

# The Areal-Typological Division of Africa: A computational approach

Harald Hammarstrom  
Uppsala University

24 Oct 2017, Berlin

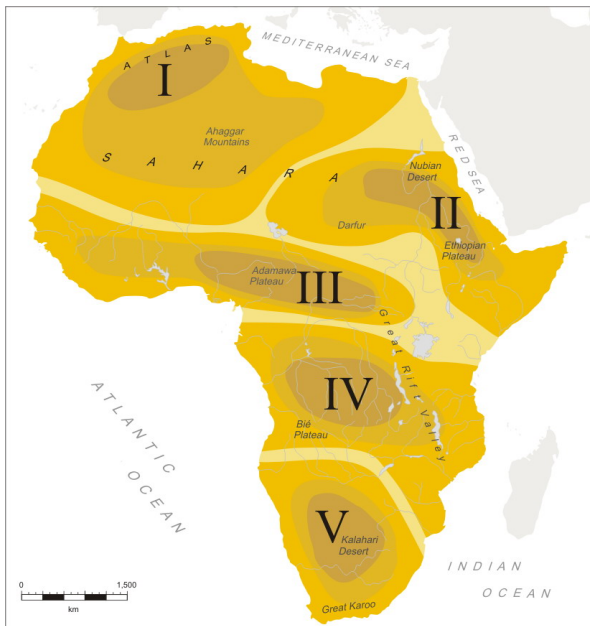
# Areas of Typological Similarity in Africa

*It has been observed since long that the languages of the African continent exhibit **typological similarities** that are **geographically conditioned***

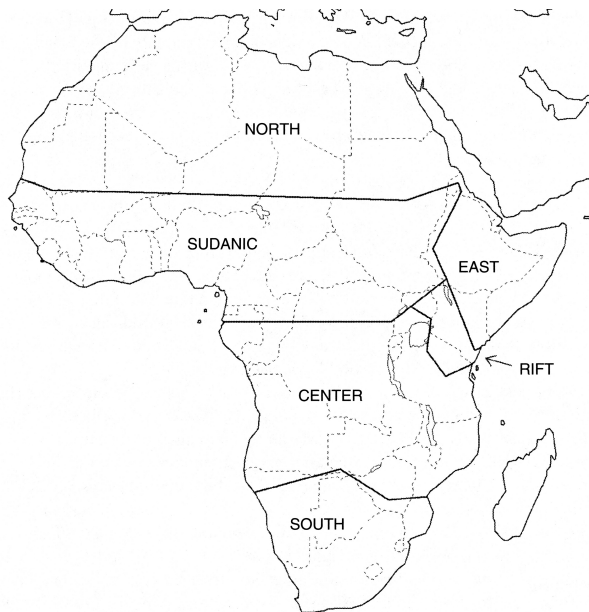
- Some researchers have distilled large-scale convergence areas (e.g., Güldemann 2008, Heine 1976, 2011, Segerer 2015)
- These areas may reflect patterns of social interaction, ethnographic similarities, geographical conduits or barriers (mountains, rivers), language family expansions, ...



# Güldemann 2010:576's African Macro-Areas



# Clements and Rialland 2008:37's African Macro-Areas



# Delimiting Macro-Areas

- Even the even the largest previous studies ultimately rely on **eyeball inspection** of features plotted on a map to define the areas
- Today we will compare with a computerized approach that  
*Given geolocated language data as input delineates the area(s) with the greatest homogeneity*
- Computational approaches have the advantage of containing **no subjectivity**, but, on the other hand, make some **simplifying assumptions**.
- Very little computational previously done on this topic (Daumé 2009, Michael et al. 2014, Muysken et al. 2015)

# What is a linguistic (macro-)area?

*... a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor. (Thomason 2001: 99)*

- Geographical region.
- Three or more languages.
- Some shared structural features.
- Result of contact.
- Not result of accident.
- Not result of inheritance from a common ancestor.

# Data Used Today

*Grambank is a database of*

- 195 **structural** (typological) features for
  - 1009 **languages** of the world
- 
- |   |                                     |
|---|-------------------------------------|
| ✓ Mainly built by student assistants reading grammars | ✓ Several dozen people contributed  |
| ✓ 87.0% inter-coder agreement                         | ✓ Release to the public in May 2018 |
| ✓ Each language is densely filled-in                  | ✓ Funded by the MPISHH, Jena        |
- 
- Centrepoint language coordinates and family/subfamily classification from Glottolog 3.0 (<http://glottolog.org>)



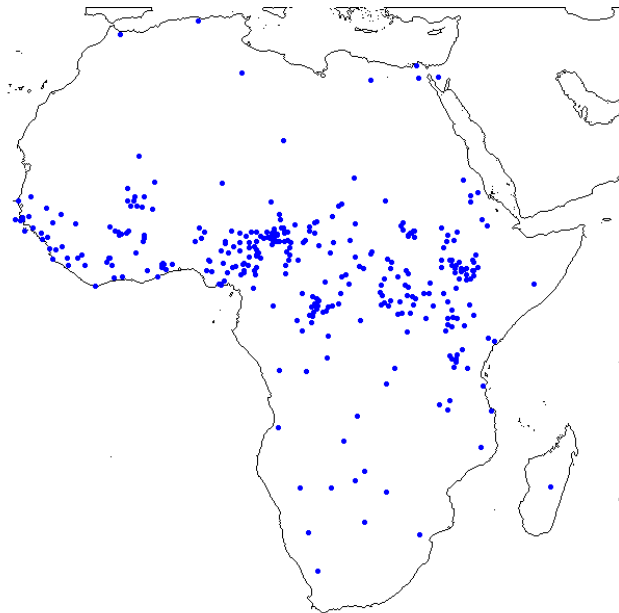
# African Languages in Grambank

- 337 languages
- 135.3 features filled on average per language
- 44 lineages represented (= all lineages with a grammar sketch as per Glottolog)

Atlantic-Congo	148
Afro-Asiatic	68
Central Sudanic	24
Mande	14

... ..

- 71 secondary lineages represented (= nearly all such)



# Homogeneity: How similar is a set of languages?

Let  $S$  be a set of languages

$$H(S) = \frac{\sum_{F_i} F_i^{majority}(S)}{\sum_{F_i} F_i(S)}$$

Homogeneity  $H(S)$  of that set is the number of datapoints with the majority value (in  $S$ )  $F_i^{majority}(S)$  for each feature  $F_i$  divided by the total number of datapoints  $\sum_{F_i} F_i(S)$

Example  $S = \{L_1, L_2, L_3, L_4\}$

	$L_1$	$L_2$	$L_3$	$L_4$	$F_i^{maj}(S)$	$F_i(S)$
$F_1$	1	?	1	0	2	3
$F_2$	?	?	1	0	1	2
$F_3$	1	?	0	0	2	3
$F_4$	0	?	?	?	1	1
$F_5$	1	1	1	1	4	4
					10	13

- $\sum_{F_i} F_i^{majority}(S) = 10$
- $\sum_{F_i} F_i(S) = 13$
- $H(S) = \frac{10}{13} = 0.769$

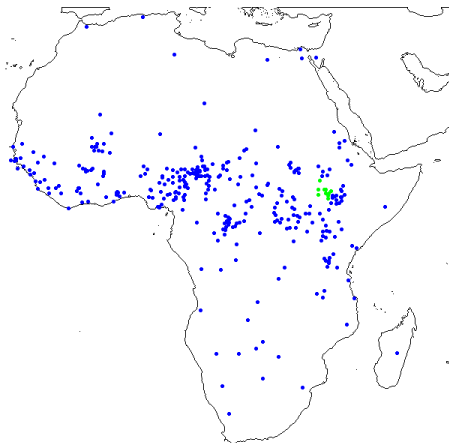
*This calculation makes sense if we assume **equal weight** and **independence** of features*

# Homogeneity Example: Shabo-10

Consider  $S$  the set of 10 languages closest to Shabo (Shabo-10)

