# Bringing together linguistic and genetic evidence to test the Bantu expansion

Cesare de Filippo[1,2], Koen Bostoen[3,4,5], Mark Stoneking[2], and Brigitte Pakendorf[1,6]

[1]Max Planck Research Group on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig; Germany.
[2]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig; Germany.
[3]Ghent University, Leuvensesteenweg 13, 3080 Tervuren, Belgium.
[4]Université libre de Bruxelles, 50 avenue F.D. Roosevelt, B-1050 Brussels, Belgium
[5]Royal Museum for Central Africa, Leuvensesteenweg 13, B-3080 Tervuren, Belgium.
[6]*current affiliation*: Laboratoire Dynamique du Langage, UMR5596, CNRS & Université Lumière Lyon 2, Lyon, France.

# Electronic supplementary material

## Genetic data and analyses

To compare the genetic differentiation of Bantu-speaking populations with that observed in other African linguistic phyla we included data from as many African populations as possible. We considered only those populations from previous studies that have a sample size of at least 10 chromosomes. With respect to mtDNA, a total of 5,018 sequences of the first hypervariable segment, consisting of positions 16024-16383, excluding the polycytosine stretch at positions 16183-16193, and with less than two uncertain nucleotide callings, were included in the analyses. For the Y chromosome, we used the data from 2,445 individuals genotyped for 11 STRs and various SNPs reported in [1]. For the autosomes, data from 2,503 individuals genotyped for 848 STRs [2] were used for the analyses. We first removed 136 STRs with more than 40% missing values over all individuals and subsequently looked for overlap of the remaining 712 STRs with those reported in [3], ending up with a total of 560 STRs. We also removed 19 Bantu-speaking individuals with more than 20% of the STR genotypes missing. In addition, the Dogon population, which showed a very high level of population differentiation [2], was removed from the analyses. Tables S1-3 include the detailed list of populations used for each marker.

We calculated pairwise population genetic distances, using variants of Wright's F-statistics appropriate for each kind of genetic data, by means of in-house R scripts. $\Phi_{ST}$ for mtDNA sequences and $R_{ST}$ for Y-chromosomal and autosomal STRs were calculated following the formulas of Michalakis and Excoffier [4], while $F_{ST}$ based on Y-chromosomal haplogroup frequencies as reported in [1] was calculated according to Weir and Cockerham [5]. Negative values of F-statistics – which result when the variation within populations is higher than the variation between populations – were set to 0.

Population-specific measures of diversity such as haplotype diversity for haploid markers or expected heterozygosity for autosomal STRs averaged across loci were calculated by means of in-house R scripts. Although these measures of genetic diversity take into account heterogeneity in sample size, the standard deviation associated with them is inversely related to the sample size. Therefore, values of genetic diversity were also calculated as the average of 1,000 bootstraps of 10 chromosomes (i.e. the population size cutoffs used in this study, 5 and 10 individuals for diploid and haploid markers, respectively). In order to minimize the effect of missing values that varied across Bantu populations, we considered only 184 autosomal STRs (table S9) having less than 20% missing values in all 33 Bantu populations.

## Linguistic data

The linguistic dataset used in this study includes the lexical data which Bastin et al. [6] compiled for their lexicostatistical study on 542 different Bantu doculects - the varieties of languages that end up in documentations [7]. For that purpose, they used the Swadesh 100-wordlist of basic vocabulary adapted to African reality, i.e. 92 different glosses. Cognacy was assigned by the original compilers for all 542 doculects, who also reported the geographic location of the language recorded, though pointing out that *"(t)he vocabularies are to be thought of as language recorded at sample-points rather than representative of language areas"* [6, p. 8]. Of this dataset, we removed 15 languages with more than 20% missing data. Because the dataset also contains languages with more than one vocabulary list recorded at different geographic locations, we further removed 117 vocabulary lists, retaining those with the fewest missing values. If the amount of missing values was the same between the vocabulary lists, we randomly chose one (table S4). We added the language Kimbundu (H21) to the dataset using data from different sources [8–13], in order to have comparative linguistic data for the genetic data of the Mbundu population [14]. See further explanation in the next section.

## Lexical distances between languages with synonymous forms for a certain meaning

Where a language had synonymous forms for a certain meaning, both forms were considered in the comparisons. We here illustrate our coding scheme with some examples concerning synonymous forms for *white*: (i) The word *white* was considered as cognate between the language Koyo (C24), with two forms for *white* coded as cognate sets 8 and 24, and languages Pama (C30) and Lingala (C36) that both have only one word for *white*, coded as 8 and 24, respectively. (ii) *white* was treated as a cognate when Koyo was compared with Liliko (D00), which has two forms coded as 8 and 27. (iii) In the comparison between Koyo and another language such as Fulunga (B83), which has two synonyms for *white* coded as 16 and 31, the item was considered as only one difference. (iv) Similarly, in the comparison between Koyo and Swahili (G42), where *white* belongs to cognate set 1, or in the comparison between Swahili and Zulu (S42), where *white* belongs to cognate set 7, one difference was scored.

## Patristic distances among languages

The rate of lexical replacement might vary across words – e.g. some of the 92 glosses used for the lexical distances might evolve faster than others – and might also change over time for the same word. Therefore, linguistic distances among languages were calculated as patristic distances from unrooted phylogenetic trees inferred with Bayesian Markov Chain Monte Carlo methods (MCMC) implemented in BayesPhylogenies V.1.1 [15]. First, the linguistic dataset of 92 words was transformed into a binary state dataset consisting of 2832 characters, with 0 indicating the absence and 1 indicating the presence of a certain cognate. Second, three different models of evolution (labeled A, B, and C) were used to infer the phylogenetic trees, although for all models the rates of gain and loss of a character were assumed to be equal. The differences among the three models are as follows. Model A considers a two-state character covarion allowing the branch length to differ along the tree, which means that for the same character there are changes in the rate of evolution (a slow and a fast rate) at different times. Model B considers a two-state character covarion (the same as in model A) and rate heterogeneity across sites (i.e. characters) with a $\gamma$ distribution set to 4. Model C considers only the rate heterogeneity across sites ($\gamma = 4$). Each model was run twice with 30,000,000 iterations, sampling the tree every 10,000 iterations. Figure S1a shows that for each model different runs have a similar likelihood, and that models A and B have a higher likelihood than model C, which considers only the rate heterogeneity across characters. Since models A and B have virtually the same likelihood, they were considered for the further analyses, where the initial 40% of trees were discarded as burn-in time. A consensus tree using the median branch-length was generated for each run per model as well as combining the two runs (a total of 3,600 trees) by means of BayesTrees V1.1 [16]. Patristic distances were calculated for each consensus tree by means of the Python library DendroPy [17]. The patristic distances were highly correlated within each model (average Mantel test Z = 0.95, all $Ps < 0.0001$) and also between the two models (Z = 0.99, $P < 0.0001$). Hence, only one model was considered for the analyses, and we chose model A, since it was computationally faster than model B and two additional runs of model A were carried out (figure S1b). Finally, we calculated patristic distances for a total of 7,200 trees constructed according to a two-state covarion evolution model (model A) to generate a distribution of linguistic distances. Each of these 7,200 linguistic distances was tested for correlations via a Mantel test with each matrix of geographic distances predicted by the different models of Bantu migrations. The patristic distances from the consensus tree were used for all other correlation tests.

## Linking linguistic and genetic data

Table S5 shows the Bantu populations having genetic data for mtDNA, Y chromosome and autosomes that were linked to languages in Bastin et al. [6]. While there was some overlap between the populations for which we had mtDNA and Y chromosomal data (21 populations), there was little overlap between the populations genotyped for the uniparentally transmitted markers and the autosomal STRs (only three and two, respectively; cf. table S5).
The closest related language (belonging at least to the same Guthrie group) was chosen as corresponding to the genetic data for those populations whose language was not included in the study of Bastin and collegues [6]. The letter-number combinations in parentheses following language names in this paper are the codes these languages were given in Guthrie's (updated) referential classification of the Bantu languages [18, 19]. In other words, the first letter refers to the Guthrie zone (e.g. A, B, C, etc...) and the number expressed in tens (e.g. 10, 20, 30, etc...) represents the Guthrie area which includes languages/dialects that are closely related.

In addition, a second letter (e.g. a, b, c, etc...) is added to these codes to further indicate relationship among languages/dialect. For instance, the Guthrie group Bubi-Benga is indicated by the code A30 ('A' for the zone, '30' for the group), which in theory includes all languages coded from A31 to A39. The Bubi-Benga group contains the languages Bubi (A31), Batanga (A32) - which includes three dialects Banoo (A32a), Bapoko (A32b) and Batanga at Fifinda (A32C) -, Yasa (A33a), Kombe (A33b) and Benga (A34).
The language Ungom (B22b) was associated with the genetic data from Akele population from Gabon [20, 21] who speak a closely related language variety, i.e. Kele (B22a). The Kuvale speakers from Angola [22] were associated with the language Herero (R31). The language Lala (M52) was used for the Lala and Bisa (M51) speakers from [1] that were merged into the Lala-Bisa population, given that both languages belong to the same group Lala-Bisa (M50). Although it was not clear whether the Mbundu, *"the second largest population in Angola (making up one-quarter of the total population)"*, from Plaza et al. [14, p. 441] were Kimbundu or Umbundu speakers, we associated them with the Kimbundu language, which is the second most common language spoken in Angola [23]. The Kwangwa, Makoma, Kwandi, Kwamulonga, Mwenyi, Mbowe, Simaa and Luyi speakers from [1] are all together considered Luyana given that they all speak closely-related speech varieties belonging to the Luyana dialect cluster (K30).

## Model-based geographic distances

In order to take into account the main differences in the spatial and temporal dynamics of the early-split and late-split models, the model-based distances were calculated by using four waypoints that correspond to the centers (or hubs) of expansion that have been previously suggested [24–26]. We placed one hub at the homeland in the Nigerian-Cameroonian border area (origin hub), one in the Great Lakes Region (eastern hub), one in the lower Congo area (western hub), and one in the Katanga region of southeastern Democratic Republic of Congo (central hub). Three hubs were considered as obligatory waypoints between two languages/populations: the origin, western and eastern hubs for the early-split model, and the origin, western and central hubs for the late-split model (figure S3a). The model-based geographic distances were calculated following the classification of Bantu languages proposed by Vansina [26], with some modifications, considering 13 different groups which were further grouped in three meta-groups: 'North', 'West' and 'East'. In details, 'North' includes the Bantoid, Boan-Lebonya, Buneya, and North-West groups; 'East' includes East-Central, East-Coastal, East-Great Lakes, East-Kilimanjaro, East-South, and East-Southeast; and 'West' includes West-Coastal, West-Congo Basin, and West-Southwest. It is worth mentioning again that the six 'East' groups and the three 'West' groups are more closely related amongst each other than to the other groups included in the 'North' meta-group. Based on this classification, the geographic distances were calculated as predicted from both models. For the early-split model, the distance between an Eastern and a Western Bantu language was calculated as the sum of the distances between the eastern, origin, and western hubs, and from the languages to the eastern and western hubs, respectively. For the late-split model, the distance between an Eastern and a Western language was calculated as the sum of the distances from the language to the central and western hubs, respectively, plus the distance between the western and central hubs (figure S3). Simple (i.e. with no waypoints) great-circle distances were calculated between two locations to represent the IBD model.

## Significance of correlation coefficients

Spearman rank correlations ($\rho$) were calculated for the filtered distances when using only the 'informative' pairwise comparisons. These are comparisons that have different values according to the late-split and early-split. Since these were not matrix-like, p-values ($P$s) were calculated by 10,000 permutations as follows. First, we indexed the 'informative' pairwise comparisons that differ between the late-split and early-split model; second, we permuted the rows and columns of one matrix; third, we considered the indexed 'informative' pairwise comparisons of the permuted matrix and tested for correlation with the other matrix (i.e. lexical distances). A Fisher r-to-z transformation was used to test for differences between correlation coefficients using the number of languages (i.e. n = 412) as independent observations, and not the total number of pairwise comparisons ($\frac{n\times(n-1)}{2}$).

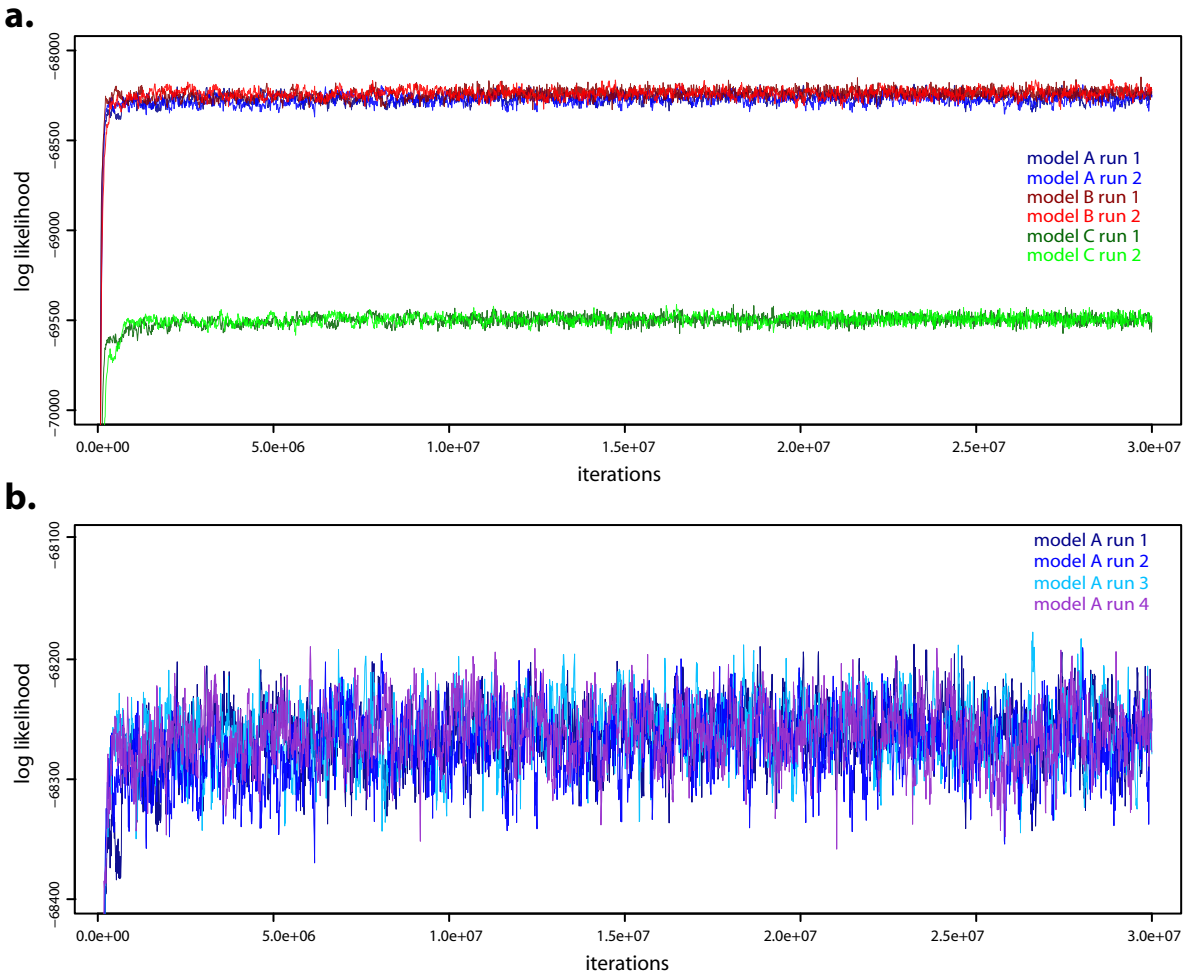


Figure S1: Log likelihood (y axis) of the tree for each language evolution model for 30,000,000 iterations. (a) Two runs (MCMC chain) per model are plotted. (b) Four runs with a two-state covarion model (model A) shows convergence of MCMC chains.

