A SINGLE FEATURE APPROACH TO LINGUISTIC AREAS: LABIAL-VELARS AND THE PREHISTORY OF THE MACRO-SUDAN BELT

Dmitry Idiatov & Mark Van de Velde

LLACAN (CNRS – INALCO)

dmitry.idiatov@cnrs.fr

mark.vandevelde@cnrs.fr

- LV stops such as k͡p, g͡b and ŋ͡m, are commonly found in NSSA languages but typologically they are known to be rare (e.g., Cahill 2008, Maddieson 2011)

- In NSSA, the languages that have them are genealogically diverse, but geographically clustered

- They are therefore used among other features to define linguistic areas in NSSA: the Macro-Sudan belt (Güldemann 2008) & the Sudanic Zone (Clements & Rialland 2008)

- Where and how did LV stops arise in NSSA and how did they spread?
  - innovation through sound change (Westermann 1911)?
  - inheritance (Greenberg 1983, Cahill 2017)?
  - borrowing of phonemes through loanwords?
  - substrate interference?
  - "diffusion"?
- In order to answer these questions, we studied the lexical distribution of LV stops in the languages that have them

Our **main findings**:

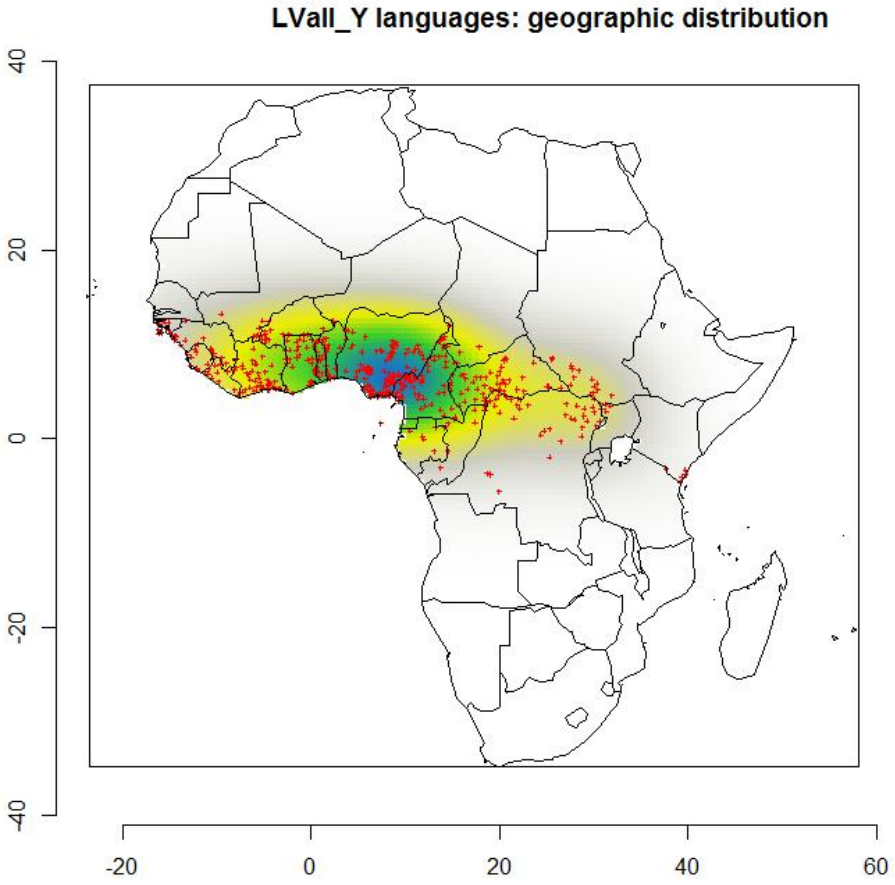
- Languages with LV vary significantly with respect to the status of LV in their phonologies and lexicons:
- In many of the languages with LV stops, they have a much lower lexical frequency than average consonant phonemes
- Languages with higher lexical frequencies of LV stops are grouped into three areal hotbeds
- LV stops have a skewed lexical distribution, both phonotactically (stem-initial position) and semantically (expressive vocabulary)

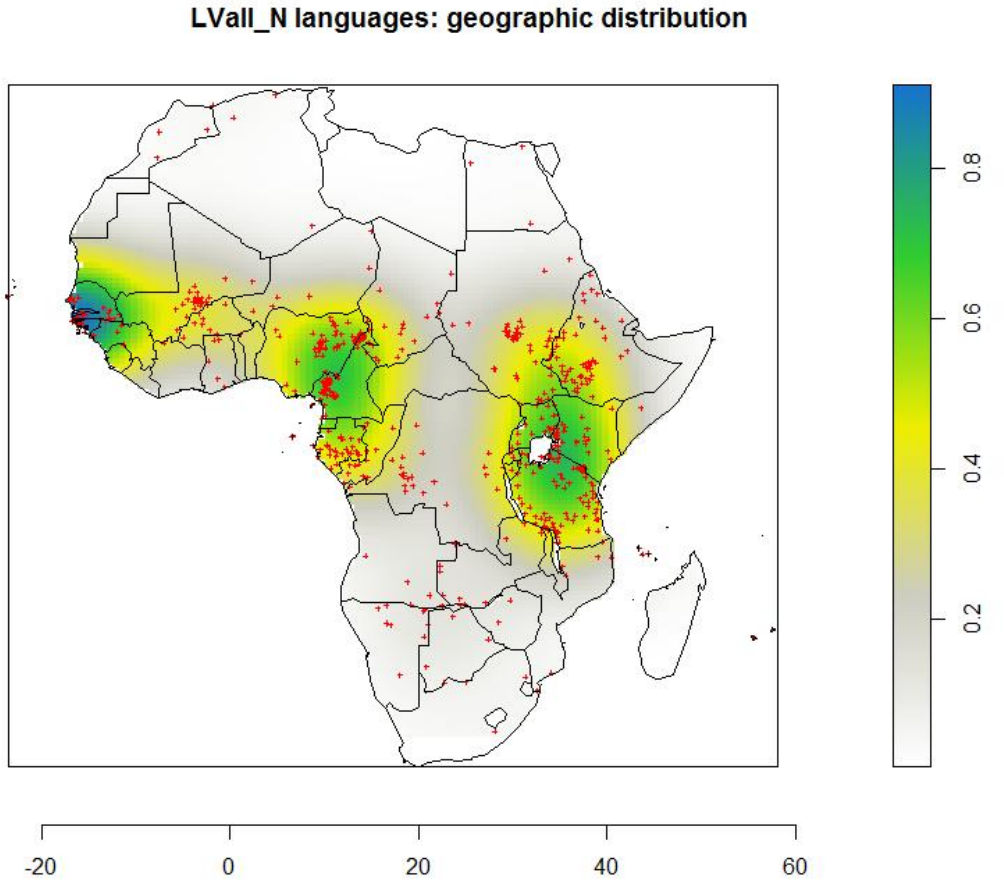
INTRODUCTION

- We argue that these findings strengthen the case against common inheritance
- The most straightforward explanation is that LV stops or the phonetic traits that can lead to their emergence were a genealogical or areal trait of the vanished languages of West and Central Africa, currently supplanted mostly by various Niger-Congo subgroups

LV data sources:

- **RefLex**, www.reflex.cnrs.fr, LVFreq data
- Phoible, www.phoible.org, YN data
- Additional LVFreq data for some Mande and Bantu languages


LVall_Y languages: geographic distribution
40
20
0
-20
-40
-20
0
20
40
60
0
0.2
0.4
0.6
0.8
1
1.2


