Axis-orientation and knowledge transmission: Evidence from the Bantu expansion

Arthur Blouin
University of Toronto

May 5, 2021*

This paper empirically examines Jared Diamond's axis-orientation technology transmission hypothesis in the context of Sub-Saharan African agriculture. Consistent with Diamond (1998), societies in southern Africa whose ancestors migrated directly south from west-central Africa - through the rainforest - engage in different types of agricultural activities than societies whose ancestors migrated east - around the rainforest. In particular, those whose ancestors could consistently keep livestock and produce traditional dry-crops throughout their multigenerational migration journey are more likely to engage in these activities today. The differential preference for wet-crop production by those with a rainforest ancestry led to settlements being established in different types of locations. The data suggests that these initial atypical settlement location preferences led to more geographically and culturally isolated societies. This may have disadvantaged these groups with respect to the adoption of any new agricultural goods or processes.

*I would like to thank Oded Galor for very clear and helpful editorial guidance, and two anonymous referees for constructive comments that improved the paper substantially. I am also grateful to Sasha Becker, Loren Brandt, Nick Crafts, Shari Eli, James Fenske, Price Fishback, Oded Galor, Bishnupriya Gupta, Victor Lavy, Rocco Macchiavello, Stelios Michalopoulos, Sharun Mukand, Elias Papaioannou, Tavneet Suri, Anand Swamy, Fabian Waldinger and Chris Woodruff for helpful comments over the life of this project. I also thank seminar and conference participants at EEA, EHS, NEUDC, CEA, Brown University and the University of Warwick. I gratefully acknowledge CAGE financial support. Correspondence: a.blouin@utoronto.ca

1. Introduction

Jared Diamond famously hypothesized that a fundamental determinant of wealth differences across the globe is whether a region is oriented primarily along an east-west or north-south axis (Diamond, 1998). His idea was that because there is less geographic variation along an east-west axis, important information about agricultural products or processes can travel more seamlessly. While there is modest empirical support for the Diamond hypothesis, (Laitin and Robinson, 2011, Laitin et al., 2012, Pavlik and Young, 2019), testing it has been a significant empirical challenge. As a result, big questions regarding mechanisms remain unanswered. For instance, why does the effect persist when information can now travel across the globe so quickly and easily? In this paper, I make some progress on these issues by first showing that (1) knowledge does seem to have historically traveled less seamlessly across more varying land conditions in Africa; (2) the effect can be seen in current agricultural production patterns, and; (3) geographic and cultural isolation may explain this persistence.

I explore all of this within the context of the Bantu expansion, one of the largest ever human migration movements. Bantu migration began along what is now the Nigeria/Cameroon border and followed two primary routes - south and east. Because migration occurred very slowly - slowly enough that land characteristics were not substantially different in any direction, for any single leg of migration - there was no selection into either route (Vansina, 1990). However, over many generations of migration large differences in the suitability of dry crops did emerge. The southern branch migrated through the rainforest in central Africa, an environment much different than Nigeria / Cameroon. The eastern branch, on the other hand, migrated around the rainforest, reaching Africa's east coast, and then veered south. Each leg along this route featured agricultural characteristics much more similar to Nigeria / Cameroon than the rainforest. Both migration branches rejoined after 1000C.E., and currently live in the same regions in some parts of southern Africa.

This paper aims to use the large difference in the ability to produce certain agricultural goods by southern migrants relative to eastern migrants, in order to explore three questions related to Diamond's hypothesis. First, did ancestral differences in crop production influence the inter-generational transmission of knowledge of some modes of agricultural production? If so, does this help to explain crop production variation today? And finally, if cropping patterns today do reflect ancestral migration routes, why?

To touch on the first, I examine the characteristics of the settlement locations of southern migrants upon exiting the rainforest. If knowledge of dry crops was not transmitted across generations when it fell out of use, then we might expect the migrants exiting the rainforest to prefer to settle on land suitable for wet crops. However, those whose ancestors were continuously able to produce both wet and dry crops may have no such preference. Consistent with this hypothesis, I find that ancestors of southern migrants were more likely to settle rainforest-similar land than ancestors of eastern migrants, even controlling for distance to the rainforest, latitude, longitude, and a wide variety of other land characteristics. Furthermore, data that distinguishes historical land use by pasture / crop land reveals that south of the rainforest, ancestors of southern migrants were historically less likely to keep livestock. This is consistent with knowledge loss, since livestock is far less productive in the rainforest because of the tsetse fly.

Perhaps surprisingly, I find similar agricultural production patterns using more detailed contemporary data. In particular, the descendants of southern migrants continue to produce less of the dry-crops that were produced in the Bantu homeland than the descendants of eastern migrants. This is true even controlling for land characteristics, including the suitability of the land for a wide variety of crops. Interestingly, however, they are also less likely to produce both wet and dry crops introduced to Africa post-Bantu settlement. This suggests that the apparent under-production of dry crops by south-Bantu is not the result of selective rainforest migration on the basis of relatively poor dry-crop ability, which might otherwise have been a concern. Instead, the result indicates that something more fundamental about adopting new farming techniques itself was influenced by the rainforest history of some societies.

This brings us to the issue of mechanisms. I explore whether isolation could be one reason for the patterns found in the data, and find some evidence to suggest that the descendants of southern migrants are both more geographically and culturally isolated.[1] This result on insularity is also consistent with the previously mentioned differences in preferences over settlement locations. Upon exiting the rainforest, ancestors of southern migrants settled away from most other societies, due to their particular preferences for land suited to wet crop production. Whether due to this isolation, or due to the underproduction of many agricultural products, the descendants of southern migrants fare worse economically today than

[1]They also have more insular institutions. Meanwhile, an examination of differences in institutions on other dimensions yields no clear patterns, and indeed it is difficult to find any evidence of other possible reasons for the observed low levels of general crop adoption.

their counterparts whose ancestors originally migrated east.

The findings primarily contribute to the literature by establishing empirical support of the Diamond (1998) hypothesis. The evidence to support the hypothesis has to date been weakly positive. For instance, Pavlik and Young (2019) find positive but imprecise correlations between axis-orientation and technology adoption. Laitin and Robinson (2011), Laitin et al. (2012) show that those along a north-south rather than east-west axis are more diverse, which we might expect if technology more easily flows east-west. One advantage of my approach is that by focusing on a particular location and industry, I can be more precise both in the main effects, and with respect to mechanisms.

The second main contribution to the literature relates to the very long-run determinants of development. Notably, Ashraf et al. (2010) show that historical isolation may foster innovation, given a diminished ability to benefit from the technological advances of others. I similarly explore whether isolated societies may be disadvantaged in knowledge adoption. My approach builds on work by Putterman and Weil (2010), who demonstrate that the economic history of a *place*, and the *people* who inhabit that place, are distinctly responsible for economic development. I extend this literature in a few ways. First, my focus on human capital sheds some light on why ancestry matters (e.g. Michalopoulos et al. (2018b)). Second, while much of the literature focuses on the economic history of a place or a society (Hibbs and Olsson, 2004, Olsson and Hibbs Jr, 2005, Comin et al., 2010), I show how a society's ancestry influenced the places they settled.

Finally, the finding that isolation is influenced by ancestral migration routes contributes to a small literature investigating the origins of ethnolinguistic diversity. The first paper in this literature, Michalopoulos (2011), shows that incentives for knowledge transfer leads to less diversity in the long-run. In this case I focus on diversity as a function of migration and settlement, and accordingly share much in common with Ashraf and Galor (2013a,b), who first established the strong relationship between diversity and migration out of Africa. Further advancing the literature on the origins of diversity is Dickens (2021), who establishes that diversity is associated with trade. Blouin and Dyer (2021) show that investments in linguistic convergence are a function of economic trade incentives. Further, Dickens (2018) investigates the nuances of the relationship between diversity and knowledge flow. One contribution of this paper is to demonstrate how ancestral location-specific knowledge can determine settlement patterns.

2. Historical Background

The Bantu homeland is close to what is now the Nigeria-Cameroon border. As early as 3,000B.C.E. Bantu migration began towards east Africa, just north of the rainforest, a route similar agriculturally to the Bantu homeland (Guthrie, 1948). In both the Bantu homeland and along this eastern route, both wet and dry crops were produced and livestock was kept. Just to the north of the eastern route were the Nilo-Saharans, from whom the eastern-migrating Bantu adopted iron. The iron working technology was diffused back along the settlements that arose along the migration route, until they arrived back at to the Bantu homeland between 0–300C.E (Oliver, 1966). The arrival of iron working allowed for a second Bantu migration branch to emerge to the south (figure 1).

Before the arrival of iron, the thicker rainforest terrain to the south was unable to be cleared. But once it was, the land was at least as profitable, largely due to lower population density and higher rainfall. The history literature argues that the land was selected almost entirely based on 2 factors: existing population density and ease of navigation. Suitability or land quality were not factors for individual migrants (Vansina, 1990). This is not because land quality was unimportant, but rather variation in land quality for a single leg of migration was very low considering the migration speeds of the time. Because of this, migrants simply chose the least populated areas that were navigable.

Even if migrants could travel far enough to reach different types of land, the arrival of iron implies that there was no selection into migration routes (Oliver, 1966). Once iron arrived in the Bantu homeland between 0-300A.D. the southern route, which had not previously been navigable, was now easily cleared, free of other farmers. This newly abundant land to the south was the principal reason for differences in routes.

By the time southern migrants reached the rainforest, dry crops and livestock had been very slowly phased-out as they were no longer suitable. Despite the relative abundance of land to the south, the substitution away from dry crops would have initially influenced productivity, and thus population growth and migration speed. “The situation suggests first very slow rate of movement as people were learning about new aquatic habitats followed by a dash once they had achieved mastery of their new environments” (Vansina, 1990).

In 1000C.E. the migration routes both emerged south of the rainforest, and the two groups were once again able to produce the same crops. It may be useful to examine this region in particular, to examine the impact of any lost knowledge on agricultural productivity within contextually similar geographic regions, where both sets of migrants shared land.

One consideration for this empirical exercise should be the extent that other factors caused differences in agricultural production methods. For example, if either branch of migration was more likely to interact with non-Bantu 'natives,' this exposure, not knowledge transmission, could account for differences observed today. This, however, does not appear to have been the case. From the start of the migration south there would not have been any other significant encounters with natives for either group, as the region was only very sparsely populated prior to the arrival of the Bantu. There were some inhabitants in the region, but they had not yet made the transition to agriculture. They were small bands of hunter-gatherers, and would not have impacted the comparatively large groups of Bantu. In fact the history literature has extensively studied whether the migrating Bantu had either scared away or fully absorbed hunter-gatherers with their relatively sophisticated tools (and weapons) and larger, stronger populations.[2] Recent evidence has put this debate to rest, suggesting that the former was much more prevalent; in fact there was very little interaction between indigenous populations and Bantu migrants (Plaza et al., 2004).

Another source of differences could be simple cultural drift. Cultural drift certainly occurred, both prior to - and during - southern migration. For instance, southern migrants could have developed a distaste for various agricultural products at any point after the eastern migration leg began. This taste-based mechanism would still have to reconcile why societies did not adopt newly introduced agricultural products at the same rate, or why they failed to re-adopt livestock. However cultural drift could theoretically have occurred on each of these dimensions, and in a direction that (by chance) was consistent with the Diamond hypothesis. One way to mitigate this concern is by bringing together a large number of outcomes, that all point to one consistent mechanism.

One final historical consideration may be interactions with traders. For instance, eastern migrants were exposed to Asian traders in a way that rainforest migrants were not, which resulted in widespread conversion to Islam (Michalopoulos et al., 2018a). However, societies that migrated south and east both established trading kingdoms. South of the Bantu homeland, there were four main trading centres. For the southern migrants, the main trading centre was the Kongo Kingdom on the coast of what is now the Democratic Republic of Congo. Along the

[2]Only Bantu had settled agriculture at this time, and accordingly could support much larger societies.

eastern migration route, there was a trading centre to the east of what is now Zimbabwe, which was the primary trade hub with south Asian traders, alluded to above. In central Africa there were two main trading centres established, Lunda state and Luba state. These are neighbouring states in central Africa, and it is unclear whether they would have been more likely to serve southern or eastern migrants, though both fall within the homelands of groups with a southern ancestry. Overall, despite the obvious difference in exposure to Asian traders by eastern migrants, the likelihood of encountering foreign traders more generally seems relatively similar. However it is important to keep in mind that the *type* of trader encountered was certainly different. To the extent that south Asian traders may have differentially influenced eastern Bantu to keep agricultural products like livestock, or caused them to be generally more open to technology adoption, the results could be confounded.

3. Data

To examine the Diamond (1998) hypothesis, and the related questions regarding mechanisms and persistence, data from a variety of sources is required, including historical and contemporary population movements and agricultural activity in Africa. The unit of analysis throughout the majority of what follows is a geographic cell, or a 'virtual country' in the words of Michalopoulos (2011), ?. Accordingly, much of the data description in this section focuses on how various sources of data were mapped to a particular geographic cell. The analysis is structured in a way to consider the average differences in various outcomes between groups following different migration paths, so I begin by examining how these migration routes are assigned to cells, and then describe the various outcomes considered throughout the analysis.

3.A. Language Trees

i) Migration Routes and Language Trees Assigning migration routes to cells was accomplished using the ethnolinguistic map of Murdock (1959). Assigning each of Murdock's groups to a geographic cell is trivial. The map was already digitized by Nunn and Wantchekon (2012), and can be simply overlaid and geographically matched to a cell using any geographic information system (GIS) software.