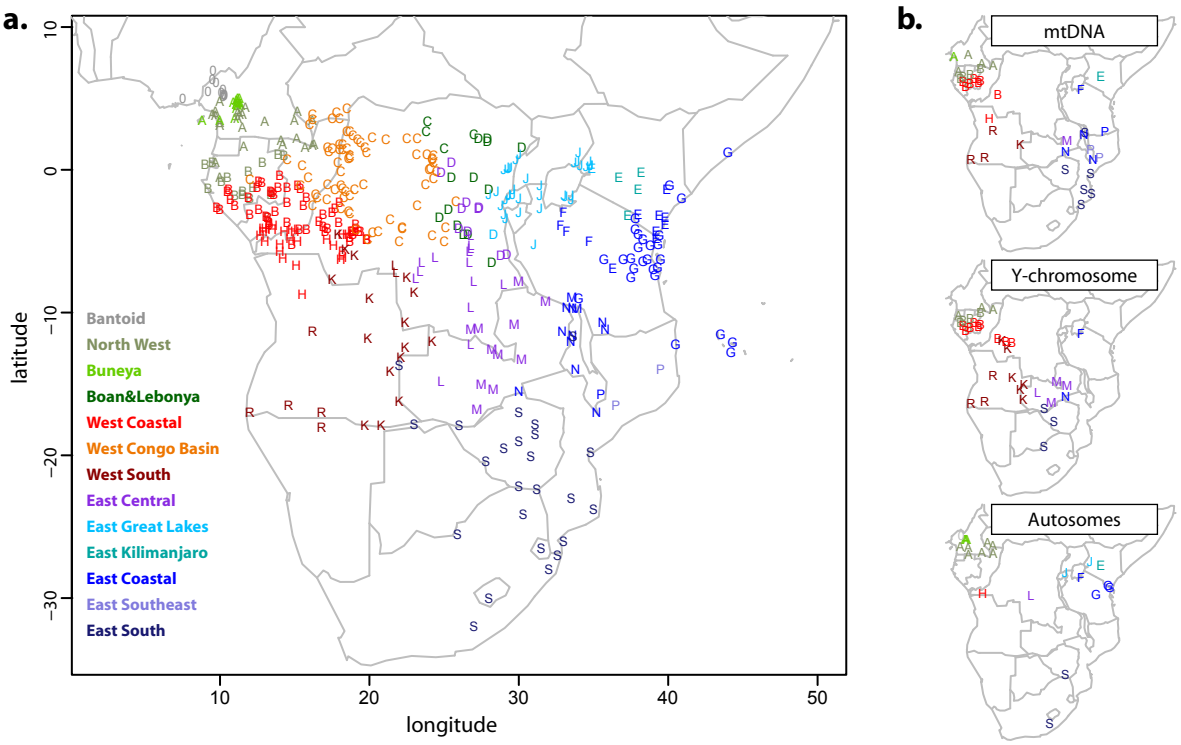


Figure S2: Sampling locations of the 412 Bantu languages considered in this study (a). The letters corresponding to Guthrie's major zones [18] are colored according to the historical linguistic classification. The small maps (b) show the location of the languages that can be associated to populations having genetic data for mtDNA, Y-chromosomal and autosomal data.

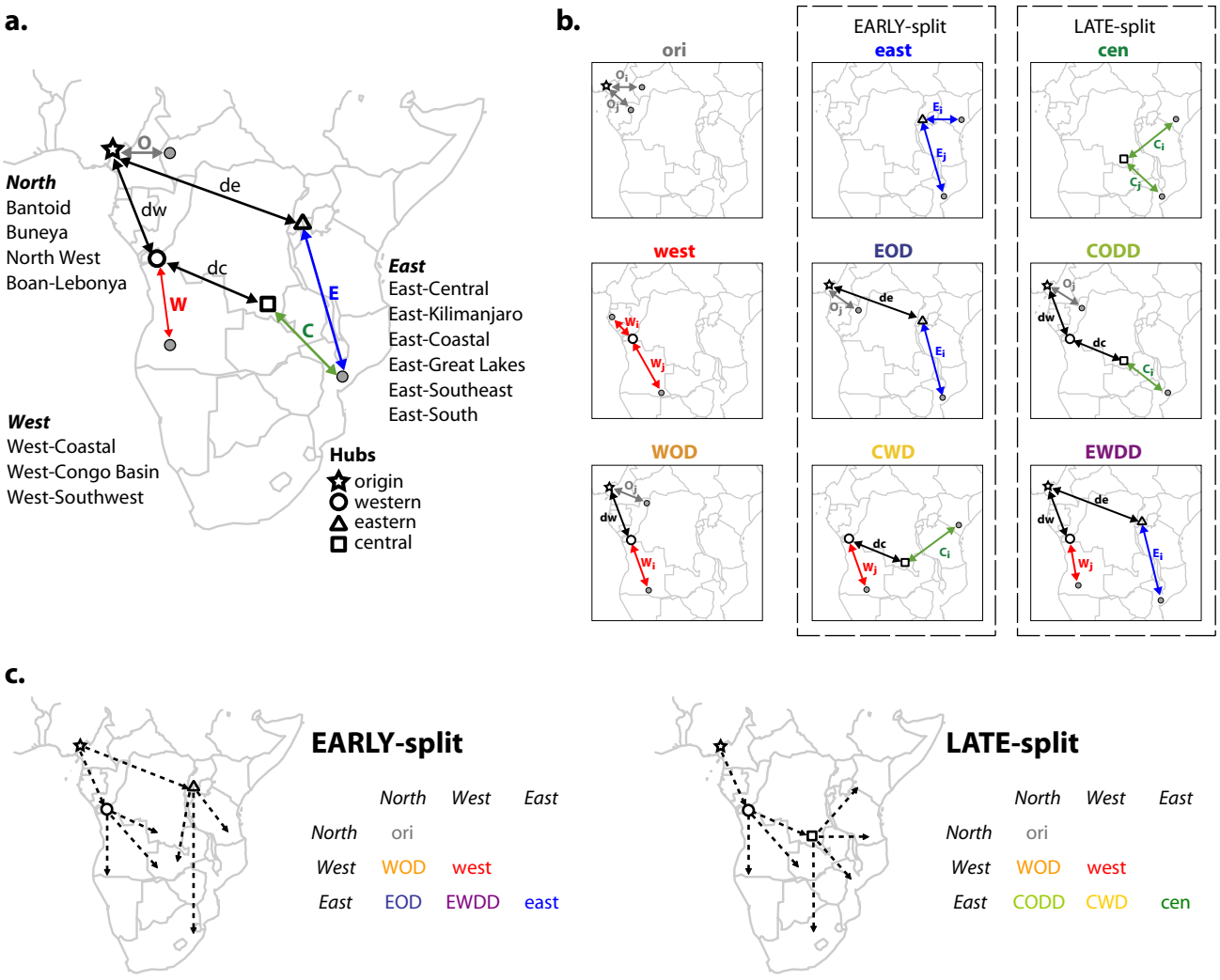


Figure S3: Schematic representation of model-based geographic distances of the late-split and early-split models. All distances were calculated between languages (gray circles) and hubs (black filled squares) of migrations using waypoints. The black arrowes correspond to constant distances between: origin and western hubs *dw*, origin and eastern hubs *de*, western and central hubs *dc*. The colored arrowes with the capital letter on top indicate the distances between hubs and 'North', 'West' or 'East' languages: in gray between a 'North' language and the origin hub; in red between a 'West' language and the western hub; in green between an 'East' language and the central hub; in blue between an 'East' language and the eastern hub (A). These lines were combined in nine different ways for the distances between the *ith* and *jth* language considered (B). The dashed lines group distances that are specific to certain models. All distances were then included in the model calculations as indicated in the matrices where each *ith* and *jth* column refers to the major groups (C). In all models 'simple' geographic distances were considered in the comparisons between languages belonging to the same historical group.

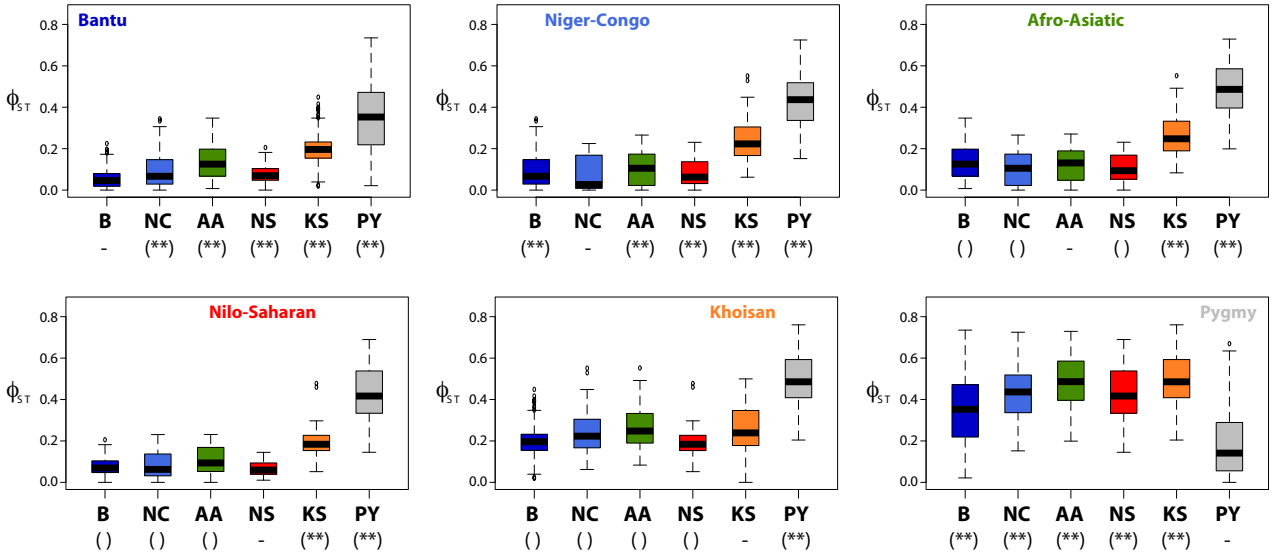


Figure S4: Box-plots of the mtDNA $\Phi_{ST}$ distances' distributions among populations grouped in six major linguistic and ethnic groups: Bantu 'B' (dark blue), Niger-Congo non-Bantu 'NC' (light blue), Afro-Asiatic 'AA' (green), Nilo-Saharan 'NS' (red), Khoisan 'KS' (orange), Pygmies 'PY' (gray). The name on each plot indicates the population that is compared to each of the other populations in the plot. Two asterisks (**) indicate that the within-group distances are significantly lower than the distances between-groups after correction for multiple tests, and one asterisk (*) indicates significance without correction for multiple tests. The dash is under the distribution of distances within the group.

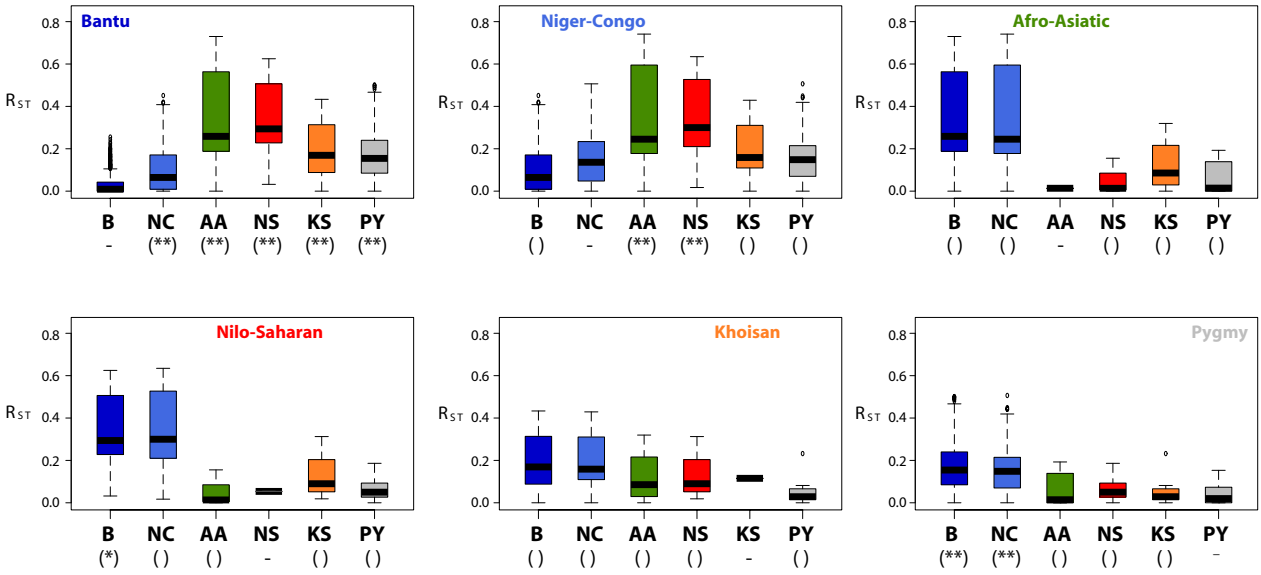


Figure S5: Box-plots as in figure S4, but with Y-chromosomal $R_{ST}$ distances.

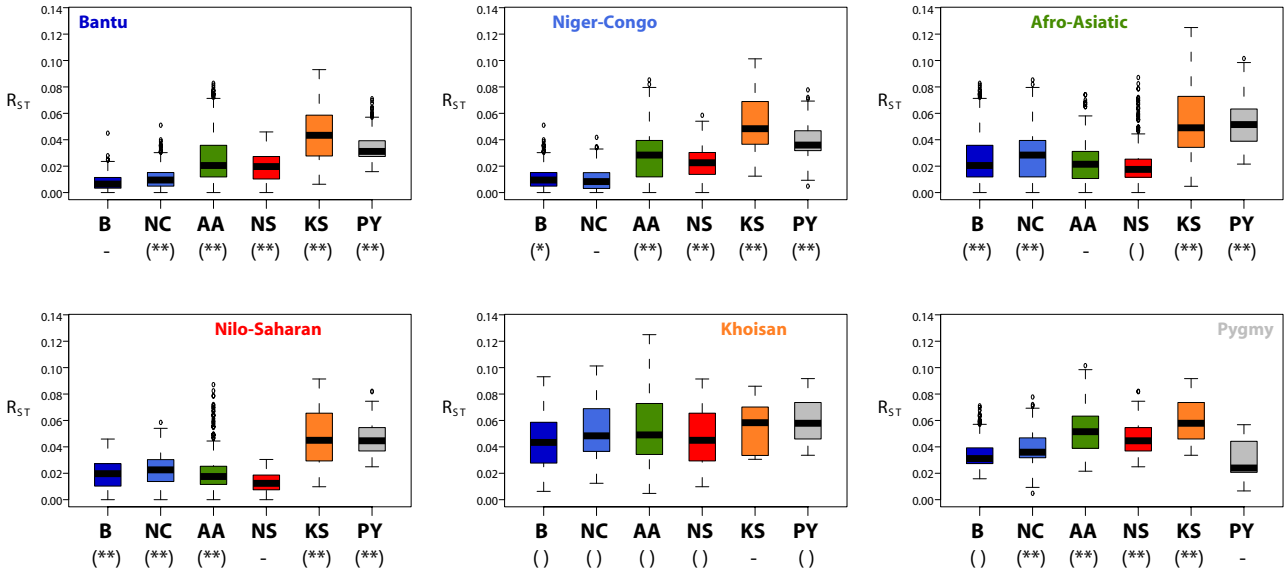


Figure S6: Box-plots as in figure S4, but with autosomal $R_{ST}$ distances.

