

# The linguistic profile and prehistory of Eastern Africa

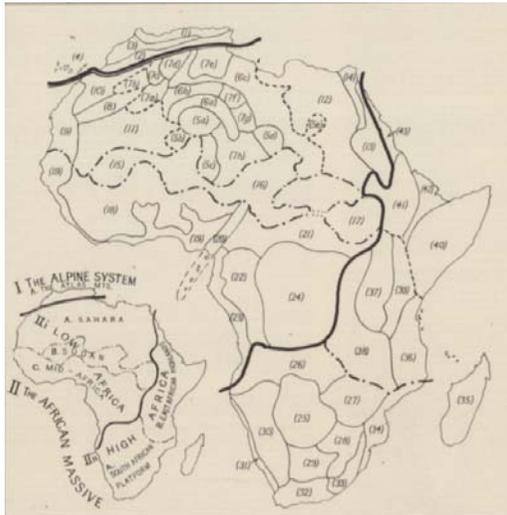
Tom Güldemann

Humboldt University Berlin and Max Planck Institute for the Science of Human History Jena

## 1 Introduction

+ "Eastern Africa" variably defined - here by geography (Map 1) and geopolitics (Map 2)

> "High Africa" (= Eastern + Southern Africa) as the domain of interest



Map 1: Physiography of Africa after Lobeck (1946)

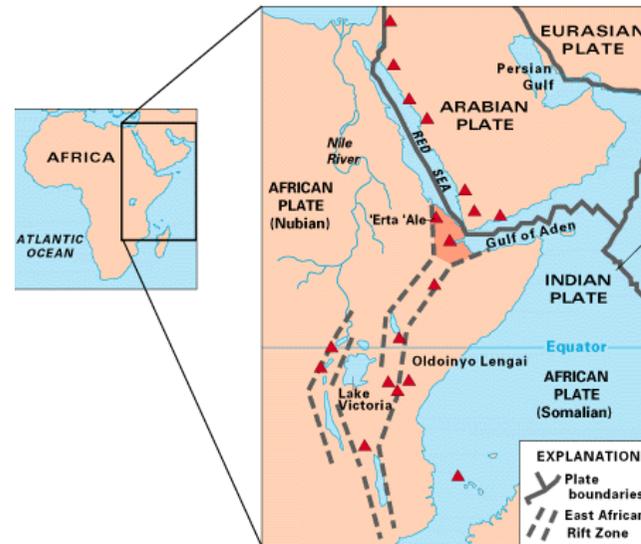


Map 2: Global geographical subregions according to UN

+ important geographical preliminaries

a) coast line and physical geography steer population movement

> assumed major impact of East African Rift zone and its landmarks (Map 3)

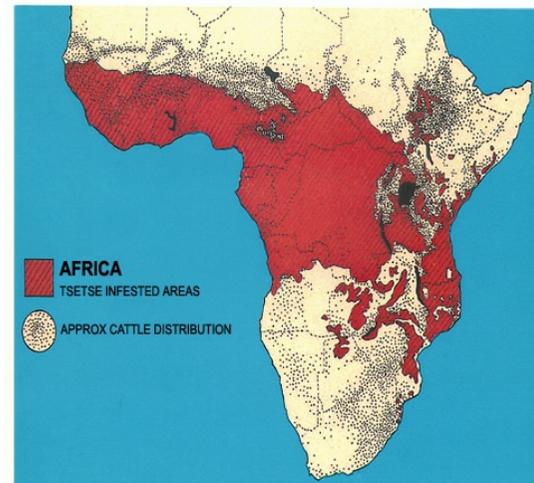


Map 3: East African Rift zone

b) environmental constraints for human settlement and individual subsistence modes

- large areas that are still today not or very thinly inhabited by humans

- disease-infested areas as major barrier for pastoralism, longitudinal corridors (Map 4)