LVall_N languages: geographic distribution
-20
0
20
40
60
0.2
0.4
0.6
0.8

LVFreq estimation

H_0: In a lexicon, all C phonemes have equal frequency (have equal probability of occurrence)

$$\boldsymbol{LVFreq} = \frac{LV_O}{LV_E} * 100\% = \frac{\sum T_{LV}}{\frac{\sum T_C}{\sum P_C} * \sum P_{LV}} * 100\%$$

LV_O - observed LV count

LV_E - expected LV count

T_{LV} - LV token

T_C - any C token

P_{LV} - LV phoneme

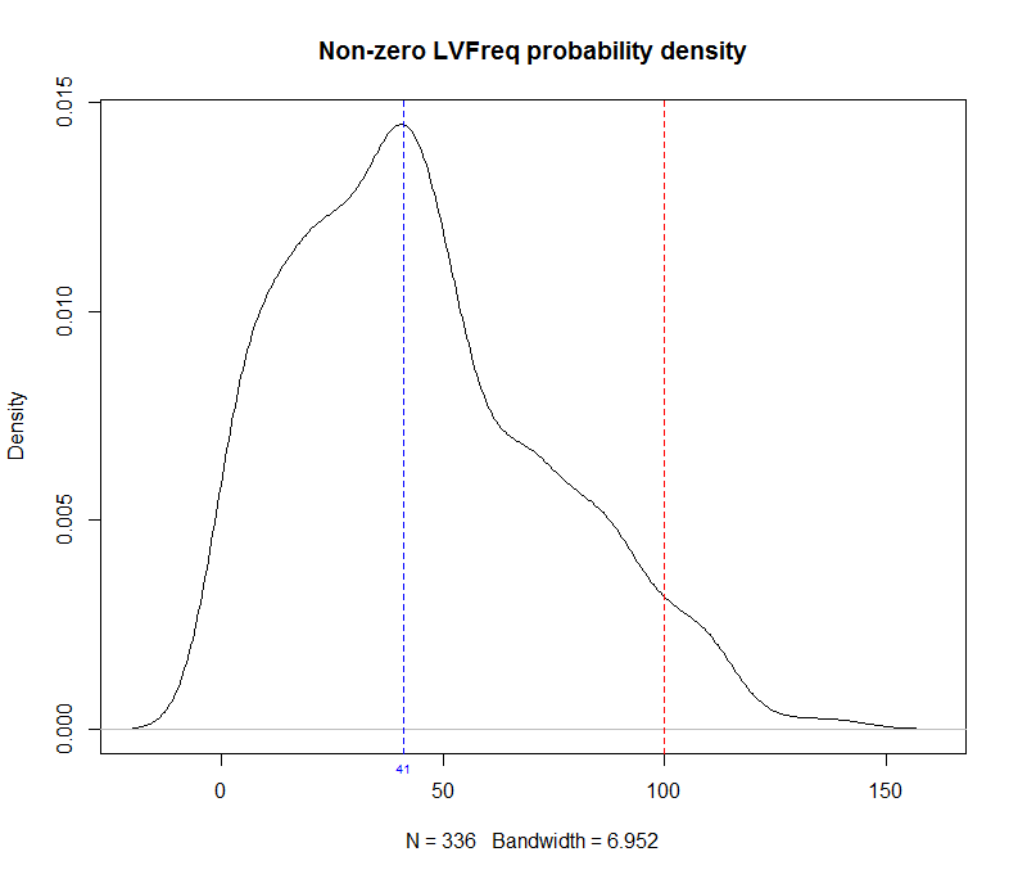
P_C - any C phoneme

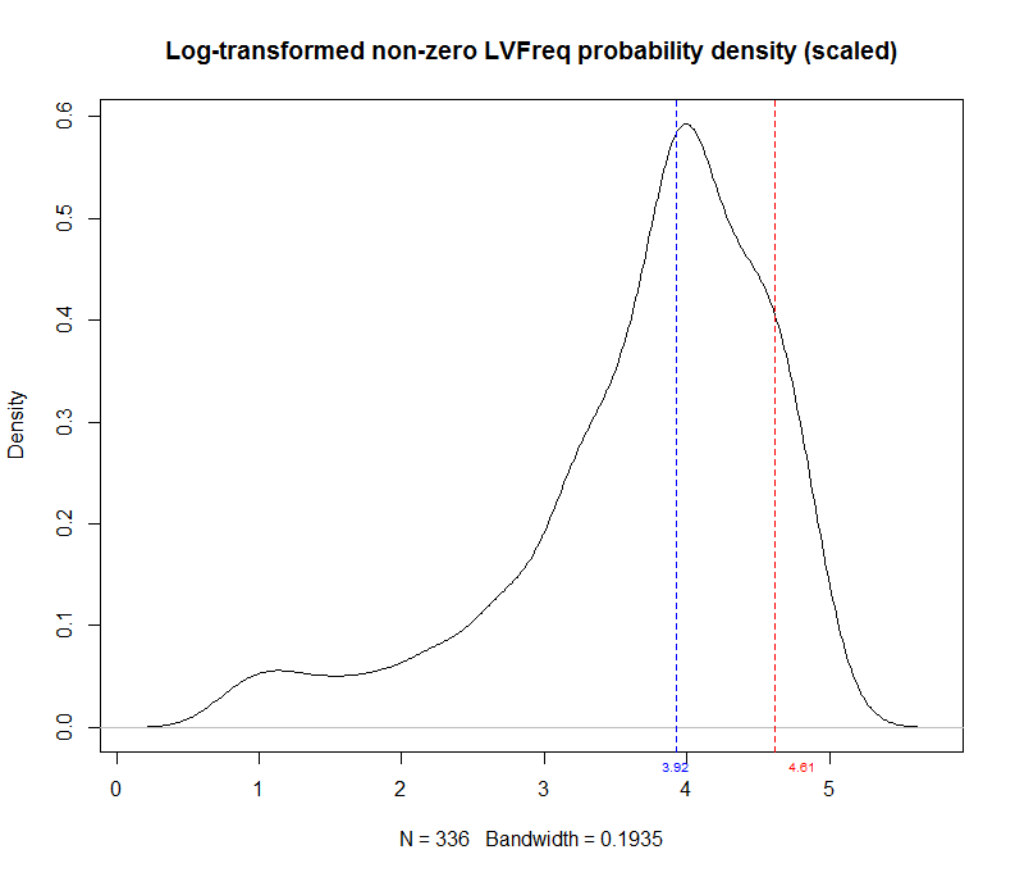
LVFreq estimation

LVFreq = **0%** no LV

LVFreq = **100%** **"reference LVFreq"** - LV are "normal" phonemes, i.e. the observed number of occurrences of LV is the same as would be expected under the H_0

LV FREQUENCY ESTIMATION





median reference LVFreq

- LV are relatively **rare phonemes** in most languages that have them, which is in accordance with their typological rarity

SKEWED LEXICAL DISTRIBUTION: SEMANTICS

Are the distributions of LV within the lexicons random?

