

A New Phylogenetic Classification for the Bantu Languages of the Northeastern Democratic Republic of the Congo

Birgit Ricquier: ULB, birgit.ricquier@ulb.be

Rebecca Grollemund: University of Missouri, grollemundr@missouri.edu

David Kopa wa Kopa: ULB – University of Kisangani, kopadaza@gmail.com

Constance Kutsch Lojenga: SIL International – University of Leiden,
c_kutsch_lojenga@sil.org

François Abuka Balabala Alumesa: University of Kisangani, francoisabuka@gmail.com

Nicolas Mombaya Liwila, Université de Kisangani, nicolasmombaya@gmail.com

Emmanuel Ngbanga Bandombele, Université de Kisangani, emmanuelngbanga@yahoo.fr

~~~~B~~~  
AN~~~~  
~~~~TU~~~  
RIV~~~~
~~~~ER~~~  
~~~~S~~~~



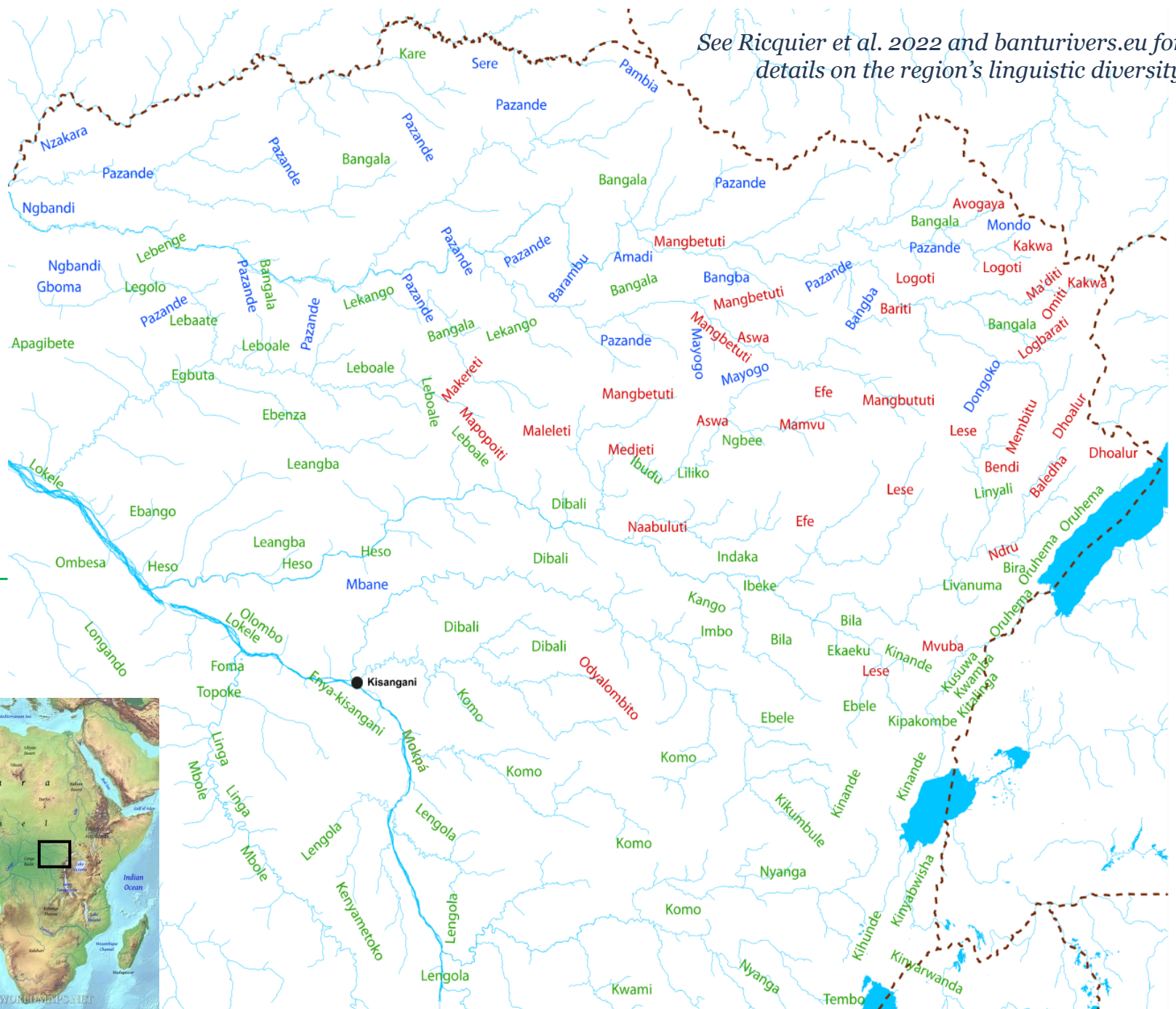
The Bantu Languages of the Northeastern Congo Basin and Their History



~~~~~B~~~~~  
 AN~~~~~  
 ~~~~~TU~~~~~  
 RIVER~~~~~
 ~~~~~ER~~~~~  
 ~~~~~S~~~~~

See Ricquier et al. 2022 and banturivers.eu for details on the region's linguistic diversity

Language families:
 Ubangi
 Central-
 Sudanic
 Bantu: C40-
 C50, D



The BANTURIVERS project

= *At a Crossroads of Bantu Expansions: Present and Past Riverside Communities in the Congo Basin, from an Integrated Linguistic, Anthropological and Archaeological Perspective* (ERC Starting Grant, ULB, 2019-2023)

= multidisciplinary, including historical linguistics, anthropology, archaeology and archaeozoology

Objectives: to unravel the early history of the northeastern DRC with a special focus on the Bantu expansions and the role of the Congo Basin river network in the early history of the concerned speech communities

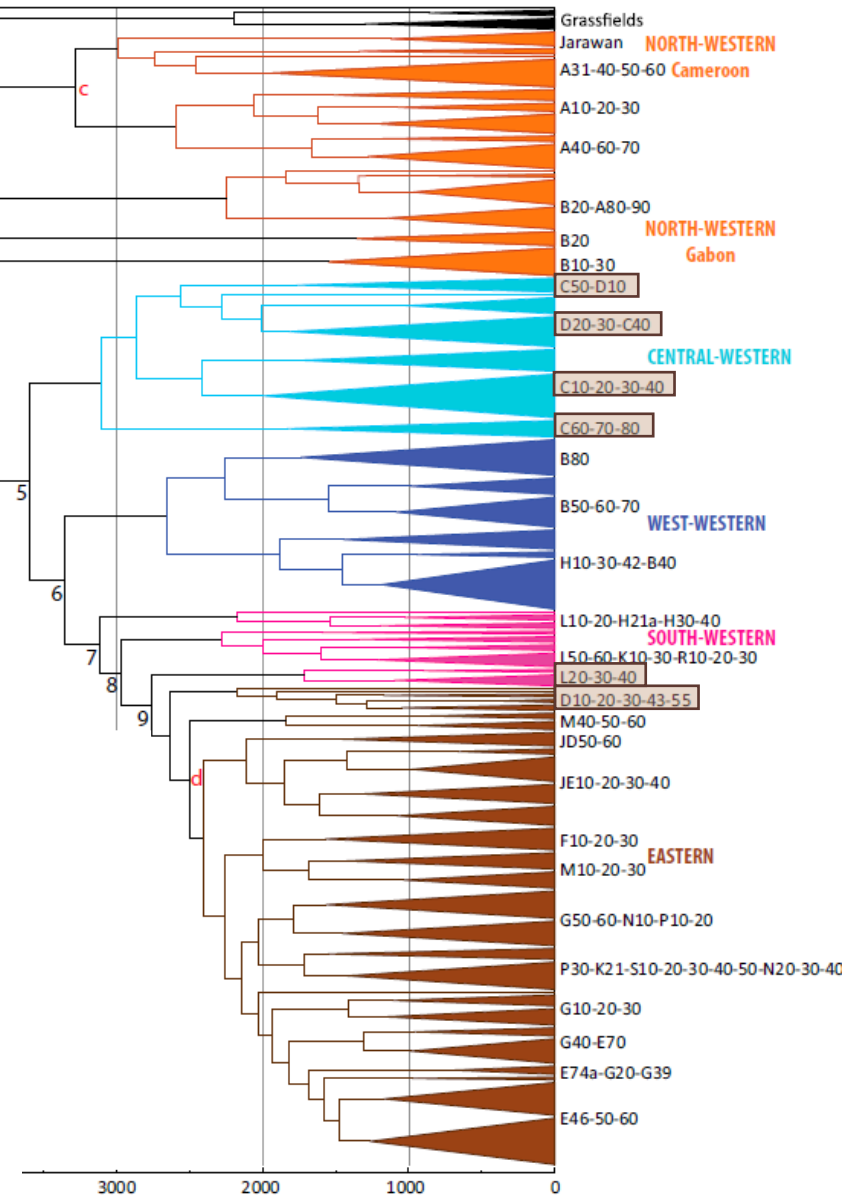
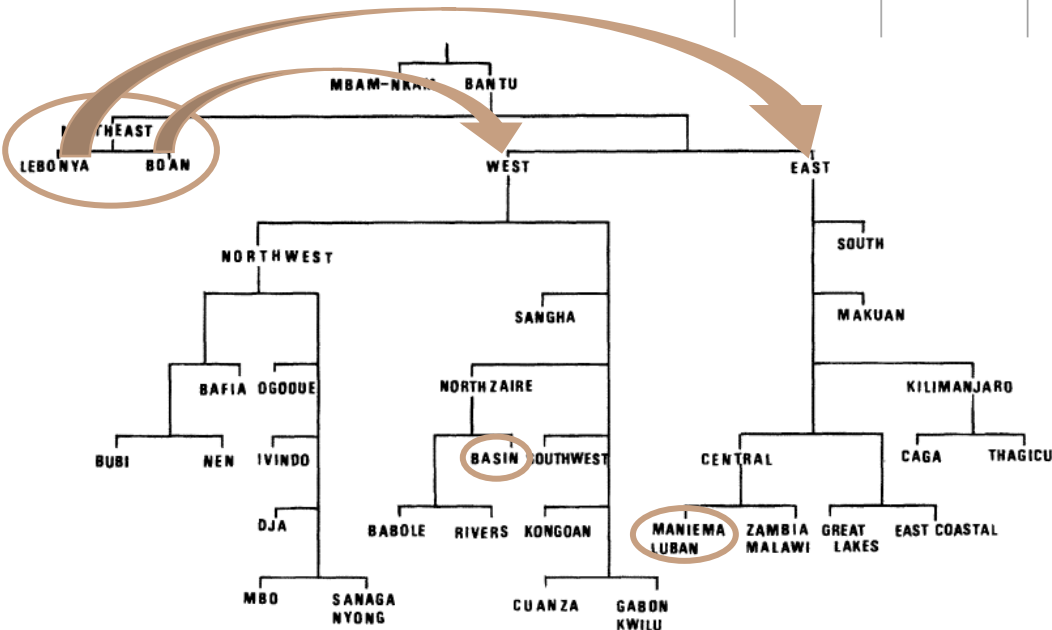
- ⇒ Language classification as one tool to discover the settlement history of Bantu speech communities
- ⇒ Language classification as a framework to interpret results from Words-and-Things research