Majang	[mpe]	Me'en	[mym]
Murle	[mur]	Sheko	[she]
Uduk	[udu]	Kafa	[kbr]
Anuak	[anu]	Shabo	[sbf]
Dizin	[mdx]	Bench	[bcq]

$$H(S) = \frac{1030}{1312} \approx 0.785$$

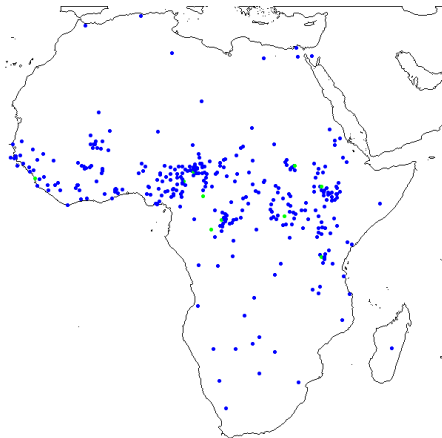


# Homogeneity Example: Random-10

Consider  $S$  the set of 10 *randomly* chosen languages

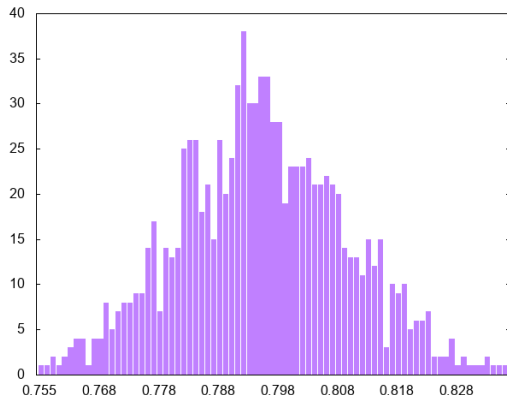
Iraqw	[irk]	Lobala	[loq]
Bullom So	[buy]	Shabo	[sbf]
Mazagway	[dkx]	Mamvu	[mdi]
Tira	[tic]	Gaa	[ttb]
Bongili	[bui]	NW Gbaya	[gya]

$$H(S) = \frac{1049}{1333} \approx 0.787$$



# Significance of Homogeneity

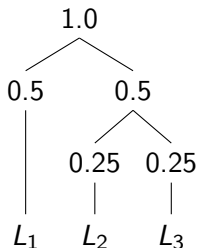
What homogeneity can one expect from a set 10 languages?



- Draw 1000 random samples of size 10 (avg homogeneity is 0.795)
- The *significance*  $H^{SIG}(S)$  is the proportion of random subsets of size  $|S|$  beaten in homogeneity by  $H(S)$

## Lineage Conditioned Homogeneity

- Intuition: Every lineage has 1 point, to be divided up according to its family tree
- This given a distribution of weights for each language, such that the weights of all languages in the same lineage sum to one point



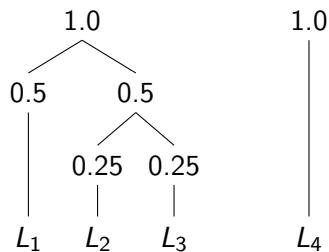
- Instead of giving one point per language, we instead give agreement points and totals corresponding to the weight of each language

$$HF(S) = \frac{\sum_{F_i} F_i^{maj}(S \cdot W_L)}{\sum_{F_i} F_i(S \cdot W_L)}$$

# Lineage Conditioned Homogeneity: Example

Given  $S = \{L_1, L_2, L_3, L_4\}$

For example, if we have two lineages



	$L_1$	$L_2$	$L_3$	$L_4$	$F_i^{maj}$ ( $S \cdot W_L$ )	$F_i$ ( $S \cdot W_L$ )
$W_L$	0.5	0.25	0.25	1.0		
$F_1$	1	?	1	0	0.75	1.75
$F_2$	?	?	1	0	0.25	1.25
$F_3$	1	?	0	0	1.25	1.75
$F_4$	0	?	?	?	0.5	0.5
$F_5$	1	1	1	1	2	2
					4.75	7.25

- $\sum_{F_i} F_i^{majority}(S) = 4.75$
- $\sum_{F_i} F_i(S) = 7.25$
- $HF(S) = \frac{4.75}{7.25} \approx 0.655$

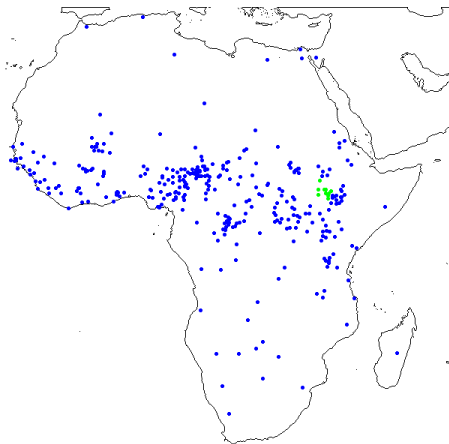
# Lineage Conditioned Homogeneity Example: Shabo-10

Consider  $S$  the set of 10 languages closest to Shabo (Shabo-10)

Language	$W_L$	Language	$W_L$
Majang	0.5	Me'en	0.25
Murle	0.25	Sheko	0.5
Uduk	1.0	Kafa	0.5
Anuak	1.0	Shabo	1.0
Dizin	0.5	Bench	0.5

$$H(S) = \frac{1030}{1312} \approx 0.785$$

$$HF(S) = \frac{595.5}{771.0} \approx 0.772$$





# What is an area? How many areas are there?

If there are  $n$  languages

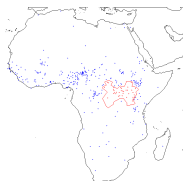
**Spherical:** Defined by a centre language and a radius

- There are no more than  $n \cdot (n - 1)$  spherical areas



**Rectangular:** Defined by a centre, breadth and width

- There are no more than  $n \cdot (n - 1) \cdot (n - 1)$  rectangular areas



**Elliptical:** Defined by centre, breadth and width

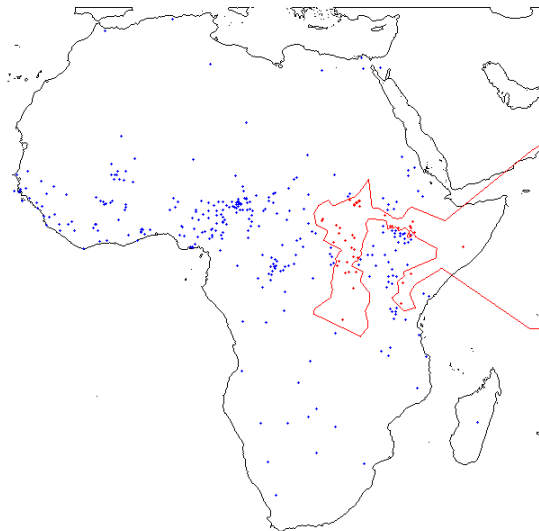
- There are no more than  $n \cdot (n - 1) \cdot (n - 1)$  elliptical areas



## But not

~~Geographically coherent but arbitrarily shaped areas~~

- These are too many to enumerate efficiently!

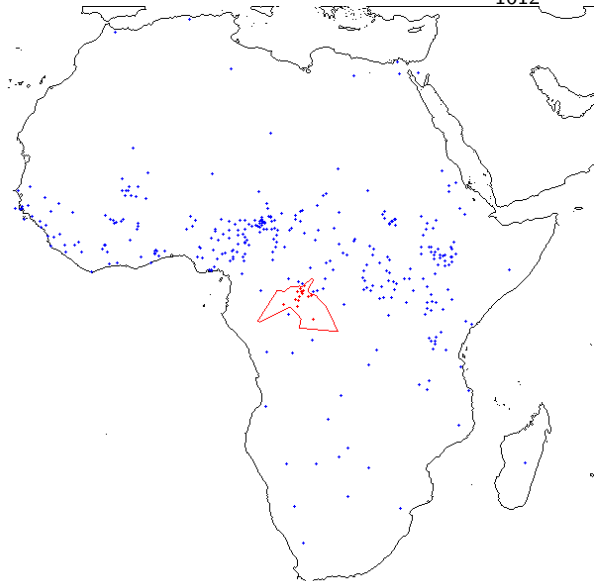


# Experiments

- Spherical, rectangular, and elliptical area shapes searched exhaustively  
*Over a million different areas checked*
- Compared in terms of *significance* (i.e., homogeneity not attainable by chance) which is comparable across areas of different sizes
- We first look at non-lineage controlled homogeneity
- And then the more interesting lineage controlled homogeneity