For each group on Murdock's map, the associated language was identified and matched to data from the Ethnologue (Lewis, 2009). The Ethnologue data include the language's position on its respective language tree. The first split in the

Bantu language tree is an east-south split that corresponds to whether the group migrated east or south. That is, the initial migration split among the Bantu generated linguistic drift. The nature of this linguistic drift can still be detected in languages today, so linguistic characteristics can be used to determine the initial migration directions of a language group's ancestors, 2,000 years ago. Once we know the placement on the Bantu language tree for a particular language, we know the migration direction, which can be matched to the Murdock map, and therefore a geographic cell.

An example is presented in figure 2. In the example, the Kissama society is located on the Murdock map. Each cell within the borders determined by this map is assigned to one of either a not-Bantu, or south-Bantu, or east-Bantu value (only cells assigned to south-Bantu and east-Bantu are considered in the analysis). These values are determined using the language trees from the Ethnologue. The language associated with the Kissama is identified and located on the Ethnologue tree. If the language is traced back to a Bantu root, it is assigned either a south-Bantu or east-Bantu value. The language tree differentiates between these values at the second level from the root of the tree. In this case, the Kissama, and all geographic cells within its boundaries, would be assigned to the southern route.

3.B. Agricultural Production

For long-run historical data I use a 1° by 1° geographic cell, while for contemporary data I use a 0.5° by 0.5° cell to reflect the higher degree of certainty in the contemporary data.

Historical data on agricultural production are rare, and to my knowledge do not exist by crop. It is however, easier to come up with estimates of historical livestock use by measuring bone sediments and fragments from archeological digs. This information allows anthropologists to generate rough estimates of how land may have been separated into crop land, livestock land or non-agricultural land. Livestock is particularly interesting in this application because, like dry-crops, livestock is less profitable in the rainforest due to the tsetse fly.

I make use of historical livestock of data, from Klein Goldewijk and Stehfest (2017), which come at a 5' by 5' resolution, and which I aggregated to the 1° by 1° level to match the resolution of other historical data. These data are based on historical sources and anthropological findings, along with probability models based on land quality and proximity to water to make historical estimates of land use and population density. They provide, for each cell, the km^2 of each cell that is devoted to either crop production or pastoral activities.

This information is supplemented with data on contemporary production of the major crops grown in Africa (Leff et al., 2004). In this dataset, the continent is divided into 0.5° by 0.5° cells and for each cell and each major crop, we observe the percentage of the cell devoted to the production of that crop. These data also include a measure of land quality, which is used as a control in each relevant specification. The land quality data also come at the 0.5° by 0.5° level. These 0.5° by 0.5° cells are used as the unit of observation for all tests involving specific crop estimates.

Information on contemporary yields is also of interest to assess the knowledge loss hypothesis. To that end, I employ contemporary crop data from Institute (2019). These data are based on sub-national administrative reports, and additionally serve as a robustness check to the (Leff et al., 2004) data. The Institute (2019) data was processed in the same way as described above.

3.C. Historical Population

I use two independent datasets to measure historical population, McEvedy (1978) and Klein Goldewijk and Stehfest (2017). The McEvedy (1978) data was geocoded for each period from 1C.E. to 1500C.E. (figure A1). It provides population data for various sub-regions of Africa. To get the data at the 1° by 1° cell level, a technique designed specifically to increase the precision of population data by Moon and Farmer (2001), was used to smooth the data. This technique smoothes the data in a two step process, first determining the most likely regions for population using land characteristics, and second smoothing the data to the level desired. The aggregate population count within the initial regions, as defined by McEvedy (1978) is preserved under this method, but the end result is that the data is disaggregated using a well known, and generally accepted methodology. Summary statistics of the resulting measures for each year can be seen in table 1.

The McEvedy (1978) data offers the advantages of having been constructed prior to Diamond's hypothesis being published, and being the most familiar and popular population dataset used in economic history research. However, the big drawback is that the dataset must be processed as described above to achieve the desired resolution, and therefore contains a lot of measurement error. An alternate dataset is available from Klein Goldewijk and Stehfest (2017), which can be used without any disaggregation, but which was published after the Diamond hypothesis was published and is less well known to economic historians.[3] My approach is to examine any overlapping patterns in estimates coming from the two datasets, to mitigate concerns with either data source individually.

The Klein Goldewijk and Stehfest (2017) population data is processed in the same manner as the previously described land-use data from the same source - both are aggregated to a 1° by 1° geographic cell. Of note is that this data produces population figures that are much larger on average than the McEvedy (1978) data (table 1), which should be kept in mind as one interprets the estimates coming from each dataset.

3.D. Crop Suitability

FAO crop suitability data were used to assess selection on land characteristics (figure A2), and as a set of controls throughout the analysis. The data provide a suitability measure for each major crop of Africa, and come at the 0.5° by 0.5° level, which is directly applicable for the contemporary outcomes (in tables 6 and 7). For the historical analysis which is less reliable and therefore typically at the 1° by 1° level, the mean of the 0.5° by 0.5° data was used to aggregate to the 1° by 1° level. In every case the data are based on low inputs and rainfed / non-irrigated growing conditions that is typical of Africa.

4. Empirical Strategy

The empirical strategy exploits ancestral migration direction to estimate the role of inter-generational knowledge loss in historical and contemporary cropping patterns. The identifying assumption throughout is that ancestral migration direction is (conditionally) independent of both historical and contemporary crop production. The primary basis for making this assumption is the assertion, by the history literature, that historical Bantu migration direction was not selected by migrants based on land characteristics. This is perhaps plausible, if migration was extremely slow, as is also suggested by the history literature. For instance, if migrants were typically travelling only 200 meters across a field, then the direction that they travel would be unlikely to matter much since all land within a 200 meter radius would likely be indistinguishable in terms of productivity characteristics. Nevertheless, this assumption is empirically assessed towards the end of this section.

[3] One concern is that Diamond's theory may have informed historical population estimates in the generation of the data.

4.A. Empirical Specification

Assuming conditional independence for the time being, the following specification would produce causal estimates of the average effect of a rainforest ancestry on those who emerged south of the rainforest, and live in southern Africa:

$$Y_i = \beta_0 + \beta_1 RF_i + \Gamma X_i + \epsilon_i \tag{1}$$

RF_i represents a rainforest ancestry. X_i is a vector of controls that includes distance to the rainforest, distance to the nearest market, latitude and longitude, suitability for all major crops, both the mean and standard deviation of elevation, the standard deviation of land quality for agriculture, the long run variability in temperature (Giuliano and Nunn, forthcoming) and an index for malaria. In some columns I additionally control for the mean land quality and a Herfindahl-Hirschman Index (HHI) of ecological regions. These are added separately due to some concern that selection may have taken place on either of these dimensions. The coefficient of interest is β_1 which identifies the effect of a rainforest ancestry on whichever outcome (Y_i) is being considered (see section 4.B). The main outcomes that I consider are the historical use of the land for livestock and the contemporary production of various crops that are both commonly produced in Africa, and for which data exists. We can estimate this regression for the set of people living together south of the rainforest. In this case (and throughout the analysis) I focus on the region south of 9°S (see figure 3 for a map of Bantu settlement intensity within this region), and only consider regions settled by those with Bantu ancestry.

4.B. Construction of Main Outcomes

The paper investigates five main outcomes: (1) historical land settlements; (2) historical reliance on livestock; (3) isolation; (4) contemporary crop production; and (5) contemporary mean night lights. All but (1) and (3) are constructed in quite a straightforward manner, either relying on the raw data or some aggregation (either mean or sum, as appropriate). For the outcomes whose construction are straightforward, I describe them as they are analyzed. For the variables that require more processing to facilitate interpretation, I describe how they are defined in this section.

i) Land Characteristics The first outcome of interest is settled land characteristics. To measure the land characteristic of interest, I construct an index for the similarity of each cell to a typical cell in the rainforest. The index is:

$$\sigma_i = C \cdot \max\{s_{ic}\} - \sum_{c}^{C} |s_{ic} - \bar{s_{cr}}| \tag{2}$$

Where s_{ic} represents suitability of crop c in cell i for each of the crops under consideration, and $\bar{s_{cr}}$ denotes the mean suitability of crop c in a rainforest cell. The index takes the maximum theoretical value of the second term ($C \cdot \max\{suit_{ic}\}$) and subtracts from that the sum of the absolute difference between the suitability of any crop in the cell of interest from the average suitability of that same crop in the rainforest. This provides a measure of the similarity of any given cell to the average rainforest cell. To facilitate interpretation by avoiding the difficulties in interpreting estimates based on an arbitrarily scaled index, the variable included in regressions is an indicator taking a value of 1 when the index is above it median, and a value of 0 otherwise.[4]

ii) Cultural Isolation The language tree data described in section 3.A can be used to derive a measure of cultural isolation. Using language trees to measure cultural differences is an idea first used by Desmet et al. (2011), and I implement a related measure. I consider the share of overlap in the language tree, and construct the share of the region that is different from group i, weighted by the similarity of the languages to each other. Formally, I construct the following:

$$C_i = \sum_{j}^{N} s_{ij} \cdot (pop_i/pop_j) \tag{3}$$

Where C_i is the cultural connectedness faced by group i. s_{ij} is the similarity between group i and j, pop_i is the population of group i and pop_j is the population of group j. In this case there are N groups in the region, including group i. I also consider a more traditional diversity measure, the HHI, which is simply:

$$D_j^{HHI} = 1 - \sum_{i}^{N} \theta_{ij}^2 \tag{4}$$

Where θ_{ij} is the share of group i in region j. This measure can be interpreted

[4]The specification was tested as a binary and a continuous variable, and it makes little difference to the direction or precision of the estimates.

as the odds that a random encounter is between two people of a different ethnic group, regardless of how similar.

In addition to these two measures of diversity, I also examine the number of links in the language tree from proto-Bantu. This measure differs somewhat by providing some suggestive evidence related to the experience of the southern Bantu during migration through the rainforest. The idea here is that when a group is geographically isolated - in the sense that they have few neighbours to trade with to facilitate independently supporting themselves - the less easy it would be to break away from a main group. So more isolation implies fewer links.

Finally, I examine the cultural borrowing measure from Blouin and Dyer (2021). This outcome measures the intensity of loanwords in a particular language, which has been traditionally interpreted as a proxy for cultural borrowing by both linguists and historians. We would expect more isolation to result in fewer loanwords in a particular language. Importantly, the measure taken from Blouin and Dyer (2021) is one that relies *only* on regional borrowing from local neighbours, and excludes all borrowing from colonial / European languages.

4.C. How plausible is the main identifying assumption?

The identifying assumption that is required is that the initial migration direction is independent of potential agricultural outcomes, conditional on the controls described in the empirical specifications above. In other words, it cannot be that the descendants of rainforest migrants would have produced less livestock or dry-crops even in the absence of ancestral migration. This might be true, for example, if those that were worst at dry-crops or livestock selected into rainforest migration. The primary argument against selection into rainforest migration is that migration initially happened extremely slowly - slowly enough that there were no observable differences between travelling east and south for any single leg of migration. Accordingly people travelled south because land was more abundant, but over many many generations the unobservable (to the migrant) differences that did exist amounted to a substantial difference in the ability to produce dry-crops and keep livestock. To try to assess this story requires information both on the differences in land characteristics and on the speed of migration. This section looks at evidence of both in an attempt to empirically validate claims made in the history literature.

To this end, I first start with migration data, to estimate migration speeds, and then move on to differences in land characteristics at the estimated differences in single-leg migration distances. Population data exists at 200-year intervals, so I begin by estimating the speed of the migration frontier for both eastern and southern migrants to ensure consistency with the historical account.

The first step in the analysis is to define the migration frontier. To do so I assign a cell to the migration frontier if in any 200-year period it has an initial population of less than 350 people per 1° x 1° cell, and more than 350 in the post-period.[5] This threshold is based on the maximum value of the population density variable in the Binford hunter-gatherer dataset, which implies a maximum possible hunter-gatherer population of just over 350 people per cell Binford (1980). The idea is that populations over that amount must represent settled agriculture, which only Bantu groups engaged in at the time.[6] I combine this information with a distance from Bantu-homeland variable, to scale the frontier variable in a more easily interpretable way. This produces a variable that may be of interest, as follows:

$$Frontier_{ct} = \mathbb{1}(p_{c,pre} < 350) \cdot \mathbb{1}(p_{c,post} > 350) \cdot d_c \tag{5}$$

Where p is population, d is distance to the Bantu homeland, c is a geographic cell, and *pre* / *post* to represent the first or second time period in any 200-year time interval in the data. With a migration frontier variable in hand, I estimate the speed of migration throughout the Bantu expansion (details of the estimation procedure, and a description of results appear in Appendix B). Interestingly, the pattern in the estimates are very similar to the historical account quoted in section 2. In both datasets we can see the "first very slow rate of movement as people were learning about new aquatic habitats followed by a dash once they had achieved mastery of their new environments" (Vansina, 1990).

The estimates in Appendix B suggest that the eastern route was travelling about 193km per 200 years, while the southern route was travelling about 100km per 200 years. The high end estimate has the eastern route around 400km per 200 years, and the southern route around 700km per 200 years. If we conservatively put a 20-40 year time gap on subsequent generations, this implies a generation migrates between about 2km and 35km. Since we are working with data that is 1° x 1° (about 100 x 100km), I can - at best - make comparisons across about 3 generations. In other words, comparing differences in land characteristics between immediate cell neighbours, represents a very conservative exercise, since it

[5] A histogram of the population per square d.d. is in figure 4.

[6] The 10th percentile of the population distribution, conditional on any inhabitants is 305 according to the Klein Goldewijk and Stehfest (2017) data.

accounts for about three legs of migration at the estimated speeds.

