CONTENTS

1 The 200,000-Year Evolution of Homo sapiens sapiens Language and Myth Families based on mtDNA Phylotree Archaeogenetics: A Thought Experiment
   James B. Harrod

87 The Kinship Term KUKU ~ KOKO ~ KAKA in the American Indigenous Languages, the Amerind Hypothesis, and the Dravidian Kinship System Section I: Linguistic Study
   Alain Matthey de l'Etang & Pierre J. Bancel

155 Indo-European and Dravidian: Some Considerations
   Stephan Hillyer Levitt

181 Burushaski and the Western Dene-Caucasian Language Family: Genetic and Cultural Linguistic Links
   John D. Bengtson

Book Reviews


229 Michael Fortescue: Language Relations across Bering Strait: Reappraising the Archaeological and Linguistic Evidence. Reviewed by Peter A. Michalove

233 News & Notices: Association for the Study of Language in Prehistory (ASLIP)
For subscription and membership information contact:

Secretary-Treasurer: Michael T. Lewis
lewismtc@rcn.com
20 Duane Avenue
Tel. 617-964-0978
West Newton, MA 02465
U.S.A.

ASLIP membership: $35.00 USD, yearly
Lifetime membership: $500.00 USD
For electronic payment see PayPal link at http://aslip.org/

For other information about ASLIP, or to inform ASLIP of important news, new scientific developments, and media relations, please contact:

Information Officer: Jonathan Sherman Morris
São Paulo, Brazil
jonathanmorris1964@gmail.com Tel: 5511-31512667

For submission of articles, notes, book reviews, or letters to MOTHER TONGUE, please contact:

Editor: John D. Bengtson palaeojdb@hotmail.com
Savage, MN U.S.A. Tel. 612-839-3649

For any other correspondence with ASLIP officers and advisors, see the contact information on the inside front cover of this volume.

ISSN 1087-0326 ASLIP Homepage: http://aslip.org
The 200,000-Year Evolution of Homo sapiens sapiens
Language and Myth Families based on the mtDNA PhyloTree, Fossil mtDNA and Archaeology: A Thought Experiment

James B. Harrod
Director, Center for Research on the Origins of Art and Religion

Abstract. A meta-pattern-analysis of the mitochondrial DNA phylotree and current distribution of language families indicates that over the last 200,000 years there are robust correspondences between mtDNA haplogroups and language macrofamilies. This study is a thought experiment, a top-down derivation of the Homo sapiens sapiens ('Proto-Human', 'Proto-World') language phylotree, which can be tested against bottom-up prehistoric linguistic reconstructions. It establishes a relative chronology for dating the emergence and branching of the global array of language macrofamilies. The language phylotree is crosschecked against archaeological data and fossil mtDNA studies, which support many of the correlations. Results imply L3M and N dispersed out-of-Africa at around 80,000 years ago with both Afrasian and Nile-Sudanic languages and mythological systems. After a 3-to-5,000-year pause in SW Asia three Borean language superfamilies emerged, Borean-N (Dené-Caucasian), Borean-M (Eurasiatic) and Borean-R, the latter including language families of SW Asia and Europe as well as SE Asia and Sahul. Alternative short-chronology hypotheses for language evolution, dating of sapiens sapiens out-of-Africa and a 'southern route fast track' from SW Asia to Sahul do not appear supported by either mtDNA genetics or archaeology. A hypothesis aligning all language families to the mtDNA phylotree yields a more differentiated and different chronology to the dyadic out-of-Africa dispersion model proposed in Fleming, Zegura, Harrod, Bengtson & Keita (2013).

Introduction.

In linguistics, a macrofamily is a hypothetical phylogenetic grouping of language families based on relationships established by lexicostatistic, multilateral ('mass') comparison or other methods (e.g., Gell-Mann, Peiros & Starostin, 2009). Proponents argue that long-range prehistoric language reconstruction must necessarily begin by applying taxonomic methods before traditional methods of proto-language reconstruction are applied. Examples of proposed macrofamilies include the superfamily Proto-Sapiens (Ruhlen, 1994; Bengtson & Ruhlen, 1994; Trombetti, 1905); models for an out-of-Africa superfamily Borean (Fleming, 2002; 1991; Fleming, Zegura et al., 2013; Starostin, 2006); and smaller groupings, such as Afroasiatic (Greenberg, 1955; 'Afrasian', Fleming, 2002), Nostratic (Bomhard, 2008; Dolgopolosky, 2008, 1998; Starostin, 1999;

Objections to acceptance of a particular macrofamily may include lack of documentation or scholarship on the constituent languages or disagreements over taxonomic methods or findings. Short-range reconstructionists may object to the general macrofamily approach due to aversion to long-range ‘mass’ comparisons or belief that an estimated time depth is too great for reconstruction, arguing a limit around 25,000 years ago for glottochronology using a traditional method of Swadesh list binary comparisons (Fleming, Zegura et al., 2013:162), or asserting that the multilateral comparative method has an inherent limit of 6-10,000 years.

With respect to the early dating of language, there is extensive evidence in the archaeological record of symbolic behavior for archaic Homo sapiens, Neanderthals and Homo sapiens sapiens, and less extensive evidence of even earlier symbolic behavior for Homo erectus. As for reconstructing Homo sapiens sapiens language evolution, Atkinson (2010) reviews glottochronology methods and observes that the standard method for glottochronology developed by Swadesh places an upper limit on language classification at around 8,000 years and a modified method (Pagel, Atkinson and Meade, 2007; Atkinson, 2010) may extend the limit to 50,000 years or so. If so, such methods appear to be incapable of dealing with predictions for out-of-Africa languages. Atkinson (2011) finds that phonemic diversity supports a serial founder effect out-of-Africa with a cline by distance similar to that of mtDNA, but provides no dating for language family evolution. On the other hand, Perreault and Mathew (2012) find that ‘proto-Sapiens sapiens’ language arose in tandem with the emergence of the species. Based on phonemic diversity they calculate that the language of sapiens sapiens emerged between 163 and 242 ka, a date range corresponding to the earliest fossil attributed to our species, Omo Kibish, ~195 ka.

One of the obstacles to long-range macrofamily phylogenetics is the absence of a method to establish a comprehensive global chronological sequence for sapiens sapiens language evolution. Fleming, Zegura et al. (2013:162) observe “the major problem with linguistic genetic taxonomy is dating or time depth.” In this study I aim to attenuate this obstacle. I propose to derive a model for a global language family phylotree drawing on mtDNA archaeogenetics. My literature review indicates some attempts have been made to detect mtDNA correlations for particular geographic areas, e.g. Europe (Soares, Achilli et al., 2010) and Central and South America (Wang, Lewis et al., 2005; Hunley, Cabana et al., 2007), but as yet there has been no global attempt.

Given the vastness of the Y-DNA and mtDNA archaeogenetic literature, I have chosen to focus on mtDNA, conducting a comprehensive review of mtDNA studies of haplogroup frequencies for particular population or tribal samples and currently spoken language. I use the standardized global mtDNA phylotree (van Oven and Kayser, 2009, Build 15, 30 Sep 2012). I use Soares, Ermini et al. (2009) and Behar, van Oven et al. (2012) with their comprehensive re-
examinations of mtDNA phylogenetics and calculations of haplogroup TMRCA dates. Archaeology and fossil mtDNA studies are used as an additional check on correlations. My method for deriving correlations for the *sapiens sapiens* language phylotree is primarily one of meta-pattern analysis across disciplines.

There are at least two models for the earliest language families that emerged with *Homo sapiens sapiens* out-of-Africa. Fleming (2002; 1991; Fleming, Zegura et al., 2013) has proposed the existence of a super-phylum, Borean, including ten different language groups, with Afrasian and Amerind as western and eastern anchor groups.

- Afrasian (Greenberg’s Afroasiatic) (Semitic, Egyptian, Berber, Chadic, Cushitic, Omotic)
- Kartvelian
- Dravidian
- Sumerian, Elamitic, and a few other extinct languages of the Near East
- Eurasian (Greenberg’s grouping including Etruscan, Indo-European, Uralic, and Altaic-Mongolian-Tungusic, Japanese-Korean-Ainu, Gilyak/Nivkh, Chukotian, Eskimo-Aleut)
- Vasco-Caucasian (Basque of Iberia, Caucasian of the Caucasus)
- Burushaski and Yeniseian
- Tibeto-Burman (Sino-Tibetan)
- Na-Dené
- Amerind

This list may be grouped into two macrofamilies with the first two clusters combined as ‘Nostratic’ (Pedersen, 1931, Illich-Svitych, 1971, Dolgopolsky, 1998, Bomhard, 2008) and the third, Dené-Caucasian (Bengtson, 2008; 1997; Nikolayev, 1991).

Gell-Mann, Peiros and Starostin (2009) hypothesizes a ‘Borean’ that included Nostratic (combining the first two major groupings above) and a second super-family Dene-Daic, subdivided into Dené-Caucasian, the third grouping above, plus Austric, which thus would add a fourth cluster to Fleming’s list, Austric, with four clades:

- Austroasiatic (Mon-Khmer)
- Hmong-Mien (Miao-Yao)
- Austronesian
- Tai-Kadai (Daic).

Compare hypothetical tree with Nostratic (Eurasian and Afroasiatic) and a supercluster termed ‘Borean’ (Dene-Caucasian and Austric) (Starostin, 2006) [http://starling.rinet.ru/images/globet.png](http://starling.rinet.ru/images/globet.png) (accessed 15 August 2014). These four clusters cover the major language families outside of Africa, with the exception of a fifth group consisting of Sunda-Sahul languages:

- Pama-Nyungan (‘Australian’)
- Gunwinyguan (‘Australian’)

3
In this study, I suggest how all five language family groups appear to correlate to distinctive mtDNA haplogroups and I infer the relative dating of the emergence of these language families based on archaeogenetic TMRCAs (Time to Most Recent Common Ancestor).

A summary of the current basic mtDNA phyloptree (after en.Wikipedia) follows:

![Evolutionary tree of human mitochondrial DNA (mtDNA) haplogroups](image)

A generally accepted map of mtDNA homelands and migratory routes out-of-Africa and across Eurasia (en.Wikipedia, citing Metspalu, Kivisild et al., 2004) follows.

![Map of mtDNA homelands and migratory routes](image)
The map appears to rely on a short-chronology model of out-of-Africa by a fast ‘southern route’. In the study that follows I present a series of maps based on mtDNA haplogroup TMRCAs that suggest the timing of stages for both in-Africa and out-of-Africa homelands and migrations, including the effects of geographic bottlenecks such as the Sinai Crossing, Transcaucasus Crossing to Europe, Zagros Crossing to India and the Ganges Crossing into SE Asia, on both haplogroup migration routes and the emergence of language families.

**Hypothesis.**

At first glance some language families appear strongly correlated to a basic mtDNA haplogroup, such as African click languages and L0-mtDNA or Eurasiatic and M-mtDNA. Given such prima facie correspondences, a hypothesis that there are no correlations between languages and genetics or mere randomness appears unsupported. Given such obvious correlations, I hypothesized that by and large each basic haplogroup might well correspond to a language family. Further, I hypothesized that the mtDNA phylotree as a whole would correlate to a global African and out-of-Africa language phylotree.

It is critical to note that when I use the prefix ‘Pre-’ before a language name, it does not designate a reconstructed protolanguage. I used this prefix to designate (a) that a specific mtDNA haplogroup emergence has occurred indicating a population has begun genetic separation and, presumably, associated geographic separation from other populations sufficient to provide a basis for the emergence of a distinct language family and (b) that in current populations this haplogroup appears differentially associated with a particular language family. My method is top-down; ‘Pre-’ does not mean a ‘proto’ as in a protolanguage reconstruction. I am not arguing that a haplogroup emergence date is the date for the emergence of a protolanguage. I am suggesting that a particular haplogroup TMRCA date corresponds to the emergence of an ancestral mtDNA lineage that is strongly associated with one particular language family. An associated protolanguage might have emerged around the time of the haplogroup’s most recent common ancestor date (TMRCA) or at a later time. Hypotheses about the timing of a language family emergence with respect to TMRCA must be checked on a case-by-case basis. In some cases, especially around 25 ka (‘ka’ = thousand years ago) or later, particular haplogroups may be associated with a dozen or more different languages, and I leave these aside under phrases such as ‘adopted various languages’. These cases may have resulted from the original language family being completely or partially lost.

To take an example of how I use the term ‘Pre-’, consider my term ‘Pre-Basque’. While HV-mtDNA and its descendent H are frequent in Europe, the population with the highest frequency of these haplogroups in Europe occurs in Basque country, depending on the study sample up to 81%. The HV-mtDNA phylotree sequence, TMRCA dates in parenthesis, is: root haplogroup R branched R0 in northern SW Asia (~40 ka). R0 branched R0 subclades across SW
Asia, especially the Arabian peninsula (~13 ka) and branched HV in the NW Fertile Crescent (~25 ka). HV dispersed westward through the Caucasus (high frequencies of HV in NE-Caucasian-speaking Dargin and Avar, 43% and 20% respectively) into Belarus-Ukraine, where it branched HV4 (~14 ka), and along the way H (~15 ka). In addition to the highest frequency of HV and H in Europe, the Basque population also has around 10% U5, which is a marker associated with Uralic languages. This suggests an extended pause in the Belarus-Ukraine area. The migration continued across southern Europe to the Franco-Cantabrian ice age refugium yielding HV4a1 (~10 ka). Genetic isolation of HV4a1a around 5.4 ka may signal the emergence of the current distinct Basque language and culture (Gómez-Carballa, Olivieri et al, 2012; Behar, Harmant et al, 2012; Yunusbayev, Metspalu et al, 2012; Abu-Amero, Larruga et al, 2008). The R0/HV diffusion route with its extended pause in Belarus-Ukraine would predict Uralic borrowings in Basque. Not being a linguist of these languages I leave this prediction aside. I mention this case to stress how I am using the prefix ‘Pre-’. When I correlate R0/HV-mtDNA under the label ‘Pre-Basque’, with a rough beginning date of 50,000 years ago, I am referring to the mtDNA lineage that is differentially that of current Basque language speakers. I leave open the question of protolanguage reconstruction and the question whether, if Basque belongs to the Dene-Caucasian family, the migrating R0/HV-bearing population maintained or adopted a linguistic predecessor of Basque when they crossed the Caucasus bottleneck.

One might speculate when a language emergence occurred subsequent to a haplogroup emergence, but I suggest it probably occurred before the next major haplogroup branching event. For instance, R mtDNA branches off R2’JT in SW Asia with a TMRCA around 54 ka and subsequently JT mtDNA, which I suggest correlates to Pre-Semitic-Egyptian, around 47 ka, and then T branches off at around 25 ka. This is roughly a 25,000-year span. When during these 25,000 years Semitic-Egyptian or Semitic and Egyptian emerged I leave as an open question. I suggest that in general language family emergences probably occurred closer to the TMRCA than later branching events. Whether or not one might some day discover a general rule for an average delay time across all major mtDNA haplogroup TMRCA—say, 5,000 years after—I would argue that the phylomtree that I have detected would still maintain its ordered pattern of branches.

The general linguists view on language continuity is that for any given geographic area, languages may have been submerged or completely lost, or abandoned in favor of another language. Worldwide evidence for such is evident in historic records. One cannot assume present-day languages continue ancestral languages over tens of thousand years or even a thousand years. The results of this study challenge this assumption. Just as highly differentiated mtDNA haplogroups are recent, so some languages are recent. However, such a claim is not universal. In the case of African click-languages, an opposite inference for long-range continuity is supported. According to mtDNA genetics and archaeology, Hadza, Sandawe and Khoisan languages differentiated externally from all other language families around 160,000 years ago. East African click peoples split from Southern African click speakers around 100,000 years ago.

6
Hadza and Sandawe appear to have genetically split by 70,000 (based on L0f lack in Hadza). The full Khoisan tool-and-art package is archaeologically evident in southern Africa by 45,000 years ago. Southern Khoisan (SAK) split from Northern Khoisan prior to 25,000 years ago. We do not know what languages may have been in southern Africa prior to the Khoisan arrival and thereafter lost. This lack of knowledge on our part does not refute genetic and archeological evidence that click language family continuity extends over at least 100,000 years. To date no alternative hypothesis for the apparent relationship between Khoisan and East African click languages has been proposed. Thus, the continuity of any given current language may be short-range or long-range; it must be established on a case-by-case basis and tested against mtDNA and Y-DNA phylotrees and archaeology.

With respect to TMRCA dating different studies sometimes suggest divergent dates for a TMRCA. This may result in linguists eschewing any attempt to use mtDNA TMRCAs. There is also the problem of selecting a molecular clock mutation rate and its dependence on the human-chimpanzee split date (‘CHCLA’), which over the decades has been steadily revised upward, so that TMRCAs must be revised to earlier and earlier dates (Harrod, 2013). Whether or not a TMRCA date is secure or ‘absolute’ is not as important in this study as the relative chronological sequence of the dates for haplogroup emergence. In this regard, Soares, Ermini et al. (2009) and Behar, van Oven et al. (2012), provide dates for haplogroups covering more or less the whole mtDNA phylotree and represent the two most reliable and useful data sets. The former study calculates dates in relation to the Cambridge Reference Sequence (CRS); the latter, from an African perspective. The relative sequence of dates in each study is roughly similar, and I reference the calculated dates for both in my database (Supplementary File, Table 1: mtDNA Database for Archaeolinguistics, online, https://originsnet.academia.edu/JamesHarrod).

I am not a linguist and I am not proposing a standard linguistic proto-language reconstruction. I am suggesting meta-pattern correlations between current population genetic haplogroup frequencies and current languages and crosschecking these against archaeological data and fossil mtDNA studies. This is basically a thought experiment. I present a hypothesis, search for patterns, propose correlations and unfold their logic. I view the overall pattern as a touchstone for linguists to use in ascertaining the sequence by which sapiens sapiens languages emerged. This paper revised some of the datings and genetics to language correlations I very tentatively proposed in Harrod (2013). Being neither a geneticist nor a linguist, I anticipate that some genetics-language family correlations may be refuted by proto-language reconstructions. With this caveat in mind, I welcome counterarguments with respect to specific correlations. If it has been said that conflating language and culture and genetics is the cardinal sin of anthropology, then felix culpa.

Method.
First, I reviewed mtDNA archaeogenetics and language family literature relevant to *Homo sapiens sapiens* both in Africa and out-of-Africa and generated from this a Master Database table (Supplementary File, Table 1: mtDNA Database for Archaeogenetic linguistics, online, https://originsnet.academia.edu/JamesHarrod). To date this database table is 181 pages long and contains at rough count 424 ethnic and population mtDNA haplogroup frequency samples and their current spoken language, and 82 fossil mtDNA studies, select archaeological sites and their references. For mtDNA phyloptree and haplogroups I used van Oven and Kayser (2009) Build 15 (30 Sep 2012). For mtDNA TMRCA (Time to Most Recent Common Ancestor) dates I relied primarily on Soares, Ermini et al. (2009), and crosschecking dates in Behar, van Oven et al. (2012). Where they do not provide a TMRCA for a particular haplogroup I have cited other studies. In addition, some relevant archaeology is incorporated into the table. As far as I am aware, Supplementary File, Table 1 is the most comprehensive inventory of mtDNA population samples, fossil mtDNA and archaeological parallels available online. It covers the full 200,000 years of *sapiens sapiens* evolution, all macrolanguage families, and many subfamilies.

For each basic mtDNA haplogroup (Hg) I asked what, if any, language family distinctively (as opposed to some other language family) corresponds to it. Generally, current ethnic or population samples are diverse; they have multiple mtDNA haplogroups. To tentatively hypothesize a correlation between an mtDNA haplogroup and language family, I employ several criteria.

**Procedure 1. Phylotree.** I order the entire master database (Supplementary File, Table 1, http://www.originsnet.org/publications.html) by the standard mtDNA Phylotree and TMRCA for each haplogroup.

**Procedure 2. Haplogroup Samples.** Under a specific haplogroup, I generally list several population or tribe haplogroup frequency samples with focus on those having highest frequencies for that haplogroup and try to rank order them by frequencies. For a given population or tribe I may list more than one study of that group's mtDNA haplogroup frequencies.

**Procedure 3. Hg Population Assignments.** In order to construct the master database, I assigned a population sample to a specific haplogroup, e.g., L1 or M or Q3. The rationale for assignment is strongest, of course, if a whole population shares one basic haplogroup, but this is rare. A second criterion for selection is to choose the Hg with the highest frequency in a sample. A third is to select the Hg with the oldest TMRCA. I may use one or both of the latter two criteria depending on contextual factors. In listing a sample in the database I generally place in bold the Hg with the highest frequency in the sample.

**Procedure 4. Homeland.** I rely on Hg homelands proposed in Soares, Ermini et al. (2009) or other archaeogenetic studies.

**Procedure 5. Language Family Correlate.** For every population or tribal mtDNA frequency sample I attempt to identify the current language and language family (grey highlight in the
Master Database). Reviewing the currently spoken languages allocated to a particular haplogroup may show a more or less clearly distinctive language family for all the peoples listed under it. In some cases if current language families are diverse, an assignment of the Hg to a language family may be supported or rejected by the logic of mtDNA Phylotree in which it is embedded, circumstantial evidence or known history. In other cases I do not hypothesize any definite correlation to one language family. This phenomenon becomes more frequent as the timeline of the mtDNA phylotree progresses.

If for populations listed under a particular haplogroup, there is a consistent current language family but it is a mismatch for the mtDNA, I assume that the current language is adopted and replaces an earlier language family. One of the most evident mismatches is the M-mtDNA derived D-mtDNA population (Han). By the logic of the phylotree Han peoples should have a Eurasian language but their current Chinese language belongs to the Dené-Caucasian macrofamily otherwise correlated to N-mtDNA.

Procedure 6. Archaeological and Fossil mtDNA Context. Archaeological site dates as well as fossil mtDNA studies are noted under particular haplogroups. They contribute to supporting or ruling out particular mtDNA and language correlations and homelands.

Procedure 7. Mythostratigraphy. Once the language phylotree and language family emergence dates are coordinated, identify mythological systems that may correspond to the phylotree.

Clearly, the correlations I propose are tentative. I reiterate that I propose only a hypothesis, a thought experiment applying the mtDNA phylotree to all sapiens sapiens languages. I am not proving or linguistically reconstructing protolanguages or emergence dates. I am developing a meta-model based on TMRCA dates along the mtDNA phylotree against which prehistoric linguistic reconstructions may be guided and tested.

Results.

Table 1 summarizes the meta-pattern-analysis of Homo sapiens sapiens language evolution phylotree crossmapped onto the mtDNA phylotree. It is derived from the comprehensive Master Database (Supplementary File, Table 1, online, https://originsnet.academia.edu/JamesHarrod). All of the identified mtDNA haplogroups and correlated language families will be discussed in this results section in chronological order beginning with the emergence of sapiens sapiens around 200,000 years ago. Table 1 also suggests very tentative mtDNA correlations to the global evolution of Homo sapiens sapiens mythological systems as proposed by Witzel (2011) and Berezkin (2010a, 2010b), and these also will be discussed in greater detail below.
### Table 1: mtDNA Archaeogenetic Phylotree Correlated with *Sapiens Sapiens* Language Families

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>~190-120 ka</td>
<td>Proto-Sapiens</td>
</tr>
<tr>
<td>L'Eve</td>
<td></td>
</tr>
<tr>
<td>L0-&gt;a'b'y:k:</td>
<td>Pre-Hadza-Sand.</td>
</tr>
<tr>
<td>L1-6-&gt;L1:</td>
<td>Pre-Niger-Congo</td>
</tr>
<tr>
<td>L2-6-&gt;L2'34'6:</td>
<td>Pre-Nilo-Saharan-Afrasian 1st out-of-Africa</td>
</tr>
<tr>
<td>L2-6-&gt;L5:</td>
<td>Pre-Nilo-Saharan-Central Sudanic</td>
</tr>
<tr>
<td>L3'4'6:</td>
<td>Pre-Nilo-S 2nd out-of-Afr</td>
</tr>
<tr>
<td>L3'4 continues</td>
<td>2nd out-of-Afr</td>
</tr>
<tr>
<td>L4: E Africa</td>
<td></td>
</tr>
<tr>
<td>L6: Yemen</td>
<td></td>
</tr>
<tr>
<td>'Pan-Gaea'</td>
<td>(Witzel)</td>
</tr>
<tr>
<td>M7: Pre-Japonic-Korean-Ainu</td>
<td></td>
</tr>
<tr>
<td>M8: Altaic-Tungusic-Mongolian</td>
<td></td>
</tr>
<tr>
<td>M9: M42'74, M33: (adopt various)</td>
<td></td>
</tr>
<tr>
<td>M27, M29'Q:</td>
<td>Q2 Pre-Papuan</td>
</tr>
<tr>
<td>N2: Pre-NE-Caucasic (NEC)</td>
<td></td>
</tr>
<tr>
<td>N9: Pre-Nivkh (adopts M2)</td>
<td></td>
</tr>
<tr>
<td>N1: Pre-Burushaski (&amp; adopt M)</td>
<td></td>
</tr>
<tr>
<td>N11: Philippines (adopts Austro?)</td>
<td></td>
</tr>
<tr>
<td>N-&gt;O:</td>
<td>(adopts Pama-Nyungan?)</td>
</tr>
<tr>
<td>U1: Pre-Elamitic (adopts IE, etc.)</td>
<td></td>
</tr>
<tr>
<td>U4: Kartvelian (ad. Samoyedic, IE)</td>
<td></td>
</tr>
<tr>
<td>U5: Pre-Finno-Ugric (Gravett.)</td>
<td></td>
</tr>
<tr>
<td>R2JT: Pre-Semitic-Egyptian</td>
<td></td>
</tr>
<tr>
<td>R0: Pre-Basque (Gravettian)</td>
<td></td>
</tr>
<tr>
<td>R24, B4'5, R9:</td>
<td>Pre-Austroic</td>
</tr>
<tr>
<td>R21: Aslian Austroasiatic</td>
<td></td>
</tr>
<tr>
<td>B4a: Austronesian</td>
<td></td>
</tr>
<tr>
<td>B5: in Tai-K, Tibeto-Burman</td>
<td></td>
</tr>
</tbody>
</table>
Derived from the literature review detailed in the Master Database (Supplementary File, Table 1, https://originsnet.academia.edu/JamesHarrod) and its inferred language family correlations, highlights of the 200,000-year prehistory of *Sapiens sapiens* language evolution follow. We first look at mtDNA evolution in Africa. It can be reduced to five basic early stages prior to out-of-Africa stages, namely L, L0, L1, L5, L2 and then L3, and they emerge roughly every twenty thousand years (Table 2).

<table>
<thead>
<tr>
<th>Table 2. Early Stages of in-Africa mtDNA PhyloTree</th>
<th>Prior to M/N Out-of-Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td>~190 ka</td>
<td>L-‘Eve’</td>
</tr>
<tr>
<td>~160 ka</td>
<td>L0</td>
</tr>
<tr>
<td>~140 ka</td>
<td>L1</td>
</tr>
<tr>
<td>~120 ka</td>
<td>L5</td>
</tr>
<tr>
<td>~100 ka</td>
<td>L2</td>
</tr>
<tr>
<td>~80 ka</td>
<td>L3</td>
</tr>
</tbody>
</table>

L-‘Eve’: *Proto-Sapiens-Sapiens*. The mtDNA phylotree begins with L-mtDNA (‘Eve’), which has a TMRCA at around 200,000 years ago (Behar, van Oven et al., 2012; Soares, Ermini et al., 2009). This closely matches dating for the first fossil evidence attributed to *Homo sapiens sapiens*, Omo Kibish, Ethiopia, KHS Early MSA, Levallois and discoidal cores, tools, 1 cordiform handaxe, exotic white-green unknapped opal silicate stone, sourced 20 km away; similar industry AHS with 9 foliates, 1 handaxe, 2 unknapped opal silicate pieces, and BNS with foliates (Shea, 2008; Shea, Fleagle et al., 2004), KHS early *Homo sapiens sapiens* ~195±5 ka (McDougall, Brown and Fleagle, 2005; Brown and McDougall, 2012); thus at border of MIS 7 (244-190 ka) and MIS 6 (190-130 ka).

With respect to language, this would correlate to ‘Proto-Sapiens Sapiens’ (‘proto-Human’). The homeland of L is generally considered to be East Africa. With respect to a trait list of symbolic behaviors (Harrod, 2014), Omo Kibish exhibits the collecting and transporting—I would say ‘curation’—of non-local exotic stones with an aesthetic quality. This continues a tradition of collection of exotic objects that begins at least by the Acheulian time period.

With respect to NE/E African archaeology circa 190 ka the earliest MSA emerges contemporaneously with Final Acheulian/Sangoan industries, the latter dating from 300 to 150 ka. Final Acheulian/Sangoan sites, more or less securely dated to around 190 ka include:

- Mieso, east-central Ethiopia, Mieso 7 and 31, bifaces, cleavers, ~212±16 ka (de la Torre, Mora et al., 2014; Benito-Calvo, Barfod et al., 2014);
- Sai Island, Nile River, northern Sudan, Site 8-B-11, Level 5, Sangoan, stone circle with 2 slabs with depressions (for grinding?), <223±19 ka, and Level 6, Sangoan, grinding stones, core axes, red and yellow ochre lumps, some with ground surfaces; 10 cm thick sandstone slab, flaked around perimeter, top pecked flat, 10x5cm depression, surrounded by 7 1cm cupules, several chert pebbles with red/yellow ochre adhering, one with black inclusions, suggests symbolic,
>182±20 ka (Van Peer, Fullagar et al., 2003); {JBH: overall shape possible zoomorphic sculpture?].

Cultural interaction between the Early MSA and Sangoan culture traditions is evident, for example the handaxes and collection of exotic stones at Omo Kibish.