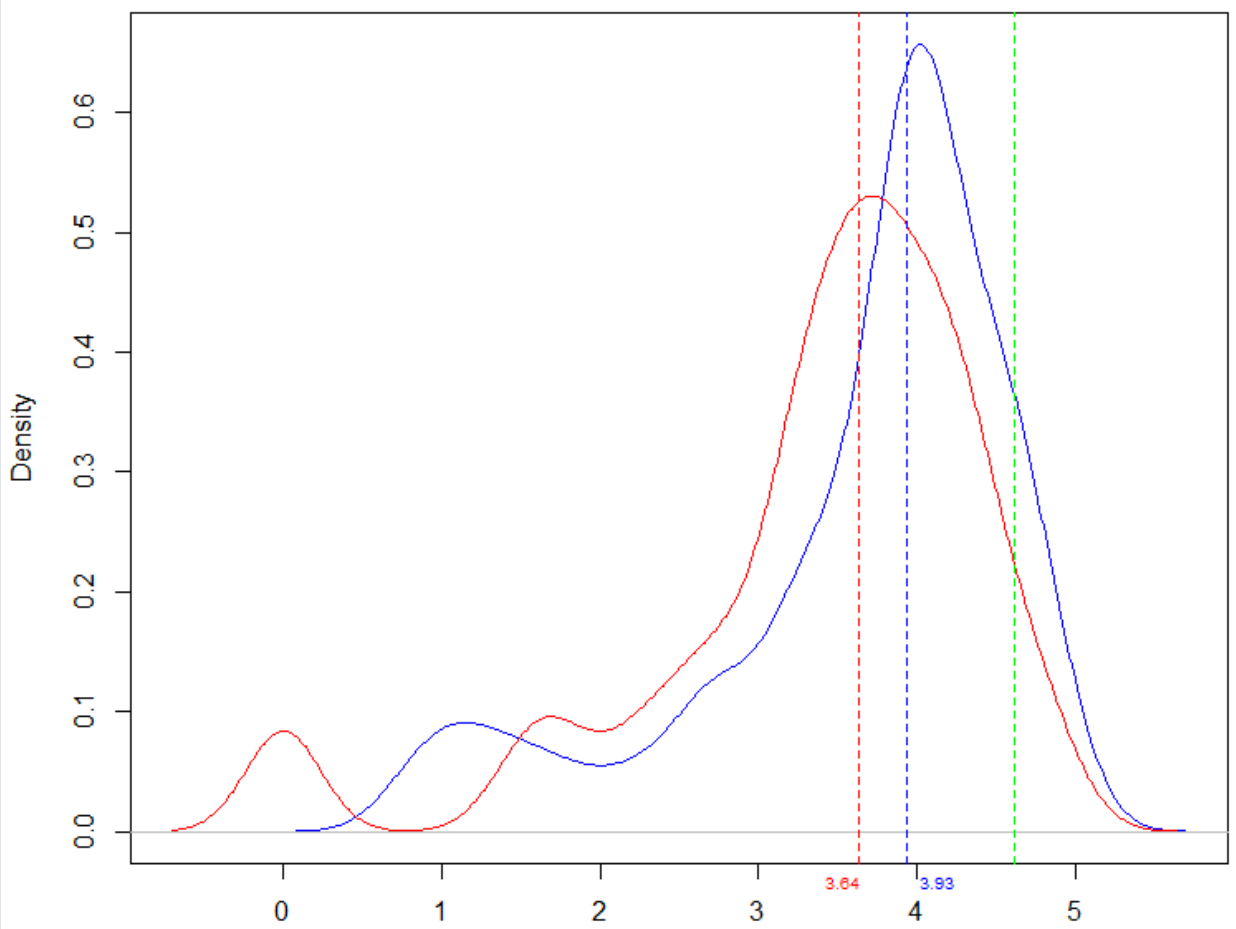
- **{H}:** LV are more common in **"expressive" parts of the lexicon**, such as ideophones or property words
- $\{H_{proxy}\}$: LV tend to be less common in "**basic vocabulary**"

SKEWED LEXICAL DISTRIBUTION: SEMANTICS

Are the distributions of LV within the lexicons random?

- A possible test: Extract a subset of entries of a "basic vocabulary" from each source of a sufficient size and compare the LVFreq pattern in the original sample with the LVFreq pattern in a "basic vocabulary" sample
- Our version of the test:
  - automatically created Swadesh-200 lists
  - the sources with $\geq$ 400 entries
  - fill the gaps with random entries
  - the result is a quasi-Swadesh-200 list

SKEWED LEXICAL DISTRIBUTION: SEMANTICS



paired U-test (Wilcoxon signed rank test):

p-value = 5.061e-13

Bootstrap (rep = 999):

100% p-values < 0.5

50% p-values ≤ p_0

original LVFreq (≥ 400 entries)

quasi-Swadesh-200 LVFreq

original median (≥ 400 entries)

quasi-Swadesh-200 median

reference LVFreq (= 100%)

SKEWED LEXICAL DISTRIBUTION: PHONOTACTICS

Are the distributions of LV within the lexicons random?

- LV tend to be less common in "basic vocabulary"
- {H}: LV are more common in "expressive" parts of the lexicon, such as ideophones or property words

- LV are largely restricted to stem-initial position (which is often also word-initial)

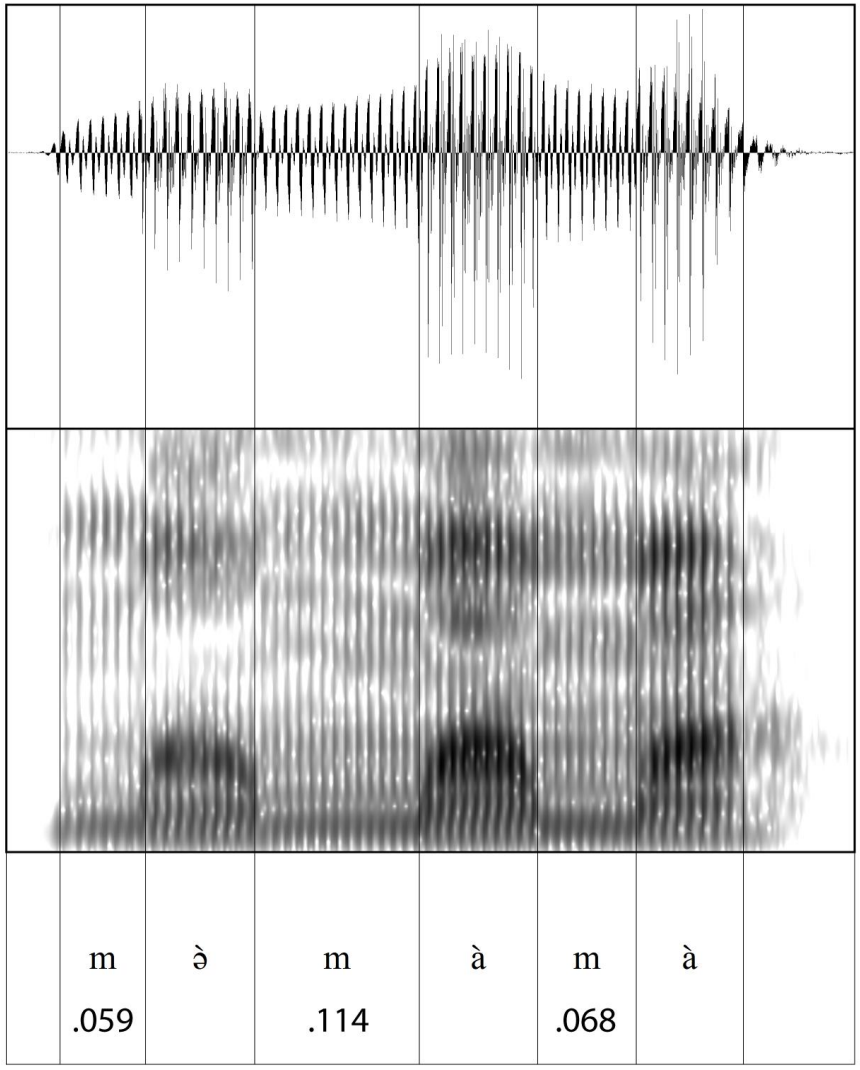
LV, SI C-ACCENT & C-EMPHASIS PROSODY

- The **two types of skewing** in the distribution of LV, semantic [LV ~ "expressive" vocabulary] and phonotactic [LV ~ stem-initial position] are **indirectly related**

- Diachronically, they are linked through the feature of **SI C-accent**, which itself is a manifestation of a more general phenomenon of **C-emphasis prosody**

