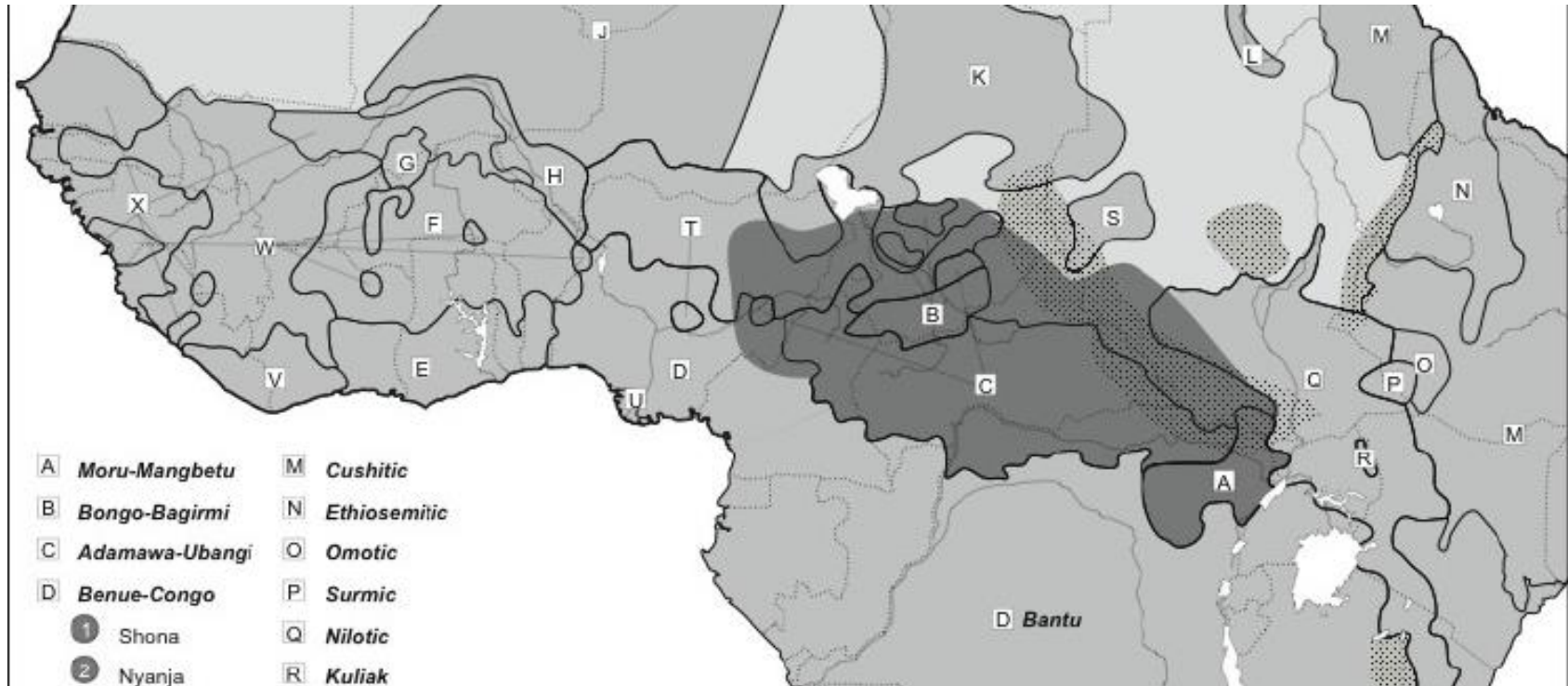George Starostin

(Russian State University for the Humanities / Russian Presidential Academy)

12th Annual Sergei Starostin Memorial Conference on Comparative-Historical Linguistics (RSUH, March 23-24, 2017)

On the issue of areal-genetic entanglement in the basic lexicon: the fate of 'moon' in the Macro-Sudanic region

A	Moru-Mangbetu	M	Cushitic
B	Bongo-Bagirmi	N	Ethiosemitic
C	Adamawa-Ubangi	O	Omotic
D	Benue-Congo	P	Surmic
❶	Shona	Q	Nilotic
❷	Nyanja	R	Kuliak

D Bantu

[*Partial map from Güldemann 2011*]