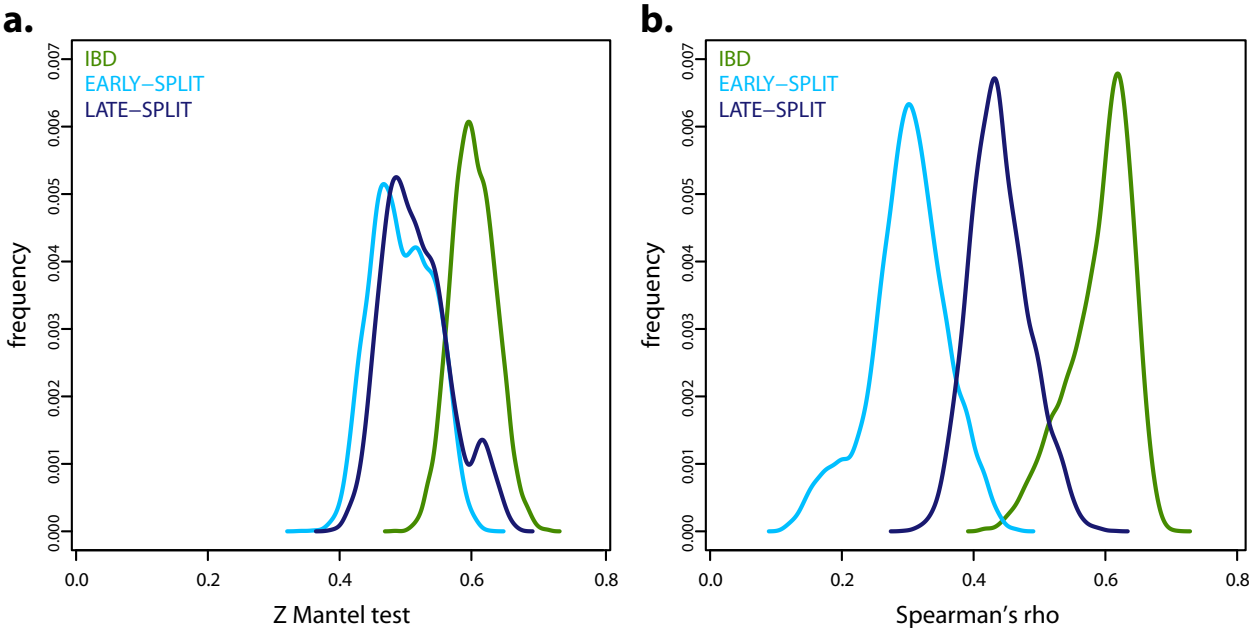


Figure S7: Distributions of correlation coefficients between the linguistic distances (7,200 trees) and the models of Bantu migration. In (a) all pairwise language comparisons were used and in (b) only the 'informative' pairwise comparisons, i.e. those with different predictions between the early-split and late-split models. All distributions among the models of migrations using both all and 'informative' comparisons are extremely different (two tails Mann-Whitney U tests with $P < 10^{-16}$).

Table S1: Details for the mtDNA data used in this study. "N" is the sample size as number of individuals, "Long" the longitude, "Lat" the latitude and "Ref" the reference study.

| **Population** | **N** | **Country** | **Group** | **Long** | **Lat** | **Ref** |
|---|---|---|---|---|---|---|
| Berber | 60 | Morocco | Afro-Asiatic | -4.7 | 33.8 | [27] |
| Burunge | 36 | Tanzania | Afro-Asiatic | 36.0 | -5.3 | [28] |
| Egyptians | 70 | Egypt | Afro-Asiatic | 31.2 | 27.0 | [29] |
| Hausa | 20 | Niger | Afro-Asiatic | 12.4 | 13.6 | [30] |
| Somali | 26 | Somalia | Afro-Asiatic | 43.5 | 3.3 | [30] |
| Tuareg | 26 | Nigeria, Niger, Mali | Afro-Asiatic | -1.8 | 17.7 | [30] |
| West-Saharan | 25 | Western Sahara | Afro-Asiatic | -13.8 | 24.5 | [27] |
| Tikar (Pygmy) | 35 | Cameroon | Bantoid | 11.0 | 5.8 | [20] |
| Akele | 46 | Gabon | Bantu | 10.5 | -0.5 | [20] |
| Ateke | 53 | Gabon | Bantu | 17.2 | -4.1 | [20] |
| Babongo (Pygmy) | 45 | Gabon | Bantu | 12.7 | -2.3 | [20] |
| Bakola (Pygmy) | 30 | Cameroon | Bantu | 10.0 | 2.8 | [28] |
| Bakola (Pygmy) | 88 | Cameroon | Bantu | 10.0 | 2.8 | [20] |
| Bakoya (Pygmy) | 31 | Gabon | Bantu | 13.4 | 1.7 | [20] |
| Benga | 50 | Gabon | Bantu | 9.7 | 0.5 | [20] |
| Biaka (Pygmy) | 17 | CAR | Bantu | 17.0 | 4.0 | [28] |
| Biaka (Pygmy) | 55 | CAR | Bantu | 17.0 | 4.0 | [20] |
| Bisa | 41 | Zambia | Bantu | 30.2 | -13.3 | [31] |
| Bubi | 45 | Equatorial Guinea | Bantu | 8.8 | 3.5 | [32] |
| Chopi | 27 | Mozambique | Bantu | 33.0 | -25.0 | [33] |
| Chwabo | 20 | Mozambique | Bantu | 37.9 | -15.8 | [33] |
| Duma | 45 | Gabon | Bantu | 12.6 | -0.8 | [20] |
| Eshira | 40 | Gabon | Bantu | 10.5 | -1.5 | [20] |
| Eviya | 38 | Gabon | Bantu | 11.0 | -1.0 | [20] |
| Ewondo | 25 | Cameroon | Bantu | 11.5 | 3.9 | [20] |
| Fang-C | 39 | Cameroon | Bantu | 14.1 | 2.1 | [20] |
| Fang-G | 66 | Gabon | Bantu | 11.0 | 2.1 | [20] |
| Galoa | 51 | Gabon | Bantu | 10.1 | -0.8 | [20] |
| Ganguela | 20 | Angola | Bantu | 21.4 | -14.1 | [22] |
| Kikuyu | 24 | Kenya | Bantu | 36.7 | -0.5 | [30] |
| Kota | 56 | Gabon | Bantu | 13.9 | 1.1 | [20] |
| Kunda | 34 | Zambia | Bantu | 30.0 | -15.5 | [31] |
| Kuvale | 54 | Angola | Bantu | 12.0 | -17.0 | [22] |
| Lomwe | 19 | Mozambique | Bantu | 36.5 | -16.5 | [33] |
| Makhuwa | 20 | Mozambique | Bantu | 34.8 | -15.1 | [33] |
| Makina | 45 | Gabon | Bantu | 12.0 | -0.5 | [20] |
| Makonde | 17 | Mozambique | Bantu | 39.4 | -12.4 | [33] |
| Mbundu | 44 | Angola | Bantu | 15.5 | -8.8 | [14] |
| Mitsongo | 58 | Gabon | Bantu | 11.5 | -1.8 | [20] |
| Ndao | 19 | Mozambique | Bantu | 34.8 | -19.8 | [33] |
| Ndumu | 37 | Gabon | Bantu | 13.5 | -1.6 | [20] |
| Ngumba | 88 | Cameroon | Bantu | 10.3 | 3.0 | [20] |
| Nguni | 12 | Mozambique | Bantu | 33.7 | -11.7 | [33] |
| Nyaneka-Nkhumbi | 153 | Angola | Bantu | 14.6 | -16.5 | [22] |

| Population | N | Country | Group | Long | Lat | Ref |
|---|---|---|---|---|---|---|
| Nyanja | 20 | Mozambique | Bantu | 33.5 | -12.0 | [33] |
| Nyungwe | 19 | Mozambique | Bantu | 32.0 | -17.0 | [33] |
| Nzebi | 62 | Gabon | Bantu | 11.9 | -1.9 | [20] |
| Obamba | 46 | Gabon | Bantu | 13.9 | -1.2 | [20] |
| Orungu | 20 | Gabon | Bantu | 9.1 | -1.0 | [20] |
| Punu | 52 | Gabon | Bantu | 11.0 | -2.5 | [20] |
| Ronga | 21 | Mozambique | Bantu | 33.0 | -26.0 | [33] |
| Sao Tome | 50 | Sao Tome&Principe | Bantu | 6.6 | 0.2 | [32] |
| Sena | 21 | Mozambique | Bantu | 35.2 | -17.0 | [33] |
| Shake | 51 | Gabon | Bantu | 11.9 | -0.1 | [20] |
| Shangaan | 21 | Mozambique | Bantu | 33.5 | -23.0 | [33] |
| Shona | 17 | Mozambique | Bantu | 30.0 | -19.0 | [33] |
| Sukuma | 32 | Tanzania | Bantu | 33.0 | -3.0 | [28] |
| Tonga | 16 | Mozambique | Bantu | 34.9 | -22.9 | [33] |
| Tswa | 19 | Mozambique | Bantu | 35.0 | -23.8 | [33] |
| Turu | 28 | Tanzania | Bantu | 32.5 | -4.5 | [28] |
| Umbundu | 91 | Angola | Bantu | 16.2 | -11.3 | [22] |
| Yao | 10 | Mozambique | Bantu | 37.5 | -11.5 | [33] |
| Hadza | 74 | Tanzania | Khoesan | 35.3 | -3.8 | [28] |
| Hadzabe | 48 | Tanzania | Khoesan | 35.3 | -3.8 | [34] |
| Khoisan | 17 | South Africa | Khoesan | 18.0 | -30.0 | [28] |
| Khwe | 31 | SouthAfrica | Khoesan | 23.5 | -28.0 | [35] |
| Kung | 16 | Botswana | Khoesan | 21.6 | -20.5 | [36] |
| Sandawe | 80 | Tanzania | Khoesan | 35.5 | -5.5 | [28] |
| Vasikela-Kung | 43 | SouthAfrica | Khoesan | 22.5 | -27.7 | [35] |
| Aghem | 113 | Cameroon | Niger-Congo | 10.1 | 6.4 | [37] |
| Akan | 151 | Ghana | Niger-Congo | -2.8 | 5.8 | [37] |
| Annang | 106 | Nigeria | Niger-Congo | 7.7 | 5.1 | [37] |
| Baka (Pygmy) | 39 | Gabon | Niger-Congo | 12.5 | 1.0 | [20] |
| Baka Central (Pygmy) | 30 | Cameroon | Niger-Congo | 12.5 | 3.5 | [20] |
| Baka West (Pygmy) | 58 | Cameroon | Niger-Congo | 10.0 | 3.5 | [20] |
| Bamun | 105 | Cameroon | Niger-Congo | 10.9 | 5.7 | [37] |
| Efik | 144 | Nigeria | Niger-Congo | 8.4 | 4.9 | [37] |
| Ejagham | 131 | Nigeria | Niger-Congo | 8.3 | 5.0 | [37] |
| Ewe | 87 | Ghana | Niger-Congo | 0.5 | 6.6 | [37] |
| Fulbe | 61 | Nigeria | Niger-Congo | 8.1 | 12.6 | [30] |
| Ibibio | 501 | Nigeria | Niger-Congo | 7.9 | 4.9 | [37] |
| Igbo | 196 | Nigeria | Niger-Congo | 7.5 | 6.1 | [37] |
| Mauritanians | 30 | Mauritania | Niger-Congo | -11.0 | 19.7 | [27] |
| Mbenzele (Pygmy) | 12 | C.A.R. | Niger-Congo | 16.1 | 3.6 | [38] |
| Oron | 98 | Nigeria | Niger-Congo | 8.2 | 4.8 | [37] |
| Senegalese | 50 | Senegal | Niger-Congo | -12.6 | 13.0 | [27] |
| Serer | 22 | Senegal | Niger-Congo | -16.0 | 13.7 | [27] |
| Tikar | 34 | Cameroon | Niger-Congo | 11.5 | 6.1 | [37] |
| Wolof | 43 | Senegal | Niger-Congo | -16.0 | 14.4 | [27] |
| Yoruba | 32 | Nigeria | Niger-Congo | 4.0 | 8.0 | [30, 36] |
| Datoga | 38 | Tanzania | Nilo-Saharan | 35.5 | -4.5 | [28] |

| Population | N | Country | Group | Long | Lat | Ref |
|---|---|---|---|---|---|---|
| Kanuri | 13 | Nigeria | Nilo-Saharan | 14.2 | 16.5 | [30] |
| Mbuti (Pygmy) | 38 | D.R.C. | Nilo-Saharan | 29.0 | 1.0 | [20] |
| Nubians | 88 | Sudan | Nilo-Saharan | 30.5 | 19.1 | [29] |
| Songhai | 10 | Niger, Mali | Nilo-Saharan | -4.3 | 10.7 | [30] |
| Turkana | 33 | Kenya | Nilo-Saharan | 37.1 | 3.4 | [30] |

Table S2: Details for the Y chromosomal data used in this study. "N" is the sample size as number of individuals, "Long" the longitude, "Lat" the latitude and "Ref" the reference study.

| **Population** | **N** | **Country** | **Group** | **Long** | **Lat** | **Ref** |
|---|---|---|---|---|---|---|
| Burunge | 23 | Tanzania | Afro-Asiatic | 36.00 | -5.30 | [28] |
| Mozabite | 20 | Algeria | Afro-Asiatic | 3.00 | 32.00 | [1] |
| Akele | 50 | Gabon | Bantu | 10.50 | -0.50 | [21] |
| Ateke | 48 | Gabon | Bantu | 17.20 | -4.10 | [21] |
| Bakola (Pygmy) | 22 | Cameroon | Bantu | 10.00 | 2.80 | [21] |
| Bantu-K | 10 | Kenya | Bantu | 37.00 | -3.00 | [1] |
| Bemba | 10 | Zambia | Bantu | 28.20 | -12.60 | [1] |
| Benga | 48 | Gabon | Bantu | 9.70 | 0.50 | [21] |
| Biaka (Pygmy) | 23 | C.A.R. | Bantu | 17.00 | 4.00 | [1] |
| Lala-Bisa | 38 | Zambia | Bantu | 30.20 | -13.30 | [1, 31] |
| Duma | 46 | Gabon | Bantu | 12.60 | -0.80 | [21] |
| Eshira | 42 | Gabon | Bantu | 10.50 | -1.50 | [21] |
| Eviya | 24 | Gabon | Bantu | 11.00 | -1.00 | [21] |
| Fang | 64 | Cameroon, Gabon | Bantu | 14.10 | 2.10 | [21] |
| Fwe | 30 | Zambia | Bantu | 23.28 | -17.50 | [1] |
| Galoa | 47 | Gabon | Bantu | 10.10 | -0.80 | [21] |
| Ganguela | 11 | Angola | Bantu | 21.40 | -14.10 | [22] |
| Kalanga | 20 | Botswana | Bantu | 27.80 | -20.45 | [1] |
| Kota | 53 | Gabon | Bantu | 13.90 | 1.10 | [21] |
| Kunda | 36 | Zambia | Bantu | 30.00 | -15.50 | [1] |
| Kuvale | 25 | Angola | Bantu | 12.00 | -17.00 | [22] |
| Kwamashi | 27 | Zambia | Bantu | 22.70 | -16.60 | [1] |
| Lozi | 94 | Zambia | Bantu | 26.00 | -17.90 | [1] |
| Luvale | 16 | Zambia | Bantu | 22.10 | -13.10 | [1] |
| Luyana | 61 | Zambia | Bantu | 22.00 | -16.20 | [1] |
| Makina | 43 | Gabon | Bantu | 12.10 | 0.00 | [21] |
| Mbala | 12 | DRC | Bantu | 17.90 | -4.50 | [1] |
| Mbugwe | 14 | Tanzania | Bantu | 35.80 | -3.80 | [28] |
| Mbukushu | 14 | Zambia | Bantu | 22.21 | -16.34 | [1] |
| Mbunda | 49 | Zambia | Bantu | 19.90 | -11.80 | [1] |
| Mbuun | 12 | DRC | Bantu | 19.80 | -4.90 | [1] |
| Mitsongo | 60 | Gabon | Bantu | 11.50 | -1.75 | [21] |
| Ndumu | 36 | Gabon | Bantu | 10.30 | 3.00 | [21] |
| Ngumba | 24 | Cameroon | Bantu | 11.90 | -1.90 | [21] |
| Nkoya | 29 | Zambia | Bantu | 24.80 | -14.80 | [1] |
| Nyaneka-Nkhumbi | 74 | Angola | Bantu | 14.60 | -16.50 | [22] |
| Nyengo | 11 | Zambia | Bantu | 22.37 | -16.47 | [1] |
| Nzebi | 57 | Gabon | Bantu | 13.90 | -1.20 | [21] |
| Obamba | 47 | Gabon | Bantu | 13.90 | -1.20 | [21] |
| Orungu | 21 | Gabon | Bantu | 9.10 | -1.00 | [21] |
| Pende | 11 | DRC | Bantu | 19.00 | -6.00 | [1] |
| Punu | 58 | Gabon | Bantu | 11.00 | -2.50 | [21] |
| Shake | 43 | Gabon | Bantu | 11.90 | -0.10 | [21] |
| Shanjo | 29 | Zambia | Bantu | 24.04 | -17.20 | [1] |