~B~
 AN~
 ~TU~
 RIV~
 ~ER~
 ~S~

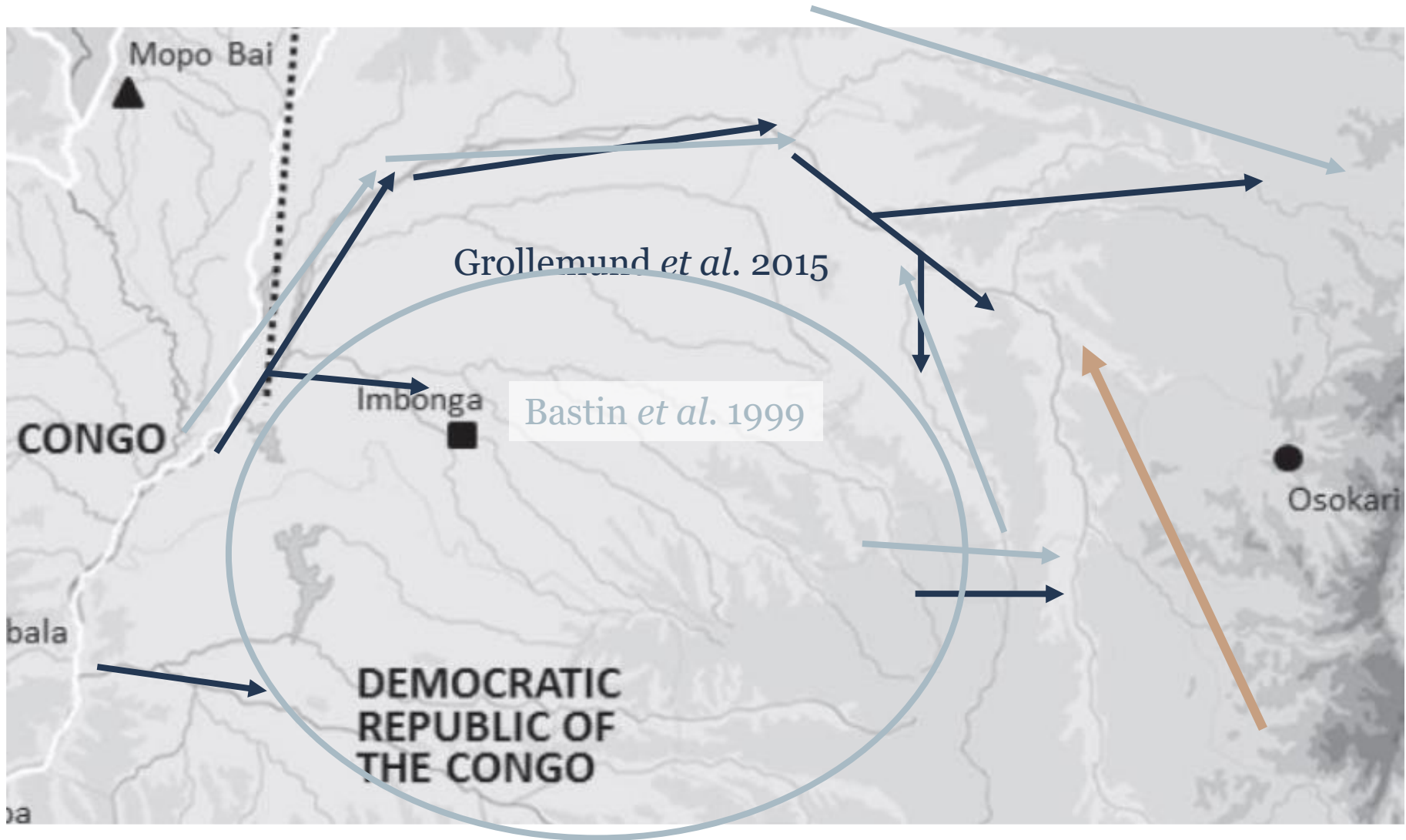
Conflicting Classifications

Grollemund et al. 2015 ↓

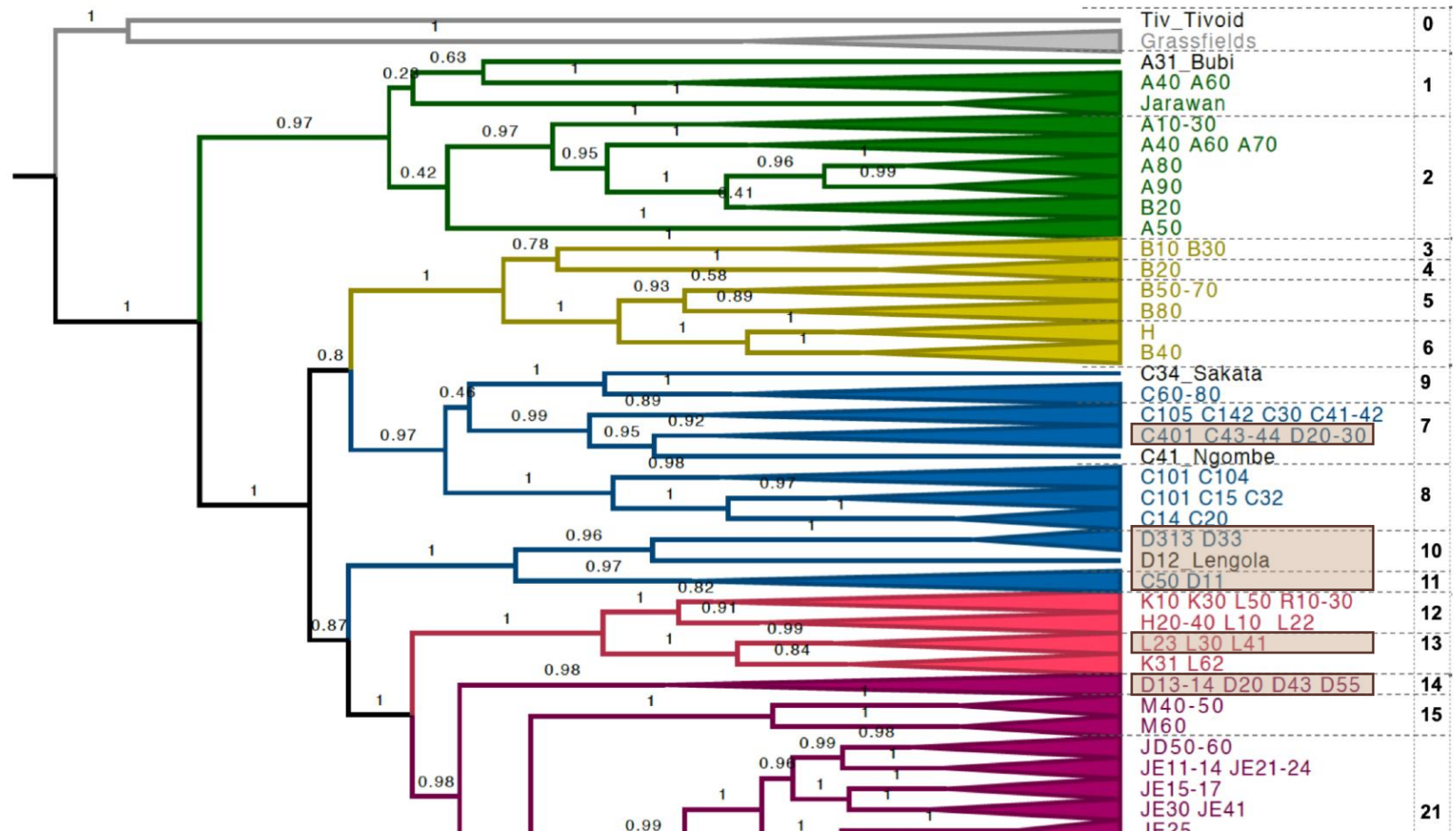
Bastin et al. 1999 (graph in Vansina 1995) ↓



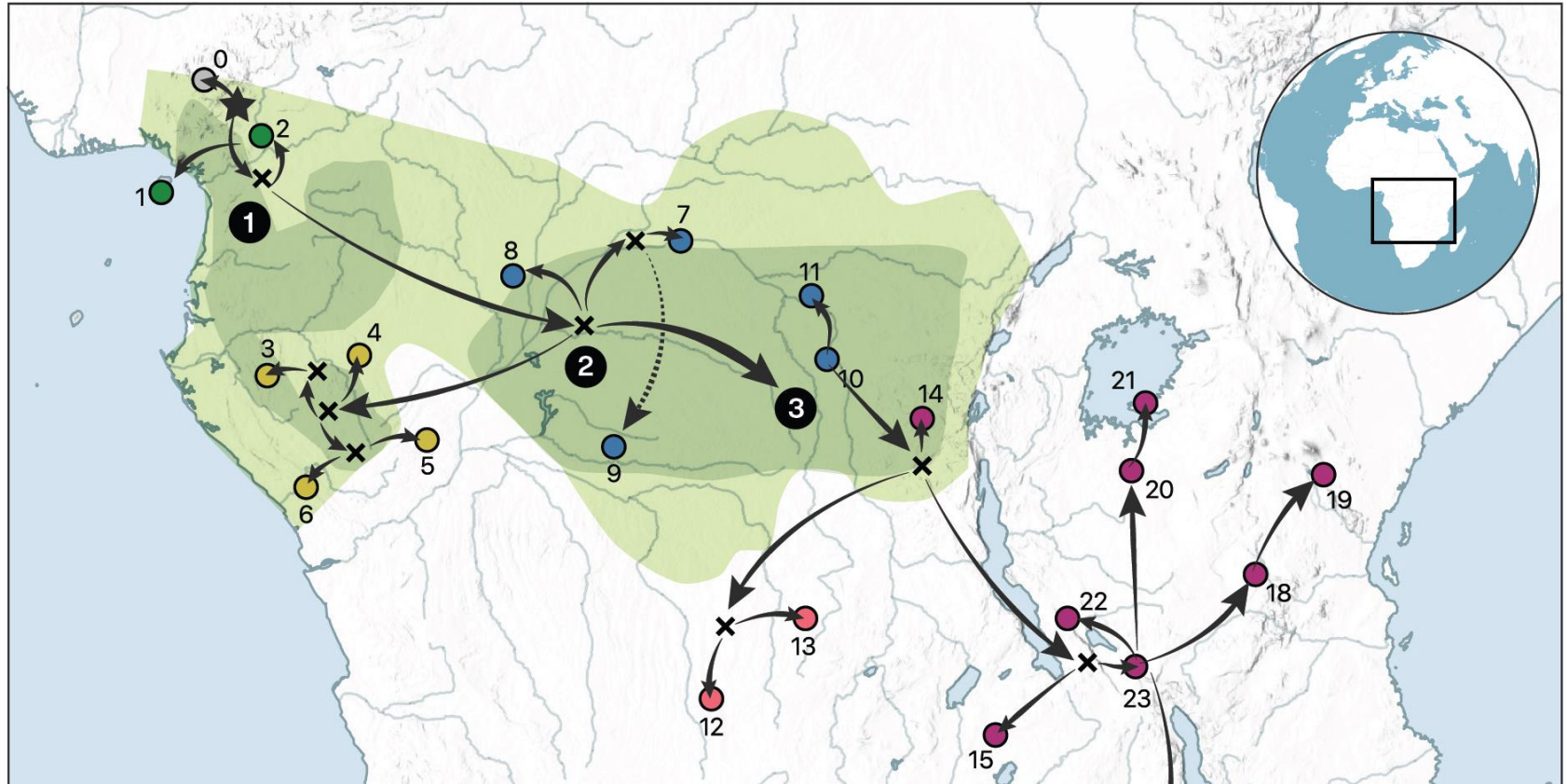
Paths through the Rainforest



Koile et al. 2022



Koile et al. 2022



Data

- field data recorded by François Abuka Balabala Alumesa, David Kopa wa Kopa, Nicolas Mombaya Liwila & Emmanuel Ngbanga Bandombele (2019-2022) for **69 doculects ~ 49 language varieties**
 - field data from SIL, shared by Constance Kutsch Lojenga (especially for southern supposedly “zone D” languages)
 - field data for Lingbe, shared by Didier Demolin
 - data from the Boyekoli Ebale Congo 2010 expedition (recorded by Birgit Ricquier and transcribed by Odette Ambouroue)
- ⇒ languages previously not included (Bokala, Mokpá, Bembe, Nonda, Lingbe)
- ⇒ concepts better covered
- Supplemented with data from previous classifications (Bastin et al. 1999 and Grollemund et al. 2015)
 - 100 concepts considered the basic vocabulary
 - coded into cognate sets

