Ethnicity, Isolation, and Knowledge: the Bantu Expansion

Arthur Blouin
University of Toronto

April 30, 2016

Question: how do ethnolinguistic groups evolve over time?

- Ethnolinguistic diversity is fundamentally important to economic development
  - Influences public goods provision
  - Influences civil conflict
  - Generally associated with poor development outcomes
- Our understanding of its evolution is surprisingly limited
  - Geographic heterogeneity plays an important role
  - 'State building' (or public goods provision) is important
- But ethnic diversity evolves dynamically over centuries and we have little understanding of these dynamics

Big picture: endogenous fractionalization

Conceptually, there are (at least) 3 mechanisms generating endogenous ethnolinguistic fractionalization:

1. Differences in societal interactions / marriage / etc
   - Michalopoulos (2012)
2. Government policy
   - Ahlerup and Olsson (2012)
   - Blouin and Mukand (2016)
3. Differences in settlement patterns and migration
   - This project

Hypothesis

- There's benefit to groups living in close proximity to each other
  - e.g. Ability to share production knowledge
- Earning this benefit requires some costs
  - Neighbours may not be near the best land for what you want to produce
  - Must invest in remaining culturally close to neighbours so that communication is possible
- Changing land preferences or cultural distance may $\rightarrow$ fractionalization
  - To do this we need to examine migration / settlement decisions

What we do

- We study Bantu migrants. Migration followed two paths:
  1. through rainforest (wet crops)
  2. around rainforest (balanced wet/dry)

- Jared Diamond - axis orientation: rainforest migration $\rightarrow$ loss of dry-production knowledge
  - did rainforest migrants 'forget' dry-land agriculture?
  - if so: does this influence settlement locations?

- Both migration paths end up in southern Africa in the same regions, and are *able* to produce wet & dry

- If rainforest migrants forgot dry-crops, they'll settle on different land and and interact less with dry-crop farmers.
  - does this influence geographic & cultural isolation?
  - is it linked to agricultural information sharing?

Context: the Bantu expansion

The empirical strategy compares the ancestors of migrants around the rainforest vs. through the rainforest

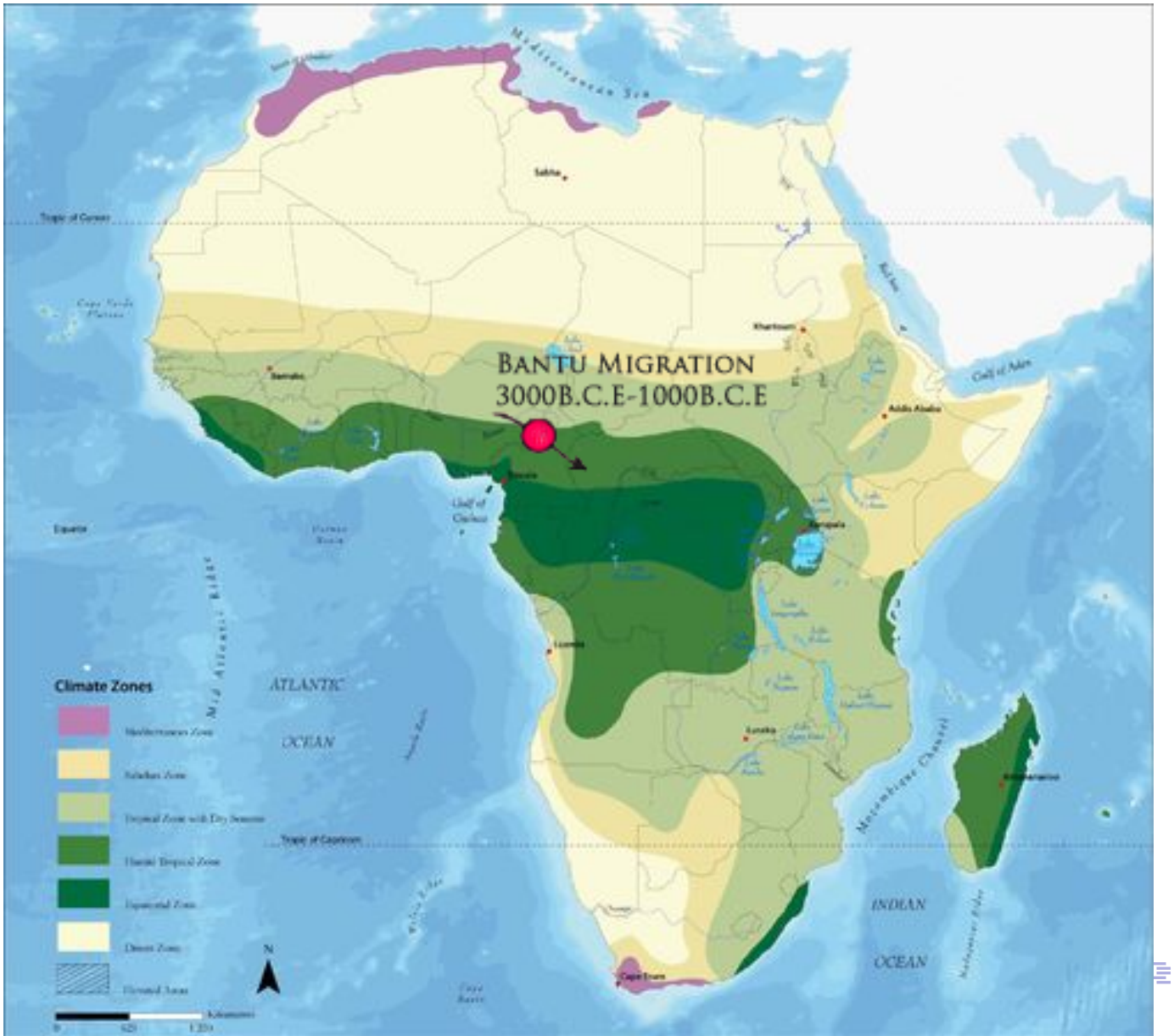
- Migration happened so slowly that for any individual migrant there was not a significant difference in quality or suitability of land in any direction
  - "...people preferred sites away from sectors where the best spots were already occupied by other farmers and trappers. To contemporaries this was just natural drift." - Jan Vansina
  - "It is likely that the expansion actually got underway by accident" - Jan Vansina

- The split was due to iron
  - "...if we consider what might have been the new development which enabled them to do this [migrate south], then I think we should look...towards the coming of the iron age." - Roland Oliver