L0: ‘Pre-Khoisan’ and L1’2’3’4’5’6: ‘Pre-Niger-Congo-Kordofanian-Nilo-Saharan-Afroasiatic’. At around 160,000 years ago, L-mtDNA (‘Eve’) branched into L0-mtDNA, which strongly correlates to the click language family of Sandawe-Hadza and later southern Africa bushmen, and L1’2’3’4’5’6-mtDNA (Figure 1). In my review I have found no current population samples still expressing the L1’2’3’4’5’6 haplogroup, nor L0, but only later subclades of the two branches. Aside: based on current population samples Cruciani, Trombetta, Massaia et al. (2011) places root Y-DNA emergence in central and northwest Africa.

Archaeologically, the TMRCA of L1’2’3’4’5’6-mtDNA would correspond to MIS 6 sites; some more or less securely dated include:

- Bir Tarfawi, Western Desert, Egypt, White Lakes, Bed 9, 150-160 ka (Wendorf, Schild et al., 1994) or ~175 ka (Van Peer, 1998)
- Benzu, near Tangier, Morocco, L3b, Mousterian Levallois, 173±10 ka (Ramos, Bernal et al., 2008)
- Ifri n’ Ammar, northeast Morocco, Lower O1, MP without tanged tools, 171±12 ka (Richter, Moser et al., 2010)
- Jebel Irhoud, southern Morocco, 4 MNI Homo sapiens sapiens, closest morphology though slightly more primitive features to Skhul-Qafzeh, associated with Levallois Mousterian tools, 160±16 ka (Smith, Tafforeau et al., 2007);

These sites suggest that L1’2’3’4’5’6 spread across both central Africa and northern Africa, though this short list of sites does not seem to suggest which area is the homeland and to which it spread.
Final Acheulian/Sangoan industries continue to occur during this time period and presumably interacting with Early MSA peoples. More or less securely dated are:

- **Herto, Bouri Formation, Ethiopia, 154±7 to 160±2 ka** (Clark, Beyene et al., 2003), Final Acheulian tools, obsidian sourced 289 km away (Negash, Brown and Nash 2011), suggesting complex social trade, exchange, gift-giving (Sahle, Hutchings et al., 2013), 3 MNI early *Homo sapiens sapiens* (McCarthy and Lucas 2014) each with defleshing cutmarks and juvenile bone polishing indicative of mortuary ritual but not cannibalism, and cutmarks similar to those on New Guinea skulls (Clark, Beyene et al., 2003);
- **Sai Island, Nile River, northern Sudan, Site 8-B-11, Level 4, Sangoan, >152±10 ka and <182±20 ka, Level 5** (Van Peer, Fullagar et al., 2003);

With respect to language, L1'2'3'4'5'6-mtDNA appears to correspond to the ancestor of all ‘non-click’ language families. Linguists have grouped languages of the central African corridor, and including some Saharan languages, as a superfamily, variously designated the ‘Macro-Sudan belt’ (Güdemann, 2011; Clements and Rialland, 2008), ‘Niger-Kongo-Nilo-Saharan’ (Dimmendaal, 2001; Bender, 2000), ‘Niger-Saharan’ (Blench, 1995); ‘Kongo-Saharan’ (Gregersen, 1972), ‘Nuclear African area’ (Greenberg, 1983, 1963, 1959), or ‘Sudansprachen’ (Westermann, 1949, 1935, 1911). For the language family correlate of L1’2’3’4’5’6 I suggest the term ‘Pre-Niger-Congo-Kordofanian-Nilo-Saharan-Afroasiatic’.

With respect to mythostratigraphy Witzel (2011) proposed a chronological sequence from Pan-Gaea to Gondwana to Laurasian myth strata. To which stratum should one assign the mythologies of click-speakers? If ‘L-Eve’ is taken as the origin-point for current world mythologies, and it is designated Pan-Gaea, then L0-‘click’ and L1’2’3’4’5’6-‘non-click’ mythologies would appear to be the earliest branches of L-mtDNA which might serve for reconstructing Pan-Gaea.

Based on limited evidence from only one site, Herto, one of the earliest traces of *Homo sapiens sapiens* symbolic (proto-religious) behavior appears to be the ritual defleshing and veneration of bones. This may be interpreted as some sort of cult of the ancestors and their bones, a behavior that in recent times has been widespread across central Africa as well as out of Africa, and especially across southern Asia into Sunda-Sahul. Herto appears to continue an Acheulian tradition, where cutmarks on the Bodo Homo erectus skull appears to have been caused by intentional postmortem defleshing (White, 1986).

Archaeogenetics further indicates that southern African click speakers (‘Khoisan’) diverged from the East African click speakers ≥ 100 ka, and subsequently the northern and southern Africa Khoisan split occurred between 25 to 43 ka (Schlebusch, Skoglund et al., 2012, compare dates in Tishkoff, Gonder et al., 2007). This correlates with their arrival in southern Africa, as evidenced by findings that San material culture (bow-and-arrow, poisons, tool-and-symbol kit) is fully represented at Border Cave, South Africa by around 44 ka (d’Errico, Backwell et al., 2012). Based on its TMRCA, L0f in Sandawe but not Hadza appears to be a token of the split between Sandawe and Hadza by around 70 ka. Those who argue that human language suddenly emerged 25,000 or even 50,000 years ago must explain how it is that southern African Khoisan and eastern African Hadza-Sandawe both have click languages, which are generally accepted as related, if in fact the two populations separated over 100,000 years ago.

(Figure 2). L2'3'4'5'6 mtDNA has a TMRCA between 149±33 ka (Soares, Ermini et al., 2009) and 139±10 (Behar, van Oven et al., 2012). L1 has TMRCAs between 140.6±33 ka to 128.5±11.1 ka. Thus, the emergence of the two branches occurred during the later MIS 6 (190-130 ka). Highest frequencies of L1 occur among Niger-Congo speakers, for example, Mbenga western pygmies of the Republic of Congo and Central African Republic, for whom DNA samples range from 74% to 100% L1c, which is the earliest dated subclade of L1 (TMRCAs ranging from ~78 to 102 ka). The highest frequency of L1c among non-pygmies occurs among Nzébi, southwestern Gabon, who are Bantu speakers (47%). Lesser rates of L1c occur among the Gabon Fang (29%) and South Cameroon Bassa (24%), Ewondo (21%) and Bakaka (14%), along with varying frequencies of L1b (TMRCA ranging from ~4 to 30 ka, homeland possibly Central Africa), all currently Bantu-speaking populations. Thus there appears to be a robust association between L1 and the Niger-Congo language family. The L1 and L2'3'4'5'6 split thus appears to correlate to the differentiation of a ‘Pre-Niger-Congo-Mande-Kordofanian (Katla-Rashad)’ family from a ‘Pre-Nilo-Saharan-Kadu)-Afroasiatic family’.

With respect to L1, Niger-Congo speaking groups having only L1c have zero L0, suggesting the exclusion of Niger-Congo origins from ‘click’ origins. Studies of Mande speaking groups indicate low frequencies of L1c and strong admixture of L2 (L2>L1cb), which may suggest this is a branch of Niger-Congo emerging later than the Atlantic branch. My review found no studies of mtDNA Hg frequency samples for Katla, Rashad or Kadu that might help ascertain their correlations to the mtDNA phylotree.

With respect to archaeology, ‘Pre-Niger-Congo’ appears to be strongly associated with dates and geographic location for the African Early MSA Lupemban tool-and-art package. Lupemban is considered the regional MSA style of central Africa (Taylor, 2011). On the role of African rainforests in early human dispersals see Mercader (2002). Sangoan Final Acheulian and Lupemban MSA sites occur in Central and West Africa including the Congo Basin, but to date none are securely dated. For the ~140 ka time period, the nearest dated Lupemban sites are:
MOTHER TONGUE

Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

- Twin Rivers Kopje, Zambia, F-block, Lupemban, choppers, lanceolate bifaces, picks, handaxes, with abundant pigments of multiple colors and a grindstone with incised grooves. ~139 to <178 ka (Barham, 2012; Barham, 2002a; Barham, 2002b; Clark & Brown 2001);

- Kabwe, Broken Hill Mine, Zambia, Lupemban or MSA, bone tools, red ochre, a red-stained 60-mm spheroid (Barham, Pinto Llona et al., 2002; Clark et al., 1947), dated by fauna to 125 ka and AAR on bone to 110 ka (Bada et al., 1974), though based on Lupemban ~130 to 300 ka, (Barham, Pinto Llona et al., 2002), MNI 3 ‘archaic’ sapiens fossils, though association to the tool industry is not yet definite (Barham, Pinto Llona et al., 2002; Trinkaus, 2009).

The TMRCA for L2'3'4'5'6-mtDNA around 140 ka corresponds to more or less securely dated early Nubian sites in E/NE Africa such as:

- Sai Island, northern Sudan, Site 8-B-11, Upper Levels 1, 2, 3, Nubian Complex MSA with Lupemban features, <152±10 ka, MIS 5, overlying Sangoan (Van Peer, Fullagar, et al., 2003);

- Taramsa 1, Qena, Upper Egypt, EMP Activity Phase I. Lupemban, lanceolates, foliates, mostly non-Levallois, also Levallois with Nubian ridge, between min. ~117 and max. ~166 ka (Vermeersch, Van Peer & Paulissen, 2010);

- K’one, Ethiopia, Nubian, Centripetal and Bidirectional Levallois, ~140 ka (Rose, Usik, et al., 2011; Kurashina, 1978);

and to the earliest evidence for Aterian in Africa:

- Ifri n’ Ammar, northeast Morocco, Upper Ol, Aterian MP with tanged tools, earliest evidence for Aterian in Africa, 145±9 ka (Richter, Moser et al., 2010);

and perhaps to the ambiguously dated:

- Gademotta, Ethiopian Rift, Ethiopia, Unit 13, ETH-72-6, Early MSA, obsidian, mostly Levallois reduction, tools mostly foliate/convergent points, denticulates, 4 Nubian Type 1 cores; 1 exotic retouched flake, the largest in the entire assemblage and the only lithic in basalt in this assemblage and all the others from Gademotta and Kulkuletti and suggesting group mobility (Douze, 2012); date between Unit 12, 260±7 ka, and Unit 15, 105±1 ka; perhaps maximum closer to Kulkuletti Unit D Tuff 185±5 ka (Sahle, Hutchings et al., 2013), though I infer that the Nubian Type 1 cores suggest a maximum ~150 ka. I note that the photo of the exotic basalt retouched flake (Douze, 2012: Fig. 70) appears to have features, whether natural or due to flaking, that appear both zoomorphic and anthropomorphic, and if so would make it even more exotic if not symbolic.

With respect to language family, I suggest that these sites would correspond to the ‘Pre-Nilo-Saharan-(Kadu)-Afroasiatic’. My review seems to indicate that there are as yet no sites in Southwest Asia around 140 ka that have tool industries evidencing affinities to Africa. On the other hand there is evidence for an early presence of robust Homo sapiens sapiens.


With respect to mythostratigraphy, given my correlation of L1 and Pre-Niger-Congo, I suggest that a search for the earliest Niger-Congo and L1-mtDNA associated myth-rituals might focus on those of Western Pygmies and the Nzebi of Gabon, who have only and high frequencies
of L1c, rather than populations with other L1 subclades, such as L1b and L1a1a, which have significantly later TMRCAs and predominate among Bantu and Mande speakers. My master database shows several Bantu-speaking tribes who have L1c and L1b and for whom L1c is greater than the Bantu markers L1a1+L1b. These include the Fang (Fan), Bassa, Ewondo and Bakaka in the Gabon and South Cameroon area, which is part of the L1 homeland. With the caveat that they have undergone some degree of Holocene Bantu-expansion genetic admixture, their mythologies may still retain survivals of the earliest myth stratum. On the other hand, I would not suggest using any other Bantu tribes in my review database to reconstruct the Middle Paleolithic myth stratum for L1-mtDNA populations. In any event, given the genetic and archaeological correlations discussed above, including the lack of evidence for out-of-Africa diffusions at this stage of sapiens sapiens evolution and an L1 homeland in Central and West Africa, I would not suggest using Bantu-expansion mythologies to reconstruct out-of-Africa mythological systems. I leave for future research mythological analysis of Western Pygmy and Nzebi mythology to reconstruct the L1 myth stratum, and only note in passing the former’s focus on assuring the benevolence of ancestor spirits and game spirits and lack of a creator or high god.

With respect to mythostratigraphy, my mtDNA and archaeology review does not suggest to me any evidence that might correlate specifically to L2’3’4’5’6 myth structures and so a minimal presumption would be that people associated with this haplogroup may have continued mortuary rituals similar to Herto, and given tool industries with Lupemban features at Sai Island and Taramsa, probably an associated ancestor cult with some similarity to that of Niger-Congo speaking Western Pygmy and Nzebi.

L5: ‘Pre-Central Sudanic’ and L2’3’4’6: ‘Pre-Nilo-Saharan-Afroasiatic (minus Central Sudanic)’. Next around 120,000 years ago, L2’3’4’5’6 bifurcated into L5-mtDNA and L2’3’4’6 (Figure 3). L5 has TMRCAs ranging from ~106 ka to 138 ka; L2’3’4’6, ~111 ka to 115 ka, and thus both lineages emerge during MIS 5e/d (130-106 ka).

L5 appears to have highest frequencies in central Africa, among several Eastern Central-Sudanic speaking populations: Mbuti pygmies (15%) (Tishkoff, 2007 citing Vigilant, 1991) or (10%) (Quintana-Murci, Quach et al., 2008) and Lugbara, West Nile, Uganda, originally from Sudan (14%) (Isabirye, 2010). While the Lugbara sample contains no L0 and may be taken as a proxy for L5, I note that their mtDNA sample shows various admixtures (14%L1*, L1b + 43%L2*, L2a + 21%L3 + 7%L4). Further, Eastern Pygmies, such as Mbuti, Sua and Asoa have zero L1-mtDNA indicating their prehistory is genetically distinct from western pygmies. With respect to Mbuti, a caveat is that the they have 25-30% L0a and thus Tishkoff, Reed et al., (2009) observe that the Mbuti share genetic ancestry with Khoisan peoples, suggesting they lost their initial click language and adopted Central Sudanic. Though linguists such as Ehret (2011) and Blench (2014; Drake, Blench et al., 2011, Bender, 1997) generally class Central Sudanic as Nilo-Saharan and reconstruct it as later than other Nilo-Saharan branches, my review suggests that ‘Pre-Central-Sudanic’ emerged first, prior to the other ancestral Nilo-Saharan languages. (Note. mtDNA samples for Western Sudanic Bongo-Bagirmi speakers in Chad, the Laka and Boulia, are predominantly L2a and show no L5, the Sara only 5%, and thus they may be viewed as genetically—as well as linguistically—distinct from the Central Sudanic Lugbara and Eastern Pygmies.)
Figure 3: ~120 ka, MIS 5e/d (130-106 ka), Stage 3 Africa mtDNA Map. L2'3'4'5'6 branches off L5 in Central Africa (120x1.1=132 ka) yielding L2'3'4'6 (115x1.1=126 ka). L2'3'4'6 appears to correlate to a first Middle Paleolithic out-of-Africa diffusion, e.g., Abdur Reef, 125±7 ka, and Jebel Faya, U.A.E., ~112 or 123 ka; and Early North African Aterian/Mousterian, Sahara wet-phase, MIS 5e, sites dating 117-125 ka, may have participated in MIS 5e dispersal out-of-Africa as well as material exchanges with Tabun C peoples at Skhul.

As for Central African archaeological sites more or less securely dated to MIS 5e/d that correspond to the L5 TMRCA, my review has only turned up one site, Mumbwa Caves, though given its location the people who left their remains may have had a genetic admixture of L1, L5 and L0.

- Mumbwa Caves, Zambia, Unit VIII: MSA, specularite and red and yellow sandstone for pigments; hearths, windbreak with probably natural ‘anthropomorphic’ dolomite piece, ~130-170 ka, OIS 5e; Unit VII, MSA with multiple worked pigments (hematite, specularite, limonite, sandstones), 2 heat treated with ‘crayon’ shape, 107±11 to 130±6.2 ka, MIS5d (Avery, 2003; Barham & Debenham 2000; Barham, Pinto-Llona & Andrews, 2000; Barham, 2000, 1995; Watts, 2009) or Bayesian reanalysis 75-148 ka (Millard, 2008) [JBH: average ~111 ka. MIS 5d].

With respect to mythostratigraphy, I suggest that Lugbara mythology along with that of Eastern Pygmies would appear to be the best bet for retaining survivals of the earliest ‘Sudanic’ myth stratum as well as retaining aspects of material symbolic behavior evident in the Mumbwa MSA, such as colored pigments and putative natural anthropomorphic stones. I note that the divine power of Lugbara mythology (Middleton, 1979, 1960) is neither a high god or deus otiosus, but a polarity of sacred power, manifesting both as Adroa, a power manifesting in the sky, weather, lightning and associated spirits, and as Adro on the earth, in rivers and streams, a being whose body is cut down the center and who has only one eye, ear, arm and leg and terrible to see. As odd as this may seem to those of a monotheistic or dualistic bent, based on Harrod (2010) I suggest
that this Lugbara theology retains strong features of the Late Acheulian symbolic meme, and conversely supports the proposed interpretation of that meme. Like the Western Pygmies, the Eastern have neither a high god nor a deus otiosus, but unlike them have no rituals for ancestors; they do have anthropomorphic game spirits who live in the forest, which is the source of life and where the dead dwell (Sawada, 2001-03). Both the Western Pygmy and Lugbara mythologies appear to place strong emphasis on ritual engagement with life-giving and life-animating forces.

As for L2’3’4’6, the logic of the mtDNA phylotree and language correlations suggest it would associate to a ‘Pre-Nilo-Saharan-Afroasiatic (minus Central Sudanic)’ language family. NE/E African archaeology sites during Stage 3 (MIS 5e/d, 130-106 ka), which might correlate to ancestral populations bearing the L2’3’4’6, include multiple regionally distinctive tool industries including Aterian, Nubian, and ‘Levallois-Mousterian’. Across North Africa the ‘Aterian sensu lato’ tool industry most frequently occurs during Sahara wet phases, MIS 5a 75-85 ka, MIS 5c (Brorup) 98-110 ka, and MIS 5e 117-125 ka, and less frequently before or after these phases. Aterian sites more or less securely dating to MIS 5e/d (130-106 ka) include

- Dar-es-Soltan II, Morocco, layer 7, Aterian, with ‘enigmatic heap of sandstone slabs 1 m in diameter and 30 cm high’, 121 ka, MIS 5e (Bouzouggar, Barton, et al., 2012; McBrearty and Brooks 2000; Debénath, 1994)
- Dar-es-Soltan I, Unit G1, Morocco, MSA, ~126-130 ka, MIS 5e (Barton, Bouzouggar et al., 2009);
- Bir Tarfawi, southwestern Egypt, Aterian, wet-phase MIS 5e and also later 5c/a (Szabo, Haynes and Maxwell, 1995).

Early North African ‘Aterian’ peoples appear to have engaged in material exchanges with Tabun C industry peoples at Skhul during MIS 5e/d, and possibly there were actual out-of-Africa migration via the Sinai during this stage.

- Skhul, Israel, Layer B, MP Tabun C industry, *Nassarius gibbosulus* shells perforated in similar manner to Aterian beads (Vanhaeren, d’Errico et al., 2006), pigments selected for yellow, orange, and red hues, 3 pieces intentionally heated to change color from yellow-orange to red (d’Errico, Salomon et al., 2010); 10 MNI H. *sapien sapiens* in shallow burials, S5 with wild boar mandible (Belfer-Cohen & Hovers, 1992; Bar-Yosef & Vandermeersch, 1993); (ESR-Useries) 100-130 ka (Grün, Stringer et al., 2005); (TL) 119± 18 ka (Mercier & Valladas, 1994).
- Oued Djebbana, Bir-el-Ater, Algeria, the Aterian type site (so far only 14C date >40 ka) evidences perforated shell beads, which were recently reanalyzed; they have the same perforation pattern as Skhul beads, inferring trade exchange or common ancestry across North Africa into SW Asia (Vanhaeren, d’Errico et al., 2006). Further, Hublin and McPherron (2012) and Hublin and Klein (2011) noting similarities in *sapiens sapiens* craniodental fossils at Skhul, Qafzeh and ~40 ka sites Peștera cu Oase, Romania, and Nazlet Khater, Egypt, argue that northern Africa must also be considered as a possible source for the modern human expansion.

Paleontology combined with the Oued Djebbana and Skhul shell bead production similarities imply a genetic and cultural exchange continuum around the Mediterranean across North Africa, the Levant and even into Eastern Europe.

In addition to Aterian, NE/E African Nubian and ‘Levallois-Mousterian’ MSA sites during Stage 3 (MIS 5e/d, 130-106 ka) would also correlate to ancestral populations bearing the L2’3’4’6. Sites more or less securely dated include:

- Sodmein Cave, Quseir, Eastern Desert, Egypt, Early Nubian MP, 118±8 ka (Mercier, Valladas et al., 1999; Van Peer, 1998);
- Nazlet Khater, Lower Nile, Upper Egypt, Nubian Complex Mid-MP, ~110 ka (Van Peer, 1998);
MOTHER TONGUE

Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

- Kharga Oasis, Western Desert, Egypt, Mata’na Site G, Bulaq Wadi 3 Locus 1 and MSA Workshop sites, with Levallois-Khargan and Nubian industries, ~110-129 ka (Smith, Hawkins, et al., 2007; Smith, Giegengack, et al., 2004);
- Bir Tarfawi, southwestern desert, Egypt, BT-A Grey Phase W1, ~125 ka, and Grey Phase W2, ~115 ka, arguably Nubian (Van Peer, 1998), or Levallois Mousterian (Schild, comment on Van Peer, 1998);
- Taramsa 1, Qena, Upper Egypt, EMP Activity Phase II, Nubian and Levallois, min. ~89 and max.117 ka (Vermeersch, Van Peer & Paulissen, 2010);
- Mumba Shelter, Tanzania, lowest level, Early MSA, with bifacial tools and possible red ochre, ~110-130 ka (McBrearty & Brooks, 2000; Mehlan, 1987);
- Abdur Reef, Eritrea, MSA with handaxes at ~125±7 ka (Walter, Buffler, et al., 2000);
- Jebel Faya, U.A.E., MSA industry with small handaxes and bifacial foliates at (OIS5e; OSL mean of 3 dates ~112 or, eliminating 2 outliers, ~123 ka, which is derived from E/NE African façonnage technique (Marks 2009; Armitage, Jasim, et al., 2011);
- Har Karkom, Negev, Israel, HK190a, HK60, HK61, HK83c, HK23 and HK204, with hutfloors, bifacial tools (handaxes) and Levallois flints, designated ‘Mousterian of Acheulian Tradition’; HK60 with a large circle constructed by alignment of stones, including four small stone piles (Anati, 2006); no dates; HK190a: possible ‘triangular female figurine’ (identified and interpreted James Harrod, during Emmanuel Anati Har Karkom Expedition, April 1996);

In addition to material exchanges between North African Aterian and the Levant (Skhul) peoples, the former bearing presumably bearing L2’3’4’6, Abdur Reef and Jebel Faya suggest a second MSA industry with handaxes, presumably bearing a similar haplogroup, participated in the first wave dispersing Homo sapiens sapiens out-of-Africa around 120,000 years ago. If the Har Karkom sites were to be dated, they may fall in the same time period as Abdur Reef and Jebel Faya, and if so, provide further evidence and confirmation for this MSA-with-handaxe diffusion out of Africa. While the Aterian material, if not genetic, exchange apparently diffused around the Mediterranean coast and/or Sinai, the MSA-with-handaxe peoples may have crossed over the Sinai, around the coast of the Red Sea and/or over the Bab el Mandeb.

While the Levantine Skhul hominins were traditionally thought to be a dispersal ‘dead end’, this seems dubious if dates for Homo sapiens sapiens ‘with robust features’ in South Asia and East Asia are accepted.

- Zhirendong, Hejiang, Guangxi, South China, no tools, H. sapiens sapiens, with robust features (Th/U-series), flowstone over hominin, min. ~106±7 ka (Liu, Jin, et al, 2010; Jin, Pan, et al, 2009)
- Huanglong Cave, Yunxi, Hubei, China, retouched flake and bone tools, H. sapiens sapiens teeth, with a few archaic features (TIMS and ICP-MS) min. ~81±1 ka to max. ~101±1 ka (Shen, Wu et al 2013; Liu, Wu, et al, 2010)
- Callao Cave, Luzon, Philippines, demonstrates ability to make open ocean crossings, gracile Homo sapiens (U-series) min. ~67±1 ka (Mijares, Détroit, et al, 2010). This date seems too early to be associated with H. sapiens sapiens with R-mtDNA in SE Asia by ~58 ka or N by ~56 ka;
and therefore may associate to the L2’3’4’6 First Diffusion Out-of-Africa or the L3’4’6 Second Diffusion Out-of-Africa?

Plausibly, if not a multi-regional evolution, the industries and fossil sapiens sapiens at these sites would seem evidence for the First Diffusion Out-of-Africa extending into East and Southeast Asia.

While there is not yet dated evidence for Nubian Complex, out-of-Africa sites during this time period, future research might verify they also participated in the first wave out-of-Africa. Thus, based on paleontology, archaeology and genetic dates, L2’3’4’6 appears to correlate to the first wave of diffusions out-of-Africa associated with at least two tool-and-symbol kits, a North African MSA Aterian and a NE Africa MSA-with-handaxes.

Given the archaeogenetic correlations, with respect to mythostratigraphy I would hypothesize that the North African Aterian and NE African MSA-with-handaxes peoples who dispersed in this first wave out-of-Africa, whether by the Sinai route, Bab route or circum-Red Sea route, bore mythological-ritual systems that would have been either distinctive or variants of each other and that would have combined features that survived among later Nilo-Saharan (non-Central Sudanic) and Afroasiatic peoples. If this First Diffusion out-of-Africa with North African Aterian-like and NE African MSA-with-handaxes industries reached South Asia, East Asia, Southeast Asia and the Sahul as the above archaeological sites suggest (see also Fleming, Zegura, et al, 2013) this supports the argument for out-of-Africa ‘Gondwana’ mythological systems across southern Asia into Sahul (Witzel 2011), and also suggests that such systems may have had two or more sub-components corresponding to these two Middle Paleolithic tool industries.

On this point I would note, tentatively, some similarity between the Dar-es-Soltan II Aterian stone construction, ‘enigmatic heap of sandstone slabs 1m in diameter and 30 cm high’, and the, yet to be dated, Har Karkom HK60 MP of Acheulian Tradition site with ‘large stone circle constructed by alignment of stones including four small stone piles’, and also the later (MIS 5a) stone pile constructed with sophisticated color and spatial symmetries at Ain El-Guettar, Tunisia, Units 16 and 17, classed as ‘Aterian’ or ‘Mousterian with bifacial foliates and rare tanged points’ (Gruet, 1954), described in more detail below. Possibly this type of Middle Paleolithic symbolic behavior, stone arrangements, is the precursor of such practices in Australia and elsewhere.

L2: ‘Pre-Lake-Palaeo-Chad-Afroasiatic’ and L3’4’6: ‘Pre-Nilo-Saharan’ or ‘Pre-Proto-Nilo-Saharan’. The next major stage of mtDNA evolution occurs around 100,000 years ago during the MIS 5c/b Saharan wet phase (106-85 ka or 106-95 ka based on nomenclature). L2’3’4’6 branched off L2 and L3’4’6. L2 is the most common mtDNA haplogroup in Africa. The L2 TMRCA ranges from ~88 to 104 ka; the L3’4’6, ~71 to 105 ka. (Note. With respect to L3’4’6, if one conservatively multiplies the TMRCA dates by 1.1 to correct for more recent dating of the chimp/human split date, the result is ~78 to 115 ka, and average 97 ka, which places it squarely in MIS 5c/b.)
The L2 subclade L2a’b’c’d emerged ~84 ka (or x1.1 ~92 ka). Its homeland appears to be West or Central Africa (Tishkoff, Gonder et al., 2007). Subsequently, listing TMRCAs by chronological order, L2a (~79 ka) has high frequencies across North Africa and in Central Africa, e.g., Mbuti, Central Africa (65%); Tuaregs, Mali (27%) and Tunisian Berbers (14%). For these population samples L2a frequencies strongly exceed L1 and later L3 frequencies. If one deletes from these samples back-migrations from SW Asia or Europe mtDNA haplogroups and considers only L haplogroups, the L2a frequencies for Mali Tuaregs is (59%) and Tunisian Berber and Arab speakers sampled in several studies (63%, 60%, 52%, 47% and 46%). While having L2a frequencies less than L3 but greater than L1, L2a also has moderately high percentages among Bongo-Bagirmi Western Central Sudanic speakers currently residing around Lake Chad, including Laka/Kabba Laka (29%), Sara (26%), Boulala (18%). Also may be mentioned here are Songhai speakers who have L2 (20%) but this frequency is less than their frequency of L1 and L3. L2b’c’d dates ~57 ka. L2b has high frequencies in West Africa and across sub-Saharan Africa, L2c in West Africa, and L2d in West Africa and into East Africa. L2e (~47 ka) shows high frequencies in around Lake Chad: Buduma/Yedina, currently Chadic speakers (17% L2e and 30% L2a, b and c) and lesser among Fulani and Fali, North Cameroon, currently Niger-Congo speakers (3-5% L2e).

Rather mixed Palaeo-Chad populations with mtDNA samples that include L2a and L2b, L2d or L2e with L2 greater than L1 or L3 include Mande-speaking groups (North Samo,
Mandinga), Senegambian speakers (Fulbe), Chadic (Buduma/Yedina, west of Lake Chad) and Western Saharan Nilo-Saharan (Kanembu, north of Lake Chad). Other mixed L2 subclade groups with L2 less than L3 include Hausa and Kanuri, southwest of Lake Chad. Groups with no L2a but only other southerly originating L2 subclades include the Dogon (67% L2b and c), who speak a language of controversial classification, perhaps an ‘early diverging’ Niger-Congo (‘Mande-like’, tonal but no noun class system) and Niger-Congo speaking Tcheboua Fulani, North Cameroon (21% L2b, c and e).

A look at the locations of subclade frequency peaks seems to me to indicate a radiation from a Saharan Lake Palaeo-Chad homeland (Table 3).

Table 3. Apparent radiation of subclades from L2 homeland, Lake Palaeo-Chad.