Basic methodological presumption

Roots X and Y in potentially related languages / language groups go back to a common ancestral state *if* and only *if*:

— their phonetic shapes are *corresponding* (form part of a pattern) or at least *compatible* (do not violate any established patterns);
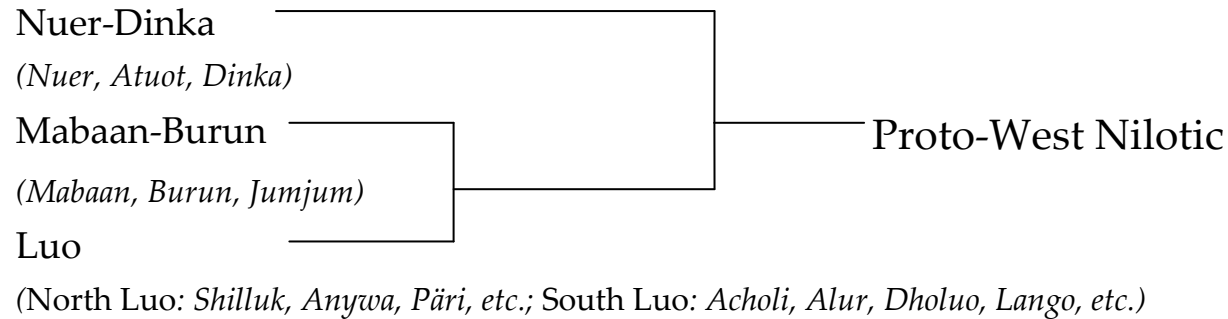
— their semantics are *identical* or follow a typologically reasonable path of semantic shifts, consistent with other etyma in the corpus;

— their topological distribution on the tree allows to reconstruct a realistic and parsimonious scenario of their evolution from original ancestral state through all the intermediate ancestral states;

— if any of these criteria are grossly violated, we deal with incidental similarities or areal diffusion (depending on specific circumstances).

1. ʻMoonʼ in West Nilotic languages.
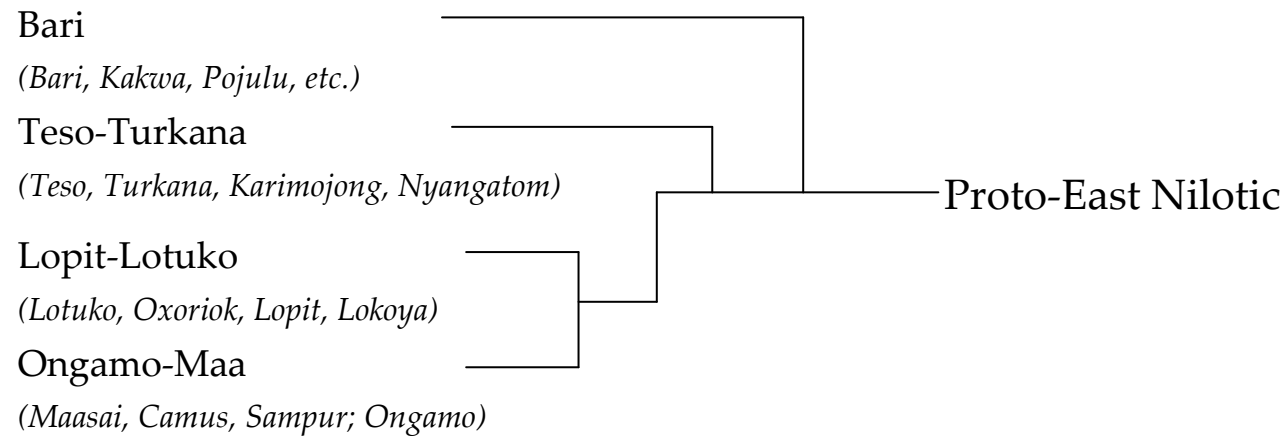
Composition of the family:

Nuer-Dinka
(Nuer, Atuot, Dinka)

Mabaan-Burun
(Mabaan, Burun, Jumjum)

Luo

Proto-West Nilotic

(North Luo: Shilluk, Anywa, Päri, etc.; South Luo: *Acholi, Alur, Dholuo, Lango, etc.)*

Forms:

Language	Form	
Nuer	pay	Cf. Proto-Luo: *dwʌy* ʻmoonʼ (no parallels in Nuer-Dinka or Mabaan-Buruun → topologically, an innovation)
Dinka	pɛːyʰ	
Mabaan	paː-nʌ̀	
Jumjum	pâː-n	
Protoform	*paːy	

2. 'Moon' in West Nilotic languages.