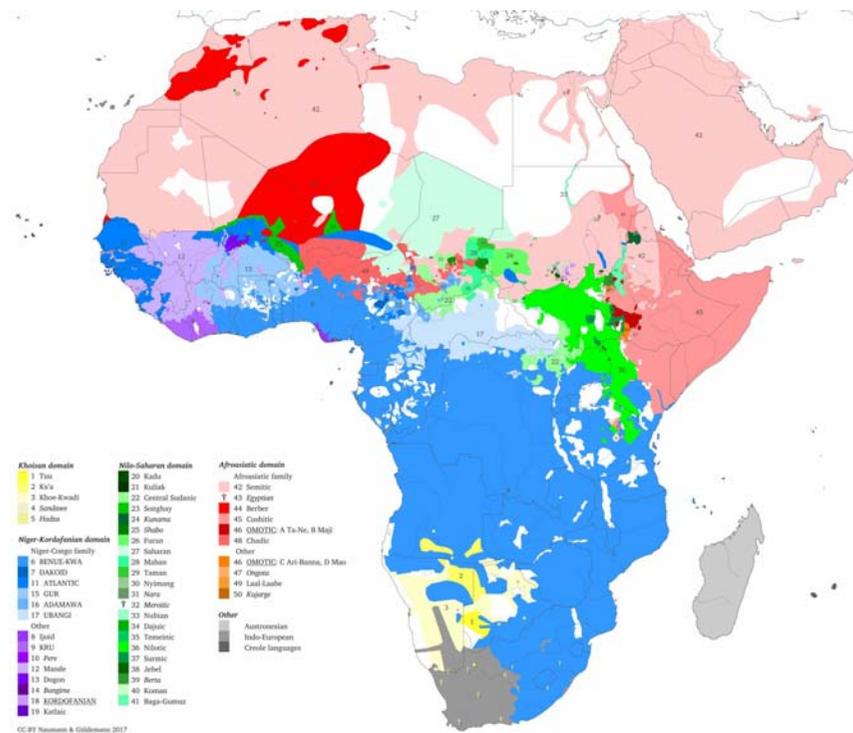
Map 4: Tsetse fly areas in Africa

## 2 Modern linguistic profile

### 2.1 Language classification in Africa

#### 2.1.1 Genealogical relationships

- + Greenberg's (1963) classification with just four African super-"families" widely accepted but methodologically and empirically not robust (cf., e.g., Campbell and Poser 2008)
- review of the state of language classification in Africa by Güldemann (forth. b), oriented to standard criteria of the general discipline (cf., e.g., Nichols 1996, Campbell 2003)
- > Africa is genealogically far more diverse - Map 5 and Appendix 1:
  - a) 2 geographically and demographically large families: Niger-Congo and Afroasiatic occupying over 2/3 of the continent and representing 80% of its languages
  - b) 3 intermediate families: Central Sudanic, Nilotic-Surmic, Mande
  - c) 35+ small units (including more than ten singletons) without convincing affiliation