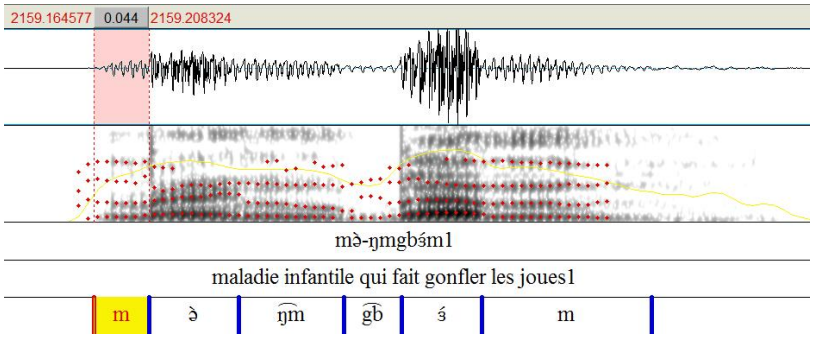
- C-emphasis prosody is a very important **factor behind the emergence of LV** in NSSA

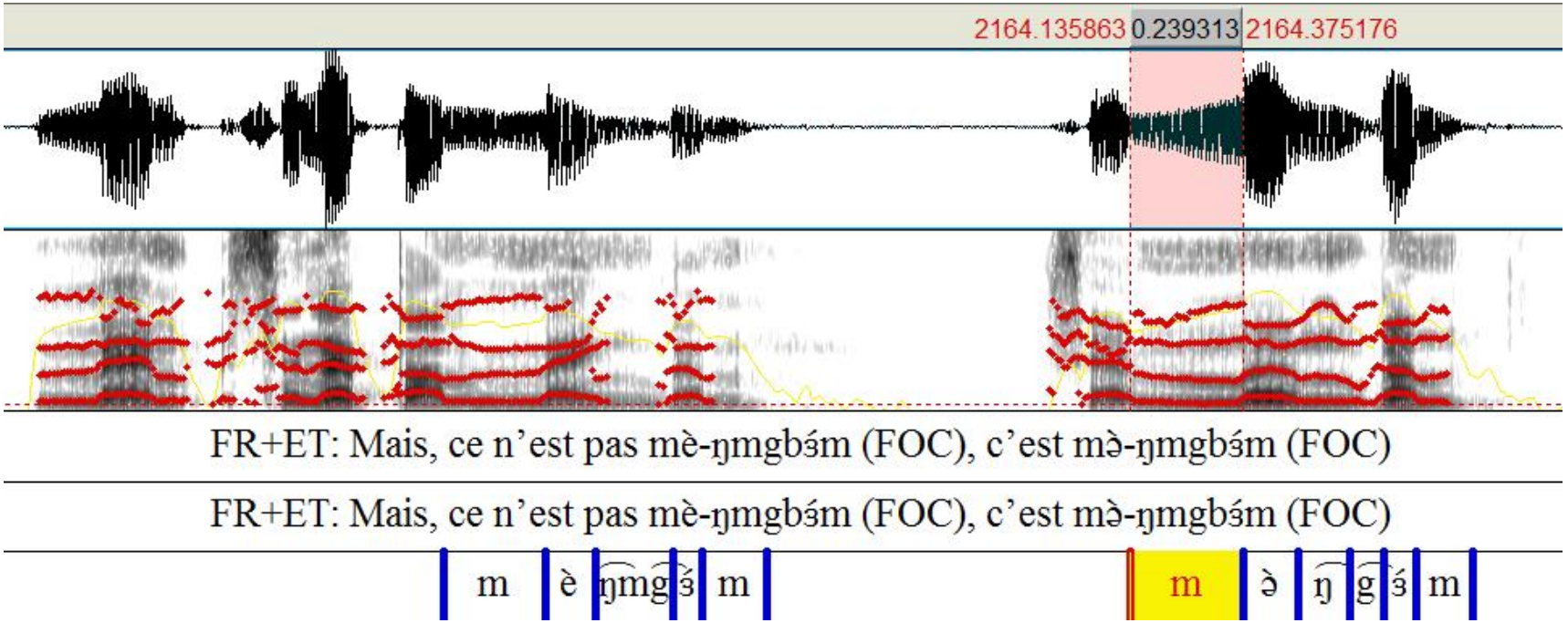
Consonant length in the nonsense word *mə-màmà* (Eton, Bantu A70)



- Corrective focus on the prefix V realized with prefix C-emphasis

Eton (A70)





EMERGENCE OF LV & SI C-ACCENT

- SI accent > round vowel diphthongization > labialization of SI C
- Accented C may also attract long-range features, such as labialization, from elsewhere in the word
- SI C-accent as longer closer duration favor the emergence & sustenance of LV:
  - by **favoring** articulatory **overshoot**: labialization > labial occlusion
  - by **inhibiting** the loss of the velar gesture
- SI C-accent is intimately linked to the "expressive" function:
  - In origin, SI C-accent is a **prosodic phenomenon**, viz. C-emphasis prosody: emphasis by exaggerating the closure duration of a C
  - **"expressive" words** are more often emphasized prosodically

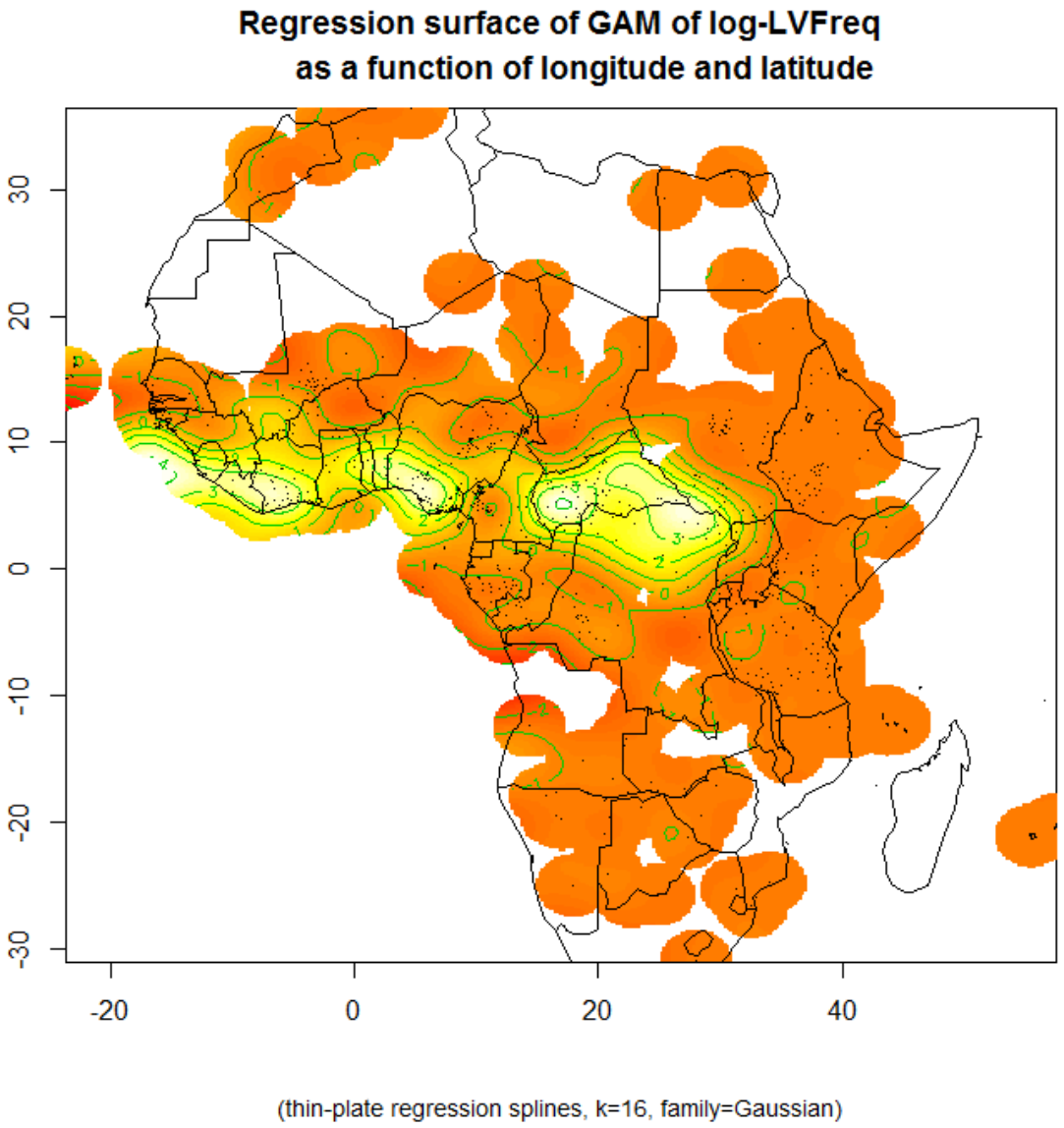
- The “expressive” function & the C-emphasis prosody as important vehicles of spread of LV through language contact (see Matras 2009, 2014… on borrowability)