Composition of the family:

Bari

(Bari, Kakwa, Pojulu, etc.)

Teso-Turkana

(Teso, Turkana, Karimojong, Nyangatom) — Proto-East Nilotic

Lopit-Lotuko

(Lotuko, Oxoriok, Lopit, Lokoya)

Ongamo-Maa

(Maasai, Camus, Sampur; Ongamo)

Forms:

Language	Form (sg.)	Form (pl.)	
Language	Form (sg.)	Form (pl.)	
Teso	ɛ̄=lápà		
Turkana	è=làp		*CVCVC structure reconstructible based on the data of plural forms (and Ongamo, where *haʙaha-ni ← *ɬapaɬa-ni, a new singulative formation).
Karimojong	ɛ̄=lāp		
Nyangatom	=lɛb		
Lotuko	ā=yáfà		
Oxoriok	ɔ̄=yɛ̄fā		Peculiarities:
Lopit	yāɸá	yāɸáʒ-ìn	
Dongotono	āfā		(a) rare root structure for a basic nominal stem, suggesting an old borrowing or compound;
Lokoya	ɔ̄=yɛ́và	ɔ̄=yɛ̀vàt-ìk	
Maasai, Camus	ɔ̄=lápà	ì=lápàt-ín	(b) rare phoneme in root-initial position (only one other example: *=ɬac- 'louse');
Ongamo	ɔ̀=hàβàhá-ni		
Bari	yápàʔ	yápál-à	(c) unique correspondence in root-final position (Vossen reconstructs *tʸ — uneconomical and unconvincing).
Kakwa	yápà		
Pojulu	yápáʔ		
Protoform	*ɬapa(ʔ)	*ɬapaɬ-	

3. ʾMoonʾ in South Nilotic languages.

Composition of the family:

Datooga

Omotik   Proto-South Nilotic
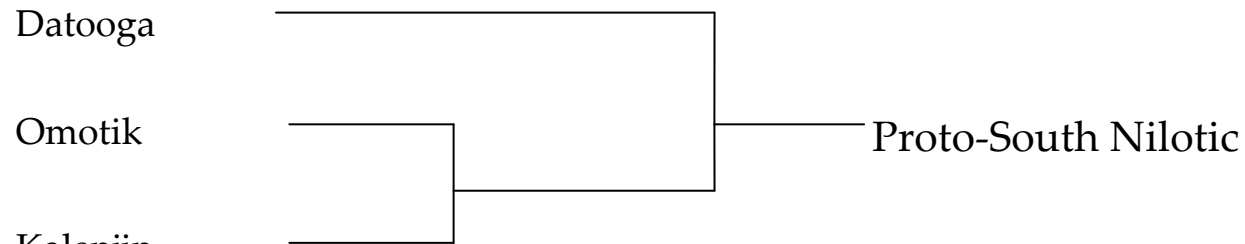
Kalenjin

(Nandi, Kipsikis, Markweta, Sabiny, Kony, Pok, Terik, Päkot, etc.)

Forms:

Language	Form	[PSN *ʟ → Omotik ɬ, Datooga š, Kalenjin *l*; cf. also
Omotik	ɬɛ̀ː-tà	PSN *ʀ → Omotik y, Datooga w, Kalenjin *r ~ x ~ y*]
Datooga	šềː-dà	Cf. Proto-Kalenjin: *araːw-* ʾmoonʾ
Protoform	*ʟɛ-	(no parallels in Datooga or Omotik; topologically, an innovation)

No possible etymological scenario!