At these migration speeds, I can comfortably examine the hypothetical difference in land characteristics if a migrant went east versus south for the set of all cells near the Bantu homeland. A visual inspection is displayed in figure 5. The top left cell shown in each panel of figure 5 represents the approximate location of the Bantu homeland in south-western Cameroon. The colour of each cell represents the land characteristic to the east minus the same characteristic to the south. The east is defined as any of the three immediately adjacent cells to the right and the south is defined as any of the three immediately adjacent cells below. I look at three main measures: in panel (a) overall land quality; in panel (b) dry land quality; in panel (c) wet land quality. The figures show no clear pattern for each, suggesting that there are no obvious observable differences across such small distances.

The more precise exercise appears in table 2, which suggests a similar conclusion. In the table I show the mean land quality to the east and the mean land quality to the south, and run a t-test for various definitions of 'entering the rainforest.' I find surprising balance in the land characteristics to the east and south, suggesting that there is little to no basis for selection on land quality, wet crop quality or dry crop quality at the estimated migration speeds. This empirical pattern reinforces the historical claims that selection effects were non-existent, and lends some credence to the idea that conditional independence may, in this case, be a reasonable assumption. The one caveat to this is that the available data, both the migration data and the suitability data is fairly imprecise. Both are based on modelling, and involve smoothing, and this inevitably introduces measurement error. In both cases we would expect this to bias estimates towards zero. But based on the information available, it does seem reasonable to cautiously proceed with the empirical exercise.[7]

5. Analysis

5.A. *Settlement South of the Rainforest*

Any agricultural data dating back thousands of years may not be as reliable as we would like, so instead, to start, I take a revealed preference approach by examining settlement preferences of eastern relative to southern migrants. If southern mi-

[7]There may be concern about selection on the southern rainforest boundary as well. A similar exercise looking at land characteristics faced by migrants as they potentially exit the rainforest is seen in table A1, and the results are nearly identical.

grants were less likely to settle on dry-land, that might imply that they preferred that particular characteristic less, which would be consistent with lost knowledge of dry crop production.

I therefore examine whether regions inhabited by south-Bantu ended up on land more similar to the rainforest. Estimates based on equation 1 appear in table 3, columns 1 and 2. In column 1 I show the results without controlling for eco-region variability or land quality, since these land characteristics may themselves have been dimensions of selection. I then show the results with these potentially endogenous controls in column 2.[8] It turns out not to matter much, in either case we see a strong and stable positive relationship suggesting that south-migrating Bantu settled land more similar to the rainforest. Note that the table presents specifications that control for distance to the rainforest in each column, so similarity is not mechanically driven by southern migrants simply not travelling as far south. Further, there is actually little heterogeneity in the effect by distance to the rainforest (as evidenced by table A3).[9]

Next, we investigate whether south-migrating Bantu performed better on rainforest similar land when they settled that land. This exercise uses population data and the Murdock map to examine whether sorting on land took place in a way consistent with the hypothesized agricultural knowledge at the time. To investigate this margin, I employ the following regression specification:

$$Population_i = \alpha_0 + \alpha_1 RF_i + \alpha_2 \sigma_i + \alpha_3 RF_i \cdot \sigma_i + \Gamma X_i + \varepsilon_i \tag{6}$$

α_3 is the variable of interest, and may be interpreted as a differential success on rainforest similar land for those with a rainforest ancestry conditional on settling that land. The rest of the variables are all as previously defined. Table 3, columns 3-5 present estimates from equation 6.[10] This is the intensive margin analogue to the extensive margin result in columns 1 and 2. Column 3 and 4 present results with and without the potentially endogenous controls outlined above, and column 5 checks a logged dependent variable since the population data is heavily skewed. The descendants of southern migrants, with skills potentially suited for rainforest crops, were more likely to settle on rainforest-similar land than eastern

[8] See table A2 columns 1 and 2 for regressions that place these variables as outcomes. In each case there appears to be no correlation between the variable and rainforest ancestry.

[9] As the table demonstrates, this is true for any of the historical outcomes examined in the paper.

[10] Robustness to using both the Klein Goldewijk and Stehfest (2017) data and the McEvedy (1978) data, as well as using different time-periods can be seen in table A4.

migrants who may have had more diversity in production skills, and therefore more profitable land options available to them.

5.B. Historical Agricultural Production

To relate these population patterns to low agricultural adoption requires an investigation into historical agricultural practices. Accordingly, this section explores land historically devoted to livestock. Livestock is not heavily kept throughout the rainforest because of the tsetse fly, but is kept outside of the rainforest in Africa, and may have been readopted by the descendants of southern migrants after they exited the rainforest between 1000C.E. and 1500C.E.

Table 4 shows estimates of the (re)adoption of livestock by the southern migrating Bantu, estimated with equation 1. Consistent with the Diamond (1998) hypothesis, less land is devoted to livestock by southern Bantu in 1500C.E (table 4, column 1). The estimate is also consistently negative over time (table A5). Again, I show robustness to potentially endogenous land quality and land-type heterogeneity (column 2 and 4 of table 4), and show results using both levels (columns 1 and 2) and logs (columns 3 and 4). The persistent lack of adoption is consistent with the Diamond (1998) hypothesis, and is consistent with the evidence presented in tables 3 and A8.

5.C. Mechanisms: Isolation / Diversity

One reason why societies with a rainforest ancestry might have been less well positioned to re-adopt crops is that they may be more geographically or culturally isolated. This is a hypothesis that has been previously advanced, for instance to explain why innovation is more likely in isolated regions (Ashraf et al., 2010).

In the same way that information seems not to travel efficiently when there are gaps in how useful it is, if there are barriers to information flow - such as geographic or cultural isolation - this could play a role in agricultural production decisions. In light of the land selection evidence already discussed, it seems plausible that initial differences in preferences over land characteristics resulted in southern migrant settlement locations that were further away from other groups. As a result of that distance, information of any type may have been less likely to reach the now more isolated descendants of rainforest migrants.[11]

[11] Another plausible explanation could be that tastes for particular crops develop. Evidence suggests that societies are willing to take-on substantial costs to produce goods they have developed a taste for (Atkin, 2016). The argument against this machanism explaining the observed cropping patterns in the Bantu context is in section Online Appendix D.

To investigate this hypothesis, I again examine again equation 1 but now with measures of cultural isolation as the independent variable. First, I construct the standard HHI used throughout the literature (from equation 4).[12] I supplement this with data on cultural isolation, by looking at a measure of linguistic connectedness (equation 3). Third, I investigate the number of links in the language tree from Proto-Bantu. Finally, I investigate cultural connectedness directly by examining the cultural borrowing measure developed in Blouin and Dyer (2021).

The estimates generated by this empirical model can be seen in table 5. Column 1 presents the estimates from a model with the HHI measuring ethnic diversity as the dependent variable. The HHI reflects the probability that two randomly drawn people from the area are from different ethnic groups. So, column 1 suggests that an individual with rainforest ancestry is about 10% less likely to randomly encounter someone from a different ethnic group than an individual without.

Column 2 shows estimates for a similar measure. It examines weighted share of people in a geographic region that speak a similar language to the group. So members of the same language group get a weight of 1, close dialects get a weight close to 1 and very distant languages get a low weight. Estimates using this measure are consistent in showing that groups with rainforest ancestry are less diverse, even when taking cultural distance into account. Next we turn to the number of links from Proto-Bantu. Here we expect fewer links from proto-Bantu to coincide with more isolation. This is precisely what we find in column 3 - south Bantu have about 3/4ths fewer language-tree links. Finally, we examine cultural borrowing directly. Blouin and Dyer (2021) construct data on language borrowing and lending from regional neighbours for all languages for which words data exists (this is just over half of all languages for Africa). Using word-borrowing from other languages, we again find evidence that southern migrating Bantu can be characterized as more culturally isolated.[13]

6. Contemporary Economic Outcomes in Southern Africa

6.A. Contemporary Crop Production

An agricultural production gap that is reinforced by both cultural and geographic isolation suggests reason to expect persistence. The analysis may therefore be

[12]This is accomplished by taking disaggregated population maps and matching them geographically to the Murdoch ethnic homelands map.

[13]Interestingly, southern migrants also adopted institutions that have more insular features than the relatively outward-oriented eastern migrants, despite that other institutional features appear relatively similar (table A6).

amenable to an examination of contemporary data. This is advantageous because contemporary data allows for an analysis of production decisions at much more detailed level. Beyond directly demonstrating the long-run implications of the Diamond hypothesis, this could be also useful as a robustness exercise, because the ability to examine different types of crops separately offers another dimension that may provide some additional clues about the reasons for the adoption decisions being made.[14]

Contemporary crops can be separated into those that were introduced before the Bantu expansion (e.g. native African crops or much of the Malaysian crop complex) or after (e.g. the Colombian exchange crops). Among the traditional crops, dry crops would have had to be adopted by the ancestors of rainforest migrants, but everyone should have had some knowledge of wet crops.[15] All post-expansion crops would have had to be adopted by everyone.

This 2 x 2 division maps directly to Diamond's hypothesis. First, since knowledge of wet-crops likely survived the journey through the rainforestbut knowledge of dry crops did not, we should expect no less production of the wet-traditional crops, but less production of dry-traditional crops. Second is the crops introduced after Bantu settlement of southern Africa, which can help to address two issues. First, if we saw the same pattern (underproduction of dry crops but not wet) then this points to selection on migration as a real concern. After all, it suggests that those with a rainforest ancestry are worse at *all* dry crops than wet crops, regardless of ancestral knowledge. Second, and conditional on the pattern being different for the contemporary crops, they can help to identify the isolation mechanism. If persistence of knowledge loss occurred because isolation hindered any acquisition of agricultural information then we should expect underproduction of all of the newly introduced crops. Finally, if some mechanism other than knowledge acquisition is the issue we might expect similar production of new crops, and underproduction of *only* dry-traditional crops.

I again estimate equation 1, but now since I examine contemporary data, I include country fixed-effects. The estimates appear in Table 6. Columns 1-4 show estimates for yields, and columns 5-8 show estimates for land allocations. The results suggest a perhaps unsurprising - given previous evidence - underproduction

[14]For instance, one reason we might expect that livestock was not readopted is that - despite the arguments put forth by the history literatures - there were in fact strong selection effects, and that today those with a rainforest ancestry simply are not physically adept at certain tasks. Alternatively, it may have nothing to do with good-specific skills or abilities. Instead, it could be that nothing is being shared by neighbours, or adopted by these groups.

[15]See table A7 for a summary of crops considered, their origins, and the timing of their introduction to Africa.

of all crop-types except for pre-Bantu expansion wet crops. The effect is not driven by the land selection result already discussed because it is not a wet-crop effect; new wet crops are significantly underproduced by ancestors of southern migrants (table 6 columns 4 & 7). The yields and land allocation results are consistent with each other. Whenever yields are low, ancestors of southern migrants devote less land to those crops. Interestingly, it does not seem as though southern migrants specialize in wet-traditional crops. While the point-estimate on yields for these crops is positive, indicating that they may do slightly better on these crops, the estimate is not statistically significant.[16]

This is entirely consistent with the evidence previously presented in table 4, albeit with newer, more precise data from a different source, and for entirely different agricultural products that fall into the same conceptual category.[17] Across several different data sources, agricultural products, and time periods, it appears as though southern migrating Bantu both stopped producing wet-traditional crops, selected more isolated land as a result, and failed to catch-up south of the rainforest by learning how best to produce in their new homelands.

6.B. *Light Density*

We might expect that this would influence the economic development of these regions. Accordingly, we now investigate whether the regions settled by the ancestors of southern migrants are less economically developed than the descendants of east-migrating Bantu. Because we are interested in sub-national levels of development, we turn to an analysis of nighttime light density, a well-established method of analyzing subnational economic development in a consistent, accurate way (e.g. Henderson et al. (2011)).

The results appear in table 7, and in particular columns 1-3 demonstrate a significant reduction in light density per person. Column 1 shows a significant reduction in nighttime light density per 1,000 people, and column 2 removes the regions with no light density at all, which may be very sparsely populated. Column 3 adds country-fixed effects, demonstrating that even within countries we see significant reductions in development levels between descendants of southern and eastern migrating Bantu.

This reduced economic development is expected in light of the evidence on the

[16] One remaining concern could be that the locations of production based on ethnic homelands are no longer the locations where people live. To address this, I also match agricultural production to current locations of people in the Afrobarometer. Still, accounting for migration, we find exactly the same patterns in the data (for details see section Online Appendix C).

[17] That is, ones that are unprofitable in the rainforest, but profitable outside of it

poorer productivity and accordingly fewer resources devoted to agriculture, but a number of questions remain. What happens instead of agriculture? Do people urbanize? Do they select out of agriculture altogether, or simply remain less productively in the agricultural sector? For the first, we can again use the nighttime lights data. If people had urbanized *en masse*, we may expect a few very large urban areas and much less wealthy rural areas. In other words, we might a high variance in nighttime lights. In columns 4-5 of table 7 we therefore investigate the variance of light density per 1,000 people. There we find a positive but insignificant effect suggesting that perhaps some of what is happening is urbanization, but the effect is not strong enough to be conclusive.

To dig deeper, we can investigate occupations data. Section Online Appendix E examines self-reported occupations from the DHS, showing that southern Bantu are indeed about 2% less likely to work in agriculture than those with an eastern Bantu ancestry. The evidence reinforces the inability of the descendants of southern Bantu to converge on agricultural production, and is consistent with the evidence on contemporary crop production, as well as the nightlights data.