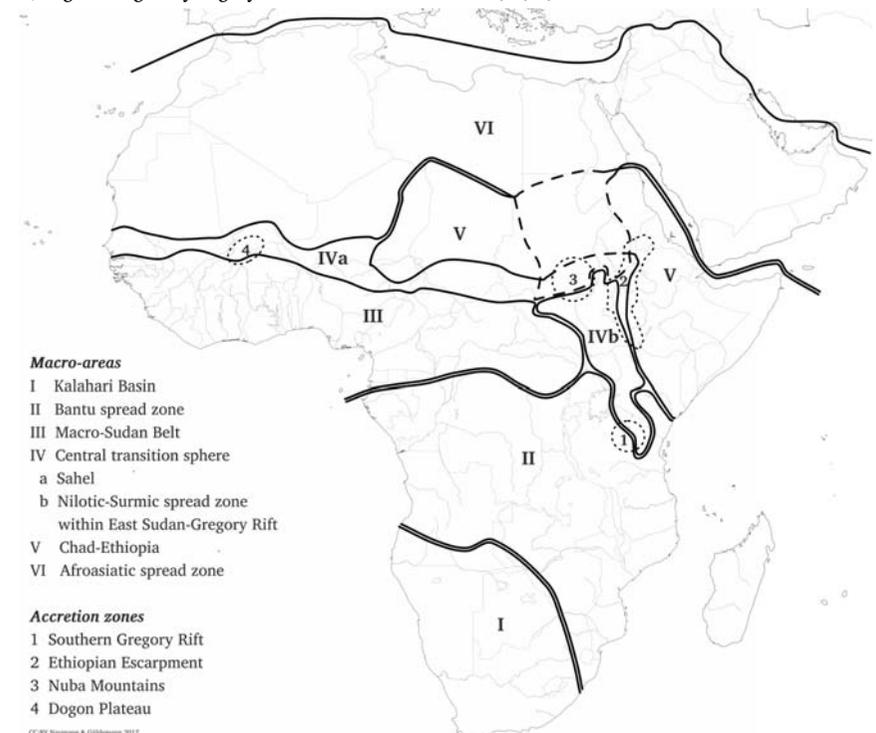


Notes: GENEALOGICAL/AREAL POOL; *Single language (complex)*

Map 5: Genealogical classification in Africa according to Güldemann (forth. b)

#### 2.1.2 Macro-areal linguistics of Afrabia

- + first more comprehensive research by Greenberg (1959, 1983) and Heine (1975, 1976)
- + resumed by Güldemann (e.g., 1998, 1999, 2001, 2003, 2005, 2008b, 2010, 2011, 2017; with Fehn 2017) and Clements and Rialland (2008) independently but with considerably similar results regarding macro-areal feature aggregations in Africa before more recent large-scale colonization from outside
- + most recent update by Güldemann (forth. a) - Map 6
- > external separation as a continent-sized unit "Afrabia" comprising Africa and Arabian Peninsula: behaves as a unit with recurrent cross-migrations of major population impact, bounded on land by Southwest Asian transition zone (Haig 2017, Haig and Khan forth.)
- > internal macro-areal partition of Afrabia into:
  - a) 2 genealogically homogeneous spread zones: II, VI - coherence due to **inheritance**
  - b) 3 genealogically diverse macro-areas: I, III, V - coherence due to **contact**
  - c) 1 non-coherent transition sphere: IV (2 subareas a and b) - separates III from V and VI
  - d) 4 genealogically highly diverse accretion zones: 1, 2, 3, 4 all within IV



Map 6: Macro-areas and accretion zones of Afrabia (Güldemann forth. a)

## 2.2 Linguistic subareas and lineages of High Africa

+ High Africa partakes in 4 of 6 macro-areas and hosts 23 of 50 basic classificatory units

Macro-area	Core classificatory unit(s)	Peripheral classificatory units
I Kalahari Basin	Tuu (U1)/ Kx'a (U2)/ Khoe-Kwadi (U3)	<b>Niger-Congo:</b> Bantu of BENUE-KWA (U6)
II Bantu spread zone	<b>Niger-Congo:</b> Bantu of BENUE-KWA (U6)	-
IVb East Sudan-Gregory Rift	<b>Nilotic-Surmic:</b> Nilotic (U36), Surmic (U37)	Jebel (U38)/ <i>Berta</i> (U39)/ Koman (U40)/ Baga (U41)/ Kuliak (U21)/ <i>Hadza</i> (U5)/ <i>Sandawe</i> (U4)/ <b>Afroasiatic:</b> Cushitic (U45)
V Chad-Ethiopia	( <b>Wadi Howar</b> ): <i>Nara</i> (U31)/ <i>Kunama</i> (U24)/ <b>Afroasiatic:</b> Ethiosemitic of Semitic (U42), Cushitic (U45), Ta-Ne (U46.A) + Maji (U46.B) of <b>OMOTIC</b> / Ari-Banna (U46.C) of <b>OMOTIC</b>	? <i>Shabo</i> (U25)/ ? <i>Mao</i> (U46.D) of <b>OMOTIC</b> / ? <i>Ongota</i> (U47)/ <b>Nilotic-Surmic:</b> Surmic (U37)

Notes: GENEALOGICAL POOL; AREAL POOL; *single-language unit*; (possible) family above basic units; / separates independent units

