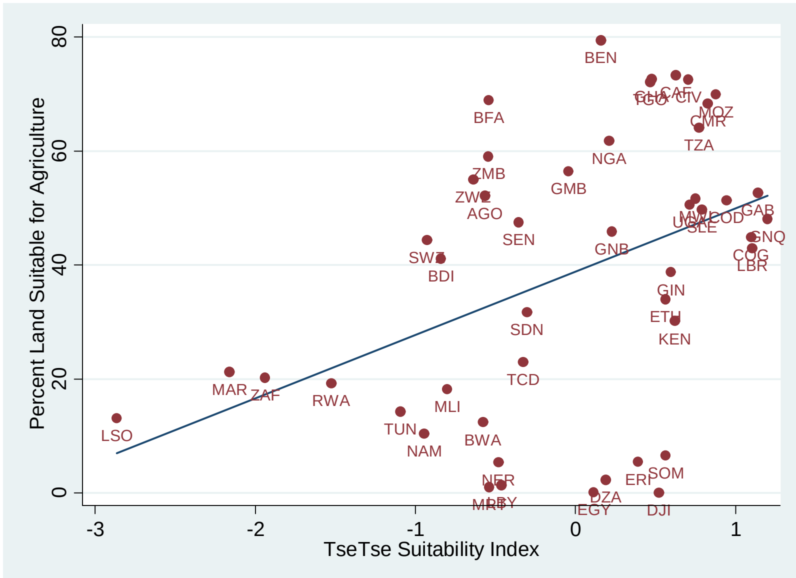
White, L. (1978). *Medieval Religion and Technology: Collected Essays*. London: University of California Press.

WHO (2001). Sponging Cattle with Insecticide Halves Malaria Rates. *Bulletin of the World Health Organization 79*.

Wint, W. and D. Rogers (2000). Report on the predicted distributions of Tsetse in Africa. DFID.

VIII Figures and Tables

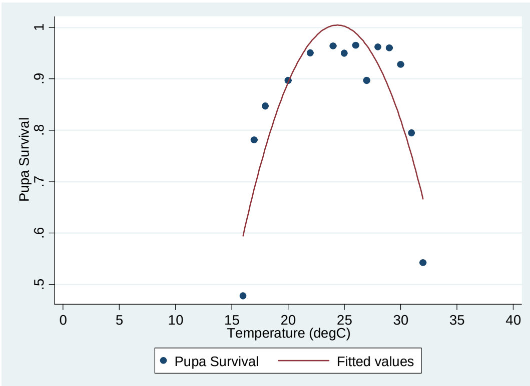
Figure I: Agricultural Suitability versus TseTse Suitability (African Countries)



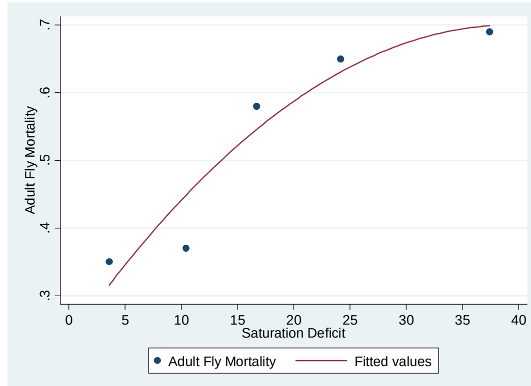
Notes: This figure demonstrates the correlation between agricultural suitability and TseTse suitability. Data on agricultural suitability for rainfed crops are from the FAO Global Agro-Ecological Zones (2002). Details on the FAO methodology can be found in Appendix B. The TseTse suitability index (TSI) is based on the author's calculations using climate data from the East Anglia Climate Research Unit. The equations for the TSI can be found in Appendix Table B.I.

Figure II: Physiology of the TseTse Fly

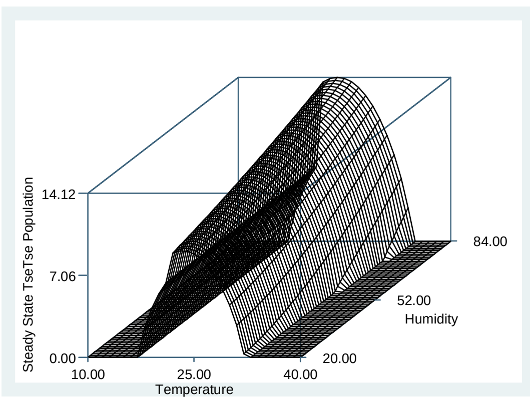
A. Pupa Survival and Temperature



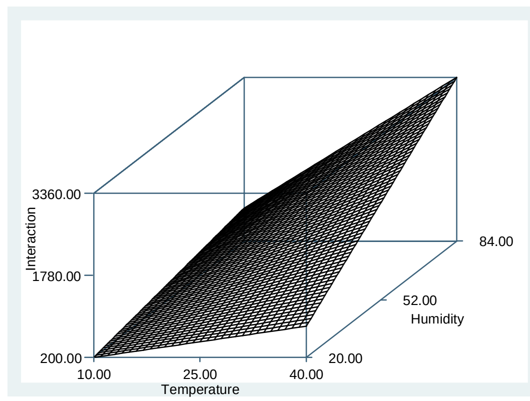
B. Adult Fly Mortality and Saturation Deficit



C. Steady State TseTse Population



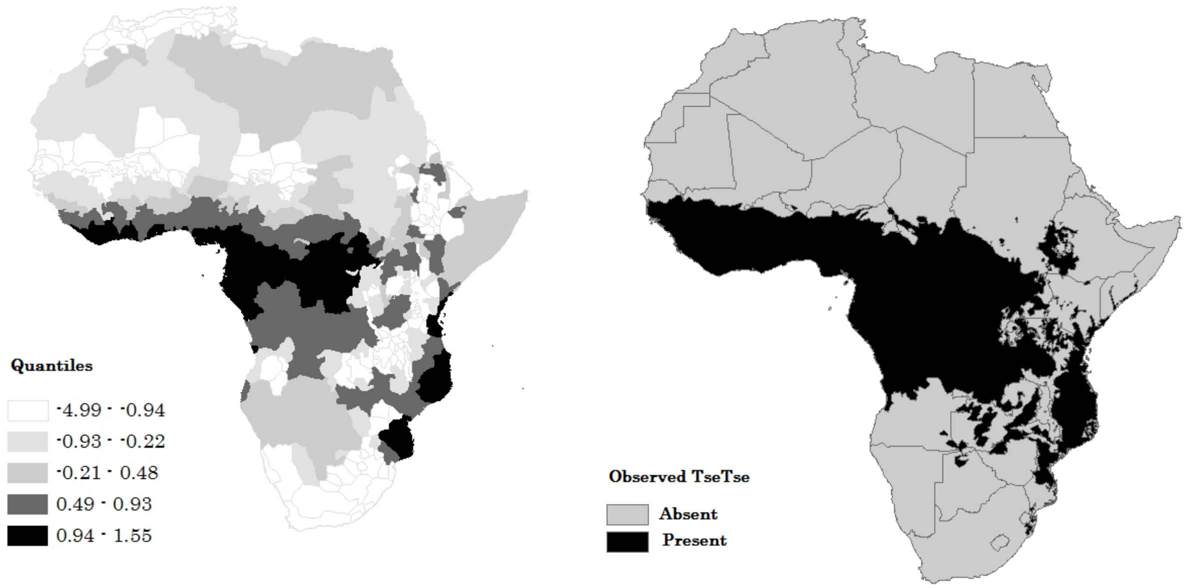
D. Linear Interaction of Climate Variables



Notes: These graphs show the relationship between TseTse physiology and climate. Panel (A) data points are from Bursell (1960) and Rajagopal and Bursell (1965). Panel (B) data points are from K. Mellanby (1937). Panel (C) depicts the steady state population of TseTse as a function of climate. Panel (D) plots the first-order interaction of humidity and temperature.