OSL dates for Lake Palaeo-Chad range from 97.7±6.5 (Fezzan Basin) to 125±12 ka (Bama Ridge, NE Nigeria), including 119±10 ka (Fezzan Basin), 114±14 ka (Bama Ridge), 108±9 ka (Fezzan Basin), and 107±8 ka (Fezzan Basin) (Drake, Blench et al., 2011). L2 appears to correlate with what might be termed ‘Pre-Lake-Palaeo-Chad-Afroasiatic’ (L2a), with later populations admixed with L2b’c’d and adopting Niger-Congo or Central Sudanic languages from south of the Sahara.

As noted earlier, the North African ‘Aterian’ industry sensu lato most frequently seems to occur during Sahara wet phases, MIS 5a 75-85 ka, MIS 5c (Brorup) 98-110 ka, and MIS 5e 117-125 ka, and less frequently >50 ka and <150 ka. Based on archaeology and genetics dating, L2 would seem to correlate to an Aterian industry during MIS 5c and L2a to the Aterian during MIS 5a (75-85 ka). Aterian sites more or less securely dated to MIS 5e/b (110-85 ka) include:

- Dar-es-Soltan I, Unit G2, Morocco, Aterian, with two ivory objects, one point-like and a plaquette, ~106-119 ka, MIS 5c (Barton, Bouzouggar et al., 2009);
- El Mnasra, Témara, Morocco, Layers 11 and 4base-7, Aterian, with fossil Homo sapiens sapiens, bone tools, hearths, worked hematite, Nassarius shells, many perforated beads, comparable to Taforalt and other North African and Levantine beads, 109±3 ka [MIS 5c]; Level 4 Upper, Aterian, ~95±9 ka [MIS 5b] (Jacobs, Roberts et al., 2012);
- Grotte des Contrebandiers, Témara, Aterian, Nassarius shells, perforated beads (d’Errico, Vanhaeren et al., 2009), ~96-107 ka, mean 103 ka, MIS 5c (Jacobs, Meyer et al., 2011);

Examples of the many North African sites dated by fauna or geology to either MIS 5c or 5a, include:
MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

- Oued Djebbana, Bir-el-Ater, Algeria, the Aterian type site, with perforated shell beads, 14C > 40 ka (Cremaschi, Di Lernia & Garcia, 1998), similar to those at Skhul, 100-135 ka (Vanhaeren, d’Errico et al., 2006);
- Uan Tabu and Uan Afuda, central Acacus, Libya, both Aterian, or Levallois, 65-90 ka (Cremaschi, Di Lernia & Garcia 1998);
- Bir Tarfawi, southwestern Egypt, Aterian, wet-phase MIS 5e and 5c/a (Szabo, Haynes & Maxwell, 1995).
- Ain El-Guettar, Tunisia, Terrace with spring, classed as ‘Final Mousterian with a few tanged points’, pile of 60 spheroids with bones, points in spring (Gruet, 1954); (14C) 47±4 and 57±7 ka, i.e., infinite (Aouadi-Abdeljaouad & Belhoucet 2006);
- Nazlet Khater, Upper Egypt, Denticulate Mousterian (K-group), ~100 ka (Van Peer, 1998);
- Bir Sahara, Eastern Desert, Egypt, BS West Lake Phase 2, BS-11 with tanged points, BS-12 and 13 with Levallois and Nubian, with emphasis on denticulates; OSL-104+10/-13, TL range 84±10 to 109±11 (Wendorf, Schild, et al., 1994; Mercier, Valladas, et al., 1999).

These North African industries include one of the most remarkable palaeoart sites at this time period in Africa. Gruet (1954) excavated at the bottom of a spring a pile of 60 spheroids. The El-Guettar construction had 1 tanged point in the base center of the pile, elongated points near the top, at the apex a flint spheroid with white cortex, flaked black one pole and painted with red ochre at its other pole; and two small stone plaques at its base, one a triangle shape the other a lozenge, and rhino and bovid bones scattered throughout the pile. I suggest that the formal symmetrical design of this circular pile of stones may be viewed as an expression parallel to the matrifocal Mali Tuareg creation myth in which a single being, both male and female, descends from above onto an undefined space, and began to move, first towards the East, then counterclockwise creating pairs of same-gendered and opposite gendered beings at each of the four cardinal directions, and finally establishing a world-axis, the male part ascending upward and the female descending downward, with the female initiating and completing the full journey of construction. Berber families and tribes trace their lineages back to ancestors nearly all women and to the ultimate Creatrix (Claudot-Hawad, 2005). This Tuareg creation myth with its high goddess/god who creates the world-cosmos and its axis using a set of complementary opposites and symmetries stands in strong contrast to the apparently earliest central African mythological world with no high god or deus otiosus and focus on ancestor and game spirits conceived in terms of an all-animating life force. I tentatively suggest that this Tuareg creation myth is a mythological survival from 100,000 years ago, and further that in being so similar to the El-Guettar stone arrangement construction, the Tuareg religion corroborates my attribution of L2-mtDNA as a distinctive signal for North African Aterian and related regionalized ethnic groups and the later pre-Afroasiatic language family.

The MIS 5e stage, in which L2 has its TMRCA, is concurrent with the site of Qafzeh, Israel:
- Qafzeh, Israel, Layer XVII-XXIV, Middle Paleolithic Tabun C industry, 18 MNI Homo sapiens sapiens fossils, 1 adult bone deposition in niche in limestone wall, Q9 adult female and Q10 6-year old b in 'double burial', Q11 in cavity in soft bedrock with fallow deer antler over hands; 10 Glycymeris marine shells not related to food acquisition, 4 perforated, worked ochre (Bar-Yosef & Vandermeersch, 1993); Q8 burial, red ochre, near broken Levallois core with triangular flat surface used as plaque incised with mostly parallel stroke marks (Hovers, Vandermeersch & Bar-Yosef, 1997); minimum of 84 ochre pieces at every level, 6 worked, specific hues selected.
I note the parallel role of the triangular shape as a design element in the symbolic behavior at both the Qafzeh and El-Guettar sites. As I discussed earlier the Skhul site with date range between 100 to 130 ka, if the lower dating holds it would be dated similar to Qafzeh, and the specific style of its perforated beads would indicate an exchange network between Northern African Aterian and Tabun C cultures. If so this would represent one of two or more Second Diffusions ‘Out-Of-Africa Before Out-Of-Africa’ of North African Aterian/Mousterian *sensu lato*. This could be called ‘Second Diffusion-A’. On the other hand there is not yet any well-dated archaeological remains attributable to the Aterian in SW Asia.

The splitting off of L2 yielded L3’4’6, which has a TMRCA around 105 ka (Soares, Ermini, et al., 2009 or ~97 ka, averaging 105x1.1=115 ka and Behar, van Oven, et al., 2012, 71x1.1=78 ka). While L3’4’6 appears lost in virtually all the current Nilo-Saharan (and Afroasiatic) population samples I could find in my review (see my Master Database), there is possibly one exception to this, the Anuak of western Ethiopia, a Luo-speaking Nilotic (or sometimes referred to as ‘Pre-Nilotic’) group (Anuak<Luo<Western Nilotic<Nilotic<Eastern Sudanic<Nilo-Saharan). The Anuak are farmers and herdsmen, unlike other Nilotics who are cattle pastoralists. They have mostly managed to continue practicing their traditional religion. Plaster (2011) contains an Anuak sample (n=108), and specifies not the haplogroup frequencies but all their haplotypes. I proceeded to manually determine the haplogroup for each of these individuals using James Lick’s mthap Version 16.0 (http://dna.jameslick.com/mthap), and then I calculated the haplogroup frequencies for the entire sample. Interestingly the result indicates that the Anuak—and here I calculate the sample frequencies for L haplogroups only—have more L2a,c,e (24%) than L3 (15%) and L2 frequency equal to L4 (24%), with lesser amounts of L0, L5 and L6. Using the mthap converter, one individual had coding that could be read as L2a or as L2a’b’c’d, and another as either L3x or L3’4’6. Granting that I am a novice at converting haplotypes to haplogroups, and relied on the Lick mthap converter, it appears to me that the Anuak sample potentially shows two very ancient haplogroup survivals. From the fact that the Plaster n=108 sample seems to show both (a) Anuak L2 exceeding L3, which is the reverse of typical southern Nilotic peoples, and (b) possible L3’4’6, I infer that the Anuak mtDNA sample reflects a genetic survival of the original divergence of L2 and L3’4’6.

In a recent study of Ethiopian population groups Pagani et al. conducted a sample (n=23) of Anuak and found that on a neighbor-joining tree analysis Anuak were closest to Gumuz, both on a branch distinct from a branch leading to Afroasiatic speaking Ari, Oromo, Amhara, Tigrai and Somali (Pagani, Kivisild et al., 2012: Fig. S2). Thus the neighbor-joining tree in Pagani et al. (2012) and the Lick mthap converter results seem to corroborate each other. Thus I suggest taking the Anuak as descendants of the earliest speakers of Nilo-Saharan and designating the language family correlating to L3’4’6 as ‘Pre-Nilo-Saharan’ or ‘Pre-Proto-Nilo-Saharan’.

I note that Ehret (2011) infers that the origin area of Nilo-Saharan is the eastern parts of the southern Sahara, equivalent to the northern Middle Nile Basin; and Ehret (2011) and Blench (2014) both consider proto-Koman (with late Koman, Gumuz and Uduk branches) as the earliest
branch of proto-Nilo-Saharan. Examining reconstructed subsistence lexicons, Ehret further suggests that proto-Koman, along with proto-Central Sudanic, do not feature horticultural terms, and thus originated with hunter-gatherers. While this may be so, the new genetic typing for a Gumuz sample, as I discuss below, seems to imply that Nilo-Saharan Saharan and Sahelian languages are older than Gumuz.

Archaeologically, during MIS 5c/b stage of evolution, Nubian tool industries occurs in NE Africa and out-of-Africa, with more or less securely dated sites:
- Aybut Auwal, Dhofar, southern Oman, MSA with Nubian Type 1 cores, industry resembling Late Nubian Complex of NE Africa, ~106 ka (Rose, Usik et al., 2011).
- JSM-1, Jubbah paleolake, Nefud Desert, Northern Arabia, Levallois recurrent centripetal, bifacial pieces, some similarity to Dungal and Dineigil Oases, Western Desert, Egypt, latter with Nubian Type 1 cores (Scerri, Groucutt, et al., 2014); Unit-B date 96±9 ka, but stratigraphy ambiguities (Petraglia, Alsharekh et al., 2012; Petraglia, Alsharekh et al., 2011).
- Katotati, Rajasthan. NE Thar Desert, Level S8, MP prepared cores, 1 Levallois-Nubian, retouched tools, bifaces (OSL) 95.6±13.1 ka (MIS 5c) (Blinkhorn, Achyuthan, Petraglia and Ditchfield 2013)

This would presumably represent a diffusion associated with L3’4’6-mtDNA, and I designate it Second Diffusion-B (Nubian Complex) ‘Out-Of-Africa Before Out-Of-Africa’. Whether this dispersal migrated to Arabia via the Bab el Mandeb or the Sinai, or most parsimoniously simply dispersed around the Red Sea coast into SW Asia and South Asia may be decided by future archaeological research.

In sum, the divergence of L2 and L3’4’6 around 100,000 years ago and associated archaeology, genetics and languages appears to correlate to the beginning divergence of some features of the Afroasiatic and Nilo-Saharan language families. It at least marks the emergence of ‘Pre-Proto-Afroasiatic’ and ‘Pre-Nilo-Saharan’, if not the emergence of proto-Afroasiatic and proto-Nilo-Saharan sensu stricto. African language experts such as Ehret (2011) and Blench (Drake, Blench, et al., 2011) argue that the Afroasiatic and Nilo-Saharan language families emerged around 15,000 years ago and possibly correlate to one or another Holocene Epipaleolithic hunter-gatherer industry. The mtDNA PhyloTree and its archaeological, linguistic and mythological correlates call into question such a low date for these two proto-languages, since it looks highly likely that aspects of proto-Afroasiatic and proto-Nilo-Saharan emerged by 100,000 years ago or at least by 80,000 years ago—when L2a separated from the other southern L2 subclades.

With respect to my hypothesis of an emergent divergence of ‘Pre-Lake-Palaeo-Chad-Afroasiatic (L2)’ and parallel branch (L3’4’6) ‘Pre-Proto-Nilo-Saharan’ 80-100,000 years ago, the following comment based on mtDNA genetic distance analyses may be relevant.

A phylogenetic tree of genetic distances from inferred ancestral clusters indicates that within Africa, the Pygmy and SAK associated ancestral clusters (AACs) form a clade, as do the Hadza and Sandawe AACs and the Nilo-Saharan and Chadic AACs, reflecting their ancient common ancestries” (4). "Afroasiatic Chadic–speaking populations from northern Cameroon cluster close to the Nilo-Saharan–speaking populations from Chad, rather than with East African Afroasiatic speakers, consistent with a language replacement among the Chadic populations (Tishkoff, Reed, Friedlaender, Ehret, Ranciaro et al., 2009)."
As noted earlier, L0 branched L0d around 100,000 years ago, found in Sandawe (5%) but not in Hadza who retain only L0a, and which is predominant in Southern African Khoisan (SAK), such as !Kung (96% L0d+4%L0k). Emergence of L0d appears to correspond to the initial split between East African and Southern African click speakers, with the latter beginning their move to southern Africa from East Africa around 100 ka (Rito et al., 2013; Behar et al., 2008) and arriving in southern Africa with a fully developed San material culture by around 45 ka. The beginning of this migration of click speakers with L0d-mtDNA appears to happen at roughly the same time as the L2 split from L3'4'6.

L3: ‘Nile-Sudanic’ (Nubian MSA and ancestral language to Nilo-Saharan Northern-Sudanic Kunama or proto-Koman, and Eastern Sahelian - Nubian language families); L4: ‘Pre-Nilotic’ and L6: ‘Pre-Yemen’. Haplogroup L3'4'6 began branching its subclades around 85,000 years ago (Figure 5) and L3'4 emerged ~83 ka (averaging Soares, Ermini, et al. [2009] 86x1.1=95 ka and Behar, van Oven, et al. [2012] 64x1.1=70 ka). These dates fall within MIS 5a (85-74 ka).

Figure 5: ~80 ka, MIS 5a (85-74 ka), Stage 5 Africa mtDNA Map. Around 83 ka (averaging Soares-Ermini 86x1.1=95 ka and Behar-van Oven L3'4 64x1.1 =70 ka) L3'4'6 branches off L3'4 and around 80 ka L3'4 branches L4, southerly, highest frequencies among pre-Nilotics/Nilotics and Yaaku hunter-gatherers; and L3 (72x1.1=79 ka, a match for 78.3 ka Fu, Mittnik etal., 2013), northerly, highest frequencies in Nilo-Saharan Saharan and Sahelian speakers from the Nile to Lake Chad. L3 branches earliest subclades L3h, ~72 ka (65x1.1=72 ka), highest frequencies among Nilotics and Omotics; out-of-Africa N ~78 (Soares-Ermini 71x1.1=78 ka) and M ~67 (60x1.1=67 ka). Nile Denticulate at Sinai-20 Split Rock, 85±13 ka and again at 62±9 ka; Jebel Qattar. L3 subclades noted at
As for contemporaneous archaeological sites at around 85 ka—and for L3 at 80 ka—this would fall within the MIS 5a (74-85 ka) wet phase. Aterian sites dated to MIS 5a include:

- El Akarit, central coast, Tunisia, Aterian, >80 ka or 90 ka (Reyss, Valladas et al., 2007; Roset & Harbi-Riahi, 2012);
- Dar-es-Soltan I, Morocco, Unit G3, Aterian with foliates, ~68-87 ka, MIS 5a (Barton, Bouzouggar, et al., 2009);
- Grotte des Pigeons, Taforalt, northeast Morocco, Aterian, both with perforated shell beads and other personal ornaments, ~82.5 ka (Bouzouggar, Barton, et al., 2007).
- Ifri n’Ammar, northeast Morocco, Upper OS, 71-95 ka, mean ~83 ka, Aterian with tanged tools, shells, personal ornaments (Richter, Moser, et al., 2010)
- Haua Fteah, Cyrenaica, Libya, lower layers, unique MP industry (Scerri 2012), associated with a fragment of a flute or whistle, initially dated as Eemian by fauna, but OSL ~68-102 ka (Douka, Jacobs, et al., 2013)

Other Mid-MSA industries occur at

- Katanda 9, Upper Semliki Valley, D.R. Congo with bone barbed and unbarbed harpoon points, ~80-90 ka (Brooks, Helgren, et al., 1995; Yellen, Brooks, et al., 1995) or 60-70 ka (Feathers & Migliorini, 2001) [but I note if the highest and lowest outliers of 7 OSL dates is removed, result is ~86 ka];
- Taramsa I-Phase III, Qena, Upper Egypt, Mid-MSA Nubian Complex, Nubian Type 1 cores and points, continuously from ~65 to 84 ka; H.s.s. child burial ~69 ka; (Vermeersch, Van Peer & Paulissen, 2010);
- Bir Tarfawi W4, Western Desert, Egypt, Middle Nubian with Nazlet Khater points and Aterian characteristics, ~70 ka (Van Peer, 1998; Vermeersch, Van Peer & Rots, 2005);
- Aduma, Middle Awash, Ethiopia, Levallois, Aduma and Nubian cores, micro-tools, ~80-100 ka (Yellen, Brooks, et al., 2005; Brooks, 2005).

MIS 5a archaeology attests to a continuation of the preceding Second Diffusion ‘Out-Of-Africa Before Out-Of-Africa’. Key sites that appear to signal this diffusion include:

- Jebel Qattar I, Jubbah paleolake, Nefud Desert, northern Saudi Arabia, Unit B, MP Levallois, ~75±5 ka, and Jebel Katefeh 1, Unit H, ~86±11 ka; classed as Tabun-C Levallois (Petraglia, Alsharekh, et al., 2012; Petraglia, Alsharekh, et al., 2011), but the latter with similarities to Dungal and Dineigil Oases. Western Desert, Egypt, latter with Nubian Type 1 cores (Scerri, Groucutt, et al., 2014), which thus appears to continue a flaking industry found at the site of JSM-1, MIS 5b/c. As Tabun-C and North African affinity sites occur at Jubbah in the same timeframe, this suggests possible cultural exchanges.
- Sinai-20 Split Rock, Zarnoq, about 30 km from Taba on the Gulf of Aqaba, ‘Nile Denticulate Mousterian’, 85±13 ka and again at 62±9 ka (Kobusiewicz, Schild, et al., 2001; Kobusiewicz, 1999; Eddy, Wendorf & Associates1999);
- Gebel Urayf an Naquah, central Sinai, Nubian Type 1 point cores, no date (Schild, 1999; Schild, comment on Van Peer, 1998); located about 24 km from Har Karkom in the Negev Desert with its many MP sites;
Qadesh Barnea, Wadi El-Qudeirat tributary to El-Arish, NE Sinai, Egypt, MP discoid cores (per se insufficient to identify industry), no date, but by analogy to U-series on mid-MP gravels at Nahal Aqev D35, Nahal Zin, Negev) ~80±10 ka (Schwarcz, Blackwell, et al., 1979); but Goldberg (1986) suggests unconformity warrants rough age ~40-90 ka, given MP/UP transition at Nahal Zin dated ~45 ka;

Har Karkom, Negev, site HK148b with MP tools including tanged points, and designated the 'Aterian Hut Site', no date (Anati, 2006).

• The site appears to me to have a floor plan, comparable in dimensions, entrance orientation and artifact arrangements to hut floor plans of Nilo-Saharan Berta, Gumuz, Mao, and other Komuz groups (González-Ruibal, 2006; González-Ruibal, and Fernández Martínez, 2007), and also of Eastern Cushitic Guji-Oromo (van de Loo, 1991).

Considering the above MIS 5a ‘out-of-Africa’ sites it seems that at least two cultures may have dispersed over the Sinai route, one a Nile ‘Denticulate’ Mousterian (at some sites classed as ‘Nile Mousterian K-group’), and the other an MSA-Nubian Complex tradition, and possibly a third tradition with Aterian features (if HK148b ‘Aterian’ were dated to this time period), and these interacted with a SW Asian indigenous tradition having a Tabun C industry. In this light, I infer that L3’4 appears to correlate to both cultures diffusing out-of-Africa around the same time, one carrying Nubian industry, the other a Nile Mousterian, and one or both populations also may have diffused an L2-mtDNA subclade (L2a).

With respect to any correlation between L3’4 and language family, my review has not turned up evidence for L3’4 in current population samples, and thus provides no mtDNA evidence for assignation to any particular language family, other than an early Nilo-Saharan or early Afroasiatic. On the other hand, from the evidence for both Nubian Complex and Nile Denticulate Mousterian tool industries out-of-Africa, an archaeogenetic hypothesis would be to associate the Nubian Complex MSA industry with an ancestor of the Nilo-Saharan proto-Northern Sudanic (Kunama) or proto-Koman (Gumuz, Uduk) language families (applying Ehret 2011 terms) and the Nile Denticulate Mousterian with an ancestor of the Afroasiatic Boreafrasian language family. It is tempting to posit proto-Koman at this stage, given the that current Gumuz have predominantly L3 and L4-mtDNA and apparently the highest frequency of these combined (76%) of Nilo-Saharan speaking tribes in NE/E Africa (see Table 4), but lack of L3h in sample makes this questionable (see further discussion of Gumuz below).

Shortly after L3’4 arose ~83 ka, at around 80 ka, L3 and L4 diverged: L3 ~79 ka, using Soares, Ermini, et al. (2009) 72x1.1=79 ka, which then matches 78 ka Fu, Mittnik et al., 2013) and L4 ~79 ka, using Behar, van Oven, et al. (2012), with no multiplier. Later L6 emerged around 20 ka (Behar, van Oven et al., 2012). L6 has peak frequencies among Yemeni (12% and 26% for L’s only) and is also frequent among eastern Ethiopians. L6 is said to be recent in Yemeni and have its homeland in East Africa (Soares, Ermini et al., 2009; Kivisild, Reitla, et al., 2004). While at first glance it might reflect L3 out-of-Africa, its late date, restricted typology and lack of N and/or M autochthonous lineages in Southern Arabia does not support a ‘southern route’ model for out-of-Africa (Abu-Amero, Larruga, et al., 2008).

L4 occurs only in East Africa, which would thus be its homeland (Soares, Ermini et al., 2009; Tishkoff, Gonder et al., 2007). L4 has highest frequencies among Nilotics, such as Nuer (18%) and Acholi (15%), some Omotic speakers, such as Hamer (18%), some Cushitic, including
Yaaku, Kenya, hunter-gatherers (32%) and Tigrai (13%). This suggests the L4 homeland may be more specifically the Southern Sudan/Omo area.

The homeland for L3, the parent of N and M and all subsequent out-of-Africa mtDNA lineages, is either East Africa or NE Africa; to date genetics does not offer a more specific location (Soares, Alshamali et al., 2012). On the other hand, I suggest the differentiation of L3 and L4 appears to correspond to L4 more southerly in East Africa and L3 more northerly along the Nile corridors. The highest frequencies of L3 occur in Nilo-Saharan Saharan and Sahelian speakers from the Nile to Lake Chad. Such a homeland and L3 TMRCA of ~79 ka would corresponds to in-Africa archaeological sites among those listed above under L3'4, such as:

- Taramsa I-Phase III, Qena, Upper Egypt, Mid-MSA Nubian Complex, Nubian Type 1 cores and points, continuously from ~65 to 84 ka; H.s.s. child burial ~69 ka; (Vermeersch, Van Peer & Paulissen, 2010);
- Bir Tarfawi W4, Western Desert, Egypt, Middle Nubian with Nazlet Khater points and Aterian characteristics, ~70 ka (Van Peer, 1998; Vermeersch, Van Peer & Rots 2005);

With respect to TMRCAs of L3 subclades there appear to be four clusters chronologically, roughly dating 65, 55, 40 and 25 ka. The earliest subclade cluster includes L3h, N and M, with L3h, ~69 ka (average Soares-Ermini 67 ka and Behar-van-Oven 59 ka = 63x1.1=69 ka). Out-of-Africa N dates about ~75-78 ka (Soares-Ermini ‘N in South Asia 71 ka x1.1=78 ka; Soares-Alshamali ~62 ka seems to be a clock violation with their date N→R ‘in South Asia’ 67 ka, and this would imply N actually emerges between 74 and 79 ka). M dates somewhat later at ~67 ka (Soares-Ermini 61x1.1=67 ka). Highest frequencies of L3h occur among Omotics and Nilotics. Root N and M occur only in SW Asia.

To discern if there were any overall patterns in genetic data and current languages that might yield some inferences about the emergence, homeland or language families associated with L3-mtDNA, I reviewed mtDNA studies of East and Northeast African population groups. I reanalyzed published haplogroup frequencies to focus on L-haplogroups only, deleting back-to-Africa haplogroups of Eurasian origin. I selected samples with highest frequencies of L3 and L3h and I note frequency ratios among L3, L2, L4 and L6. Grouping ethnic groups by language, six clusters became more or less apparent, each with fairly distinct haplogroup frequency patterning (Table 4). Boattini, Castri, et al. (2013), using far more sophisticated principle component analysis and model-based clustering techniques, and focusing on Afroasiatic speakers, appear to me to have independently confirmed four of the six clusters.
Table 4: Combined L3≥30% and/or L3h>9% Frequency (based on L5 only, subtracting Eurasian)

<table>
<thead>
<tr>
<th>L0</th>
<th>L1</th>
<th>L5</th>
<th>L2</th>
<th>L4</th>
<th>L3</th>
<th>L6</th>
<th>ratio L3 / L2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early Nilo-Saharan:</strong> L2&gt; L3, L4&gt; L6 (L3h= ~65ka, N and M= ~60ka, ergo o-o-A?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anuak nilotic (White Nile)</td>
<td>12%</td>
<td>15%</td>
<td>24%</td>
<td>24%</td>
<td>15%</td>
<td>8%</td>
<td>inverse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxeb (15%)</td>
<td>1 L3'A'6 or x?</td>
<td></td>
</tr>
<tr>
<td><strong>Saharan and Sahelian:</strong> L3&gt; L2, no or rare L5, L4, L6 (L3h= ~65ka, N and M= ~60ka, ergo o-o-A?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kanuri pro-shr</td>
<td>8%</td>
<td>3%</td>
<td>15%</td>
<td>2%</td>
<td>66%</td>
<td>4x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfedb (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>h (20%), h (17%)d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laal nc+chad</td>
<td>18%</td>
<td>9%</td>
<td>9%</td>
<td>63%</td>
<td>6x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfed (9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(96%, 36%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sara wc sud</td>
<td>16%</td>
<td>5%</td>
<td>5%</td>
<td>26%</td>
<td>37%</td>
<td>1.5x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfedb (5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(11%, 6%, 6%, 6%, b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nubian e shl</td>
<td>26%</td>
<td>7%</td>
<td>7%</td>
<td>22%</td>
<td>1%</td>
<td>35%</td>
<td>1.5x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfed (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(20%, 16%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[JBH: This is the Nile Corridor for out-of-Africa as well as E-W Sahel/Sudanian Savanna Corridor.]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Komuz: L3, L4 but no L1, L2 or L6 (L4= ~80ka; L3f= ~50ka; L3i= ~44ka; L3x= ~35ka; ergo, not o-o-A?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gumuz komuz</td>
<td>12%</td>
<td>12%</td>
<td>24%</td>
<td>52%</td>
<td>infinite</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxdb (no h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(20%), 24%, 8%b</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Semitic, Chadic (Boreafraision):</strong> L3&gt; L2, no L4, L5, L6; no or low L0 (L3h= ~65ka; N, M= ~60ka, ergo o-o-A?) (compare Boattini cluster C) (L3f= ~50ka; L3e= ~40ka; L3x= ~35ka)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egyptian sm Alexandria</td>
<td>10%</td>
<td>25%</td>
<td>16%</td>
<td>50%</td>
<td>3x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxebc (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2%)x, 20%, 12%e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guma</td>
<td>6%</td>
<td>12%</td>
<td>53%</td>
<td>infinite</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(no h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[JBH: This is Northern Africa/Sinai corridor for out-of-Africa.]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cushitic, Semitic:</strong> L3≥L2, L4&gt;L6 (L3h= ~65ka, N, M= ~60ka, ergo o-o-A?) (compare Boattini cluster A) (L4= ~80ka; L3a= ~55 ka; L3f= ~50ka; L3i= ~40ka; L3x= ~35ka; L3d= ~30ka; L3b= ~20ka; L6= ~20 ka)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agaw cc</td>
<td>16%</td>
<td>25%</td>
<td>26%</td>
<td>8%</td>
<td>43%</td>
<td>1.5x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxed (4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(14%), 12%, 6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yemeni sem</td>
<td>15%</td>
<td>15%</td>
<td>35%</td>
<td>38%</td>
<td>26%</td>
<td>2.5x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxed (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afar ec</td>
<td>20%</td>
<td>3%</td>
<td>2%</td>
<td>36%</td>
<td>3%</td>
<td>=</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxed (8%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amhara sem</td>
<td>17%</td>
<td>2%</td>
<td>2%</td>
<td>31%</td>
<td>2%</td>
<td>=</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxd (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[JBH: This is region of Blue Nile Corridor for out-of-Africa and/or via Afar and Bab-el-Mandeb.]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Omotic:</strong> L3&gt;L2, L6&gt;L4 (L3h= ~65ka, N, M= ~60ka, but probably not o-o-A) (compare Boattini cluster B1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maale n om</td>
<td>6%</td>
<td>4%</td>
<td>6%</td>
<td>12%</td>
<td>49%</td>
<td>4x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hafxb (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3%), 36%x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dawro n om</td>
<td>13%</td>
<td>1%</td>
<td>17%</td>
<td>46%</td>
<td>5%</td>
<td>2.5x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hafxb (11%), (4%), 18%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hamer s om</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>30%</td>
<td>10%</td>
<td>1.5x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(30%), (5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[JBH: This is SW ETH, Omo River Basin to Lake Turkana with no corridor for out-of-Africa.]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>L3 Nilotics:</strong> L3≥L2, L4≥L5, no L6 (L3h= ~65ka; L3 subclades moving south) (compare Boattini cluster B2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkana</td>
<td>24%</td>
<td>7%</td>
<td>11%</td>
<td>13%</td>
<td>48%</td>
<td>4x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hafxb (9%), 2%a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samburu</td>
<td>22%</td>
<td>8%</td>
<td>14%</td>
<td>12%</td>
<td>45%</td>
<td>3x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfxdc (9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luo k</td>
<td>22%</td>
<td>8%</td>
<td>14%</td>
<td>8%</td>
<td>45%</td>
<td>3x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hafxb (6%), (2%)a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuer w nilot</td>
<td>10%</td>
<td>15%</td>
<td>20%</td>
<td>20%</td>
<td>36%</td>
<td>1.5x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfeb (15%), 10%e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acholi u</td>
<td>9%</td>
<td>9%</td>
<td>27%</td>
<td>18%</td>
<td>36%</td>
<td>1.5x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>heb (9%), (18%)h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maasai k</td>
<td>27%</td>
<td>12%</td>
<td>11%</td>
<td>17%</td>
<td>33%</td>
<td>3x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hafxb (4%), 7%ax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dinka</td>
<td>12%</td>
<td>8%</td>
<td>6%</td>
<td>30%</td>
<td>14%</td>
<td>4x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfe (14%), 9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Datoga tz</td>
<td>56%</td>
<td>6%</td>
<td>6%</td>
<td>13%</td>
<td>26%</td>
<td>4x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hfb (20%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[JBH: Nuer, northern-central Sudan, and Dinka, Southern Sudan: this is the White Nile Corridor.]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. L3h has geographic diverse sub-lineages requiring caution in relying only on HVS-1 (Rosa & Brehm, 2011)
Note. Gumuz Hgs (with gratitude, Pagani personal communication Dec 2014 re Pagani et al. 2015 in preparation)
Given Table 4 and all the preceding L3 tables, I then searched each cluster for a distinctive subclade above and beyond L3h, such as subclade high frequency or feature distinguishing that cluster from other clusters, which would serve as its distinctive marker. I suggest the markers summarized in Table 5.