# Rank #1: Raw Homogeneity $H$ (Not Lineage Conditioned)

Litoka-14: Significance 1.000, Homogeneity:  $\frac{1483}{1612} \approx 0.919$



# Litoka-14: Features

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
	-	biz	bkt	bmj	bui	bws	lie
GB020. Are there definite or specific articles?	0	1	1	0	0	0	0
GB021. Do indefinite nominals commonly have indefinite articles?	0	0	0	0	0	0	0
GB022. Are there pronominal articles?	0	0	1	0	0	?	0
GB023. Are there postnominal articles?	0	1	0	0	0	?	0
GB024. What is the order of numeral and noun in the NP?	N-Num	N-Num	N-Num	N-Num	N-Num	N-Num	N-Num
GB025. What is the order of adnominal demonstrative and noun?	N-Dem	N-Dem	N-Dem	N-Dem	N-Dem	N-Dem	N-Dem
GB026. Can adnominal property words occur discontinuously?	?	?	0	?	?	?	?
GB027. Are nominal conjunction and comitative expressed by different elements?	0	0	0	0	0	0	0
GB028. Is there an inclusive/exclusive distinction?	0	0	0	0	0	0	0
GB030. Is there a gender distinction in independent 3rd person pronouns?	1	1	1	1	1	1	1
GB031. Is there a dual or unit augmented form (in addition to plural or augmented) for all person categories in the pronoun system?	0	0	0	0	0	0	0
GB035. Are there three or more distance contrasts in demonstratives?	0	0	1	0	0	?	1
GB036. Do demonstratives show an elevation distinction?	0	0	0	0	0	0	0
GB037. Do demonstratives show a visible-nonvisible distinction?	0	0	0	0	0	0	0
GB038. Are there demonstrative classifiers?	0	0	0	0	0	0	0
GB039. Is there nonphonological allomorphy of noun number markers?	1	1	1	1	1	1	?
GB041. Are there several nouns (more than three) which are suppletive for number?	?	?	?	?	?	?	?

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB042. Is there productive overt morphological singular marking on nouns?	1	1	1	1	1	1	1
GB043. Is there productive morphological dual marking on nouns?	0	0	0	0	0	0	0
GB044. Is there productive morphological plural marking on nouns?	1	1	1	1	1	1	1
GB046. Is there an associative plural marker for nouns?	?	?	?	?	?	?	?
GB047. Is there a productive morphological pattern for deriving an action/state noun from a verb?	1	1	1	1	1	?	?
GB048. Is there a productive morphological pattern for deriving an agent noun from a verb?	?	?	1	?	?	?	?
GB049. Is there a productive morphological pattern for deriving an object noun from a verb?	?	?	1	?	?	?	?
GB051. Is there a noun class/gender system where sex is a factor in class assignment?	0	0	0	0	0	0	0
GB052. Is there a noun class/gender system where shape is a factor in class assignment?	?	0	0	0	1	0	0
GB053. Is there a noun class/gender system where animacy is a factor in class assignment?	1	1	1	1	1	1	1
GB054. Is there a noun class/gender system where plant status is a factor in class assignment?	?	0	0	0	0	0	0
GB057. Are there numeral classifiers?	0	0	0	0	0	0	0
GB058. Are there possessive classifiers?	0	0	0	0	0	0	0
GB059. Is the adnominal possessive construction different for alienable and inalienable nouns?	0	0	0	0	0	0	0
GB068. Do core adjectives (defined semantically as property concepts such as value, shape, age, dimension) act like verbs in predicative position?	1	?	1	?	1	?	?

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobalo
GB069. Do core adjectives (defined semantically as property concepts; value, shape, age, dimension) used attributively require the same morphological treatment as verbs?	1	1	1	1	1	?	1
GB070. Are there morphological cases for non-pronominal core arguments (i.e. S/A/P)?	0	0	0	0	0	0	0
GB071. Are there morphological cases for pronominal core arguments (i.e. S/A/P)?	0	0	0	0	0	0	0
GB072. Are there morphological cases for oblique non-pronominal NPs (i.e. not S/A/P)?	0	1	1	0	0	0	0
GB073. Are there morphological cases for oblique independent personal pronouns (i.e. not S/A/P)?	0	?	1	0	0	0	0
GB074. Are there prepositions?	1	1	1	1	1	1	1
GB075. Are there postpositions?	0	0	0	0	0	0	0
GB081. Is there productive infixation in verbs?	0	0	0	0	0	0	0
GB082. Is there overt morphological marking of present tense on verbs?	0	?	0	?	0	1	1
GB083. Is there overt morphological marking on the verb dedicated to past tense?	1	?	0	1	0	1	1
GB084. Is there overt morphological marking on the verb dedicated to future tense?	1	?	0	?	0	1	1
GB086. Is a morphological distinction between perfective and imperfective aspect available on verbs?	?	?	?	?	?	1	?
GB089. Can the S argument be indexed by a suffix/enclitic on the verb in the simple main clause?	0	0	0	0	0	0	0
GB090. Can the S argument be indexed by a prefix/proclitic on the verb in the simple main clause?	1	1	1	1	1	1	1
GB091. Can the A argument be indexed by a suffix/enclitic on the verb in the simple main clause?	0	0	0	0	0	0	0
GB092. Can the A argument be indexed by a prefix/proclitic on the verb in the simple main clause?	1	1	1	1	1	1	1

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB093. Can the P argument be indexed by a suffix/enclitic on the verb in the simple main clause?	0	0	0	0	0	0	0
GB094. Can the P argument be indexed by a prefix/proclitic on the verb in the simple main clause?	1	1	1	1	0	1	1
GB095. Are variations in marking strategies of core participants based on TAM distinctions?	0	0	0	0	0	0	0
GB096. Are variations in marking strategies of core participants based on verb classes?	0	0	0	0	0	0	0
GB098. Are variations in marking strategies of core participants based on person distinctions?	0	0	0	0	0	0	0
GB099. Can verb stems alter according to the person of a core participant?	?	0	0	0	?	0	0
GB103. Is there a benefactive applicative marker on the verb (including indexing)?	0	?	1	1	0	1	0
GB104. Is there an instrumental applicative marker on the verb (including indexing)?	0	?	0	?	0	?	0
GB105. Can the recipient in a ditransitive construction be marked like the monotransitive patient?	?	?	1	?	?	?	?
GB107. Can standard negation be marked by an affix, clitic or modification of the verb?	1	1	1	1	1	1	1
GB108. Is there directional or locative morphological marking on verbs?	0	?	0	?	0	?	0
GB109. Is there verb suppletion for participant number?	?	0	0	0	?	0	0
GB110. Is there verb suppletion for tense or aspect?	?	?	?	?	?	0	?
GB111. Are there conjugation classes?	?	0	0	0	0	0	?
GB113. Are there verbal affixes or clitics that turn intransitive verbs into transitive ones?	?	?	1	?	1	?	?
GB114. Is there a phonologically bound reflexive marker on the verb?	1	1	1	1	?	1	1



# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobalo
GB115. Is there a phonologically bound reciprocal marker on the verb?	?	?	1	?	1	?	?
GB116. Do verbs classify the shape, size or consistency of absolutive arguments by means of incorporated nouns, verbal affixes or suppletive verb stems?	0	0	0	0	0	0	0
GB117. Is there a copula for predicate nominals?	?	?	1	?	0	1	?
GB118. Are there serial verb constructions?	?	?	0	?	?	?	?
GB119. Can mood be marked by an inflecting word ("auxiliary verb")?	?	0	?	?	?	0	?
GB120. Can aspect be marked by an inflecting word ("auxiliary verb")?	?	0	?	?	?	1	?
GB121. Can tense be marked by an inflecting word ("auxiliary verb")?	?	?	?	?	?	0	?
GB122. Is verb compounding a regular process?	?	0	0	0	0	0	0
GB123. Are there verb-adjunct (aka light-verb) constructions?	?	?	0	?	?	?	?
GB124. Is incorporation of nouns into verbs a productive intransitivizing process?	0	0	0	0	0	0	0
GB126. Is there an existential verb?	?	?	1	0	1	0	?
GB127. Are different posture verbs used obligatorily depending on an inanimate locatum's shape or position (e.g. 'to lie' vs. 'to stand')?	?	0	?	?	?	0	?
GB129. Is there a notably small number, i.e. about 100 or less, of verb roots in the language?	?	?	0	?	?	0	?
GB130. What is the pragmatically unmarked order of S and V in intransitive clauses?	SV	SV	SV	SV	SV	SV	SV
GB131. Is a pragmatically unmarked constituent order verb-initial for transitive clauses?	0	0	0	0	0	0	0
GB132. Is a pragmatically unmarked constituent order verb-medial for transitive clauses?	1	1	1	1	1	1	1

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB133. Is a pragmatically unmarked constituent order verb-final for transitive clauses?	0	0	0	0	0	0	0
GB134. Is the order of constituents the same in main and subordinate clauses?	?	?	?	?	?	?	?
GB135. Do clausal objects usually occur in the same position as nominal objects?	?	?	1	?	?	?	?
GB136. Is the order of core argument (i.e. S/A/P) constituents fixed?	1	1	1	1	1	1	1
GB137. Can standard negation be marked clause-finally?	0	1	0	0	0	1	1
GB138. Can standard negation be marked clause-initially?	1	0	1	1	1	1	1
GB139. Is there a difference between imperative (prohibitive) and declarative negation constructions?	1	?	1	1	0	1	?
GB140. Is verbal predication marked by the same negator as all of the following types of predication: locational, existential and nominal?	?	0	0	?	?	1	?
GB146. Is there a morpho-syntactic distinction between predicates expressing controlled versus uncontrolled events or states?	?	?	?	?	1	?	?
GB147. Is there a morphological passive marked on the lexical verb?	?	?	1	1	0	?	?
GB148. Is there a morphological antipassive marked on the lexical verb?	?	?	1	?	1	?	?
GB149. Is there a morphologically marked inverse on verbs?	?	0	0	0	?	0	?
GB150. Is there clause chaining?	?	?	0	?	?	?	?