Figure III: TseTse Suitability Index Versus the Observed TseTse Distribution

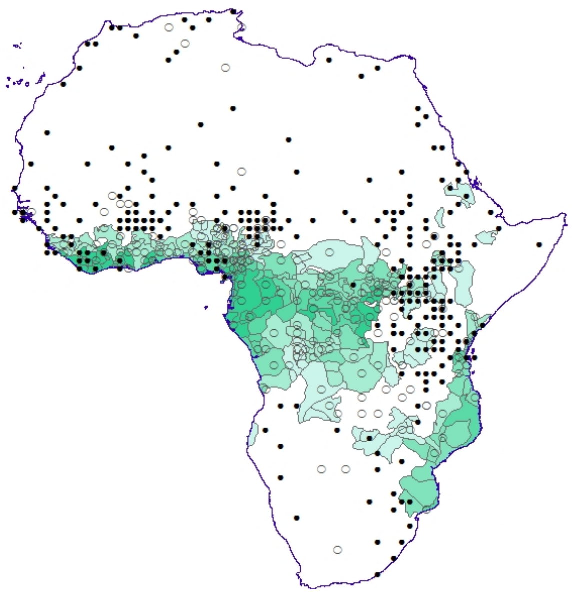
A. TseTse Suitability Index (1871) B. TseTse Distribution (1973)



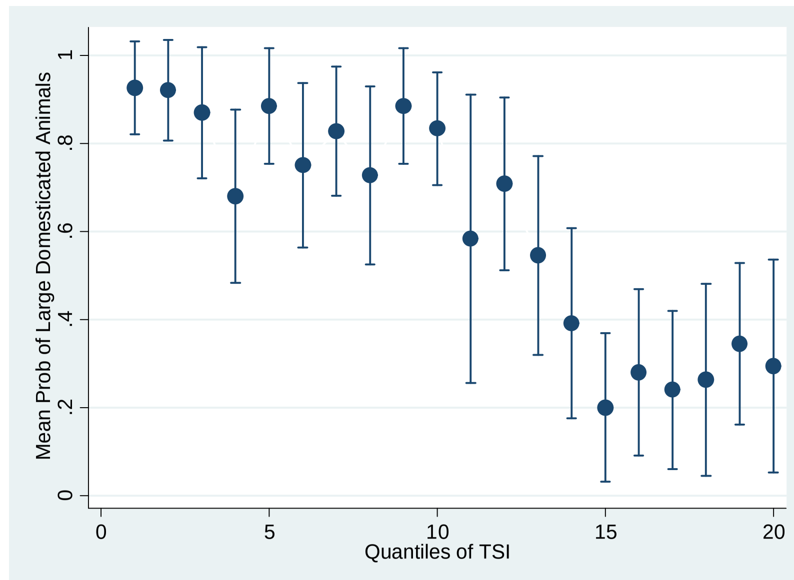
Notes: Panel (A) shows the historical TseTse suitability index created using climate data from NOAA's 20th century reanalysis for the year 1871. Panel (B) shows the observed TseTse distribution in 1973 (Ford and Katondo, 1977).