Table 5. Hypothetical Correlation of L3 Subclades in Current Population samples to Nilo-Saharan and Afroasiatic Language Families with TMRCA-based Emergence Dates

<table>
<thead>
<tr>
<th>Subclades</th>
<th>Current Population Samples</th>
<th>Relevant Language Families</th>
<th>TMRCA Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>L3h (~65ka)</td>
<td>EGYPT (Semitic A-A)</td>
<td>(Note. No L3h in Beja Northern Cushitic A-A)</td>
<td>(Red Sea Coastal Corridor)</td>
</tr>
<tr>
<td>OMO (Omotic A-A)</td>
<td>(Omo River Basin)</td>
<td>(Saharan and Sahelian N-S)</td>
<td>(Nile Corridor)</td>
</tr>
<tr>
<td>L3h+L3f (~50 ka)+(i)(~40 ka)</td>
<td>NUBIA-SAHARA</td>
<td>CUSHITE</td>
<td>(Central Cushitic, Semitic A-A)</td>
</tr>
<tr>
<td>L3h+L3d (~30 ka)+(i)(~40 ka)</td>
<td></td>
<td>(Note. No L3h in Beja, Oromo, Yaaku Eastern Cushitic, Tigray Semitic or Ongota A-A)</td>
<td></td>
</tr>
<tr>
<td>L3h+L3e (~40 ka)+b (~20 ka)</td>
<td>NILOTIC (Eastern, Western and Southern Nilotic N-S)</td>
<td></td>
<td>(White Nile Corridor)</td>
</tr>
<tr>
<td>L3h+L3a (~55ka)</td>
<td>RIFT</td>
<td>(Eastern Cushitic Burunge and Southern Cushitic Yaaku A-A)</td>
<td>(Rift Corridor)</td>
</tr>
</tbody>
</table>

Method: (a) examine the L3h population sample clusters, identify additional subclade(s) that appear to be markers signaling that cluster, e.g., predominant subclade, high frequency of subclade or other feature distinguishing it from the other clusters.

TMRCA dates are based on Soares, Ermini, et al. (2009) and/or Behar, van Oven, et al. (2012). None of these dates are corrected by a 1.1x or 1.2x multiplier suggested by recent studies arguing for redating of the CHCLA.

Afroasiatic and Nilo-Saharan language associated clusters are supported by different methods of cluster analyses in Boattini, Castri, et al. (2013) and phylotree in Pagani, Kivisild et al. (2012).
Every sample in Table 4, except Guma Egypt has L3f, but L3f’s highest frequencies appear to occur in samples from Nilo-Saharan Kanuri and Nubians. Frequencies of L3i are not very dissimilar between Nubians and Cushitic speakers. L3a appears especially distinctive for Eastern Cushitic Afar at 2% versus 0% for other Cushitic and Semitic populations, and for Southern Cushitic Burunge (16%) and originally Eastern Cushitic Yaaku (26%). Raising a question about the main linguistic speculations for Ongota, Ongota genetics suggests they were originally northern ‘click-language’ speakers, who appear to have later adopted Nilo-Saharan and subsequently some features of Afroasiatic, thus causing considerable exasperation for linguists.

Neither Table 5 nor Table 4 appears to provide any definite clue as to the homeland of L3. On the other hand, if we consider that L3’4 breaks into L3 and L4, the latter in East Africa, this may suggest that L3 has a homeland north of East Africa. Table 5 focusing on L3h suggests that the Omo Basin is either the homeland or remnant of the oldest L3 subclade L3h. If so, I suggest that we may consider that L3M and L3N branched off further north along the Nile Corridor or Sinai Crossing out-of-Africa. The tables appear to provide no strong support for a Bab route out-of-Africa for L3M or L3N.

L3M and L3N out-of-Africa with admixture of L2a: variable combinations of ‘Nile-Sudanic’ and ‘Boreafrasian’ languages (Nubian Complex and Nile Denticulate MSA industries). As their TMRCAs are only a few thousand years apart during MIS 5a (L3’4 ~83 ka; L4 ~80 ka; L3 ~79 ka), what I inferred earlier based on archaeology, genetics and language with respect to L3’4 seems to me equally valid for L3-M and N out-of-Africa. There is evidence for both Nubian Complex and Nile Denticulate Mousterian tool industries dispersing out-of-Africa. A likely hypothesis would be that the Nubian Complex MSA industry correlates to the Nilo-Saharan proto-Northern Sudanic (Kunama) or proto-Koman (Gumuz, Uduk) language families, and/or, by the time of L3 ~79 ka, the Proto-Saharan-Sahelian and Proto-Eastern Sahelian (Nubian) languages, applying linguistic terms from the language phylotree in Ehret (2011). To designate the ancestral language family that appears to correlate to the Nubian Complex industry out-of-Africa bearing L3 or its subclades M and N, whether by a Sinai or circum-Red Sea Crossing, I find it necessary to invent a term, and I suggest ‘Nile-Sudanic’. The Nile Denticulate Mousterian out-of-Africa may correlate to the Afroasiatic Boreafrasian language family, perhaps with a strong Northern African L2a component. In inferring these two population dispersals out-of-Africa and correlated archaeology and language families, I am not equating L3M and L3N respectively to each of the two archaeological tool traditions. Current mtDNA population genetics shows that Nubian populations and Northern African/Egyptian populations both have high frequencies of L2a and L3 mixture, so I suggest inferring that the populations who made the Nubian Complex and Nile Denticulate Mousterian tool industries already had some degree of such haplogroup diversity and perhaps even language diversity.

With respect to the evolution of mythological structures, I would similarly infer that populations dispersing out-of-Africa, whether by the Sinai route, Bab route or circum-Red Sea route, dispersed with myth-ritual systems that combined to greater or lesser extent components of the two major northern and central African myth-systems, which evolved prior to the TM RCA of L3 and which may be termed ‘North African’ (correlating to L2, L2a) and ‘Sudanic’ (correlating to L5, L4) or ‘Nile-Sudanic’ (correlating to L3 and its subclades). The Sudanic system emphasizes ancestral and game spirits, where ‘spirit’ means a life-giving, life-animating and life-enhancing forces or energies, with neither a high god nor deus otiosus; the North African, a creative power,
female or androgynous, who organizes a cosmos with four directions, complementarities and world-axis and establishes the nature of life’s unfolding as one of balance and complementarity of polar or gender opposites. Archaeology and mtDNA genetics suggests that both of these mythological systems dispersed out-of-Africa during MIS 5a around 80 ka.

Again, I do not necessarily associate these two mythological traditions with populations predominantly either M or N-mtDNA, but I do suggest that myth-ritual admixtures may have occurred among these populations, especially as they resided in SW Asia prior to dispersals through the Transcaucasus and Zagros bottlenecks. How and to what extent these two myth-ritual systems might correspond to Gondwana and Laurasian mythological structures (Witzel 2011) is a topic for future research.

M, N and R-mtDNA: Early Borean. The period around 75,000 years ago (early MIS 4 ~74-59 ka) witnessed the divergence of M, N, and R-mtDNA lineages in Southwest Asia. As noted earlier, I date N about ~75-78 ka (Soares-Ermini ‘N in South Asia 71x1.1=78 ka; Soares-Alshamali ~62 ka seems to be a clock violation with their date N→R ‘in South Asia’ 67 ka, and this would imply N actually emerges between 74 and 79 ka). I suggest the date for the branching of R-mtDNA from N at around 74 ka (Soares-Alshamali R ‘in South Asia’ 67x1.1=74 ka). I date M somewhat later at ~67 ka (Soares-Ermini 61x1.1=67 ka). These dates I propose, around 75 ka, using the x1.1 multiplier produce excellent agreement with the archaeology of NE Africa and Arabia, and are similar to recent conclusions of archaeologists and geneticists (Petraglia, Haslam, Fuller, et al., 2010; Reyes-Centeno, Ghirotto, Détroit, et al., 2014).

These dates suggest that the *sapiens sapiens* dispersal out-of-Africa endured a roughly 5,000 year pause in SW Asia, during which N and M differentiated subclades and N branched off R-mtDNA. Two geographic bottlenecks, the Transcaucasus and Zagros Crossing, extant Neanderthal and other archaic species, the Toba supereruption (~74 ka), as well as the positive subsistence landscape of the ‘Persian Gulf Oasis’ (Rose, 2010; Rose, 2007) presumably contributed to this delay. A *sapiens sapiens* ‘fast track’ on some sort of ‘southern route’, e.g., around 60-50 ka (Mellars, Gori et al., 2013) and earlier studies proposing and African LSA out-of-Africa ~45 ka do not appear supported. Recent genetic studies using various methods argue for such a pause in SW Asia (Xing, Watkins, Hu et al., 2010; Amos & Hoffman, 2009). Such a pause in SW Asia before diffusion north and east through the bottlenecks might be compared to the pause at the later Beringia crossing bottleneck.

Reflecting on Fleming’s Borean hypothesis, my meta-review of archaeogenetics, archaeology and language correlations suggests that the three major Borean language families map fairly precisely onto the earliest out-of-Africa mtDNA branches. M, N and R. Eurasian languages are strongly associated with M-mtDNA and its branches; Burushaski-Caucasic-Déné languages with N-mtDNA, and Afroasiatic, Kartvelian, Dravidian, Elamitic, etc. with R-mtDNA. Interestingly, Austric and Pama-Nyungan seem most closely associated with R-mtDNA. Noting the controversial hypotheses for a so-called ‘southern route’ to the Sahul, I suggest that the current mtDNA phylogrete appears to me to imply that Austric and Pama-Nyungan descendants are actually on a descendent branch of R, and that I term Borean-R. Here I leave aside for future discussion questions regarding the extent of Denisovan or erectus genetic admixture, technological regime hybridizing, or admixing with languages of archaic *sapiens*.

The earliest branching of subclades of Borean-N, M and R, which occurred around 70 ka to 60 ka, I designate as ‘Early Borean’. Using the relative chronological order of TMRCAs in Soares, Ermini, et al. (2009) [‘SE09’] or Behar, van Oven, et al. (2012) [‘BO12’], uncorrected for
proposed redating of a CHCLA based mutation clock, the 70-60 ka earliest differentiations of Early Borean (each listed in approximate chronological order) include:

**Early Borean-N** (homeland SW Asia)
- N in South Asia (71.2±16 ka, *SE09*); N → N1’5 (57±5 ka, *BO12*)
- N in W Eurasia (61.9±11 ka, *SE09*)
- N in East Asia (58.2±14 ka, *SE09*)
- N in SE Asia/Sahul. N → N11 (56±4 ka, *BO12*)

Based on these dates N appears to have taken about 15,000 years to disperse over South Asia, Western Eurasia, East Asia and SE Asia/ Sahul, and in that geographic order. No simple ‘rapid coastal southern route’ is evident here, and, apparently ironically for that hypothesis, dispersal into Western Eurasia occurs 5,000 years prior to SE Asia. I explore more evidence for this and for my hypothesis that N correlates predominantly with the Burushaski-Caucasic-Dene macrofamily when I further discuss N subclades below.

**Early Borean-R.** N → R (geographical homeland not yet determined, 56.5±2.1 ka, *BO12*; but see earlier TMRCAs below) (Fossil DNA, R*, Ust’Ishim, Irtysh River, Russia, ~43-47 cal ka; see below)
- R in South Asia (66.6±14 ka, *SE09*)
  - R → R31 (64.5±14 ka, *SE09*; 54.9±3.1 ka, *BO12*), with frequency peaks in Rajasthan and Sri Lanka (Karmin, thesis, 2005)
  - R → R30 (64.0±15 ka, *SE09*; 53.6±4.0 ka, *BO12*), with frequency peak in Gujarat, NW to Central India (Karmin, thesis, 2005); R30 → R30b (51.0±4.7 ka, *BO12*), with highest frequency in Vedda, Sri Lanka (Ranaweera, Kaewsuthi et al., 2014)
- R in W Eurasia (59.1±12 ka, *SE09*)
  - R → R2’JT (54.7±12 ka, Near East, *SE09*; 53.7±5.7 ka, *BO12*)
  - R → U (54±11 ka, *SE09*)
- R in SE Asia/Sahul (Australia/Melanesia, 58.4±8.4 ka, *SE09*; or 64.6 ka, Jinam, Hong, et al., 2012)
  - R → P (54.9±3.1 ka, *BO12*), Sahul: Australia, Papua NG, Melanesia
- R in East Asia (54.3±13 ka, *SE09*)
  - R → T16189C! (Caspian-Baikal; as ‘R→B’ 51±13 ka, *SE09*), highest frequency in Mongol, Southern China, Japan (rare)

Based on these dates R appears to have taken about 13,000 years to disperse over South Asia, Western Eurasia, SE Asia/ Sahul and East Asia, and in that geographic order. Again no simple ‘rapid coastal southern route’ is evident here. This view is supported by a remarkable recent discovery (Fu, Li, et al., 2014), which provides evidence for the earliest fossil mtDNA out-of-Africa, at Ust’Ishim, Irtysh River, Russia, bearing root-R with a novel SNP not in any other R
subclades in current populations, and dating ~43-47 ka. This study notes that in terms of current subclades of root-R, the R haplotype of the fossil is most closely related to mtDNA subclades P, B, F, T and J. When I entered the 12 SNPs in this study into the online James Lick mt haplogroup assigner, it showed the best matches were root R (11 matches, 1 extra SNP) and next best were R30, P, R2 JT and R(T16189C, which is the parent to B4’5) (also 11 matches, but 1 mismatch, and the 1 extra). Apparently, this individual’s lineage with root-R, prior to these subclades evolving, migrated into Western Siberia, and presumably taking one of the likeliest routes to get there, such as from the Persian Gulf Oasis north to the main ‘Silk Route’ to Central Asia or the ‘Silk Route Indus Spur’ to Central Asia, and from there migrating further north via the Aral Basin or Irtysh or other tributary to the Ob River. This is further confirmation of my hypothesis herein that the Borean-R language macrofamily had a homeland around the Persian Gulf Oasis and/or NW South Asia and then branched subclades into Western Eurasia, SW Asia, SE Asia/Sahul and Central Asia/East Asia; and a further refutation of the single rapid coastal route to Sahul hypothesis.

Interestingly, the early dispersals of N-mtDNA and R-mtDNA occur in roughly the same west to east sequence over roughly the same period of time. (N disperses in 15,000 years, R in 13,000; N disperses to East Asia, then SE Asia/Sahul, while R the converse; however, given the standard deviation of the TMRCAs these differences seem insignificant.) If this was the case, it raises the question of whether peoples with N and R migrated together? Did the earliest tribal group(s) out-of-Africa have a dual organization of its society that segregated and mixed these two haplogroups, or were they simply two separate diffusions?

I explore more evidence for my hypothesis that R correlates predominantly with speakers of Afroasiatic, Kartvelian, Dravidian, Elamitic, Austro and Pama-Nyungan families when I further discuss R subclades below.

Early Borean-M

- M in East Asia (60.6±13 ka, SE09)
  - M → M12’G (57±14 ka East Asia SE09; or 47.3±3.6 ka, BO12)
- M in South Asia
  - M → M42’74 (~55 ka, India/Australia divergence (Kumar, Ravuri, et al., 2009) or 49.6±8.1 ka, BO12)
- M in SE Asia/Sahul (49.4±10 ka, SE09; or ~64 kya, Jinam, Hong et al., 2012)

(Note. The wide variance in TMRCA in these studies of the earliest M dispersal reflect variously proposed problems in dating the M lineage compared to N and R lineages, such as tuning the molecular clock, variable mutation rates, possible impacts of the Toba supereruption, and need for more population samples from SE Asia, etc.)

Based on these TMRCAs, the Early M dispersal appears to begin from East Asia, from there to South Asia, and later into SE Asia/Sahul, and taken about 12,000 years. It is possible that the Toba supereruption attenuated an earlier presence of M in South Asia. According to these TMRCAs, M in South Asia, which is associated with current Austro-Asiatic speakers, appears to be a back-migration from SE Asia. Once again, no simple ‘rapid coastal southern
route’ is evident here. I explore more evidence for this and for my hypothesis that M correlates predominantly with speakers of the Eurasian macrofamily when I further discuss M subclades below.

Keeping in mind the caveat that the date for the emergence of a particular protolanguage may likely be later than the mtDNA TMRCA, the genetic and language family correlations I hypothesize as ‘Early Borean’ at around 70 ka are summarized in Table 6.

During this same stage, circa 70 ka, additional subclades of L3 emerge in Africa, which

<table>
<thead>
<tr>
<th>Table 6: ~70 ka (MIS 4) ‘Early Borean’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early Borean-N</strong></td>
</tr>
<tr>
<td>~71 ka (S Asia); ~62 ka (W Eurasia);</td>
</tr>
<tr>
<td>~58 ka (E Asia); ~56 ka (SE Asia/Sahul)</td>
</tr>
<tr>
<td>Dené-Caucasic (Burushaski)</td>
</tr>
<tr>
<td>N1'5</td>
</tr>
<tr>
<td>~57 ka</td>
</tr>
<tr>
<td>(N1: Central Asia, S Asia, SW Asia,</td>
</tr>
<tr>
<td>Europe; N5: South Asia)</td>
</tr>
<tr>
<td><strong>Early Borean-R</strong></td>
</tr>
<tr>
<td>~67 ka (S Asia); ~59 ka (W Eurasia);</td>
</tr>
<tr>
<td>~58 ka (SE Asia/Sahul); ~54 ka (E Asia)</td>
</tr>
<tr>
<td>Afroasiatic, Dravidian, Hattic, Austric, P-N</td>
</tr>
<tr>
<td>R31</td>
</tr>
<tr>
<td>~65 ka</td>
</tr>
<tr>
<td>(adopts Dravidian or I.E.)</td>
</tr>
<tr>
<td>R30</td>
</tr>
<tr>
<td>~64 ka</td>
</tr>
<tr>
<td>(adopts Dravidian or I.E.)</td>
</tr>
<tr>
<td><strong>Early Borean-M</strong></td>
</tr>
<tr>
<td>~61 ka (E Asia); ~55 ka (S Asia); ~49</td>
</tr>
<tr>
<td>ka (SE Asia/Sahul)</td>
</tr>
<tr>
<td>Eurasian</td>
</tr>
</tbody>
</table>

are associated with subsequent emergences of Sahara-Sahelian, Nilotic, Omotic, Cushitic and Chadic languages families. (For more details for L3-mtDNA subclades, TMRCAs and genetic-language family correlations, see Master Database, Supplementary File, Table 1, https://originsnet.academia.edu/JamesHarrod).

With respect to the on-going debates about ‘Australoid’ groups in southeastern Asia, these groups do not appear to show any similarities in extant mtDNA lineages. Their diversity is characterized by distinct markers among population samples, namely M31 and M32 in the Andamanese, N11b in the Mamanwa of the Philippines, M21a and R21 in the Jehai and Kensiu from West Malaysia, and haplogroups P, Q, S, and O in the Melanesians and Australian Aboriginals (Jinam, Hong, et al., 2012). Thus so-called ‘negrito’ populations are not a single out-of-Africa dispersal, but multiregional convergences under selection for tropical environments, which is supported by paleontology, fossils, and a variety of DNA studies using different techniques (see e.g., 2013, Special Issue on Revisiting the “Negrito” Hypothesis, Human Biology, 85(1)).

M, N and R-mtDNA: Middle Borean. The By around 50,000 years ago (MIS 3c/b), based on mtDNA phylotree and its TMRCAs, it appears that roughly 28 new mtDNA haplogroups had emerged. I list them below, noting dates, populations with high Hg frequencies, and current languages [in brackets] associated with those populations.
Middle Borean-N [N correlating to Burushaski-Caucasic-Dene macrofamily] (Fossil DNA, ‘macro-N’ (branch for N1a,b,c, X, I, W) in Gravettian Paglicci12, ~29 cal ka.)

- N in South Asia (71.2±16 ka, SE09)
  - N → N1'5 (57±5 ka, BO12)
  - N1'5 → N1 (54.2±13 ka, SE09; or 52±6 ka, BO12) Central Asia, SW Asia, S Asia, Europe [currently mostly I.E. speakers] [4.5%N1 and 2.3%N2 in Hunza, Burushaski speakers, but also 25%M]
  - N1'5 → N5 (37±8 ka, BO12) India, Madhya Pradesh Sahariya [I.E.]

- N in W Eurasia (61.9±11 ka, SE09)
  - N → N2 (44±7 ka, BO12) SW Asia, W Eurasia, includes Dargin, Chechen [Northeast Caucasian]
  - N → X (31.8±13 ka, SE09) Avar (15%X), Druze (16%X1'3+11%X2), Georgian (8%X)

- N in East Asia (58.2±14 ka, SE09)
  - N → N9 (49.1±12 ka, SE09; or 45.7±7.9 ka, BO12) root SW Asia, then Central Asia to E Asia, Udegey (29%N9+8%N9b), Japan (7%N9a,b)

- N in SE Asia/Sahul
  - N → N11 (56.3±3.6 ka, BO12); N11b (6.7±4.6 ka) in Mamanwa, Philippines [Austronesian]
  - N → S (53.5±5.5, BO12; or 25.4±5.2, Hudjashov, Kivisild, et al., 2007) only in Australia, Tasmania [Pama-Nyungan and non-Pama-Nyungan]
  - N → O in Australia (48.0 ka, van Holst Pellekaan, Ingman et al., 2008) (Fossil DNA, O1a-mtDNA in aboriginal hair, southern West Australia burial, 100 BP) [adopts Pama-Nyungan]

As observed earlier, N appears to have taken about 14,000 years to disperse over South Asia, Europe, East Asia and SE Asia into Sahul, and in that geographic order. No simple ‘rapid coastal southern route’ is evident here, and, apparently ironically for that hypothesis, dispersal into Western Eurasia occurs prior to SE Asia.

N1'5 branched N1 at ~54 or ~52 ka. In current population samples it has highest frequencies among the Mazandar, SW Capsian, northern Iran (14.3%); Sindhi, Pakistan (8.7%); Caucasus (8.7%); Lebanon (9%); Finland (8.3%) and Central Asian Kurds (7.3%), populations that currently are mostly I.E. speakers. In addition to 25%M, 4.5%N1 and 2.3%N2 occur in Hunza, who are Burushaski speakers. I suggest correlating N1 to an ancestor of Burushaski. These frequencies appear to reflect a dispersal encountering the Zagros Crossing bottleneck, and taking northerly migrations over the Silk Route Indus Spur to the Central Silk Road, thence west to Southern Caspian, and later splitting (a) back into Mesopotamia and (b) taking the Caucasus Silk Road Spur into Europe and (c) from Indus/Central Silk Road east into Tajikistan, Uzbekistan.
N1′5 branched N5 around 17,000 years later in India, ~37 ka, with highest frequency in the Sahariya, northern Madhya Pradesh (23%). This may reflect a successful Zagros bottleneck crossing into India and taking either a Ganges route across northern India or turning south along the coast and dispersing east over the Narmada River corridor. This hypothesis has archaeological support, as the TMRCA roughly correlates to the recent discovery of India’s earliest microblade industry along the Narmada River, Mehtakheri, Madhya Pradesh, Unit 2 ~48 ka, Unit 1 ~35 ka (Mishra, Chauhan & Singhvi, 2013). Maji, Krithika and Vasulu (2008) and Palanichamy et al. (2004) argue that N5 may be autochthonous for India.

N branched N2 around ~44 ka, with high frequencies among Sindhi, Pakistan (17.4%); Kurds in Middle East (10%) and in Asia (8.5%); Finland (9%); Gujarati, India (8.8%); Romania (6.4%); Caucasus (5%); Slovenia (4.8%); Shugnan Tajikistan (4.5%); Pathan, Pakistan (4.5%); Iran (4.1%); Uzbek (2.4%); Hunza (2.3%), many of which populations are currently I.E. speakers. Since I.E. is generally accepted by linguistics as very recent, even Holocene in origin, for the ancestral language family of N2, I suggest taking as a clue the trace N2a (~24 ka) in Daghestan Dargin (1%N2a or 4%N) and Chechens (<1%N2a), who are speakers of Northeast Caucasian. If so, N2 correlates to Northeast Caucasian. Similarly to N1’s haplogroup frequency samples, N2 appears to reflect Zagros Crossing bottleneck, taking the Indus Spur to the Central Silk Road and thence (a) west to Southern Caspian, then the Caucasus Spur into Europe and (b) east into Tajikistan and Uzbekistan.

N disperses into East Asia (~58 ka), and appears to branch a Middle Borean subclade N9 (~49 ka), and its later Y branch. Having high frequency among Nivkh, I suggest correlating N9-Y to a ‘Pre-Nivkh’. At some point in time, the Nivkh ancestors appear to have abandoned their Borean-N language and adopted a language from the Eurasian language family. Early N dispersed into SE Asia (~56 ka), with its N11 branch, the descendents of which seem to have adopted Austriic languages. S and O-mtDNA diffused into Sahul, ~53 ka and ~48 ka respectively, where S seems to have adopted non-Pama-Nyungan languages and O adopted Pama-Nyungan languages.

**Middle Borean-R** [R correlating to speakers of Afroasiatic, Kartvelian, Dravidian, Northwest Caucasian, Austric and Pama-Nyungan families]

- R in South Asia (66.6±14 ka, SE09)
  - R → R31 (64.5±14 ka, SE09; 54.9±3.1 ka, BO12), with frequency peaks in Rajasthan and Sri Lanka (Karmin thesis 2005) [adopt I.E. or Dravidian]
  - R → R30 (64.0±15 ka, SE09; 53.6±4.0 ka, BO12), with frequency peak in Gujarat, NW to Central India (Karmin, thesis, 2005); generally N/NW South Asia,
    - R30 → R30b (51.0±4.7 ka, BO12), with highest frequency in Vedda (39%R30b+R8a1a3), other Sri Lankans (Ranaweera, Kaewsutthi et al., 2014);
    - R30 → R30a (19.3±6.9 ka, BO12) in Central Tharu tribes, Nepal (range: 20% to
R \rightarrow R6 (51.1±16 ka, \text{SE09}; 43.9±9.6 ka, \text{BO12}), with frequency peak in Bharia, Madhya Pradesh [Dravidian] (18.4%); Kashmir (5.3%); Tamil Nadu (3.9%); Rajasthan (2.8%); Tharu (2.5%) [adopt I.E. or Dravidian]

R \rightarrow R8 (42±16 ka, \text{SE09}; 32.8±6.9 ka, \text{BO12}), in Vedda (see R30b above)

R in W Eurasia (59.1±12 ka, \text{SE09}) (Fossil DNA, ‘root-R’ in fossil Ust-Ishim, Irtysh, ~45 cal ka; R in fossil Fumane, ~41 cal ka.)

R2’JT (54.7±12 ka, \text{SE09}; 53.7±5.7 ka, \text{BO12}), homeland SW Asia

- R2’JT \rightarrow JT (50.3±12 ka, \text{SE09}; 47.0±6.5 ka, \text{BO12}) Egypt (27%J+T), Palestine; Iraq;
  - JT \rightarrow J (32.6±11 ka, \text{SE09}; 34.3±4.9 ka, \text{BO12}, but 43.3/34.5 ka, Pala, et al., 2012), SW Asia general (12%), Iran (13%); Europe (11%); Egypt (9%); J in Solutrean Nerja. Málaga fossil DNA, 17-20 ka;
  - JT \rightarrow T (26.8±9 ka, \text{SE09}; 25.1±4.7 ka, \text{BO12}) homeland Caspian?; Swanetia, Georgia [Svan<Kartvelian] (10.4%T+4.2%T1); Egypt (15%); Palestine (13%); Syria (12%); Iran (10%)

R2’JT \rightarrow R2 (41.0±16 ka (Metspalu et al., 2004); 13.7±6.9 ka, \text{BO12}) Al-Mahra, East Yemen (12%); Mazandar, SW Caspian [I.E.] (9.5%), Brahu (Dravidian) (7.9%)

R \rightarrow U (54±11 ka, \text{SE09}) homeland SW Asia (Fossil DNA, U near basal-R, not related to any current subclade of U, in Mal’ta MA-1, child burial, ~24 cal ka.)