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB151. Is there an overt verb marker dedicated to signalling coreference or noncoreference between the subject of one clause and an argument of an adjacent clause ("switch reference")?	?	?	?	?	?	?	?
GB152. Is there a morphologically marked distinction between simultaneous and sequential clauses?	?	?	1	?	?	?	?
GB155. Are causatives formed by affixes or clitics on verbs?	?	?	1	?	1	?	?
GB156. Is there a causative construction involving an element that is unmistakably grammaticalized from a verb for 'to say'?	?	?	0	?	?	?	?
GB158. Are verbs reduplicated?	?	?	1	?	?	?	?
GB159. Are nouns reduplicated?	?	?	?	?	?	?	?
GB160. Are elements apart from verbs or nouns reduplicated?	?	?	1	?	?	?	?
GB165. Is there productive morphological trial marking on nouns?	0	0	0	0	0	0	0
GB166. Is there productive morphological paucal marking on nouns?	0	0	0	0	0	0	0
GB167. Is there a logophoric pronoun?	0	0	0	0	0	0	0
GB170. Can an adnominal property word agree with the noun in gender/noun class?	1	1	1	1	1	1	1
GB171. Can an adnominal demonstrative agree with the noun in gender/noun class?	1	1	1	1	1	1	1
GB172. Can an article agree with the noun in gender/noun class?	0	0	1	0	0	?	0
GB177. Can the verb carry a marker of animacy of argument, unrelated to any gender/noun class of the argument visible in the NP domain?	0	0	0	0	0	0	0

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobalo
GB184. Can an adnominal property word agree with the noun in number?	1	1	1	1	1	1	1
GB185. Can an adnominal demonstrative agree with the noun in number?	1	1	1	1	1	1	1
GB186. Can an article agree with the noun in number?	0	0	1	0	0	?	0
GB187. Is there any productive diminutive marking on the noun (exclude marking by system of nominal classification only)?	?	0	0	?	?	?	?
GB188. Is there any productive augmentative marking on the noun (exclude marking by system of nominal classification only)?	?	?	0	?	?	?	?
GB192. Is there a gender system where a noun's phonological properties are a factor in class assignment?	?	0	0	0	0	0	0
GB193. What is the order of adnominal property word and noun?	N-ANM	N-ANM	N-ANM	N-ANM	N-ANM	N-ANM	N-ANM
GB196. Is there a male/female distinction in 2nd person independent pronouns?	0	0	0	0	0	0	0
GB197. Is there a male/female distinction in 1st person independent pronouns?	0	0	0	0	0	0	0
GB198. Can an adnominal numeral agree with the noun in gender/noun class?	1	1	1	1	1	1	1
GB203. What is the order of the adnominal collective universal quantifier ('all') and the noun?	N-UQ	?	N-UQ	N-UQ	N-UQ	N-UQ	N-UQ
GB204. Do collective ('all') and distributive ('every') universal quantifiers differ in their forms or their syntactic positions?	?	?	?	?	?	?	?
GB250. Can predicative possession be expressed with a transitive 'habeo' verb?	0	0	0	?	0	?	?
GB256. Can predicative possession be expressed with an S-like possessor and a possessum that is coded like a comitative argument?	?	1	1	1	?	1	1

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB265. Is there a comparative construction that includes a form that elsewhere means 'surpass, exceed'?	?	?	1	?	?	?	?
GB266. Is there a comparative construction that employs a marker of the standard which elsewhere has a locational meaning?	?	?	0	?	?	?	?
GB270. Can comparatives be expressed using two conjoined clauses?	?	?	0	?	?	?	?
GB273. Is there a comparative construction with a standard marker that elsewhere has neither a locational meaning nor a 'surpass/exceed' meaning?	?	?	0	?	?	?	?
GB275. Is there a bound comparative degree marker on the property word in a comparative construction?	?	?	0	?	?	?	?
GB276. Is there a non-bound comparative degree marker modifying the property word in a comparative construction?	?	?	0	?	?	?	?
GB291. Can polar interrogation be marked by tone?	?	?	?	?	?	1	?
GB296. Is there a phonologically or morphosyntactically definable class of ideophones that includes ideophones depicting imagery beyond sound?	1	?	1	?	1	?	?
GB297. Can polar interrogation be indicated by a V-not-V construction?	?	?	?	?	?	?	0
GB298. Can standard negation be marked by an inflecting word ("auxiliary verb")?	0	0	0	1	0	1	0
GB299. Can standard negation be marked by a non-inflecting word ("auxiliary particle")?	0	1	1	1	0	0	0
GB300. Does the verb for 'give' have suppletive verb forms?	?	?	?	?	?	?	?
GB301. Is there an inclusory construction?	?	?	?	?	?	?	?
GB302. Is there a phonologically free passive marker ('particle' or 'auxiliary')?	0	?	0	?	0	?	0
GB303. Is there a phonologically free antipassive marker ('particle' or 'auxiliary')?	0	?	0	?	0	?	0

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB304. Can the agent be expressed overtly in a passive clause?	?	?	1	?	1	?	?
GB305. Is there a phonologically independent reflexive pronoun?	?	?	1	?	?	?	?
GB306. Is there a phonologically independent non-bipartite reciprocal pronoun?	?	?	0	?	?	?	?
GB309. Are there multiple past or multiple future tenses, distinguishing distance from Time of Reference?	1	1	1	1	1	1	1
GB312. Is there overt morphological marking on the verb dedicated to mood?	0	?	1	1	1	1	1
GB313. Are there special adnominal possessive pronouns that are not formed by an otherwise regular process?	0	0	1	0	1	0	1
GB314. Can augmentative meaning be expressed productively by a shift of gender/noun class?	1	0	0	0	0	0	0
GB315. Can diminutive meaning be expressed productively by a shift of gender/noun class?	0	0	0	0	0	0	0
GB316. Is singular number regularly marked in the noun phrase by a dedicated phonologically free element?	0	0	0	0	0	0	0
GB317. Is dual number regularly marked in the noun phrase by a dedicated phonologically free element?	0	0	0	0	0	0	0
GB318. Is plural number regularly marked in the noun phrase by a dedicated phonologically free element?	0	0	0	0	0	0	0
GB319. Is trial number regularly marked in the noun phrase by a dedicated phonologically free element?	0	0	0	0	0	0	0
GB320. Is paucal number regularly marked in the noun phrase by a dedicated phonologically free element?	0	0	0	0	0	0	0

# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB321. Is there a large class of nouns whose gender/noun class is not phonologically or semantically predictable?	1	1	1	1	1	1	1
GB322. Is there grammatical marking of direct evidence (perceived with the senses)?	0	0	0	0	0	0	0
GB323. Is there grammatical marking of indirect evidence (hearsay, inference, etc.)?	0	0	0	0	0	0	0
GB324. Is there an interrogative verb for content interrogatives (who?, what?, etc.)?	?	0	0	0	?	0	?
GB325. Is there a count/mass distinction in interrogative quantifiers?	?	0	0	?	0	?	?
GB326. Do (nominal) content interrogatives normally or frequently occur in situ?	?	?	?	?	?	?	?
GB327. Can the relative clause follow the noun?	1	1	1	1	?	1	1
GB328. Can the relative clause precede the noun?	?	?	?	?	?	?	?
GB329. Are there internally-headed relative clauses?	?	?	?	?	?	?	?
GB330. Are there correlative relative clauses?	?	?	?	?	?	?	?
GB331. Are there adjoined relative clauses?	?	?	?	?	?	?	?
GB333. Is there a decimal numeral system?	?	?	1	?	?	?	?
GB334. Is there synchronic evidence for any element of a quinary numeral system?	?	?	0	?	?	0	?
GB335. Is there synchronic evidence for any element of a vigesimal numeral system?	?	?	0	?	?	?	?
GB336. Is there a body-part tallying system?	?	?	0	?	?	?	?
GB400. Are all person categories neutralized in some voice, tense, aspect, mood and/or negation?	?	0	0	0	?	0	?
GB401. Is there a class of patient-labile verbs?	?	?	?	?	?	?	?
GB408. Is there any accusative alignment of flagging?	?	0	?	?	?	0	?
GB409. Is there any ergative alignment of flagging?	?	0	?	?	?	0	?
GB410. Is there any neutral alignment of flagging?	?	1	?	?	?	1	?