Figure IV: Visual Reduced Form

A. Map of Large Domesticated Animals

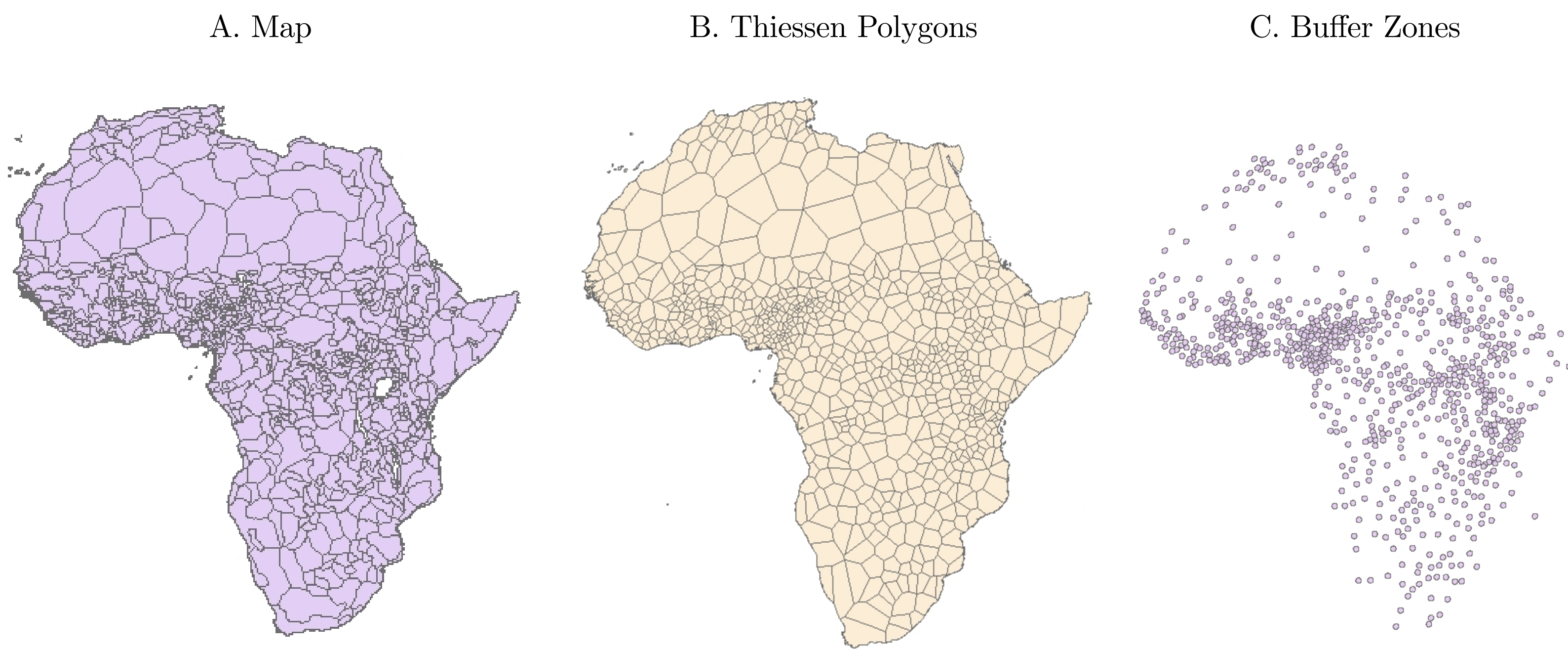


B. Quantiles of TSI and the Mean of Large Domesticated Animals



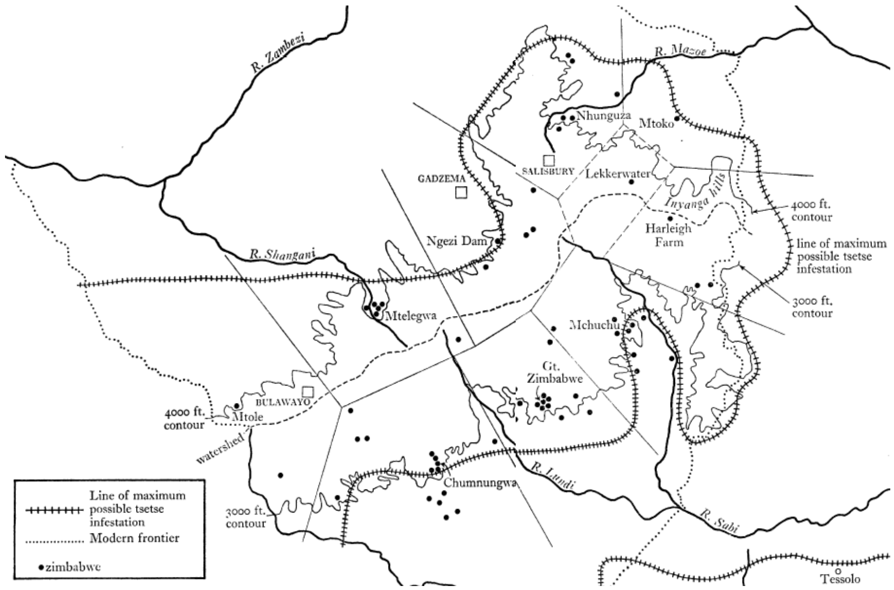
Notes: These figures demonstrate the reduced form relationship between the historical Tsetse suitability index (TSI) and large domesticated animals. Panel (A) depicts the upper 50th percentile of the TSI with darker shades denoting higher values. Black circles represent ethnicity-level presence of the outcome. Hollow circles represent ethnicity-level absence of the outcomes. The plot in panel (B) was created by binning the TSI into twenty equally-sized bins (5 percentile points) and plotting the mean probability of using large domesticated animals within each bin. Error bars represent 95 percent confidence intervals.

Figure V: Thiessen Polygons



Notes: This figure compares the Murdock map of ethnic groups in panel (A) with the constructed Thiessen polygons in panel (B) and with the buffer zone approach, panel (C).

Figure VI: Map of Great Zimbabwe



Notes: This figure demonstrates the overlap between the boundaries of Great Zimbabwe and the line of maximum possible TseTse infestation. The figure is used with permission from Garlake (1978). The hatched line indicates the line of maximum possible TseTse infestation as described by Oxford zoologists David Rogers and Sarah Randolph. The black dots represent the *zimbabwe*–stone enclosures indicative of human settlement.

Table I

REDUCED FORM ESTIMATES OF THE RELATIONSHIP BETWEEN HISTORICAL AFRICAN DEVELOPMENT AND TSETSE SUITABILITY

Dependent Variable	(1)	(2)	(3)	(4)	No. Observations	No. Clusters	Sample mean
Agriculture					[subsample]	[subsample]	[subsample]
Large Domesticated Animals	–0.216***	–0.212***	–0.219***	–0.255***	484	44	0.62
	(0.046)	(0.047)	(0.040)	(0.047)	[443]	[43]	[0.60]
Intensive Agriculture	–0.104***	–0.095***	–0.087***	–0.106***	468	43	0.32
	(0.030)	(0.033)	(0.028)	(0.033)	[427]	[42]	[0.28]
Plow Use	–0.078***	–0.084***	–0.067***	–0.040**	476	44	0.06
	(0.025)	(0.027)	(0.021)	(0.019)	[442]	[43]	[0.03]
Female Participation in Agriculture	0.262***	0.234***	0.195***	0.238***	316	43	0.49
	(0.050)	(0.048)	(0.055)	(0.067)	[292]	[40]	[0.49]
Urbanization							
Log Population Density (1700)	–0.859***	–0.777***	–0.615**	–0.405*	522	44	0.76
	(0.274)	(0.270)	(0.228)	(0.214)	[478]	[43]	[0.90]
Presence of Urban Center (1800)	–0.044**	–0.041*	–0.033**	–0.026**	521	44	0.04
	(0.021)	(0.020)	(0.016)	(0.013)	[477]	[43]	[0.03]
Institutions							
Indigenous Slavery	0.110***	0.124***	0.117***	0.102**	444	44	0.85
	(0.037)	(0.041)	(0.039)	(0.040)	[411]	[43]	[0.87]
Centralization	–0.066*	–0.085**	–0.082**	–0.095**	468	44	0.33
	(0.037)	(0.036)	(0.033)	(0.041)	[429]	[43]	[0.31]
Climate controls	Y	Y	Y	Y			
Malaria control	N	Y	Y	Y			
Geography controls	N	N	Y	Y			
Tropical Africa subsample	N	N	N	Y			

Notes: OLS estimates of Equation (9) are reported. The dependent variable is listed in the leftmost column of the upper panel. Climate controls refer to temperature, relative humidity and the first-order interaction between temperature and humidity. Malaria refers to the malaria ecology index developed by Kiszewski et al., (2004). Geography controls include log mean altitude, the FAO's agricultural suitability index, longitude, absolute latitude and access to waterways. Coefficients are reported with robust standard errors clustered at the level of cultural province in parentheses. Subsample means and clusters in brackets.* ** *** Signficant at 10, 5 and 1 percent levels.

Table II

OLS AND IV ESTIMATES OF THE RELATIONSHIP BETWEEN OBSERVED TSETSE AND HISTORICAL AFRICAN DEVELOPMENT

Dependent Variable	Large Domesticated Animals	Intensive Agriculture	Plow Use	Female Participation in Agriculture	Indigenous Slavery	Centralization	Log Population Density (1700)	Presence of Urban Center (1800)
				Panel A: Two-Stage Least Squares				
Observed TseTse (1973)	–2.202***	–0.869***	–0.639**	1.908**	1.190***	–0.850*	–6.316*	–0.337*
	(0.776)	(0.358)	(0.272)	(0.752)	(0.433)	(0.430)	(3.316)	(0.188)
				Panel B: First Stage				
TseTse Suitability Index (1871)	0.100***	0.100***	0.105***	0.102***	0.098***	0.096***	0.097***	0.097***
	(0.027)	(0.025)	(0.027)	(0.033)	(0.024)	(0.028)	(0.027)	(0.027)
F-statistic	11.99	11.94	12.64	12.69	10.53	11.34	10.47	10.43
				Panel C: Ordinary Least Squares				
Observed TseTse (1973)	–0.177	0.042	–0.055	–0.021	0.180	–0.114	0.256	–0.009
	(0.112)	(0.121)	(0.036)	(0.118)	(0.134)	(0.074)	(0.391)	(0.016)
No. obs.	484	465	476	316	444	468	522	521
Clusters	44	43	44	43	44	44	44	44

Notes: Panel (A) reports the two-stage least squares estimates. The dependent variable is noted in row (1). Panel (B) reports the corresponding first stage. Panel (C) reports the OLS coefficient from a regression of the outcome in row (1) on the fraction of tribal land observed to be TseTse-infested in 1973 (Ford and Katondo, 1977). Control variables (not reported to save space) include climate controls (temperature, relative humidity and the first-order interaction of temperature and humidity), the malaria ecology index developed by Kiszewski et al., (2004) and other geography controls (log mean altitude, the FAO's agricultural suitability index, longitude, absolute latitude and access to waterways). Coefficients are reported with robust standard errors clustered at the level of cultural province in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table III

ROBUSTNESS TESTS OF THE RELATIONSHIP BETWEEN THE REDUCED FORM ESTIMATES OF TSETSE SUITABILITY AND HISTORICAL AFRICAN DEVELOPMENT