- U \rightarrow U2 (54±13 ka, \text{SE09}; 43±4 ka, \text{BO12}); South Asia autochthonous (Metspalu et al., 2004), Kubachi, Daghestan [NE Caucasian] (24%); Iran (2.4%). (Fossil DNA, U2 in Eastern Aurignacian Kostenki14 burial, ~36-39 cal ka;
  - U2a,b,c (=U2i) (~23, 29, 39 ka, \text{BO12}), Irula, Tamil Nadu [Dravidian] (50%); Gondi [Dravidian] (36%); Uttar Pradesh Brahmins [I.E.] (27%); Sindhi (17.3%); Pathan (15.9%); Uttar Pradesh (15.3%); Sri Lanka (12%); Pakistan (mixed) (11%); Karnataka (10.6%); Kashmir (10.5%) [adopt I.E. or Dravidian]

- U \rightarrow U8 (50.2±11 ka, \text{SE09}; 43±4 ka, \text{BO12}); Asia and Europe; U8a (37±14 ka, \text{SE09}) SW Europe (2.8%); U8b’K (46±11 ka, \text{SE09}) (Asia/Europe). (Fossil DNA, U8 in Gravettian-Pavlovian Dolní Vestonice DV13, ~31 cal ka.)

- U \rightarrow U4’9 (~43±12 ka, \text{SE09}; 37±6 ka, \text{BO12}), Central Asia, Caucasus, Europe; U4’9 \rightarrow U9 (25.7±6.6 ka, Pakistan, \text{SE09}) \rightarrow U9a Andhra Pradesh, Ethiopia, U9b, Pakistan, with possible African origin; and U4’9 \rightarrow U4 (21±10 ka, \text{SE09}) Kalash [I.E.] (34%), Ket [Dené-Yeniseian] (29%), Nganasan [Samoyedic] (21%), Tundra Nenets [Samoyedic] (13%’U’), Tubalar [Altai] (15-18%), Western Siberia
(Mansi, Nentsi, Nganasan, Ket average) (17%), Mansi [Ob-Ugric] (16%),
Pakistan (14%), Volga-Ural peoples (9.7%), Hazara [I.E.] (9%), Swanetia (8.3%) and
Georgians [Kartvelian] (8%)

- U ➔ U3 (~11±6 ka, SE09; 34±6 ka, BO12), SW Asia; Lur Zagros Iran (18%)
  [I.E.], Jordan (15%); Adygei (14 or 5%), Abazins (12%), and Kabardin (10%)
  [latter three NW Caucasian speakers]

- R ➔ R0 (~39±15 ka, SE09; 40±11 ka, BO12), SW Asia; Marsh Arabs (0.7% R0 +
  6.9%R0a) [descendents of Sumerians]; Persians (2%). (Fossil DNA, ‘either R0 or
  HV’ in Gravettian Paglicci25, ~27 cal ka) [Pre-Basque]

- R in SE Asia/Sahul (Australia/Melanesia, 58.4±8.4 ka, SE09; or 64.6 ka, Jinam, Hong, et al.,
  2012)

  - R ➔ P (54.9±3.1 ka, BO12), Sahul: Australia, Papua NG, Melanesia
  
    - P ➔ P4 (53.0±4.4 ka, BO12) Australia, Melanesia

      P4b (40.1±7.1 ka, BO12) SW and No Australia [Greater-Pama-Nyungan]

    - P ➔ P8 (~39.8 ka, van Holst Pellekaan, Ingman et al., 2006), Yuendumu
      Warlpiri, central desert, Northern Territory [Pama-Nyungan]

    - P ➔ P1 (32.9±6.2 ka, BO12) Melanesia: Irian Jaya highlands [adopts Trans-New-
      Guinea, Papuan or Austronesian]

  - R ➔ R9 (47±12 ka, SE09; or 46.7±7.7 ka, BO12) Hmong-Mien (20%)

  - R9 ➔ R9c (46.7±6.3 ka, BO12) Philippine Batek Negrito [Austronesian] (58%)

    - R9c ➔ F (43±11 ka, SE09; or 42.79±5.6 ka, BO12) Temiar Senoi
      [Aslian<Austroasiatic] (43%F1a1a); Hue Vietnam [Mon-Khmer<Austroasiatic]
      (29%F*/F1a); Thailand [Tai-Kadai] (20%) and oddly Ket [Yeniseian] (24%)

    - R9 ➔ R9b (38.5±8.7 ka, BO12) Semelai Aboriginal Malay [Aslian Austro-Asiatic]
      (28%)

  - R ➔ R21 (47±12 ka, SE09) Jehai and Kensiu Semang and Temiar Senoi [Aslian<Austro-
    Asiatic];

  - R in East Asia (54.3±13 ka, SE09) (Fossil DNA, fossil B* in Tianyuan, China, ~40 cal ka.)

    - R ➔ B4’5 (49.5±6.6 ka, BO12) ➔ B4 (44±12 ka, SE09), Hmong-Mien; Vietnam [Mon-
      Khmer]; Polynesia, Micronesia, Madagascar; e.g., Vanatu [Austronesian] (40%)

As observed above in the discussion of Early Borean-R, R-mtDNA appears to have taken
about 13,000 years to disperse over South Asia (~67 ka, earliest branches R31 ~65 ka and R30
~64 ka, suggesting a Ganges or Narmada corridor east), Western Eurasia (~59 ka), SE Asia into
Sahul (~58 ka), and East Asia (~54 ka) and in that geographic order. Again no simple ‘rapid coastal
 southern route’ appears to be evident in these Early Borean R subclades and their dates.
Considering the above Middle Borean haplogroups and their TMRCAs around 50,000 years ago,
it seems that R-mtDNA has a much more complex pattern of differentiation and dispersal than
does N or M-mtDNA. Given the TMRCAs, R emerges in South Asia and its primary Middle Borean subclades, R6 (~51 ka), R30b (~51 ka), and R8 (~42 ka), occur with high frequency among Sri Lankan Vedda as well as Madhya Pradesh Bharia, that is, in two rather distant geographic regions.

In SW Asia, R branches R2'JT (~55 ka), which I would correlate to Pre-Egyptian-Semitic, R0 (~39 ka) correlating to Pre-Basque, and U (~54 ka), and their subsequent subclades. Based on haplogroup frequencies and distinctiveness, I suggest Middle Borean-U subclade U2 correlates to Pre-Dravidian, U8 to Pre-Hattic, U4'9 to Pre-Kartvelian and U3 to Pre-Northwest-Caucasic.

As linguists currently believe Kartvelian emerged in the Holocene, I here give my rationale for the label ‘Pre-Kartvelian’. This will also show how I have found it necessary to use the non-linguistic term ‘Pre’. The U4’9 homeland is variously postulated as Central Asia/Caucasus/Europe. In my review I found no mtDNA studies identifying U4’9; it appears it has not survived in extant population. There are extensive studies on U4 (~28 ka). Given its frequency cline, the U4 homeland appears to be somewhere along the Eurasian LGM ice-free zone paleolakes, around the southern side of Paleolake Mansi, including the Turgay Spillway to Paleolake Aral and eastward to the headwaters of the Irtysh, Ob, and Yenesi/Angara Rivers, on the Western Siberian Plain. It appears that from there four dispersals occurred:

(a) Northward following the ice retreat along the Yenesi (Ket, presently near confluence of the Tunguska River, and further north Sel'kup and Nenet along the Arctic Kara Sea; Nganasan on the Taymyr Peninsula);
(b) Northward following the ice retreat along the Ob (southern Sel'kup, Mansi, Khant, and northern Mansi along the Arctic Kara Sea) and southward to headwaters of Ob in the Altai (Tubalar);
(c) From Lake Mansi (across the Urals) northward toward the Barents Sea, White Sea, Kola Peninsula area (high frequency U4 in Mesolithic fossil DNA); into the Volga basin and north end of Paleolake Caspin and down to Caucasus (Dargin); and westward into the Dnieper and Danube basins north of Paleolake Black; southward along the eastern Black into Transcaucasia (present day Georgia);
(d) South along the Mansi-Aral spillway and Aral River into Central Asia (Koreimein Uzbekistan) and the Indus River Valley (Kalash, Pakistanis).

Peoples with high frequency of U4 appear to have adopted languages from various families, including Indo-European (Kalash), Dene-Yeniseian (Ket), Turkic (Tubalar), Uralic Ob-Ugric (Mansi) and Uralic Samoyedic (Nganasan, Nenet’s). Kartvelian (Swanetians, Georgians) is generally considered an isolate. If one assumes that Nganasan and Nenet’s peoples adopted a Uralic precursor and innovated Samoyedic, that leaves Kartvelian as the only distinctive language correlation for U4 peoples. So, I tentatively take Kartvelian to be a late innovation on an earlier U4’9 language family, and label that earlier U4’9 family ‘Pre-Kartvelian’.

Meanwhile, R in SE Asia/Sahul branches P (~55 ka), which seems strongly associated with Greater-Pama-Nyungan, and subsequently East Asian B4’5 (~50 ka) and Southeast Asian R9 (~47 ka), R21 (~47 ka) and F (~43 ka), the latter three all correlating to Pre-Austric. The dates for P match earliest archaeological dates for Australia ~52 ka.
Malakunanja II, Kakadu, Northern Territory, Australia, bipolar ‘horsehoof’ cores, flake lithics, pigments (OSL) 52±8 ka, confirmed; but lowest level artifacts and pigments (OSL) 61±10 ka disputed as disturbed; and Nauwalabila I, Kakadu, lithics, striated ochre pigment, bipolar ‘horsehoof’ cores (OSL) ~53±5 ka, though disputed (Roberts, Jones and Smith, 1990; Bird, Turney, et al 2002; Flood, 1990; O’Connell and Allen, 2004)

Middle Borean-M [M correlating to speakers of Eurasian macrofamily]

- M in East Asia (60.6±13 ka, SE09)
  - M → M12'G (57±14 ka East Asia SE09; or 47.3±3.6 ka, BO12) [Chukotko-Kamchatkan]
  - M → M7 (54.8±13 ka East Asia, SE09; or 44.9±3.5 ka, BO12) Ryukyu (36%); Ainu (20%); Japan (13%); Korea (11%) [Japonic-Korean-Ainu]; Han Chinese (4-8%); Buryat (5%); Tibet (3%); Kalmyk (2%)
  - M → M80'D → D (48.3±13 ka, East Asia, SE09; or 38.4±4.7 ka, BO12) → D4 (40.4/34.1 ka, Perego, Achilli et al., 2009) Han Chinese (62% to 32% D depending on location); Tibet (40%); Mongolian Buryat (35%) and Kalmyk (30%); Toto, Sikkim [Tibeto-Burman] (63%)
  - M → M13'46'61 (46.7±3.7 ka, BO12) Shannan, southern Tibet (7%M13a,b), Shertukpen, Arunachal Pradesh (23%M61), Lachungpa, Sikkim (12%M61) [all three groups Tibeto-Burman];
  - M → M8 (42.7±12 ka East Asia, SE09; or 36.4±7.3 ka, BO12)

- M in South Asia (see date for M42'74)
  - M → M42'74 (~55 ka, India/Australia divergence (Kumar, Ravuri, et al., 2009) or 49.6±8.1 ka, BO12)
    - M42'74 → M42 (47.7±8.1 ka, BO12) → M42b (40.3±7.0 ka, BO12 or 44.5±12.3 ka, Kumar, Ravuri, et al., 2009) → M42b1a (7.0±6.2 ka, BO12) in Madia-Gond, Maharashtra [Dravidian/I.E.]; Munda [Austro-Asiatic]; M74 (35.1±7.2 ka, BO12) in Pauri Bhuiya, Orissa [Dravidian/I.E.]
  - M → M33 (44.9±12 ka, SE09; 42.3±8.2 ka, BO12) multiple India tribes, e.g. Lepcha [Tibeto-Burman] (23%); Mal Paharia, Jharkhand [Austro-Asiatic] (11%); Dungri Bhil, Rajasthan/Gujarat [I.E.] (10%)
  - M → M5 (40±12 ka, SE09; 37.1±14.8 ka, BO12) multiple India tribes, e.g. Dungri Bhil (25%); Andh, Maharashtra, Andhra Pradesh, Chhattisgarh [I.E.] (18%); Kamar, Chhattisgar [Dravidian] (15%); Nihal [Nihali isolate/I.E.] (8%)
  - M → M4'67 (34.5±4.2 ka, BO12)
As noted above, the Early M dispersal appears to begin from East Asia, from there to South Asia, and later into SE Asia/Sahul, and taken about 12,000 years. According to the Middle Borean mtDNA branches and their TMRCAs, it appears that M12’G-mtDNA is strongly associated Chukotko-Kamchatkan and M7 with Japonic-Korean-Ainu. Since D-mtDNA is a branch of M, one might predict populations with high frequencies of D to be speakers of a language in the Eurasian family. Unexpectedly high frequency of D is found in Han Chinese speakers, and thus I infer that the ancestors of the Han originally spoke a Eurasian language but at some later time abandoned it and adopted or developed Chinese from a Borean-N language in the Dené-Caucasian family. The M13’46’61 branch seems to correlate to Pre-Tibeto-Burman; M29’Q to Pre-Papuan; and M8 (CZ->C) to Pre-Altaic-Mongolian-Tungusic and the Yukaghir isolate. M in South Asia, which is associated with current Austro-Asiatic speakers, appears to be a back-migration from SE Asia.

Thus, by around 50,000 years ago (MIS 3c/b), based on mtDNA phyloretree TMRCAs, it appears that roughly 28 new haplogroups had emerged, and based on current languages associated with them, I suggest that at least 17 of these haplogroups correlate to the emergence of 17 language families from Europe to East Asia and the Sahul. I suggest the term ‘Middle Borean’ to designate the ancestors of these language families. Keeping in mind the caveat that the date for the emergence of a particular protolanguage may likely will be later than the mtDNA TMRCAs, the genetic and language family correlations I hypothesize at around 50 ka are summarized in Table 7 and Figure 6. Figure 6 circles are placed on the map to illustrate
approximate homelands; original mtDNA homelands, migration routes and destinations (current homelands) may have covered wider or more irregular regions.

<table>
<thead>
<tr>
<th>Table 7: ~50 ka (MIS 3c/b) 'Middle Borean'</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early Borean-N</strong></td>
</tr>
<tr>
<td><strong>Middle Borean-N</strong></td>
</tr>
<tr>
<td>N1'5</td>
</tr>
<tr>
<td>N1</td>
</tr>
<tr>
<td>N2</td>
</tr>
<tr>
<td>N9, Y</td>
</tr>
<tr>
<td>N11</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>O</td>
</tr>
</tbody>
</table>

**Early Borean-M** ~61 ka (E Asia); ~55 ka (S Asia); ~49 ka (SE Asia/Sahul) Eurasian

<p>| <strong>Middle Borean-M</strong>                        | |
| M12'G                                      | ~57 ka Pre-Chukotko-Kamchatkan |
| M7                                         | ~55 ka Pre-Japonic-Korean-Ainu |
| M9                                         | ~53 ka (adopt Tibeto-Burman, Sinitic, Austric, Papuan) |
| M42'74                                     | ~52 ka (adopts various Dravidian, Austroasiatic, I.E.) |
| D                                          | ~48 ka (adopts Borean-N Dené-Caucasic-Burushaski as Sino-Tibetan?) |
| M13'46'61                                  | ~47 ka Pre-Tibeto-Burman |
| M29'Q (Q2)                                 | ~47 ka Pre-Papuan |
| M9ab'E                                     | ~47 ka (adopt Sinitic, T-B, Austric, Papuan) |
| M21                                        | ~46 ka (adopts? Aslian Austro-Asiatic) |
| M33                                        | ~45 ka (adopts various Tibeto-Burman, Austroasiatic, I.E.) |
| M27                                        | ~45 ka (adopts Papuan?) |
| M8                                         | ~43 ka Pre-Altaic-Mongolian-Tungusic |
| M22                                        | ~41 ka (adopts Malayan, Austronesian) |
| M42a                                        | ~41 ka (adopts Pama-Nyungan) |
| M5, M2                                     | ~40 ka (adopts various Dravidian, Austroasiatic, I.E.) M5 in Nihal |</p>
<table>
<thead>
<tr>
<th>Early Borean-R</th>
<th>~67 ka (S Asia); ~59 ka (W Eurasia); ~58 ka (SE Asia/Sahul); ~54 ka (E Asia) Afroasiatic, Dravidian, Hattic, Austric, P-N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Middle Borean-R</strong>&lt;br&gt;South Asian</td>
<td><strong>Middle Borean-R</strong>&lt;br&gt;Western Eurasian</td>
</tr>
<tr>
<td>R6</td>
<td>~51 ka Bharia (adopts Dravidian)</td>
</tr>
<tr>
<td>R30b</td>
<td>~51 ka Vedda substrate</td>
</tr>
<tr>
<td>R8</td>
<td>~42 ka in Vedda root-R in fossil Ust-Ishim, ~45 ka; R in fossil Fumane, ~41 ka</td>
</tr>
<tr>
<td>R2':JT</td>
<td>~55 ka</td>
</tr>
<tr>
<td>R→JT</td>
<td>~50 ka Pre-Semitic-Egyptian</td>
</tr>
<tr>
<td>R→U</td>
<td>~54 ka basal U in fossil Mal'ta MA1, ~24 ka</td>
</tr>
<tr>
<td>U2</td>
<td>~54 ka Pre-Dravidian fossil U2 in Aurignacian Kostenki14, ~38 ka</td>
</tr>
<tr>
<td>U8</td>
<td>~50 ka Pre-Hattic-Kaskian? fossil U8 in Gravettian Dolni Vestonice13, ~31 ka</td>
</tr>
<tr>
<td>U4'9</td>
<td>~43 ka Pre-Kartvelian</td>
</tr>
<tr>
<td>U3</td>
<td>~41 ka Pre-Northwest-Caucasic</td>
</tr>
<tr>
<td>R0</td>
<td>~39 ka Pre-Basque (also in Marsh Arabs, Iraq, presumed descendents of Sumerians) fossil R0/HV in Gravettian Paglicci25, ~27 ka</td>
</tr>
<tr>
<td>U6</td>
<td>~36 ka (back migration into Northern Africa, adopts Berber) Early to Late UP Dabban 43 to 17 ka, Haua Fteah</td>
</tr>
<tr>
<td><strong>Middle Borean-R</strong>&lt;br&gt;SE and E Asian</td>
<td><strong>Middle Borean-R</strong>&lt;br&gt;South Asian</td>
</tr>
<tr>
<td>P</td>
<td>~55 ka Greater-Pama-Nyungan fossil B* in Tianyuan, ~40 ka</td>
</tr>
<tr>
<td>B4'5</td>
<td>~50 ka Pre-Austric</td>
</tr>
<tr>
<td>B4</td>
<td>~44 ka (Austronesian, Austroasiatic, adopts Trans-New-Guinea)</td>
</tr>
<tr>
<td>B5</td>
<td>~42 ka (Kam-Tai, Tibeto-Burman)</td>
</tr>
<tr>
<td>R9</td>
<td>~47 ka Pre-Austric (Malay Aslian Austroasiatic, Kam-Tai, Austronesian, Hmong-Mien)</td>
</tr>
<tr>
<td>R21</td>
<td>~47 ka Pre-Austric (Aslian Austroasiatic)</td>
</tr>
<tr>
<td>R9→F</td>
<td>~43 ka Pre-Austric (in Mon-Khmer, Austroasiatic, Tai, Aslian)</td>
</tr>
</tbody>
</table>
Figure 6. Out-of-Africa mtDNA Haplogroups Ancestral ('Pre') to Language Macrofamilies at circa 50,000 Years Ago

Blue: Borean-R (Pre-Afroasiatic, etc.); Green: Borean-N (Pre-Caucasic-Burushaski-Dene); Red: Borean-M (Pre-Eurasian)
M. N and R-mtDNA: Late Borean. By around 25,000 years ago another dozen language families emerged—which I term ‘Late Borean’—and this, for the most part, appears to have completed the development of the major language families of the world.

Late Borean-R  [R correlating to speakers of Afroasiatic, Kartvelian, Dravidian, Elamitic, Austric and Pama-Nyungan families]

- R in W Eurasia
  - JT → J (32.6±11 ka, SE09; 34.3±4.9 ka, BO12; 34.5/43.3 ka, Pala, Olivieri, et al., 2012) Saudi Arabia (21%); SW Asia general (12%); Iran (13%); Europe (11%); Egypt (9%); Basque; Caucasus (8%); South Asia
    - J → J1 (24.1±5.8 ka, SE09; 26.9±5.3 ka, BO12; 25.0/33.3 ka, Pala, Olivieri, et al., 2012) Eastern Europe, Ukraine, Balkans, arrived post-LGM; in Neolithic and Mesolithic fossil DNA
    - J → J2 (28.3±4.6 ka, BO12; 32.9/36.8 ka, Pala, Olivieri, et al., 2012) SW Asia; J2a and J2b1 in Europe ~15-16 ka (Pala, Olivieri, et al., 2012). (Fossil J or less likely JT in Solutrean, Nerja, Málaga, ~20-24 cal ka (Fernández, thesis, 2005) [JBH: Oven & Kayser Build 14 has some Nerja SNPs in JT, J1 and J2; and Pala et al., 2012) has 3 SNPs for J2b1 similar to Nerja fossils.]
  - R0 HV (27.1±7.5 ka, SE09; 21.9±2.8 ka, BO12; Near East) Syria (24%R0s/HV+16Hs), Basque (16%HV+52%H+7%V or 3%to14%HV+43%to67%Hs); Iraqis (13%HV+4%R0a+16%Hs), Persians (2%R0+11%HV+17%Hs); Dargin (9%HV+23%Hs) and Avar (8%HV+23%Hs) [NE Caucasian]; Marsh Arabs (8%R0/R0a+4%HV+12%Hs) [Sumerian?]; fossil Minoan (3%R0+8%HV+32%Hs)
    - HV → HV0 (‘pre-V’) (19±7 ka, SE09; 13.5±3.2 ka, BO12; before LGM, perhaps Eastern Europe, spreading E to W along Gravettian axis); HV0a, HV4a1 Basque
    - HV → H (18.6±4 ka, SE09; 12.8±0.8 ka, BO12) Basque (16%HV+52%H+7%V or 3%to14%HV+43%to67%Hs); Scandinavia (49%); Sicily (49%); Germany (49%), France (47%), North Italy (47%); European Russians (42%); Dargin (24%), Chechen (24%), Avar (23%). (Fossil H in Magdalenian La Pasiega; in Mesolithic Villabruna, ~14 cal ka.)
  - U → U1 (37±11 ka, SE09; 32.0±5.4 ka, BO12) Kubachi [NE Caucasian] (48%), Azeri [Turkic] (8%), Lur Zagros Iran [IE] (6%), Persians [IE] 3%. [Pre-Elamitic?]
  - U → U5 (36±11 ka, SE09; 30.2±5.3 ka, BO12) Finno-Lapic Saami (48%U5b1b1), Seto (23%), Finns (19%), and Karelians (17%); Finno-Perm Mordvin (16%) European Russia [I.E.] (14%); Basque (11%); Hungarian (8%). (Fossil DNA, U5 mutations in Gravettian Dolní Vestonice 14 and 15, ~31 cal ka; U5 n Magdalenian Cantabria; U5b1 in Late UP Federmesser, ~14 cal ka.) [Pre-Finno-Ugric/Uralic]
MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

- U8→U8b (35±12 ka, SE09; 38.4±4.9 ka, BO12) Kurdish, W Iran [IE] (10%); Qatar (10%); U.A.E (8%); Gilaki, N Iran, SW Caspian [IE] (2.7%); Konya, Anatolia [Turkish] (2.0%); Jordan, Italy; U8→U8a (37±14 ka, SE09; 18.5±5.2 ka, BO12) SW Europe (2.8%); Basque (1.1%); north-central Europe (0.2%), Anatolia (0.2%); Basque since 28±9 ka, U8a in Basque expansion 23±14 ka, U8a1 expansion 14±5 ka;

- U8b→K (31.9±11 ka, SE09; 26.7±4.3 ka, BO12) Druze (16%); Swanetia [Kartvelian] (12.5%); Kurdish Turkmenistan [IE] (12.5%); Georgian (10%); Kurdish, Iran [IE] (10%); U.A.E (10%); Caucasus (8.6%); Palestine, and Iran (8%); Syria, Iraq and Lur Zagros Iran (6%); Turkey (5%); Jordan (4%); Hazara [IE]; Baluch [IE]; Europe [IE], Basque (K1a1), Ashkenazi (K1a1b1a). (Fossil DNA, K-mtDNA in LBK Germany (14%); is a Neolithic marker.)

- U → U7 (22±10 ka, SE09, homeland SW Asia between Gujarat and Iran; 18.1±3.7 ka, BO12) Iran Kurds [I.E.] (20%); Gilaki, SW Caspian [I.E.] (11%); Brahui [Dravidian] (10.5%); Gujarati [I.E.] (9%); Sindhi [I.E.] (9%); Hunza [Burushaski] (7%);


- R in SE Asia, East Asia, Beringia and Americas

- B4′5 → B4 → B4b (28±9 ka, SE09), Tubalar (6%), Mongol (15%B4), W. Evenk (4%B4), Altai-Kizhi, Tuva, Tofalar

- B4′5 → B4 → B4a (26±8 ka, SE09) China, Thailand, Indonesia, Taiwan, Philippines [Austronesian], B4a1a (10±5 ka, SE09) ‘Polynesian motif’; Karkan Islanders [Austronesian and Trans-New-Guinea] (11%B4a+23%B4a1a1);

- B4′5 → B5 → B5a (27.6±5.9 ka, BO12) Musuo, Yunnan [Na<Tibeto-Burman] (13%B4+17.4%B5a)

- B4′5 → B4 → B2 (21.2±2.4 ka, Achilli, Perego, et al., 2008) in North American tribes such as Kiliwa (Baja) [Yuman] (100%); Jemez Pueblo [Tanoan] (89%); Zuni [isolate] (77%); Penutian Wintuan, Utian, Miwok, Costanoan, Yokut (56%) and Salish-Sahaptian-Yakama-Wishram (50-67%); South American tribes such as Matsiguenga, Peru [Arawak] (92%); Aché, Paraguay [Guarani<Tupian] (90%); Xavante, Mata Grosso [Ge] (84%); Uros [Uruquilla isolate] (73%); Aymara [Greenberg: Andean] (72%); Quechua [Andean] (61%) [Question: does this grouping correspond to Y. Berezkin (2010a, 2010b) ‘Indo-Pacific’ in the Americas?]