7. Discussion

Understanding the underproduction of African crops, and particularly New World crops has been a puzzle for development economists for at least the past 25 years (Morris and Byerlee, 1993). Longstanding hypotheses such as Diamond (1998) have historically been able to explain the low use of some seemingly profitable technologies, but seemed unable to explain the underproduction of crops like wheat, which were not introduced along a north-south axis. This paper shows that the Diamond (1998) axis-orientation hypothesis does seem relevant for explaining crop production in Africa, but also provides some suggestive evidence to show that in the specific context of sub-Saharan Africa, axis-orientation may have broader implications for crop production than just the north-south diffusion.

I do this by pulling together information from a number of datasets, to examine the land characteristics of population settlements, migration patterns, historical and contemporary agricultural production, diversity and institutions. This exercise allows me to document empirical patterns consistent with agricultural knowledge loss throughout the multi-generation migration from west-Africa to southern Africa. Lineages that took a migration route through the rainforest would have almost certainly abandoned dry crops that were not suitable, and their descendants in southern Africa (where dry-crops are again suitable) never fully readopted these crops. Part of the reason for this appears to be that their newfound preference for wet-crops led to non-standard preference over settlement locations. This led them to become less diverse and more geographically isolated, leaving them disadvantaged regarding re-adoption. The effect appears to persist through this adoption channel rather than a direct dry-land suitability channel since both wet and dry New World crops are underproduced by descendants of rainforest migrants.

References

Quamrul Ashraf and Oded Galor. Genetic diversity and the origins of cultural fragmentation. *American Economic Review*, 103(3):528–33, 2013a.

Quamrul Ashraf and Oded Galor. The 'out of africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46, 2013b.

Quamrul Ashraf, Ömer Özak, and Oded Galor. Isolation and development. *Journal of the European Economic Association*, 8(2-3):401–412, 2010.

David Atkin. The caloric costs of culture: Evidence from indian migrants. *American Economic Review*, 106(4):1144–1181, April 2016.

Lewis R Binford. Willow smoke and dogs' tails: hunter-gatherer settlement systems and archaeological site formation. *American antiquity*, pages 4–20, 1980.

Arthur Blouin and Julian Dyer. How cultures converge: An empirical investigation of trade and linguistic exchange. *University of Toronto Working Papers*, (691), February 2021.

Fabrizio Colella, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. Inference with arbitrary clustering. 2019.

Diego Comin, William Easterly, and Erick Gong. Was the wealth of nations determined in 1000 bc? *American Economic Journal: Macroeconomics*, 2(3): 65–97, 2010.

K. Desmet, I. Ortuño-Ortín, and R. Wacziarg. The political economy of linguistic cleavages. *Journal of Development Economics*, 2011.

Jared Diamond. *Guns, Germs and Steel: a short history of everybody for the last 13,000 years*. Vintage, 1998.

Andrew Dickens. Population relatedness and cross-country idea flows: evidence from book translations. *Journal of Economic Growth*, 23(4):367–386, 2018.

Andrew Dickens. Understanding ethnolinguistic differences: The roles of geography and trade. *Unpublished manuscript: Brock University*, 2021.

Colin Flight. Malcolm guthrie and the reconstruction of bantu prehistory. *History in Africa*, 7:pp. 81–118, 1980. ISSN 03615413. URL http://www.jstor.org/stable/3171657.

Paola Giuliano and Nathan Nunn. Understanding cultural persistence and change. *Review of Economic Studies*, forthcoming.

Malcolm Guthrie. *The classification of the Bantu languages.* Pub. for the International African Institute by the Oxford Univ. Press, London; New York, 1948.

Vernon Henderson, Adam Storeygard, and David N Weil. A bright idea for measuring economic growth. *American Economic Review*, 101(3):194–199, may 2011. doi: 10.1257/aer.101.3.194. URL https://doi.org/10.1257%2Faer.101.3.194.

Douglas A Hibbs and Ola Olsson. Geography, biogeography, and why some countries are rich and others are poor. *Proceedings of the national Academy of sciences*, 101(10):3715–3720, 2004.

International Food Policy Research Institute. Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 2.0, 2019. URL https://doi.org/10.7910/DVN/PRFF8V.

A. Beusen J.Doelman Klein Goldewijk, K. and E. Stehfest. Anthropogenic land use estimates for the holocene; hyde 3.2. *Earth System Science Data*, 9:927–953, 2017.

David Laitin and Amanda Robinson. The continental axis theory revisited. In *APSA 2011 Annual Meeting Paper*, 2011.

David D Laitin, Joachim Moortgat, and Amanda Lea Robinson. Geographic axes and the persistence of cultural diversity. *Proceedings of the National Academy of Sciences*, 109(26):10263–10268, 2012.

Billie Leff, Navin Ramankutty, and Jonathan A. Foley. Geographic distribution of major crops across the world. *Global Biochemical Cycles*, 18(GB1009), 2004.

Paul M. Lewis. *Ethnologue: Languages of the World.* SIL International, Dallas, Texas, sixteenth edition, 2009.

Colin McEvedy. *Atlas of world population history*. Penguin, Harmondsworth ;;New York [etc.], 1978. ISBN 9780140510768.

Stelios Michalopoulos. The origins of technolinguistic diversity. Economics Working Papers 0095, Institute for Advanced Study, School of Social Science, May 2011. URL http://ideas.repec.org/p/ads/wpaper/0095.html.

Stelios Michalopoulos. The origins of ethnolinguistic diversity. *American Economic Review*, 102(4):1508–1539, jun 2012. doi: 10.1257/aer.102.4.1508. URL https://doi.org/10.1257%2Faer.102.4.1508.

Stelios Michalopoulos, Alireza Naghavi, and Giovanni Prarolo. Trade and geography in the spread of islam. *The Economic Journal*, 128(616):3210–3241, 2018a.

Stelios Michalopoulos, Louis Putterman, and David N Weil. The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa. *Journal of the European Economic Association*, 17(4):1186–1231, 09 2018b. ISSN 1542-4766. doi: 10.1093/jeea/jvy029. URL https://doi.org/10.1093/jeea/jvy029.

ZK Moon and FL Farmer. Population density surface: A new approach to an old problem. *Society & Natural Resources*, 14(1):39–49, January 2001. ISSN 0894-1920. doi: 10.1080/08941920120013.

Michael L Morris and Derek Byerlee. Narrowing the wheat gap in sub-saharan africa: a review of consumption and production issues. *Economic Development and Cultural Change*, 41(4):737–761, 1993.

George Peter Murdock. *Africa : its peoples and their culture history*. McGraw-Hill, New York [u.a.], 1959.

Nathan Nunn and Leonard Wantchekon. The slave trade and the origins of mistrust in africa. *American Economic Review*, 101(7):3221–52, 2012. doi: 10.1257/aer.101.7.3221. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.7.3221.

R. Oliver. The problem of the bantu expansion. *Journal of African History*, 7(3): 361–376, 1966.

Ola Olsson and Douglas A Hibbs Jr. Biogeography and long-run economic development. *European Economic Review*, 49(4):909–938, 2005.

Jamie Bologna Pavlik and Andrew T Young. Did technology transfer more rapidly east–west than north–south? *European Economic Review*, 119:216–235, 2019.

Stéphanie Plaza, Antonio Salas, Francesc Calafell, Francisco Corte-Real, Jaume Bertranpetit, Ángel Carracedo, and David Comas. Insights into the western bantu dispersal: mtdna lineage analysis in angola. *Human genetics*, 115(5): 439–447, 2004.

Louis Putterman and David N. Weil. Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality*. *The Quarterly Journal of Economics*, 125(4):1627–1682, 11 2010. ISSN 0033-5533. doi: 10.1162/qjec.2010.125.4.1627. URL https://doi.org/10.1162/qjec.2010.125.4.1627.

Jan Vansina. *Paths in the rainforests : toward a history of political tradition in equatorial Africa.* University of Wisconsin Press, Madison Wis., 1990. ISBN 9780299125707.

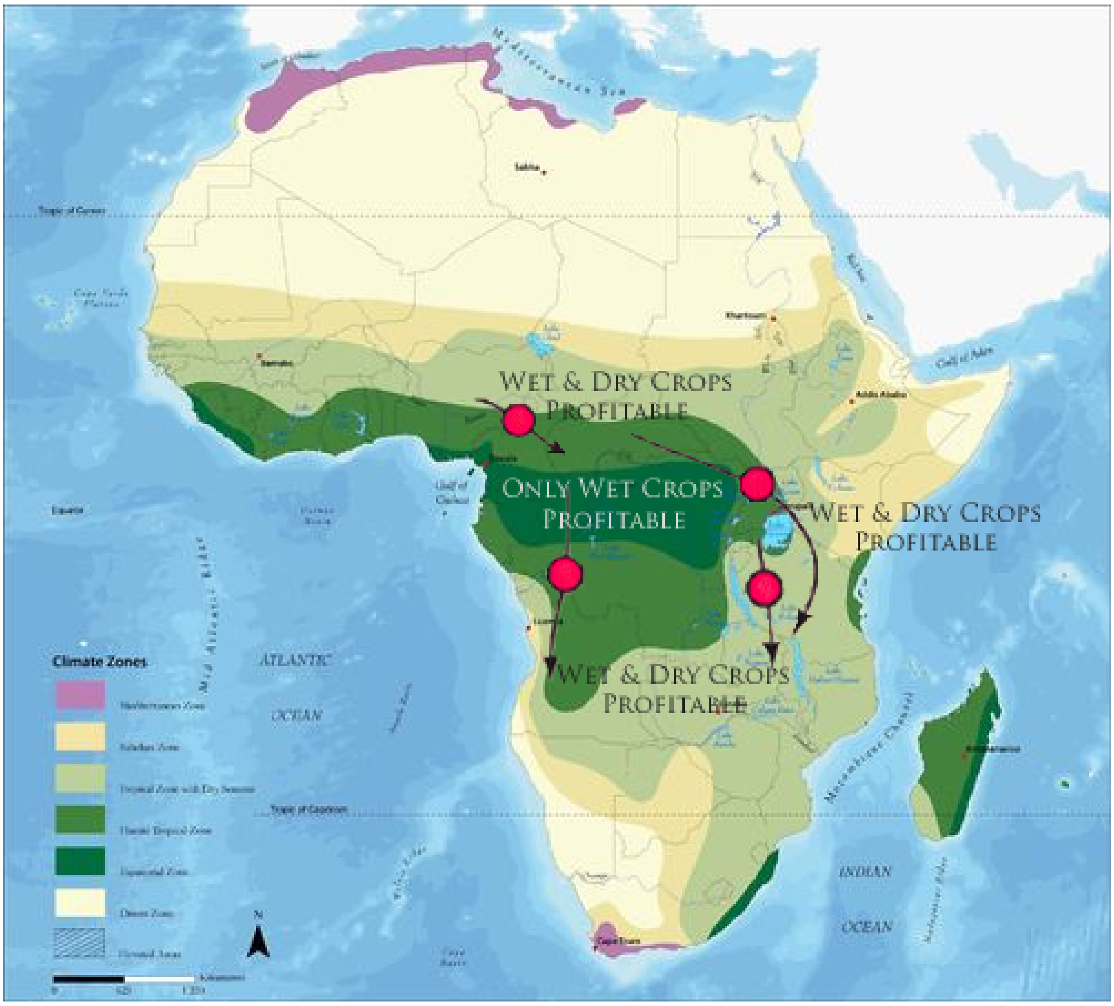


Figure 1: Migration routes during the settlement of Africa. Arrows denote the two branches of the Bantu expansion

Notes: The map shows the two main branches of migration. The branch that goes directly south migrates through the rainforest, while the branch that goes east then south avoids the rainforest. Both migration routes end up in the southern region of Africa which is the study region for much of the analysis in the paper.

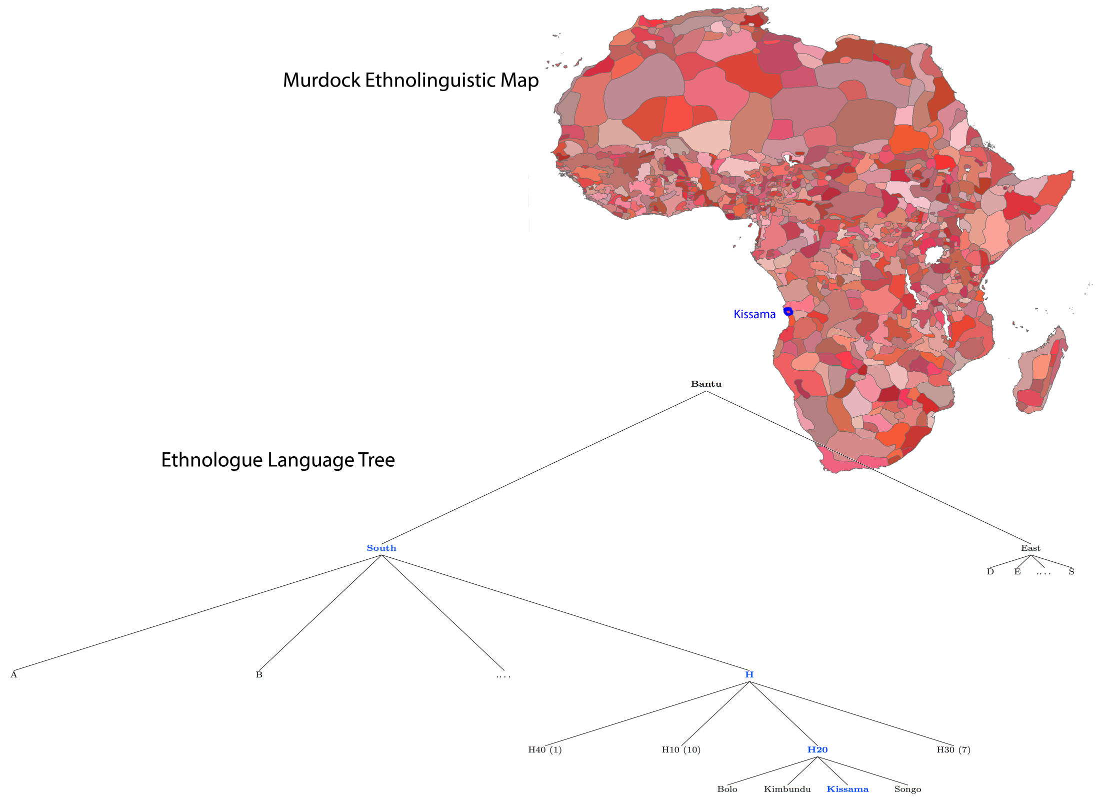


Figure 2: Migration routes during the settlement of Africa. Arrows denote the two branches of the Bantu expansion

Notes: The figure provides an example of how a society is matched to a migration route. Societies are hand-matched to Ethnologue groups, which are matched to a language tree. The language tree is followed back to the point where we observe the east-south split, and the group's location with respect to this split is used to determine the ancestral migration direction.