> Functions that serve to negotiate attitudes among the participants in the interaction and which convey evaluations, assessments, the processing of presuppositions, or emotions, are particularly prone to borrowing: This includes information structuring at the level of the discourse and clause, [...], prosody in phonetics and phonology, discourse particles [...] They represent bilingual speakers’ need to align the emotional and presupposition-oriented side of negotiating communicative interaction across interaction settings.
>
> (Matras 2014:5)

LV, SIC-ACCENT & C-EMPHASIS PROSODY

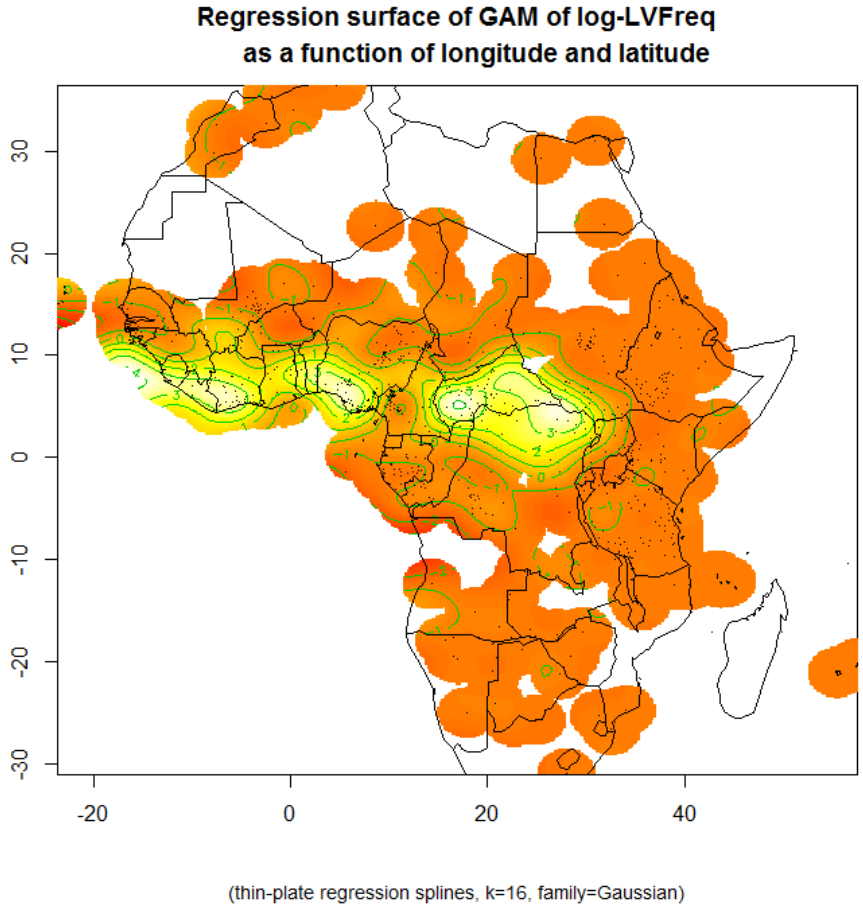
- In a broader perspective, **C-emphasis prosody** is a very good candidate for the role of a **major driving force** behind the emergence of several other types of sounds, such as labial flaps, bilabial trills, and clicks

SPATIAL DISTRIBUTION



- 2 clearly separated clusters
  - Coastal West Africa (composed of 2 sub-clusters: Upper Guinea and Lower Guinea Hotbeds)
  - Central Africa (Ubangi Basin Hotbed)
- possibly, +1 less prominent cluster
  - SE Mali & SW Burkina-Faso
- 1 major spatial discontinuity
  - NE Nigeria & Cameroon
- 1 minor spatial discontinuity
  - Ghana

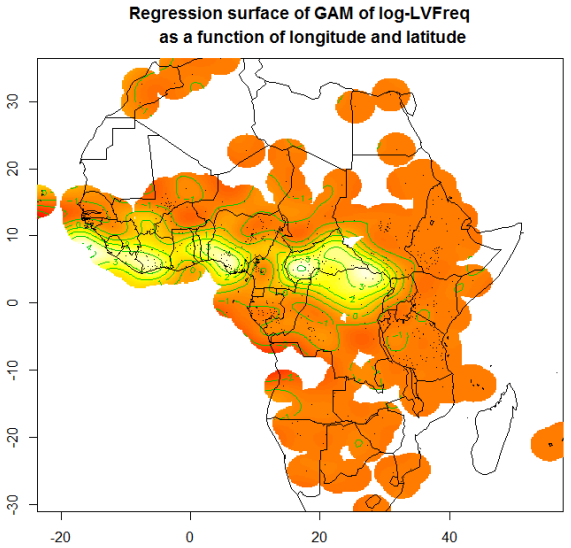
HISTORICAL IMPLICATIONS: HOTBEDS OF RETENTION

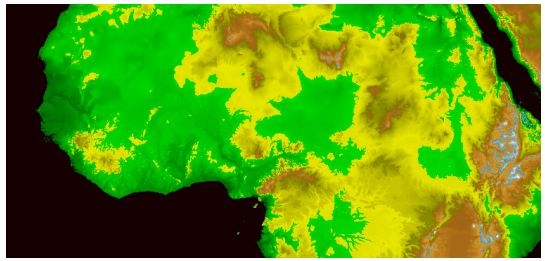


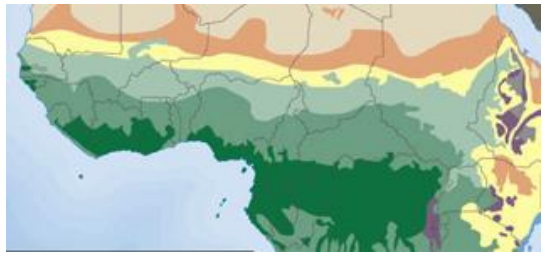
- Logically, the 3 hotbeds of high LVFreq reflect the **retention** of LV (and ultimately of C-emphasis prosody and SI C-accent), present in the original population:

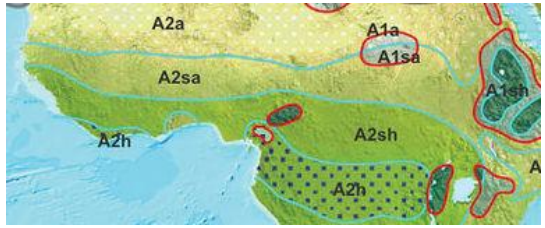
  - Typologically, LV are rare
  - Several hotbeds of independent emergence of high LVFreq are unlikely