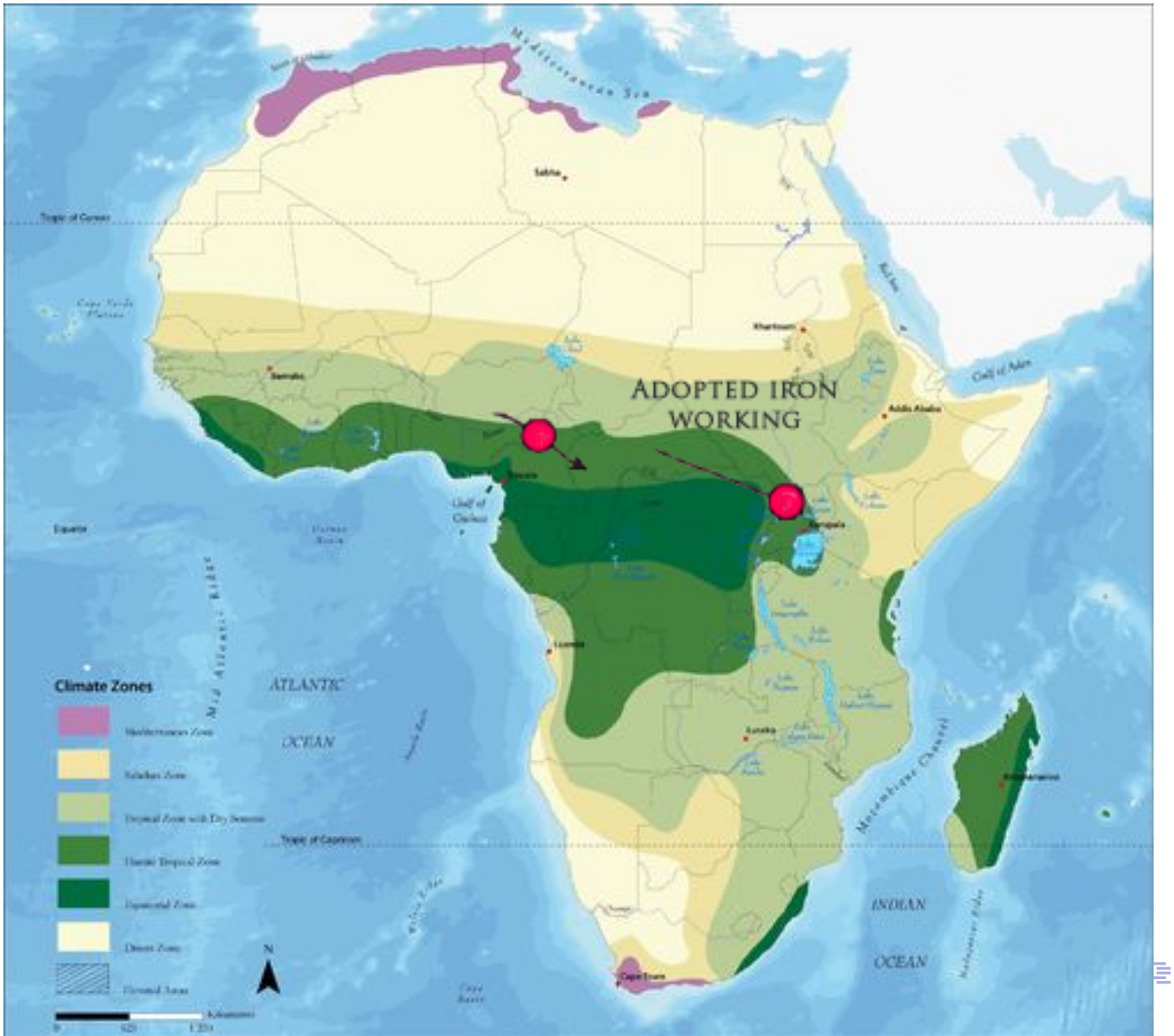
Arthur Blouin
University of Toronto

Quick history: Bantu expansion

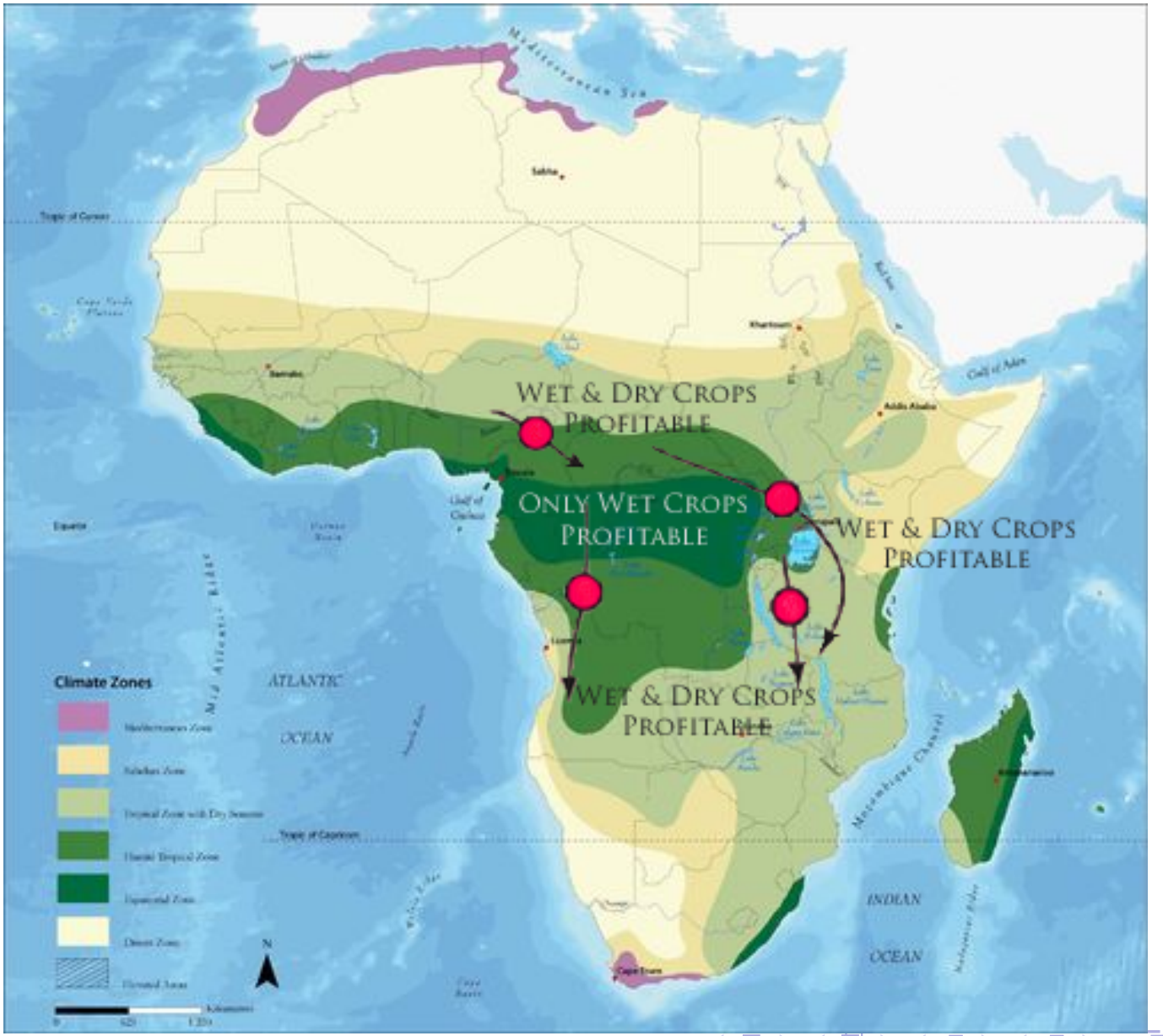
B.C.E.