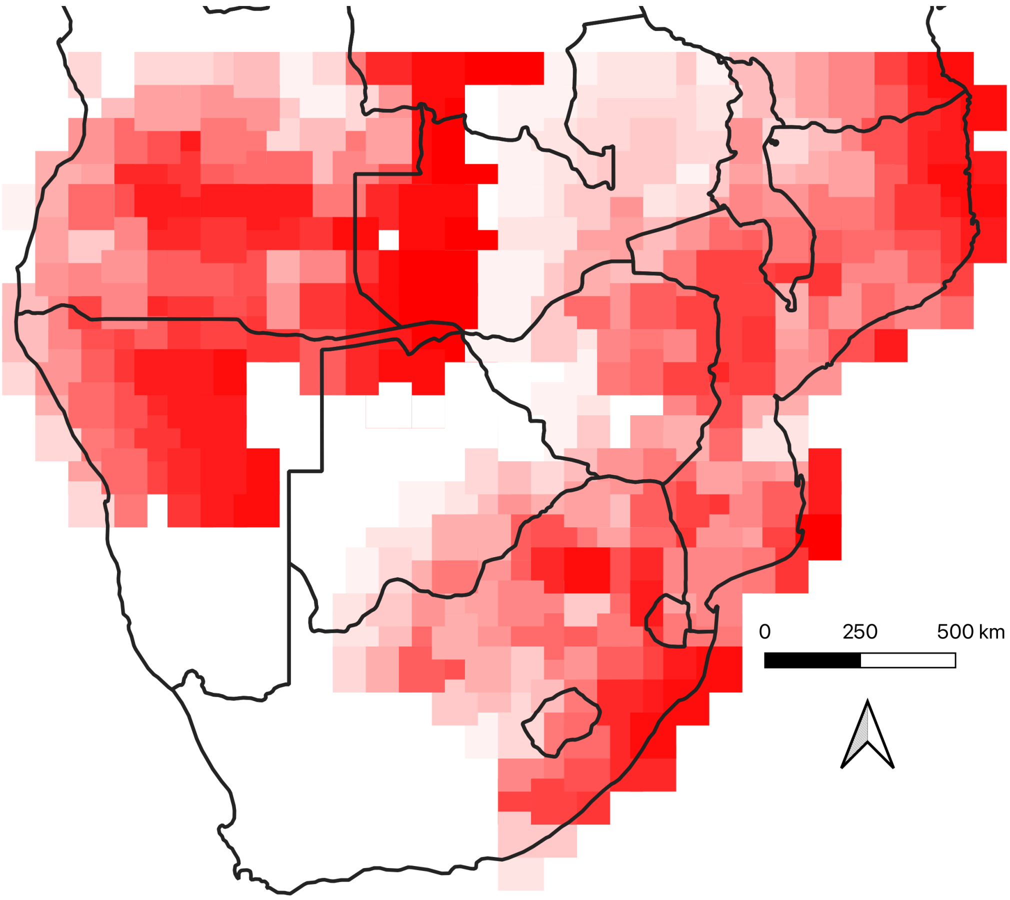


Figure 3: Variation in Bantu settlement location used in the empirical analysis

Notes: The map shows the residual variation in southern Bantu settlement locations to be able to visualize the geographic distribution of the variation exploited in the main empirical analysis. Darker red means more bantu settlement.

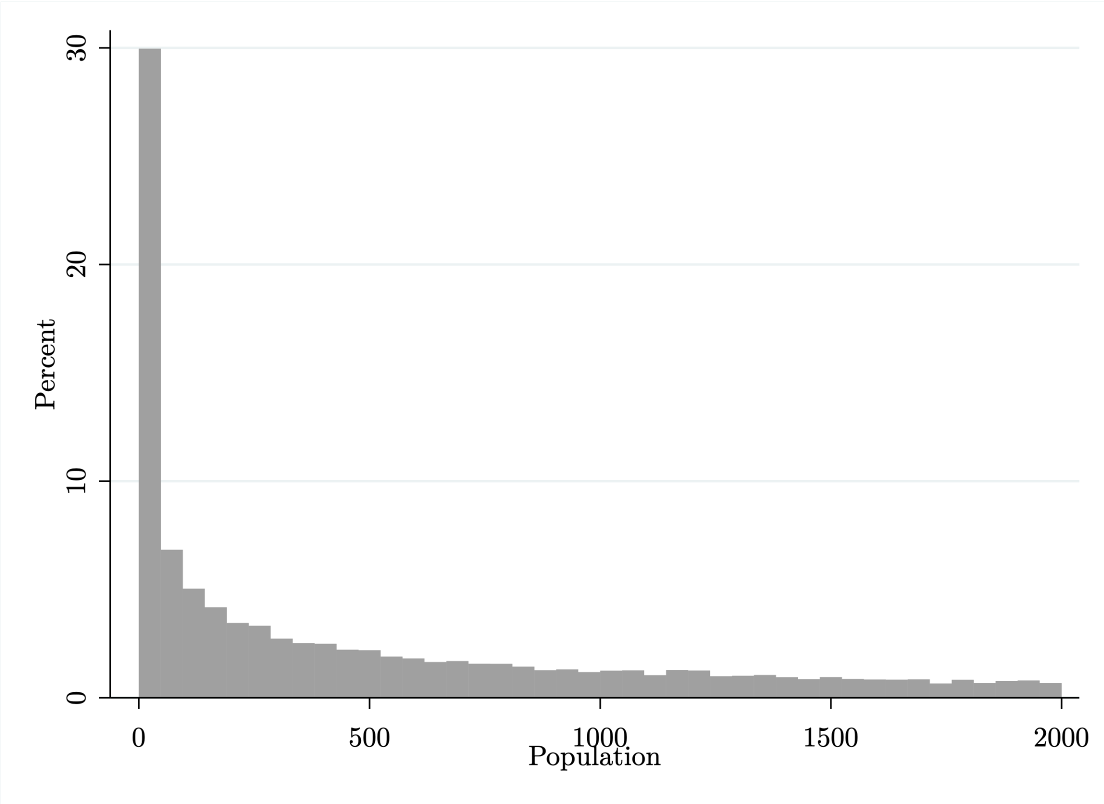
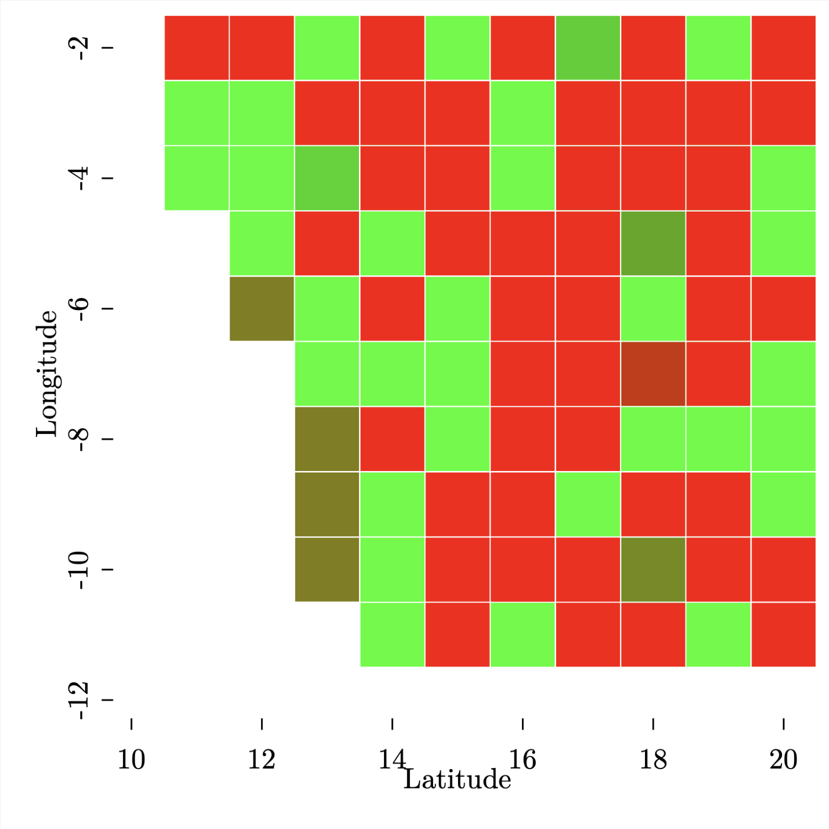
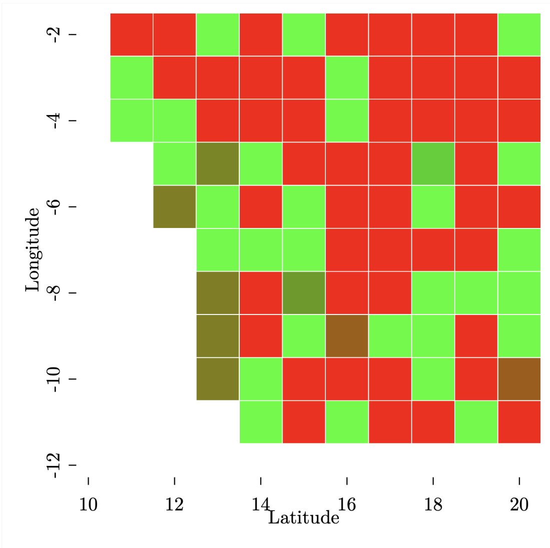


Figure 4: Population histogram (censored at 2,000 people) per square d.d.

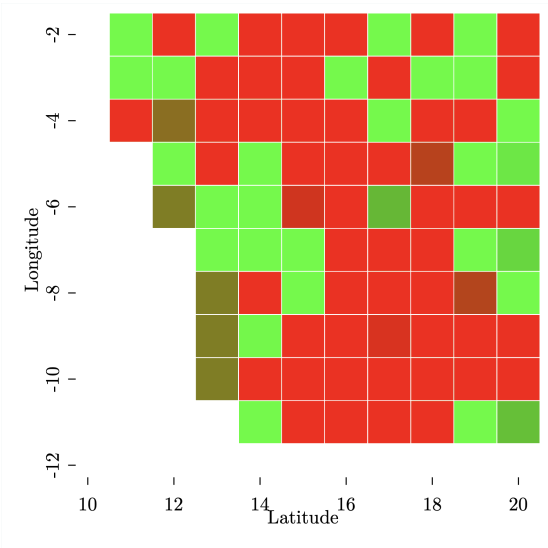
Notes: The figure plots the distribution of population. A 350 person threshold is used to distinguish between hunter-gatherer populations and groups relying on settled agriculture.



(a) Overall Land Quality



(b) Wet Land Quality



(c) Dry Land Quality

Figure 5: Land quality: east minus south (red: east better; green: south better)

Notes: This figure displays relative land quality to the east relative to the south. Each cell shows the land quality on a particular dimension if the migrant travelled to the east relative to if they travelled south. Red indicates positive numbers and green indicates negative numbers, so red indicates that east would be preferred if selection was based on that dimension of land quality and green indicates that south would be preferred if selection was based on that dimension. Subfigure (a) shows patterns in selection on overall land quality; subfigure (b) shows patterns in selection on wet-crop land quality; subfigure (c) shows patterns in selection on dry-crop land quality. Only crops that are traditional to Africa are used to compute relative land quality.