SNil * *ɬE-:

may be the same as **ɬa-* in ENil **ɬa-paɬ*,

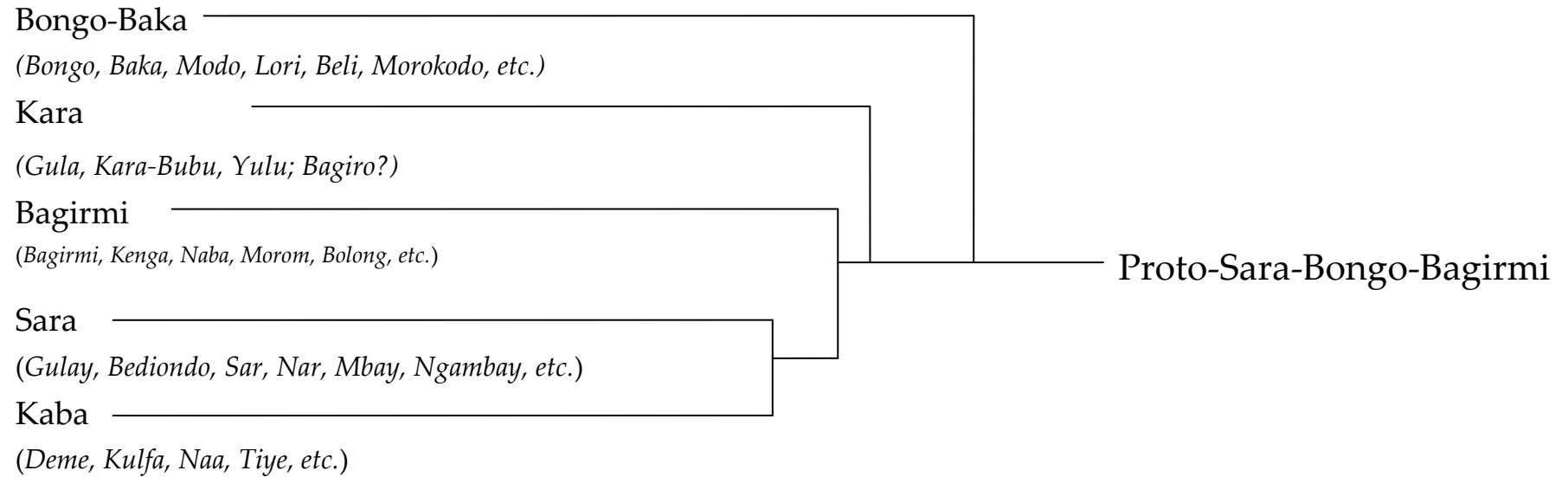but how to explain the second syllable?

ENil **ɬa-paɬ*:

second syllable may be the same as WNil *paːɣ*,

but how to explain the first syllable?

[Cf. a similar strange case of WNil CVC corresp. to ENil/SNil CVCVC:

ENil **ŋa-ʒep* : SNil **ŋa-ɬep* : WNil **lɛp* ʿtongueʾ]

4. ʾMoon' in Sara-Bongo-Bagirmi languages (Central Sudanic).

Composition of the family:

Bongo-Baka

(Bongo, Baka, Modo, Lori, Beli, Morokodo, etc.)

Kara

(Gula, Kara-Bubu, Yulu; Bagiro?)

Bagirmi

(Bagirmi, Kenga, Naba, Morom, Bolong, etc.)

Sara

(Gulay, Bediondo, Sar, Nar, Mbay, Ngambay, etc.)

Kaba

(Deme, Kulfa, Naa, Tiye, etc.)

Proto-Sara-Bongo-Bagirmi

Forms:

Language	Form		Language	Form
Baka	ɛ́fɛ́		Bagiro	nap ~ napo
Bongo	ɲíhí		Kenga	lāːpā ~ nāːpā
Modo	ɲɛ́pɛ́		Naba	naːfɛ
Lori	ɲɛ́fɛ́		Bolong	lapa
Beli	ɲɪpɪ			
Morokodo	ɲɛhɛ		Proto-Sara	*nãy
Bagiro	nāfɛ̄		Deme	nɔ̄hē
Fer	līf		Tiye	nɔ̀fé
Kara-Bubu	lihi		Naa	nóhé
Yulu	ɲɛ̄ːp		Kulfa	náfé
Gula	lēhē ~ nɔ̄hɔ̄			

Proto-Sara-Bongo-Bagirmi:

*napɛ ~ *lapɛ, with a unique correspondence series (Boyeldieu 2000: **ʔ-f-*)