Quick history: Bantu expansion 100C.E. - 300C.E.



Quick history: Bantu expansion



Summary of the story

- Southern migrants travel through the rainforest over several generations. Two things may happen:
  1. 'forget' dry-agriculture $\rightarrow$ strong preference for wet-crop land
  2. possible cultural drift over generations of migration through rainforest

- due to both (1) & (2) $\rightarrow$ living close to neighbours is less beneficial

- They arrive in southern Africa where they *should* produce some dry-crops

- Prefer to branch off from group to live on land they prefer
  - causes geographic isolation (lower ELF)

Data requirements

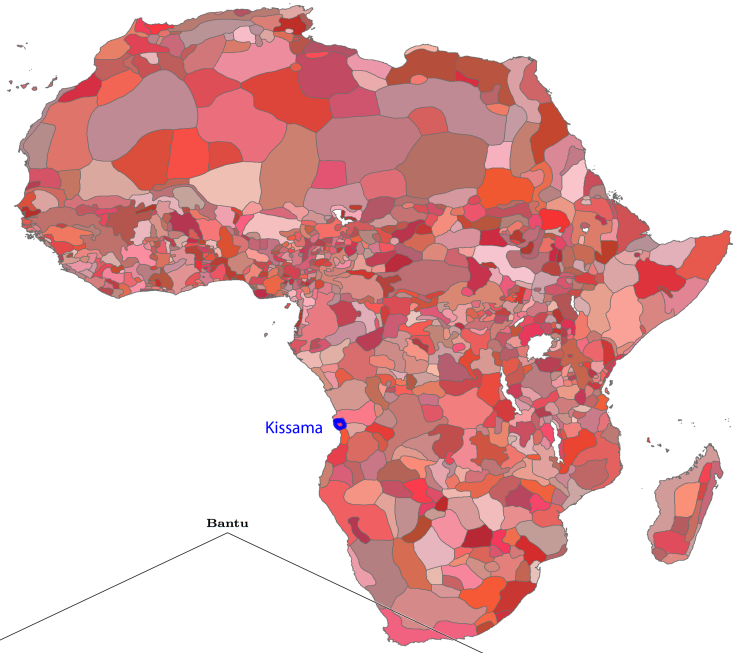
The unit of observation throughout is a geographic cell (0.5° by 0.5°). For each cell we want information on:

1. Migration and settlement (who, when, how much?)
2. Agricultural Production and information sharing
   - Information sharing is tough one to get
   - We argue: if rainforest migrants produce less of *all* crops that required adoption since settlement and *none* of the crops that didn't they must share less information
3. Cultural similarity and ethnolinguistic fractionalization

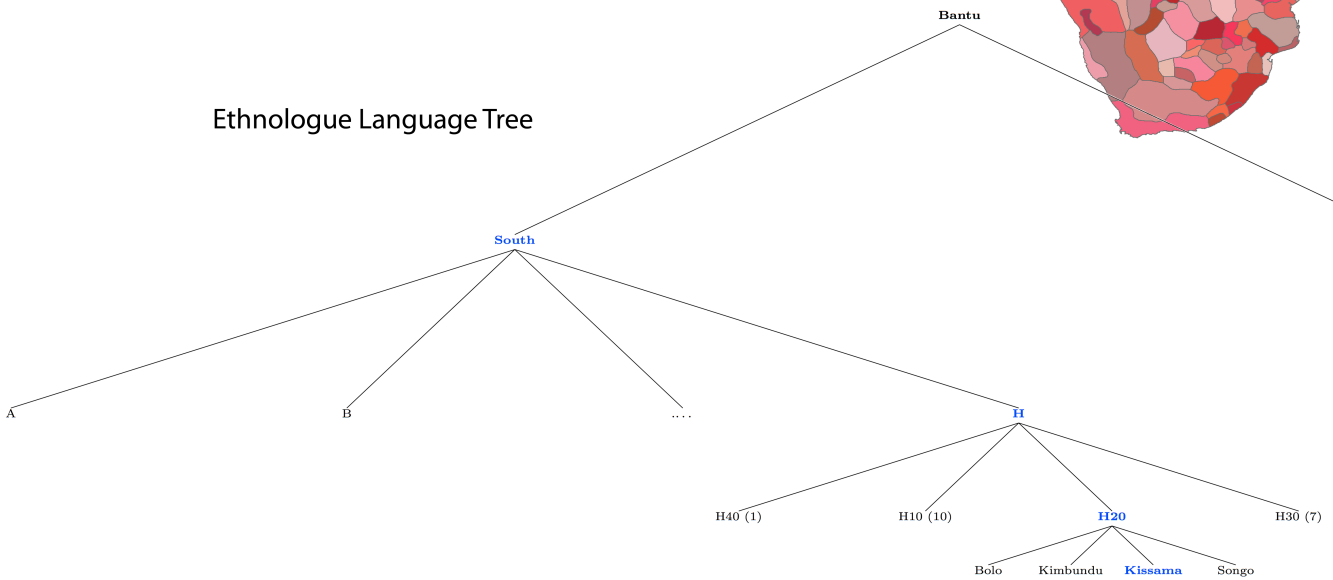
1. Migration data

- When/How much migration?
  - Migration is inferred from changes in historical population data (2 sources: Klein, Goldewijk et al. and McEvedy)
  - Works in this context because there were initially only nomadic tribes south and east of the Bantu homeland
  - Look at 'migration frontier' as the year a cell has a population $>$ nomadic population
- Who? How are migration routes assigned?
  - Each ethnicity in Murdock's Ethnolinguistic map of Africa was matched to the Ethnologue database
  - The Ethnologue data include a language tree which splits in correspondence with the migration split.
  - Each Bantu settled cell is assigned to either a direct south migration route or an east route

Murdock Ethnolinguistic Map



Ethnologue Language Tree



2. Cultural similarity and fractionalization data

We need measures of geographically/cultural isolation

- Historical fractionalization is straightforward to compute using the population and Murdoch data
  - Calculated within 5° x 5° 'virtual countries'
- 'Cultural similarity' is more difficult:
  - We rely again on the Bantu language tree data
  - We take some fixed level of the language tree and compute the number of languages that share that root
  - We argue that languages that share the root are 'close'
  - Then: straightforward to compute the number of close neighbours as a measure of 'cultural connectedness'