Table 1: Summary Statistics

Sample	N	Mean	St.D	N	Mean	St.D
	Rainforest Ancestors			Eastern Ancestors		
Distance to market (km)	287	489	236	536	377	203
Similarity of land to rainforest	287	0.467	0.499	536	0.287	0.452
Suitability of cereals	287	5.36	1.23	536	5.12	1.63
Suitability of maize	287	5.91	1.96	536	5.61	1.61
Suitability of cotton	287	6.39	1.03	536	5.97	1.63
Suitability of oil	287	5.19	1.32	536	5.25	1.64
Suitability of pulses	287	5.61	1.09	536	5.41	1.60
Suitability of rice	287	6.62	0.99	536	6.63	1.21
Suitability of roots	287	6.41	0.94	536	6.20	1.39
Suitability of sugar	287	7.44	0.71	536	7.29	0.96
Suitability of wheat	287	6.98	0.79	536	7.13	1.31
Land Quality index	287	0.327	0.189	536	0.434	0.222
Pasture Land 1900	287	12	81	536	23	120
Population 1500 (Klein Goldewijk and Stehfest, 2017)	287	11,160	16,718	536	10,005	10,403
Population 1500 (McEvedy, 1978)	287	2,797	2,759	536	4,435	3,950

Notes: Population statistics are for Bantu regions only. Crop production statistics are for the southern region of Africa (south of -9 degrees latitude)

Table 2: Mean differences in land characteristics along southern versus eastern migration directions near northern rainforest boundary

	Observations (1)	Mean to the East (2)	Mean to the South (3)	Difference (4)
		Panel A: Overall Land Quality		
Mean north of 0°	110	8.18	8.87	-0.69
		(0.79)	(0.82)	(0.66)
Mean north of -1°	144	9.34	9.25	0.08
		(0.76)	(0.69)	(0.57)
Mean north of -2°	174	9.82	9.34	0.49
		(0.67)	(0.59)	(0.50)
Mean north of -3°	198	9.59	9.30	0.29
		(0.60)	(0.53)	(0.46)
		Panel B: Dry Land Quality		
Mean north of 0°	110	2.39	2.90	-0.51
		(0.34)	(0.40)	(0.39)
Mean north of -1°	144	2.45	2.68	-0.23
		(0.31)	(0.34)	(0.31)
Mean north of -2°	174	2.70	2.53	0.17
		(0.27)	(0.30)	(0.29)
Mean north of -3°	198	2.59	2.61	-0.02
		(0.25)	(0.28)	(0.26)
		Panel B: Wet Land Quality		
Mean north of 0°	110	5.78	5.96	-0.19
		(0.54)	(0.54)	(0.41)
Mean north of -1°	144	6.89	6.58	0.32
		(0.57)	(0.48)	(0.41)
Mean north of -2°	174	7.12	6.81	0.32
		(0.51)	(0.42)	(0.38)
Mean north of -3°	198	6.99	6.69	0.31
		(0.47)	(0.38)	(0.36)

Notes: *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being east of 20°E and south of 8°E to account only for rainforest boundaries and Bantu settlement locations respectively. Each row reports means for the adjacent cells to the south and to the east, as well as a t-test for the difference between them.

Table 3: Sorting into Regions Similar to African Rainforest

	Rainforest Similar		Population		log Population
	(1)	(2)	(3)	(4)	(5)
South x Rainforest Similar			1,179**	922.4*	0.531*
			(505.0)	(502.2)	(0.309)
South	0.267***	0.261***	250.0	392.0	-0.0188
	(0.0881)	(0.0881)	(559.7)	(561.7)	(0.287)
rainforest similar			33.43	163.1	0.0390
			(361.3)	(360.1)	(0.148)
Distance to rainforest	✓	✓	✓	✓	✓
Distance to market	✓	✓	✓	✓	✓
Suitability of major crops	✓	✓	✓	✓	✓
Latitude	✓	✓	✓	✓	✓
Longitude	✓	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓	✓
Elevation	✓	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓	✓
Herfendal index ecological zones		✓		✓	✓
Mean Land Quality		✓		✓	✓
Observations	823	823	823	823	823
R-squared	0.274	0.276	0.238	0.242	0.450
Dependent Variable Mean	0.350	0.350	3864	3864	8.217

Notes: Standard errors are clustered according to Colella et al. (2019). *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table 4: Historical Land Use

	Historical Land for Pasture			
	Mha/person		log Mha/person	
	(1)	(2)	(3)	(4)
South	-71.72**	-70.28**	-0.708***	-0.686**
	(30.89)	(30.82)	(0.271)	(0.269)
Distance to rainforest	✓	✓	✓	✓
Distance to market	✓	✓	✓	✓
Suitability of major crops	✓	✓	✓	✓
Latitude	✓	✓	✓	✓
Longitude	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓
Herfendal index ecological zones		✓		✓
Mean Land Quality		✓		✓
Observations	823	823	823	823
R-squared	0.176	0.182	0.316	0.327
Dependent Variable Mean	19.19	19.19	0.402	0.402

Notes: Standard errors are clustered according to Colella et al. (2019). *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table 5: Isolation by ancestral migration route

	Fractionalization	Linguistic Similarity	Links from Proto-Bantu	Loanwords from Neighbours
	(1)	(2)	(3)	(4)
South	-0.0953*	0.287***	-0.671***	-0.587**
	(0.0544)	(0.0993)	(0.115)	(0.274)
Suitability of major crops	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓
HHI: ecological zones	✓	✓	✓	✓
Observations	820	823	823	414
R-squared	0.130	0.150	0.460	0.434
Dependent Variable Mean	0.539	0.885	9.295	0.0792

Notes: Standard errors are clustered according to Colella et al. (2019).*** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table 6: Contemporary crop production by ancestral migration route

	Yields				Land allocation (share of land allocated to X)			
	Native crops		New crops		Native crops		New crops	
	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
South	-90.68***	594.8	-985.0**	-231.4***	-0.000674***	-0.000148	-0.00202**	-0.438***
	(31.41)	(461.5)	(407.9)	(63.31)	(0.000211)	(0.000111)	(0.000948)	(0.0571)
Suitability of major crops	✓	✓	✓	✓	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓	✓	✓	✓	✓
Herfendal index ecological zones	✓	✓	✓	✓	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	3,285	3,285	3,285	3,285	3,538	3,538	3,538	3,538
R-squared	0.316	0.557	0.788	0.802	0.342	0.729	0.449	0.651
Dependent Variable Mean	266.9	7371	6972	265.6	0.00292	0.00180	0.0140	1.741

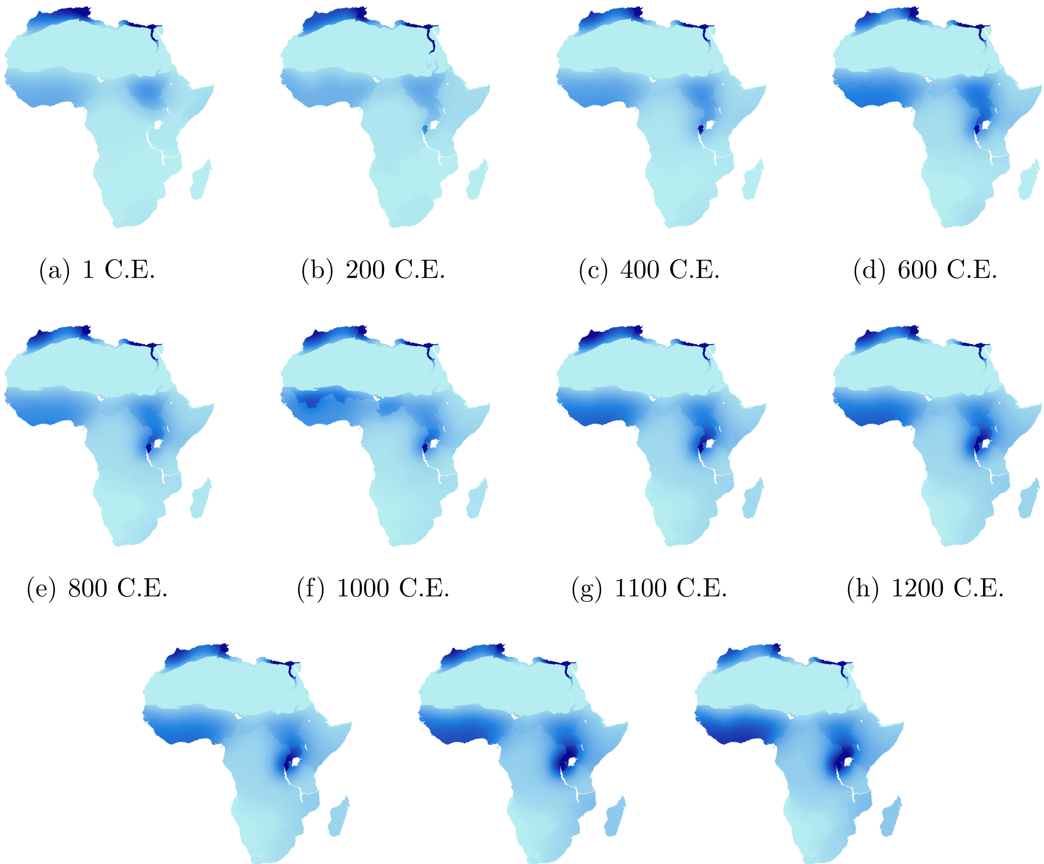
Notes: Standard errors are clustered according to Colella et al. (2019). *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The unit of observation is a 0.5° x 0.5° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table 7: Light density by ancestral migration route

	Light Density per 1,000 people			Variance LD / cap.	
	All	light density > 0			
	(1)	(2)	(3)	(4)	(5)
South	-0.908*	-2.753**	-1.042*	0.474	0.0702
	(0.543)	(1.288)	(0.540)	(0.349)	(0.407)
Suitability of major crops	✓	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓	✓
HHI: ecological zones	✓	✓	✓	✓	✓
Country Fixed Effects			✓		✓
Observations	792	388	272	402	285
R-squared	0.134	0.212	0.162	0.343	0.401
Dependent Variable Mean	0.772	1.577	1.577	2.505	2.505

Notes: Standard errors are clustered according to Colella et al. (2019).*** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Appendix A. Online Appendix

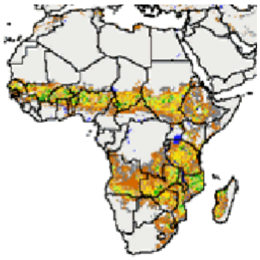


(a) 1 C.E. (b) 200 C.E. (c) 400 C.E. (d) 600 C.E.

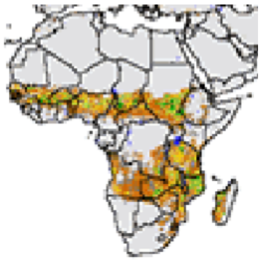
(e) 800 C.E. (f) 1000 C.E. (g) 1100 C.E. (h) 1200 C.E.

(i) 1300 C.E. (j) 1400 C.E. (k) 1500 C.E.

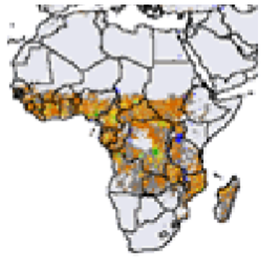
Figure A1: Population data at each time period. Darker regions are more populated



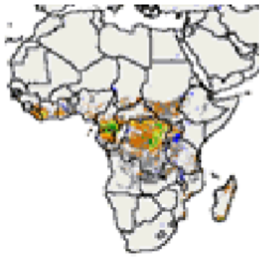
(a) Sorghum



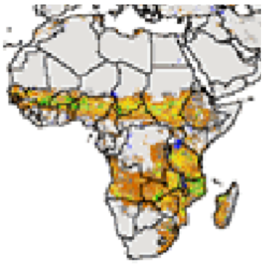
(b) Cotton



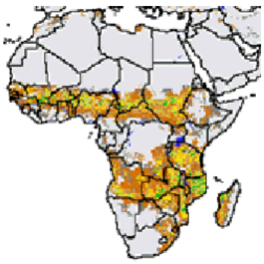
(c) Rice



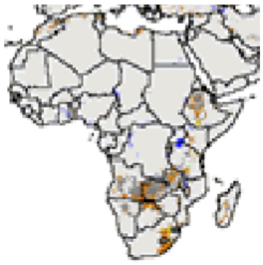
(d) Oilpalm



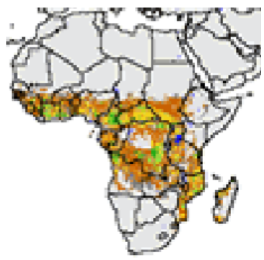
(e) Maize



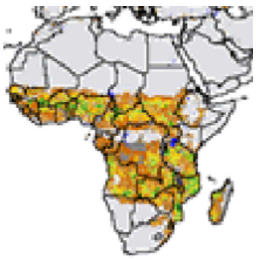
(f) Soy



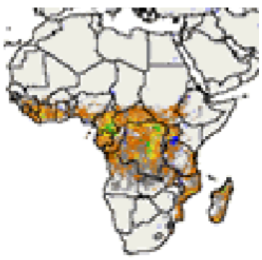
(g) Sunflower



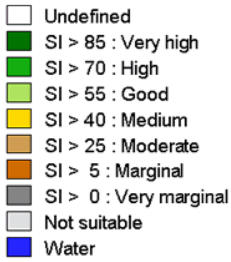
(h) Cassava



(i) Groundnut



(j) Sugarcane



(k) Legend

Figure A2: FAO Suitability Data

Table A1: Mean differences in land characteristics along southern versus eastern migration directions near southern rainforest boundary

	Observations (1)	Mean to the East (2)	Mean to the South (3)	Difference (4)
		Panel A: Overall Land Quality		
Mean south of -9°	166	9.62	10.05	-0.42
		(0.71)	(0.63)	(0.76)
Mean south of -8°	179	9.35	9.84	-0.49
		(0.67)	(0.59)	(0.71)
Mean south of -7°	197	9.31	9.67	-0.35
		(0.67)	(0.59)	(0.50)
Mean south of -6°	212	9.47	9.67	-0.19
		(0.60)	(0.51)	(0.63)
		Panel B: Dry Land Quality		
Mean south of -9°	166	2.97	2.95	0.02
		(0.35)	(0.29)	(0.33)
Mean south of -8°	179	2.93	2.86	0.06
		(0.32)	(0.27)	(0.31)
Mean south of -7°	197	2.90	2.76	0.15
		(0.30)	(0.25)	(0.29)
Mean south of -6°	212	2.93	2.73	0.20
		(0.29)	(0.24)	(0.29)
		Panel B: Wet Land Quality		
Mean south of -9°	166	6.66	7.09	-0.44
		(0.47)	(0.46)	(0.49)
Mean south of -8°	179	6.42	6.98	-0.56
		(0.44)	(0.43)	(0.46)
Mean south of -7°	197	6.41	6.91	-0.50
		(0.41)	(0.40)	(0.42)
Mean south of -6°	212	6.57	6.93	-0.39
		(0.40)	(0.38)	(0.41)

Notes: *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being east of 20°E to account for rainforest boundaries. Each row reports means for the adjacent cells to the south and to the east, as well as a t-test for the difference between them.