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Higher Order Terms	Malaria (1900)	Alternative TseTse Index	Conley S.E.	Probit	Ethnic Group Land Area
Agriculture						
Large Domesticated Animals	–0.223***	–0.242***	–0.349***	–0.219***	–0.216***	–0.212***
	(0.053)	(0.035)	(0.060)	(0.042)	(0.040)	(0.038)
Intensive Agriculture	–0.107***	–0.112***	–0.121***	–0.087**	–0.109**	–0.079**
	(0.053)	(0.034)	(0.056)	(0.035)	(0.044)	(0.036)
Plow Use	0.013	–0.082***	–0.043*	–0.067**	–0.010*	–0.062***
	(0.039)	(0.022)	(0.023)	(0.024)	(0.006)	(0.021)
Female Participation in Agriculture	0.145*	0.243***	0.272***	0.195***	0.216***	0.179***
	(0.075)	(0.046)	(0.074)	(0.043)	(0.050)	(0.051)
Urbanization						
Log Population Density (1700)	–0.419**	–0.641***	–0.865***	–0.615**	N.A.	–0.458***
	(0.199)	(0.158)	(0.207)	(0.229)		(0.147)
Presence of Urban Center (1800)	–0.046*	–0.036***	–0.016	–0.033**	–0.036**	–0.039***
	(0.016)	(0.013)	(0.017)	(0.014)	(0.017)	(0.014)
Institutions						
Indigenous Slavery	0.081	0.111***	0.099**	0.117***	0.080**	0.115***
	(0.065)	(0.035)	(0.049)	(0.042)	(0.034)	(0.037)
Centralization	–0.139**	–0.065	–0.001	–0.082**	–0.087*	–0.107**
	(0.058)	(0.041)	(0.060)	(0.041)	(0.046)	(0.042)

Notes: The dependent variable is noted in the leftmost column. All specifications include temperature, relative humidity and the first-order interaction between temperature and humidity as well as the log mean altitude, the FAO's agricultural suitability index, access to waterways, the malaria ecology index developed by Kiszewski et al., (2004), absolute latitude and longitude. In addition, column (1) includes second-order terms in temperature, humidity and their interaction. Column (2) uses a different malaria measure: malaria prevalence in 1900 (Hay et al., 2004). Column (3) reports the results from an alternative measure for TseTse suitability developed by Rogers and Randolph (1986). Column (4) uses Conley standard errors with cutoffs at 5 degrees latitude and 5 degrees longitude. Column (5) includes results from a probit model and a negative binomial model for the last row (number of cities); marginal effects are reported. Column (6) includes the log of tribal area in square kilometers as a control. For columns (1)-(4) and column (6) OLS coefficients are reported. Robust standard errors in parentheses for columns (1)-(3) and columns (5)-(6). * ** *** Signficant at 10, 5 and 1 percent levels.

Table IV
COMPARISION OF MURDOCK MAP AND THIESSEN POLYGON COEFFICIENTS

Dependent Variable	(1) Map	(2) Polygon	(3) Difference
Agriculture			
Large Domesticated Animals	–0.219***	–0.274***	0.06
	(0.040)	(0.085)	
Intensive Agriculture	–0.087***	–0.184***	1.51
	(0.028)	(0.057)	
Plow Use	–0.067***	–0.088*	0.43
	(0.021)	(0.044)	
Female Participation in Agriculture	0.195***	0.338***	–1.31
	(0.055)	(0.094)	
Urbanization			
Log Population Density (1700)	–0.615**	–0.947***	0.92
	(0.228)	(0.282)	
Institutions			
Indigenous Slavery	0.117***	0.173**	–0.80
	(0.039)	(0.058)	
Centralization	–0.082**	–0.055	–0.30
	(0.033)	(0.081)	

Notes: OLS estimates of Equation (9) using the Murdock *Map* sample in column (1) and the Thiessen Polygons sample in column (2). The dependent variable is noted in the leftmost column. All specifications include temperature, relative humidity and the first-order interaction between temperature and humidity as well as the log mean altitude, the FAO's agricultural suitability index, access to waterways, the malaria ecology index developed by Kiszewski et al., (2004), absolute latitude and longitude. Coefficients are reported in columns (1) and (2) with robust standard errors clustered at the level of cultural province in parentheses. The *t*-statistic on the difference between the two coefficients is reported in column (3). * ** *** Signficant at 10, 5 and 1 percent levels..

Table V
PLACEBO TEST

Dependent Variable	(1) Main Effect TSI (β)	(2) Africa Interaction TSI (δ)	(3) Africa Total TSI (β+δ)
Agriculture			
Large Domesticated Animals	0.030	–0.304***	–0.275***
	(0.032)	(0.070)	(0.054)
Intensive Agriculture	–0.039	–0.212***	–0.251***
	(0.044)	(0.051)	(0.013)
Plow Use	–0.021	–0.062	–0.083**
	(0.020)	(0.039)	(0.035)
Female Participation in Agriculture	–0.098**	0.364***	0.266***
	(0.043)	(0.071)	(0.053)
Urbanization			
Log Population Density (1700)	–0.158	–0.792**	–0.950***
	(0.262)	(0.309)	(0.186)
Institutions			
Indigenous Slavery	–0.078	0.229**	0.151***
	(0.065)	(0.080)	(0.052)
Centralization	–0.025	–0.105**	–0.130***
	(0.037)	(0.044)	(0.028)

Notes: OLS estimates of Equation (10). The dependent variable is noted in the leftmost column. The sample includes ethnic groups from all over the world. Ethnic group boundaries are constructed using Thiessen polygons. Column (1) reports the coefficient on TSI. Column (2) reports the coefficient on the TSI*Africa interaction. Column (3) reports the coefficient of column (1) + column (2). Standard errors clustered at the ethnic language family in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table VI
SIMULATION

Dependent Variable	(1) Africa Baseline TseTse	(2) Africa Reduced TseTse	(3) Eurasia
Agriculture			
Large Domesticated Animals	0.62	0.93	0.83
Intensive Agriculture	0.32	0.54	0.75
Plow Use	0.07	0.13	0.72
Female Participation in Agriculture	0.46	0.10	0.13
Urbanization			
Log Population Density (1700)	0.85	1.63	1.75
Institutions			
Indigenous Slavery	0.84	0.61	0.23
Centralization	0.34	0.44	0.63

Notes: This table is created by using OLS to predict the outcomes for the Africa sample using Equation (10). The mean values of these predicted outcomes are shown above in column (1). Then each element in the TSI*Africa vector is reduced by one standard deviation. This change forces the mean value of the TSI within Africa to be approximately equivalent to the mean value of the TSI outside of Africa. The mean values of the predicted outcomes with the new TSI*Africa vector are shown in column (2).