**Table 1: Classificatory units and macro-areas of High Africa (Güldemann forth. a)**

+ drastic regional differences in terms of linguistic diversity

Area	Lineages	Diversity
Kalahari Basin	Bantu/ Khoe-Kwadi/ Kx'a/ Tuu	moderate
Southeast Africa	Bantu	low
East Africa	Bantu/ Cushitic/ Nilotic/ Sandawe/ Hadza	moderate
Horn of Africa	Ethiosemitic, Cushitic, Ta-Ne (all Afroasiatic)	moderate~low
Ethiopian Escarpment	<i>Nara</i> / <i>Kunama</i> / <i>Jebel</i> / <i>Berta</i> / <i>Koman</i> / <i>Baga</i> / <i>Shabo</i> / <i>Mao</i> / <i>Maji</i> / <i>Surmic</i> / <i>Ongota</i> / <i>Ari-Banna</i>	high

**Table 2: Linguistic regions and lineage inventory of High Africa**

## 3 Linguistic prehistory of High Africa

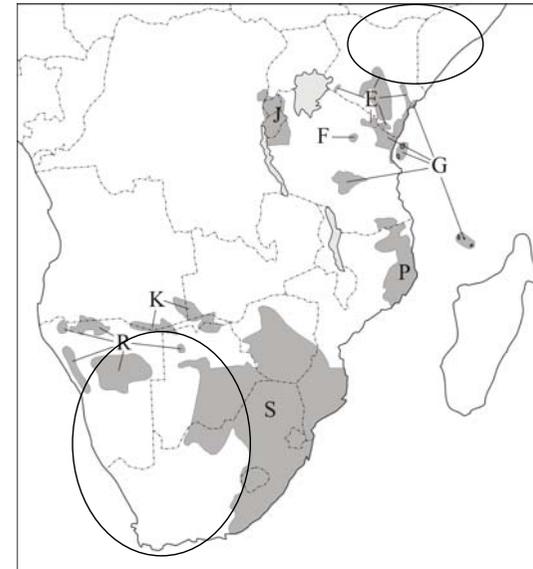
+ linguistic dynamics can be studied by means of:

- non-trivial areal feature distributions crossing genealogical boundaries > §3.1
- language contact phenomena, particularly shift-induced substrates > §3.2
- historical trajectories of linguistic lineages > §3.3

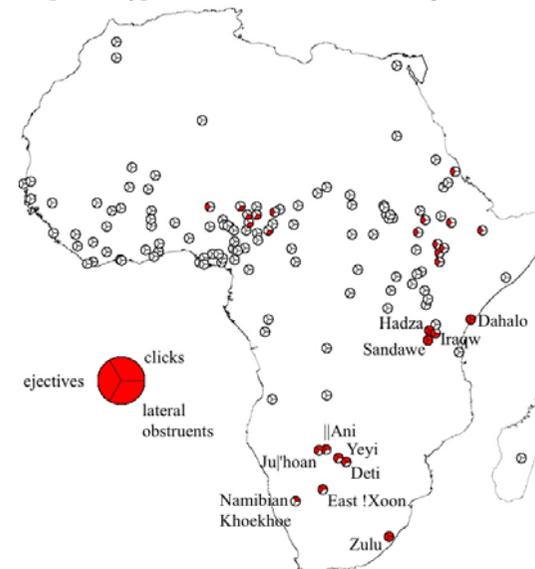
+ beware the time depth: so far largely restricted to (second half of) **Holocene!!!**

## 3.1 Long-distance linguistic affinities across High Africa

+ two linguistic feature distributions diagnostic for some earlier coherence of High Africa



Note: Circled areas = areas of Non-Bantu lineages with head-final noun marking  
**Map 7: Untypical head-final noun marking in Bantu (Güldemann 1999)**



**Map 8: Globally and continentally rare obstruent consonants (Güldemann 2010)**

> Naumann (2016):

- confirms and extends phonological affinities across High Africa
- identifies Southern Africa (including both the Kalahari Basin and Southeast Africa settled today by Southern Bantu) as a weaker but larger phonological area