Table A2: Robustness checks: endogeneity of controls

VARIABLES	Land Quality	HHI: ecological regions
	(1)	(2)
South	0.0378	0.0135
	(0.0339)	(0.0530)
Suitability of major crops	✓	✓
Malaria Index	✓	✓
Standard deviation of elevation	✓	✓
Standard deviation of temperature	✓	✓
Observations	823	823
R-squared	0.584	0.474
Dependent Variable Mean	0.397	0.565

Notes: Standard errors are clustered according to Colella et al. (2019). *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table A3: Robustness checks: Heterogeneity by distance to rainforest

VARIABLES	Rainforest similar land	Pasture land	Linguistic similarity
	(1)	(2)	(3)
Rainforest x South	-0.000186	-0.00952	3.58e-05
	(0.000161)	(0.00787)	(0.000118)
South	0.522**	-2.075	0.235
	(0.240)	(8.179)	(0.199)
Suitability of major crops	✓	✓	✓
Malaria Index	✓	✓	✓
Standard deviation of elevation	✓	✓	✓
Standard deviation of temperature	✓	✓	✓
Observations	823	823	823
R-squared	0.277	0.173	0.131
Dependent Variable Mean	0.350	3.085	0.885

Notes: Standard errors are clustered according to Colella et al. (2019). *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table A4: Sorting into Regions Similar to African Rainforest

	Panel A: Hyde et al. population in year:					
	1000	1100	1200	1300	1400	1500
	(1)	(2)	(3)	(4)	(5)	(6)
South x rainforest similar	3,285**	3,407**	4,397**	4,380**	5,689**	5,813**
	(1,441)	(1,615)	(1,941)	(2,104)	(2,554)	(2,787)
rainforest similar	-847.3	-629.5	-994.8	-729.9	-1,146	-766.5
	(595.6)	(669.6)	(779.8)	(857.4)	(1,001)	(1,164)
South	285.4	649.0	879.6	1,328	1,546	1,380
	(961.6)	(1,136)	(1,339)	(1,522)	(1,783)	(1,957)
Observations	823	823	823	823	823	823
R-squared	0.296	0.318	0.317	0.339	0.337	0.301
Dependent Variable Mean	5008	6053	7201	8226	9576	10583
	Panel B: McCrary et al. population in year:					
South x rainforest similar	280.0	466.2*	620.6**	786.3**	985.2**	1,179**
	(194.4)	(256.2)	(311.3)	(370.4)	(438.2)	(505.0)
rainforest similar	101.9	91.01	71.23	54.25	44.68	33.43
	(143.2)	(185.5)	(228.5)	(272.3)	(316.7)	(361.3)
South	-48.66	18.77	53.39	100.2	178.3	250.0
	(239.5)	(300.2)	(354.3)	(414.6)	(487.6)	(559.7)
Observations	823	823	823	823	823	823
R-squared	0.201	0.206	0.221	0.232	0.235	0.238
Dependent Variable Mean	1593	2059	2487	2925	3398	3864
Distance to rainforest	✓	✓	✓	✓	✓	✓
Distance to market	✓	✓	✓	✓	✓	✓
Suitability of major crops	✓	✓	✓	✓	✓	✓
Latitude	✓	✓	✓	✓	✓	✓
Longitude	✓	✓	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓	✓	✓
Elevation	✓	✓	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓	✓	✓

Notes: Standard errors are clustered according to Colella et al. (2019). *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table A5: Land allocated to pasture by ancestral migration and year

	Pasture land in year:				
	1500	1600	1700	1800	1900
	(1)	(2)	(3)	(4)	(5)
South	-15.09**	-15.97**	-18.12**	-34.19**	-71.72**
	(7.521)	(7.909)	(8.990)	(16.86)	(30.89)
Distance to rainforest	✓	✓	✓	✓	✓
Distance to market	✓	✓	✓	✓	✓
Suitability of major crops	✓	✓	✓	✓	✓
Latitude	✓	✓	✓	✓	✓
Longitude	✓	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓	✓
Elevation	✓	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓	✓
Observations	823	823	823	823	823
R-squared	0.169	0.165	0.160	0.160	0.176
Dependent Variable Mean	3.085	3.347	3.805	7.368	19.19

Notes: Standard errors are clustered according to Colella et al. (2019). *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table A6: Impact of a rainforest history on institutions

	Jurisdictional Hierarchy Beyond Local Level (1)	Jurisdictional Hierarchy At Local Level (2)	Land Rights (3)	Inheritance Norms (4)	Divine Appointment (5)	Democracy (6)
Rainforest Ancestry	-0.980***	0.477	0.0235	-0.0466	-0.0400	-0.00925
	(0.160)	(0.371)	(0.0967)	(0.0784)	(0.0897)	(0.00926)
Distance to rainforest	✓	✓	✓	✓	✓	✓
Distance to market	✓	✓	✓	✓	✓	✓
Suitability of major crops	✓	✓	✓	✓	✓	✓
Latitude	✓	✓	✓	✓	✓	✓
Longitude	✓	✓	✓	✓	✓	✓
Malaria Index	✓	✓	✓	✓	✓	✓
Elevation	✓	✓	✓	✓	✓	✓
Standard deviation of elevation	✓	✓	✓	✓	✓	✓
Standard deviation of temperature	✓	✓	✓	✓	✓	✓
Standard deviation of land quality	✓	✓	✓	✓	✓	✓
Observations	379	428	428	428	428	428
R-squared	0.672	0.434	0.661	0.321	0.399	0.150
Dependent Variable Mean	2.649	2.451	0.780	0.911	0.0748	0.00467

Notes: Standard errors are clustered according to Conley (1999), using a 150km radius, to adjust for spatial correlation *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a 1° x 1° geographic cell. The partition used to define South vs Eastern migration route is {ABCHKLR:DEFGMNPS.}.That is Guthrie's partition as described in Flight (1980). The sample is conditioned on being south of 9°S to account only for non-rainforest settlements and avoid a mechanical relationship. Suitability controls include controls for cereal, maize, cotton, oil, pulses, rice, roots, sugar and wheat.

Table A7: Crops Studied and their Origins

Crop	Origin	Introduced to Africa
	Panel A: Dry pre-settlement	
Sorghum/Millet	West Africa	Native
Cotton	West Africa	Native
	Panel B: Wet pre-settlement	
Rice	Africa & South East Asia	Malaysian Complex
Pulses	West Africa	Native
Oil Palm	West Africa	Native
Banana	New Guinea	4,500 years ago
	Panel B: Dry post-settlement	
Soy	East Asia	1857
Maize	Americas	Colombian Exchange
Potato	Americas	Colombian Exchange
	Panel B: Wet prost-settlement	
Yams	Americas	Colombian Exchange
Cassava	Americas	Colombian Exchange
Groundnut	Americas	Colombian Exchange
Sweet Potato	Americas	Colombian Exchange

Notes: While there are pulses that are both native to Africa, Asia and America, the American crops are included separately in the data (ie. cassava, groundnut). Rice and Sugarcane are treated as traditional African crops as they war part of the Malaysian agricultural complex that arrived in Africa prior to southern expansion. The source for each crop is Murdock (1959) except sunflower, and soy (not included). The source for the introduction of sugarcane is: Ponting, Clive (2000). World history: a new perspective. London: Chatto & Windus. p. 353. ISBN 0-701-16834-X. The source for soybean introduction is: William Shurtleff, Akiko Aoyagi. History of Soybeans and Soyfoods in Africa (1857-2009), 2009 Sept. 6. The source for sunflower is: University of Cincinnati (2008, April 29). Ancient Sunflower Fuels Debate About Agriculture In The Americas. ScienceDaily. The source for banana is: https://www.sciencemag.org/news/2006/01/early-africans-went-bananas

Appendix B. Historical Migration Patterns

Reliable data on agricultural productivity from 2000 years ago does not exist, so as is typical in the economic history literature I use data on the slightly more reliable population as a proxy for productivity.[18] The standard logic for the approach is that in a Malthusian environment productivity generates higher births, more population pressure and faster rates of migration. My application of that concept is that migration speed is related to productivity. The hypothesis is that a change in environment caused a change in agricultural knowledge, which influenced productivity, population pressure and migration speed. Documenting this fact serves as the basis for the validity of the experiment in the first place, but still falls short of demonstrating anything relating to the Diamond hypothesis. Nevertheless, an important step.

The specification used to test this hypothesis is:

$$Frontier_{ct} = \beta_0 + \beta_1 SouthRoute_c \cdot Post_t + \Lambda_t + \Gamma_c + \epsilon_{ct} \tag{7}$$

c is a cell and t is a unit of time. Λ_t refers to a time fixed-effect, and essentially captures the speed of the eastern-migrants in the way the data is set-up. So I examine, for instance, the change in migration speed upon entering the rainforest between year 200-400C.E. I construct a dataset of population by cell which includes year 200 and 400. Λ_t is a time fixed effect, with the initial period always omitted, so it refers to observations taking place at year 400 in this case - in other words it's always a '*Post*' term. The outcome *Frontier* measures whether a geographic cell belongs to the migration frontier, and if so, takes the distance of this cell from the Bantu homeland. The frontier is defined as any cell with hunter/gatherer density in the 'before' period, and settler level population density in the 'post' period.[19] β_1 therefore measures the difference migration speed by southern migrants relative to eastern migrants in $km/200 - years$. Faster relative migration speeds are interpreted as implying higher levels of productivity.

i Results Table A8 shows the speed of the migration frontier throughout time in km/200-years. Due to the noisy nature of the data, individual estimates may be less informative than the pattern formed by the estimates at different periods

[18]Population is more credibly estimated using archeological findings than agricultural productivity, where there is really little hope of credible estimates.

[19]I allow some nominal population in a cell to account for the existence of a nomadic population prior to Bantu arrival

in time, which is most easily seen in figure A3. Recall the expected pattern based on the account of the history literature is "...first very slow rate of movement as people were learning about new aquatic habitats followed by a dash..." (Murdock, 1959). The top and bottom panels of figure A3 use different datasets but also produce very similar patterns. The Klein Goldewijk and Stehfest (2017) data shows the migration into the rainforest occurring about 100 years earlier than the McEvedy (1978) data, however in both cases there is a much slower migration speed within the 1C.E.-300C.E. time range that iron working is known to have induced Bantu to enter the rainforest. This implies an adaptation period where dry-crops were phased out, as described in the history literature. In both cases the southern migration speed recovers until 1000A.D. where there is a second drop in migration speed at the time the migration frontier is known to have exited the rainforest. Convergence occurs much faster in the Klein Goldewijk and Stehfest (2017) data, which even suggests that southern migrants 'dash' faster than their eastern migrating counterparts, as described by Murdock.

The decline in speed of southern migrants relative to eastern migrants upon exiting the rainforest is meaningful in both datasets. This is consistent with Diamond (1998). In contrast to the pattern observed upon entering the rainforest, convergence does not occur immediately in either dataset when we consider the adjustment to the new agricultural environment south of the rainforest. This general trend can be seen in table A8. Migration speeds dropped by between 1-4km per year on average upon entering the rainforest (column 2), small but precise effects. Upon exiting the rainforest, speeds dropped again, but by less. The speed of the frontier declined by roughly 1/4 to 1/2 a km per year exiting the rainforest (columns 6-7). Again, the effect was not large, but is at least consistent with knowledge loss.

Table A8: Population Change of Southern Migration Branch Relative to Eastern Migration Branch Throughout Bantu Expansion (people per 110km^2 cell)

Panel A: Klein Goldewijk, Beusen, and Janssen Data

	0-200	200-400	400-600	600-800	800-1000	1000-1200	1200-1400
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post x Rainforest	-92.59*	42.72	36.41	-118.8*	26.75	-37.24	-65.14*
	(55.53)	(47.18)	(63.53)	(65.69)	(42.23)	(45.15)	(33.29)
R^2	0.039	0.036	0.045	0.010	0.001	0.000	0.009
N	3,950	3,950	3,950	3,950	3,950	3,950	3,950

Panel B: McEvedy and Jones Data

	0-200	200-400	400-600	600-800	800-1000	1000-1200	1200-1400
Post x Rainforest	320.6***	307.7***	-415.6***	-143.5***	41.33	-22.63	-95.53***
	(54.82)	(110.9)	(77.55)	(40.39)	(41.81)	(39.00)	(28.00)
R^2	0.039	0.036	0.045	0.010	0.001	0.000	0.009
N	3,950	3,950	3,950	3,950	3,950	3,950	3,950
Cell Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Notes: Standard errors are clustered by region. Migration partition as described in Flight (1980) is used ({ABCHKLR:DEFGMNPS.}). Year dummies and cellFE are also included as controls.***p<0.01,**p<0.05,*p<0.1.
Description: This table shows that the migration pattern is consistent with imperfect vertical transmission of knowledge. When entering the rainforest there is an adjustment period as societies learn to produce wet-crops exclusively. As they learn, they converge to the speed of the other migration route. Because knowledge of dry-crops is imperfectly vertically transmitted, there is a second period exiting the rainforest where migration speed slows, since societies need to re-adjust to the new (old) agricultural environment.