Table VII

PERSISTENCE OF HISTORICAL INSTITUTIONS

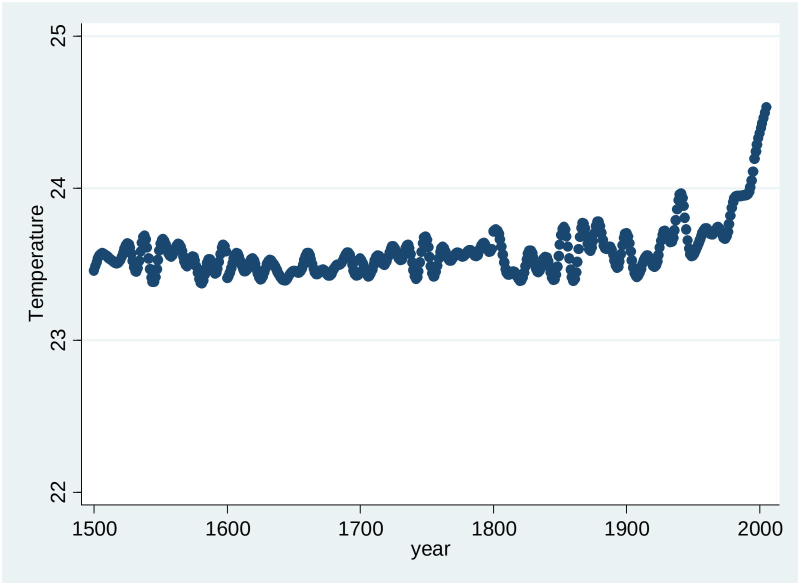
Dependent Variable is the Log GDP per capita (2005)					
Panel A: Two-Stage Least Squares					
	(1)	(2)	(3)	(4)	(5)
Observed TseTse (1973)	–0.761**	–0.798**	–0.526*	–0.179	–0.164
	(0.335)	(0.340)	(0.298)	(0.262)	(0.283)
Landlocked		–0.644***	–0.522**	–0.474***	–0.456***
		(0.341)	(0.169)	(0.141)	(0.140)
Log Suitable Land Agriculture (10 km^2)		0.022	–0.021	0.032	0.032
		(0.049)	(0.040)	(0.043)	(0.043)
Temperature		–0.058***	–0.084***	–0.068***	–0.066***
		(0.022)	(0.025)	(0.022)	(0.023)
Historical Centralization				0.710***	0.706***
				(0.256)	(0.252)
Historical Indigenous Slavery				–0.854***	–0.840***
				(0.262)	(0.261)
Colonizer and Legal Origins Fixed Effects	N	N	Y	Y	Y
Historical Slave Exports	N	N	N	N	Y
Panel B: First Stage for Observed TseTse (1973)					
TseTse Suitabiliey Index (1961-1999)	0.344***	0.385***	0.406***	0.423***	0.391***
	(0.036)	(0.050)	(0.054)	(0.078)	(0.087)
F-statistic	93.61	60.04	62.13	32.18	22.42
Panel C: Ordinary Least Squares					
Observed TseTse (1973)	–0.645**	–0.629**	–0.386*	–0.177	–0.163
	(0.266)	(0.241)	(0.213)	(0.214)	(0.220)
No. obs.	43	43	43	43	43

Notes: Panel (A) reports the two-stage least squares estimates with log GDP per capita in 2005 as the dependent variable. Panel (B) reports the corresponding first stage. Panel (C) reports the OLS coefficient from a regression of Log GDP per capita on the fraction of a country's land area that is TseTse infested. Control variables are not reported in panels (B) and (C) to save space. Legal Origins are from La Porta et. al., (1999). The sample is either of French or British legal origin. Log {(slave exports+1)/land area} by ethnicity from Nunn (2008) are included in column (6). Coefficients are reported with robust standard errors in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

IX Appendices

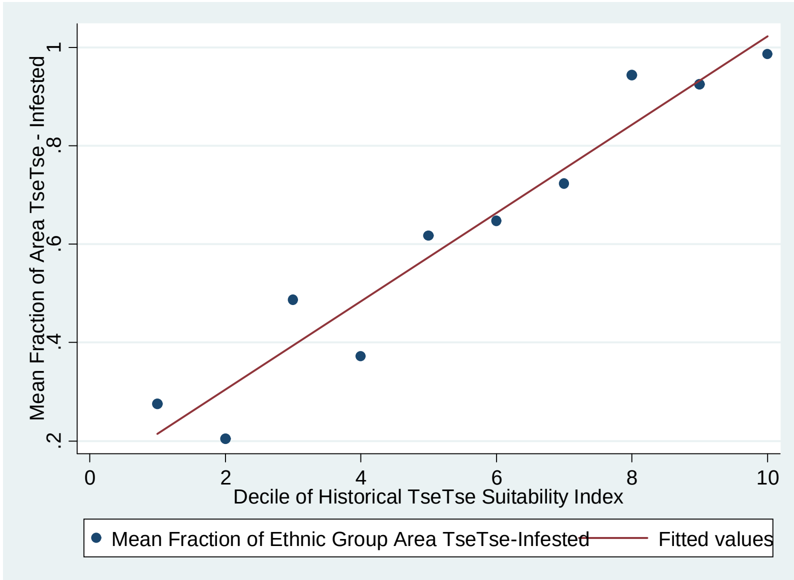
IX.1 Appendix A: Figures and Tables

Figure A.I: Africa's Temperature Over the Long Run



Notes: Figure constructed using paleoclimatic data on temperature from Mann et al., (2008).

Figure A.II: Fraction of Land Area of Ethnic Group TseTse-Infested versus the Historical TSI



Notes: This figure plots the correlation between the average fraction of an ethnic group's land area TseTse-infested per decile of the historical TseTse suitability index. The fraction of area TseTse-infested is calculated from the Ford and Katondo maps (1973).

Figure A.III: Correlation Matrix of the Historical TSI with Other Geographic and Climate Covariates

	TSI	Temp	RH	Itx	Mal 1900	Mal Ecology	Abs Lat	Longitude	River	Coast	Log Alt	Ag Suit
Historical TSI	1.00											
Historical Temperature	0.38	1.00										
Historical RH	0.36	–0.45	1.00									
Interaction	0.66	0.04	0.87	1.00								
Malaria 1900	0.42	0.25	0.41	0.59	1.00							
Malaria Ecology Index	0.38	0.70	–0.01	0.37	0.46	1.00						
Absolute Latitude	–0.39	–0.08	–0.52	–0.60	–0.52	–0.25	1.00					
Longitude	–0.21	–0.46	0.15	–0.10	–0.21	–0.39	–0.22	1.00				
River	0.10	–0.05	0.19	0.19	0.18	0.12	–0.10	0.02	1.00			
Coast	0.04	0.01	0.06	0.07	–0.11	–0.16	0.29	–0.20	–0.03	1.00		
Log Mean Altitude (meters)	–0.23	–0.40	–0.05	–0.27	–0.21	–0.33	0.05	0.41	0.15	–0.33	1.00	
Agricultural SI	0.32	0.16	0.38	0.52	0.57	0.50	–0.43	–0.08	0.18	–0.15	–0.26	1.00

Figure A.IV: Sample and Clusters

Notes: The sample of ethnicities from the *Ethnographic Atlas* employed in the analysis is shown above. Shading is used to represent the 44 clusters.