Language tree

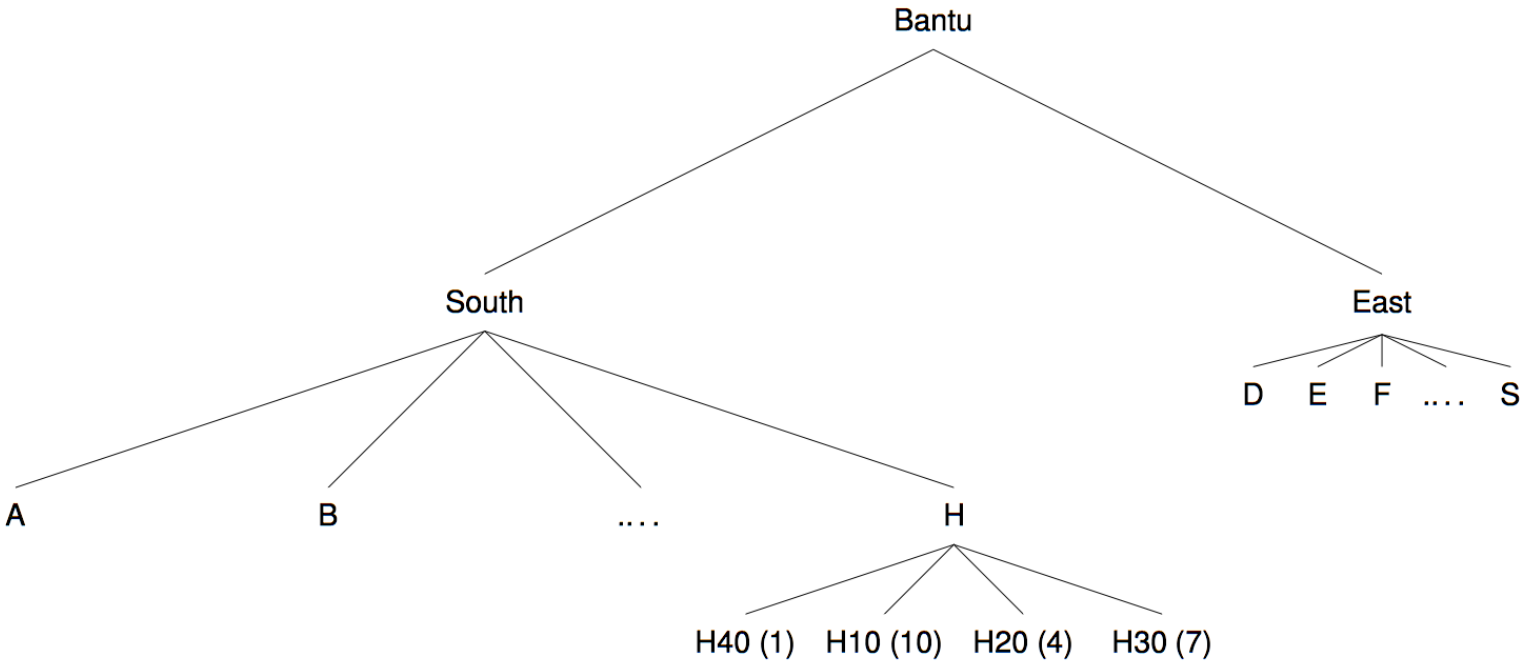


Figure 3: Bantu Language Tree. The numbers in brackets denote the number of tribes with that point as their root.

3. Agricultural production data

Ag. data may help us say something about any link between information sharing and diversity. 2 related hypotheses:

1. Diamond hypothesis: did rainforest migrants 'forget' production techniques they couldn't use in the rainforest?
   - Historical data is ideal: data on land devoted to livestock
   - livestock not kept in rainforest because of tsetse fly.

2. Information transfer: did migrants adopt new production techniques (here, crops) after settlement?
   - Here have to mainly rely on contemporary data
   - Do south-Bantu underproduce crops that would have, at some point since settlement, be adopted from someone else?
   - Focus on crops that are native to Africa versus crops introduced during the slave trade (after Bantu settlement)

3. Agricultural production data

Historical Data:

- We can get historical data on pasture land from Klein, Goldewijk et al.

Contemporary Data:

- We have mid 1990s crop production for all major crops, measured using GIS data (Leff et al. (2004))
  - Provides, for each cell and each major crop, the percentage of the cell devoted to the production of that particular crop
- To test robustness, contemporary production is also measured using You et al. (2006)
  - It combines rougher GIS estimates with admin. data.

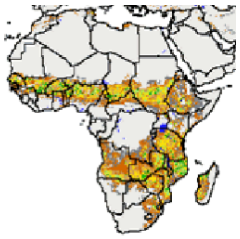
3. Agricultural production data

Since we want to try to use production data to measure information transfer we need to control for production potential

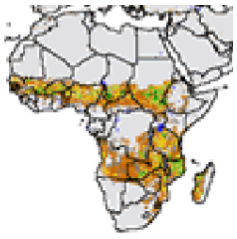
- FAO crop suitability data was used to measure land type
  - A potential yield for each crop in each cell comes in a suitability category ranging from 1-8
  - Regressions include crop specific yield-band FE to control for what people *should* be producing
  - Beyond that we can control for region-specific effects (or pick up any residual measurement error) with region FE

Suitability of wet and dry crops

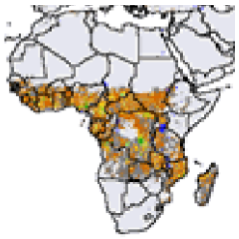
Sorghum



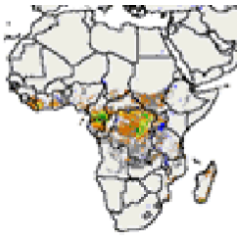
Cotton



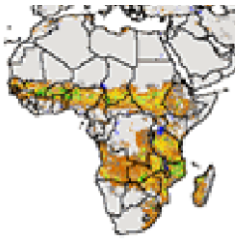
Rice



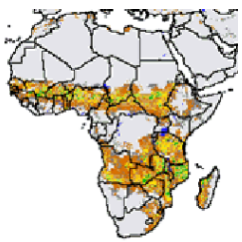
Oilpalm



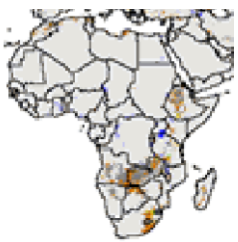
Maize



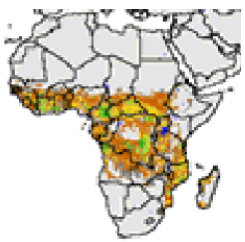
Soy



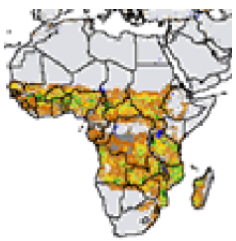
Sunflower



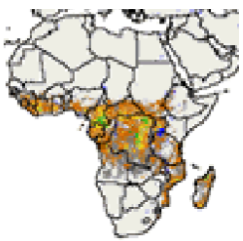
Cassava



Groundnut



Sugar



Legend

General plan

With this data there is a set of empirical questions we can test:

1. Did migrants change their production techniques because they were in the rainforest?
2. Did they 'forget' any production techniques they would have had prior to rainforest migration?
3. When they exited the rainforest did production knowledge influence settlement patterns?
4. Were differences in settlement pattern large enough to influence geographic and cultural connectedness?
5. Can we link 'connectedness' to information flows?