| Population | N | Country | Group | Long | Lat | Ref |
|---|---|---|---|---|---|---|
| Subiya | 11 | Zambia | Bantu | 25.18 | -17.39 | [1] |
| Sukuma | 30 | Tanzania | Bantu | 33.00 | -3.00 | [28] |
| Tonga | 30 | Zambia | Bantu | 27.20 | -16.80 | [1] |
| Totela | 15 | Zambia | Bantu | 24.17 | -16.72 | [1] |
| Tswana | 20 | Botswana | Bantu | 25.90 | -25.50 | [1] |
| Turu | 20 | Tanzania | Bantu | 32.50 | -4.50 | [28] |
| Umbundu | 94 | Angola | Bantu | 16.20 | -11.30 | [22] |
| Yansi | 23 | D.R.C. | Bantu | 18.70 | -4.70 | [1] |
| Hadza | 54 | Tanzania | Khoesan | 35.30 | -3.80 | [28] |
| Sandawe | 67 | Tanzania | Khoesan | 35.50 | -5.50 | [28] |
| Baka (Pygmy) | 33 | Gabon | Niger-Congo | 12.50 | 1.00 | [21] |
| Bissa | 40 | Burkina Faso | Niger-Congo | -0.56 | 11.32 | [1] |
| Kassena | 33 | Burkina Faso | Niger-Congo | -1.03 | 11.15 | [1] |
| Lyela | 40 | Burkina Faso | Niger-Congo | -2.57 | 12.29 | [1] |
| Mandenka | 15 | Senegal | Niger-Congo | -12.00 | 12.00 | [1] |
| Marka | 33 | Burkina Faso | Niger-Congo | -3.28 | 12.79 | [1] |
| Mossi | 36 | Burkina Faso | Niger-Congo | -0.91 | 12.43 | [1] |
| Nuna | 29 | Burkina Faso | Niger-Congo | -2.01 | 11.62 | [1] |
| Pana | 19 | Burkina Faso | Niger-Congo | -3.27 | 13.26 | [1] |
| Samo-South | 41 | Burkina Faso | Niger-Congo | -2.93 | 12.67 | [1] |
| Samo-North | 37 | Burkina Faso | Niger-Congo | -3.25 | 13.20 | [1] |
| Samoya | 21 | Burkina Faso | Niger-Congo | -3.04 | 13.56 | [1] |
| Yoruba | 12 | Nigeria | Niger-Congo | 5.00 | 8.00 | [1] |
| Datoga | 31 | Tanzania | Nilo-Saharan | 35.50 | -4.50 | [28] |
| Karimojong-JieDodos | 118 | Uganda | Nilo-Saharan | 34.60 | 2.50 | [39] |
| Mbuti (Pygmy) | 11 | D.R.C. | Nilo-Saharan | 29.00 | 1.00 | [1] |

Table S3: Details for the autosomal data used in this study. "N" is the sample size as number of individuals, "Long" the longitude, "Lat" the latitude and "Ref" the reference study.

| **Population** | **N** | **Country** | **Group** | **Long** | **Lat** | **Ref** |
|---|---|---|---|---|---|---|
| Baggara | 23 | Cameroon | Afro-Asiatic | 14.5 | 12.5 | [2] |
| Beja-Banuamir | 23 | Sudan | Afro-Asiatic | 36.0 | 21.0 | [2] |
| Beja-Hadandawa | 19 | Sudan | Afro-Asiatic | 36.0 | 21.0 | [2] |
| Beta-Israel | 17 | Ethiopia | Afro-Asiatic | 38.0 | 12.0 | [2] |
| Borana | 32 | Kenya | Afro-Asiatic | 38.0 | 3.0 | [2] |
| Burji | 24 | Ethiopia | Afro-Asiatic | 37.8 | 5.5 | [2] |
| Burunge | 22 | Tanzania | Afro-Asiatic | 36.0 | -5.3 | [2] |
| ElMolo | 16 | Kenya | Afro-Asiatic | 36.8 | 2.8 | [2] |
| Fiome-Gorowa | 22 | Tanzania | Afro-Asiatic | 35.8 | -4.3 | [2] |
| Gabra | 17 | Kenya | Afro-Asiatic | 37.5 | 3.0 | [2] |
| Giziga | 24 | Cameroon | Afro-Asiatic | 14.3 | 10.3 | [2] |
| Hausa-C | 27 | Cameroon | Afro-Asiatic | 14.5 | 10.5 | [2] |
| Hausa-N | 16 | Nigeria | Afro-Asiatic | 8.0 | 12.0 | [2] |
| Iraqw | 46 | Tanzania | Afro-Asiatic | 35.5 | -4.0 | [2] |
| Konso | 14 | Ethiopia | Afro-Asiatic | 37.5 | 5.5 | [2] |
| Kotoko | 17 | Cameroon | Afro-Asiatic | 14.8 | 11.8 | [2] |
| Mada | 28 | Cameroon | Afro-Asiatic | 14.1 | 10.8 | [2] |
| Mandara | 26 | Cameroon | Afro-Asiatic | 14.0 | 11.3 | [2] |
| Massa | 15 | Cameroon | Afro-Asiatic | 15.3 | 10.3 | [2] |
| Mbugu | 22 | Tanzania | Afro-Asiatic | 38.5 | -4.8 | [2] |
| Mozabite | 29 | Algeria | Afro-Asiatic | 3.0 | 32.0 | [2] |
| Ouldeme | 26 | Cameroon | Afro-Asiatic | 14.3 | 11.0 | [2] |
| Podokwo | 30 | Cameroon | Afro-Asiatic | 12.1 | 11.0 | [2] |
| Rendille | 28 | Kenya | Afro-Asiatic | 37.5 | 2.3 | [2] |
| Wata | 6 | Kenya | Afro-Asiatic | 37.0 | 3.5 | [2] |
| Yaaku | 19 | Kenya | Afro-Asiatic | 37.0 | 0.5 | [2] |
| Zime | 30 | Cameroon | Afro-Asiatic | 14.5 | 9.0 | [2] |
| Zulgo | 22 | Cameroon | Afro-Asiatic | 14.0 | 10.8 | [2] |
| Bafia | 30 | Cameroon | Bantu | 11.0 | 4.8 | [2] |
| Baluba | 6 | D.R.C. | Bantu | 25.0 | -9.0 | [2] |
| Bamoun | 31 | Cameroon | Bantu | 10.8 | 5.5 | [2] |
| Banen | 25 | Cameroon | Bantu | 10.8 | 4.8 | [2] |
| Bantu-K | 11 | Kenya | Bantu | 37.0 | -3.0 | [2] |
| Bantu-SA | 8 | South Africa | Bantu | 24.3 | -25.6 | [2] |
| Batanga | 20 | Cameroon | Bantu | 10.0 | 3.0 | [2] |
| Batie | 16 | Cameroon | Bantu | 11.0 | 4.3 | [2] |
| Bulu | 22 | Cameroon | Bantu | 11.0 | 3.0 | [2] |
| Fang | 19 | Cameroon | Bantu | 13.0 | 2.5 | [2] |
| Gogo | 13 | Tanzania | Bantu | 36.0 | -6.0 | [2] |
| Iyassa | 37 | Cameroon | Bantu | 9.8 | 2.5 | [2] |
| Kikuyu | 22 | Kenya | Bantu | 37.0 | -1.0 | [2] |
| Kongo | 17 | D.R.C. | Bantu | 15.0 | -5.5 | [2] |
| Lemande | 26 | Cameroon | Bantu | 11.0 | 4.5 | [2] |
| Luhya | 17 | Kenya | Bantu | 34.5 | 0.5 | [2] |

| Population | N | Country | Group | Long | Lat | Ref |
|---|---|---|---|---|---|---|
| Mabea | 13 | Cameroon | Bantu | 10.3 | 2.9 | [2] |
| Mbugwe | 21 | Tanzania | Bantu | 35.8 | -3.8 | [2] |
| Mvae | 24 | Cameroon | Bantu | 12.0 | 3.0 | [2] |
| Ngumba | 27 | Cameroon | Bantu | 10.3 | 3.0 | [2] |
| Tikar-North | 13 | Cameroon | Bantu | 11.5 | 6.3 | [2] |
| Ntumu | 11 | Cameroon | Bantu | 10.5 | 2.3 | [2] |
| Pare | 23 | Tanzania | Bantu | 38.0 | -4.5 | [2] |
| Rangi | 36 | Tanzania | Bantu | 36.0 | -5.0 | [2] |
| Sambaa | 18 | Tanzania | Bantu | 38.3 | -4.5 | [2] |
| Tikar-South | 21 | Cameroon | Bantu | 11.5 | 5.5 | [2] |
| Sukuma | 10 | Tanzania | Bantu | 33.5 | -3.0 | [2] |
| Turu | 32 | Tanzania | Bantu | 35.0 | -5.0 | [2] |
| Tutsi-Hutu | 8 | Rwanda | Bantu | 30.0 | -2.0 | [2] |
| Venda | 13 | South Africa | Bantu | 30.0 | -22.5 | [2] |
| Wimbum | 15 | Cameroon | Bantu | 10.8 | 6.5 | [2] |
| Xhosa | 28 | South Africa | Bantu | 28.0 | -32.0 | [2] |
| Yambassa | 17 | Cameroon | Bantu | 11.3 | 4.8 | [2] |
| Bakola (Pygmy) | 42 | Cameroon | Bantu | 10.0 | 2.8 | [2] |
| Bedzan (Pygmy) | 17 | Cameroon | Bantu | 11.6 | 5.5 | [2] |
| Hadza | 63 | Tanzania | Khoesan | 35.3 | -3.8 | [2] |
| San | 6 | Namibia | Khoesan | 20.0 | -21.0 | [2] |
| Sandawe | 51 | Tanzania | Khoesan | 35.5 | -5.5 | [2] |
| XunKxoe | 8 | South Africa | Khoesan | 18.0 | -30.0 | [2] |
| Ashanti | 15 | Ghana | Niger-Congo | -1.0 | 6.0 | [2] |
| Bassange | 20 | Nigeria | Niger-Congo | 5.5 | 9.0 | [2] |
| Brong | 26 | Ghana | Niger-Congo | -2.0 | 7.5 | [2] |
| Dioula | 5 | Ivory Coast | Niger-Congo | -4.5 | 9.5 | [2] |
| Dogon | 9 | Mali | Niger-Congo | -3.0 | 14.0 | [2] |
| Fulani-Adamawa | 41 | Cameroon | Niger-Congo | 13.5 | 9.0 | [2] |
| Fulani-Mbororo | 13 | Cameroon | Niger-Congo | 14.8 | 11.8 | [2] |
| Gbaya | 15 | C.A.R | Niger-Congo | 15.0 | 5.0 | [2] |
| Gwari | 22 | Nigeria | Niger-Congo | 7.0 | 10.0 | [2] |
| Igala | 17 | Nigeria | Niger-Congo | 7.0 | 7.0 | [2] |
| Igbo | 28 | Nigeria | Niger-Congo | 7.0 | 6.0 | [2] |
| Koma | 12 | Nigeria | Niger-Congo | 12.7 | 8.5 | [2] |
| Mandenka | 22 | Senegal | Niger-Congo | -12.0 | 12.0 | [2] |
| Mbum | 13 | C.A.R | Niger-Congo | 13.5 | 5.5 | [2] |
| Tupuri | 22 | Cameroon | Niger-Congo | 14.8 | 10.3 | [2] |
| Yakoma | 6 | C.A.R | Niger-Congo | 22.3 | 4.3 | [2] |
| Yoruba | 25 | Nigeria | Niger-Congo | 4.0 | 8.0 | [2] |
| Yoruba (CEPH) | 22 | Nigeria | Niger-Congo | 5.0 | 8.0 | [2] |
| Baka (Pygmy) | 48 | Cameroon | Niger-Congo | 13.5 | 2.5 | [2] |
| Biaka (Pygmy) | 23 | C.A.R | Niger-Congo | 17.0 | 4.0 | [2] |
| Akie | 23 | Tanzania | Nilo-Saharan | 37.5 | -5.0 | [2] |
| Bulala | 15 | Chad | Nilo-Saharan | 18.0 | 13.0 | [2] |
| Datoga | 54 | Tanzania | Nilo-Saharan | 35.5 | -4.5 | [2] |
| Dinka | 17 | Sudan | Nilo-Saharan | 30.0 | 8.0 | [2] |

| Population | N | Country | Group | Long | Lat | Ref |
|---|---|---|---|---|---|---|
| Dorobo | 10 | Tanzania | Nilo-Saharan | 37.0 | -5.0 | [2] |
| Ilchamus | 27 | Kenya | Nilo-Saharan | 37.1 | 1.5 | [2] |
| Kaba | 27 | Chad | Nilo-Saharan | 16.8 | 8.0 | [2] |
| Kanembou | 5 | Chad | Nilo-Saharan | 15.0 | 14.0 | [2] |
| Kanuri | 31 | Cameroon | Nilo-Saharan | 14.3 | 11.3 | [2] |
| Laka | 33 | Chad | Nilo-Saharan | 16.0 | 8.0 | [2] |
| Luo | 28 | Kenya | Nilo-Saharan | 34.5 | -0.5 | [2] |
| Maasai-Ilgwesi | 21 | Kenya | Nilo-Saharan | 36.8 | 0.3 | [2] |
| Maasai-Mumonyot | 12 | Kenya | Nilo-Saharan | 37.0 | 0.6 | [2] |
| Maasai-Tanzania | 36 | Tanzania | Nilo-Saharan | 37.0 | -4.0 | [2] |
| Marakwet | 14 | Kenya | Nilo-Saharan | 35.5 | 1.3 | [2] |
| Nandi | 11 | Kenya | Nilo-Saharan | 35.5 | 0.0 | [2] |
| Ngambaye | 30 | Chad | Nilo-Saharan | 16.0 | 9.0 | [2] |
| Nuer | 18 | Sudan | Nilo-Saharan | 31.0 | 8.5 | [2] |
| Nyimang | 12 | Sudan | Nilo-Saharan | 29.5 | 12.3 | [2] |
| Okiek | 22 | Kenya | Nilo-Saharan | 36.0 | 0.3 | [2] |
| Pokot | 23 | Kenya | Nilo-Saharan | 35.5 | 1.5 | [2] |
| Sabaot | 20 | Kenya | Nilo-Saharan | 34.8 | 1.0 | [2] |
| Samburu | 18 | Kenya | Nilo-Saharan | 37.0 | 1.5 | [2] |
| Saravarious | 27 | Chad | Nilo-Saharan | 17.5 | 8.0 | [2] |
| Sengwer | 21 | Kenya | Nilo-Saharan | 35.0 | 1.0 | [2] |
| Shilook | 15 | Sudan | Nilo-Saharan | 32.0 | 10.0 | [2] |
| Tugen | 22 | Kenya | Nilo-Saharan | 35.8 | 0.8 | [2] |
| Turkana | 26 | Kenya | Nilo-Saharan | 36.0 | 3.0 | [2] |
| Mbuti (Pygmy) | 13 | D.R.C. | Nilo-Saharan | 29.0 | 1.0 | [2] |

Table S4: Details for the languages used in this study. "Guthrie" is the Guthrie's zone, "Hist-Groups" corresponds to the historical classification proposed by Vansina [26], "G-code" is the group's code, "L-code" is the language code, "missing" is the number of missing items.