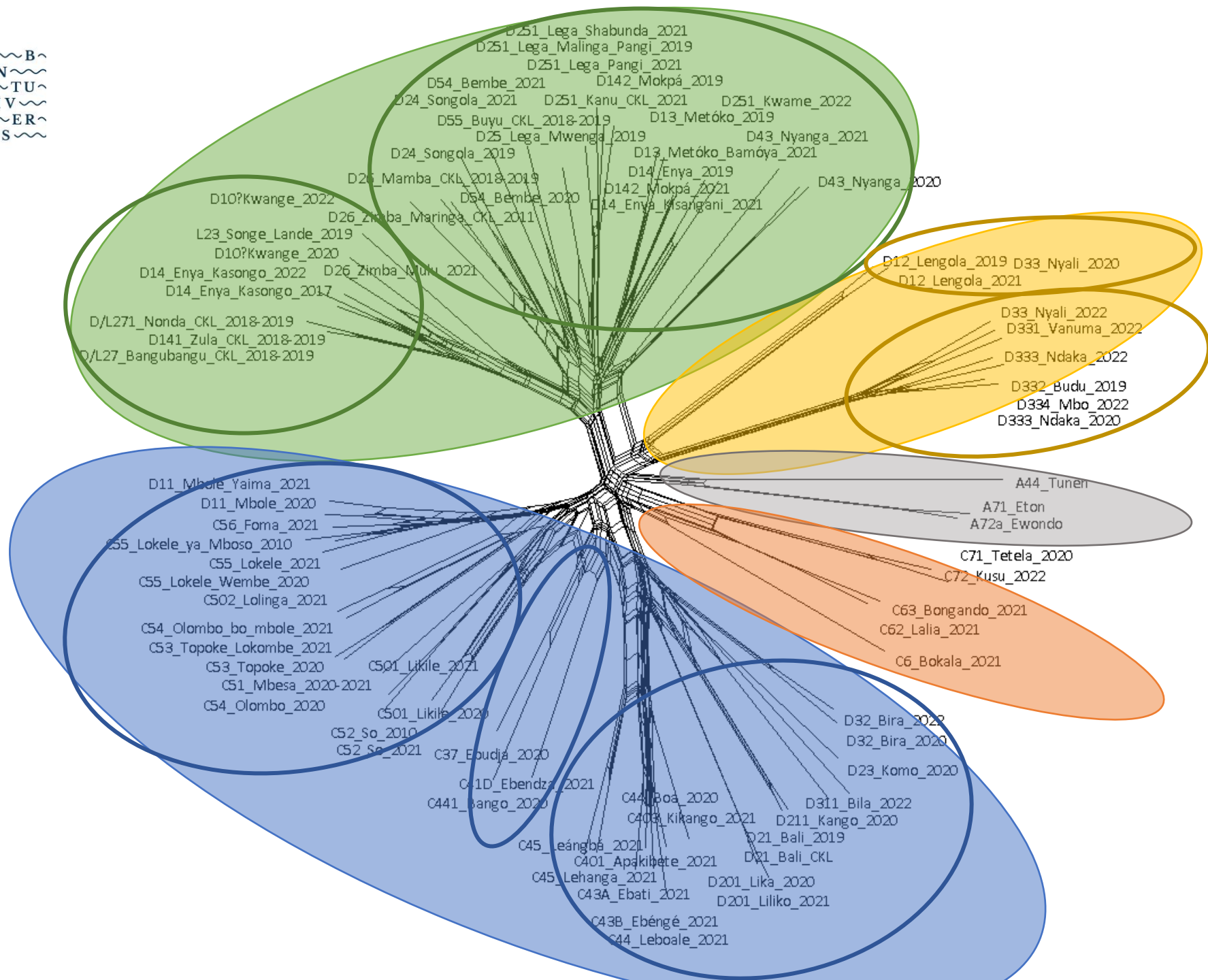
Methods

- **Databases: 2 databases were built**
 - a selection of 84 languages: only recent data + 3 Northwest Bantu
 - a selection of 163 languages: also data from previous classifications, including a representative set of other Bantu languages
- **Classifications** achieved by using two types of phylogenetics methods:
 1. **distance based method:** Neighbour-Net algorithm (with Splitstree software). Neighbour-Net method uses a multistate encoding where each cognate set receives a number. Neighbour-Net allows the reconstruction of phylogenetic networks.
 2. **Character-based method** using Bayesian inference of phylogeny (using a Markov chain Monte Carlo -MCMC) with a covarion model (CV2) that allows the rate of evolution to vary across the tree. Bayesian method uses a binary encoding where we mark the presence or the absence of a particular cognate set.