General plan

1. Did migrants change their production techniques because they were in the rainforest?
2. Did they 'forget' any production techniques they would have had prior to rainforest migration?

Migration and knowledge

- It's impossible to get reliable data on yields in, for example, the year 200.
- However it's reasonable to assume a Malthusian environment
- Migration occurred to relieve population pressures, so migration speed is associated with agricultural productivity
- If migrants had to 'learn' about wet-crop agriculture expect dip in migration speed as they enter the rainforest
- If they 'forgot' dry agriculture, expect another dip in migration speed as they exit the rainforest

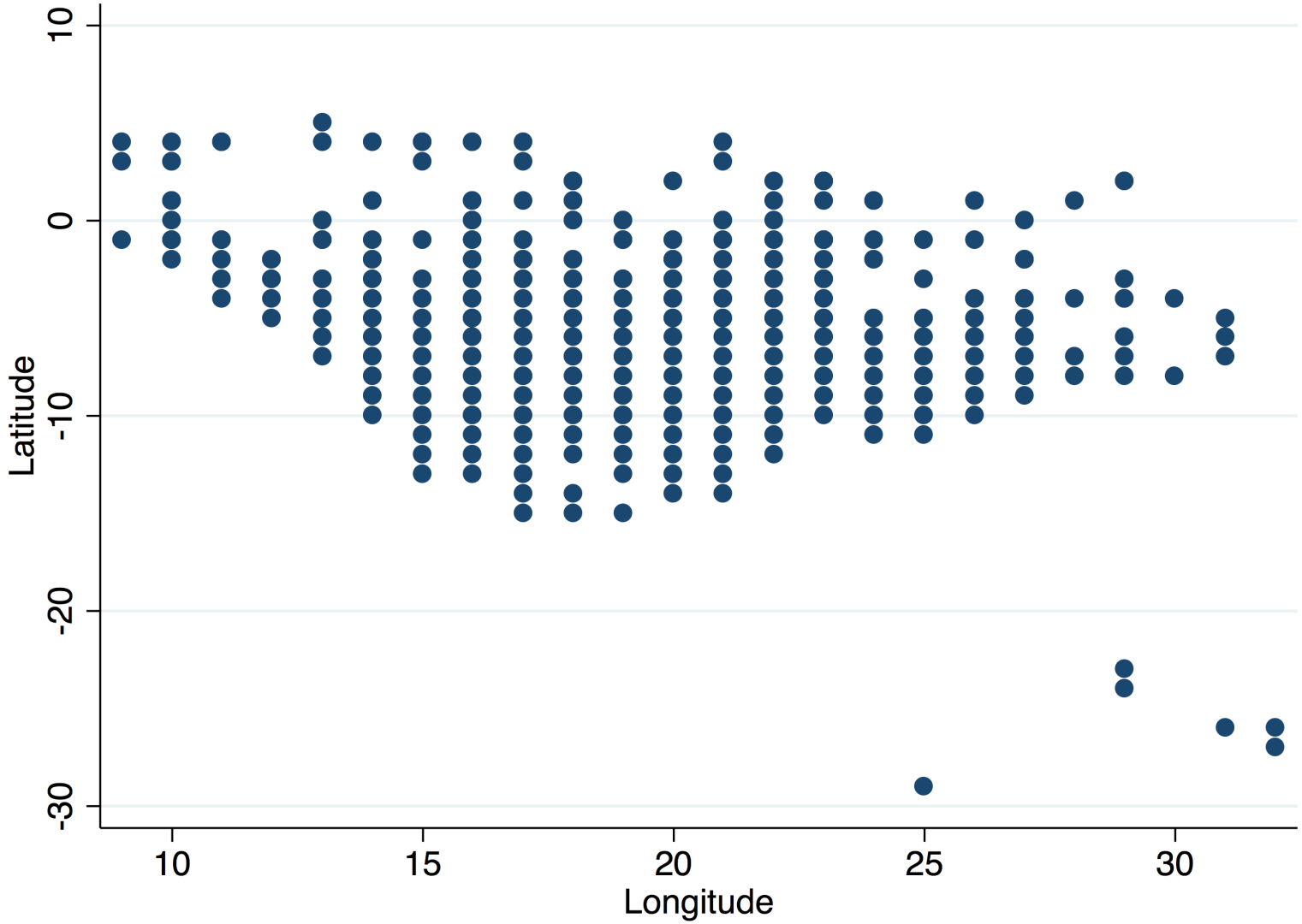
Migration and knowledge loss

1st step in Diamond hypothesis: some ag. techniques 'abandoned' (i.e. migrants had to adjust to rainforest)