Cf. ʾmoonʾ in other subgroups of Central Sudanic:
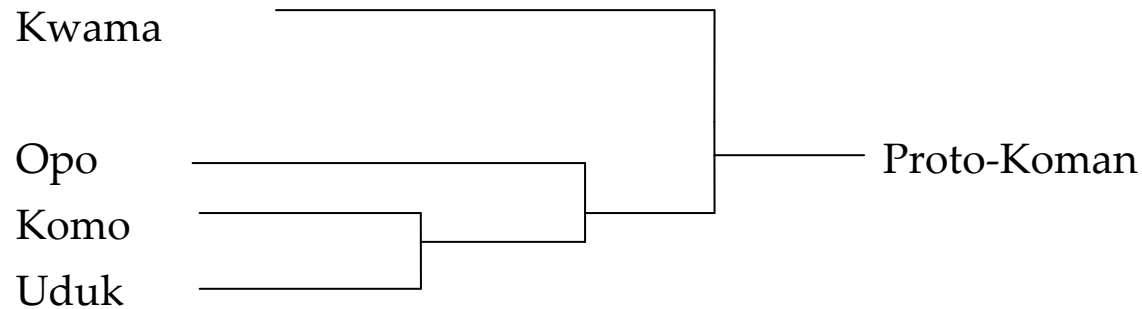
Proto-Moru-Maɗi: *ìmbā

Proto-Lendu-Ngiti: *àbī

Proto-Mangbetu-Asoa: *=aŋgwɛ

Proto-Mangbutu-Efe: *=imba ~ *=emba

Comparison between Moru-Maɗi and Mangbutu-Efe (most distant members of East Central Sudanic) allows to reconstruct PECS **(i)mba** ʾmoonʾ. No traces of *napɛ ~ *lapɛ.

5. 'Moon' in Koman languages (Ethiopia / Sudan).

Composition of the family:

Language	Form	
Kwama	sʔiawan	
Opo	a=dɔy	Same as in Proto-Luo (*dwʌy). Borrowed from Anywa?
Komo	pái	Same as in Proto-West Nilotic (*paːy). Borrowed from WNil into Komo-Uduk, or from Proto-Koman into Proto-WNil?..
Uduk	à=ppéː	

Cf. a strange scenario in Ehret 2001:

701. *àp'á:y   "moon"
  Koman: UDUK àppéé "moon"
  Koman: KOMO [pai "moon": LOAN from WNil]
  CSud: WCSud *εpε "moon"   [regressive V-fronting assim.]
  Kir-Abb: Nil: PWNil *pai "moon"   [possible ancient loan from Koman?]

1) Uduk and Komo forms are inherited from a common ancestor (Uduk *a=* is a standard nominal prefix); if borrowed from WNil, then *both* forms are borrowed, not just one.

2) "WCSud **εpε* is actually just Baka *εfε* (see above), should not be extracted from SBB **napε ~ *lapε*. Loss of word-initial resonants in Baka is common.

3) If PWNil is an "ancient loan from Koman", and Komo is a "loan from WNil", then the entire etymology is essentially an Uduk-Baka isogloss. But Uduk cannot be separated from Komo, meaning that the etymology is non-existent.

Adding data from other families:

(A) Niger-Congo: Ubangi (Moñino et al. 1988)

Language	Form	
Sere-Ngbaka		Very similar to West Nilotic, Koman.
Ngbaka	pɛ̄	May have been borrowed into some small, geographically adjacent subbranches of Central Sudanic (Lendu-Ngiti *àbī, Birri *afi*).
Mayogo	=pɛ́	
Mundu	fɛ́	
Ndungale	hɛ́-	
Dongo-ko	pɛ̀-	
Sere	fɪ	
Banda		Very similar to SBB *napɛ ~ *lapɛ. Since *CVCV forms are normal for Banda, but not very normal for SBB (especially if the first C is represented by a unique series of correspondences), may ultimately be a Ubangism in SBB.
Linda	yípɨ̄	
Yangere	yúpō	
Ngao	yípī	
Mbanza	ɲépɨ̄	

(B) East Cushitic: Lowland East Cushitic (Konsoid, Arbore, Elmolo) + Dullay + Yaaku