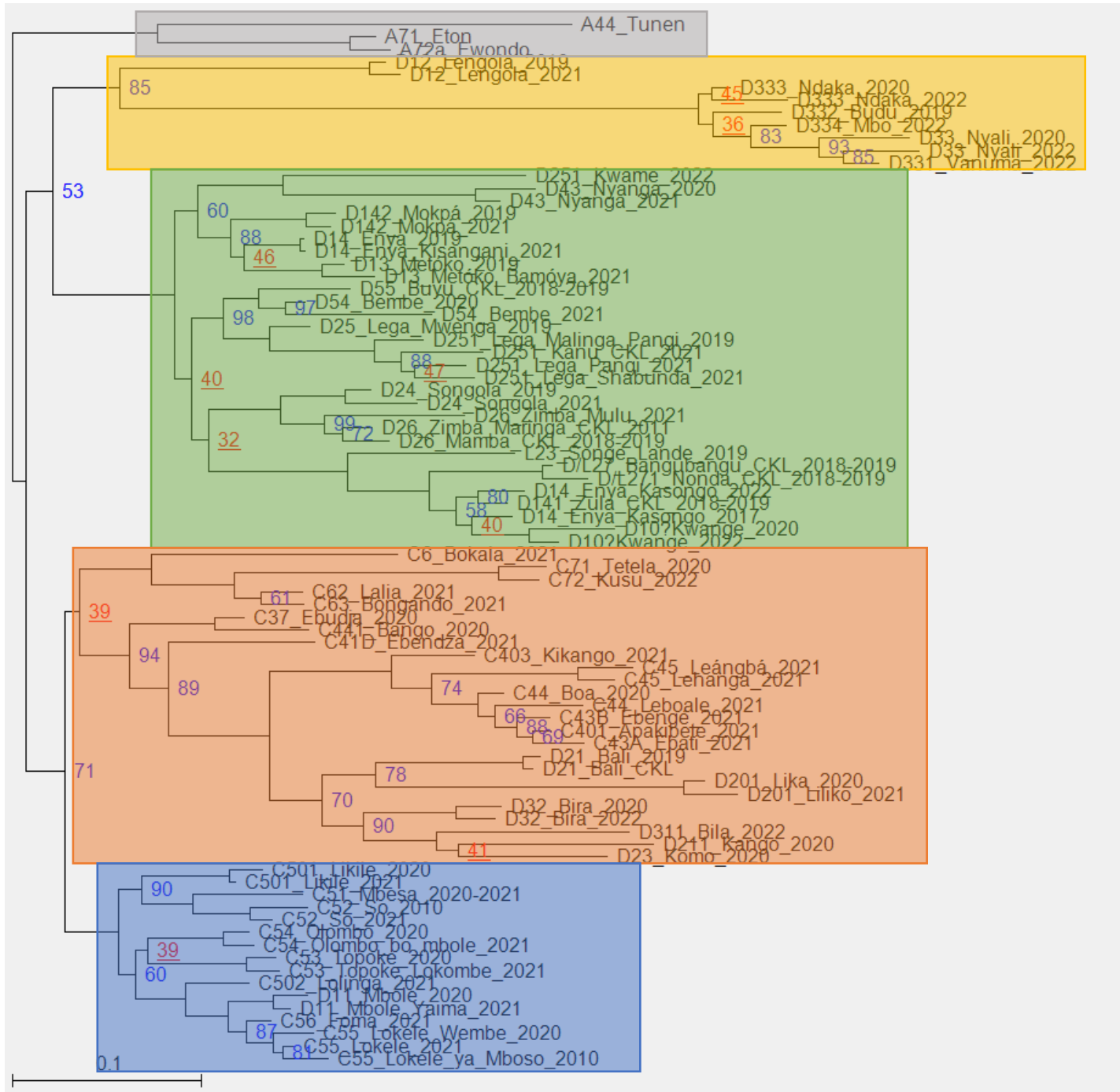
The Results

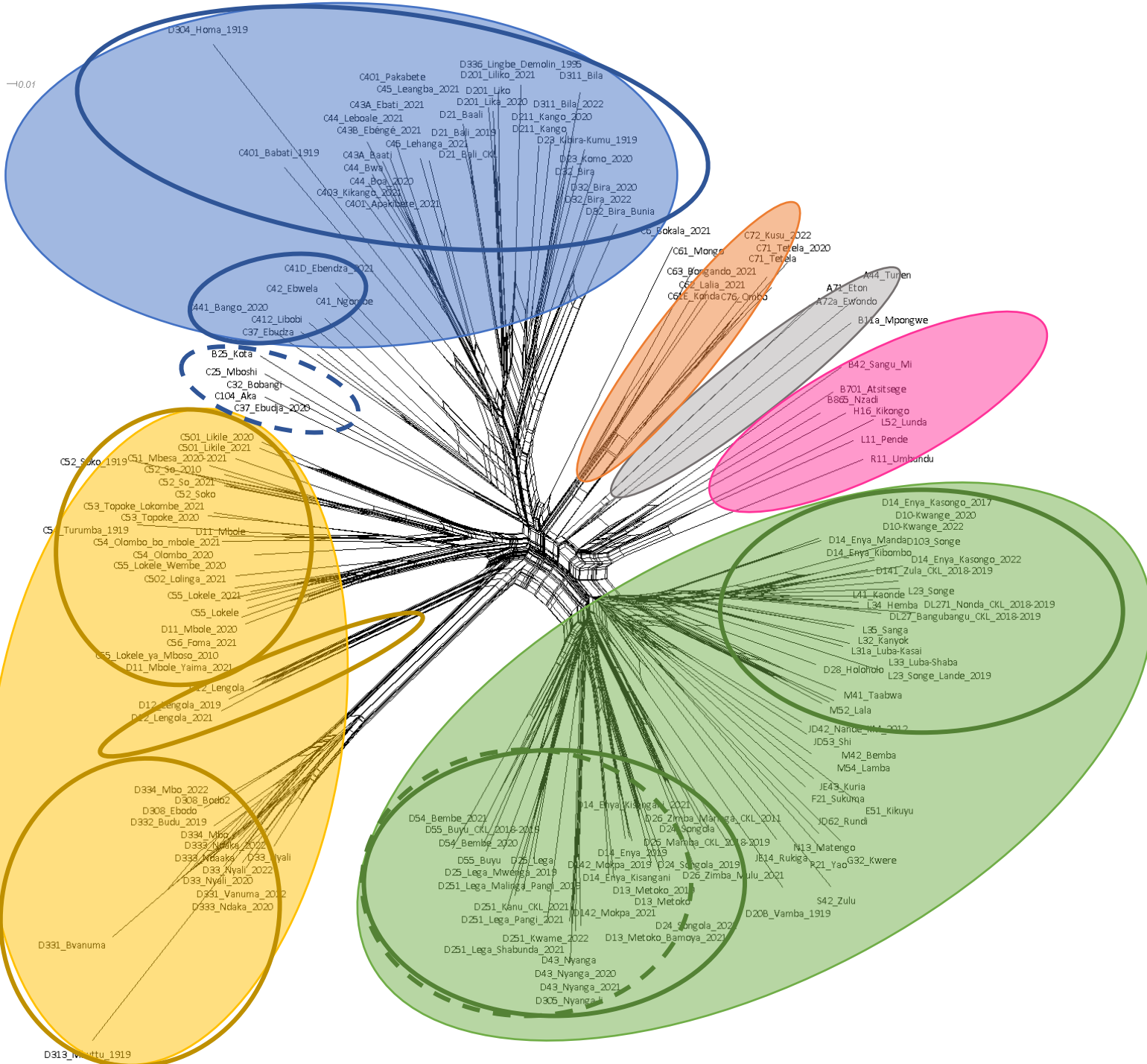


~~~~B~  
 AN~~~~  
 ~~~~TU~  
 RIVER~~~
 ~~~~S~~~~

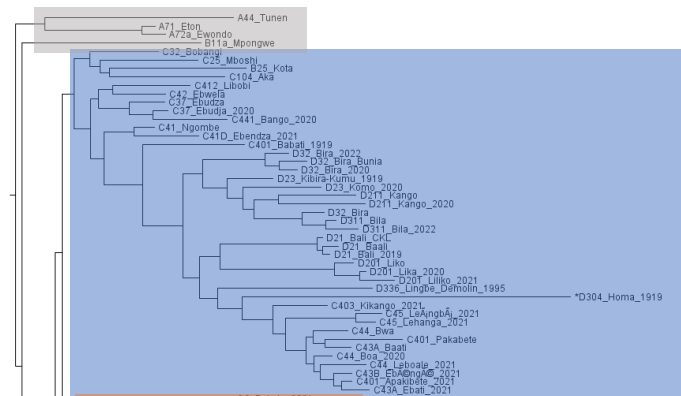


~~~~~B~~~~~  
 AN~~~~~
 ~~~~~TU~~~~~  
 RIV~~~~~  
 ~~~~~ER~~~~~  
 ~~~~~S~~~~~

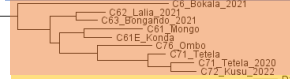




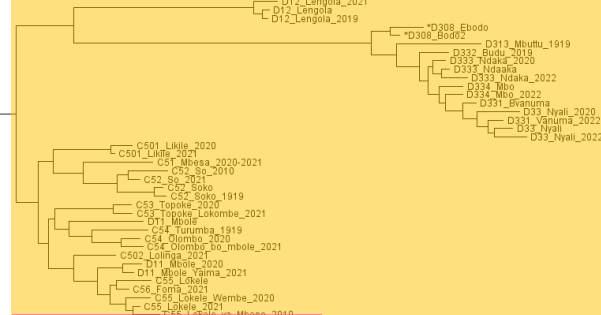
A phylogenetic tree showing the evolutionary relationships between 100 HIV-1 CRF01\_AG sequences. The tree is rooted on the left and branches out to the right. The sequences are color-coded by year: blue (1919-2022), orange (1919-2022), yellow (1919-2022), pink (1919-2022), and green (1919-2022). The tree is divided into five main clusters, each highlighted with a colored background. The clusters are: 1) Blue cluster (top), 2) Orange cluster, 3) Yellow cluster, 4) Pink cluster, and 5) Green cluster (bottom). The tree shows a clear temporal progression, with older sequences (1919) at the top and newer sequences (2022) at the bottom. Bootstrap values are indicated at the nodes. A scale bar of 0.1 is shown at the bottom left.



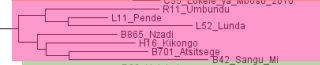
Congo Rivers = Rivers + Boa  
+ westernmost C



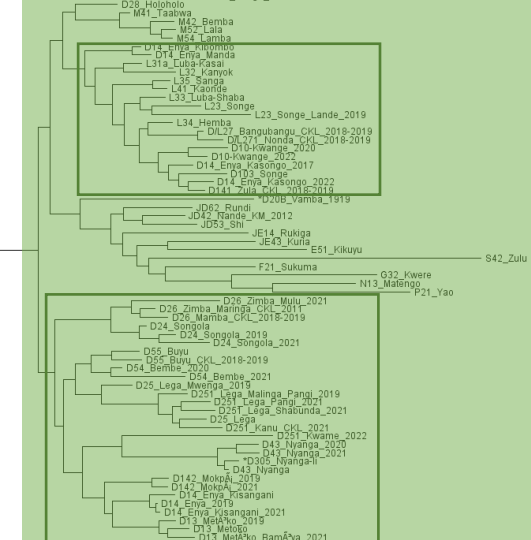
(Inner) Congo Basin



Upper Congo +  
Lengola + D33



West-Coastal + Southwest



East!  
Integrates Luba & Lega

# The Main Bantu Subgroups of East Congo



# Boan Bantu

Not in the periphery, but a member of the Congo Rivers group.

A two-fold split between West- and East Boan, but Bali+Liko form a subgroup that alternates, no doubt due to its central position.

Congo Basin Bantu:

> Inner Congo Basin

> Congo Rivers

...

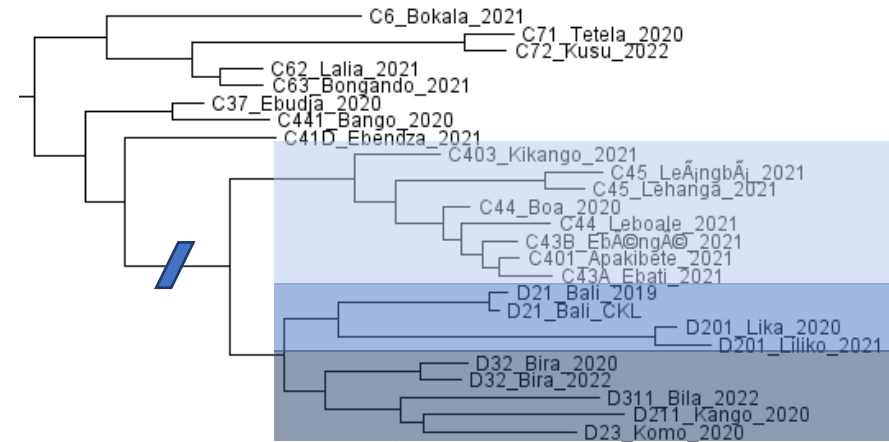
> Boan

> West-Boan

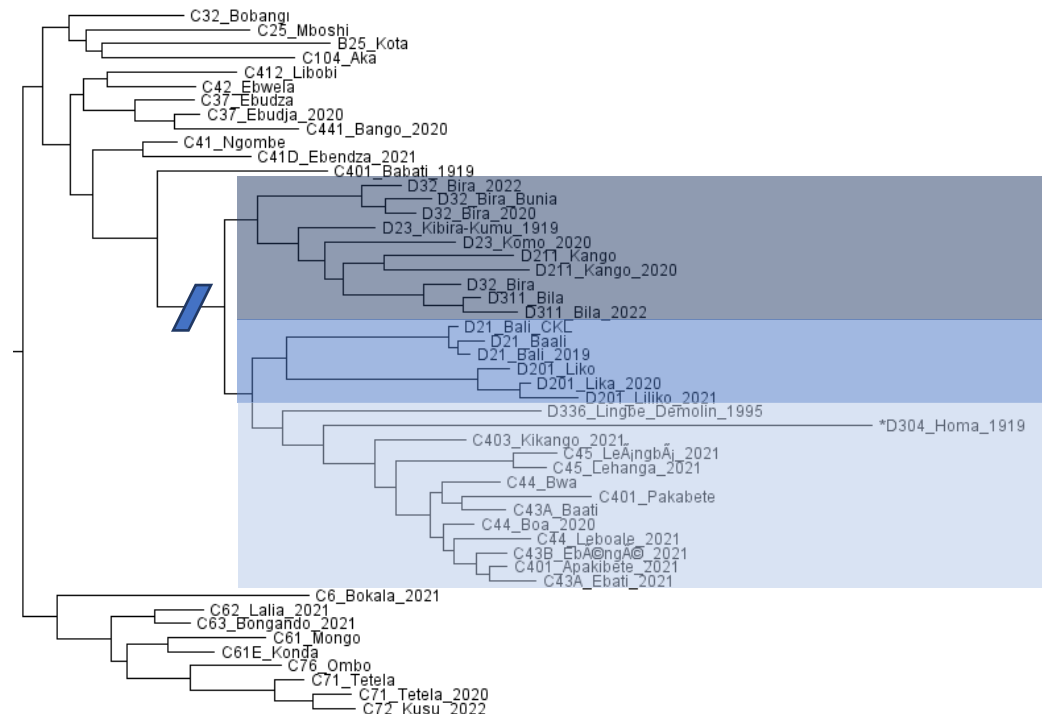
> Bali-Liko

> East-Boan

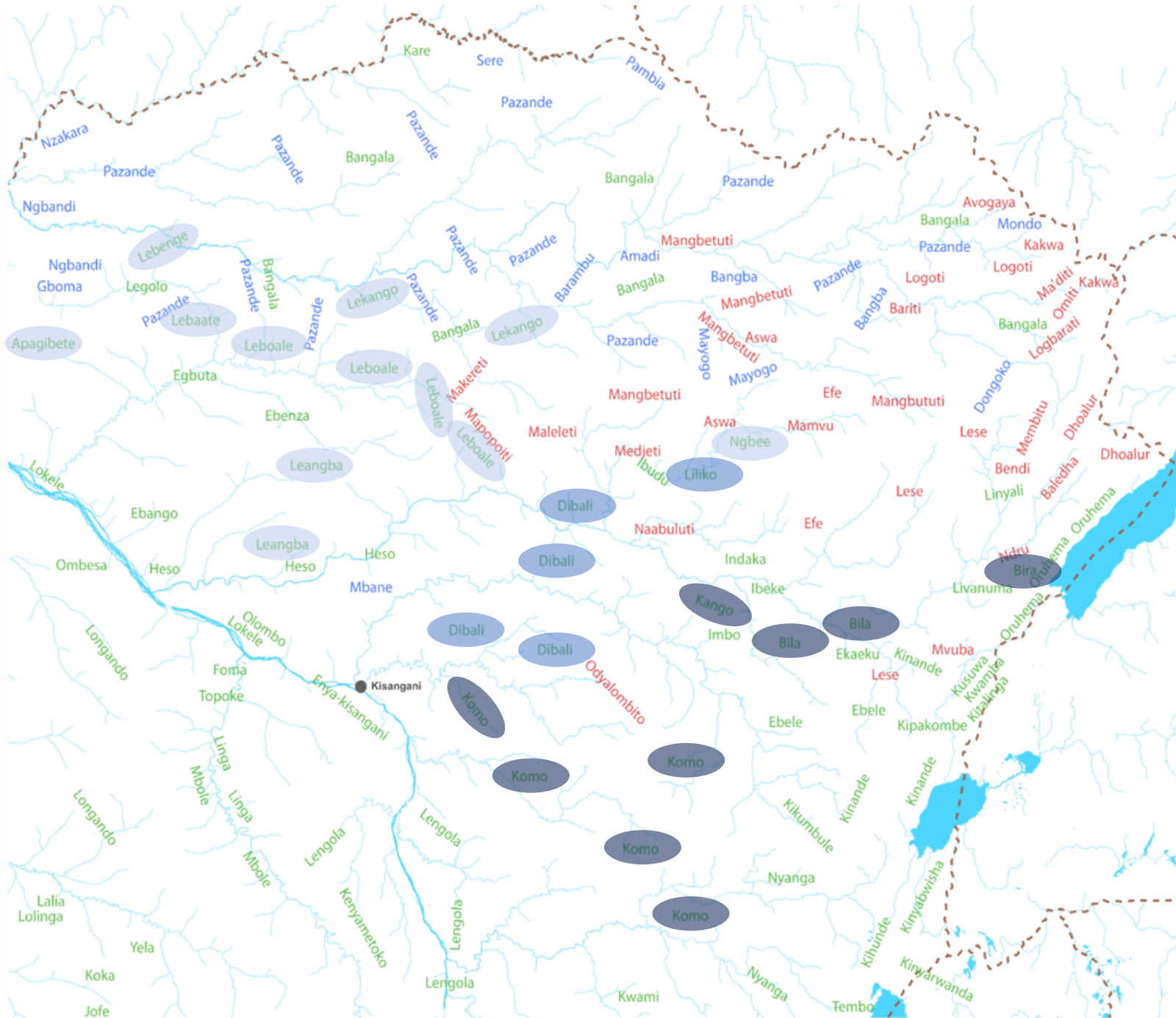
Recent only:



All:



~~~~~B~~~~~  
~~~~~AN~~~~~  
~~~~~TU~~~~~  
~~~~~RIVER~~~~~  
~~~~~S~~~~~



Upper Congo

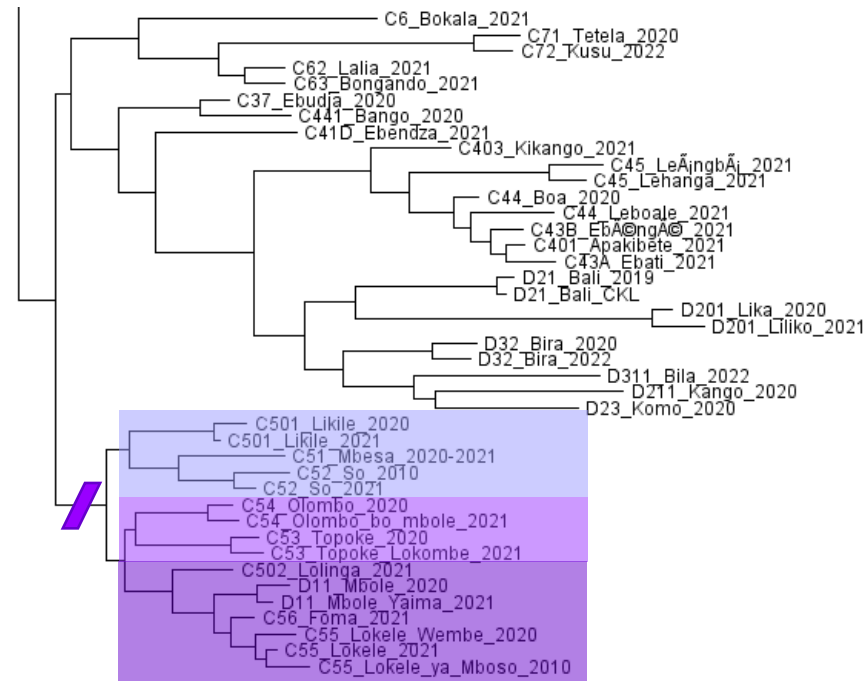
Forms a separate branch of Congo Basin, sometimes paired to Lebonya.

Two main subgroups can be detected, the southern also split in two.

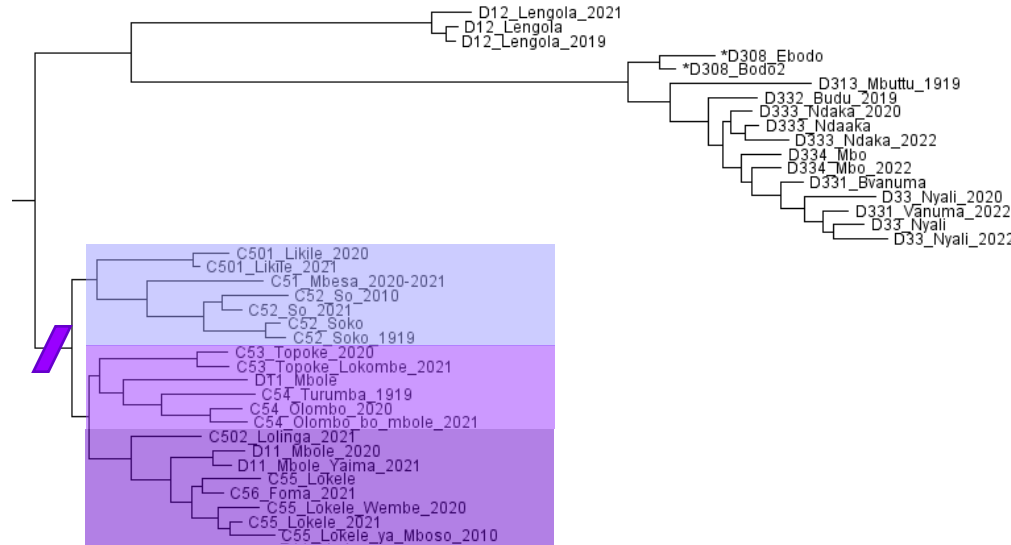
Congo Basin Bantu:

- > Inner Congo Basin + Congo Rivers
- > Upper Congo (+ Lebonya)
 - > North Upper Congo
 - > Central + South Upper Congo

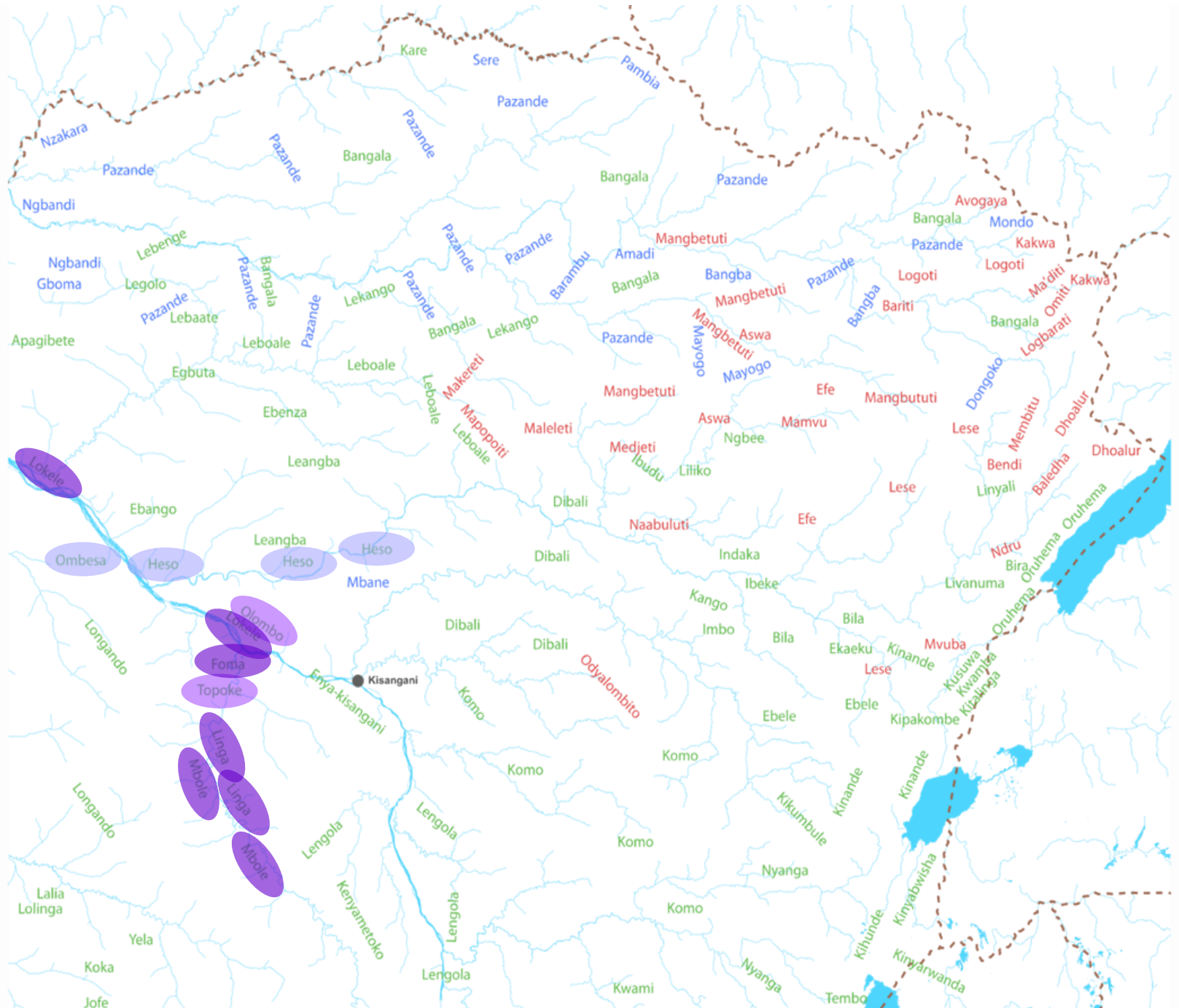
Recent only:



All:



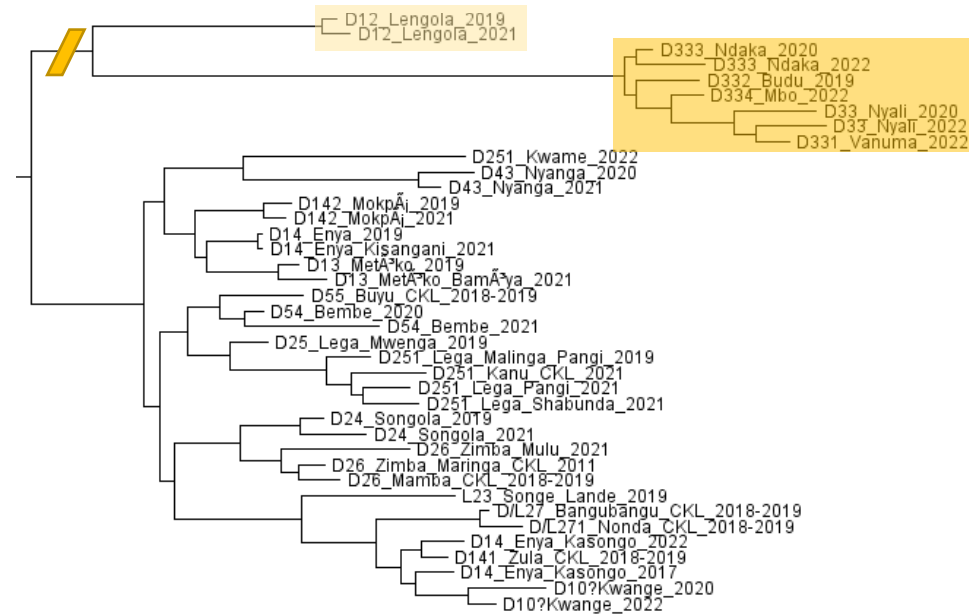
~~~~~B~~~~~  
AN~~~~~  
~~~~~TU~~~~~  
RIV~~~~~
~~~~~ER~~~~~  
~~~~~S~~~~~