Table A.I

SUMMARY STATISTICS FOR HISTORICAL ANALYSIS					
Variable	*Mean*	*Std. Dev.*	*Min*	*Max*	*N*
Large Domesticated Animals	0.62	0.48	0.00	1.00	484
Female Participation in Agriculture	0.49	0.50	0.00	1.00	316
Intensive Agriculture	0.32	0.47	0.00	1.00	485
Plow Use	0.06	0.24	0.00	1.00	476
Indigenous Slavery	0.85	0.36	0.00	1.00	444
Centralization	0.33	0.47	0.00	1.00	468
Log Population Density (1700) inhab/km^2	0.75	1.97	−6.91	4.89	522
Presence of Urban Center (1800)	0.04	0.19	0.00	1.00	521
Mean Historical Temperature	24.45	3.08	14.69	29.45	522
Mean Historical Relative Humidity	57.52	14.60	20.89	83.96	522
Longitude	17.82	15.76	–17.00	48.00	522
Absolute Latitude	9.75	7.48	0.00	37.00	522
Log Mean Altitude (meters)	5.19	1.44	–3.71	7.46	522
Agricultural Suitability Index	0.53	0.20	0.00	0.91	522
Coast	0.15	0.36	0.00	1.00	522
River	0.57	0.50	0.00	1.00	522
Malaria Ecology Index	13.67	9.58	0.00	34.49	522
Malaria Prevalence (1900)	3.84	0.90	0.00	5.00	522
Historical TSI (1871)	0.00	1.00	–3.16	1.49	522
Fraction of Tribal Land TseTse Infested (1973)	0.00	1.00	0.62	0.43	522

Table A.I

SUMMARY STATISTICS FOR HISTORICAL ANALYSIS					
Variable	*Mean*	*Std. Dev.*	*Min*	*Max*	*N*
Large Domesticated Animals	0.62	0.48	0.00	1.00	484
Female Participation in Agriculture	0.49	0.50	0.00	1.00	316
Intensive Agriculture	0.32	0.47	0.00	1.00	485
Plow Use	0.06	0.24	0.00	1.00	476
Indigenous Slavery	0.85	0.36	0.00	1.00	444
Centralization	0.33	0.47	0.00	1.00	468
Log Population Density (1700) inhab/km^2	0.75	1.97	−6.91	4.89	522
Presence of Urban Center (1800)	0.04	0.19	0.00	1.00	521
Mean Historical Temperature	24.45	3.08	14.69	29.45	522
Mean Historical Relative Humidity	57.52	14.60	20.89	83.96	522
Longitude	17.82	15.76	−17.00	48.00	522
Absolute Latitude	9.75	7.48	0.00	37.00	522
Log Mean Altitude (meters)	5.19	1.44	−3.71	7.46	522
Agricultural Suitability Index	0.53	0.20	0.00	0.91	522
Coast	0.15	0.36	0.00	1.00	522
River	0.57	0.50	0.00	1.00	522
Malaria Ecology Index	13.67	9.58	0.00	34.49	522
Malaria Prevalence (1900)	3.84	0.90	0.00	5.00	522
Historical TSI (1871)	0.00	1.00	−3.16	1.49	522
Fraction of Tribal Land TseTse Infested (1973)	0.00	1.00	0.62	0.43	522

Table A.II

AES coefficients

Grouping	Base	Tropics Only
Agriculture	–0.343***	–0.386***
	(0.042)	(0.062)
Urbanization	–0.227***	–0.178***
	(0.076)	(0.066)
Institutions	–0.217***	–0.219***
	(0.053)	(0.055)

Notes: Columns give AES estimates for the base and tropics only subsample specifications. The AES averages the normalized treatment effects obtained from a seemingly unrelated regression in which each dependent variable is a subheading in Table I. *Agriculture* includes the binary dependent variables: large domesticated animals plow use, female participation in agriculture and intensive agriculture. *Urbanization* includes the ln population density in 1700 and the presence of a city in 1800 and *Institutions* includes indicator variables for political centralization and the practice of indigenous slavery. Standard errors are clustered at the level of cultural province. The control group is defined as all ethnic groups with a TSI<0. The signs of female participation and indigenous slavery are reversed. * ** *** Significant at 10, 5 and 1 percent levels.

Table A.III

SENSITIVITY ANALYSIS

Dependent Variable	(1)	(2)	(3)	(4)
	ψ=0.40	ψ=0.80	ψ=1.20	ψ=1.60
Agriculture				
Large Domesticated Animals	–0.189***	–0.203***	–0.216***	–0.224***
	(0.039)	(0.039)	(0.049)	(0.041)
Intensive Agriculture	–0.075**	–0.084***	–0.086**	–0.084**
	(0.032)	(0.028)	(0.027)	(0.029)
Plow Use	–0.041**	–0.054**	–0.065**	–0.076***
	(0.019)	(0.021)	(0.020)	(0.021)
Female Participation in Agriculture	0.182***	0.186***	0.193***	0.193***
	(0.052)	(0.052)	(0.055)	(0.058)
Urbanization				
Log Population Density (1700)	–0.548**	–0.579**	–0.608**	–0.615**
	(0.203)	(0.217)	(0.226)	(0.229)
Presence of Urban Center (1800)	–0.030**	–0.033**	–0.033**	–0.030*
	(0.014)	(0.015)	(0.016)	(0.016)
Institutions				
Indigenous Slavery	0.097**	0.107***	0.155***	0.119***
	(0.037)	(0.037)	(0.038)	(0.041)
Centralization	–0.068**	–0.078**	–0.081**	–0.081**
	(0.031)	(0.031)	(0.032)	(0.034)

Notes: OLS estimates of Equation (9). The dependent variable is noted in the leftmost column. Each cell is the coefficient on the historical TSI as the the slope parameter ψ in Equation (3) is varied over the feasible range (May et al., 1974). All specifications include temperature, relative humidity and the first-order interaction between temperature and humidity as well as the log mean altitude, the FAO's agricultural suitability index, access to waterways, the malaria ecology index developed by Kiszewski et al., (2004), absolute latitude and longitude. Robust standard errors clustered at the level of cultural province in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table A.IV
REDUCED FORM ESTIMATES OF THE RELATIONSHIP BETWEEN CEREAL CULTIVATION, MILKING AND TSETSE SUITABILITY

Dependent Variable	(1)	(2)	(3)
Cereal	–0.143**	–0.133**	–0.135**
	(0.063)	(0.061)	(0.064)
Milking	–0.125**	–0.148***	–0.186***
	(0.046)	(0.045)	(0.032)
Climate controls	Y	Y	Y
Malaria control	N	Y	Y
Geography controls	N	N	Y

Notes: OLS estimates of Equation (9) are reported. The dependent variables are listed in the leftmost column of the upper panel and include whether the ethnic group practiced cereal cultivation and whether the ethnicity milked their livestock frequently. Climate controls refer to temperature, relative humidity and the first-order interaction between temperature and humidity. Malaria refers to the malaria ecology index developed by Kiszewski et al., (2004). Geography controls include log mean altitude, the FAO's agricultural suitability index, longitude, absolute latitude and access to waterways. Coefficients are reported with robust standard errors clustered at the level of cultural province in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table A.V
BALANCED PANEL

Dependent Variable	(1)	(2)
Agriculture		
Large Domesticated Animals	–0.214***	–2.053***
	(0.046)	(0.731)
Intensive Agriculture	–0.095**	–0.910**
	(0.037)	(0.440)
Plow Use	–0.073***	–0.699**
	(0.026)	(0.314)
Urbanization		
Log Population Density (1700)	–0.552**	–5.300*
	(0.241)	(3.071)
Presence of Urban Center (1800)	–0.035**	–0.333*
	(0.016)	(0.188)
Institutions		
Indigenous Slavery	0.106***	1.020***
	(0.038)	(0.392)
Centralization	–0.096**	–0.923*
	(0.045)	(0.512)
Num Obs.	381	381
F-stat	N.A.	11.32