HISTORICAL IMPLICATIONS: REFUGE ZONES



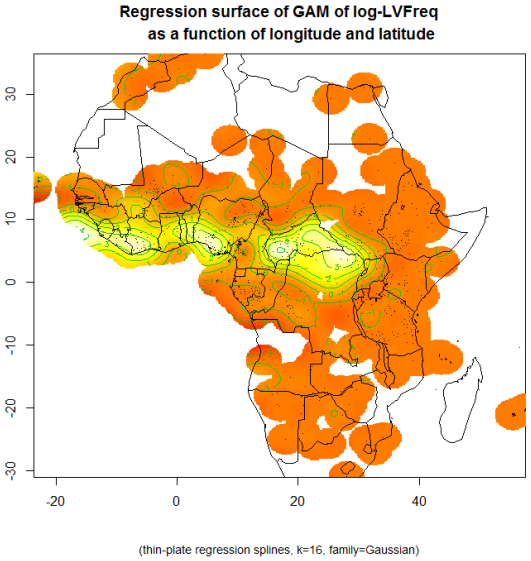






- Geographically, the 3 hotbeds of high LVFreq are typical **refuge zones**: mostly tropical forests delimited by natural boundaries (sea, savanna, mountain ranges)
- Ghana discontinuity ≈ Dahomey forest gap
- NE Nigeria & Cameroon discontinuity ≈ Adamawa Plateau, Cameroon mountains

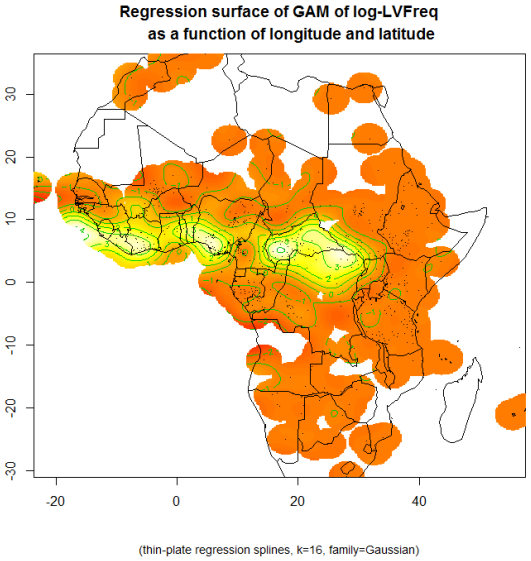
HISTORICAL IMPLICATIONS: SHIFT VS CONTACT



- Savanna populations speaking languages without LV migrated southwards
- They encountered Primary LV Populations
- Migration was easiest and fastest in the savanna areas
- The marginal LVFreq outside of the hotbeds is most likely due to borrowing in the first place

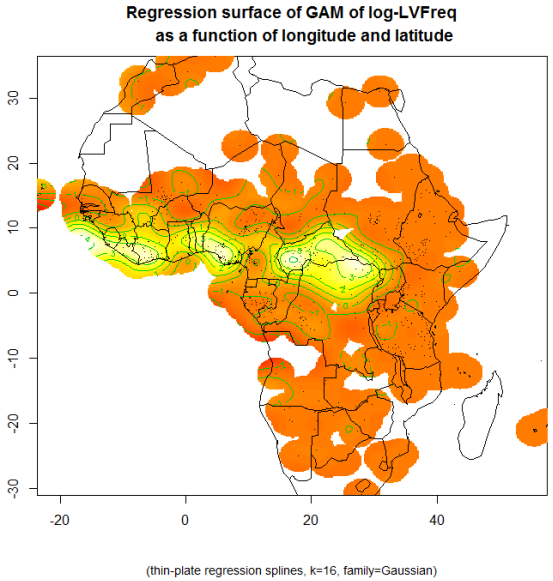
- The migration was much slower in the tropical forest refuge zones
- The high LVFreq in the hotbeds is most likely the result of a language shift of Primary LV Populations to the languages of the newcomers
- shift-induced substrate interference (Thomason & Kaufman 1988)

HISTORICAL IMPLICATIONS: WHO IS WHO?



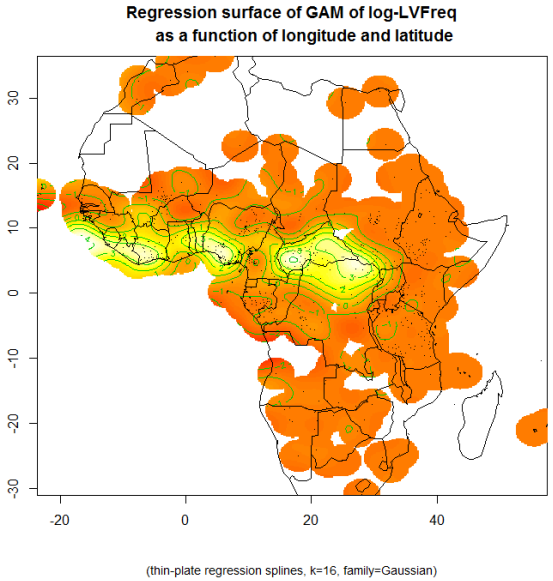
- The language family(s) of Primary LV populations are likely to have disappeared today
- The overwhelming majority of the language groups in the hotbeds are either "robust" or "promising" Niger-Congo members (cf. Güldemann 2018)
- Members of the Central Sudanic family on the eastern fringes of the Ubangi Basin Hotbed
- Numerous NC groups and some CS groups are also found outside of the hotbeds and often lack LV
- The savanna populations without LV that migrated south were by and large speakers of Niger-Congo languages

HISTORICAL IMPLICATIONS: RECONSTRUCTION OF LV



- LV should not be reconstructed in Proto Niger-Congo or its major branches:
  - NC expanded into the hotbeds from the north
  - The further contemporary NC languages are from the hotbeds, the less likely they are to have LV stops
- The same applies to Central Sudanic

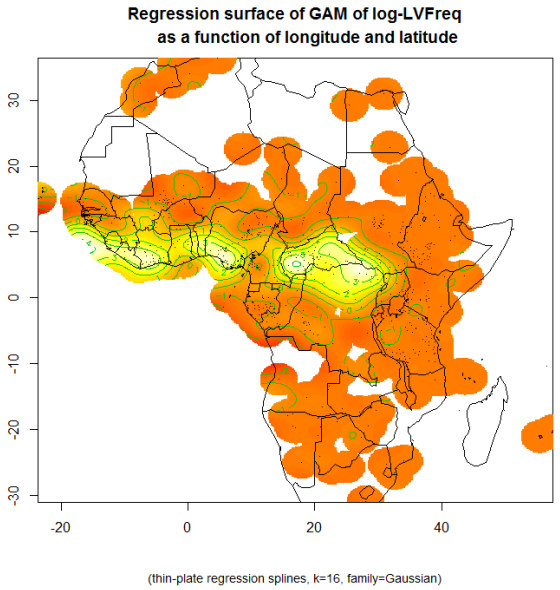
HISTORICAL IMPLICATIONS: RECONSTRUCTION OF LV



- Comparative evidence confirms that LV should not be reconstructed in Proto Niger-Congo or its major branches:
  - Typical cognacy: LV || velar (labialized or followed by a rounded vowel)
  - Reconstruction of a LV would imply lenition and loss of the labial release
  - Highly unlikely for perceptual reasons (the labial release follows the velar release)
  - Highly unlikely for phonotactic reasons (SI position of prosodic prominence