Late Borean-N [N correlating to Burushaski-Caucasic-Déné macrofamily]

50
**Mother Tongue**

*Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014*

- **N in East Asia, Beringia and Americas**
  - **N → A (29.2±10 ka, SE09; Caspian-Baikal homeland; or 24.2±4.9 ka, BO12)**
  - **A → A2 (15.6±1.8 ka, BO12; or 18.1±1.8 ka, Achilli, Perego, et al., 2008); Na Dené (95%); Dogrib, Tlingit (100%); Haida (85-96%); Maya, Quintana Roo [Mayan] (80%); Bella Coola [Salishan] (65%); Mixtec-Alta [Otomanguean] (73%); Mixe [Mixe-Zoquean] (63%); Apache (63%); Navaho [Dené] (52-58%A2+38-41%B); Chumash (53-60%); Nahua (57%A2+28%B2) and Huichol (56%A2+25%B2) [both Uto-Aztecan]; Kuna (100%), Ijka (90%), Arsario (68%), Kogi (65%) [all four Chibchan-Paezan]; Guaraní, Brazil [Tupian] (84%); Barasano [Equatorial-Tucanoan] (80%); Kaingang [Ge] (62%); I would call this 'Pre-Dené' or 'A-Amerind'**
  - **N9'Y Y (22±11 ka, SE09; or 24.6±7.1 ka, BO12) Nivkh [isolate] (66%); Ulchi SE Tungusic I, 24%+11%N9b); Ainu [isolate or Japonic-Korean-Ainu] (20%+8%N9b)**

- **N in W Eurasia**
  - **N1→N1ae’I → N1e’I → I (26.3±10 ka, SE09) Northern Europe (2-4%), Ukraine (11%), Pakistan (9%) [I.E.], El Molo [Cushitic] (22%); N1a (19.3-22.3 ka, Fernandes, Alshamali et al., 2012), with branches in Central Asia, SW Asia, Europe, in LBK Neolithic fossil**
  - **N2 → N2a (24±8 ka, BO12) Caucasus, Eastern Europe; N2 → W (21±8 ka, SE09) Indus Valley, SW Asia, Europe, NW Africa, India; Neolithic marker; Sindhi [I.E.] (17%); Finns [I.E.] (10%); Kurds [I.E.] (10%); Mazandar, Iran [I.E.] (9.5%); Gujarati [I.E.] (9%); Swandan [Kartvelian] (8%)**
  - **N → X (31±13 ka, SE09) Avar (15%X [basal X]+6%N) → X1’2’3 (~28.8 ka, Fernandes, Alshamali et al., 2012) → X1’3 (~21 ka, Fernandes, Alshamali et al., 2012) Druze (16%X1’3+11%X2); Georgian (8%X) and → X2 (20.9±9 ka, SE09; 19.2±2.6 ka, BO12) populations across SW Asia, Europe, North Africa, Central Asia, Siberia**
    - **X2 → X2a’j (17.1±3.1 ka, BO12; ~19.4 ka, Fernandes, Alshamali et al., 2012)**
    - **X2a (12.8±7 ka, SE09; 12.7±3.5 ka, BO12; ~14.1 ka, Fernandes, Alshamali et al., 2012) Micmaw (50%X+33%A), Anishinabe/Ojibwa, Minnesota (50%), Cheyenne (18%+50%A) [all three Algonquian]; Nuu-chah-nulth (7%) and Yakama (5%) [both Penutian]; Navaho [Na-Dené] (3%) ['X-Amerind']. (Fossil X2a in ‘Kennewick Man’, ~8.5 cal ka.)**

- **N in South Asia**
  - **N1’5 → N5 (37±8 ka, BO12) India, Madhya Pradesh Sahariya [I.E.]**

- **N in SE Asia/Sahul**
  - **N → N22 (25.2±8.8 ka, BO12) Temuan Aboriginal Malay [Malayan, Austronesian] (12%N22+15%N21 or17%N22+22%N21) and → N21 (22.4±9.0 ka, BO12) Semelai Aboriginal Malay (31%N21); Temiar Senoi (31%N21). M21 in Temuan appears derived from ancestral type found in Cham of Vietnam (Jinam, Hong, et al., 2012)**

51
MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

- N → N12,13,14 (~17 ka, Hudjashov, Kivisild et al., 2007) in Kalumburu, others Australia [Pama-Nyungan and non-Pama-Nyungan speakers]
- N → S → S2 (38±7 ka, BO12; or 15.2±5.1, Hudjashov, Kivisild et al., 2007) and → S1 (22.0±7.7 ka, Hudjashov, Kivisild et al., 2007) Paakintji, Lower Darling River (7%S1+1%S2) and Yuendumu-Warlpiri, Central Desert (10%S1+8%S2+1%S5) [both Pama-Nyungan] and Kalumburu, Kimberly [non-Pama-Nyungan] (28%S2+6%S5)

Late Borean-M [M correlating to speakers of Eurasian macrofamily]
- M in East Asia, Beringia and Americas
  - D4 → D1 (16.8±2.9 ka, BO12; 16.9±1.6 ka, Bodner, Perego et al., 2012) → D1g (11.6±4.4 ka, BO12) Chile and Argentina, esp. Mapuche, thus earliest stage of rapid coastal route, Beringia to Southern Cone dispersing in less than 2,000 years (Bodner, Perego et al., 2012) Yahgan [Yahgan/Yamana isolate] (10%D1+33%D1g+10%D4h3a); Huilliche, central Chile (4%D1+37%D1g+4%D4h3a) and Mapuche, central Argentina (4%D1+24%D1g+1%D1j) [both Araucanian isolate<Andean]; Mayo, Sinaloa [Uto-Aztecan] (33%); Kawésqar [Alacalufan isolate], nomadic seafarers (8%);
    D (subclade not specified): Shoshone/N. Paiute [Uto-Aztecan] (48%); Yuk-Utian [Penutian] (47%); Wapishana, Brazil (67%). Zoró (60%) and Tucano (57%) [all three Equatorial-Tucanoan]
  - D4 → D4h (21.5±2.4 ka, BO12) → D4h3 (18.3±2.9 ka, BO12) → D4h3a (13.0±2.6 ka, BO12) mostly South America, Chile, less so in Mexico, California, confirms a coastal route (Perego, Achilli et al., 2009) Kawésqar [Alacalufan isolate], nomadic seafarers (46%); Tehuelche, central southern Argentina [Chonan] (28%D4h3a+16%D1g); Cayapa, Ecuador (22%); Chumash (16%); Aonikenk, Fuego [Chonan] (73%D). (Fossil D4h3a in Anzick-1, ~12.5 cal ka.)
  - M12’G → G (35.7±10 ka East Asia SE09; or 31.3±5.7 ka, BO12) Itelmen (68%) and Koryak (42%) [both Chukotko-Kamchadalan]
  - M8 → CZ → Z (24±9 ka Caspian-Baikal, SE09; or 21.7±8.4 ka, BO12) and CZ → C (28±9 ka Caspian-Baikal, SE09; or 23.9±4.8 ka, BO12) Yukaghir [isolate language] (66-72%); E. Evenk (62%), W. Evenk (50%) and Even (43%) [all three N. Tungusic]; Tofalar (61%), Tuvan (51%) and Yakut (47%) [all three Turkic]; Nganasan (51%), Tundra Nenets (32%) and Sel’kup (22%) [all three Samoyedic]; Altai-Kizhi [Altai] (34%); Mongol (19%)
  - M8 → M8a (26±10 ka, China, Japan, SE09), Dirang Monpa, Arunachal Pradesh [Tibeto-Burman] (24%)
  - C → C1 (17.1±5 ka, SE09; 18.3±4.2 kya, BO12) Ayoreo [Zamucoan] (84-100%); Ancient Maya-Copan (89%); Baja Seri [isolate] (88%); Pima (82%) and Tarahumare
MOTHER TONGUE

Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

(54%) [both Uto-Aztecan]; Taino [Arawakan] (75%); Makiritare [Ge-Pano-Carib] (70%);
Movina [Equatorial-Tucanoan] (64%); Yanomama [Chibchan-Paezan] (54-72%);
Mapuche, central Argentina [Araucanian isolate<Andean] (42%)

o M in South Asia
  o M → M32′56, M31, M39′70, M36, M49, M3, ~35 ka; and M60, M6, M19′53, M41,
    M44, <30 ka [associated with/adopt various language families]

o M in SE Asia/Sahul
  o M → M75, M21a,b, M76, M45, M26, M51, M20, M1, M9a,b, E; all <30 ka [associated
    with/adopt various language families]
  o M29′Q → Q (37.5±5.6 ka, BO12) Melanesia, Australia → Q1 (18.2±7.3 ka, BO12)
    Muyu, Irian Jaya Highlands [Ok] (88%), Lowland Riverine (Mandobo), West Papua
    (85%); Asmat, Irian Jaya SW Coast [Asmat] (84%), Dani, Irian Jaya Highlands [Dani]
    (76%), Una (pygmyoid), Irian Jaya Highlands [Mek] (62%); Bandi [Chimbu-Wangi]
    (29%) [all Trans-New-Guinea]; Aita, North Bougainville [Papuan] (82%)

Thus, by around 25,000 years ago, based on mtDNA phylotree TMRCAs, it appears that
roughly two dozen new haplogroups had emerged, and based on current languages associated
with them, I suggest that at least a dozen of these haplogroups correlate to the emergence of
distinct language families in Europe, East Asia and the Sahul. I suggest the term ‘Late
Borean’ to designate the ancestors of these language families. Keeping in mind the caveat
that the date for the emergence of a particular protolanguage may likely will be later than the
mtDNA TMRCA, the genetic and language family correlations I hypothesize at around 25 ka
are summarized in Table 8.
### Table 8: ~25 ka (MIS 2) ‘Late Borean’

<table>
<thead>
<tr>
<th>Late Borean-N</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>~29 ka</td>
<td>(Caspian-Baikal)</td>
</tr>
<tr>
<td>A2</td>
<td>~16 ka</td>
<td>A-Amerind (Pre-Dené)</td>
</tr>
<tr>
<td>W</td>
<td>~21 ka</td>
<td>(adopts I.E.)</td>
</tr>
<tr>
<td>Y</td>
<td>~22 ka</td>
<td>Nivkh (adopts Pre-Eurasiatic language?)</td>
</tr>
<tr>
<td>X, X1'2'3</td>
<td>~31 ka</td>
<td>(SW Asia, Central Asia, Siberia)</td>
</tr>
<tr>
<td>X2a’j</td>
<td>~17 ka</td>
<td>X-Amerind (Algonquian)</td>
</tr>
<tr>
<td>X2a</td>
<td>~13 ka</td>
<td></td>
</tr>
<tr>
<td>N22</td>
<td>~25 ka</td>
<td>(in Aboriginal Malay, Senoi)</td>
</tr>
<tr>
<td>N12, N13, N14</td>
<td></td>
<td>(in Pama-Nyungan and non-Pama-Nyungan)</td>
</tr>
<tr>
<td>S1</td>
<td>~22 ka</td>
<td>(in Pama-Nyungan and non-Pama-Nyungan)</td>
</tr>
<tr>
<td>S2</td>
<td>38/15 ka</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Late Borean-M</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>~35 ka</td>
<td>Chukoto-Kamchatkan</td>
</tr>
<tr>
<td>C</td>
<td>~28 ka</td>
<td>Pre-Altaic-Tungusic-Mongolian and Yukaghir (isolate)</td>
</tr>
<tr>
<td>E</td>
<td>~23 ka</td>
<td>(adopts Papuan, Austronesian)</td>
</tr>
<tr>
<td>Q1</td>
<td>~18 ka</td>
<td>Trans-New-Guinea</td>
</tr>
<tr>
<td>D1</td>
<td>~17 ka</td>
<td>D-Amerind (wave 1?)</td>
</tr>
<tr>
<td>D4h3a</td>
<td>~13 ka</td>
<td>D-Amerind (wave 2?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fossil D4h3a in Anzick-1, 12.5 ka</td>
</tr>
<tr>
<td>C1</td>
<td>~17 ka</td>
<td>C-Amerind</td>
</tr>
<tr>
<td>M9a</td>
<td>~15 ka</td>
<td>(adopts Tibeto-Burman)</td>
</tr>
<tr>
<td>M9b</td>
<td>~19/3 ka</td>
<td>(adopts various Austric)</td>
</tr>
<tr>
<td>Late Borean-R Western Eurasian</td>
<td>~37 ka</td>
<td>Pre-Elamitic (?)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------</td>
<td>----------------</td>
</tr>
<tr>
<td>U8→U8b</td>
<td>~35 ka</td>
<td>(in SW Asia, Iran, Qatar, UAE, Anatolia, Italy)</td>
</tr>
<tr>
<td>U8b→K</td>
<td>~30 ka</td>
<td>(in SW Asia, Druze, Kurds, Caucasus, Georgia, Iran, Turkey, Europe)</td>
</tr>
<tr>
<td>U8→U8a</td>
<td>~37 or 18 ka</td>
<td>(in SW and N Cen Europe, Basque, Anatolia)</td>
</tr>
<tr>
<td>J</td>
<td>~33 ka</td>
<td>(in SW Asia, Europe, S Asia)</td>
</tr>
<tr>
<td>→J2</td>
<td>~28 ka</td>
<td>(in SW Asia, Europe)</td>
</tr>
<tr>
<td>HV</td>
<td>~27 ka</td>
<td>(in SW Asia, Caucasus, Basque)</td>
</tr>
<tr>
<td>→HV subclades</td>
<td>~19 ka</td>
<td>(in Basque, Saami, E. Eurasia)</td>
</tr>
<tr>
<td>→H</td>
<td>~19b ka</td>
<td>(in Basque, E. Eurasia)</td>
</tr>
<tr>
<td>U7</td>
<td>~22 ka</td>
<td>Pre-Harappan</td>
</tr>
<tr>
<td>U4</td>
<td>~21 ka</td>
<td>(in Nganasan, Nenets, Ket, Kalash, Mansi, Dargin, Swanetia, Georgians)</td>
</tr>
</tbody>
</table>

Late Borean-R SE Asian, E Asian Beringia, Americas

| B4b                           | ~28 ka | (in Mongolia, Siberia, various languages) | |
| B5a                           | ~28 ka | (in Tai-Kadai, Tibeto-Burman) | |
| B4a                           | ~26 ka | Austronesian | |
| B4a1a                         | ~10 ka | (Polynesian motif) | |
| B2                            | ~21 ka | B-Amerind | ‘Indo-Pacific’ in Americas (Berezkin) |
Discussion.

This discussion section focuses on out-of-Africa geographic bottlenecks and derives correlations of mtDNA and language family dispersals for each bottleneck. To provide a detailed discussion of each geographic region of the world is beyond the scope of this paper. For this level of detail the reader is encouraged to peruse the Master Database (Supplementary File, Table 1, https://originsnet.academia.edu/JamesHarrod), which to a large extent is chronologically ordered and ordered by region. In addition Appendix 1 consists of out-of-Africa mtDNA and correlated language family dispersal maps. Five maps are posited for each of the major stages of mtDNA phylotree haplogroup evolution and TMRCA dates, illustrating apparent dispersal routes and correlations to language families at each stage. In drawing these maps I superimposed my dispersal hypothesis onto maps illustrated in Mishra, Chauhan & Singhvi (2013: Fig. 2), for which I am grateful.

Based on the results of my meta-pattern-analysis it appears that basic mtDNA haplogroup and language family emergences and differentiations occur each of seven (7) major geographic bottlenecks that had to be crossed from Africa through Asia to Sahul and the Americas.

- Africa to SW Asia Crossing (circa 75 ka, differentiating L3 and Early Borean-N, R and M)
- SW Asia to S Asia, the Zagros Crossing (circa 65 ka, differentiating Middle Borean-R
- Ganges Delta Crossing to SE Asia (circa 50 ka, differentiating Middle Borean-R and Middle Borean-M)
- SW Asia to W Eurasia, the Transcaucasus Crossing (circa 50 ka, differentiating Middle Borean-U)
- ‘Silk Route Crossing’, Persian and Indus Spurs bifurcation west to Europe, east to Central and East Asia (circa 25 ka, differentiating Late Borean-N)
- Sunda/Sahul Crossing (circa 50 ka, differentiating Middle Borean-N, M and R)
- Siberia to the Americas, the Beringia Crossing (circa 25 ka, differentiating Late Borean-N, R and M)

The mtDNA phylotree branchings at these seven bottlenecks seem to display a common pattern. On the proximal side of a bottleneck, one or more haplogroups seem to swirl backward in the opposite direction from the narrow passage; on the distal side, haplogroups spurt out in divergent directions. At E-W crossings (Africa-SW Asia, Zagros and Ganges), they diverge in northerly, central and southerly directions. At N-S crossings (Transcaucasus and Silk Route), they diverge in easterly, central and westerly directions. They can be visualized schematically as having the shape of a scorpion, and so I call them ‘scorpion bottlenecks’.
Based on my meta-pattern-analysis results I diagram eight (8) correlations of mtDNA haplogroups and language families at the seven geographic bottleneck crossings. TMRCA dates and original or current homelands are those identified in the results section of this study.

1. Africa to SW Asia Crossing (circa 75 ka, differentiating L3 and Early Borean-N, R and M). (TMRCA dates for this crossing are multiplied 1.1x to account for earlier revised CHCLA date, see narrative.)

   L3h subclade (~69 ka, Omo, Egypt, Nubia-Sahara, Cushites, Nilotics): (adopts Afroasiatic and Nilo-Saharan)

   N (~75-78 ka, S Asia): Borean-N: Pre-Caucasian-Burushaski-Dene
   M (~67 ka, E or S Asia): Borean-M: Pre-Eurasiatic
   N→R (~74 ka, S Asia): Borean-R: SW Asia and ‘northern route’ South, East and SE Asia
   N (~63 ka, E Asia)→N11 (~61 ka, SE Asia)

L3 dates around 80 ka. Distal branches N and M occur only out-of-Africa in SW Asia, with N dating about ~75-78 ka; R-mtDNA from N at around 74 ka; and M somewhat later at ~67 ka. N arrives in E Asia around 63 ka and in SE Asia a couple thousand years later. There is insufficient well-dated archaeology to evaluate whether the L3 dispersal route into SW Asia was via the Sinai, the Bab el Mandeb or around the Red Sea, or some combination of these. The scorpion diagram suggests taking as its proximal ‘backward movement’ the oldest L3 subclade, L3h, which has a TMRCA later than N and which has highest frequencies among Egyptians, Nubian-Saharans, Omotics, and Nilotics. As noted earlier, these dates suggest that the *sapiens* dispersal out-of-Africa endured a roughly 5,000-year pause in SW Asia. A *sapiens* ‘fast track’ on some sort of ‘southern route’ sometime between 60 to 45 ka does not appear supported.
2. **SW Asia to S Asia, the Zagros Crossing (circa 65 ka, differentiating Middle Borean-R)**

('mixed' = Dravidian, Indo-European, Austro-Asiatic)

- R\(\rightarrow\)R30 (~64 ka, NW/N India, Nepal: mixed
- R\(\rightarrow\)R31 (~65 ka, NW/N India): mixed
- R (~38 ka/~66 ka\(^{MK}\), Central, N. India): mixed
- R6 (~52 ka. Central. N. India. Tamil Nadu): mixed
- R (~59 ka, W Eurasia) \(\rightarrow\)R2'JT (~55 ka): Pre-Semitic-Egyptian

The emergence of R-mtDNA around 74 ka appears to have occurred in the Persian Gulf Oasis before the Zagros Crossing to South Asia and it underwent multiple branchings around 65 ka. A correlation for these Early and Middle Borean branches to any surviving language family does not appear evident, as it seems the early mtDNA lineages on the distal side of the Crossing in South Asia have adopted Dravidian, Indo-European or Austro-Asiatic languages. A case may be made on the ‘back movement’ proximal side for R2’JT (~55 ka), which seems to be associated with Pre-Semitic-Egyptian. Remarkably, the highest frequency locations of South Asian R30, R31, R5 and R6 lineages occur across northern and central India, and this suggests that rather than some sort of ‘southern route’ across India to E/SE Asia dispersals followed the more direct Ganges Basin and Narmada Basin routes, with outliers dispersing into Nepal and Tamil Nadu.

3. **Ganges Delta Crossing to SE Asia (circa 50 ka, differentiating Middle Borean-R)**

R (~54 ka, E Asia) \(\rightarrow\)B4'5 (~50 ka) \(\rightarrow\)B4 (~44 ka) SE Asia in Austronesian, A-A)

- B4'5 \(\rightarrow\)B5 (~42 ka, in Kam-Tai, Tibeto-Burman)
- R \(\rightarrow\)R9,21 (~47 ka), R9 \(\rightarrow\)F (~43 ka): Pre-Austric
- R9 in Hmong-Mien, Batek, Aslian Malay, Semang; R21 in Semang, Senoi; F in Mon-Khmer, Tai A-A

- R (~58 ka, Melanesia/Australia) \(\rightarrow\)P (~55 ka)
- P \(\rightarrow\)P4 (~53 ka, Australia): Pama-Nyungan

thus = ‘southern route’ via East/SE Asia

Around 50 ka R-mtDNA encountered the geographic bottleneck at the Ganges Delta Crossing to SE Asia, and this appears to correlate to a further differentiation of Middle Borean-R macrofamilies. Out of the bottleneck’s distal side language families diverge into East Asia, SE Asia and Sahul. Current languages spoken by populations with high frequency and distinctive B4'5, R9 and R21-mtDNA subclades all appear to correlate to languages that have been reconstructed to the Austric macrofamily. Additionally R and its P subclades appear to arrive in
the Melanesia, New Guinea and Australia around 58-53 ka and current populations distinctively P4 are currently speakers of Greater-Pama-Nyungan languages. (I discuss this in more detail under #7 Sunda/Sahul Crossing below.) On the 'back movement' proximal side of the Ganges bottleneck R7 and R8-mtDNA emerge in southern and eastern India, with high or distinctive frequency in Vedda and Munda, currently speaking diverse (perhaps adopted) languages.


Meanwhile, around 50 ka, haplogroup M appears to have encountered the Ganges Delta bottleneck. On the distal side of the Crossing, branching southerly into SE Asia and northerly into E Asia, M subclades appear distinctively associated with several Middle Borean-M language macrofamilies. In the northerly dispersal, M subclades 12’G has high frequency and distinctiveness that correlates to Pre-Chukotko-Kamchatkan; M7 to Pre-Japonic-Korean-Ainu; and M8 to Pre-Altaic-Tungusic-Mongolian. In the southerly direction, M9 later adopts various SE and E Asian language families. M21a’b in SE Asia occurs in Semang, currently speakers of Aslian Austro-Asiatic. In East Asia D-mtDNA (frequent in Han Chinese) is a remarkably anomalous case (the exception that proves the rule?). Since this clade is currently associated with the Sino-Tibetan language family, this does not match the M subclade genetics. I hypothesize that this population group dropped an initial Eurasian language and adopted a Borean-N (Pre-Dené-Caucasic-Burushaski) language, which evolved into Chinese. On the proximal side of the Ganges Delta Crossing, M42’74, M33, M5 and M2 have TMRCA dates around 50 ka. They may be viewed as the ‘back movement’ at the geographic bottleneck, and high or distinctive frequencies of these haplogroups occur in populations that appear to have adopted various languages, including Dravidian, Austroasiatic, Indo-European and Tibeto-Burman. In part, it appears to be a Pre-Austroasiatic back migration from SE Asia. Interestingly, M5 is frequent in Nihal speakers.
5. SW Asia to W Eurasia, the Transcaucasus Crossing (circa 50 ka, differentiating Middle Borean-U; R→U ~54 ka).

R's subclade U emerged in SW Asia with a TMRCA ~54 ka. Major branchings of its subclades occurred between 50 and 40 ka apparently in response to the geographic bottleneck at the Transcaucasus Crossing. On the distal side of the crossing an array of subclades dispersed easterly and westerly, with some peoples remaining in the Caucasus area. Each of these mtDNA clades has high or distinctive frequency in current populations that appear to correlate to a major language family. Based on mtDNA haplogroup frequencies in various current population samples, recent genomic analyses of fossil hominins, and archaeology, I infer—in chronological order—U2 (~54 ka) correlates to Pre-Dravidian, U8 (~50 ka) to Pre-Hattic, U4'9 (~43 ka) to Pre-Kartvelian, U3 (~41 ka) to Pre-Northwest-Caucasic, RO (~39 ka) to Pre-Basque, U1 (~37 ka) to Pre-Elamitic (?) and U5 (~36 ka) to Pre-Finno-Ugric. Again I use the designation ‘Pre’ to indicate a genetic TMRCA dating for a language family that may have emerged at that date or sometime thereafter. I acknowledge that linguists may view Kartvelian as emerging tens of millennia later than U4'9's date around 43 ka. On the proximal side of the Transcaucasus bottleneck I suggest U6 (~36 ka), which represents a back-migration into Africa, across Northern Africa, has archaeological correlates and in term of language may correlate to an early Afroasiatic Berber.

6. 'Silk Route Crossing', Persian and Indus Spurs then bifurcation west to Europe, east to Central and East Asia (circa 25 ka, differentiating Late Borean-N; N→N1'5 ~57 ka; N1 ~53 ka; N2 ~44 ka).

N1b (~21 ka): Iran, Marsh Arab Iraq (adopt various) N→A (~29 ka, Caspian-Baikal): Pre-Dené
N5 (~37 ka): Madhya Pradesh Sahariya (adopts I.E.) N9'Y→Y (~22 ka): Nivkh (and in Ulchi, Ainu)
X1'3 (~32 ka)→X1 (>25 ka, Near East, N. and E. Africa): adopts various local languages

N1a (~21 ka, SW Asia, Europe, C Asia): adopt various

N→X (~31 ka)→X2 (~21 ka, SW Asia, Europe, Caucasus, Hunza, Evenk): adopt various languages
At about 25,000 years ago, the differentiations of N and N1-mtDNA appears to have occurred in response to northerly dispersal of N on routes later known as the Persian and Indus Spurs of the Silk Route, and, when arriving at the main Silk Route, bifurcating to disperse westerly toward and into Europe and easterly toward Central and East Asia. N2 offspring N2a and W have high frequency, if not homelands, in Eastern and Western Europe, and given current language association seem to have adopted various languages as they arrived at their current lands. In contrast A has a Caspian-Baikal homeland, and I would correlate it to Pre-Dené, and N9’Y’s offspring Y has highest frequency in Nivkh. X2 as well as N1a and N1e occur in high or distinctive frequencies among current populations residing along all the east, west, north and south ‘Silk’ routes, and appear to have adopted their current languages. On the proximal side of the Persian and Indus Spurs geographic bottleneck, possibly to be viewed as southerly back migrations, I suggest positing the emergence of N1b with high or distinctive frequencies in Iran and Iraq Marsh Arabs, who have been proposed as descendents of the earlier Sumerians (Al-Zahery, Pala et al., 2011) and X1’3 and offspring X1, which occur predominantly in the Near East and Northern and Eastern Africa. Tentatively, I suggest that N5-mtDNA, distinctive in the Sahariya of Madhya Pradesh and considered autochthonous for India, also may represent a proximal back movement from the Indus Spur geographic bottleneck.


M29’Q (~44 ka, Australia/Melanesia)
→Q2: (~30 or 45 ka, Irian Jaya, New Britain): Pre-Papuan (Kimberley Kalumburu adopt non-P-N)
→Q (~32 or 46 ka, Irian Jaya Ok-Muyu, Asmat, Una, Dani)→Q1 (~21 or 27 ka, PNG Bandi): Trans-New-Guinea.

P→P1 (~33 or 43 ka, Irian Jaya ‘pygmy’ Una, Ketengba; Bandi): adopt TNG, Austronesian, Papuan
P→P4 (~53 ka, Austr./Melan.)→P4b (~40/47 ka, Desert Warlpiri and NSW); P→P8 (~40 ka, Warlpiri): Pama-Nyungan

M→M21 (~46 ka, South Asia, Bangladesh, SE Asia, Semang): adopt Austronesian, Asian Austro-Asiatic
M→M22 (~41 ka, Temuan Malay): adopt Austronesian, Malayan

M (~53 ka, Australia/Melanesia)
→M42a (~41 ka or 33 ka, NSW Australia, Tasmania): in Pama-Nyungan speakers

N→S (~53 ka, Australia, Tasmania): in Pama-Nyungan and non-PN
N→O (~48 ka, SW & Desert Australia): in Pama-Nyungan speakers
Around 50 ka N, R and M-mtDNA encountered the geographic bottleneck at the Sunda/Sahul Crossing to New Guinea and Australia. Out of the bottleneck’s distal side multiple mtDNA branches emerge as peoples disperse throughout the Sahul in northerly, centrally and southerly directions. It appears that all three major mtDNA lineages, N, R and M, enter the Sahul region. R subclade P4 appears to arrive first ~53 ka, dispersing into the Central Desert and southeasterly to the New South Wales area. P4 is strongly associated with speakers of Pama-Nyungan. Around the same time N subclades S (~53 ka) and O (~48 ka) disperse into the Australian Central Desert and southwest area and into Tasmania. In current populations O-mtDNA is frequent or distinctive in Pama-Nyungan speakers, S-mtDNA in both Pama-Nyungan and non-Pama-Nyungan speakers. Whether this suggests it was people bearing R subclade P4 that spread Pama-Nyungan and N subclades S and O later adopted it, or another alternative, I leave open. While the TMRCA for M in Australia/Melanesia has a similar date, it does not appear that a specific subclade arrives in Australia until ten thousand years later, as M42a (~41 ka). By around 20 ka, M again (Q1-mtDNA) arrives in northern Australia, apparently speakers of Trans-New-Guinea languages and later, N (N12-mtDNA) in the north, apparently speakers of Gunwinyguan languages. Circa 40 ka there were also northerly dispersals out of the Sunda/Sahul bottleneck into New Guinea and Melanesia of both M and R subclades. M29’Q offspring Q2 in Irian Jaya and New Britain is predominantly found among Papuan speakers and Q1, Trans-New-Guinea. Populations with R subclade P1 in Irian Jaya and Papua New Guinea are currently speakers of Austronesian, Papuan and Trans-New-Guinea languages, perhaps one or all are adopted. In sum, it appears there was a complex diffusion across the ‘Southern Route’ to Sahul, with 3 or more waves crossing the geographic bottleneck over a time spanning at least 35,000 years.
8. Siberia to the Americas, the Beringia Crossing (circa 25 ka, differentiating Late Borean-N, R and M).

N→X (~31 ka)→X1'2'3 (~29 ka, Caucasus, SW and Central Asia)→X2 (~21 ka, SW Asia to Siberia)→X2a'j (~17 ka; X2a ~13 ka, Americas, following glacial ice edge): X-Amerind (Algonquian)

N→A (~29 ka, Caspian-Baikal)→A2 (~17 ka, N, C and S America): A-Amerind (Dené)

M8 (~43 ka, Altaic-Tungusic-Mongolian)→CZ→C (~28 ka, Caspian-Baikal, Yukaghir, Evenk, Tuvan)→C1 (~17 ka, esp. Central America): C-Amerind


M→M80'D→D (~48 ka, adopts 'N' language family, Sino-Tibetan)→D4→D4h (~21 ka, mostly South America by coastal route)→D4h3 (~18 ka, D4h3a ~13 ka) and D1 (~17 ka, mostly Chile, Argentina by coastal route): D-Amerind

Around 25 ka N, R and M-mtDNA encountered the geographic bottleneck at the Beringia Crossing from Siberia to the Americas. While Greenberg (Greenberg, 1960, 1987; Greenberg and Ruhlen, 2007; Ruhlen, 1994a, 1994b) group all Amerindian languages into three groups Amerind, Na-Dene and Eskimo-Aleut, the taxon Amerind has been challenged pro and con. I do not enter this particular debate as a linguist but with respect to the archaeogenetics I suggest it is possible to differentiate by the primary mtDNA haplogroups what I term X, A, C, B and D-Amerind. Those with the most evident correlation to a language family are X-mtDNA and Algonquian speakers and A-mtDNA and Dené speakers. From a mythological perspective, B-mtDNA, which in the Old World is found in SE Asia into E Asia, I suggest correlates with Berezkin’s discovery of ‘Indo-Pacific’ mythology in the Americas, and especially South America (2010a, 2010b). In any case, out of the Beringia bottleneck’s distal side mtDNA subclades and associated language families emerge to disperse occur across North America and along the coast from North to Central to South America, and thence inland. X2a-mtDNA with the Algonquian language family appears to have followed the northern North American glacial
ice edge. B-Amerind and D-Amerind appear to have taken the southerly coastal route all the way to Tierra del Fuego. With respect to the scorpion diagram I place A2 and C1 in the middle since A2 (Dené) and C1 are each found across North, Central and South America, while C1 has highest frequencies in populations in Central America. On the ‘back movement’ proximal side of the Beringia bottleneck M12’G gives rise to G around 35 ka and G1 around 22 ka, with highest frequencies in Itelmen and Koryak, speakers of Chukotko-Kamchatkan.

Conclusions.

- A meta-pattern-analysis of the mitochondrial DNA phylotree and current distribution of language families indicates that over the last 200,000 years there are robust correspondences between mtDNA haplogroups and language macrofamilies. This study is a thought experiment, a top-down derivation of the Homo sapiens sapiens (‘Proto-Human’, ‘Proto-World’) language phylotree, which can be tested against bottom-up prehistoric linguistic reconstructions. It establishes a relative chronology for dating the emergence and branching of the global array of language macrofamilies. The language phylotree is crosschecked against archaeological data and fossil mtDNA studies, which support many of the correlations.