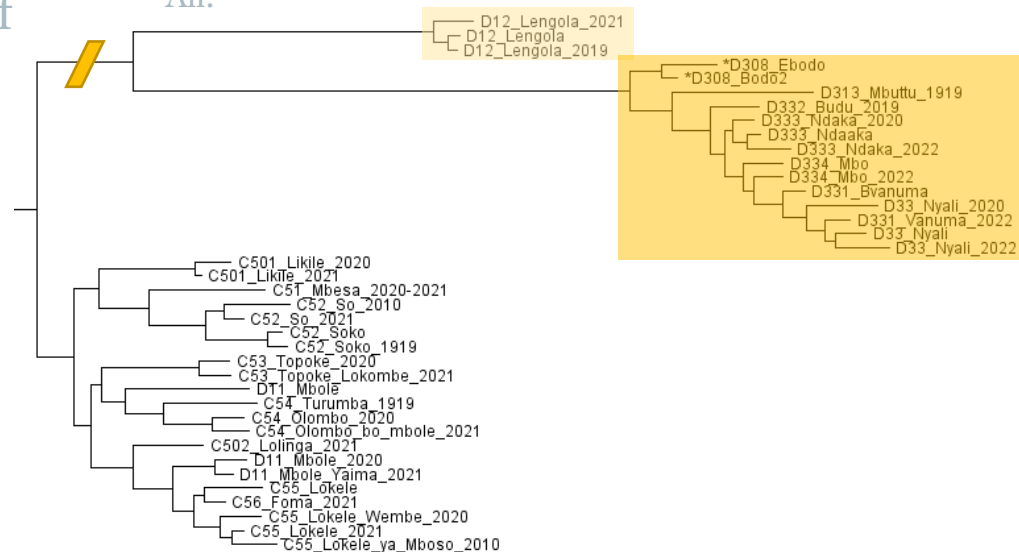
“Lebonya”

The “Lebonya” group was forwarded by Bastin et al. 1999 but considered “peripheral Bantu”. The subgroup seems clearly defined, consisting of two branches, but the higher nodes remain problematic: either grouped with East Bantu, or paired to Upper Congo as part of Congo Basin.

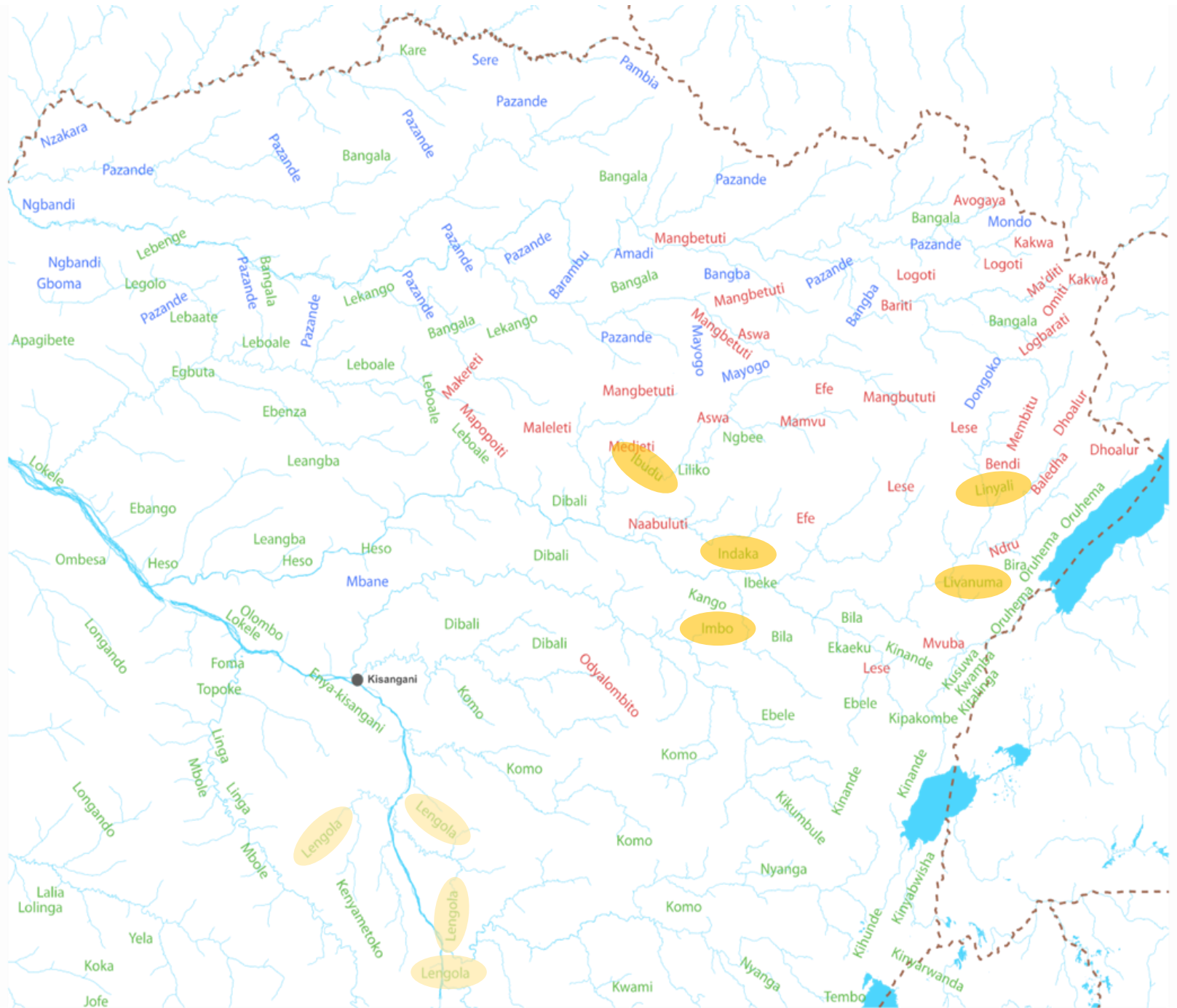
Recent only:



All:



~~~~~B~~~~~  
AN~~~~~  
~~~~~TU~~~~~  
RIV~~~~~
~~~~~ER~~~~~  
~~~~~S~~~~~



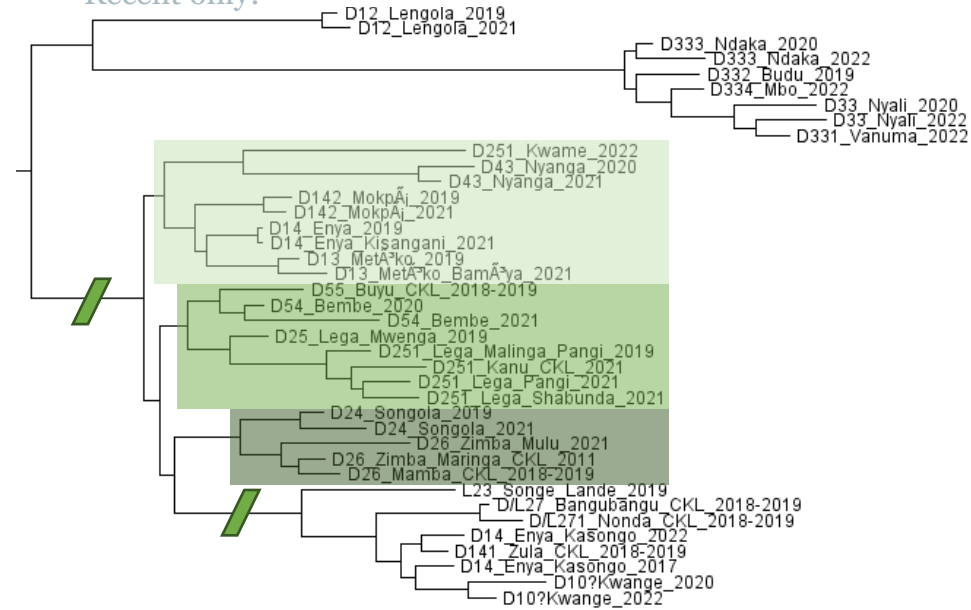
Lega

A well-defined subgroup of East Bantu, with three subgroups:

- Songola + Zimba
- Bembe-Buyu + Lega-Kanu
- Nyanga-Kwame + Mokpa-Enya-Metoko

Songola and Zimba display more similarities with other East-Bantu languages than the remaining Lega languages.

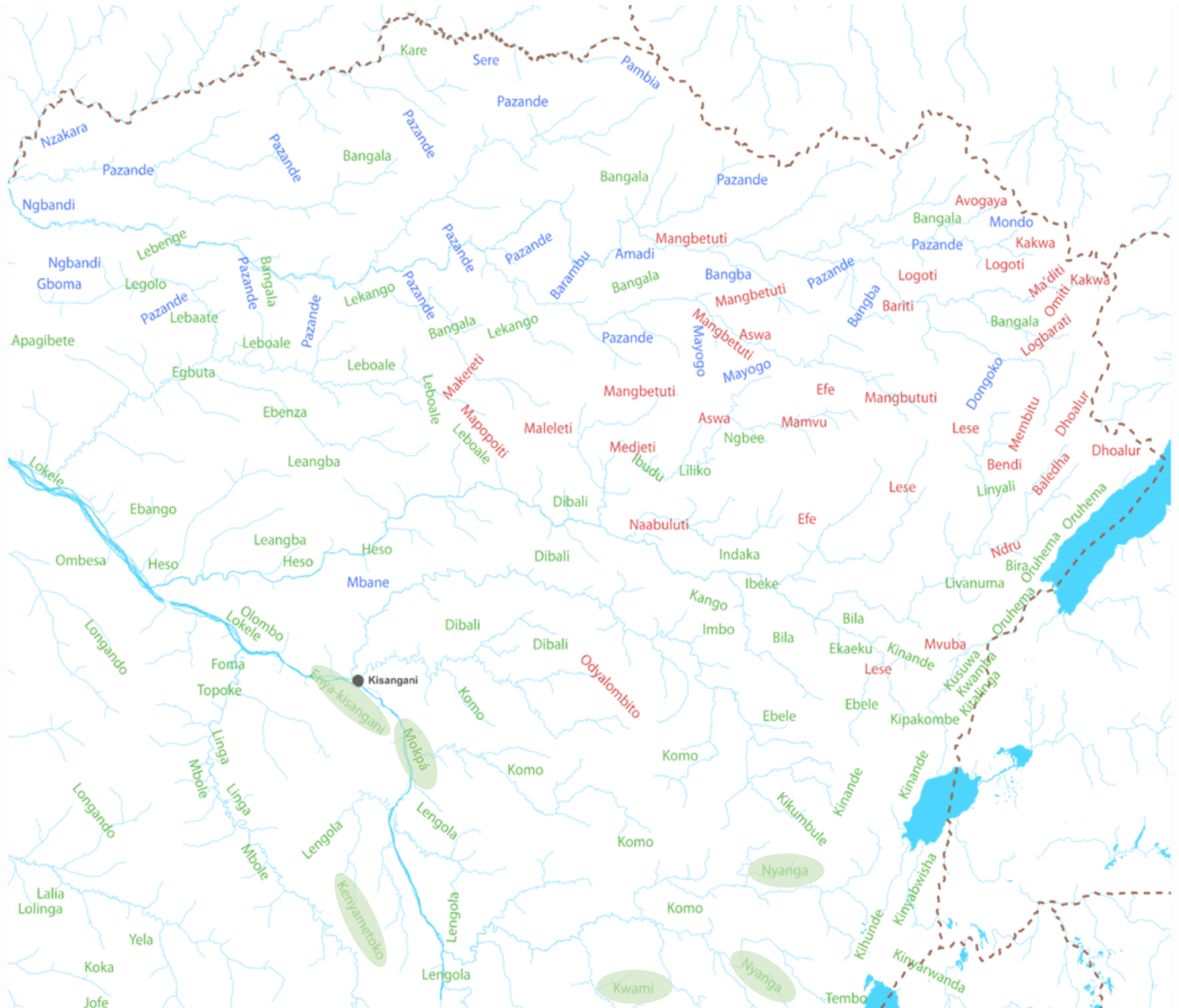
Recent only:



All:



~~~~~B~~~~~  
AN~~~~~  
~~~~~TU~~~~~  
RIV~~~~~
~~~~~ER~~~~~  
~~~~~S~~~~~

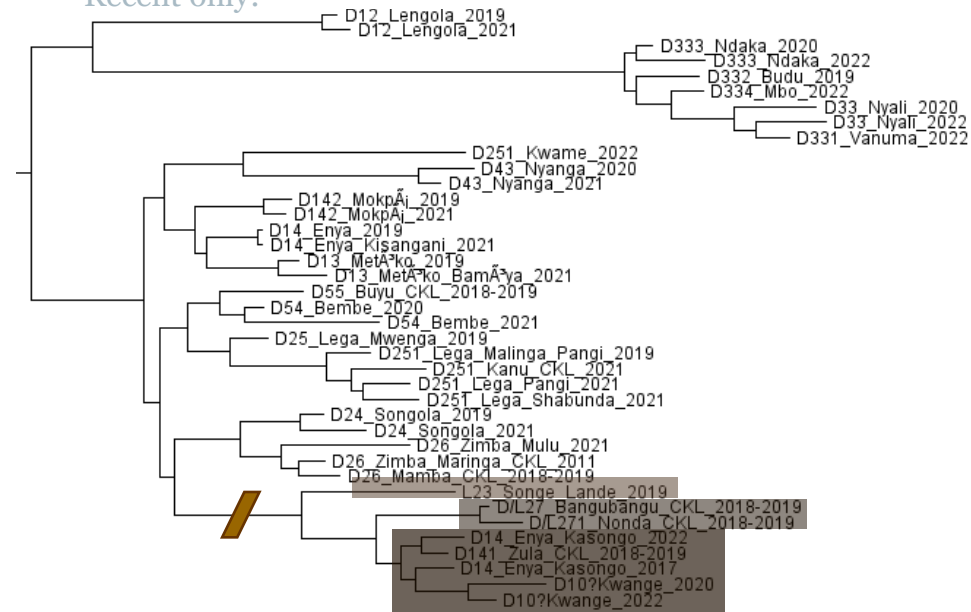


Luba

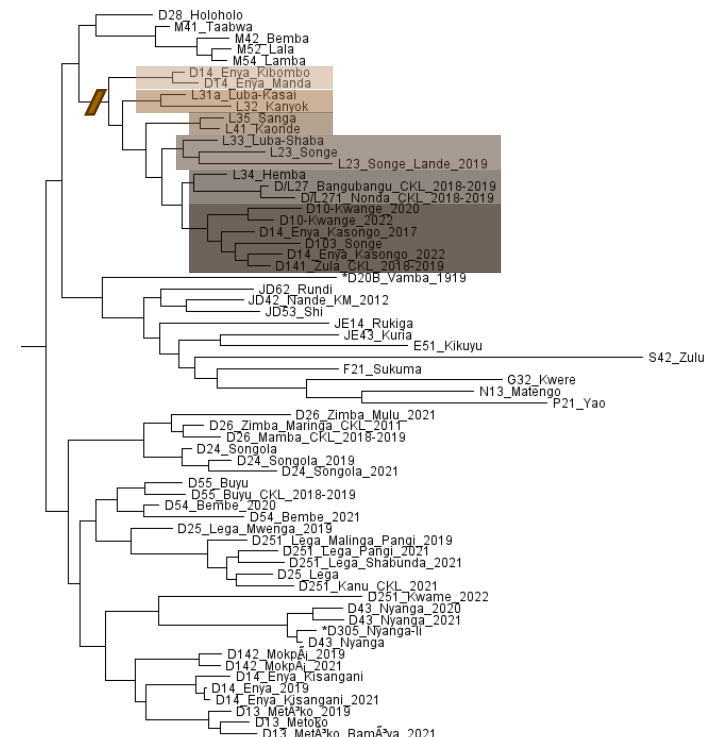
The languages of the southern Maniema province classified in zone D belong to the Luba subgroup. Rather than diverging into splits, this subgroup appears as a continuum.

The Luba group is most closely linked to Holoholo, Bemba and other M40 and M50.

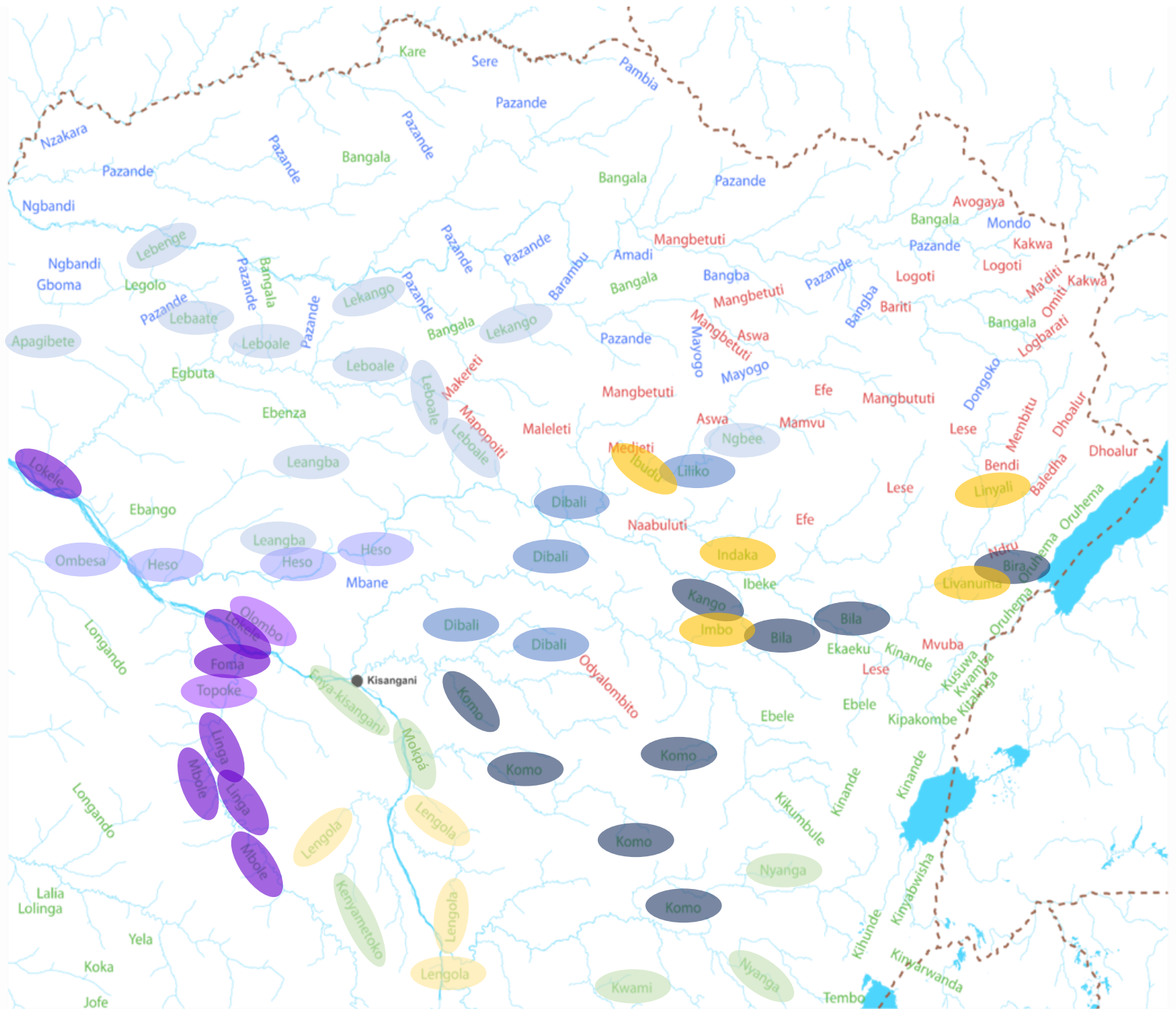
Recent only:



All:



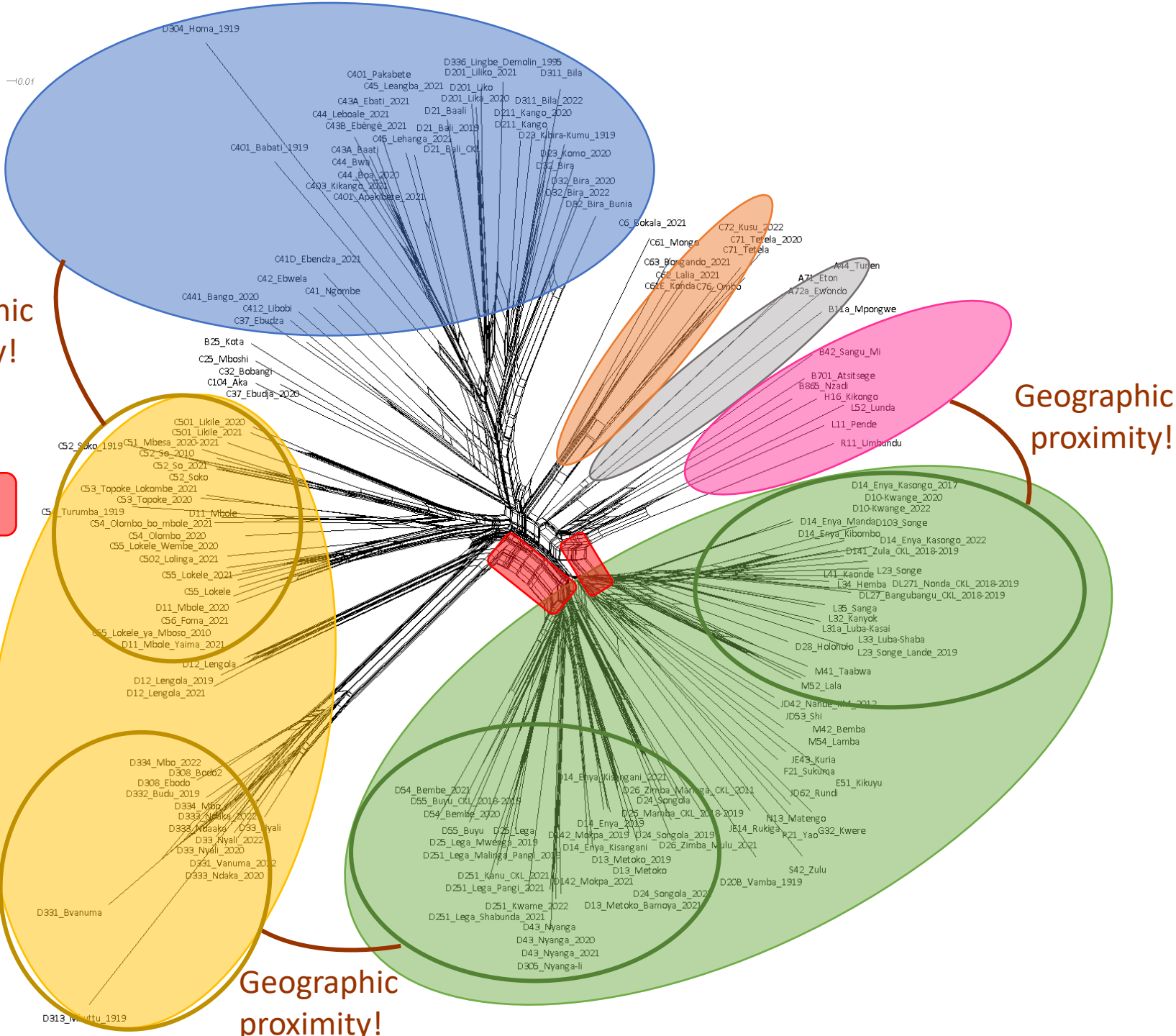
~~~~~B~~~~~  
AN~~~~~  
~~~~~TU~~~~~  
RIV~~~~~
~~~~~ER~~~~~  
~~~~~S~~~~~