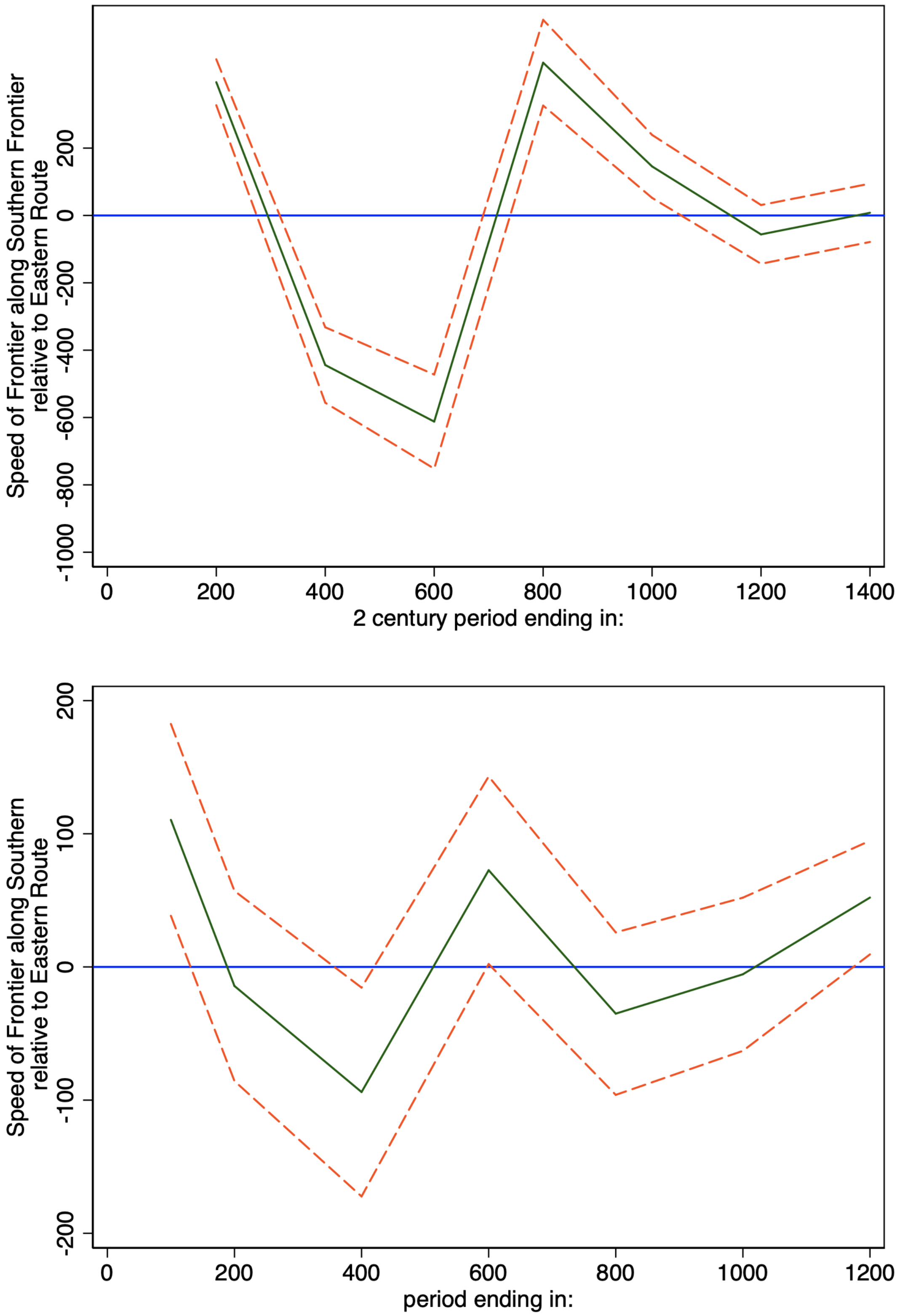


Figure A3: Speed of Migration Frontier

Notes: This figure shows observed population dynamics using two different datasets. Consistent with the historical account there is a slowing in migration as migrants enter the rainforest, followed by a "dash", which slows as migrants exit the rainforest. Data is presented applying the same model to two datasets. The Klein Goldewijk, Beusen, and Janssen is on the top and McEvedy and Jones data is on the bottom.

Appendix C. Alternate Contemporary Production Data

In many respects the available data is still questionable. For instance, the settlement and population data does not account for migration, and the production data is based on remote sensing images which are known to have a difficult time identifying some crops, and is almost certainly subject to some measurement error. In each case, classical measurement error would work against the findings, not with them, but it still seems prudent to test the robustness of the findings with other sources of data. Because the robustness checks are themselves using even worse data (otherwise they would be presented as the main results) the statistical significance is of less concern, but it would be disconcerting if the same patterns in the estimates did not emerge.

The robustness check uses the Afrobarometer data. The entire Afrobarometer was geocoded, and matched to an alternative source of crop production data, Institute (2019). The unit of observation is now at the individual level. A settlement group is assigned to the observation based on the language spoken by the respondent in the Afrobarometer survey. All respondents in the same village are assigned to the same Institute (2019) cell, so all analyses are now clustered at the village level. It is now assumed that villages are sufficiently separated that they can be assumed to be independent agro-climactically.

There are a number of advantages and disadvantages to this approach over the estimates used in the main results. First, information exists on individual farmers, so additional individual level controls become available. These include gender, education, exposure to the slave trade (taken from Nunn who also uses the Afrobarometer) and age. The ability to control for education is the main advantage. A number of other controls are added to this framework, to control for other potential explanations in the literature. Unfortunately data on the range of crops used in the main analysis is not available here, but one crop from each category is available. For dry-traditional crops, I have sorghum data; rice is available for wet-traditional; maize is used to for the dry recently introduced crops; and groundnut is available for the wet-recent crops.

Table A9 shows the results of this robustness check. Sorghum is still significantly less produced by individuals whose ancestors migrated south. Each of the two recently introduced crops are significantly less produced as well, while the rice estimate is statistically insignificant and fairly close to 0. This general pattern is very similar to the results found in the main analysis.

These estimates should not be considered in isolation however, there are a

Table A9: Contemporary crop production by crop-type

VARIABLES	(1) log(Sorghum) (dry-traditional)	(2) log(Rice) (wet-traditional)	(3) log(Maize) (dry-post colonial)	(4) log(Groundnut) (wet-post colonial)
Bantu x South	-0.74***	0.63***	-0.59**	-1.92***
	(0.15)	(0.18)	(0.26)	(0.17)
Latitude and Longitude controls	✓	✓	✓	✓
Education Fixed Effects	✓	✓	✓	✓
Suitability Fixed Effects	✓	✓	✓	✓
Age and Gender Controls	✓	✓	✓	✓
Distance to Metropolitan Centre	✓	✓	✓	✓
Exposure to Slavery	✓	✓	✓	✓
N	3194	3194	3194	3194
R^2	0.28	0.46	0.49	0.28

Notes: Robust standard errors in parentheses clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1. The partition is ABCHKLR:DEFGMNPS.That is Guthrie's partition as described in Flight (1980).

number of concerns with the exercise. Institute (2019) is based on a very high level of interpolation and are expected to be very high in measurement error. Furthermore, the sample is far from representative in the Afrobarometer. Only 3 sub-branches out of 7 are represented in the southern migrant ancestry, while 8 out of 8 sub-branches of the eastern migrants are represented. This itself seems to suggest isolation, since the Afrobarometer is simply not reaching the ancestors of southern migrants. In any case, the main point is that all estimates are consistent with the main estimates.

Furthermore, none of the robustness estimates are sensitive to the additional controls that are able to be included in the robustness check data. Education has little to no effect on any of the estimates, nor does exposure to the slave trade which could impact the adoption of recently introduced crops as most of these were introduced through the slave trade. This robustness check also reassures that the main results are not driven by recent migration. It might have initially been worrisome that initial settlement data was being matched with current production data. Using survey data alleviates this concern, as it takes the current location of respondents rather than their historical settlement. There are positives and negatives to this choice however. Current location is a choice variable of the individuals currently producing, so this leaves the estimates more open to selection criticisms. The initial settlement location is arguably a better measure to use as it is outside the control of current producers and so it suffers from fewer endogeneity issues. Nevertheless, the results are robust to using either current location or initial settlement location so it is a non-issue.

Appendix D. Tastes and Nutrition

One alternate explanation for different crops being produced by different societies may be based on tastes. Atkin (2016) provides evidence that Indian farmers consume fewer calories per rupee because of a cultural preference for certain foods. In the Bantu context, it could be that dry-traditional crops are under produced because of a cultural preference for wet-traditional crops rather than an inability to adopt new crops. This argument does not hold for the New World crops however. There could not have been a cultural preference for these crops since all societies were introduced to them at the same time. Accordingly, taste seemingly could explain the pattern in the estimates we observe.

By adopting New World crops, both the eastern and southern migrants would be substituting away from a traditional crop that they had developed a taste for. It therefore comes down to which one would take-on the higher opportunity cost of substitution. These substitutions, in most cases, were more costly for the non-rainforest migrants. As an example, consider the introduction of wheat, which is a New World crop that is less heavily produced by rainforest migrants. Wheat, nutritionally, is a substitute for traditional grains like sorghum and rice. However, sorghum has significantly more nutritional advantages over wheat than rice does, and would therefore have been costlier for non-rainforest migrants to switch based on nutritional profile. The only advantage that rice has over sorghum is that protein from rice is of higher quality, although this benefit is eliminated once rice is cooked, and is fairly insignificant since each of rice, wheat and sorghum are very poor protein sources. Sorghum contains the highest levels of fat, and has far more calories per cup (650 sorghum; 250 rice; 600 wheat).[20]

[20] Nutritional information from wolframalpha.com accessed August 21, 2013

Appendix E. Occupational Choice

One robustness check is that if agriculture is less efficient for one migration branch, it should be that these people substitute away from the agricultural industry over time. This can be examined for a subset of countries using the DHS. The DHS contains information on the occupations of individuals of various ethnicities.The dates of the surveys range from 1988 to 2006. Seven surveys from southern Africa (south of the rainforest) were combined: Malawi, Namibia, Zambia, Zimbabwe, Swaziland and Lesotho. South Africa was not used because some of the required data was missing. Unfortunately many of the controls previously available are not available here because the DHS is not always geocoded. Because of this the specification is extremely simple, and should be considered only as a suggestive, supporting piece of evidence to that above. A binary dependent variable for agricultural occupation is used, and tested against a bare-bones model indicating whether respondents had a rainforest history.

$$AgrOccupation = \beta_0 + \beta_1 Bantu + \beta_2 Bantu \cdot South + \Gamma CountryFE + \epsilon \tag{8}$$

Here, β_2 is the variable of interest. The expectation is that if the previously identified barriers to adoption existed, there would be differential exit from the agricultural sector among southern migrants.

The results can be found in table A10. The results indicate that Bantu speakers with a rainforest history are about 2% less likely to become agriculturalists than Bantu speakers whose ancestors went around the rainforest. This is consistent with the previous results, using a different dataset and suggests at least some of the effect seen previously is due to extensive margin differences.

E.1. Institutions

Another possible explanation for poor information flow could be that some societal structures are not conducive to cross-societal interaction. If this is also a function of rainforest migration, it could account for poor information flow. Accordingly, I check for differences in institutional features between the societies that descended from rainforest migrants, and eastern migrants. Precolonial institution data is available from (Murdock, 1959). Data exists on institutional hierarchy, both at and beyond the local level. There is also other information on institutions, such as property rights and inheritance norms. Again, I use a version of equation 1,

Table A10: Likelihood of agriculture being occupational choice by migration route

	Dependent Variable: Agriculture as Occupational Choice			
	(1)	(2)	(3)	(4)
Bantu x South	-0.00326***	-0.00584**	-0.0234**	-0.0239**
	(5.51e-05)	(0.00194)	(0.00914)	(0.00921)
Country Fixed Effects	Y	Y	Y	Y
Age	N	Y	Y	Y
Education	N	N	Y	Y
Gender of HH Head	N	N	N	Y
Observations	18447	18447	18443	18440
R-squared	0.131	0.139	0.168	0.168

Notes: Notes: Standard errors are clustered by region. Migration partition as described in Flight (1980) is used ({ABCHKLR:DEFGMNPS.}).***p<0.01,**p<0.05,*p<0.1.
Description: This table shows the impact of rainforest history on contemporary occupational choice. Southern migrants are up to 2.5% less likely to be agriculturalists today, suggesting that a lot of the lack of adoption may occur at the extensive margin rather than the intensive margin. If regions that are unsuitable for new world crops or dry-traditional crops are driving this result, it explains how the knowledge loss could persist for so long.

using one of the institutions variables from the Murdoch dataset as the dependent variable.

The results are presented in table A6 and **??**.[21] Interestingly the estimates in differences in institutions are fairly consistent with the idea that societies with a rainforest history are more geographically isolated. We see that these groups are less likely to have a hierarchy beyond the local level (column 1); which may be consistent with a lower return to travelling beyond the local level. This seems fairly consistent with a more isolated group, as suggested by table 5. Beyond that there are few differences on the institutions dimension. Local institutions are similarly complex (column 2), land rights are similarly enforced (column 3) and inheritance norms are similar (column 4). Looking more closely at inheritance norms suggests that patrilineal inheritance may be slightly more likely among southern groups, but the effect is not too precise, and not particularly robust so I do not make too much of that.

[21] Inheritance rights are relegated to the appendix as it seems harder to think of how they would be related to a rainforest ancestry.