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| 0 | Bantoid | 800 | Eja | Ejagham | 0 | 5.50 | 9.20 |
| 0 | Bantoid | 802 | Tiv | Tiv | 0 | 5.00 | 7.50 |
| 0 | Bantoid | 805 | Ama | Amasi | 9 | 6.10 | 9.75 |
| 0 | Bantoid | 806 | Amb | Ambele | 9 | 6.95 | 9.55 |
| 0 | Bantoid | 894 | Asu | Asumbo | 14 | 6.30 | 9.50 |
| 0 | Bantoid | 951 | Bang | Bangangte | 1 | 5.60 | 10.20 |
| 0 | Bantoid | 900a | Mif | Mifi | 3 | 5.20 | 10.30 |
| 0 | Bantoid | 900b | Band | Bandjoun | 0 | 5.30 | 10.30 |
| 0 | Bantoid | 900c | Dsc | Dschang | 0 | 5.40 | 10.10 |
| 0 | Bantoid | 970a | Fef1 | Fefe | 0 | 5.30 | 10.20 |
| 0 | Bantoid | 970b | Baf | Bafang | 1 | 5.10 | 10.20 |
| A | North-West | A15g | A-Mbo | Mbo | 0 | 4.80 | 10.00 |
| A | North-West | A24 | A-Dua | Duala | 0 | 3.80 | 9.70 |
| A | North-West | A26 | A-Pon | Pongo | 0 | 4.10 | 9.70 |
| A | North-West | A27 | A-Lim | Limba | 0 | 3.90 | 9.50 |
| A | Buneya | A31 | A-Bub | Bubi | 4 | 3.50 | 8.80 |
| A | North-West | A32a | A-Noh | Noho | 9 | 3.10 | 10.00 |
| A | North-West | A32c | A-Tan | Tanga | 0 | 3.30 | 10.00 |
| A | North-West | A34 | A-Be1 | Benga | 0 | 0.50 | 9.70 |
| A | Buneya | A43a | A-Bas | Basaa | 0 | 3.60 | 11.00 |
| A | Buneya | A43b | A-Kok | Koko | 0 | 3.50 | 10.00 |
| A | Buneya | A44 | A-Nen | Nen | 0 | 4.70 | 10.90 |
| A | Buneya | A51 | A-Lef | Lefa | 17 | 5.00 | 11.20 |
| A | Buneya | A51 | A-Maj | Maja | 6 | 4.80 | 11.20 |
| A | Buneya | A51 | A-Tum | Tumi | 4 | 4.90 | 11.10 |
| A | Buneya | A53 | A-Kpa | Kpa | 1 | 4.60 | 11.10 |
| A | Buneya | A53 | A-Rop | Rope | 2 | 4.60 | 11.10 |
| A | Buneya | A54 | A-Nja | Njanti | 6 | 5.00 | 11.30 |
| A | Buneya | A61a | A-Tuk | Tuki | 3 | 4.50 | 11.40 |
| A | Buneya | A62a | A-Gun | Gunu | 11 | 4.60 | 11.30 |
| A | Buneya | A62b | A-Mma | Mmala | 7 | 4.50 | 11.20 |
| A | Buneya | A62d | A-Kal | Kalonge | 0 | 4.30 | 11.10 |
| A | North-West | A71 | A-Eto | Eton | 0 | 4.00 | 11.20 |
| A | North-West | A71a | A-Ndo | Ndongo | 0 | 3.90 | 11.30 |
| A | North-West | A72a | A-Ewo | Ewondo | 0 | 3.90 | 11.50 |
| A | North-West | A72a | A-Yao | Yaounde | 0 | 3.60 | 11.20 |
| A | North-West | A72d | A-Yas | Yasem | 3 | 4.40 | 11.70 |
| A | North-West | A74a | A-Bul | Bulu | 1 | 3.00 | 11.50 |
| A | North-West | A75 | A-Fan | Fang | 2 | 2.10 | 14.10 |
| A | North-West | A75 | A-Mak | Make | 3 | -0.50 | 10.00 |
| A | North-West | A75 | A-Ntu | Ntum | 1 | 1.62 | 11.60 |
| A | North-West | A80 | A-Kw1 | Kwejo | 0 | 3.50 | 16.20 |
| A | North-West | A80 | A-Kw2 | Kweso | 5 | 1.80 | 16.30 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| A | North-West | A84 | A-Koo | Koozime | 0 | 3.40 | 13.00 |
| A | North-West | A85b | A-Bek | Bekwil | 0 | 1.80 | 14.20 |
| A | North-West | A86c | A-Mpy | Mpyemo | 0 | 3.50 | 15.00 |
| A | North-West | A87 | A-Bom | Bomwali | 3 | 1.55 | 15.70 |
| A | North-West | A87 | A-San | Sangasanga | 5 | 2.20 | 15.15 |
| A | North-West | A92 | A-Pom | Pomo | 2 | 1.70 | 16.20 |
| A | North-West | A93a | A-Ka2 | Kako | 0 | 4.10 | 15.20 |
| B | North-West | B11a | B-Pon | Pongwe | 1 | 0.35 | 9.00 |
| B | North-West | B11C | B-Ga2 | Galwa | 0 | -0.80 | 10.10 |
| B | North-West | B11d | B-Him | Himba | 6 | -1.48 | 11.54 |
| B | North-West | B11e | B-Nko | Nkomi | 1 | -1.30 | 9.22 |
| B | North-West | B20 | B-Sha | Shama | 0 | 0.55 | 12.85 |
| B | North-West | B21 | B-Sek | Seki | 0 | 0.50 | 9.50 |
| B | North-West | B22b | B-Ung | Ungom | 0 | -0.50 | 10.50 |
| B | North-West | B22c | B-Pov | Pove | 2 | -1.20 | 12.20 |
| B | North-West | B23 | B-Mbh | Mbahouin | 16 | -1.70 | 13.40 |
| B | North-West | B24 | B-Wu1 | Wumbu | 0 | -2.00 | 12.00 |
| B | North-West | B25 | B-Mah | Mahongwe | 0 | 1.00 | 14.30 |
| B | North-West | B25 | B-Sa1 | Sake | 7 | -0.10 | 11.90 |
| B | North-West | B25b | B-Ko3 | Kota | 0 | 1.10 | 13.90 |
| B | North-West | B26 | B-Nd1 | Ndasa | 2 | -3.30 | 13.20 |
| B | North-West | B31 | B-Tsg | Tsogo | 4 | -1.75 | 11.50 |
| B | North-West | B33 | B-Pin | Pinji | 1 | -1.80 | 11.00 |
| B | West-Coastal | B40 | B-Bon | Bongo | 2 | -3.10 | 12.00 |
| B | West-Coastal | B40a | B-Vu2 | Vungu | 1 | -2.10 | 10.70 |
| B | West-Coastal | B41 | B-Bwa | Bwali | 2 | -1.40 | 10.60 |
| B | West-Coastal | B41 | B-Shi | Shira | 0 | -1.50 | 10.50 |
| B | West-Coastal | B42a | B-Sg2 | Sango | 1 | -1.20 | 12.50 |
| B | West-Coastal | B43 | B-Pu1 | Punu | 1 | -2.50 | 11.00 |
| B | West-Coastal | B44 | B-Lu1 | Lumbu | 0 | -2.80 | 10.00 |
| B | West-Coastal | B45 | B-Bwi | Bwisi | 1 | -3.25 | 13.10 |
| B | West-Coastal | B46 | B-Bar | Barama | 1 | -2.60 | 9.80 |
| B | West-Coastal | B51 | B-Dum | Duma | 5 | -0.80 | 12.60 |
| B | West-Coastal | B51 | B-Wan | Wanji | 2 | -0.90 | 12.80 |
| B | West-Coastal | B52 | B-Nze | Nzebi | 1 | -1.90 | 11.90 |
| B | West-Coastal | B53 | B-Tsa | Tsangi | 2 | -2.90 | 13.70 |
| B | West-Coastal | B61 | B-Mbe | Mbede | 1 | -0.40 | 14.00 |
| B | West-Coastal | B62 | B-Bam | Bamba | 0 | -1.20 | 13.90 |
| B | West-Coastal | B62 | B-Mpi | Mpini | 3 | -1.20 | 14.50 |
| B | West-Coastal | B63 | B-Kuy | Kuya | 10 | -1.50 | 13.40 |
| B | West-Coastal | B63 | B-Ndu | Ndumu | 11 | -1.60 | 13.50 |
| B | West-Coastal | B64 | B-Ngw | Ngwi | 2 | -4.30 | 19.40 |
| B | West-Coastal | B65a | B-Ng2 | Ngoli | 1 | -4.15 | 19.10 |
| B | West-Coastal | B66 | B-Kan | Kaningi | 5 | -1.75 | 13.60 |
| B | West-Coastal | B70 | B-Nun | Nunu | 3 | -3.10 | 16.90 |
| B | West-Coastal | B70 | B-Tgu | Tegue | 3 | -1.10 | 15.20 |
| B | West-Coastal | B70 | B-Tek | Teke | 7 | -4.10 | 17.20 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| B | West-Coastal | B71a | B-Tkk | Teke-Kale | 2 | -1.30 | 15.40 |
| B | West-Coastal | B72a | B-Gun | Gungwel | 4 | -1.90 | 15.80 |
| B | West-Coastal | B73 | B-Teg | Tege-W | 3 | -3.30 | 13.20 |
| B | West-Coastal | B73A | B-Ts2 | Tsayi | 2 | -2.80 | 12.60 |
| B | West-Coastal | B73B | B-Lal | Lali | 2 | -3.70 | 13.30 |
| B | West-Coastal | B73c | B-Iya | Iyaa | 3 | -3.80 | 13.30 |
| B | West-Coastal | B74b | B-Bo | Bo | 0 | -2.50 | 15.60 |
| B | West-Coastal | B74C | B-Ti2 | Tio | 3 | -3.70 | 14.50 |
| B | West-Coastal | B75 | B-Ite | Iteke | 3 | -3.60 | 16.65 |
| B | West-Coastal | B77A | B-Ku2 | Kukuya | 0 | -2.30 | 14.50 |
| B | West-Coastal | B77b | B-Fum | Fumu | 0 | -4.20 | 15.30 |
| B | West-Coastal | B78 | B-Iwu | Iwum | 1 | -4.30 | 17.05 |
| B | West-Coastal | B80 | B-Gon | Gongo | 2 | -5.70 | 18.20 |
| B | West-Coastal | B80 | B-Kem | Kempee | 2 | -2.30 | 18.00 |
| B | West-Coastal | B80 | B-Kes | Kesaa | 2 | -3.00 | 17.50 |
| B | West-Coastal | B80 | B-Mi2 | Mpiin | 4 | -5.00 | 18.70 |
| B | West-Coastal | B80a | B-Zad | Zadi | 2 | -3.80 | 19.10 |
| B | West-Coastal | B81a | B-Te2 | Tiene | 1 | -2.70 | 17.70 |
| B | West-Coastal | B83 | B-Ful | Fulunga | 3 | -4.20 | 15.80 |
| B | West-Coastal | B83a | B-Fin | Finu | 4 | -5.20 | 14.80 |
| B | West-Coastal | B84a | B-Mpu | Mpuono | 3 | -3.95 | 16.90 |
| B | West-Coastal | B85b | B-Iyn | Iyanz | 2 | -4.70 | 18.70 |
| B | West-Coastal | B85B | B-Kiy | Kiyey | 6 | -3.70 | 17.95 |
| B | West-Coastal | B85d | B-Nso | Nsongo | 0 | -4.90 | 19.85 |
| B | West-Coastal | B85e | B-Imp | Impur | 0 | -4.30 | 18.80 |
| B | West-Coastal | B85E | B-Mp2 | Mput | 2 | -4.30 | 18.90 |
| B | West-Coastal | B85f | B-Tm2 | Tsambaan | 3 | -4.50 | 18.90 |
| B | West-Coastal | B86 | B-Dzi | Dzing | 2 | -4.30 | 19.30 |
| B | West-Coastal | B87 | B-Mb1 | Mbuun | 1 | -4.90 | 19.80 |
| C | West-Congo Basin | C00 | C-Bam | Bambomba | 8 | 1.30 | 17.30 |
| C | West-Congo Basin | C00 | C-Bwa | Bwamba | 0 | 3.60 | 18.00 |
| C | West-Congo Basin | C00 | C-Lik | Likau | 5 | 2.40 | 18.90 |
| C | West-Congo Basin | C00 | C-Mok | Mokiba | 2 | 0.70 | 14.50 |
| C | West-Congo Basin | C00a | C-Ba2 | Babole | 3 | 1.40 | 18.00 |
| C | West-Congo Basin | C01 | C-Gan | Gando | 7 | 3.60 | 17.50 |
| C | West-Congo Basin | C02 | C-Kot | Kota | 1 | -0.70 | 12.00 |
| C | West-Congo Basin | C10 | C-Iso | Isongo | 7 | 4.00 | 17.80 |
| C | West-Congo Basin | C10 | C-Lke | Leke | 4 | 1.60 | 17.20 |
| C | West-Congo Basin | C11 | C-Kod | Kondi | 12 | 3.90 | 16.20 |
| C | West-Congo Basin | C12 | C-Pan | Pande | 7 | 3.50 | 16.20 |
| C | West-Congo Basin | C12b | C-Gon | Gongo | 7 | 3.20 | 16.00 |
| C | West-Congo Basin | C15 | C-Bon | Bongili | 0 | 0.30 | 15.50 |
| C | West-Congo Basin | C16 | C-Lob | Lobala | 4 | 2.70 | 18.50 |
| C | West-Congo Basin | C22 | C-Aka | Aka | 7 | 4.30 | 18.50 |
| C | West-Congo Basin | C24 | C-Koy | Koyo | 0 | -0.50 | 15.90 |
| C | West-Congo Basin | C25 | C-Bo1 | Bosi | 0 | -1.10 | 15.40 |
| C | West-Congo Basin | C27 | C-Lkb | Likuba | 2 | -1.70 | 16.40 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| C | West-Congo Basin | C30 | C-Bab | Babale | 3 | -2.10 | 17.15 |
| C | West-Congo Basin | C30 | C-Blk | Boloki | 1 | 0.90 | 18.70 |
| C | West-Congo Basin | C30 | C-Gya | Gyando | 1 | 1.80 | 18.20 |
| C | West-Congo Basin | C30 | C-Lkk | Likoka | 2 | 2.70 | 18.70 |
| C | West-Congo Basin | C30 | C-Mot | Motembo | 1 | 2.10 | 21.10 |
| C | West-Congo Basin | C30 | C-Ndo | Ndobo | 0 | 1.00 | 18.70 |
| C | West-Congo Basin | C30 | C-Pam | Pama | 2 | -1.05 | 17.20 |
| C | West-Congo Basin | C30a | C-Mo1 | Moya | 2 | -2.50 | 16.20 |
| C | West-Congo Basin | C31 | C-Dok | Doko | 4 | 1.60 | 19.70 |
| C | West-Congo Basin | C31a | C-Lo1 | Loi | 1 | 0.50 | 17.90 |
| C | West-Congo Basin | C31e | C-Zam | Zamba | 2 | 1.40 | 18.20 |
| C | West-Congo Basin | C31f | C-Mbz | Mbonzo | 0 | 1.80 | 18.00 |
| C | West-Congo Basin | C31r | C-Lib | Libinza | 0 | 1.50 | 18.50 |
| C | West-Congo Basin | C31s | C-Bal | Balobo | 1 | 1.90 | 19.40 |
| C | West-Congo Basin | C31y | C-Mwe | Mwe | 8 | 2.10 | 19.20 |
| C | West-Congo Basin | C32 | C-Bob | Bobangi | 0 | -1.40 | 17.70 |
| C | West-Congo Basin | C32 | C-Kin | Kinunu | 4 | -1.90 | 16.50 |
| C | West-Congo Basin | C32 | C-Mbm | Mbompo | 2 | -0.10 | 18.10 |
| C | West-Congo Basin | C33 | C-Sen | Sengele | 1 | -1.80 | 18.70 |
| C | West-Congo Basin | C34 | C-Sak | Sakata | 3 | -2.70 | 18.10 |
| C | West-Congo Basin | C34a | C-Kba | Kibai | 6 | -2.90 | 18.60 |
| C | West-Congo Basin | C34B | C-Kes | Kesha | 2 | -3.05 | 18.70 |
| C | West-Congo Basin | C35 | C-Bol | Bolia | 1 | -1.50 | 18.40 |
| C | West-Congo Basin | C35 | C-Nt1 | Ntomba-B | 0 | -1.10 | 18.10 |
| C | West-Congo Basin | C36 | C-Lol | Lolia | 1 | -1.30 | 18.20 |
| C | West-Congo Basin | C36c | C-Bdz | Budza | 1 | 2.20 | 22.50 |
| C | West-Congo Basin | C36d | C-Lin | Lingala | 1 | 1.50 | 17.50 |
| C | West-Congo Basin | C36g | C-Lko | Leko | 2 | 0.80 | 18.60 |
| C | West-Congo Basin | C37 | C-Bdj | Budja | 4 | 2.20 | 23.20 |
| C | West-Congo Basin | C41 | C-Kul | Kula | 3 | 2.30 | 20.20 |
| C | West-Congo Basin | C41 | C-Ngo | Ngombe | 3 | 1.20 | 19.80 |
| C | Boan-Lebonya | C43 | C-Ben | Benge | 5 | 3.40 | 23.90 |
| C | Boan-Lebonya | C44a | C-Boa | Boa | 3 | 2.70 | 23.80 |
| C | Boan-Lebonya | C45 | C-Leb | Lebeo | 6 | 2.50 | 27.00 |
| C | Boan-Lebonya | C45 | C-Nge | Ngelema | 1 | 1.50 | 25.00 |
| C | West-Congo Basin | C50 | C-Ela | Elambo | 1 | -1.00 | 23.75 |
| C | West-Congo Basin | C51 | C-Mbe | Mbesa | 3 | 1.10 | 23.15 |
| C | West-Congo Basin | C52 | C-Hes | Heso | 3 | 0.70 | 24.20 |
| C | West-Congo Basin | C52 | C-Lon | Lolinga | 1 | 1.05 | 23.85 |
| C | West-Congo Basin | C52 | C-Lyo | Lyombo | 2 | 0.50 | 24.30 |
| C | West-Congo Basin | C53 | C-Bau | Bauma | 6 | 0.00 | 24.00 |
| C | West-Congo Basin | C53 | C-Wen | Wenge | 7 | 0.10 | 24.10 |
| C | West-Congo Basin | C54 | C-Ilo | Ilombo | 1 | 1.00 | 24.30 |
| C | West-Congo Basin | C55 | C-Kle | Kele | 1 | 0.80 | 24.20 |
| C | West-Congo Basin | C61 | C-Kun | Kundo | 0 | -1.00 | 19.60 |
| C | West-Congo Basin | C61 | C-Mng | Mongo | 1 | -0.60 | 24.30 |
| C | West-Congo Basin | C64 | C-Kon | Konda | 0 | -1.40 | 19.00 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| C | West-Congo Basin | C65a | C-Nt4 | Ntomba-I | 2 | -1.90 | 18.30 |
| C | West-Congo Basin | C68a | C-Ml2 | Mbole-K | 0 | -0.10 | 24.50 |
| C | West-Congo Basin | C69 | C-Omb | Ombo | 0 | -2.20 | 25.80 |
| C | West-Congo Basin | C71 | C-Tet | Tetela | 0 | -5.00 | 25.00 |
| C | West-Congo Basin | C71a | C-Yyo | Yyondo | 0 | -4.60 | 24.20 |
| C | West-Congo Basin | C72a | C-Ku2 | Kushu | 0 | -4.00 | 25.10 |
| C | West-Congo Basin | C75 | C-Kla | Kela | 4 | -2.50 | 23.20 |
| C | West-Congo Basin | C80 | C-Ile | Ilebo | 6 | -4.30 | 20.60 |
| C | West-Congo Basin | C80 | C-Mbi | Mbingi | 6 | -4.05 | 22.20 |
| C | West-Congo Basin | C80 | C-Shu | Shuwa | 4 | -4.10 | 22.20 |
| C | West-Congo Basin | C81 | C-Nde | Ndengese | 6 | -3.50 | 21.40 |
| C | West-Congo Basin | C83 | C-Won | Wongo | 0 | -4.90 | 19.90 |
| C | West-Congo Basin | C83a | C-Lu1 | Luhilel | 0 | -4.35 | 20.20 |
| C | West-Congo Basin | C83B | C-Lor | Lori | 2 | -4.20 | 19.10 |
| C | West-Congo Basin | C90 | C-Cwa | Cwao | 6 | -5.00 | 22.00 |
| D | Boan-Lebonya | D00 | D-Kwa | Kwange | 0 | -3.90 | 25.90 |
| D | Boan-Lebonya | D00 | D-Lil | Liliko | 0 | 2.20 | 27.30 |
| D | Boan-Lebonya | D00 | D-Lum | Lumbwe | 6 | -6.50 | 28.20 |
| D | Boan-Lebonya | D00 | D-Nge | Ngengele | 1 | -2.80 | 25.30 |
| D | Boan-Lebonya | D00 | D-Son | Songe | 0 | -4.50 | 26.30 |
| D | Boan-Lebonya | D00 | D-Zur | Zura | 0 | -4.50 | 26.60 |
| D | Boan-Lebonya | D00a | D-La2 | Langa | 0 | -3.30 | 24.70 |
| D | Boan-Lebonya | D12 | D-Len | Lengola | 0 | -0.50 | 25.50 |
| D | East-Central | D13 | D-Mit | Mituku | 18 | -0.20 | 24.80 |
| D | East-Central | D14 | D-Eny | Enya | 6 | 0.50 | 25.50 |
| D | East-Central | D14 | D-Ge1 | Genia | 0 | -4.10 | 26.00 |
| D | East-Central | D24 | D-Bin | Binja-N | 0 | -2.70 | 26.20 |
| D | East-Central | D24 | D-Sng | Songola | 1 | -2.30 | 26.50 |
| D | East-Central | D25 | D-Le1 | Lega | 0 | -2.70 | 27.30 |
| D | East-Central | D25 | D-Reg | Rega | 0 | -2.60 | 27.35 |
| D | East-Central | D26 | D-Zim | Zimba | 0 | -4.30 | 26.60 |
| D | East-Central | D28 | D-Hol | Holoholo | 1 | -5.90 | 29.20 |
| D | East-Central | D28b | D-Tum | Tumbwe | 5 | -6.00 | 28.80 |
| D | Boan-Lebonya | D32 | D-Bi1 | Bira | 0 | 1.60 | 30.20 |
| D | Boan-Lebonya | D33 | D-Ny1 | Nyali | 0 | 2.00 | 27.90 |
| D | Boan-Lebonya | D35 | D-Bod | Bodo | 0 | 2.20 | 27.90 |
| D | Boan-Lebonya | D37 | D-Kum | Kumu | 0 | -0.50 | 27.00 |
| D | Boan-Lebonya | D43 | D-Nyg | Nyanga | 0 | -1.40 | 28.00 |
| D | East-Great Lakes | D55 | D-Buy | Buyu | 17 | -4.50 | 28.30 |
| E | East-Great Lakes | E41 | E-Log | Logoli | 3 | 0.10 | 34.90 |
| E | East-Kilimanjaro | E51 | E-Gik | Gikuyu | 0 | -0.50 | 36.70 |
| E | East-Kilimanjaro | E54 | E-Tha | Tharaka | 0 | -0.20 | 38.20 |
| E | East-Kilimanjaro | E55 | E-Kam | Kamba | 6 | -1.40 | 38.00 |
| E | East-Kilimanjaro | E62b | E-Cha | Chaga | 2 | -3.20 | 37.30 |
| E | East-Coastal | E71 | E-Po1 | Pokomo | 10 | -1.40 | 39.90 |
| E | East-Coastal | E72a | E-Gi1 | Giryama | 0 | -3.20 | 39.40 |
| E | East-Coastal | E72c | E-Cho | Choni | 9 | -3.80 | 39.80 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| E | East-Coastal | E72e | E-Rab | Rabai | 10 | -3.30 | 39.80 |
| E | East-Coastal | E73 | E-Di1 | Digo | 9 | -4.30 | 39.20 |
| E | East-Coastal | E73 | E-Seg | Segeju | 11 | -4.80 | 39.20 |
| E | East-Coastal | E74a | E-Daw | Dawida | 10 | -3.10 | 38.00 |
| E | East-Coastal | E74b | E-Sag | Sagala | 11 | -6.90 | 36.30 |
| F | East-Coastal | F21 | F-Suk | Sukuma | 18 | -3.00 | 33.00 |
| F | East-Coastal | F22 | F-Nya | Nyamwezi | 0 | -4.30 | 33.20 |
| F | East-Coastal | F23 | F-Sum | Sumbwa | 0 | -3.90 | 32.80 |
| F | East-Coastal | F31 | F-Nyi | Nyiramba | 10 | -5.00 | 34.70 |
| G | East-Coastal | G11 | G-Gog | Gogo | 10 | -6.30 | 35.70 |
| G | East-Coastal | G12 | G-Kag | Kaguru | 10 | -6.30 | 37.00 |
| G | East-Coastal | G21 | G-Tav | Taveta | 12 | -3.40 | 37.80 |
| G | East-Coastal | G22 | G-Cas | Casu | 6 | -4.20 | 37.80 |
| G | East-Coastal | G22 | G-Par | Pare | 1 | -4.50 | 38.00 |
| G | East-Coastal | G23 | G-Sha | Shambala | 9 | -4.90 | 38.30 |
| G | East-Coastal | G24 | G-Bon | Bondei | 9 | -5.30 | 38.80 |
| G | East-Coastal | G31 | G-Zig | Zigula | 9 | -5.40 | 38.00 |
| G | East-Coastal | G32 | G-Ngw | Ngwele | 10 | -6.40 | 38.30 |
| G | East-Coastal | G33 | G-Doe | Doe | 9 | -6.30 | 38.60 |
| G | East-Coastal | G33 | G-Zar | Zaramo | 10 | -7.00 | 39.00 |
| G | East-Coastal | G34 | G-Ngu | Ngulu | 9 | -6.20 | 37.50 |
| G | East-Coastal | G35 | G-Lug | Luguru | 10 | -6.90 | 37.70 |
| G | East-Coastal | G36 | G-Kam | Kami | 11 | -7.40 | 39.10 |
| G | East-Coastal | G37 | G-Kut | Kutu | 10 | -7.50 | 37.50 |
| G | East-Coastal | G40 | G-Elw | Elwana | 10 | -1.10 | 40.10 |
| G | East-Coastal | G40 | G-Had | Hadimu | 10 | -6.30 | 39.50 |
| G | East-Coastal | G40 | G-Mii | Miini | 0 | 1.20 | 44.00 |
| G | East-Coastal | G40 | G-Pem | Pemba | 10 | -5.20 | 39.30 |
| G | East-Coastal | G40 | G-Vum | Vumba | 12 | -4.60 | 39.40 |
| G | East-Coastal | G41 | G-Tik | Tikuu | 10 | -2.00 | 40.90 |
| G | East-Coastal | G42 | G-Sw1 | Swahili | 0 | -6.80 | 39.30 |
| G | East-Coastal | G44a | G-Nga | Ngazija | 4 | -11.50 | 43.50 |
| G | East-Coastal | G44b | G-Mao | Maore | 1 | -12.80 | 44.20 |
| G | East-Coastal | G44b | G-Nju | Njuani | 10 | -12.10 | 44.30 |
| G | East-Coastal | G44c | G-Mwl | Mwali | 0 | -12.20 | 40.50 |
| G | East-Coastal | G65 | G-Kin | Kinga | 1 | -9.00 | 34.00 |
| H | West-Coastal | H11a | H-Be2 | Bembe | 0 | -3.90 | 13.55 |
| H | West-Coastal | H12 | H-Vi1 | Vili | 3 | -4.60 | 12.70 |
| H | West-Coastal | H12b | H-Su3 | Suundi | 1 | -6.30 | 18.10 |
| H | West-Coastal | H12b | H-Yo1 | Yombe | 0 | -5.00 | 13.00 |
| H | West-Coastal | H12B | H-Yo2 | Yombe | 1 | -4.10 | 12.90 |
| H | West-Coastal | H13a | H-Kun | Kunyi | 0 | -4.10 | 13.05 |
| H | West-Coastal | H13b | H-Su4 | Suundi | 0 | -4.30 | 12.90 |
| H | West-Coastal | H16 | H-Snd | Sonde | 8 | -6.10 | 18.20 |
| H | West-Coastal | H16B | H-Man | Manyanga | 1 | -5.20 | 14.00 |
| H | West-Coastal | H16b | H-Mbo | Mboma | 2 | -5.70 | 14.00 |
| H | West-Coastal | H16f | H-Laa | Laadi | 4 | -4.30 | 15.00 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| H | West-Coastal | H16f | H-Lar | Lari | 2 | -4.20 | 14.80 |
| H | West-Coastal | H16g | H-Nta | Ntandu | 0 | -5.00 | 15.20 |
| H | West-Coastal | H16h | H-Sik | Sikongo | 0 | -6.20 | 14.20 |
| H | West-Coastal | H16i | H-Su5 | Suundi | 1 | -4.80 | 14.70 |
| H | West-Coastal | H16j | H-Han | Hangala | 0 | -4.30 | 14.30 |
| H | West-Coastal | H16k | H-Zoo | Zoombo | 1 | -6.70 | 15.10 |
| H | West-Coastal | H17b | H-Kam | Kamba | 4 | -4.15 | 13.50 |
| H | West-Coastal | H21 | H-Kim | Kimbundu | 0 | -8.75 | 15.51 |
| H | West-Coastal | H24 | H-Sng | Songo | 6 | -4.50 | 18.20 |
| H | West-Coastal | H31 | H-Ya1 | Yaka | 0 | -4.90 | 17.15 |
| H | West-Coastal | H32 | H-Suk | Suku | 1 | -5.30 | 17.60 |
| H | West-Coastal | H33 | H-Hu1 | Hungana | 1 | -4.50 | 18.50 |
| J | East-Great Lakes | J00 | J-Reg | Regi | 4 | -1.90 | 33.10 |
| J | East-Great Lakes | J00 | J-Yaa | Yaaka | 0 | -1.80 | 29.30 |
| J | East-Great Lakes | J11 | J-Nyo | Nyoro | 0 | 1.10 | 30.00 |
| J | East-Great Lakes | J13 | J-Him | Hima | 0 | -0.60 | 30.50 |
| J | East-Great Lakes | J15 | J-Gan | Ganda | 1 | 0.50 | 33.70 |
| J | East-Great Lakes | J16 | J-Sog | Soga | 3 | 0.60 | 33.90 |
| J | East-Great Lakes | J17 | J-Gwe | Gwere | 6 | 1.20 | 34.00 |
| J | East-Great Lakes | J21 | J-Tal | Talinga | 4 | 0.70 | 29.80 |
| J | East-Great Lakes | J22 | J-Hay | Haya | 6 | -1.80 | 31.40 |
| J | East-Great Lakes | J22 | J-Zib | Ziba | 3 | -1.10 | 31.40 |
| J | East-Great Lakes | J23 | J-Zin | Zinza | 3 | -2.70 | 31.40 |
| J | East-Great Lakes | J24 | J-Ker | Kerebe | 5 | -2.10 | 33.50 |
| J | East-Great Lakes | J25 | J-Jit | Jita | 4 | -1.80 | 33.00 |
| J | East-Great Lakes | J25 | J-Kwa | Kwaya | 4 | -1.80 | 33.50 |
| J | East-Great Lakes | J30 | J-Kis | Kisa | 17 | 0.30 | 34.80 |
| J | East-Great Lakes | J30 | J-Mar | Maraba | 12 | 0.25 | 34.50 |
| J | East-Great Lakes | J31c | J-Buk | Bukusu | 5 | 0.60 | 34.70 |
| J | East-Great Lakes | J32 | J-Isu | Isuka | 6 | 0.30 | 34.80 |
| J | East-Great Lakes | J34 | J-Sam | Samia | 6 | 0.30 | 34.10 |
| J | East-Great Lakes | J42 | J-Nan | Nande | 0 | 0.20 | 29.20 |
| J | East-Great Lakes | J42 | J-Shu | Shu | 0 | 0.15 | 29.40 |
| J | East-Great Lakes | J42 | J-Swa | Swaga | 2 | 0.00 | 29.00 |
| J | East-Great Lakes | J42 | J-Yir | Yira | 0 | 0.10 | 29.30 |
| J | East-Great Lakes | J51 | J-Hun | Hunde | 12 | -1.50 | 28.50 |
| J | East-Great Lakes | J52 | J-Hav | Havu | 1 | -2.10 | 29.10 |
| J | East-Great Lakes | J53 | J-Shi | Shi | 1 | -2.50 | 28.80 |
| J | East-Great Lakes | J54 | J-Ful | Fuliru | 1 | -3.10 | 29.00 |
| J | East-Great Lakes | J55 | J-Vir | Vira | 1 | -3.40 | 29.10 |
| J | East-Great Lakes | J57 | J-Tem | Tembo | 0 | -1.70 | 28.00 |
| J | East-Great Lakes | J61 | J-Kig | Kigoyi | 3 | -1.70 | 29.30 |
| J | East-Great Lakes | J61 | J-Nig | Nigi | 4 | -1.40 | 29.60 |
| J | East-Great Lakes | J61 | J-Rer | Rera | 0 | -1.70 | 29.60 |
| J | East-Great Lakes | J61 | J-Rwa | Rwanda | 0 | -2.00 | 30.00 |
| J | East-Great Lakes | J62 | J-Run | Rundi | 0 | -3.00 | 30.00 |
| J | East-Great Lakes | J67 | J-Vin | Vinza | 4 | -5.20 | 31.00 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| K | West-Southwest | K01 | K-Hol | Holu | 5 | -7.70 | 17.50 |
| K | West-Southwest | K11 | K-Cio | Ciokwe | 0 | -9.00 | 20.00 |
| K | West-Southwest | K14 | K-Luv | Luvale | 0 | -13.10 | 22.10 |
| K | West-Southwest | K14 | K-Lwe | Lwena | 0 | -12.40 | 22.40 |
| K | West-Southwest | K15 | K-Mbu | Mbunda | 0 | -11.80 | 19.90 |
| K | West-Southwest | K19 | K-Gan | Gangela | 1 | -14.10 | 21.40 |
| K | West-Southwest | K21 | K-Sal | Salampasu | 2 | -7.50 | 22.50 |
| K | West-Southwest | K22 | K-Nde | Ndembu | 5 | -10.70 | 22.40 |
| K | West-Southwest | K23 | K-Lun | Lunda | 1 | -12.00 | 24.20 |
| K | West-Southwest | K23 | K-Ruu | Ruund | 9 | -8.60 | 23.00 |
| K | West-Southwest | K31 | K-Luy | Luyi | 1 | -16.20 | 22.00 |
| K | West-Southwest | K33 | K-Kwa | Kwangali | 0 | -17.90 | 19.70 |
| K | West-Southwest | K39 | K-Dci | Dciriku | 0 | -17.90 | 20.80 |
| K | West-Southwest | K51B | K-Mb4 | Mbala | 0 | -4.50 | 17.90 |
| K | West-Southwest | K52 | K-Phe | Pheende | 0 | -6.00 | 19.00 |
| K | West-Southwest | K53 | K-Kwe | Kwezo | 1 | -5.60 | 18.40 |
| L | East-Central | L00 | L-Bwi | Bwile | 7 | -8.00 | 29.00 |
| L | East-Central | L00 | L-Keb | Kebwe | 4 | -6.50 | 26.70 |
| L | East-Central | L00 | L-Yaz | Yazi | 5 | -5.30 | 26.80 |
| L | East-Central | L21 | L-Kei | Kete-I | 3 | -7.60 | 23.10 |
| L | West-Southwest | L22a | L-Mba | Mbagani | 0 | -6.70 | 21.70 |
| L | West-Southwest | L22b | L-Lw1 | Lwalwa | 2 | -7.20 | 21.80 |
| L | East-Central | L23 | L-Son | Songye | 1 | -6.15 | 24.40 |
| L | East-Central | L27a | L-Ba2 | Bangubangu | 4 | -4.60 | 26.80 |
| L | East-Central | L31 | L-LuK | Luba-Ks | 2 | -6.50 | 23.50 |
| L | East-Central | L32 | L-Kan | Kanyok | 2 | -7.10 | 23.40 |
| L | East-Central | L33 | L-LuS | Luba-Sh | 1 | -7.80 | 27.00 |
| L | East-Central | L34 | L-He1 | Hemba | 0 | -5.70 | 26.70 |
| L | East-Central | L34 | L-He2 | Hemba-m | 1 | -5.50 | 26.80 |
| L | East-Central | L35 | L-San | Sanga | 0 | -9.60 | 26.80 |
| L | East-Central | L41 | L-Kao | Kaonde | 1 | -12.20 | 26.80 |
| L | East-Central | L62 | L-Nko | Nkoya | 1 | -14.80 | 24.80 |
| M | East-Central | M15 | M-Mam | Mambwe | 2 | -9.20 | 31.80 |
| M | East-Coastal | M31 | M-Kon | Konde | 0 | -8.95 | 33.55 |
| M | East-Coastal | M31 | M-Ny1 | Nyakyusa | 0 | -9.70 | 33.90 |
| M | East-Central | M40 | M-Tem | Temba | 1 | -11.10 | 27.30 |
| M | East-Central | M41 | M-Ta1 | Taabwa | 0 | -7.80 | 30.00 |
| M | East-Central | M42 | M-Bem | Bemba | 1 | -12.60 | 28.20 |
| M | East-Central | M43 | M-Aus | Aushi | 1 | -11.20 | 26.80 |
| M | East-Central | M50 | M-Ngu | Ngumbo | 0 | -10.80 | 29.70 |
| M | East-Central | M52 | M-Lal | Lala | 0 | -13.30 | 30.20 |
| M | East-Central | M54 | M-Lam | Lamba | 0 | -12.90 | 28.60 |
| M | East-Central | M61 | M-Len | Lenje | 0 | -15.00 | 27.50 |
| M | East-Central | M62 | M-Sol | Soli | 1 | -15.40 | 28.30 |
| M | East-Central | M64a | M-To2 | Tonga | 0 | -16.80 | 27.20 |
| N | East-Coastal | N14 | N-Mpo | Mpoto | 2 | -10.70 | 35.60 |
| N | East-Coastal | N21a | N-Tu2 | Tumbuka | 1 | -11.60 | 33.60 |