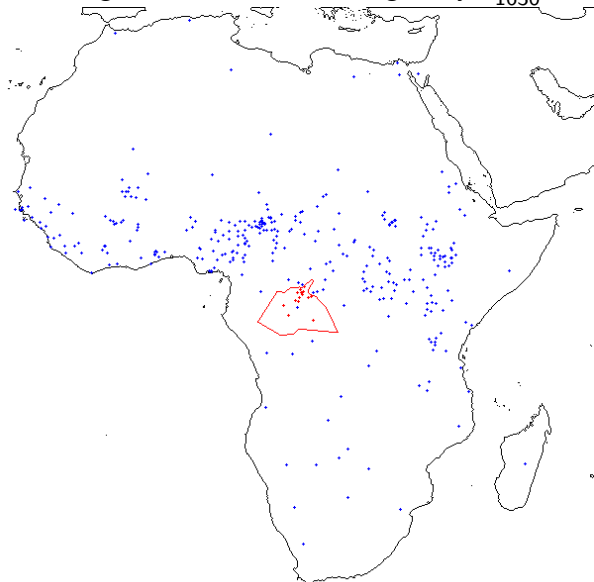
# Litoka-14: Features Cont'd

	Litoka	Loi-Likila	Boloki	Bamwe	Bongili	Bomboma	Balobo
GB415. Is there a politeness distinction in 2nd person forms?	?	0	?	?	?	0	?
GB430. Can adnominal possession be marked by a prefix on the possessor?	0	0	1	0	0	0	0
GB431. Can adnominal possession be marked by a prefix on the possessed noun?	0	0	0	0	0	0	0
GB432. Can adnominal possession be marked by a suffix on the possessor?	0	0	0	0	0	0	0
GB433. Can adnominal possession be marked by a suffix on the possessed noun?	0	0	0	0	1	0	0
GB519. Can mood be marked by a non-inflecting word ("auxiliary particle")?	?	0	?	?	?	0	?
GB520. Can aspect be marked by a non-inflecting word ("auxiliary particle")?	?	0	?	?	?	0	?
GB521. Can tense be marked by a non-inflecting word ("auxiliary particle")?	?	0	?	?	?	0	?
GB522. Can the S or A argument be omitted from a pragmatically unmarked clause when the referent is inferable from context ("pro-drop" or "null anaphora")?	?	1	?	?	?	1	?



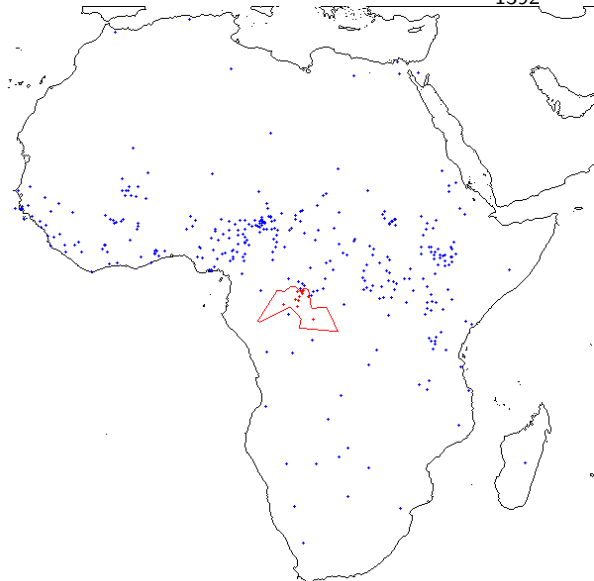
## Rank #2: Raw Homogeneity $H$ (Not Lineage Conditioned)

Litoka-13: Significance 0.999, Homogeneity:  $\frac{1497}{1630} \approx 0.904$



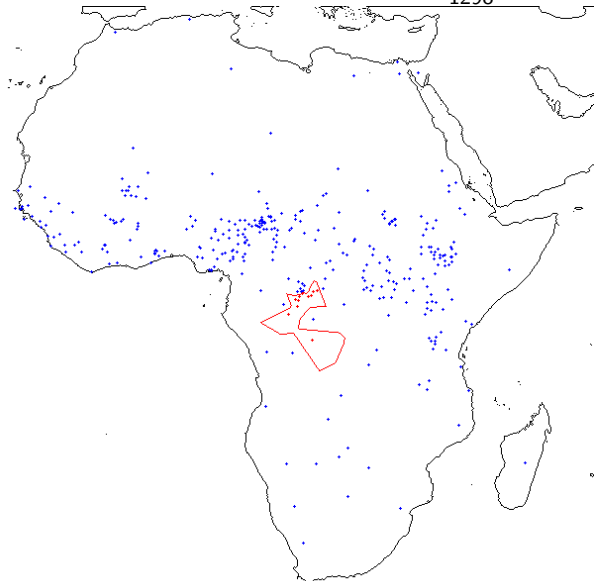
## Rank #3: Raw Homogeneity $H$ (Not Lineage Conditioned)

Litoka-11: Significance 0.999, Homogeneity:  $\frac{1282}{1392} \approx 0.922$



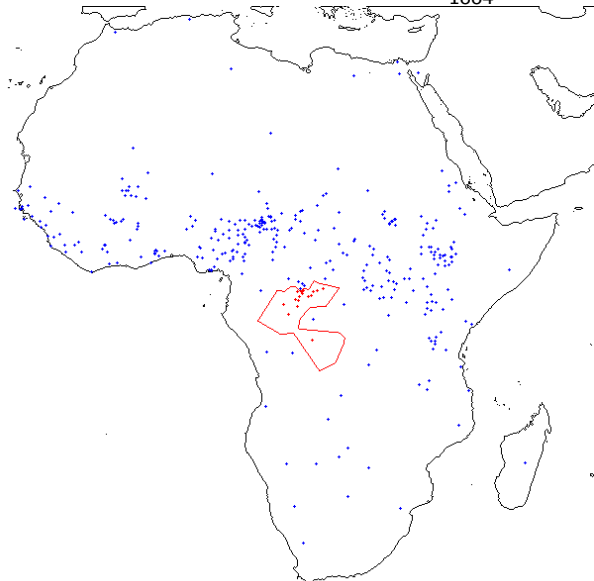
## Rank #4: Raw Homogeneity $H$ (Not Lineage Conditioned)

Nzadi-11: Significance 0.998, Homogeneity:  $\frac{1175}{1298} \approx 0.905$ , 1298



## Rank #59: Raw Homogeneity $H$ (Not Lineage Cond.)

Nzadi-16: Significance 0.951, Homogeneity:  $\frac{1696}{1884} \approx 0.900$ , 1298



# Purging Variant Areas

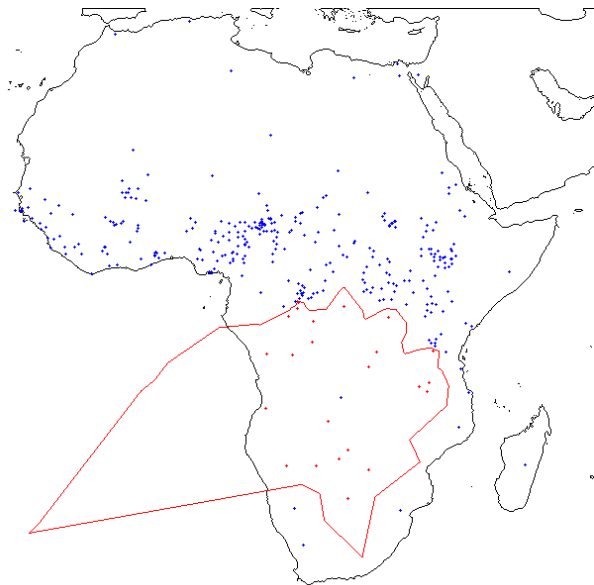
- If an area  $S'$  extends an area  $S$  but has lower significance than  $S$ , we kick out  $S'$