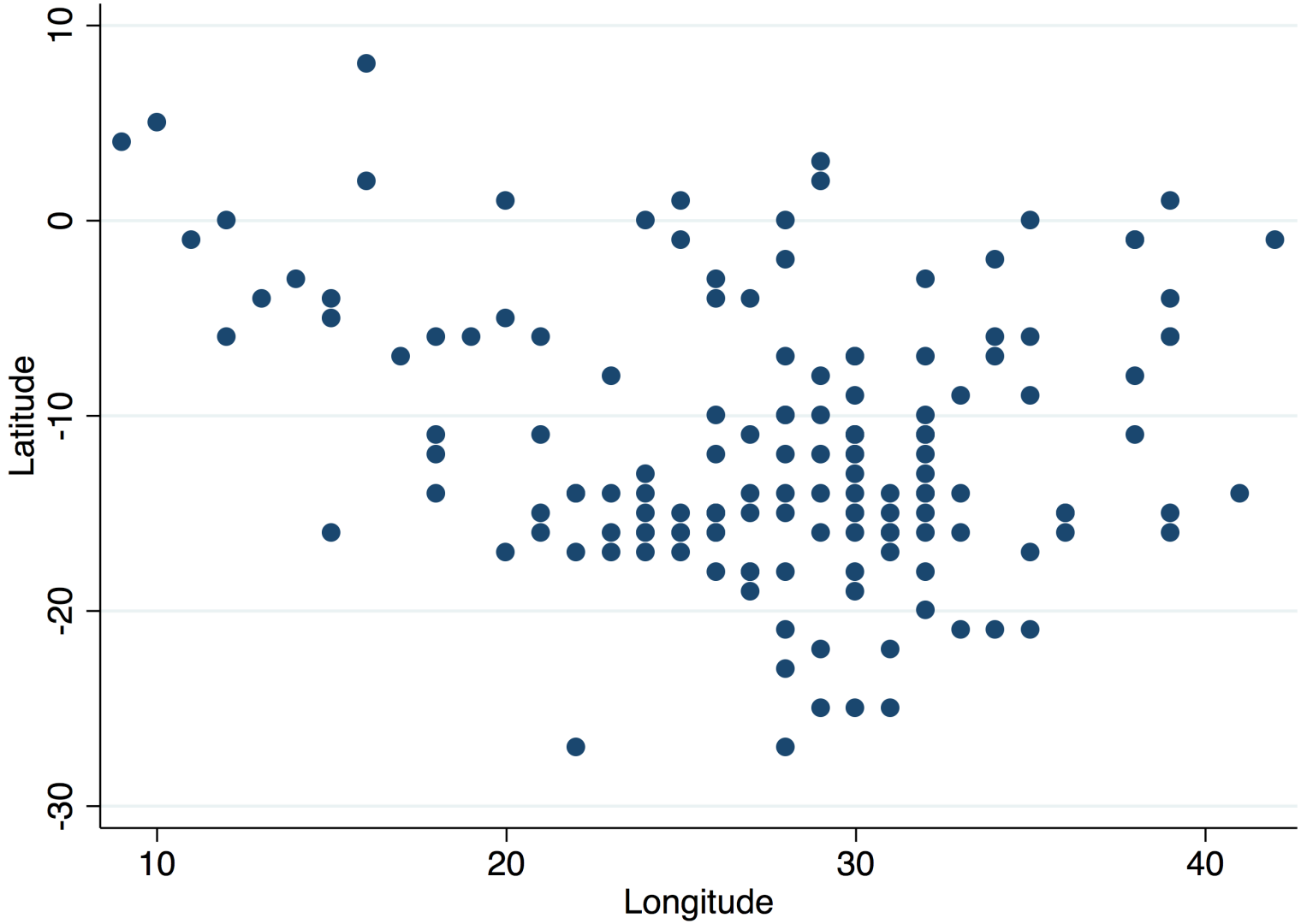
$$MigrationFrontier_{it} = \beta_0 + \beta_1 South_i \cdot Post_t + \beta_2 Post_t + \gamma CellFE_i + \epsilon_{it}$$

- i is a geographic cell; t is a period of time (increments of 200 years)
- *MigrationFrontier* provides the distance in km from the Bantu homeland of the migration frontier
  - Defined as a settled agriculture population in the post period and a population density consistent with nomadic tribes in the pre-period
- $South_i$ denotes a cell with southern migrants
- β_1 provides the speed of rainforest migration in km/200 years relative to migration around the rainforest.

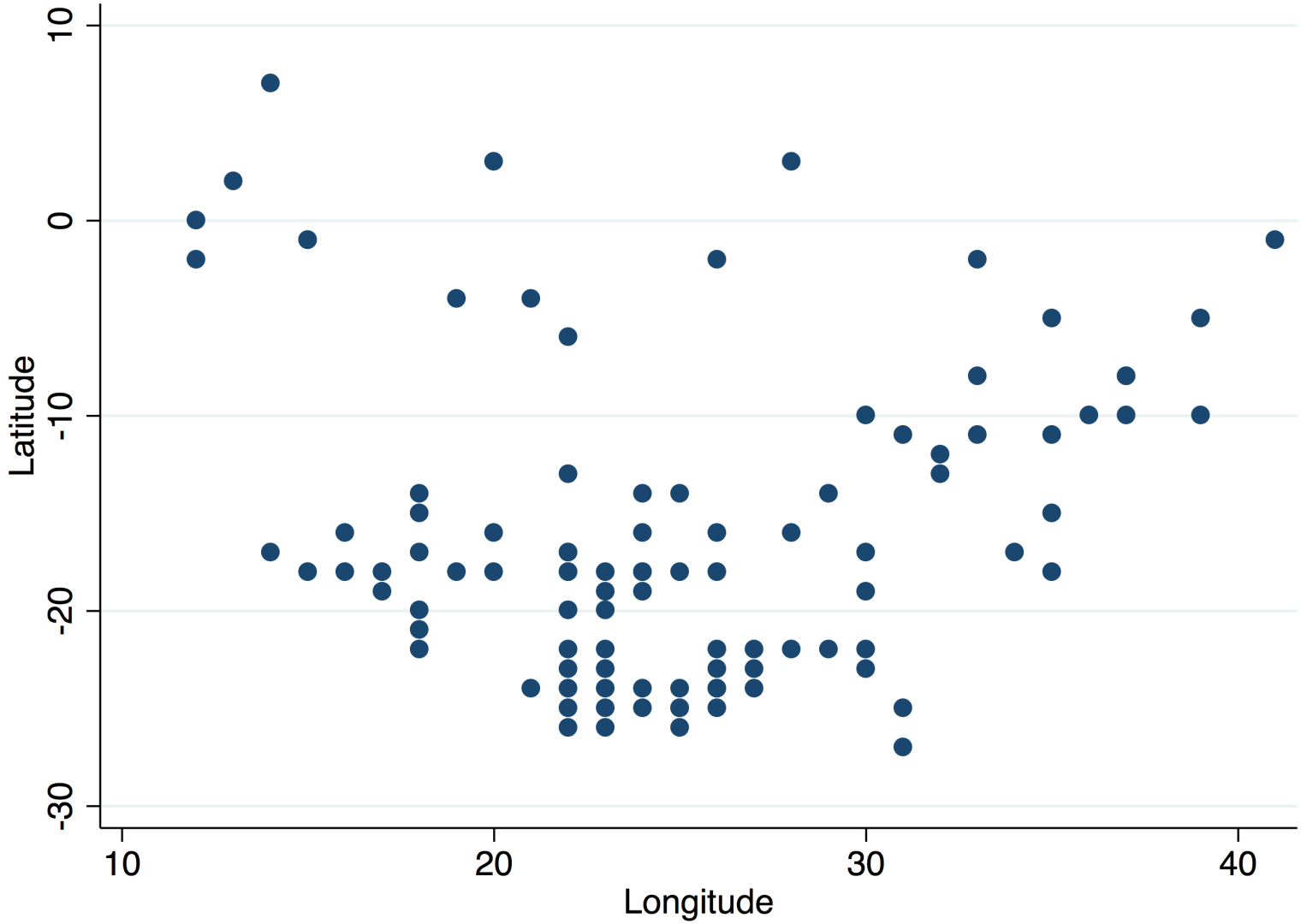
New Bantu settlements (migration frontier) Year 0