| Guthrie | Hist-Groups | G-code | L-code | Language | missing | Long | Lat |
|---|---|---|---|---|---|---|---|
| N | East-Coastal | N25 | N-Suk | Sukwa | 0 | -9.60 | 33.20 |
| N | East-Coastal | N26 | N-Lam | Lambya | 0 | -9.70 | 33.60 |
| N | East-Coastal | N31 | N-Nys | Nyasa | 7 | -11.20 | 35.80 |
| N | East-Coastal | N31a | N-Nyj | Nyanja | 0 | -12.00 | 33.50 |
| N | East-Coastal | N31b | N-Cew | Cewa | 1 | -14.00 | 33.80 |
| N | East-Coastal | N41 | N-Sng | Senga | 0 | -11.30 | 32.90 |
| N | East-Coastal | N42 | N-Kun | Kunda | 0 | -15.50 | 30.00 |
| N | East-Coastal | N44 | N-Sen | Sena | 2 | -17.00 | 35.20 |
| P | East-Coastal | P21 | P-Ya1 | Yao | 0 | -15.70 | 35.50 |
| P | East-Southeast | P31c | P-Ma4 | Makua | 0 | -14.00 | 39.50 |
| P | East-Southeast | P32 | P-Lom | Lomwe | 0 | -16.50 | 36.50 |
| R | West-Southwest | R11 | R-Um1 | Umbundu | 0 | -11.30 | 16.20 |
| R | West-Southwest | R14 | R-Nku | Nkumbi | 13 | -16.50 | 14.60 |
| R | West-Southwest | R21 | R-Kwa | Kwanyama | 0 | -17.00 | 16.80 |
| R | West-Southwest | R22 | R-Ndo | Ndonga | 0 | -18.00 | 16.80 |
| R | West-Southwest | R31 | R-Her | Herero | 0 | -17.00 | 12.00 |
| S | East-South | S00 | S-Mba | Mbalangwe | 0 | -17.80 | 23.00 |
| S | East-South | S10 | S-Sho | Shona | 0 | -19.00 | 30.00 |
| S | East-South | S11 | S-Kor | Korekore | 4 | -17.00 | 30.00 |
| S | East-South | S12 | S-Zez | Zezuru | 0 | -17.80 | 31.10 |
| S | East-South | S13 | S-Man | Manyika | 0 | -18.50 | 31.10 |
| S | East-South | S14 | S-Kar | Karanga | 0 | -20.10 | 30.80 |
| S | East-South | S15 | S-Nda | Ndau | 0 | -19.80 | 34.80 |
| S | East-South | S16 | S-Kal | Kalanga | 1 | -20.45 | 27.80 |
| S | East-South | S21 | S-Ven | Venda | 1 | -22.20 | 30.00 |
| S | East-South | S31 | S-Twn | Tswana | 0 | -25.50 | 25.90 |
| S | East-South | S31d | S-Kha | Khalaxadi | 4 | -13.70 | 22.00 |
| S | East-South | S32 | S-Sot | Sotho-N | 0 | -30.00 | 28.00 |
| S | East-South | S34 | S-Loz | Lozi | 0 | -17.90 | 26.00 |
| S | East-South | S41 | S-Xho | Xhosa | 0 | -32.00 | 27.00 |
| S | East-South | S42 | S-Zul | Zulu | 0 | -28.00 | 32.00 |
| S | East-South | S43 | S-Swa | Swati | 0 | -26.50 | 31.50 |
| S | East-South | S44 | S-Nde | Ndebele | 0 | -19.50 | 29.00 |
| S | East-South | S45 | S-Ngo | Ngoni | 3 | -11.65 | 33.65 |
| S | East-South | S51 | S-Tsw | Tswa | 6 | -23.80 | 35.00 |
| S | East-South | S52 | S-Nku | Nkuna | 6 | -24.10 | 30.30 |
| S | East-South | S53 | S-Tso | Tsonga | 0 | -23.00 | 33.50 |
| S | East-South | S54 | S-Ron | Ronga | 6 | -26.00 | 33.00 |
| S | East-South | S54a | S-Kon | Konde | 6 | -27.00 | 32.60 |
| S | East-South | S54b | S-Lul | Luleke | 9 | -22.40 | 31.20 |