Supplementary materials

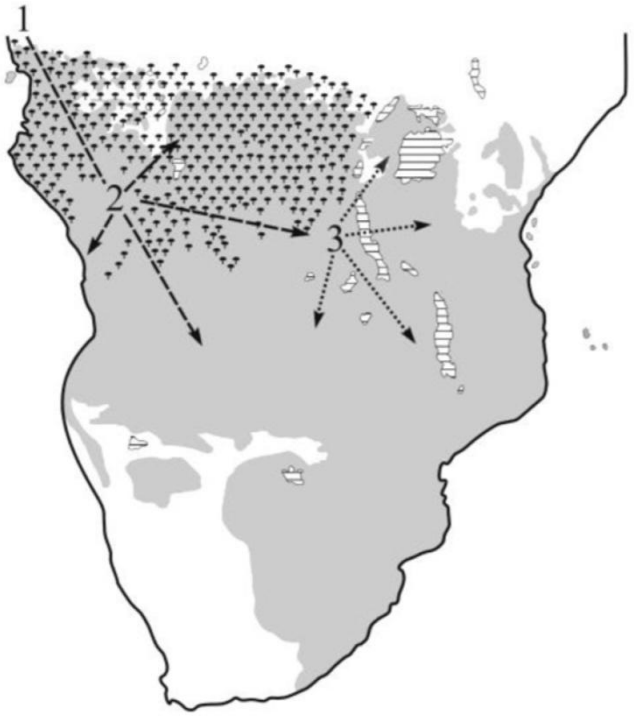
HISTORICAL IMPLICATIONS: BANTOID



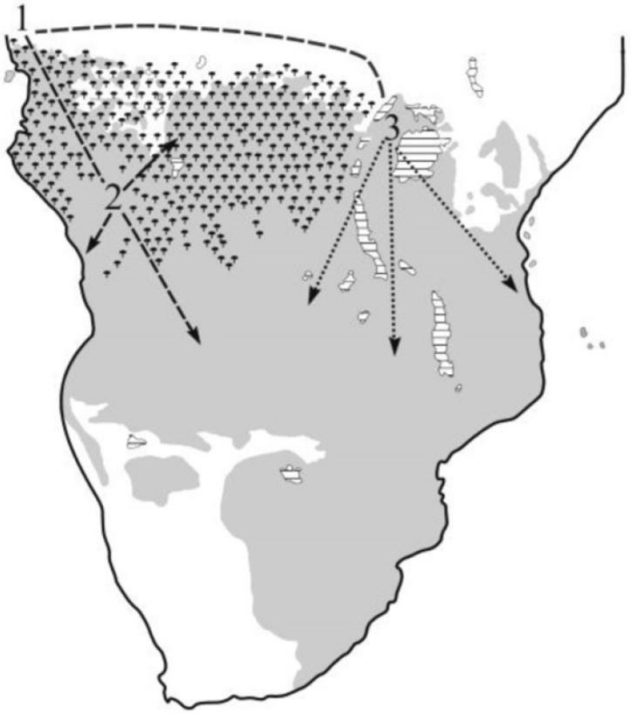
- The majority of Bantoid languages are spoken outside of the hotbeds of high LVFreq

HISTORICAL IMPLICATIONS: BANTU EXPANSION

Two principal models of Bantu expansion



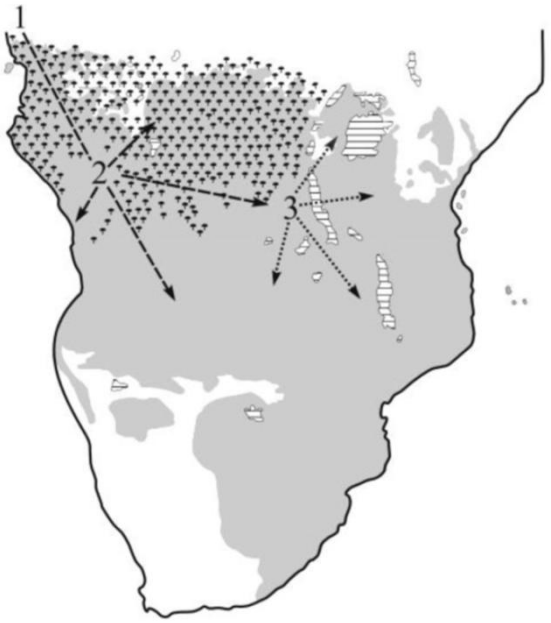
a. East out of West

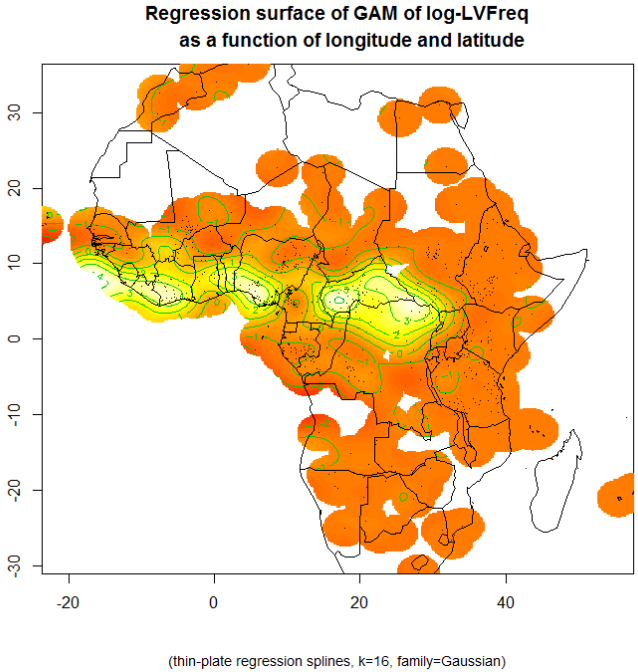


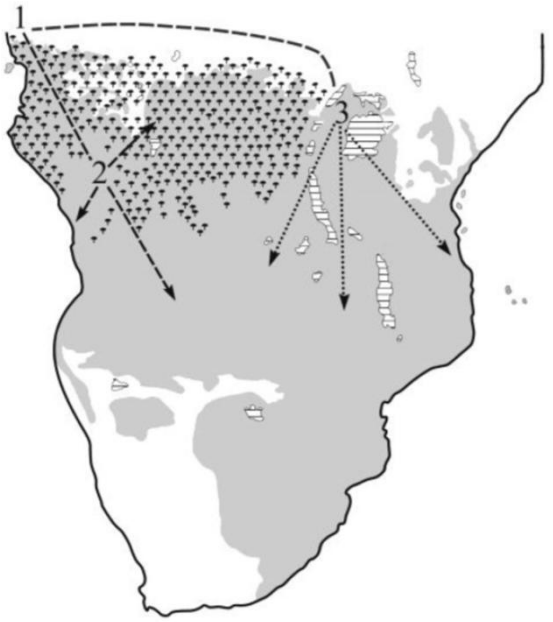
b. East separate from West

(adapted from Pakendorf et al. 2011 :8).