Notes: The sample is balanced and includes all outcomes except female participation in agriculture. (Including female participation in agriculture reduces the sample size to 260 though the results are qualitatively similar). Column (1) reports OLS estimates of Equation (9) and Column (2) reports two-stage least squares estimates. The fraction of tribal area TseTse-infested is instrumented with the historical TSI. The dependent variable is noted in the upper panel leftmost column of the upper panel. All specifications include temperature, relative humidity and the first-order interaction between temperature and humidity as well as log mean altitude, the FAO's agricultural suitability index, access to waterways, the malaria ecology index developed by Kiszewski et al., (2004), absolute latitude and longitude. Standard errors clustered at the level of cultural province in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table A.VI
ROBUSTNESS TEST FOR THE EXPORT SLAVE TRADE

Dependent Variable	(1)
Urbanization	
Log Population Density (1700)	–0.634***
	(0.228)
Presence of Urban Center (1800)	–0.035**
	(0.016)
Institutions	
Indigenous Slavery	0.112***
	(0.041)
Centralization	–0.084**
	(0.033)

Notes: OLS estimates of Equation (9). The dependent variable is noted in the leftmost column. All specifications include temperature, relative humidity and the first-order interaction between temperature and humidity as well as log mean altitude, the FAO's agricultural suitability index, access to waterways, the malaria ecology index developed by Kiszewski et al., (2004), absolute latitude and longitude. In addition, log {(slave exports+1)/land area} by ethnicity from Nunn and Wantchekon (2011), are also included. Robust standard errors clustered at the level of cultural province in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table A.VII
VIRTUAL COUNTRIES

Dependent Variable	(1)
Agriculture	
Large Domesticated Animals	–0.118**
	(0.051)
Intensive Agriculture	–0.148***
	(0.048)
Plow Use	–0.040
	(0.038)
Females in Agriculture	0.126*
	(0.065)
Urbanization	
Log Population Density (1700)	–0.504*
	(0.289)
Presence of Urban Center (1800)	–0.043*
	(0.023)
Institutions	
Indigenous Slavery	0.110**
	(0.046)
Centralization	–0.103*
	(0.061)
Num Obs.	133

Notes: OLS estimates of Equation (9) shown. The sample is comprised of virtual countries. Each country is the shape of a square approximately 160,000 kilometers^2 in size. Ethnic group outcomes are averaged within the virtual country boundaries. All regressions include temperature, relative humidity and the first-order interaction between temperature and humidity as well as log mean altitude, the FAO's agricultural suitability index, access to waterways, the malaria ecology index developed by Kiszewski et al., (2004), absolute latitude and longitude. Standard errors in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table IV
COMPARISION OF MURDOCK MAP AND THIESSEN POLYGON COEFFICIENTS

Dependent Variable	(1) Map	(2) Polygon	(3) Difference
Agriculture			
Large Domesticated Animals	–0.219***	–0.274***	0.06
	(0.040)	(0.085)	
Intensive Agriculture	–0.087***	–0.184***	1.51
	(0.028)	(0.057)	
Plow Use	–0.067***	–0.088*	0.43
	(0.021)	(0.044)	
Female Participation in Agriculture	0.195***	0.338***	–1.31
	(0.055)	(0.094)	
Urbanization			
Log Population Density (1700)	–0.615**	–0.947***	0.92
	(0.228)	(0.282)	
Institutions			
Indigenous Slavery	0.117***	0.173**	–0.80
	(0.039)	(0.058)	
Centralization	–0.082**	–0.055	–0.30
	(0.033)	(0.081)	

Notes: OLS estimates of Equation (9) using the Murdock *Map* sample in column (1) and the Thiessen Polygons sample in column (2). The dependent variable is noted in the leftmost column. All specifications include temperature, relative humidity and the first-order interaction between temperature and humidity as well as the log mean altitude, the FAO's agricultural suitability index, access to waterways, the malaria ecology index developed by Kiszewski et al., (2004), absolute latitude and longitude. Coefficients are reported in columns (1) and (2) with robust standard errors clustered at the level of cultural province in parentheses. The *t*-statistic on the difference between the two coefficients is reported in column (3). * ** *** Signficant at 10, 5 and 1 percent levels.

Table A.IX
ADDITIONAL PLACEBO TESTS

Panel A: Dropping Europe

Dependent Variable	(1) Main Effect TSI (β)	(2) Africa Interaction TSI (δ)	(3) Africa Total TSI (β+δ)
Agriculture			
Large Domesticated Animals	0.030	–0.304***	–0.275***
	(0.032)	(0.067)	(0.054)
Intensive Agriculture	–0.039	–0.212***	–0.251***
	(0.044)	(0.051)	(0.013)
Plow Use	–0.021	–0.061	–0.083**
	(0.020)	(0.039)	(0.035)
Female Participation in Agriculture	–0.098**	0.363***	0.266***
	(0.044)	(0.071)	(0.053)
Urbanization			
Log Population Density (1700)	–0.157	–0.823***	–0.980***
	(0.264)	(0.313)	(0.192)
Institutions			
Indigenous Slavery	–0.079	0.231***	0.151***
	(0.066)	(0.081)	(0.052)
Centralization	–0.024	–0.106**	–0.130***
	(0.037)	(0.044)	(0.028)

Panel B: Africa to Other Tropical Areas

Dependent Variable	(1) Main Effect TSI (β)	(2) Africa Interaction TSI (δ)	(3) Africa Total TSI (β+δ)
Agriculture			
Large Domesticated Animals	–0.064	–0.195**	–0.259***
	(0.050)	(0.078)	(0.060)
Intensive Agriculture	–0.085	–0.155**	–0.239***
	(0.056)	(0.062)	(0.019)
Plow Use	0.032	–0.110*	–0.068**
	(0.041)	(0.054)	(0.033)
Female Participation in Agriculture	–0.038	0.292***	0.254***
	(0.072)	(0.093)	(0.061)
Urbanization			
Log Population Density (1700)	–0.330	–0.620*	–0.950***
	(0.348)	(0.365)	(0.189)
Institutions			
Indigenous Slavery	0.016	0.148	0.158***
	(0.081)	(0.098)	(0.051)
Centralization	–0.001	–0.116***	–0.118***
	(0.054)	(0.064)	(0.031)

Notes: OLS estimates of Equation (10). The dependent variable is noted in the leftmost column. The sample in Panel (A) includes all continents except Europe. The sample in Panel (B) compares Africa to other tropical parts of Asia, South America and Oceania/Australia. Column (1) reports the coefficient on the TSI. Column (2) reports the coefficient on TSI*Africa. Column (3) reports the coefficient of column (1) + column (2). Standard errors clustered at the language family in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

Table A.X
PERSISTENCE OF HISTORICAL INSTITUTIONS

Dependent Variable is Government Effectiveness (2005)

	(1)	(2)	(3)	(4)	(5)
Panel A: Two-Stage Least Squares					
Observed TseTse (1973)	–0.772**	–0.829**	–0.738**	–0.402	–0.442
	(0.312)	(0.340)	(0.329)	(0.359)	(0.452)
Landlocked		–0.383**	–0.314**	–0.306*	–0.306*
		(0.180)	(0.198)	(0.161)	(0.161)
Log Suitable Land Agriculture (10 km^2)		0.062	0.034	0.056	0.049
		(0.054)	(0.055)	(0.046)	(0.044)
Temperature		–0.036*	–0.045*	–0.040*	–0.043*
		(0.020)	(0.023)	(0.021)	(0.022)
Historical Centralization				0.672***	0.663*
				(0.316)	(0.336)
Historical Indigenous Slavery				–0.521*	–0.533*
				(0.298)	(0.300)
Colonizer and Legal Origins Fixed Effects	N	N	Y	Y	Y
Historical Slave Exports	N	N	N	N	Y
Panel B: First Stage for Observed TseTse (1973)					
TseTse Suitabiliey Index (1961-1999)	0.340***	0.374***	0.405***	0.423***	0.381***
	(0.036)	(0.049)	(0.056)	(0.078)	(0.098)
F-statistic	86.88	57.16	57.35	25.24	22.42
Panel C: Ordinary Least Squares					
Observed TseTse (1973)	–0.531**	–0.518**	–0.423	–0.174	–0.162
	(0.215)	(0.264)	(0.288)	(0.267)	(0.302)
No. obs.	43	43	43	43	43