- The hypothesis of this study is that there is a rough 1:1 correspondence between the 200,000-year mtDNA phylotree and its TMRCA haplogroup dates and the emergence of language macrofamilies. A meta-pattern-analysis of the mtDNA phylotree, archaeogenetics and archaeology appears to support this hypothesis.

- The analysis provides a relative timeline for the emergence and branching of all the language macrofamilies of Homo sapiens sapiens language (‘Proto-Human’, ‘Proto-World’), which may prove useful for linguistic reconstructions of proto-Sapiens-Sapiens and for reconstructions of the prehistory of mythological and ritual systems both within-Africa and out-of-Africa.

- Proto-Sapiens-Sapiens appears to have emerged with the earliest stage of fossil Homo sapiens sapiens at Early Middle Stone Age Omo Kibish, Ethiopia, around 195,000 years ago. Late dating of ‘human’ language origins to 45, 60 or even 100 ka is contradicted by mtDNA archaeogenetics as well as archaeology.

- In this proposed timeline, click languages, strongly associated with L0-mtDNA, diverged from all other languages around 160,000 years ago. The Niger-Congo language family, robustly correlated to L1-mtDNA, emerged around 140,000 years ago. Around 120,000
years ago the ancestor of Central Sudanic, apparently correlating to L5-mtDNA, diverged from a pre-Nilo-Saharan-Afroasiatic macrolanguage. At this time period the First Wave Dispersal out-of-Africa occurs bearing L2³4⁶-mtDNA and Pre-Nilo-Saharan-Afroasiatic.

- Around 100,000 years ago during the period of the Lake Paleo-Chad and central Sahara corridor the ancestral divergence occurred between Afroasiatic and Nilo-Saharan macrolanguage families, the former associated with L2-mtDNA and the latter, L3⁴⁶. Archaeology at this stage evidences a Second Wave Dispersal out-of-Africa into SW Asia at Aybut Auwal, Oman, which would have carried one or both of these ancestral languages, proto-Austroasiatic and/or proto-Nilo-Saharan.

- Archaeology in SW Asia around 85 ka provides evidence for at least two out-of-Africa industries, including two sites in the Sinai, with Nubian Complex and Nile Denticulate Mousterian, and one in northern Saudi Arabia at the Jubbah paleolake, with Nubian Complex affinity, perhaps correlating to L3⁴. This appears to be a continuation of the Second Wave dispersals out of Africa. The Nubian Complex MSA industry could have been bearers of an ancestor of the Nilo-Saharan proto-Northern Sudanic (Kunama) or proto-Koman (Gumuz, Uduk) language families (in Ehret 2011 terms) and the Nile Denticulate Mousterian, an ancestor of the Afroasiatic Boreafrasian language family.

- Around 80,000 years ago (late MIS 5a) a Third Wave Dispersal out-of-Africa occurred bearing the L3 subclades M and N. Archaeology and mtDNA genetics again imply at least Nubian Complex and Nile Denticulate Mousterian cultural traditions diffused into SW Asia, correlatable to Northern Sudanic and/or Koman and Boreafrasian language families, and also to Proto-Saharan-Sahelian (Kanuri) or Proto-Eastern Sahelian (Nubian), and these interacted with SW Asian indigenous populations having Tabun C industries. In inferring this I am not equating M and N respectively to these two traditions, but dispersing populations probably had varying admixtures of northeastern African L3M and L3N as well as northern African L2a. A similar argument would apply to their ritual-myth-and-art traditions.

- There is currently insufficient archaeological evidence to determine whether routes out-of-Africa were via the Sinai, the Bab-el-Mandeb or circum-Red Sea or some combination thereof.

- In terms of mythostratigraphy out-of-Africa, based on archaeology and mtDNA genetics, I suggest the best inference would be that the dispersals out of Africa would have carried
myth-ritual systems that combined to greater or lesser extent components of the two major North and East African myth-systems, which evolved prior to the TMRCA of L3 and which may be termed ‘North African’ (correlating to L2, L2a) and ‘Sudanic’ (correlating to L5, L4) or ‘Nile-Sudanic’ (correlating to L3 and its subclades). The Nile-Sudanic myth-system would have emphasized ancestral and game spirits, where ‘spirit’ means a life-giving, life-animating and life-enhancing forces or energies. This is a religious system having neither a high god nor deus otiosus. The North African religious system would appear to have had a creative power, female or androgynous Creatrix, who organized the cosmos with a world-axis, four directions and their associated complementarity principles, and thereby established the nature of life’s unfolding as one of balance and complementarity of polar or gendered pairs in all their variant combinations and recombinations.

- Around 75,000 years ago L3-mtDNA’s out-of-Africa dispersal into SW Asia had a pause of up to 5,000 years, during which N and M differentiated subclades and N branched off R-mtDNA and its subclades. Geographic bottlenecks at the Transcaucasus Crossing to Western Eurasia and Zagros Crossing to South Asia, extant Neanderthals in both directions and other archaic species, and possibly the Toba supereruption (~74 ka) and positive subsistence landscape of the ‘Persian Gulf Oasis’ presumably contributed to this extended delay.

- Three major Borean language families appear robustly associated with the differentiation of N, R and M-mtDNA clades, and I term these Borean-M, Borean-N and Borean-R. Respectively, M-mtDNA corresponds strongly to the Eurasianic language family; N-mtDNA to the Dené-Caucasian, and R-mtDNA to Afroasiatic. Kartvelian, Dravidian, etc. I suggest that Austric and Pama-Nyungan are most closely associated with R-mtDNA. Borean-R appears to have taken a ‘northern route’ diffusing branches across northern South Asia and into Southeast Asia and Sahul, and also into Europe, Central and East Asia. While at first glance this might appear contradictory to current language macrofamily reconstructions, it seems supported by Berezkin’s discovery using principle component analysis (2010a, 2010b) of similar mythological motifs shared by peoples in three disparate geographic regions: Indo-Pacific, South American and western Eurasia/Europe.

- The hypothesis for a sapiens sapiens ‘southern route fast track’ of a few thousand years from SW Asia to Sahul, e.g., leaving Africa around 50, 60 or even 70 ka is not supported by mtDNA genetics or archaeology. On the contrary the Third Wave Dispersal from
Africa began circa 80 ka and arriving in Sahul circa 55 ka apparently took around 25,000 years, including an up to 5,000 year delay in SW Asia.

- By around 50,000 years ago (MIS 3c/b), based on mtDNA phylotree and its TMRCAs, it appears that roughly 28 new mtDNA haplogroups had emerged, and based on current languages associated with them, at least 17 of them, from Europe to East Asia and Sahul, correlate to the emergence of 17 language families—which I term ‘Middle Borean’.

- By around 25,000 years ago another dozen language families emerged—which I term ‘Late Borean’—and this, for the most part, appears to have completed the development of the major language families of the world.

- The 1:1 correlation of mtDNA haplogroups and language families generally appears more robust at earlier stages of the phylotree than recent stages. In the latter language replacements and adoptions appear more frequent and current population genetic samples sometimes more admixed. There is one notable anomaly to the 1:1 correlations for Middle Borean languages around 50,000 years ago. Current Han people are especially associated with mtDNA haplogroup D, which belongs to the M-mtDNA clade. They would be expected to have a Borean-M Eurasian language. Instead their Chinese language is classified as belonging to the Dené-Caucasian family (Borean-N). Thus the genetics suggests that the Han may have initially been speakers of a Eurasian language and later adopted a Dené-Caucasian language, which evolved into Chinese. Linguists might explore this possibility further. While the case of Han Chinese might be taken to invalidate my basic hypothesis of a 1:1 correlation of major mtDNA haplogroups and language macrofamilies, I suggest that the correlation appears to hold in general and this Han Chinese exception appears to be the exception that proves the rule.

- Finally, it appears that Fleming’s Borean model (Fleming, 2002; 1991; 1987; Fleming, Zegura et al., 2013) with its 3 major subclades maps almost precisely onto the 3 primary branches of mtDNA out-of-Africa and their correlated language macrofamilies. Fleming’s cluster of Afrasian (Afroasiatic), Kartvelian, Dravidian, Elamitic, and other SW Asia extinct languages maps onto descendents of R-mtDNA and more precisely its U-mtDNA branches; Caucasian-Burushaski-Déné maps onto N-mtDNA and Eurasian onto M-mtDNA. Archaeogenetics further supports adding SE Asian/Sahul languages, including Austroic, Trans-New-Guinea, Papuan and Pama-Nyungan, as a fourth cluster to Fleming’s Borean, as argued for by Gell-Mann, Peiros and Starostin (2009) and correlating these languages to the Borean-R language phylum.
Limitations.

Evidence for correlating some languages to mtDNA haplogroups appears more robust than others; it seems by and large that correlations become weaker the closer we approach recent times. As a nonlinguist, I am aware that there are complex language replacement issues pertaining to mismatches between the languages populations currently speak and those they may have spoken in historic or prehistoric times. The language correlations that I have proposed are more or less tentative. I am open to counterarguments with respect to which language to correlate to particular branches in the mtDNA phylotree.

In this study I have chosen to first look at the archaeogenetics of mtDNA—or so to speak, the ‘mother tongue’. The archaeogenetics of Y-DNA remains for future research.

Because of the complexity of cross-mapping databases for mtDNA archaeogenetics, language macrofamilies, and archaeology and the constant advances in each field, the metatransource proposed in this study is necessarily tentative and open to revision.

Acknowledgments.

I offer my gratitude to Hal Fleming, and colleagues of the *Mother Tongue* journal, for the challenge and encouragement to bring my interest in mythology, archaeology and archaeogenetics to the question of long-range language origins. I thank Luca Pagani for generating the haplogroup assignments and frequencies for his Gumuz sample that I might refer to them in this study. Also thanks to the other geneticists who answered over the years emails in which I asked so many neophyte questions. May this study also honor a mentor, the archaeologist, linguist and folklorist Marija Gimbutas, who inspired and encouraged me to enter the field of paleolithic myth and symbol.

Supplementary Information.

The mtDNA Database for Archaeogeneticlinguistics containing approximately 181 pages, 424 ethnic and population mtDNA haplogroup frequency samples, their current spoken languages, 82 fossil mtDNA studies, and select archaeological sites, with references, is available online at [https://originsnet.academia.edu/JamesHarrod](https://originsnet.academia.edu/JamesHarrod)
APPENDIX 1: Out-of-Africa mtDNA and Correlated Language Family Dispersal Maps

Figure 1. MIS 6 (~190-130 ka). Early Homo sapiens sapiens in East Africa and other archaics in Africa. Neanderthals in Europe and Denisovans in Central and Eastern Asia, Indian archaics in the Indian Subcontinent and Sundaland.

MIS 5e/d (~130-106 ka) and MIS 5c/b (~106-85 ka). Based on archaeogenetics and archaeology Homo sapiens sapiens 'with robust archaic features' and Middle Paleolithic technologies expand across Northern Africa and out-of-Africa into SW Asia, in what appears to be two successive waves. L2'3'4'6 diffuses across Northern Africa (Magreb Levallois Mousterian and Aterian; Bir Tafawi Aterian and Early Nubian Complex; Abur MSA with handaxes) and into SW Asia (Jebel Faya with handaxes). Fossil Zhirendong, south China, at minimum ~106 ka, suggests expansion across South Asia into East and SE Asia. A second diffusion, probably L3'4'6, into SW Asia occurs during MIS 5c/b (Mousterian Complex; Sinai Split Rock-Lower Denticulate Mousterian; Aybut Auwal Nubian Complex). Indian archaics retreated from Sundaland submerged by higher sea level, Mishra, Chauhan and Singhvi (2013).

mtDNA haplogroup TMRCAs from Soares, Ermini, et al (2009), with caveat that for MIS 6 and MIS 5e/d Soares TMRCAs are multiplied by a factor of 1.1x to accommodate redating of chimpanzee/human split (Langergraber, Prüfer et al 2012; their redating is 1.1x CHLCA used by Soares; to be conservative I use a 1.1x multiplier). Also for L3'4'6 and L3'4 I have averaged Soares TMRCAs with Behar, van Oven, et al (2012) TMRCAs times 1.1. Correlations of haplogroups to ancestors of associated language families are author's tentative hypotheses. Maps are author's archaeogenetic overlays and modifications of maps in Mishra, Chauhan and Singhvi (2013 fig.2), which provide a summary display of global population movements with respect to South Asia and challenge the single-diffusion southern route fast track out-of-Africa to SE Asia model.
'Out-of-Africa-into-SW-Asia'

L3 - 79 ka 'Pta-Mroasiatic-Bonan'  

Figure 2. MIS 5a (~85-74 ka). Around 80 ka a third wave of *Homo sapiens sapiens* with haplogroup L3 and Middle Paleolithic technologies diffuses across Northern Africa (Ifri n'Ammar Aterian; Pigeons Taforalt Aterian; El-Guetta 'Final Mousterian'; Bir Tarfawi Aterian; Taramsa 1-Phase III Levallois and Nubian) and out-of-Africa into SWAsia (possibly Jebel Qattar Tabun C; Sinai Split Rock Upper Denticulate). Bottlenecks due to geography, Neanderthal movement into SWAsia and the Toba supereruption (~74 ka) prevented further diffusion. The latter also contributed to fading away of the Indian archaics.

During MIS 4, L3 subclades expand in Africa, while in SW Asia, after about a 4,000 year stasis and post-Toba, modern humans disperse out-of-SW Asia in at least three waves along three northern routes. M and N diverge as they migrate mostly north of the Himalayas along the 'silk route'. Shortly thereafter remaining N in SW Asia branches off R which further branches off U clades in SW Asia and moves south of the Himalayas across northern India, branching off R31-30, and from there into SE Asia and eventually Australia. Presumably all three of these MIS 4 waves carry Middle Paleolithic tool-and-art kits. Archaic populations contribute some admixture into these dispersing modern human lineages.

Figure 3  MIS 4 (~74-59 ka) and MIS 3c/b (~59-40 ka). During MIS 4, L3 subclades expand in Africa, while in SW Asia, after about a 4,000 year stasis and post-Toba, modern humans disperse out-of-SW Asia in at least three waves along three northern routes. M and N diverge as they migrate mostly north of the Himalayas along the 'silk route'. Shortly thereafter remaining N in SW Asia branches off R which further branches off U clades in SW Asia and moves south of the Himalayas across northern India, branching off R31-30, and from there into SE Asia and eventually Australia. Presumably all three of these MIS 4 waves carry Middle Paleolithic tool-and-art kits. Archaic populations contribute some admixture into these dispersing modern human lineages.

TMRCAs for MIS 4 for the most part are from Soares, Ermini, et al (2009) conservatively multiplied by 1.1 for the revised CHCLA, with several caveats. Soares, Ermini, et al (2009) dates N at 71.1 = 70 ka, but Soares, Alshamali, et al (2013) reduces this to 63.1 = 60 ka. This reduction results in a clock violation with the earlier Soares TMRCAs for R mtDNA (67.1 = 74 ka) and all subsequent branches. Thus, I suggest the range 75-78 ka between R at 74 ka and L3 at 79 ka. For M4*87 and M2 in South Asia, I select the TMRCAs from Thangaraj, Chauhan, et al (2008, compare Kumar, Padmanabhan, et al 2009).

TMRCAs for MIS 3c/b are all taken from Soares, Ermini, et al (2009) and I have applied no multiplier, as TMRCAs during this more recent time period seem satisfactory with respect to archaeology and use of multiplier appears to result in numerous clock violations. For N2 Soares provides no TMRC, I use Fernandes, Alshamali, et al (2012). Similarly, for N5 and M29’Q I use Behar, van Oven, et al (2012); for M42’74 I use Kumar, Ravuri, et al (2009); and for M in Southeast Asia I use Jinam, Hong, et al (2012).

From the beginning of the next stage, MIS 3a ~40 ka, Later Stone Age/Upper Paleolithic microblade technologies occur across Africa, SW Asia, South Asia, Europe and beyond. Each region likely saw innovations by a mix of indigenous groups and cross-region diffusion. For example South Asia microblade industries may have been innovated by R31-30 and M4*87 mtDNA descendents out of their Mode 3 technologies along with diffusions via N2/N5 and U2 lineages from the northwest.
### Derivation of Dates for mtDNA and Correlated Language Family Dispersal Maps

<table>
<thead>
<tr>
<th>MIS 6 (190-130)</th>
<th>Approximate TMRCA</th>
<th>Sources and Calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eve</td>
<td>195</td>
<td>Soares: 192 ( x_1.1 = 211 ), Gonder: 194, BvO: 177 ( x_1.1 = 195 )</td>
</tr>
<tr>
<td>L1-6</td>
<td>184</td>
<td>167 ( [L(3-7)] ) ( x_1.1 = 184 ), 153 ( [L1-5] )</td>
</tr>
<tr>
<td>L0</td>
<td>165</td>
<td>150 ( x_1.1 = 165 ), 146, 136</td>
</tr>
<tr>
<td>L1</td>
<td>155</td>
<td>141 ( x_1.1 = 155 ), 128.5</td>
</tr>
<tr>
<td>L5</td>
<td>132</td>
<td>120 ( x_1.1 = 132 ), 129, 111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIS 5e/d (130-106)</th>
<th>Approximate TMRCA</th>
<th>Sources and Calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2346</td>
<td>126</td>
<td>115 ( [L34762] ) ( x_1.1 = 126 ), 111 ( x_1.1 = 122 )</td>
</tr>
<tr>
<td>L3’4’6</td>
<td>97</td>
<td>105 ( [L3476] ) ( x_1.1 = 115 ), 71 ( x_1.1 = 78 )</td>
</tr>
<tr>
<td>L3’4</td>
<td>83</td>
<td>86 ( [L347] ) ( x_1.1 = 95 ), 64 ( [L34] ) ( x_1.1 = 70 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIS 5a (85-74)</th>
<th>Approximate TMRCA</th>
<th>Sources and Calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>L3</td>
<td>79</td>
<td>72 ( [S2009] ) ( x_1.1 = 79 ), Gonder: 97, OGA: 84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIS 4 (74-59)</th>
<th>Approximate TMRCA</th>
<th>Sources and Calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>75-78</td>
<td>71 ( \text{SoAs} ) ( x_1.1 = 78 ), 626 ( x_1.2 = 74 ), 49 ( x_1.1 = 68 )</td>
</tr>
<tr>
<td>M</td>
<td>67</td>
<td>60.6 ( x_1.1 = 67 ), 51260.5 ( x_1.2 = 73 ), 49 ( x_1.1 = 67 )</td>
</tr>
<tr>
<td>N→R</td>
<td>74</td>
<td>675 ( \text{SoA} ) ( x_1.2 = 80 ), 56.5 ( \text{geo indet} ) ( x_1.2 = 68 ), 56 ( x_1.1 = 62 )</td>
</tr>
<tr>
<td>R31 R30</td>
<td>71</td>
<td>64 ( 56 \text{NNNSa} ) ( x_1.2 = 77 ), 55 ( R31 ), 53 ( R30 )</td>
</tr>
<tr>
<td>U</td>
<td>59</td>
<td>54 ( x_1.2 = 65 ), 47 ( x_1.2 = 66 ), 52 ( A )</td>
</tr>
<tr>
<td>R2J T</td>
<td>60</td>
<td>55 ( x_1.2 = 66 ), 54 ( x_1.2 = 65 )</td>
</tr>
<tr>
<td>B</td>
<td>56</td>
<td>64 ( B ) ( 51 ), 56 ( x_1.1 = 56 ), 54 ( R = 54 )</td>
</tr>
<tr>
<td>P</td>
<td>64</td>
<td>58 ( \text{Aus/Mel} ) ( x_1.2 = 70 ), 53 ( B ) ( 49.5 ), 54 ( x_1.2 = 59 )</td>
</tr>
<tr>
<td>M4’67</td>
<td>62</td>
<td>40.5 ( x_1.2 = 49 ), 35 ( x_1.2 = 42 ), 38 ( T = 60 )</td>
</tr>
<tr>
<td>M2</td>
<td>60</td>
<td>34 ( x_1.2 = 46 ), 36 ( x_1.2 = 43 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MIS 3c/b (59-40)</th>
<th>Approximate TMRCA</th>
<th>Sources and Calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>54</td>
<td>54 ( N = 57 ), 44 ( N5 = 37 ), 49 ( FA = 50-63 )</td>
</tr>
<tr>
<td>N2</td>
<td>44-51</td>
<td>49 ( 46 ), 48 ( SV = 39 )</td>
</tr>
<tr>
<td>N9</td>
<td>49</td>
<td>55 ( 45 )</td>
</tr>
<tr>
<td>M7</td>
<td>55</td>
<td>48 ( 38 )</td>
</tr>
<tr>
<td>D</td>
<td>48</td>
<td>43 ( 36 )</td>
</tr>
<tr>
<td>M8/CZ</td>
<td>43</td>
<td>43 ( 36 )</td>
</tr>
<tr>
<td>M42’74</td>
<td>55</td>
<td>43 ( 36 )</td>
</tr>
<tr>
<td>M33</td>
<td>45</td>
<td>50 ( KR = 55 )</td>
</tr>
<tr>
<td>M-SEA</td>
<td>44-52</td>
<td>45 ( 42 ), 47 ( H = 53 )</td>
</tr>
<tr>
<td>M29’Q</td>
<td>47</td>
<td>55 ( 43 )</td>
</tr>
<tr>
<td>U2</td>
<td>55</td>
<td>50 ( 43 )</td>
</tr>
<tr>
<td>U8</td>
<td>50</td>
<td>45 ( 37 )</td>
</tr>
<tr>
<td>U4’9</td>
<td>45</td>
<td>50 ( 47 )</td>
</tr>
<tr>
<td>JT</td>
<td>50</td>
<td>50 ( P = 56-58 )</td>
</tr>
</tbody>
</table>
OLD L3 calculation from MT 2013
If all 4 dates L3 and L4 ave = 71x1.2= 85; if elim outliers, 67 and 72 ave=69.5x1.2 = 83;
if 65x1.2=78 or 66x1.2=79 or 72x1.2=86

A. Achilli, Rengo, Magri et al., ‘H Franco-Cantabrian Refuge’ 2004
F. Fernandes, Alshamali et al., 2012
G. Gonder, Mortensen et al., 2007. Table 2. Uses Sanderson 1997, 2002, 2003; and CHLCA
6.0+0.5 MYA.
H. Hudjashov, Kivisild et al., 2007 using Kivisild 2006 dating
J. Jinam, Hong, Phipps et al., 2012
K. Kumar, Padmanabham, Ravuri et al., 2008
K. Kumar, Ravuri, Koneru et al., 2009
P. Pala, Olivieri, Achilli, et al., 2012
S. Soares, Alshamali et al., 2012
S. Soares, Ermini et al., 2009. Uses complete genome clock; average of methods, 1 mutation per
7884 years; CHLCA 6.5+0.5 =7 MYA. I have taken Point Estimate dates from Sup. Info. Table
S5, which uses complete genome and maximum likelihood analysis; and also indicate Table 3
dates for L3, M, N, R and U, which use overlapping interval (OvInt) of complete genome rho
from two different clocks and maximum likelihood analysis estimates.
S. Sukernik, Volodko, Mazuin et al., 2012
T. Thangaraj, Chaubey, Singh et al., 2006
References

(Only references directly mentioned in this narrative; references for the Supplementary File, Table 1 Master Database table run into the hundreds and can be consulted in that table, http://www.originsnet.org/publications.html.)


The Kinship Term KUKU ~ KOKO ~ KAKA in the American Indigenous Languages, the Amerind Hypothesis, and the Dravidian Kinship System

Part I: Linguistic Study

Alain Matthey de l’Etang & Pierre J. Bancel

Abstract. The study of kinship terminologies in American indigenous languages unequivocally shows that the terms KOKO ~ KUKU ~ KAKA ‘MB, EF’ or ‘GM, FZ’ are present in most regions of the double continent, in a vast majority of linguistic families as well as in many languages not yet classified or considered isolated. The distribution of this term cannot ultimately be explained in terms of areal diffusion, as the continental distribution of the linguistic reconstructions demonstrates, but points to an ancestry within each linguistic family with time depths of several millennia. The antiquity of this term, its global distribution across the Americas, its phonetic properties and its semantic consistency, along with the fact that, in all probability, the initial peopling of the Americas was the result of a major colonizing event by a single source population, are consistent with the hypothesis that one or several KOKO ~ KAKA terms were present in the kinship lexicon of this founding population, which on genetic, archeological and geographical grounds may have entered in America as early as 16,000 years ago.

Abbreviations: P ‘parent,’ G ‘grand,’ M ‘mother,’ F ‘father,’ Z ‘sister,’ B ‘brother,’ U ‘uncle,’ A ‘aunt,’ E ‘spouse,’ e/y ‘elder/younger,’ W ‘wife,’ H ‘husband,’ C ‘child,’ S ‘son,’ D ‘daughter,’ inL ‘-in-law,’ Sib ‘sibling.’ Other relationships are obtained by combination of these primary symbols: MB ‘mother’s brother,’ GM ‘grandmother,’ etc. The symbols (♀) and (♂) found before kin type abbreviations indicate the sex of the person speaking; ad.: vocative or address term; ref.: referential term; AM & PB: Matthey de l’Etang & Bancel; EHL: Evolution of Human Languages project (ehl.santafe.edu/introl.htm); ASJP: Automated Similarity Judgment Program (http://wwwstaff.eva.mpg.de/~wichmann/ASJPHomePage.htm); R: ‘reconstruction done by’; LDRC: University of Alberta, Language Documentation Research Cluster; P-: ‘Proto-’; dial.: dialect.

1. Presentation

Our first goal in this paper is to make a general appraisal of the distribution in the Americas of a term long and well known to Amerindianists, which can be labeled KOKO and has the general semantic scope of GF, MB, EF for the masculine side and GM, FZ and EM for the feminine side. Our second goal is to give an explanation for its amazing distribution, which was first remarked on

---

1 This paper (including the anthropological developments) was presented at the annual meeting of the American Anthropological Association (kinship session) held in San Francisco on November 14-18, 2012.

1 Association d’études linguistiques et anthropologiques préhistoriques, Paris. First author’s mail: <a.matthey@free.fr>.
by von Martius (1867a & b) during the first half of the 19th century in South America², and by other authors at the turn of the 20th century for Meso-America. Moreover, more than a hundred years after von Martius, Landar (1977: 423) stated that, although uncharted, “The areal distribution of koko-type words for kinmeu ... will involve North as well as South America.” Still more recently, Ruhlen (1994b: 218-19) confirmed Landar’s claim, arguing in favour of the Proto-Amerind antiquity of KAKA, and furthermore proposing a joint Amerind-Eurasiatic etymology: Amerind *(k)aka eB, eZ = Eurasiatic *aka eB. The same year, Ruhlen (1994a: 122-124) published the global etymology kaka ‘U, eB,” showing that this root was not only widespread in the Americas, but also in Eurasia and Oceania, a distribution that we later widened to many more families from these areas as well as from Africa, Southern Asia, Australia, and New Guinea. As a result, KAKA or KOKO words are attested in most language families worldwide, with reflexes in more than 1,300 languages in our database of some 2,400 kinship terminologies (Bancel & Matthey de l’Etang 2002; Bancel, Matthey de l’Etang & Bengtson 2011; Matthey de l’Etang & Bancel 2002, 2008; Matthey de l’Etang, Bancel & Ruhlen 2011)³.

More recently, Dixon & Aikhenvald (1999: 8), Aikhenvald (2002: 294), Dixon (2004a: 13, 2004b: 17), apparently ignoring Ruhlen’s publications, explained the South American lowland (Amazonian) distribution of KOKO in terms of areal diffusion. The two authors believe that what triggers such a chain of borrowings is the need, for societies practising intermarriage, to agree on the term referring to the father-in-law (EF), hence KOKO. Even if it is difficult to figure out exactly what their “across-Amazonia” distribution means in terms of the languages concerned, their conclusions oblige us to reconsider the distribution, meaning and origin of this term. Do KOKO terms found all over the Americas, beyond the particular destinies of their multiple forms, ultimately descend from a common source, from multiple sources, or are they the result of some gigantic cascade-borrowing process? In order to reach convincing answers, we will then specify the distribution of KOKO terms in various language families of South, Central and North America.⁴ This will confirm, amplify and specify the presentations given by Ruhlen in 1994.

Our third goal is to unveil the nature of the kinship system of the group or groups that first reached the American Continent. This particular issue will be the object of the forthcoming second (anthropological) part of our paper.

Sections 2-4 present the geolinguistic distribution of KOKO forms and illustrate how reconstructions have already been worked out by linguists in a substantial number of linguistic families all over the Americas, some of which are at a time depth of over 5,000 years.⁵ We also present potential cognate sets from other language families. In doing so, we will bring to light some deviant forms within a number of cognate sets, explaining them by phonetic or morphological

---

² Von Martius, while pointing out the distribution of forms such as ghuk or koko across the continent – most of the time referring to the uncle or the mother's brother –, classified languages using this term in what he called the "coco or guck group" (Martius 1867a: 346-347).

³ We have now collected data from about 3,400 languages and dialects. See our database at language-kinship.org.

⁴ Languages have been arrayed according to the classifications adopted by Campbell (2012) for South America, by Kaufman in the Atlas of the World’s Languages (2nd edition 2007) for Middle (Meso)America, and by Golla, Campbell, Mithun, Mixco and Goddard, in the same atlas, for North America.

⁵ The dates given for proto-languages are taken from Kaufman (1994 [2nd edition 2007]), from the Automated Similarity Judgment Program consortium (http://wwwstaff.eva.mpg.de/~wichmann/ASJPHomePage.htm) or from the authors referenced in the appropriate columns of our tables.
properties, or sometimes by their probable borrowing from languages belonging to other linguistic stocks. Section 5 presents a synthesis of the reconstructions. Section 6 discusses the origin of KOKO in the Americas linking linguistic and genetic studies. Our conclusion will be that the distribution of these terms with highly consistent phonetic forms and meanings in numerous reconstructed proto-languages of the whole continent definitely precludes that they might have resulted from cascade borrowings, especially as most reconstructed families display neatly individualized forms that have consistently been preserved in an overwhelming majority of their respective descendants, to the exception of a small number of easily identifiable borrowings. This distribution is consistent and best explained by KOKO being a retention from Proto-Amerind, a hypothetical ancestral language, nowadays substantially supported both linguistically and genetically.