$$H^{SIG}(S) > H^{SIG}(S') \text{ and } S \subset S'$$

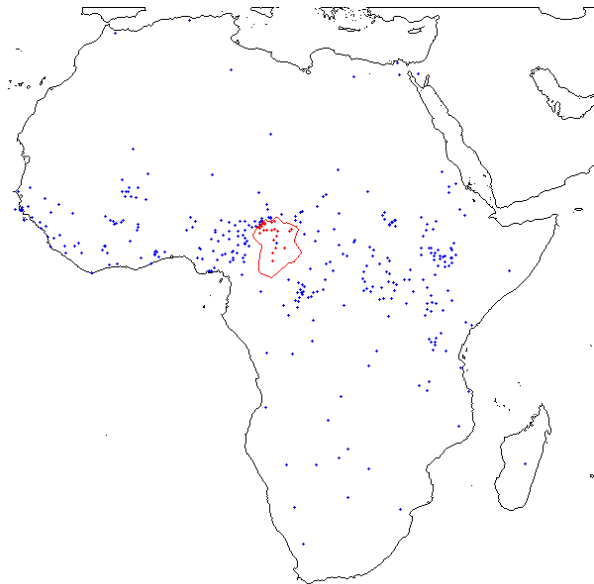
- If an area  $S'$  is encompassed in a larger area  $S$  but has lower significance than  $S$ , we kick out  $S'$

$$H^{SIG}(S) > H^{SIG}(S') \text{ and } S' \subset S$$

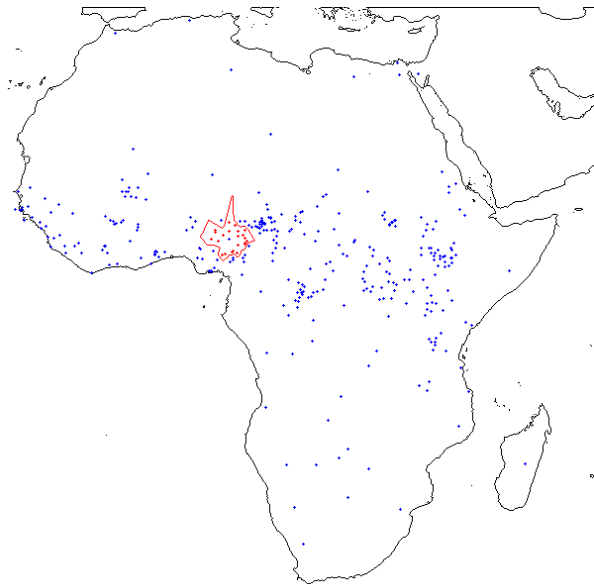
## After Purging Rank #2: Raw Homogeneity $H$



## After Purging Rank #3: Raw Homogeneity $H$

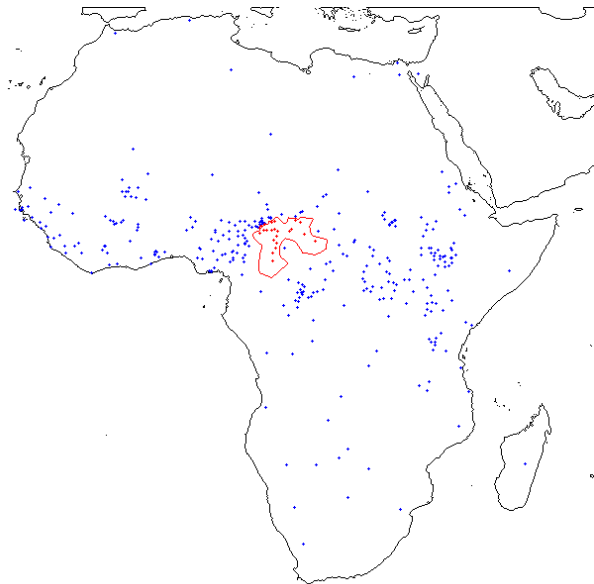


## After Purging Rank #4: Raw Homogeneity $H$

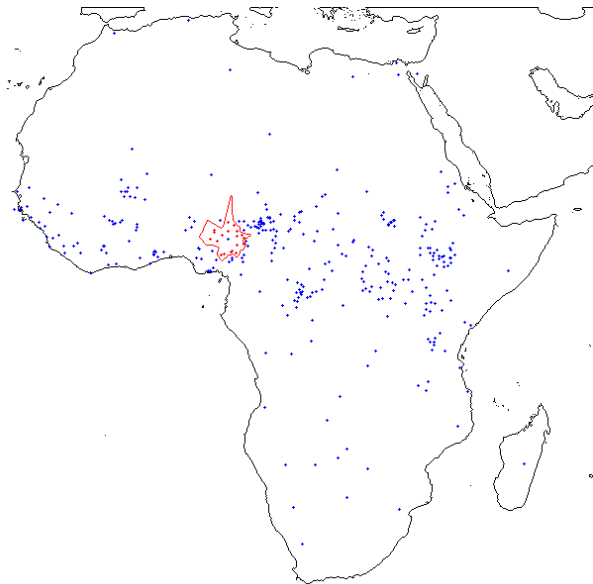




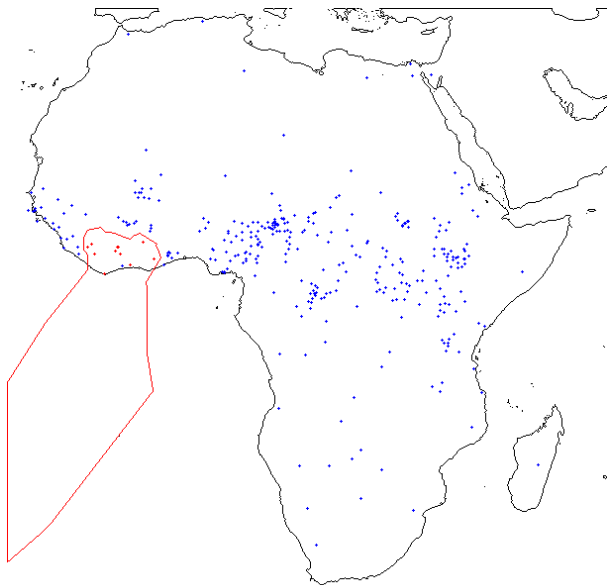
## After Purging Rank #5: Raw Homogeneity $H$



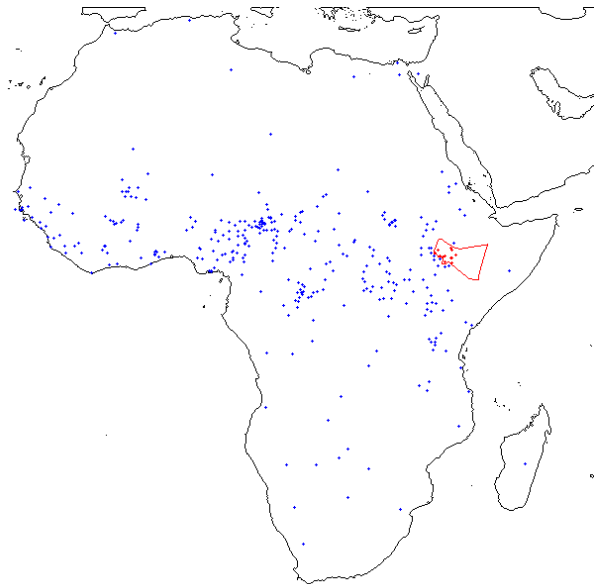
## After Purging Rank #6: Raw Homogeneity $H$



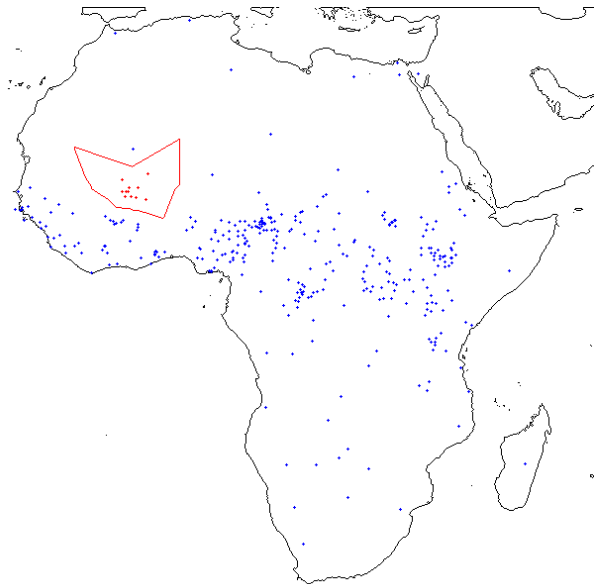
## After Purging Rank #7: Raw Homogeneity $H$