Notes: Panel (A) reports the two-stage least squares estimates with Government Effectiveness in 2005 as the dependent variable. Government Effectiveness (from Kaufmann et al., 2010) is constructed using perceptions from various stakeholders on the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation and the credibility of the government's commitments to such policies. Panel (B) reports the corresponding first stage. Panel (C) reports the OLS coefficient from a regression of Log GDP per capita on the fraction of a country's land area that is TseTse infested. Control variables are not reported in panels (B) and (C) to save space. Legal Origins are from La Porta et. al., (1999). The sample is either of French or British legal origin. Log {(slave exports+1)/land area} by ethnicity from Nunn (2008) are included in column (6). Coefficients are reported with robust standard errors in parentheses. * ** *** Signficant at 10, 5 and 1 percent levels.

IX.2 Appendix B: Variable Definitions and Sources

Climate Data and the TseTse Suitability Index. The TSI is constructed using global, gridded daily climate variables from the National Oceanic and Atmospheric Administration's 20th Century Reanalysis:

http://www.esrl.noaa.gov/psd/data/20thC_Rean.

The earliest year of available data, 1871, was used in the historical analysis. The single level daily data file for air temperature and relative humidity was used in the analysis. The interested reader is also referred to Compo et al., (2011) for more information on the reanalysis technique. Modern climate data used in the analysis span the period 1961-1990 and are from the East Anglia Climate Research Unit:

http://www.cru.uea.ac.uk/cru/data/hrg/tmc.

The equations used in the model of TseTse fly population growth are provided in the table below:

Table B.I

FORMULAE FOR THE POPULATION GROWTH MODEL OF THE TSETSE FLY

Variable	*Formula*
Birth Rate (B)	(1-(0.0058 Temperature²-.2847 Temperature+3.467))
Saturation Deficit	(6.1078*exp(17.2694*(Temperature)/(Temperature+237)))-((RH/100) (6.1078*exp(17.2694*(Temperature)/(Temperature+237))))
Adult Fly Mortality (M)	(-.0003 (Saturation Deficit)²+0.0236 Saturation Deficit+.235)[£]
Rate of Increase (λ)	max((B-M) ,0)
Steady State Population	$N^* = \left(\frac{\lambda}{\phi}\right)^{\frac{1}{\psi}}$
TseTse Suitability Index	Z-score of N*

[£]Mortality was replaced by its maximum value if the mean tempearture was less than 22 degrees Centigrade to reflect the "chill coma" phenomenon. Temperature is in degrees Celsius. In the baseline model ? =0.025 and ψ=1.25.

Population and Urbanization Data. City location geospatial data are from Chandler (1987) for the year 1800. Cities are defined by Chandler as locations with over 20,000 inhabitants. Gridded population data for the world are available from the History Database of the Global Environment (HYDE, version 3.0) from 1700 to 2000 (Goldewijk 2005). Historical population estimates are obtained through the collection of subnational historical population data and hindcasting to the base year of 1700 with regional population growth rates from (Grigg, 1980) when no other historical data were available. Population density is defined as logarithm (.001+ inhabitants in 1700 per square kilometer).

The gridded population data for the year 2000 used in Equation (12) are taken from LandScan™ `http://www.ornl.gov/sci/landscan`. The LandScan algorithm uses remote sensing spatial data and imagery analysis technologies to disaggregate census counts within an administrative boundary and estimate an ambient population (e.g., an average population over 24 hours).

Geographic Variables: GTOPO 30 is a digital elevation model of the world, developed by the United States Geological Survey (USGS). `http://eros.usgs.gov`. Elevation is calculated in meters. The logarithm of mean elevation per tribal area is used in the analysis. Data on the location of rivers were obtained from the Harvard GIS Center: `http://www.gis.harvard.edu/icb/icb.do`

The malaria ecology index by Kiszewski et al., (2004) is used to approximate the prevalence of severe forms of malaria. This index is derived from an equation relating the human-biting tendency of the mosquito to the daily mortality rate. The parameters for the index are taken from field studies. Biting force is proxied for by the number of mosquitoes in a given area that have evidence of a human bloodmeal and mortality is based on the climatic limits for *Anopheles* survival. There is also an adjustment for the type of mosquito most prevalent in a particular region. A coarser malaria map from the year 1900 is used in a robustness check (Hay et al., 2004).

Agricultural suitability refers to suitability of land for rainfed crops and is accessible from the FAO's *Global Agro-Ecological Zones* (2000) website: `http://www.iiasa.ac.at/Research/LUC/GAEZ/index.htm`. The FAO's measure of agricultural suitability is based on climate, soil and terrain slope constraints. The suitability index is normalized to range from 0 to 1 with higher values indicating greater suitability.

Slave Exports: Tribe-level data are from Nunn and Wantchekon (2011) and measure the

number of slaves taken from each ethnic group between years t-1 and t. Estimates begin in the 1400. The logarithm of 1+ the total slaves exported per ethnic group, normalized by tribal land area, is used as a measure of slave export intensity. Data on country-level slave exports are from Nunn (2008). Both data sets can be found at:
`http://www.economics.harvard.edu/faculty/nunn/data_nunn.`

Government Effectiveness. Government Effectiveness is a standard normal variable that combines perceptions of government officials, NGOs, private industry and individual citizens regarding the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. The data set can be found at: `http://info.worldbank.org/governance/wgi/index.asp.`

Income per Capita. GDP per capita in 2005 is available from Angus Maddison's website: `http://www.ggdc.nl/maddison.` The logarithm of income per capita is used in the analysis.

Precolonial Ethnographic Outcomes. Ethnographic outcomes are from the *Ethnographic Atlas.* The variables used in this analysis and their definitions are provided in the table below:

Table B.II
ETHNOGRAPHIC ATLAS VARIABLE NAMES AND DEFINITIONS

Varible Name (No.)	*Definition*
Centralization (v33)	Indicator variable equal to one if >=2 levels of hierarchy above the local authority
Cereal Cultivation (v29)	Indicator variable equal to one if major crop type is a cereal grain
Cultural province (chapter heading)	Grouping based on common cultural/genealogical attributes and spatial proximity
Female participation in agriculture (v54)	Indicator variable equal to one if females perform the majority of agricultural tasks
Indigenous slavery (v70)	Indicator variable equal to one for incipient/reported/hereditary slavery
Language family (v98)	Linguistic affiliation
Milking (v41)	Indicator variable equal to one for milking more often than sporadically
Plow use (v39)	Indicator variable equal to one for aboroginal plow use
Presence of Draft Animals (v40)	Indicator variable equal to one for presence of bovines, deer, camelids or equines
Intensive agriculture (v28)	Indicator variable equal to one for intensive, irrigated agriculture, and zero for extensive/shifting casual or no agriculture