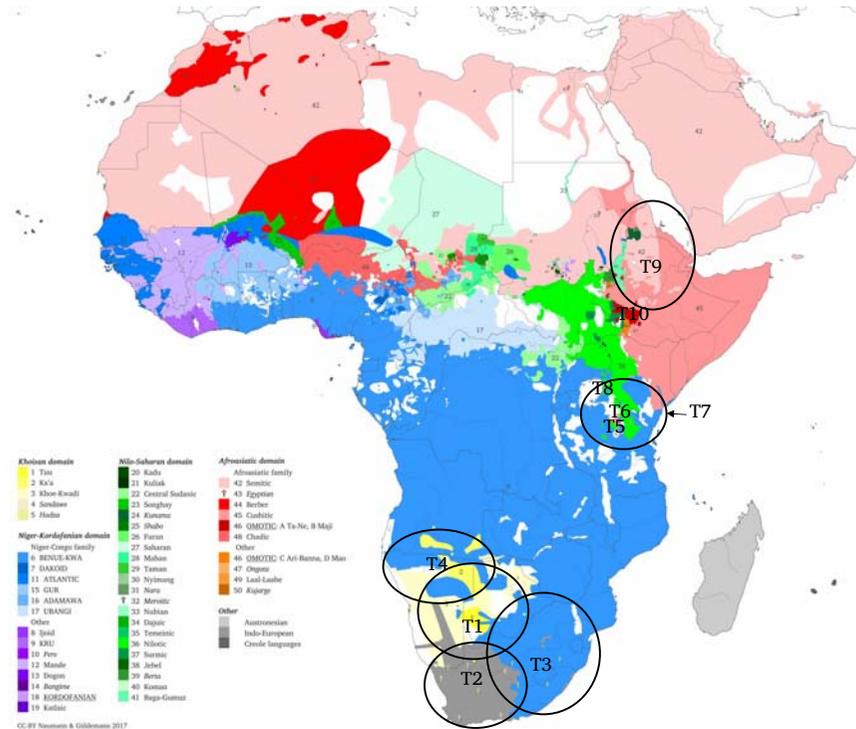
**3.2 Linguistic substrates**

+ substrate = layer of linguistic features in a language/lineage that is due to shift-induced interference from another defunct language/lineage (cf., e.g., Thomason and Kaufman 1988)

> identified in particular by differential linguistic-structural profiles within a lineage as a function of the areal context

- considerable amount of evidence for contact and substrate in all genealogically non-homogeneous areas of High Africa (cf. Table 2 above) - especially relevant for Khoe, Bantu, Nilotic, Surmic, Cushitic, Semitic - Map 9, Table 3

- linguistic arguments often supported by anthropological, genetic, historical, and/or archeological evidence



Notes: Tn = Key to Table 3

**Map 9: Linguistic substrate (candidates) across High Africa**

Resultant languages	Colonized area	Source area	Substrate	References
<b>Kalahari Basin</b>				
1 Khoe	(I)	(IVb)	Kx'a	Voßen (1992), Güldemann and Loughnane (2012), Elderkin (2014), Güldemann (forth.)
2 Khoekhoe	(I) Cape	(I) North	Tuu	Güldemann (2002, 2006, 2013a)
3 Bantu S	(I) East	(II)	Khoe-Kwadi, Tuu	Meinhof (1905), Engelbrecht (1925), Maingard (1933), Bourquin (1951a, b), Lanham (1962), Bill (1974), Louw (1974, 1976, 1977a, b, c, 1979, 1986, 2013), Lickey (1985), Argyle (1986), Herbert (1987, 1990a, b), Voßen (1997), Güldemann (1999), Pakendorf et al. (2017)
4 Bantu R, K	(I) North	(II)	Khoe-Kwadi, Kx'a	Engelbrecht (1925), Kubik (1984), Sommer and Voßen (1992), Voßen (1997), Güldemann (1999), Sommer (2013), Möhlig (2013), Barbieri et al. (2013), Gunnink et al. (2015), Pakendorf et al. (2017)
<b>East Africa</b>				
5 South Cushitic	(IVb)-1	(V)	? (Hadza, Sandawe)	Ten Raa (1969), Kießling (2002), Kießling and Mous (2003), Kießling, Mous, and Nurse (2008)
6 South and East Nilotic	(IV)b South	(IV)b North	(East) Cushitic	Heine, Rottland and Voßen (1979), Winter (1979), Heine and Vossen (1983), Brenzinger (1992)
7 Bantu E, F, G	(IV)b	(II)	Cushitic	Ehret and Nurse (1981), Nurse and Rottland (1991/2), Nurse (1988, 1994, 2000a, b), Mous (2003)
8 South and West Nilotic	(IV)b	(IVb)	Bantu	Adhiambo (1991), Nurse and Rottland (1991/92), Rottland and Okombo (1992), Dimmendaal (1995, 2001b), Reh (2000), Kuteva (2000), Wrigley (2001), Hieda (2011)
<b>Horn of Africa and Ethiopian Escarpment</b>				
9 Ethiosemitic	(V) Horn of Africa	(VI) Arabia	Cushitic	Leslau (1945, 1952), Palmer (1974), Raz (1989), Hudson (1994), Crass (2002), Meyer (2002, 2009), Crass and Meyer (2008, 2011), Lucas (2013: 419-423)
10 Southeast Surmic	(V) fringe	(IVb)-2	OMOTIC	Hieda (1991a, 1993, 1996), Dimmendaal (1998a, b)