## After Purging Rank #8: Raw Homogeneity $H$

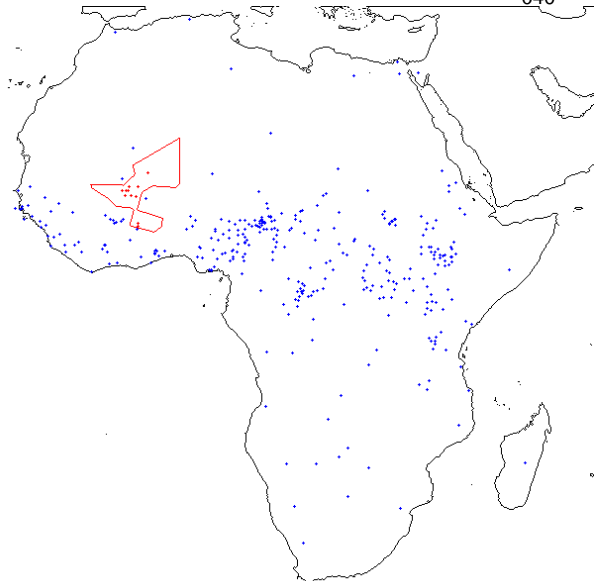


## After Purging Rank #9: Raw Homogeneity $H$



# Rank #1: Controlled Homogeneity $HF$ (After Purging)

Bangi Me-11: Significance 0.845, Homogeneity:  $\frac{547.83}{640} \approx 0.855$



# Features of Bangi Me-11

	Bangime	Ben Tey	Jamsay	Tengou-Togo	Tommo So	Humburi	Korom
GB433. Can adnominal possession be marked by a suffix on the possessed noun?	dba ?	dbt 0	djm 0	dtk 0	dto 0	hmb 0	kfz 0
GB432. Can adnominal possession be marked by a suffix on the possessor?	?	0	0	0	0	0	0
GB431. Can adnominal possession be marked by a prefix on the possessed noun?	?	0	0	0	0	0	0
GB430. Can adnominal possession be marked by a prefix on the possessor?	?	0	0	0	0	0	0
GB400. Are all person categories neutralized in some voice, tense, aspect, mood and/or negation?	0	0	0	?	?	0	0
GB336. Is there a body-part tallying system?	?	0	0	0	0	0	0
GB334. Is there synchronic evidence for any element of a quinary numeral system?	?	0	0	0	0	0	0
GB325. Is there a count/mass distinction in interrogative quantifiers?	?	0	0	0	0	0	0
GB324. Is there an interrogative verb for content interrogatives (who?, what?, etc.)?	?	0	0	0	0	0	0
GB320. Is paucal number regularly marked in the noun phrase by a dedicated phonologically free element?	?	0	0	0	0	0	0
GB319. Is trial number regularly marked in the noun phrase by a dedicated phonologically free element?	?	0	0	0	0	0	0
GB317. Is dual number regularly marked in the noun phrase by a dedicated phonologically free element?	?	0	0	0	0	0	0

# Features of Bangi Me-11 Cont'd

	Bangime	Ben Tey	Jamsay	Tengou-Togo	Tommo So	Humburi	Korom
GB316. Is singular number regularly marked in the noun phrase by a dedicated phonologically free element?	?	0	0	0	0	0	0
GB305. Is there a phonologically independent reflexive pronoun?	?	1	1	1	?	1	1
GB304. Can the agent be expressed overtly in a passive clause?	?	0	0	0	0	0	0
GB303. Is there a phonologically free antipassive marker ('particle' or 'auxiliary')?	?	0	0	0	0	0	0
GB302. Is there a phonologically free passive marker ('particle' or 'auxiliary')?	?	0	0	0	0	0	0
GB301. Is there an inclusory construction?	?	0	0	0	0	0	0
GB300. Does the verb for 'give' have suppletive verb forms?	?	0	?	?	?	0	0
GB297. Can polar interrogation be indicated by a V-not-V construction?	0	0	0	0	0	0	0
GB291. Can polar interrogation be marked by tone?	0	0	0	0	0	?	0
GB286. Can polar interrogation be indicated by overt verbal morphology only?	0	0	0	0	0	0	0
GB285. Can polar interrogation be marked by a question particle and verbal morphology?	0	0	0	0	0	0	0
GB275. Is there a bound comparative degree marker on the property word in a comparative construction?	0	0	0	0	0	0	0
GB270. Can comparatives be expressed using two conjoined clauses?	0	?	?	0	0	0	0
GB264. Is there a polar interrogative particle that most commonly occurs neither clause-initially nor clause-finally?	0	0	0	0	0	0	0
GB260. Can polar interrogation be indicated by a special word order?	0	0	0	0	0	0	0
GB203. What is the order of the adnominal collective universal quantifier ('all') and the noun?	N-UQ	N-UQ	N-UQ	N-UQ	N-UQ	N-UQ	N-UQ



# Features of Bangi Me-11 Cont'd

	Bangime	Ben Tey	Jamsay	Tengou-Togo	Tommo So	Humburi	Korom
GB197. Is there a male/female distinction in 1st person independent pronouns?	0	0	0	0	0	0	0
GB196. Is there a male/female distinction in 2nd person independent pronouns?	0	0	0	0	0	0	0
GB188. Is there any productive augmentative marking on the noun (exclude marking by system of nominal classification only)?	?	0	0	0	0	0	0
GB177. Can the verb carry a marker of animacy of argument, unrelated to any gender/noun class of the argument visible in the NP domain?	0	0	0	0	0	0	0
GB166. Is there productive morphological paucal marking on nouns?	0	0	0	0	0	0	0
GB165. Is there productive morphological trial marking on nouns?	0	0	0	0	0	0	0
GB156. Is there a causative construction involving an element that is unmistakably grammaticalized from a verb for 'to say'?	0	0	0	0	0	0	0
GB149. Is there a morphologically marked inverse on verbs?	0	0	0	0	0	0	0
GB129. Is there a notably small number, i.e. about 100 or less, of verb roots in the language?	0	0	0	0	0	0	0
GB126. Is there an existential verb?	1	1	1	1	1	1	1
...							

# Features of Bangi Me-11 Cont'd

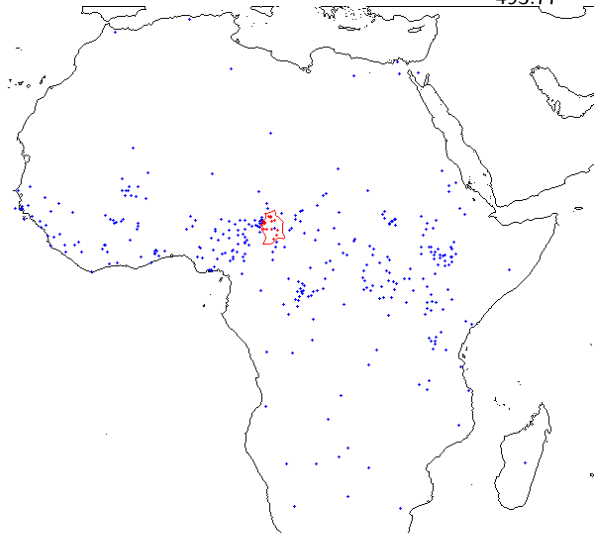
	Bangime	Ben Tey	Jamsay	Tengou-Togo	Tommo So	Humburi	Korom
...							
GB326. Do (nominal) content interrogatives normally or frequently occur in situ?	?	1	1	1	1	0	0
GB299. Can standard negation be marked by a non-inflecting word ("auxiliary particle")?	?	0	0	0	0	1	1
GB140. Is verbal predication marked by the same negator as all of the following types of predication: locational, existential and nominal?	1	0	1	1	0	0	0
GB185. Can an adnominal demonstrative agree with the noun in number?	1	1	1	1	0	0	1
GB137. Can standard negation be marked clause-finally?	0	1	1	1	1	0	0
GB132. Is a pragmatically unmarked constituent order verb-medial for transitive clauses?	1	0	0	0	0	1	1
GB042. Is there productive overt morphological singular marking on nouns?	0	1	1	0	1	0	1
GB522. Can the S or A argument be omitted from a pragmatically unmarked clause when the referent is inferrable from context ("pro-drop" or "null anaphora")?	?	?	?	?	?	?	0
GB521. Can tense be marked by a non-inflecting word ("auxiliary particle")?	?	?	?	?	?	?	0
GB520. Can aspect be marked by a non-inflecting word ("auxiliary particle")?	?	?	?	?	?	?	0
GB422. Is there a postposed complementizer in complements of verbs of thinking and/or knowing?	?	?	?	?	?	?	1
GB421. Is there a preposed complementizer in complements of verbs of thinking and/or knowing?	?	?	?	?	?	?	0
GB415. Is there a politeness distinction in 2nd person forms?	?	?	?	?	?	?	1
GB401. Is there a class of patient-labile verbs?	?	?	?	0	?	?	?
GB257. Can polar interrogation be marked by intonation only?	?	0	0	0	1	?	0