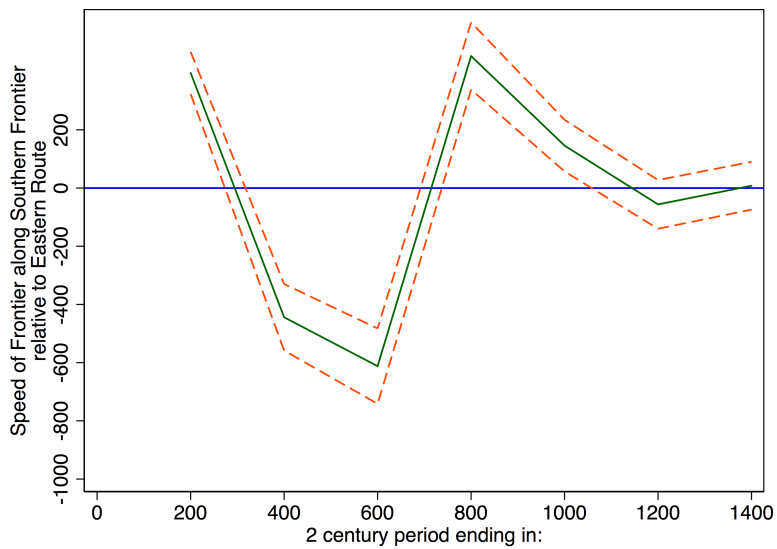
New Bantu settlements (migration frontier) Year 600

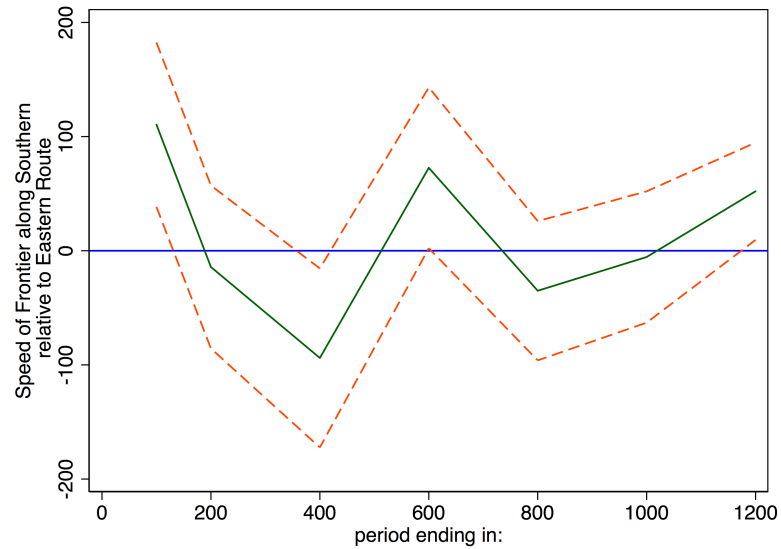


New Bantu settlements (migration frontier) Year 1000



Speed of migration frontier over time for southern migrants relative to eastern migrants





Production and knowledge loss

The 2nd step in the Diamond argument is testing whether ag. info was 'forgotten'

- Bantu couldn't keep livestock in the rainforest because of the tsetse: what happens when they settle in the south?

$$Livestock_i = \beta_0 + \beta_1 South_i + \gamma LandCharacteristics_i + \lambda RegionFE_i + \epsilon_i$$

- $Livestock_i$: share of land in cell i devoted to livestock
- $South_i$: a cell settled by southern Bantu migrants
- $LandCharacteristics_i$: FE for quality of each crop
- Sample: southern Africa where both groups live together
- Sample: Separate regressions for year 1500, 1600, 1700, 1800

If $\beta_1 < 0$: even for similar plots of land, rainforest Bantu are less likely to keep livestock

Historical land use

	Historical Use of Land for Pasture											
	1500C.E.			1600C.E.			1700C.E.			1800C.E.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
South Bantu	-0.545*	-0.637**	-0.666**	-0.556**	-0.668**	-0.697***	-0.561**	-0.692***	-0.720***	-0.490*	-0.675**	-0.711**
	(0.279)	(0.260)	(0.264)	(0.279)	(0.260)	(0.265)	(0.280)	(0.261)	(0.266)	(0.288)	(0.271)	(0.277)
Region Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Latitude & Longitude	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Land Quality	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Crop Suitability Score FE	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
Distance to Market	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Rainforest Similarity Index	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Desert Index	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Observations	3,763	3,763	3,763	3,763	3,763	3,763	3,763	3,763	3,763	3,763	3,763	3,763
R-squared	0.653	0.744	0.745	0.662	0.753	0.754	0.667	0.759	0.760	0.677	0.774	0.775

General plan

3. When they exited the rainforest did production knowledge influence settlement patterns?

Settlement

If rainforest Bantu lost the ability to produce dry-agriculture and re-adoption is costly, they may prefer to settle land where they don't need to adopt new techniques

- If so: they would want to settle land most similar to the rainforest
- We therefore construct an index for similarity to the rainforest for each cell.

$$Index_i = 1 - \frac{\sum_c^8 (suit_{ic} - \bar{suit}_{cr})^2}{max\{\sum_c^8 (suit_{ic} - \bar{suit}_{cr})^2\}}$$

- $suit_{ic}$: The suitability of crop c in cell i
- $\bar{suit}_{cr}$: The average suitability of crop c in the rainforest

Sorting on land characteristics

$$\begin{aligned} Population_i = &\beta_0 + \beta_1 South_i \cdot Index_i + \beta_2 South + \beta_3 Index_i \\ &+ \beta_3 LandQuality_i + \Upsilon SuitabilityFE + \theta f(Coordinates) + \epsilon_i \end{aligned}$$

- Tests whether the index determines settlement patterns differentially for southern migrants

Sorting into regions similar to rainforest

Panel A: Klein Goldewijk, Beusen, and Janssen Data

Dependent Variable: Population per cell

	1000	1100	1200	1300	1400	1500
	(1)	(2)	(3)	(4)	(5)	(6)
Bantu South x Similarity	4011.1***	4132.9***	5127.4***	5102.1***	6466.6***	7194.6***
	(1274.1)	(1421.6)	(1681.2)	(1818.4)	(2183.8)	(2456.1)
Similarity Index	-687.61	-509.3	-918.2	-676.7	-1134.61	-330.5
	(591.6)	(654.1)	(756.8)	(829.23)	(966.5)	(1090.6)
R-squared	0.2280	0.2543	0.2667	0.2895	0.2934	0.2494
Observations	935	935	935	935	935	935

Sorting into regions similar to rainforest

Panel B: McEvedy and Jones Data

Dependent Variable: Population per cell

	1000 (1)	1100 (2)	1200 (3)	1300 (4)	1400 (5)	1500 (6)
Bantu South x Similarity	196.1	351.6	474.5*	608.6*	776.0*	937.8**
	(171.7)	(226.1)	(278.3)	(330.7)	(392.8)	(453.4)
Similarity Index	167.9	181.0	185.2	191.0	202.5	212.4
	(108.9)	(140.2)	(172.2)	(204.8)	(237.8)	(271.1)
R-squared	0.126	0.138	0.156	0.170	0.175	0.181
Observations	935	935	935	935	935	935

General plan

4. Were differences in settlement pattern large enough to influence geographic and cultural connectedness?

Settlement patterns $\longrightarrow$ isolation?