The Puzzle of the Higher Nodes: Discussion

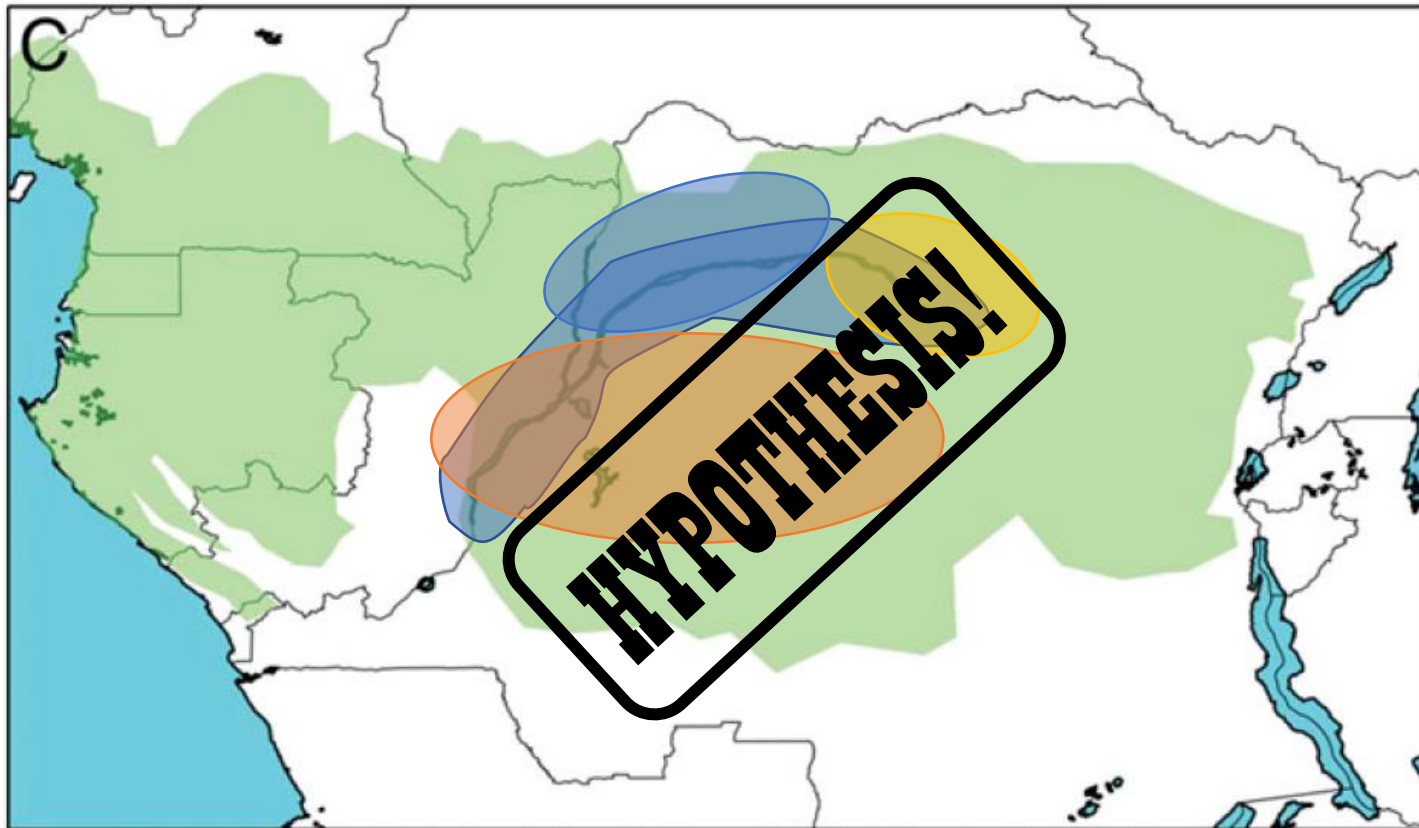




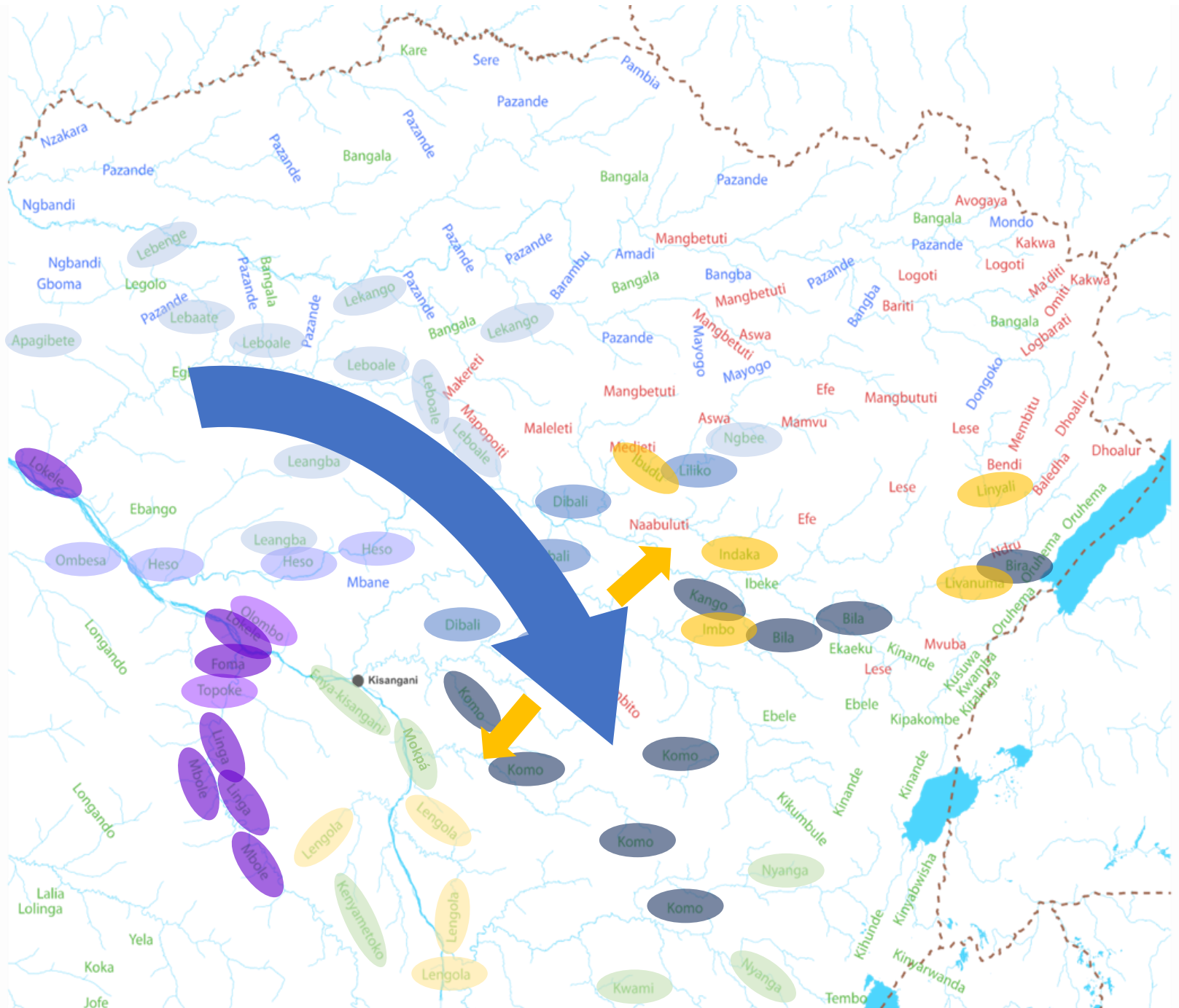


Upper Congo + Lebonya: What Kind of History?

- Upper Congo as original end of Congo Basin continuum? Implying Boan arrived later and separated Upper Congo from Lebonya



~~~~~B~~~~~  
AN~~~~~  
~~~~~TU~~~~~  
RIV~~~~~
~~~~~ER~~~~~  
~~~~~S~~~~~



Upper Congo + Lebonya: What Kind of History?

- “Lebonya”: East, West (Congo Basin) or Periphery?!
 - Influence of contact between Lengola and Mbole?
 - Time depth: divergence of Upper Congo and Bodo-Nyali seems relatively recent
 - Observation: the two subgroups are characterized by an important number of (different) lexical innovations
- => contact history? Similar context?

How to Solve the Puzzle?

- What if Lengola is highly influenced? Can leaving out Lengola lead to different results?
- Unravelling the sound and morphological changes of Lebonya may show more shared lexicon ~ less lexical innovations
- Contact: analyse the origins of lexical innovations in Upper Congo and Lebonya > Central-Sudanic or other origins?
- Time trees: can we identify which subgroups are relatively young?

**(Ø-)péle* ‘snake’: Upper Congo + Lebonya

| | |
|--------------------------|--------|
| C51_Mbesa_2020-2021 | ǫéle |
| C52_Soko | fele |
| C52_Soko_1919 | fele |
| C52_So_2010 | félé |
| C52_So_2021 | féle |
| C53_Topoke_2020 | péle |
| C54_Olombo_2020 | féle |
| C54_Olombo_bo_mbole_2021 | féle |
| D11_Mbole_2020 | féle |
| D11_Mbole_Yaima_2021 | féle |
| D12_Lengola_2019 | nnyéle |
| D12_Lengola_2021 | nyéli |
| D331_Nyali_2020 | ĩlí |
| D33_Nyali_2022 | yili |
| D331_Vanuma_2022 | hǎlí |
| D332_Budu_2019 | zéé |

~ B ~
~ AN ~
~ TU ~
~ RIV ~
~ ER ~
~ S ~



Thanks!

To Jacky Maniacky for the linguistic map.

To Didier Demolin for the data on Lingbe.

To the University of Kisangani, in particular prof. Charles Kumbatulu, for supporting active collaborations in the field.

This presentation is part of the BANTURIVERS project which benefits from an ERC Starting Grant (2019-2023, ULB, PI: Birgit Ricquier).