Language	Form
Elmolo	lêʔ
Arbore	lèh
Konso!	léyaː
Mashile!	léya
Gidole!	léhaː
Gawwada	leʔo
Gobeze	leːʔo
Werize	leːʔo
Tsamai	lɛʔu
Yaaku	lɛɛʔ
Proto-LEC (Black 1971)	*leʕ

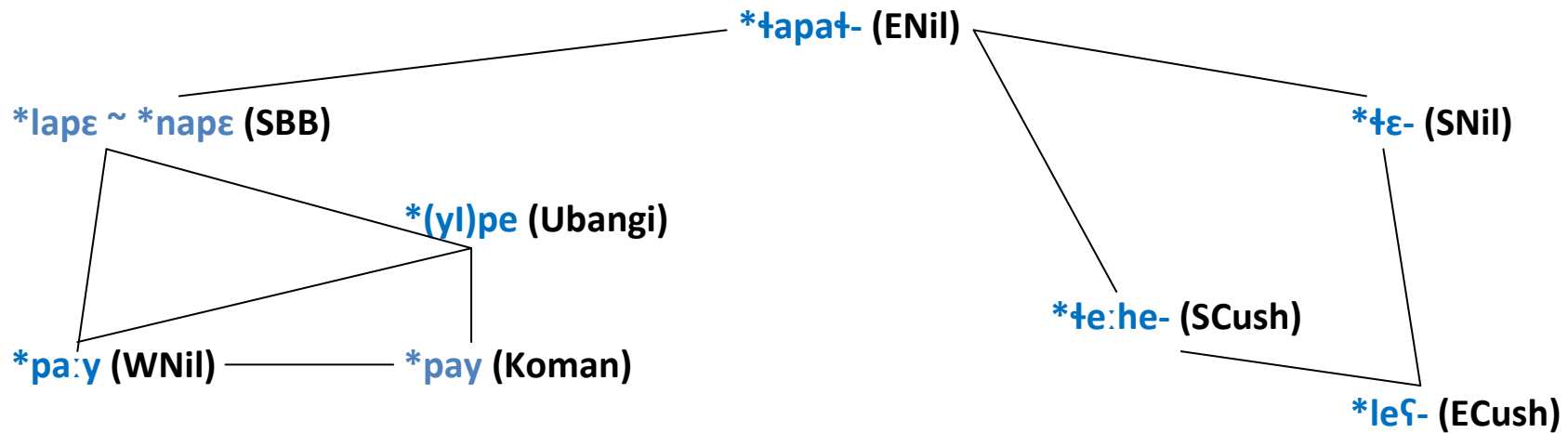Note: absent in the Northernmost LEC languages (Oromo, Somali), but present in Ongota: **léʕa**

(C) South Cushitic:

Language	Form
Ma'a	m=ɬíhe
Qwadza	ɬahay-iko
Aasax	lehe-k
Iraqw	ɬahaŋ
Burunge	ɬeheŋ
Alagwa	ɬehe
Proto-SC (Ehret 1980)	*ɬeːhe-

Observations:

1) EC ***leʕ** and SC ***ɬeh-** do not correspond to each other regularly under any of the existing correspondence models for Cushitic.

2) Both forms (especially the second) are strikingly similar to Proto-South Nilotic ***ɬE-**. Is it a coincidence that Omotik, Datooga, and South Cushitic languages are spoken in adjacent regions? Who borrowed from whom?..

Final hypotheses:

*ɬapaɬ- (ENil)

*lapɛ ~ *napɛ (SBB)

*ɬɛ- (SNil)

*(yI)pe (Ubangi)

*ɬeːhe- (SCush)

*paːy (WNil) ——————— *pay (Koman)

*leʕ- (ECush)

1) An original root with an approximate structure like *ɬape or *ɬapay may be ultimately responsible for this entire areal cluster. The language to which it originally belonged remains unknown, but it could hardly have been the direct ancestor of any of the listed groups.

2) Typical paths of development were either *ɬapay → *ɬafay → *ɬahay (from whence most of the Cushitic and South Nilotic forms), or *ɬapay → *apay → *pay (from whence most of the Nilotic, Central Sudanic, and Ubangian forms).

3) Despite the similarities, these forms cannot be used as arguments in favor of long-range hypotheses — e. g. the link between Nilotic, Koman, and Central Sudanic crumbles under the weight of phonetic and topological arguments.