# Features of Bangi Me-11 Cont'd

	Bangime	Ben Tey	Jamsay	Tengou-Togo	Tommo So	Humburi	Korom
GB121. Can tense be marked by an inflecting word ("auxiliary verb")?	1	?	?	?	?	?	1
GB086. Is a morphological distinction between perfective and imperfective aspect available on verbs?	0	1	1	1	1	0	1
GB074. Are there prepositions?	0	0	0	0	0	1	1
GB071. Are there morphological cases for pronominal core arguments (i.e. S/A/P)?	0	1	0	0	1	1	0
GB059. Is the adnominal possessive construction different for alienable and inalienable nouns?	?	1	1	0	1	1	0

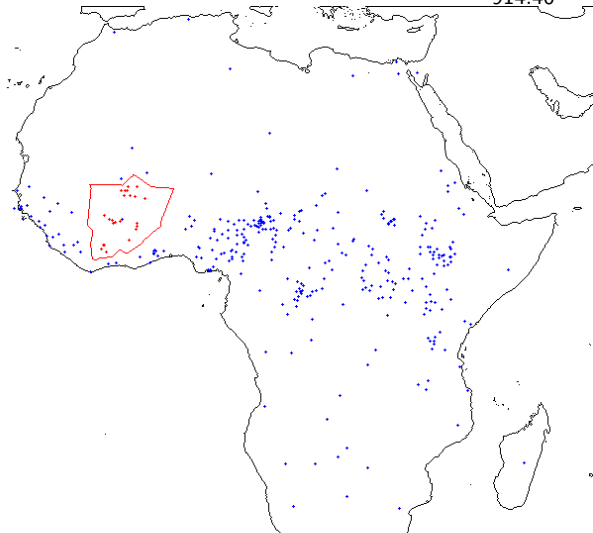
## Rank #2: Lineage Controlled Homogeneity $HF$ (After Purging)

Bagirmi-15: Significance 0.838, Homogeneity:  $\frac{422.58}{493.77} \approx 0.856$



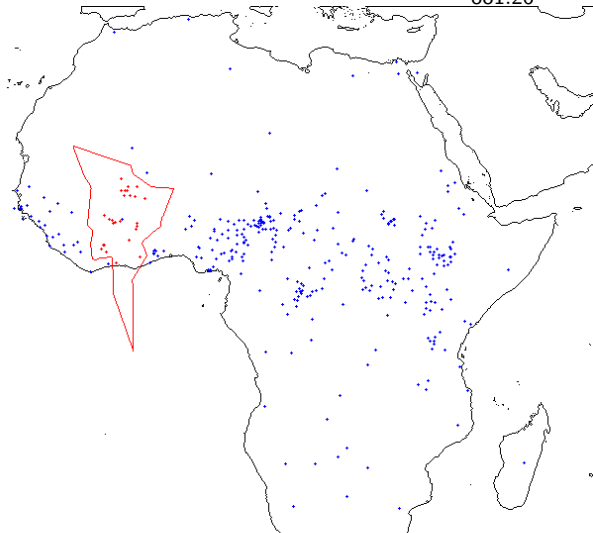
## Rank #3: Lineage Controlled Homogeneity $HF$ (After Purging)

Chakali-23: Significance 0.827, Homogeneity:  $\frac{775.05}{914.40} \approx 0.848$



## Rank #4: Lineage Controlled Homogeneity $HF$ (After Purging)

Abé-26: Significance 0.827, Homogeneity:  $\frac{747.08}{881.26} \approx 0.848$



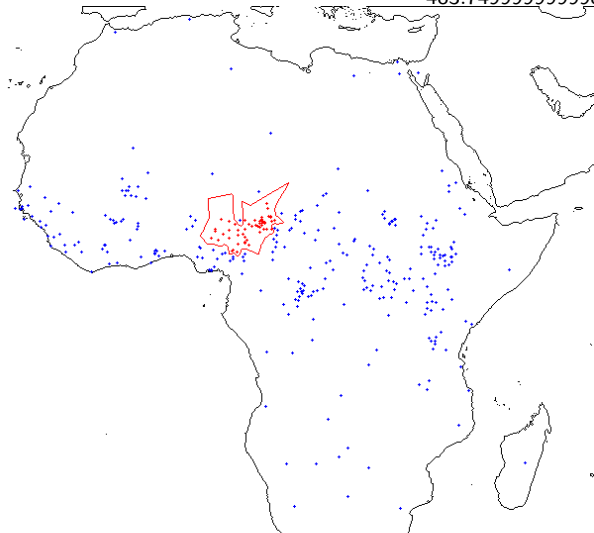
## Rank #5: Lineage Controlled Homogeneity $HF$ (After Purging)

Bench-13: Significance 0.826, Homogeneity:  $\frac{273.87}{323.97} \approx 0.845$



## Rank #6: Lineage Controlled Homogeneity *HF* (After Purging)

Goemai-53: Significance 0.817, Homogeneity:  $\frac{405.711805555553}{483.7499999999824} \approx 0.838$





# Conclusions

- Presented one automated technique for testing which linguistic areas show the most homogeneity
- Still no automated technique for transforming a long list of more or less significant areas into a distilled picture that can be compared to human divisions
- Many assumptions made in the quest for objectivity
- Ideas on how to weigh/combine features greatly appreciated
- Ideas on how to incorporate different time-depth of families much appreciated

# Assumptions

- Features have equal weight
- Features are independent
- Absences count as much as presences
- The entire typological profile counts, not sections of it
- Linguistic areas are of regular shape type
- Genealogical dependence can be captured by a naive decay formula
- Language family classification is sufficiently correct
- Centrepoin coordinate adequately represent geographical location
- Grambank contains adequate sample of features and languages

Thank you



- Clements, N. and Rialland, A. (2008). Africa as a phonological area. In Heine, B. and Nurse, D., editors, *A linguistic geography of Africa*, pages 36–87. Cambridge: Cambridge University Press.
- Daumé, III, H. (2009). Non-parametric bayesian areal linguistics. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL '09*, pages 593–601, Morristown, NJ, USA. Association for Computational Linguistics.
- Güldemann, T. (2008). The macro-sudan belt: towards identifying a linguistic area in northern sub-saharan africa. In Heine, B. and Nurse, D., editors, *A linguistic geography of Africa*, pages 151–185. Cambridge: Cambridge University Press.
- Güldemann, T. (2010). "sprachraum" and geography: Linguistic macro-areas in africa. In Lameli, A., Kehrein, R., and Rabanus, S., editors, *Language and Space: An International Handbook of Linguistic Variation Volume 2: Language Mapping*, volume 30/2 of *Handbooks of Linguistics and Communication Science*, pages 561–585. Berlin: Mouton de Gruyter.

- Heine, B. (1976). *A Typology of African Languages*, volume 4 of *Kölner Beiträge zur Afrikanistik*. Berlin: Dietrich Reimer.
- Heine, B. (2011). Areas of grammaticalization and geographical typology. In Hieda, O., König, C., and Nakagawa, H., editors, *Geographical Typology and Linguistic Areas: With Special Reference to Africa*, pages 41–66. Amsterdam: John Benjamins, Amsterdam.
- Michael, L., Chang, W., and Stark, T. (2014). Exploring phonological areality in the circum-andean region using a naive bayes classifier. *Language Dynamics and Change*, 4(1):27–86.
- Muysken, P., Hammarström, H., Birchall, J., van Gijn, R., Krasnoukhova, O., and Müller, N. (2015). Linguistic areas, bottom up or top down? the case of the guaporé-mamoré region. In Comrie, B. and Golluscio, L., editors, *Language Contact and Documentation*, pages 205–238. Berlin: DeGruyter Mouton.
- Segeer, G. (2015). How databases shape research: labial-velars distribution in africa. Paper presented at the Workshop Areal Phenomena in Northern Sub-Saharan Africa (8th World Congress of African Linguistics), August 20-24, 2015, Kyoto, Japan.