Introducing a state-of-the-art phylogenetic classification of the Kikongo language cluster

Gilles-Maurice de Schryver,
Rebecca Grollemund,
Simon Branford &
Koen Bostoen

Supplementary material

The supplementary material consists of:

0. This **Read me** file
1. **Cognates**: an Excel sheet with the multistate encoding underlying our phylogenetic classification of the Kikongo language cluster (KLC)
2. **Nexus**: the nexus file for BayesPhylogenies
3. **Distance Matrix for Fig. 24**
4. **Distance Matrix for Fig. 25**
5. **Covarion**: Results of the model testing to decide on the number of covarion categories to use

The data in the Excel sheet is an extract from a larger database, and should be interpreted as follows:

*   The 95 West Bantu varieties used in the present study are shown in Column A. For ease of interpretation, we have post-coloured the name of each variety in column A, so it reflects the final outcome.
*   From columns B and onwards up to column GY, the 92 so-called 'Tervuren concepts' are listed (… YB 13, YB 14, …). Given our method allows for multiple synonyms to be included for a single concept, extra columns were inserted in-between (e.g., YB 13b and YB 13c).
*   The cognation judgements are indicated with background colours in each cell. So, for each concept (say 'burn *intr. v.*' = YB 13 + YB 13b + YB 13c), material in the same background colour is cognate. Those colours are random choices per concept, and there is of course no relation with colours used for the other 91 concepts.
*   More information on our cognation judgements is included at certain concepts, in pop-up 'Comment boxes'. See for instance at YB 01, YB 16, YB 21, YB 23, YB 28, etc.
*   Those background colours were subsequently turned into a multistate format, i.e. plain numbers (1, 2, 3, 4, …) per concept, and in a next step the latter were turned into a binary format consisting of either 0 or 1, meaning a certain cognate is either absent or present in a certain variety.
*   An equal sign '=' means there are no (further) synonyms for a certain concept in a certain variety, and a question mark '?' means that we were unable to collect the form of the said concept. Note that because this data is an extract from a much larger database, some columns in the attached extract only consist of equal sings. (At that point either other language varieties (not included in the 95 West Bantu varieties

reported on) have material in those columns, or historical language varieties that belong to the KLC (to be used in a forthcoming article) have data in there.)

- Column GZ shows the source(s) for each variety, with for the KongoKing fieldwork material further information from column HA onwards up to column HG.