Table S5: Details for the populations with linguistic data. "G-code" is the group's code, "L-code" the language's code, "$N_A$" the number of individuals for autosomal data, $N_{mt}$ the number of individuals for mtDNA, "$N_y$" the number of individuals for Y-chromosome, and "Ref" the reference study.

| **Population** | **Language** | **G-code** | **L-code** | $N_A$ | $N_{mt}$ | $N_y$ | **Ref.** |
|---|---|---|---|---|---|---|---|
| Akele | Ungom | B22b | B-Ung | - | 48 | 50 | [20, 21] |
| Ateke | Teke | B70 | B-Tek | - | 54 | 48 | [20, 21] |
| Bafia | Kpa | A53 | A-Kpa | 60 | - | - | [2] |
| Baluba | Luba-Ks | L31 | L-LuK | 12 | - | - | [2] |
| Banen | Nen | A44 | A-Nen | 50 | - | - | [2] |
| Barega | Rega | D25 | D-Reg | 8 | - | - | [2] |
| Batanga | Tanga | A32c | A-Tan | 40 | - | - | [2] |
| Batie | Bandjoun | 900b | Band | 32 | - | - | [2] |
| Bemba | Bemba | M42 | M-Bem | - | - | 10 | [1] |
| Benga | Benga | A34 | A-Be1 | - | 50 | 48 | [20, 21] |
| Bubi | Bubi | A31 | A-Bub | - | 45 | - | [32] |
| Bulu | Bulu | A74a | A-Bul | 44 | - | - | [2] |
| Duma | Duma | B51 | B-Dum | - | 47 | 46 | [20, 21] |
| Eshira | Shira | B41 | B-Shi | - | 40 | 42 | [20, 21] |
| Eton | Eton | A71 | A-Eto | 8 | - | - | [2] |
| Ewondo | Ewondo | A72a | A-Ewo | 6 | 25 | - | [2, 20] |
| Fang | Fang | A75 | A-Fan | 38 | 105 | 60 | [2, 20, 21] |
| Galoa | Galwa | B11C | B-Ga2 | - | 51 | 47 | [20] |
| Ganguela | Gangela | K19 | K-Gan | - | 20 | 11 | [22] |
| Gogo | Gogo | G11 | G-Gog | 26 | - | - | [2] |
| Kalanga | Kalanga | S16 | S-Kal | - | - | 20 | [1] |
| Kikuyu | Gikuyu | E51 | E-Gik | 44 | 25 | - | [2, 30] |
| Kongo | Sikongo | H16h | H-Sik | 34 | - | - | [2] |
| Kota | Kota | B25b | B-Ko3 | - | 56 | 53 | [20, 21] |
| Kunda | Kunda | N42 | N-Kun | - | 36 | 36 | [31] |
| Kuvale | Herero | R31 | R-Her | - | 54 | 25 | [22] |
| Lala-Bisa | Lala | M52 | M-Lal | - | 42 | 38 | [1, 31] |
| Lomwe | Lomwe | P32 | P-Lom | - | 20 | - | [33] |
| Lozi | Lozi | S34 | S-Loz | - | - | 94 | [1] |
| Luhya | Kisa | J30 | J-Kis | 34 | - | - | [2] |
| Luvale | Luvale | K14 | K-Luv | - | - | 16 | [1] |
| Luyana | Luyi | K31 | K-Luy | - | - | 61 | [1] |
| Mabea | Kwejo | A80 | A-Kw1 | 26 | - | - | [2] |
| Makhuwa | Makwa | P31c | P-Ma4 | - | 20 | - | [33] |
| Mbala | Mbala | K51B | K-Mb4 | - | - | 12 | [1] |
| Mbunda | Mbunda | K15 | K-Mbu | - | - | 49 | [1] |
| Mbundu | Kimbundu | H21 | H-Kim | - | 44 | - | [14] |
| Mbuun | Mbuun | B87 | B-Mb1 | - | - | 12 | [1] |
| Mitsongo | Tsogo | B31 | B-Tsg | - | 64 | 60 | [20, 21] |
| Mvae | Kako | A93 | A-Ka2 | 48 | - | - | [2] |
| Ndao | Ndau | S15 | S-Nda | - | 19 | - | [33] |
| Ndumu | Ndumu | B63 | B-Ndu | - | 39 | 36 | [20, 21] |