Notes: GENEALOGICAL/AREAL POOL, Remnant languages, ?/(...) = speculative

**Table 3: Linguistic substrate (candidates) across High Africa**

### 3.3 Movement of linguistic populations

+ immigration into High Africa:

- Semitic from northeast (Arabian Peninsula)
- ?Nilotic-Surmic from northwest
- **Bantu from (north)west into East Africa, Southeast Africa and Kalahari Basin:**
  - > hypothesis about entrance from northeast of rainforest today largely abandoned
  - > incremental submergence of apparent earlier coherence across High Africa (different areal layers of replacement; see, e.g., Naumann 2016)

+ emigration out of High Africa:

- Nilotic westward into Macro-Sudan belt
- Bantu eastward into Indian Ocean rim islands

+ migration within High Africa:

a) southward:

- Khoe-Kwadi likely from East Africa via Southeast Africa into Kalahari Basin (responsible for first PRE-BANTU spread of pastoralism - Güldemann 2008a)
- Cushitic from Horn of Africa into East Africa
- Nilotic from Sudan into East Africa
- Early Savannah Bantu from East Africa via Southeast Africa into Kalahari Basin
- ?Interlacustrine Bantu from East Africa via Southeast Africa into Kalahari Basin (reflected in ethnographic “cattle complex” within Bantu Zone S)
- Swahili Bantu “Sprachbund” from East Africa to Southeast Africa

b) northward:

- ?Swahili Bantu “Sprachbund” from Southeast Africa to East Africa
- Nguni Bantu from Southeast Africa into East Africa

### 4 General implications for prehistoric modelling

+ prehistorical hypotheses should work with:

- a) linguistic entities realistic for the relevant time depth
- b) overall higher population diversity than previously assumed (some of which submerged and unknown today), involves multiple movements of different Bantu-speaking groups
- b) relative rather than absolute linguistic chronologies
- c) linguistic research that appears in canonical disciplinary and peer-reviewed outlets

- + linguistic findings attest to an overall strong north-south trajectory of population dynamics, as opposed to recurrent east-west signals elsewhere in Africa (cf. Güldemann 2010, forth. a; Güldemann and Hammarström forth.)
- this axis of dynamics presumably induced by physical geography and environment (cf. §1)

### Appendix 1: African language groups (Güldemann forth. b)

No.	Classificatory unit	Internal	External		
01	Tuu	A, C	Tuu-Kx'a:	South African	Khoisan (domain): D, F
02	Kx'a	B	D, F	Khoisan:	
03	Khoe-Kwadi	A, C	Khoe-Kwadi-	D, F	
04	<i>Sandawe</i>	n.a.	Sandawe: D, F		
05	<i>Hadza</i>	n.a.			
06.A	BANTOID	D	BENUE-KWA: D	Niger-Congo: A, C	Niger-Kordofanian (domain): D
06.B	CROSS-RIVER	D			
06.C	KAINJI-PLATOID	D			
06.D	Igboid	C, E			
06.E	Idomoid	C, E			
06.F	Nupoid	C, E			
06.G	Edoid	A, B			
06.H	Akpes	C, E			
06.I	<i>Ukaan</i>	n.a.			
06.J	<i>Oko</i>	n.a.			
06.K	Owon-Arigidi	C, E			
06.L	Ayere-Ahan	C			
06.M	Yoruboid	B			
06.N	Gbe	B			
06.O	GHANA-TOGO M.	D			
06.P	Potou-Akanic	B			
06.Q	Ga-Dangme	B			
06.R	LAGOON	D			
06.S	<i>Ega</i>	n.a.			
07	DAKOID	D	ATLANTIC: D, F		
11.A	(CORE) ATLANTIC	D			
11.B	Mel	A, B			
11.C	<i>Gola</i>	n.a.			
11.D	<i>Limba</i>	n.a.			
11.E	<i>Sua</i>	n.a.			
11.F	<i>Nalu</i>	n.a.	GUR: D, F		
11.G	Rio Nunez	C, E			
15.A	(Central) Gur	A, B			
15.B	Kulangoic	C			
15.C	<i>Miyobe</i>	n.a.			