Different land preferences may or may not lead to diversity:

- IF: no cultural drift in the rainforest vs. around
  - rainforest migrants may cluster together on wet-crop land
  - others cluster together on dry-crop land
  - Neither is more or less geographically isolated

- IF: $\uparrow$ cultural drift in rainforest
  - cost to information sharing higher in general
  - geographic drift

Isolation: empirical specification

$$\begin{aligned} C_i = &\beta_0 + \beta_1 Bantu \cdot South_i + \beta_2 Bantu_i + \beta_3 LandQuality_i \\ &+ \Gamma SuitabilityFE_i + \lambda RegionFE_i + \theta f(Coordinates) + \epsilon_i \end{aligned}$$

- C: refers to either the linguistic closeness of neighbours or ethnolinguistic diversity
  - Meant to capture expected effect on cultural similarity and fractionalization

Impact of axis orientation on language

Dependant Variable:	Number of Linguistically Close (general)				Percentage of Neighbours Linguistically Close			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bantu x South	-1.690**	-1.524*	-1.800**	-2.445***	-0.226***	-0.209***	-0.164***	-0.178***
	(0.780)	(0.789)	(0.767)	(0.562)	(0.0620)	(0.0613)	(0.0555)	(0.0498)
Suitability (all crops)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Latitude and Longitude polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land Quality	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance to Major Market	No	No	Yes	Yes	No	No	Yes	Yes
Region FE	No	No	No	Yes	No	No	No	Yes
Observations	968	968	968	968	995	995	995	995
R-squared	0.149	0.163	0.171	0.322	0.193	0.194	0.194	0.310

Impact of axis orientation on fractionalization

Dependant Variable:	Fractionalization 1960				Fractionalization 1500			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bantu x South	-0.0496*	-0.0538**	-0.0539**	-0.0457*	-0.0667**	-0.0688**	-0.0704**	-0.0577*
	(0.0252)	(0.0251)	(0.0251)	(0.0268)	(0.0283)	(0.0284)	(0.0284)	(0.0320)
Suitability (all crops)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Latitude and Longitude polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land Quality	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance to Major Market	No	No	Yes	Yes	No	No	Yes	Yes
Region FE	No	No	No	Yes	No	No	No	Yes
Observations	968	968	968	968	995	995	995	995
R-squared	0.149	0.163	0.171	0.322	0.193	0.194	0.194	0.310

General plan

5. Can we link 'connectedness' to information flows?

Isolation $\longleftrightarrow$ information flows?

Rainforest migrants chose geographic isolation because of information sharing was costly:

- If true: after settlement we should see them be less likely to adopt new agricultural products
    - We look at 4 categories of crops: wet vs. dry and traditional to Africa versus New World crops (introduced through the slave trade)
    - For rainforest migrants wet-traditional do not need to be adopted, everything else did
    - If the mechanism is really costly info sharing then all crops requiring adoption should be less produced, but none not requiring adoption should be

Using production data to capture information flows

$$log(CropProduction)_i = \beta_0 + \beta_1 Bantu \cdot South_i + \beta_2 Bantu_i + \beta_3 LandQuality_i + \Gamma SuitabilityFE + \lambda RegionFE_i + \theta f(Coordinates) + \epsilon_i$$

- i is a 'cell'
- *Bantu*: a cell was settled by either a southern or eastern migrating Bantu society
- *South*: a cell was settled by a southern migrating Bantu society
- *LandQuality*: An index of overall quality of land
- *SuitabilityFE*: The crop specific suitability controls. 8 categories per crop included as FE

Production by ancestral migration route and crop type

	Crops Traditional to Africa						
	Dry Crops				Wet Crops		
	log(Sorghum) (1)	log(Cotton) (2)	log(Barley) (3)	log(Wheat) (4)	log(Rice) (5)	log(Pulses) (6)	log(Oil Palm) (7)
South x Bantu	-0.348***	-0.111**	-0.0208**	-0.00239**	0.00161	0.0394	0.0309
	(0.0770)	(0.0461)	(0.00948)	(0.000950)	(0.0349)	(0.0358)	(0.0209)
Bantu	0.0564	0.0868***	0.0154	0.00171	-0.000802	-0.0119	-0.00839
	(0.0723)	(0.0308)	(0.0104)	(0.00107)	(0.0166)	(0.0218)	(0.00899)
Region Fixed Effects	Y	Y	Y	Y	Y	Y	Y
Latitude & Longitude	Y	Y	Y	Y	Y	Y	Y
Land Quality	Y	Y	Y	Y	Y	Y	Y
Crop Suitability Score FE	Y	Y	Y	Y	Y	Y	Y
Distance to Market	Y	Y	Y	Y	Y	Y	Y
Rainforest Similarity Index	Y	Y	Y	Y	Y	Y	Y
Desert Index	Y	Y	Y	Y	Y	Y	Y
Observations	4524	4524	4524	4524	4524	4524	4524
R-squared	0.547	0.613	0.474	0.473	0.563	0.671	0.511

Production by ancestral migration route and crop type

	Crops Introduced During the Slave Trade				
	Wet Crops			Dry Crops	
	log(Sunflower)	log(Groundnut)	log(Sugarcane)	log(Maize)	log(Soy)
	(1)	(2)	(3)	(4)	(5)
South x Bantu	-0.114***	-0.0942*	-0.0536**	-0.325***	-0.0358*
	(0.0362)	(0.0560)	(0.0220)	(0.0953)	(0.0197)
Bantu	0.102***	0.0927***	0.0417*	0.315***	0.0374***
	(0.0333)	(0.0337)	(0.0218)	(0.0761)	(0.0135)
Region Fixed Effects	Y	Y	Y	Y	Y
Latitude & Longitude	Y	Y	Y	Y	Y
Land Quality	Y	Y	Y	Y	Y
Crop Suitability Score FE	Y	Y	Y	Y	Y
Distance to Market	Y	Y	Y	Y	Y
Rainforest Similarity Index	Y	Y	Y	Y	Y
Desert Index	Y	Y	Y	Y	Y
Observations	4524	4524	4524	4524	4524
R-squared	0.566	0.491	0.341	0.780	0.524

Conclusion

The origins of diversity are rooted in land heterogeneity, but it's a complex dynamic process

- When we consider endogenous migration over time we see a more nuanced relationship between land diversity and ethnic diversity
- Information sharing is crucial both to determining fractionalization and as an outcome of it
- Cultural isolation and fractionalization seem to co-evolve - they respond similarly to production and information incentives, but may also contribute to low information flow and therefore lower productivity