| **Population** | **Language** | **G-code** | **L-code** | $N_A$ | $N_{mt}$ | $N_y$ | **Ref.** |
|---|---|---|---|---|---|---|---|
| Nguni | Ngoni | S45 | S-Ngo | - | 12 | - | [33] |
| Nkoya | Nkoya | L62 | L-Nko | - | - | 29 | [1] |
| Ntumu | Ntum | A75 | A-Ntu | 22 | - | - | [2] |
| Nyaneka-Nkhumbi | Nkumbi | R14 | R-Nku | - | 153 | 74 | [22] |
| Nyanja | Nyanja | N31a | N-Nyj | - | 20 | - | [33] |
| Nzebi | Nzebi | B52 | B-Nze | - | 63 | 57 | [20, 21] |
| Obamba | Bamba | B62 | B-Bam | - | 47 | 47 | [20, 21] |
| Pare | Pare | G22 | G-Par | 46 | - | - | [2] |
| Pende | Pheende | K52 | K-Phe | - | - | 11 | [1] |
| Punu | Punu | B43 | B-Pu1 | - | 52 | 58 | [20, 21] |
| Ronga | Ronga | S54 | S-Ron | - | 21 | - | [33] |
| Sambaa | Shambala | G23 | G-Sha | 36 | - | - | [2] |
| Sena | Sena | N44 | N-Sen | - | 21 | - | [33] |
| Shake | Sake | B25 | B-Sa1 | - | 51 | 43 | [20, 21] |
| Shangaan | Tsonga | S53 | S-Tso | - | 22 | - | [33] |
| Shona | Shona | S10 | S-Sho | - | 18 | - | [33] |
| Sukuma | Sukuma | F21 | F-Suk | 20 | 32 | 30 | [2, 28, 34] |
| Tonga | Tonga | M64a | M-To2 | - | - | 28 | [1] |
| Tswa | Tswa | S51 | S-Tsw | - | 19 | - | [33] |
| Tswana | Tswana | S31 | S-Twn | - | - | 20 | [1] |
| Tutsi-Hutu | Rwanda | J61 | J-Rwa | 16 | - | - | [2] |
| Ovimbundu | Umbundu | R11 | R-Um1 | - | 92 | 94 | [22] |
| Venda | Venda | S21 | S-Ven | 26 | - | - | [2] |
| Xhosa | Xhosa | S41 | S-Xho | 56 | - | - | [2] |
| Yambassa | Gunu | A62a | A-Gun | 34 | - | - | [2] |
| Yansi | Yanz | B85b | B-Iyn | - | - | 23 | [1] |
| Yao | Yao | P21 | P-Ya1 | - | 10 | - | [33] |

Table S6: Correlations among models of Bantu migration using all (Mantel test Z) and only 'informative' (Spearman rank-correlation test $\rho$) pairwise comparisons. Note that Z is for matrix-like and $\rho$ is for vector-like variables, respectively. All $P < 0.001$.

| Model 1 | Model 2 | Z | $\rho$ 'informative' |
|---|---|---|---|
| IBD | late-split | 0.668 | 0.471 |
| IBD | early-split | 0.660 | 0.230 |
| late-split | early-split | 0.781 | 0.314 |

Table S7: Correlations (Mantel tests Z) between lexical distances (i.e. the consensus tree) and models of Bantu migration using only languages for which genetic data are available. All $P < 0.001$.

| genetic marker | migration model | | |
|---|---|---|---|
| | IBD | late-split | early-split |
| mtDNA | 0.702 | 0.647 | 0.551 |
| Y chromosome | 0.531 | 0.505 | 0.392 |
| Autosomes | 0.782 | 0.764 | 0.806 |

Table S8: Correlations (Mantel tests Z with uncorrected *P*s in parentheses) of mtDNA and Y-chromosomal distances with the models of migration using only 21 populations that have data for both markers.

| genetic distance | migration model | | |
|---|---|---|---|
| | IBD | late-split | early-split |
| mtDNA $\Phi_{ST}$ | 0.697 (0.001) | 0.656 (0.001) | 0.581 (0.001) |
| Y-chrom $F_{ST}$ | 0.164 (0.094) | 0.257 (0.010) | 0.073 (0.248) |
| Y-chrom $R_{ST}$ | 0.108 (0.158) | 0.147 (0.078) | 0.052 (0.287) |

Table S9: List of the 184 autosomal STRs [2] that minimize the missing data across Bantu populations.

X4PTEL04, AAAT126, AAT071, AAT200, AAT203, AAT226, AAT238, AAT247, AAT257, AAT262, AATA019, AATA045, ACT3F12, AGAT060, AGAT113Z, AGAT116P, AGAT117, AGAT127, AGAT140P, AGAT142P, ATA063, ATA080M, ATA11D10M, ATA12D05P, ATA16D09, ATA19H08, ATA20G07M, ATA22D02, ATA24A08, ATA24F10, ATA26D07, ATA27A06P, ATA27D04P, ATA27F01, ATA28B11, ATA29E07M, ATA3A07, ATA47D07, ATA4E02, ATA55A05, ATA57D10M, ATA5A09N, ATA67B07P, ATA73A08M, ATA7D07, ATA82B02N, ATAA009, ATAA018P, ATAG042, ATAG053P, ATAG089P, ATC033, ATC3D09, ATC6A06M, ATCT018, ATCT035, ATGT006Z, CTAT016, GAAT2C03P, GATA060, GATA115E01N, GATA117D01N, GATA11C12, GATA11H10, GATA129F05, GATA12H10, GATA134B03, GATA137A12M, GATA143C02, GATA149B10M, GATA151F03N, GATA152F05L, GATA163G03, GATA168F06, GATA173A03, GATA178G09M, GATA185C06Z, GATA194A05M, GATA194B06P, GATA196C10P, GATA197B10P, GATA22F09P, GATA22G05M, GATA22H02, GATA23A02, GATA25A04, GATA25C10M, GATA26D02M, GATA27E01, GATA28F03, GATA29A01, GATA29B11, GATA30A03N, GATA30E06P, GATA30G01, GATA31H11P, GATA32C12, GATA41A01, GATA41G07M, GATA42H02P, GATA43A04, GATA43H03N, GATA44F10P, GATA48E02, GATA4A10, GATA50C03N, GATA50D10, GATA50G06, GATA51A07P, GATA51B02M, GATA51F04P, GATA52A12, GATA63D12P, GATA63F08P, GATA63G01, GATA64B04P, GATA64F05, GATA65E01, GATA68A03, GATA68F07, GATA6A05, GATA6E05, GATA6F05P, GATA6F06, GATA70B08, GATA70E11, GATA70F12M, GATA72G09N, GATA73B08M, GATA73D01, GATA73E11, GATA7D12, GATA7G10, GATA81B01, GATA81D12M, GATA85A04M, GATA85D10, GATA86C08P, GATA87D11, GATA87E02N, GATA88F09, GATA89A11, GATA89G08z, GATA8A05, GATA8G10M, GATA8H05, GATA90E02, GATA91H06M, GATA92B06P, GGAA10F06M, GGAA17G05P, GGAA21G11L, GGAA22C05, GGAA23C05N, GGAA2A03, GGAA2f11N, GGAA3G05, GGAA4A12, GGAT1A4, GGAT3G09M, Mfd238, MFD442-GTTT002, PAH, SE30, SraP, TAAA014P, TACA003, TAGA049, TAT024Z, TATC012, TATC046, TATG002P, TCAT006ZP, TCTA017M, TCTA025, TTA015P, TTAT023Z, UT1772, UT2095M, UT5029, UT6540, UT7129L, UT7136, UT7544.

# References


1 de Filippo, C., Barbieri, C., Whitten, M., Mpoloka, S. W., Gunnarsdóttir, E. D., Bostoen, K., Nyambe, T., Beyer, K., Schreiber, H., de Knijff, P. et al. 2011 Y-chromosomal variation in Sub-Saharan Africa: Insights into the history of Niger-Congo groups. *Mol. Biol. Evol.* **28**, 1255–1269. (doi:10.1093/molbev/msq312)

2 Tishkoff, S. A, Reed, F. A, Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J. B., Awomoyi, A. A., Bodo, J.-M, Doumbo, O. et al. 2009 The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044. (doi:10.1126/science.1172257)

3 Pemberton, T. J., Sandefur, C. I., Jakobsson, M. & Rosenberg, N. A. 2009 Sequence determinants of human microsatellite variability. *BMC Genomics* **10**, 612. (doi:10.1186/1471-2164-10-612)

4 Michalakis, Y. & Excoffier, L. 1996 A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**, 1061–1064.

5 Weir, B. S. & Cockerham, C. C. 1984 Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358–1370.

6 Bastin, Y., Coupez, A. & Mann, M. 1999 *Continuity and divergence in the Bantu languages from a lexicostatistic study.* Tervuren: Annales, Sciences humaines. Royal Museum for Central Africa.

7 Bowern, C. 2008 *Linguistic fieldwork: a practical guide.* Hampshire: Palgrave Macmillan.

8 Chatelain, H. 1888-1890 Die Grundzüge des Kimbundu oder der Angola-Sprache. *Zeitschrift für Afrikanische Sprachen* **2**, 265–314.

9 Chatelain, H. 1964 *Grammatica elementar do Kimbundu ou lingua de Angola.* New Jersey: The Gregg Press Incorporated

10 Arvanites, L. 1976 Kimbundu tones: Tone pattern in two contexts. In *Studies in Bantu tonology* (eds L. Hyman), pp 131–140. Los Angeles: Department of Linguistic, University of Southern California.

11 Maia, A. d. S. 1964 Dicionário rudimentar português-kimbundu, lingua nativa de Luanda a Malange, Angola. Cucujães: Editorial Missões.

12 Maia, A. d. S. 1994 Dicionário Complementar Português-Kimbundu-Kikongo (Linguas nativas do centro e norte de Angola). Cucujães: Editorial Missões.

13 Vansina, J. 2001 Portuguese vs Kimbundu: Language Use in the Colony of Angola (1575-c. 1845). *Bulletin des Séances de l'Académie royale des Sciences d'Outre-Mer* **47**, 267–281.

14 Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, A. & Comas, D. 2004 Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum. Genet.* **115**, 439–447. (doi:10.1007/s00439-004-1164-0)

15 M. Pagel and A. Meade. 2010 A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* **53**, 571–581

16 ***BayesTrees V1.1*** http://www.evolution.reading.ac.uk/BayesTrees.html

17 Sukumaran, J. & Holder, M. T. 2010 Dendropy: A python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571.

18 Guthrie, M. 1971 Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages, Bantu Prehistory, Inventory and Indexes. London: Gregg International.

19 Maho, J. F. 2003 A classification of the Bantu languages: an update of Guthrie's referential system. In *The Bantu languages* (eds D. Nurse & G. Philippson), pp. 639–651. London: Routledge language family series 4.

20 Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., Sica, L., Mouguiama-Daouda, P., Comas, D., Tzur, S. et al. 2008 Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc. Natl. Acad. Sci. USA* **105**, 1596–601. (doi:10.1073/pnas.0711467105)

21 Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., van der Veen, L., Hombert, J.-M., Quintana-Murci, L. & Comas, D. 2009 Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol. Biol. Evol.* **26**, 1581–1589. (doi:10.1093/molbev/msp069)

22 Coelho M., Sequeira F., Luiselli D., Beleza S. & Rocha J. (2009) On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol. Biol.* **9**, 80. (doi:10.1186/1471-2148-9-80)

23 Lewis M. 2009 *Ethnologue: Languages of the World*. Dallas: SIL International.

24 Heine, B., Hoff, H. & Vossen, R. 1977 Neuere Ergebnisse zur Territorialgeschichte der Bantu. In *Zur Sprachgeschichte und Ethnohistorie in Afrika* (eds. W. J. Möhlig, F. Rottland & B. Heine), pp. 52–72. Dietrich Reimer: Berlin.

25 Möhlig, W. J. 1981 Stratification in the history of the Bantu languages. *Sprach. Gesch. Afr.* **3**, 251–317.

26 Vansina, J. (1995) New linguistic evidence and the Bantu expansion. *J Afr Hist* **36**, 173–195.

27 Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H. J. 1998 Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and Sub-Saharan populations. *Ann. Hum. Genet.* **62**, 531–550. (doi:10.1046/j.1469-1809.1998.6260531.x)

28 Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T. B., Ramakrishnan, U., Reed, F. A. & Mountain, J. L. 2007 History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.* **24**, 2180–2195. (doi:10.1093/molbev/msm155)

29 Krings, M., Salem A. E., Bauer K., Geisert H., Malek A. K., Chaix L., Simon C., Welsby D., Di Rienzo A., Utermann G. et al. 1999 mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am. J. Hum. Genet.* **64**, 1166–1176. (doi:doi:10.1086/302314)

30 Watson, E., Bauer, K., Aman, R., Weiss, G., von Haeseler, A. & Pääbo, S. 1996 mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.* **59**, 437–444.

31 de Filippo, C., Heyn, P., Barham, L., Stoneking, M. & Pakendorf, B. 2010 Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. *Am. J. Phys. Anthropol.* **141**, 382–394. (doi:10.1002/ajpa.21155)

32 Mateu, E., Comas, D., Calafell, F., Pérez-Lezaun, A., Abade A. & Bertranpetit, J. 1997 A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann. Hum. Genet.* **61**, 507–518. (doi:10.1046/j.1469-1809.1997.6160507.x)

33 Salas, A., Richards, M., De la Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V. & Carracedo, A. 2002 The making of the African mtDNA landscape. *Am. J. Hum. Genet.* **71**, 1082–1111. (doi:10.1086/344348)

34 Knight, A., Underhill, P. A., Mortensen, H. M., Zhivotovsky, L. A., Lin, A.A., Henn, B. M., Louis, D., Ruhlen, M. & Mountain, J. L. 2003 African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol.* **13**, 464–473.

35 Chen, Y. S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. 2000 mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am. J. Hum. Genet.* **66**, 1362–1383. (doi:10.1086/302848)

36 Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. 1991 African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507.

37 Veeramah, K. R., Connell, B. A., Pour, N. A., Powell, A., Plaster, C. A., Zeitlyn, D., Mendell, N. R., Weale, M. E., Bradman, N. & Thomas, M. G. 2010 Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evol. Biol.* **10**, 92. (doi:10.1186/1471-2148-10-92)

38 Destro-Bisol, G., Coia, V., Boschi, I., Verginelli, F., Cagliá, A., Pascali, V., Spedini, G. & Calafell, F. 2004 The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *Am. Nat.* **163**, 212–226. (doi:10.1086/381405)

39 Gomes, V., Sánchez-Diz, P., Amorim, A., Carracedo, A. & Gusmão, L. 2010 Digging deeper into East African human Y chromosome lineages. *Hum. Genet.* **127**, 603–13. (doi:10.1007/s00439-010-0808-5)