15.D	Tiefo	C			
15.E	<i>Viemo</i>	n.a.			
15.F	Tusian	C			
15.G	Samuic	C			
15.H	Senufo	C, F			
16.A	Tula-Waja	C			
16.B	<i>Longuda</i>	n.a.			
16.C	Bena-Mboi	C			
16.D	Bikwin-Jen	C			
16.E	Samba-Duru	C			
16.F	Mumuyic	B			
16.G	Maya	C			
16.H	Kebi-Benue	C			
16.I	Kimic	C			
16.J	Buaic	A, C			
16.K	<i>Day</i>	n.a.			
16.L	<i>Baa~Kwa</i>	n.a.			
16.M	<i>Nyingwom~Kam</i>	n.a.			
16.N	<i>Fali</i>	n.a.			
17.A	Gbayaic	A, B			
17.B	Zandic	C, E			
17.C	Mbaic	A, B			
17.D	Mundu-Baka	A, B			
17.E	Ngbandic	C, E			
17.F	Bandaic	C, E			
17.G	NDOGOIC	D			
09.A	(Narrow) Kru	A, C			
10	<i>Pere</i>	n.a.			
13	Dogon	C, E			
14	<i>Bangime</i>	n.a.			
18.A	Heibanic	A, B			
18.B	Talodic	A, B			
18.C	<i>Lafofa</i>	n.a.			
18.D	Rashadic	C			
19	Katlaic	C			
08	Ijoid	A, B			
09.B	<i>Siamou</i>	n.a.			
12	Mande	C, E			

ADAMAWA: D

UBANGI: D

KORDOFANIAN:  
DNiger-  
Kordofanian  
(domain): D

20	Kadu	C, F			
21	Kuliak	B			
22	Central Sudanic	A, B			
23	Songhay	C			
24	<i>Kunama</i>	n.a.			
25	<i>Shabo</i>	n.a.			
26	Furan	C			
27	Saharan	A, C			
28	Maban	A, B			
29	Taman	A, B			
30	Nyimang	C			
31	<i>Nara</i>	n.a.			
32	<i>Meroitic</i>	n.a.			
33	Nubian	A, B			
34	Dajuic	A, B			
35	Temeinic	C			
36	Nilotic	A, B			
37	Surmic	A, B			
38	Jebel	(C), F			
39	<i>Berta</i>	C			
40	Koman	B			
41	Baga	C			
48	Chadic	A, B			
42	Semitic	A, B			
43	<i>Egyptian</i>	n.a.			
44	Berber	A, B			
45	Cushitic	A, C			
46.A	Ta-Ne	B, C			
46.B	Maji	B, C			
46.C	Ari-Banna	A, C			
46.D	Mao	C			
47	<i>Ongota</i>	n.a.			
49	Laal-Laabe	C			
50	<i>Kujarge</i>	n.a.			

Northern East  
Sudanic ~  
"Wadi Howar":  
D, E, F

East Sudanic: D

Nilotic-Surmic:  
C, FJebel-Berta:  
D, FKoman-Baga:  
D, F

Afroasiatic: A, C

OMOTIC: D, F

Nilo-  
Saharan  
(domain): DAfroasiatic  
domain: D

Notes: GENEALOGICAL/AREAL POOL; *Single language (complex)*; n.a. = not applicable;  
A = Reconstructed morpheme paradigms; B = Regularly reconstructed lexicon;  
C = Strong resemblances of bona fide reconstructibility; D = Scattered  
resemblances; E = Lexicostatistic calculations; F = Structural similarities.