HISTORICAL IMPLICATIONS: BANTU EXPANSION





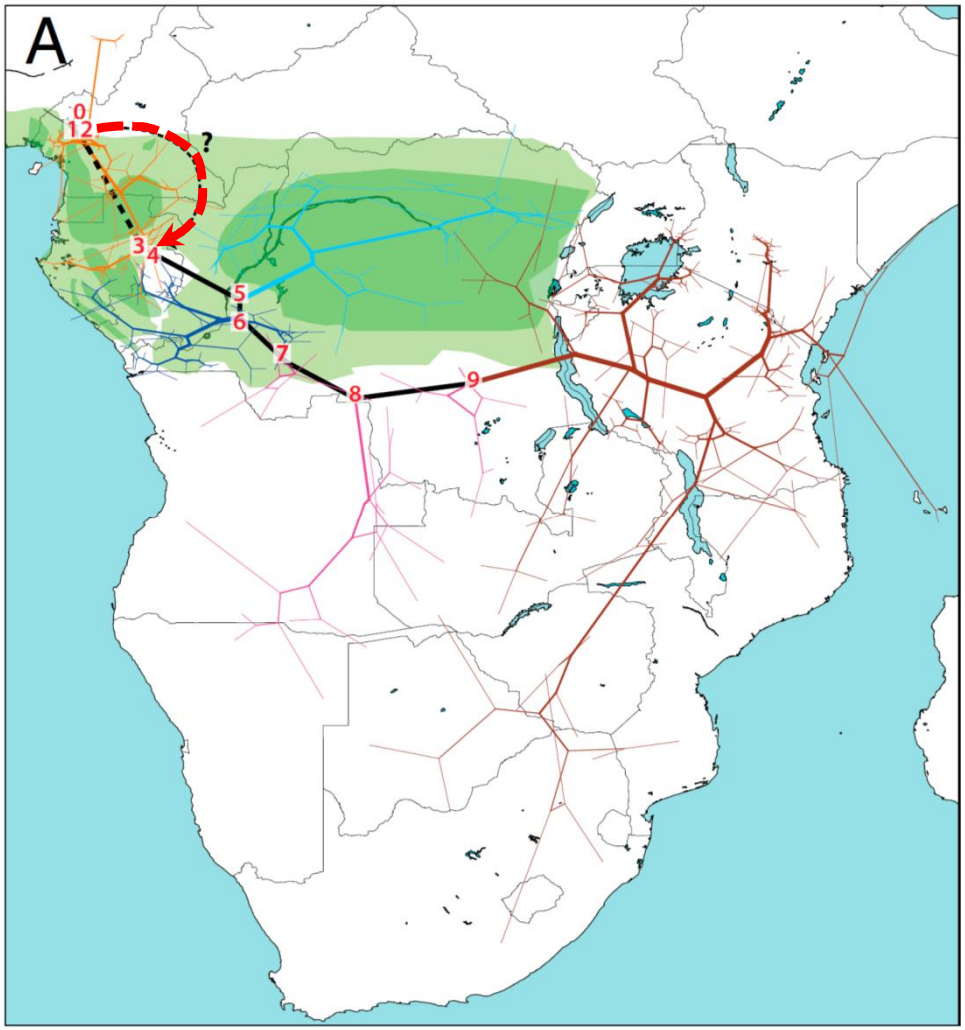


a. East out of West

b. East separate from West

- Our model supports the "East-out-of-West" hypothesis of the E Bantu emergence with the E Bantu break-off point somewhere south of the rainforest

HISTORICAL IMPLICATIONS: BANTU EXPANSION



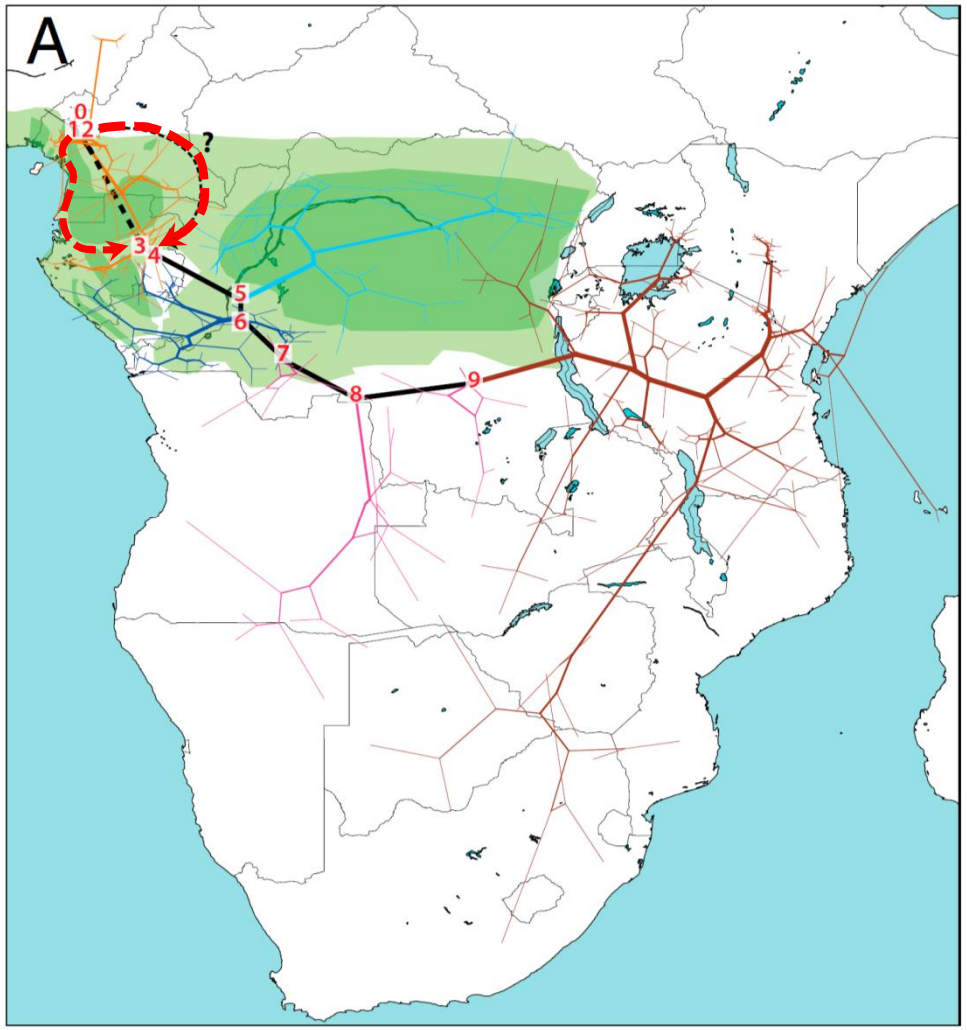
Bantu migration route reconstructed by Grollemund et al. (2015) on consensus tree by using geographical locations of contemporary languages and connecting ancestral locations by straight lines (true route will differ).

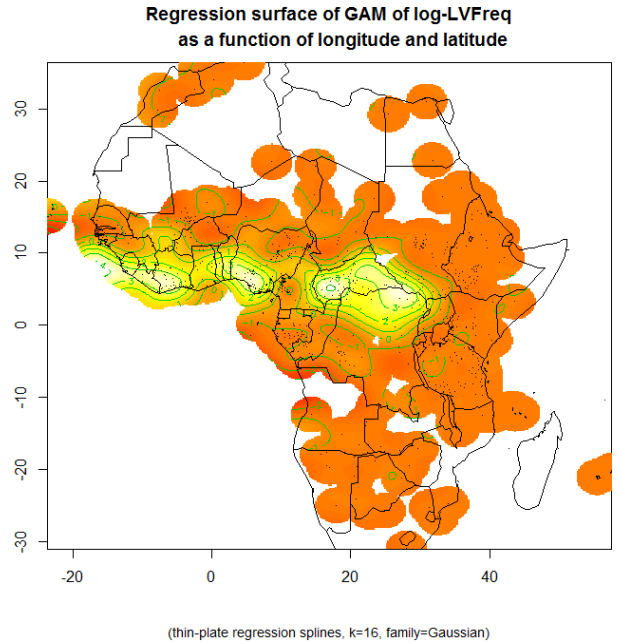
Numbered positions correspond to major diversification nodes on the consensus tree.

Curved dashed line indicates suggested migration route through savannah corridors (**Sangha River Interval**)

Lighter green shading corresponds to the delimitation of the rainforest at 5.000 B.P.; the darker green corresponds to the delimitation of the rainforest at 2.500 B.P.

HISTORICAL IMPLICATIONS: BANTU EXPANSION





- Our model suggests that the migration between nodes 2 and 3 is more likely to have happened through a **coastal route** rather than the Sangha River Interval.