2. The kinship term KOKO in South America

The first records of KOKO kinship terms in South America date from the 16th and 17th centuries: Tarascan cucu [kuku] GM (Gilberti 1559), Quechua caca [kaka] MB (Santo Tomás 1560a & b), Island Carib neiékécayem [no-kaka-jem] ‘FZH, MZH, MB’ (Breton 1665, 1666; with a 1st person possessive prefix no-), Kipea (Kariri) i-cucu [kuku] MB (Mamiani 1698: 197), Muisca (Chibcha) caca [kaka] GM (anonymous manuscripts dating back to the beginning of the 17th century: ms. 158, voc. fol. 9r, at the National Library of Colombia, Bogota, and ms. 2923, fol. 4r, at the library of the Palacio Real, Madrid). The first author to mention the existence of the term KOKO on a large scale was von Martius (1867a & b). Others authors since then (Koch-Grünberg 1911; de Créqui-Montfort & Rivet 1921, 1922; Rivet 1948; de Goeje 1909, 1946; Schuller 1928; Dixon 2004a & b) collected KOKO terms in various linguistic families such as Arawakan (Maipuran), Takanan, Panoan, Cariban, Chibchan, Quechuan, Aymaran, Mondé, Nambikuaran, and Karirian. We also present the Lencan and Misumalpan material (not included in Campbell 2012) in section 2, because there is convincing

2.1. The geographic and linguistic distribution of KOKO in South America

Campbell’s (2012) classification of South American indigenous languages comprises 108 language units: 53 families comprising at least 2 languages, and 55 language isolates. Arraying our data according to this classification does not mean that we necessarily adhere to its preamble, but its conservative formulation allows us to present data without embarking into unnecessary controversy.

KOKO forms are found in 53 units out of the 108 considered in Campbell (2012): 35 families out of 53 comprising at least 2 languages, and in 18 language isolates out of the 30 for which we have data (63.9% of groupings). The families where one or several KOKO forms have previously been reconstructed or suggested are, in the order where they are presented below: Arawakan (Maipuran), Pano-Takanan (Panoan, Takanan), Arawan, Cariban, Guahiboan, Chibchan, Quechuan, Aymaran, Mondé, Nambikuaran, and Karirian. We also present the Lencan and Misumalpan material (not included in Campbell 2012) in section 2, because there is convincing

---

6 This wide distribution of KOKO may be underestimated, as this word is primarily an address form, a category sometimes neglected by field linguists and anthropologists.
evidence that both these language families are related to Chibchan (Constenla Umaña 2012). The other language units displaying KOKO forms are presented in subsection 2.5.

KOKO forms have been looked for in kinship terminologies from some 407 languages and dialects of South America and found in all the geolinguistic units defined by Kaufman (2007) (map1): North-Western Western Amazonia, Northern Foothills, Andes, Southern Foothills, the Cone, Eastern Brazil, North-Eastern Brazil, and Central Amazonia. The Lenca-Misumalpan forms present in Nicaragua, Honduras, and El Salvador are shown on map 2.

2.2. The phonetic series KOKO in South American linguistic families

Our list of potential cognates has been established taking KOKO ~ KUKU as a starting point, because of the obvious and long-known existence of such forms in South American indigenous languages. We admitted to our list kinship terms presenting no or almost no phonetic difference with this KOKO ~ KUKU model, as well as forms displaying vocalic differences, like kuka, koka or kaka. We also admitted non-reduplicated or partially reduplicated forms like ghuk, uka, gu, ko, and finally forms likely to display a phonetic evolution like axa, axu, etc., with velar fricatives most likely reflecting an earlier velar stop k or g. Finally, we also took into account forms prefixed with a 1st person singular possessive, like Curripaco (Arawakan) nu-kui-ru ‘my FZ’, or nu-ki-ri ‘my MB’, because their roots can be easily identified and related to their appellative counterparts kuku FZ and kiiki MB. Out of the 407 kinship terminologies under examination, 300 (73%) exhibit 448 likely cognates. Table 1 recapitulates the various subtypes encountered. This phonetic series is remarkably consistent and homogeneous, as 188 out of 448 reflexes (42%) are CVCV (kVkV) forms. Among these, reduplicated koko and kuku forms are predominant.

<table>
<thead>
<tr>
<th>Phonetic forms</th>
<th>KOKO</th>
<th>KUKU</th>
<th>KAKA</th>
<th>Other kVkV forms</th>
<th>Other forms</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>57</td>
<td>41</td>
<td>25</td>
<td>65</td>
<td>260</td>
<td>448</td>
</tr>
<tr>
<td>Percentage</td>
<td>12.7</td>
<td>9.2</td>
<td>5.6</td>
<td>14.5</td>
<td>58</td>
<td>100</td>
</tr>
</tbody>
</table>


8 This figure includes the Lenca-Misumalpan languages. Out of these 407 languages, 335 dialects and languages are still alive (299 languages, 36 dialects).

9 Kaufman’s (2007) classification arrays languages and language families along geolinguistic regions.

10 A language can exhibit more than one reflex of KOKO referring to one or several relationships. Whenever a language displays variants of KOKO to refer to the same relationship, only one has been taken into consideration for our general count. In some cases, one reflex can also refer to both feminine and masculine relationships: GP = GF and GM. Such reflexes are counted twice when dealing with the semantic series.
2.3. The reconstruction of KOKO in South American linguistic families

Speaking about the presence of the form *koko in every language of the Arawan family, Dixon (2004b: 17) says that “a similar form is found in many other Amazonian languages [from other families, AM & PB], constituting a clear areal feature. It could have been borrowed from anyone of a number of sources – in Proto-Arawan or else into individual languages.” At the same time, the same author also says that “the Proto-Arawá term is koko” (Dixon 2004a: 13). The idea which emerges from his remarks is that the distribution of KOKO terms throughout Amazonia might result from an ongoing but already ancient process.

But can one maintain such an idea, if similar forms can be reconstructed at the Proto-language stage of other indigenous language families of America – some of them very ancient —, especially when their respective homelands are far from one another, some of them even far from Amazonia?

We will review and discuss the reconstructions already achieved by previous authors, and will also propose obvious cognate sets from other language families (section 2.4.).

2.3.1. Reconstruction of KOKO in Proto-Arawakan (Maipuran): before 3,000 BP?

The Proto-Arawakan homeland is currently posited between the Upper Amazon (Solimões) in Brazil and the Middle Orinoco in Venezuela (Heckenberger 2002: 103; see Map 1). The initial split of Proto-Arawakan is generally estimated to date back to before 3,000 BP (Heckenberger 2002: 106-9, quoting Noble 196512), while Kaufman (2007: 65) reports Swadesh’s dating of 4,500 BP. At the time of European contact, this family occupied a territory spanning from the Bahamas to Northern Argentina, that is to say a space extending far beyond the Amazonian Basin. According to Heckenberger (2002: 107), its maximal extent was reached by ca. 500 AD.


❖ Proto-Arawakan *kuhko U (MB), EF (Payne’s cognate sets numbered from 1 to 13 in boxed table 2).

A number of remarks have to be made regarding Payne’s reconstruction of *kuhko U, EF which is founded on a contrast between u and o, only pervasive in 3 languages belonging to 3 different branches of Arawakan (1991: 476): Terena (Southern Arawakan), Chamicuro (Chamicuro), and Guajiro (Caribbean Arawakan).

11 When two different time depths are mentioned between brackets, the first one is from glottochronology, the second one from ASJP. When only one date BP is mentioned, round numbers are from glottochronology, the others from ASJP.

12 Heckenberger’s acception of Arawakan matches Aikhenvald’s (1999a) classification and includes Southern Arawak, Paressi-Xingu, South-Western Arawak, Campa, Amuesha, Chamicuro, Rio Branco, Palikur, Caribbean, and Northern Amazonian. This grouping corresponds to Noble’s (1965) Arawakan subgroups less the Arawan and Uru-Chipaya branches. It also matches Payne’s (1991) Arawakan (Maipuran). According to Noble, Arawakan (Maipuran) is just one of the seven branches derived from Proto-Arawakan. It does not include Taino, Chamicuro, Amuesha and Apolista. Noble lexicostatistically estimated the split of Proto-Arawakan into its seven branches ca. 3300 BP (1965: 111), and that of its Arawakan (Maipuran) branch ca. 2500 BP (1965: 109-111).
Map 1. Approximate locations of koko across South America. Proto-forms (P-) appear in bold.
In order to explain the lack of contrast in the other branches, the author states (1991: 478-79) that: “It must have been the case that the contrast between the rounded vowels in Proto-Maipuran carried a low functional load and was already beginning to be eroded.” Besides, the presence of consonant -h- in the form reconstructed by Payne is based on the existence of h in the Chamicuro word kohka MB, which is certainly an exaggerated importance given to a single language, while Payne does not document regular loss of word-internal *h in any other Arawakan language. As a result, we propose that Chamicuro be considered irregular, a more economical solution than positing a Proto-Arawakan root *kukho and (implicitly) assuming that all languages but Chamicuro are irregular in having lost it. Another remark concerns the Campa languages where the kin term’s second vowel depends on the sex of the speaker (-o- is feminine in Arawakan, while -i- is masculine). This remark could perhaps also apply to the North Amazonian languages, whose MB terms of address regularly display a second vowel i. Other comments will be found in the appropriate column of table 2. We need to mention finally the rather unexpected replacement of the back vowel o or u by a high front vowel -i-, found in South Arawak (Muçóxone ni-kiko U, Bauré -kik MB, EF), which is explained by Payne (1991: 477, 479) by the regular correspondence - at least in Bauré - *u > i. All that being said, these remarks are unlikely to cast doubt on the validity of the Proto-Arawakan root *kuko U, EF, if one admits the contrast between u and o. In any case, the bulk of evidence presented in the following augmented cognate sets warrants beyond any reasonable doubt that, in accordance with what occurs in a great many regions of the Americas, there existed a root form for MB and EF in Proto-Arawakan, presenting a reduplicated phonetic shape *koko ~ *kuku, with the possibility of the final vowel i indicating a male speaking.

> Comparative data for MB terms in Arawakan including possessive forms

Southern Arawakan 1. Terena eiko MB, EF; Kinikinau euko (cq)U; Muçóxone nikiko U; 2. Ignaciano (Moxo) nekuka U; 3. Bauré -kik MB, EF; Southwestern Arawakan Piro nukoxiru MB, FZH, HF; 4. Piro koko ad. MB, FZH, EF; Apuriná wôkkuru U; Cuniba kuku MB, EF; Cushiñineri koko U; Kanamaré ghughu MB; Mashko kokoa U; Sirineiri kokoa U; Paresi-Xingu Yawalapiti kukujú (cq)MB; Mehínáku kuku (cq)MB; 5. Paresi koko-re U, EF; Saraveca koko-re-ki U;

---

13 Payne (1991: 478) attributes the final /u/ in Palikur kuku to some old unreliable source, and infers that it “should have been /o/.”
14 Alternatively, the Chamicuro form kohka MB, EF can be explained as a loan from Panoan koka MB, EF.
15 We did not retain a number of KOKO terms such as Baré nukáka MB, nukáka EF, and Mandawaka kaka MB, EF or even Island Carib nokoka-yem in our cognate set, in the absence of any supporting evidence for a phonetic correspondence *u/o > a. These terms could have been borrowed from other languages, although we have no direct evidence for this. Let us also mention that the loss of (initial) *k in Terena (and Kinikinau) is only supported by five examples (Payne 1991: 441). Besides, together with the root form, some cognates also display the 1sg possessive prefix, reconstructed by Matteson (1972: 164) as *n(V)-, and by Payne (1987: 62) as *mu-.
16 Languages are arrayed according to Aikhenvald’s classification (1999a: 67-71). Together with the root form, cognates also display the 1sg possessive prefix (see note 15), as well as the masculine relative suffix reconstructed as *-ri by Matteson (1972: 164) and Payne (1987: 63).
Table 2. Arawakan $kVkV$ forms (MB)

<table>
<thead>
<tr>
<th>Languages</th>
<th>$k$</th>
<th>$V$</th>
<th>$k$</th>
<th>$V$</th>
<th>Denotata</th>
<th>Linguistic comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Terena</td>
<td>e</td>
<td>u</td>
<td>k</td>
<td>o</td>
<td>MB, EF</td>
<td>Payne 1991: *u &gt; u, *o &gt; o</td>
</tr>
<tr>
<td>2 Moxo</td>
<td>n</td>
<td>e</td>
<td>k</td>
<td>u</td>
<td>k</td>
<td>U</td>
</tr>
<tr>
<td>Mucoxeone</td>
<td>n</td>
<td>e</td>
<td>k</td>
<td>i</td>
<td>k</td>
<td>a</td>
</tr>
<tr>
<td>3 Baure</td>
<td>-</td>
<td>k</td>
<td>i</td>
<td>k</td>
<td>MB, EF</td>
<td>Payne 1991: *u &gt; i</td>
</tr>
<tr>
<td>4 Piro</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>MB, EF, FZH</td>
<td>Payne 1991: *u &gt; o</td>
</tr>
<tr>
<td>Cuniba</td>
<td>k</td>
<td>u</td>
<td>k</td>
<td>u</td>
<td>MB, EF</td>
<td></td>
</tr>
<tr>
<td>Cushichineri</td>
<td>k</td>
<td>u</td>
<td>k</td>
<td>u</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>Kanamaré</td>
<td>g</td>
<td>h</td>
<td>u</td>
<td>g</td>
<td>h</td>
<td>u</td>
</tr>
<tr>
<td>Mashko, Sirineiri</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>a</td>
<td>U</td>
</tr>
<tr>
<td>Apurína</td>
<td>u</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>u</td>
<td>- U</td>
</tr>
<tr>
<td>Yawalapiti</td>
<td>k</td>
<td>u</td>
<td>k</td>
<td>u</td>
<td>-</td>
<td>MB</td>
</tr>
<tr>
<td>Mehinaku</td>
<td>k</td>
<td>u</td>
<td>k</td>
<td>u</td>
<td>MB</td>
<td></td>
</tr>
<tr>
<td>5 Paresi</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>-</td>
<td>U, EF</td>
</tr>
<tr>
<td>Saraveca</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>-</td>
<td>U</td>
</tr>
<tr>
<td>Ashaninka (Campa)</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>-</td>
<td>(♀)MB</td>
</tr>
<tr>
<td>Ashaninka (Campa)</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>i</td>
<td>(♂)MB</td>
<td>Final -i masculine marker?</td>
</tr>
<tr>
<td>Ashaninka (Tambo dialect)</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>(♂)MB</td>
<td>Final -o feminine marker?</td>
</tr>
<tr>
<td>Ashaninka (Tambo dialect)</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>i</td>
<td>(♂)MB</td>
<td>Final -i masculine marker?</td>
</tr>
<tr>
<td>Asheninka</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>(♀) MB</td>
<td>Final -o feminine marker?</td>
</tr>
<tr>
<td>Ashéninka</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>i</td>
<td>(♂) MB</td>
<td>Final -i masculine marker?</td>
</tr>
<tr>
<td>Matsiguenga</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>i</td>
<td>MB, EF</td>
<td>Final -i masculine marker?</td>
</tr>
<tr>
<td>Nomatsiguenga</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>i</td>
<td>(♂)MB, EF ad.</td>
<td>Final -i masculine marker?</td>
</tr>
<tr>
<td>7 Chamicuro</td>
<td>k</td>
<td>o</td>
<td>h</td>
<td>k</td>
<td>a</td>
<td>MB, EF</td>
</tr>
<tr>
<td>Manaó</td>
<td>g</td>
<td>h</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>MB</td>
</tr>
<tr>
<td>Wainuma</td>
<td>g</td>
<td>h</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>i</td>
</tr>
<tr>
<td>Waraiçu</td>
<td>g</td>
<td>h</td>
<td>u</td>
<td>k</td>
<td>MB</td>
<td></td>
</tr>
<tr>
<td>8 Palikur</td>
<td>k</td>
<td>u</td>
<td>k</td>
<td>u</td>
<td>MB, EF</td>
<td>Payne 1991: *u &gt; u, *o &gt; u ad.</td>
</tr>
<tr>
<td>Marawa</td>
<td>u</td>
<td>k</td>
<td>i</td>
<td>-</td>
<td>U</td>
<td>-i masculine marker?</td>
</tr>
<tr>
<td>9 Yucuna</td>
<td>o</td>
<td>k</td>
<td>u</td>
<td>MB, EF</td>
<td>Payne 1991: *u &gt; o,</td>
<td></td>
</tr>
<tr>
<td>10 Cabiyari</td>
<td>ā</td>
<td>k</td>
<td>u</td>
<td>MB, EF</td>
<td>*u &gt; u, or &lt; Proto-Arawakan *-aku FZ?</td>
<td></td>
</tr>
<tr>
<td>Resigaro</td>
<td>k</td>
<td>h</td>
<td>ā</td>
<td>g</td>
<td>i</td>
<td>MB ad.</td>
</tr>
<tr>
<td>Achagua</td>
<td>k</td>
<td>ū</td>
<td>w</td>
<td>i</td>
<td>U ad.</td>
<td>-i masculine marker? See kāu FZ</td>
</tr>
<tr>
<td>11 Piapoco</td>
<td>k</td>
<td>u</td>
<td>i</td>
<td>MB, EF</td>
<td>-i masculine marker? See kiuu FZ ad.</td>
<td></td>
</tr>
<tr>
<td>Curripaco</td>
<td>k</td>
<td>ā</td>
<td>k</td>
<td>i</td>
<td>MB, EF</td>
<td>-i masculine marker? See kuuku ad. FZ</td>
</tr>
<tr>
<td>Baniwa</td>
<td>n</td>
<td>ā</td>
<td>ū</td>
<td>k</td>
<td>u</td>
<td>MB</td>
</tr>
<tr>
<td>12 Tariana</td>
<td>n</td>
<td>ā</td>
<td>ū</td>
<td>k</td>
<td>h</td>
<td>i</td>
</tr>
</tbody>
</table>

94
<table>
<thead>
<tr>
<th>Language</th>
<th>Vowel</th>
<th>Consonant</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guarequena</td>
<td>$n$</td>
<td>$u$</td>
<td>MB</td>
</tr>
<tr>
<td>Guinau</td>
<td>$n$</td>
<td>$u$</td>
<td>EF</td>
</tr>
<tr>
<td>Yavitero</td>
<td>$n$</td>
<td>$ó$</td>
<td>MB</td>
</tr>
</tbody>
</table>

**Campa**

- Axininca: *koko*ini ad. ($\varnothing$)MB, HF, ($\varnothing$)FZH, *koNki* ad. ($\varnothing$)MB, MF, ($\varnothing$)FZH, no-*koNkiri* my WF, no-*koNki*oriri my distant U; Asháninka-*koNkiri* MB, EF, *koko* ($\varnothing$)MB, my ($\varnothing$)EF; Asháninka (Tambo dialect) *koko* my MB (ego $\varnothing$), HF *koki* my MB (ego $\varnothing$), WF; Ashéninka (Apurucayali) *nokoNkiri* my MB, my EF, my FZH; 6. Ashéninka (Apurucayali) koko-*ini* (-*ini* = masc. suffix) ($\varnothing$)U, HF, ($\varnothing$)FZH (ad.), *koNki* ~ *koki* ad. ($\varnothing$)U, WF, ($\varnothing$)FZH; Ashéninka Peréné *koki* ad. MB (ego $\varnothing$), WF, *koko* ad. MB (ego $\varnothing$), *nokoNkiri* MB, EF; Caquinte köNk-*ini* ‘tio’; Nomatsiguenga *nokoNgiri* my MB, my EF, my FZH, my ($\varnothing$)EF; Ashéninka (Tambo dialect) *koko* ($\varnothing$)MB (ego $\varnothing$), HF, koki ($\varnothing$)MB, my ($\varnothing$)EF; Ashéninka (Apurucayali) *nokoNkiri* my MB, my EF, my FZH, koki (ad.) ($\varnothing$)U, WF, ($\varnothing$)FZH; Ashéninka Peréné *koki* ad. MB (ego $\varnothing$), WF, *koko* ad. MB (ego $\varnothing$), *nokoNkiri* MB, EF; Amuesha *nego?* MB, EF, FZH;

**Chamicuro**

- Chamicuro: *ukojka* [*ukohka*] MB, EF; *kohka* MB, EF (ad.);

**Palikur**

- Palikur: *kuku* ad. MB, EF, *nukukrin* (my) MB, (my) EF; Marawa *uki* U;

**Caribbean Island Carib**

- Caribbean Island Carib: *nokska-yem* MB, FZH, MZH;

**Ta-Arawakan**

- Guajiro: *tausí* HF; Paraupano *táurt* EF;

**North Amazonian**


**Proto-Arawakan *aku(ro)* A (FZ), EM**

In the cognate sets below, Payne’s material is numbered from 1 to 5. As in the case of the previous cognate sets, we have eliminated consonant *h* from the reconstruction. The reconstruction of vowel *u* is founded on regular correspondences. The final ro is the feminine pronominal suffix, reconstructed in Proto-Arawakan (Maipuran) by Matteson (1972: 164) and Payne (1987: 63).

The cognate set *aku~ [*(n)aku(ru)]* my FZ/EM/(GM) in Arawak (Maipuran) (for details of sound correspondences see Payne 1991)

**Southern Arawakan**

1. Bauré- *aki* EM; Ignaciano *naka* A, GMZ;

**Campa**

- Asháninka: *áiro* - FZ, WM; Ashéninka *nayiro* my FZ, my EM; Ashéninka Peréné *aïronsí* ad. FZ (ego $\varnothing$), WM, *nayiro* EM; Nomatsiguenga *nágro* my FZ, my EM; Matsiguenga -*agiro* FZ, MBW, EM; Nanti *pagiro* ad. EM, *obagiro* her HM;

2. Chamicuro: *ajka* [*ahka*] my A (ego $\varnothing$);

---

17 The Campa languages (Asháninka, Ashéninka, Nomatsiguenga, etc.) display forms with an additional internal nasal archiphoneme (traditionally noted $N$) whose phonetic realization varies according to context. This feature belongs to the phonemic inventory of the Campa languages: it occurs after a vowel and it “homorganically assimilates to a following obstruent” (Payne 1981: 62, repeated in Mihas 2010: 62).
There is a near consensus on recognizing the genetic relatedness of the Panoan and Takanan language families. Proto-Pano-Takanan has been dated back to ca. 4,700 BP by Swadesh (Kaufman 2007: 70).

The Panoan languages are now spread across the Peruvian lowlands, notably on the Ucayali River basin, the lower Urubamba River, the Upper Purús in the western part of the state of Acre (Brazil), the southwestern part of the state of Amazonas (Brazil), and finally in northern Bolivia. Our knowledge of this language group’s prehistory relies on linguistics and archaeology. Using glottochronology, d’Ans (1973: 364) stated that, by AD 100 (±300 years), the Panoan language family which, according to Lathrap, originated in northern Bolivia, had just differentiated into three subgroups: Proto-Preandino (Cashibo), Proto-Ucayali-Cabeceras (Proto-Ucayali-Yaminahua), and Proto-Beniano (Proto-Chacobo). Myers (1974: 135) and other archaeologists assume that the arrival of Panoan speakers on the Ucayali corresponds to the appearance of ceramics of the Pacacocha tradition in the archeological record, ca. AD 300. Chacobo, Kaxariri, Pacahuara groups still live in Bolivia close to Takanan-speaking groups (Erikson 1993: map p. 46).

As far as historical documentation can trace them back (500 years or so), Takanan language groups, among which the Araona, Eparamona, Uchupiamona are known for having been established within the northern part of the angle composed by the Madre de Dios-Manu Rivers and the Beni River. This region is posited as the Proto-Takanan homeland (Wichmann et al. n.d.: 20).

Key (1968: 73) listed a few Takanan and Panoan KOKO terms in her Takanan cognate set glossed ‘uncle’, but did not formally reconstruct a proto-form. Girard (1971: 90) also established a cognate set and reconstructed *ku- or *kuku MB in Proto-Takanan. He agreed with Shell’s (1965: 144, items 190 and 192) Proto-Panoan reconstructions *koka MB and *koko ‘nephew’, and proposed *kuku ~ *kuka MB in Proto-Pano-Takanan (1971: 165). Table 3 below gives fairly complete cognate sets in Panoan and Takanan languages. Panoan languages are presented according to Amarante Ribeiro’s (2006) classification.

18 Speaking of her reconstructions, Shell (1965-1975: 11) used the term ‘reconstructed Panoan’ (‘Pano reconstruído’) instead of ‘Proto-Panoan’, an expression that she wanted to keep for future, more advanced reconstructive works.
Proto-Takanan *kuku

Key (1968: 35, 37; 1992: 98, 100) and Girard (1971: 23-25) disagree on which consonant phoneme is to be reconstructed in Proto-Takanan: *k for Girard, *x for Key. Girard convincingly argues for *k > h, x or k in Takanan languages, pointing to its relationship to Proto-Panoan *k, and the evidence of Cavinena k (Girard 1971: 25). As noted in table 3, Ese Ejja and Huarayo display irregular reflexes, respectively tsotso and toto MB.19

Table 3. Pano-Takanan cognate sets

<table>
<thead>
<tr>
<th>Proto-Pano-Takanan</th>
<th>Proto-Panoan</th>
<th>Proto-Panoan</th>
<th>Proto-Panoan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k u k u MB, EF</td>
<td>k o k a MB, EF, FZH</td>
<td>k o k o 'nephew'</td>
</tr>
<tr>
<td>Amahuaca</td>
<td>k o k á MB, (?)F</td>
<td>Group I</td>
<td></td>
</tr>
<tr>
<td>Capanahua</td>
<td>k o k a U (MB), EF</td>
<td>Group II</td>
<td></td>
</tr>
<tr>
<td>Cashibo</td>
<td>k u k u MB, EF, FZH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huaripano (Panobo)</td>
<td>k o k a U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pano (Navarro)</td>
<td>k u k u MB, HF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shipibo-Cunibo</td>
<td>k o k a MB, WF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shipibo-Cunibo</td>
<td>k o k o- (?)(?)BS</td>
<td>Group III</td>
<td></td>
</tr>
<tr>
<td>Cashinahua</td>
<td>k u k a MB, FZH, EF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catuquina</td>
<td>k o k a MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isconahua</td>
<td>k o k a MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaxariri</td>
<td>k u k u U, EF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marubo</td>
<td>k o k a (?)(?)MeB, (?)(?)cZS, (?)(?)MB, (?)(?)ZS</td>
<td>Group III</td>
<td></td>
</tr>
<tr>
<td>Marunahua (Shell)</td>
<td>k ó k á U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poyanahua</td>
<td>k u k a Ad. U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shanenahua</td>
<td>k u k a U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharanahua</td>
<td>k o k a MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yaminahua</td>
<td>k o k a MB, EF, FZH, &amp; ad. id. + (?)(?)ZS, (?)(?)S</td>
<td>Group IV</td>
<td></td>
</tr>
<tr>
<td>Yawanahua</td>
<td>k u k a U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chacobo</td>
<td>k o k o U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korubo</td>
<td>k o k o MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matis</td>
<td>k u k u MyB, FZH, EF, yZS.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayoruna Tabatinga (Spix: 1820)</td>
<td>k u k u U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May. (Matses: Erikson)</td>
<td>k u k u MyB, EF, yZS.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May. (Matses: Fields)</td>
<td>k o k a MB, EF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May. (Matses Fields)</td>
<td>k a k o (?)(?)BS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paehuara (Armentia)</td>
<td>k u k o U, EF, 'sobrino'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paehuara (d'Orbigny)</td>
<td>k o k o U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Araona</td>
<td>h u h u U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cavinėna</td>
<td>k o k o MB (ad.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ese Ejja (ts irregular)</td>
<td>ts o ts o MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huarayo (t irregular)</td>
<td>t o t o MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reyesano</td>
<td>h o h o U</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tacana</td>
<td>h u h u MB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proto-Panoan *koka and *koko

Shell (1965-1975: 144, items 190 and 192), followed by Girard (1971: 165), reconstructs two roots *koka and *koko in Panoan on the basis of regular sound correspondences.

Girard (1971: 165) glossed *koka 'maternal uncle' upon the overwhelming evidence of the reflexes referring to this relationship in groups I, II and III.20 The same (1971: 165) glossed *koko 'nephew' upon the evidence presented by Shell (2008: 144), namely Pano (Navarro) kuku 'sobrina' [niece], Shipibo-Cunibo koko (?)ZS, 'sobrino,' 'hijo de la hermana de una mujer' [nephew], Cashibo koko EF, Chacobo koko 'tio' [uncle], Mayoruna kuku 'tio [uncle], suegro [EF], sobrino

19 In Ese Ejja, a few words display an apparent evolution *k > x, e.g. *kunu > Takana hunu 'liana,' Cavinėna kono 'liana,' Ese Ejja xono 'liana,' *kuř > Takana huri 'ocelot,' Cavinėna ho-kori 'badger,' Ese Ejja xoxi 'ocelot,' etc. (Girard 1971: 90).

20 Koka MB was also elicited in group IV (Matses) by Fields in 1970 (Erikson 1986: 192, 201), but has never been recorded since.

97
[nephew] (details in table 3), Pacahuara koko ‘tío, suegro, sobrino’. To these must been added Matis kuku which refers to MB, EF and yZS. But there are also a few koka forms meaning both ‘nephew’ and MB, EF, such as Marubo koka MeB, (🕶)eZS, (🕶)ZS, Yaminahua koko ref. MB, EF, FZH & ad. MB, EF, FZH, (🕶)ZS, (🕶)S. As a matter of fact, the opposition between koka MB, EF on the one hand, and kuku ‘nephew’ on the other hand is not so clear-cut, although the presence of reflexes of both forms in all the groups including more than one language indicates, without much possible doubt, that they have their origin in Proto-Panoan. It is important to note that all the languages of group IV, except Matses, display kuku ~ koko forms while most languages from the other groups display koka forms.

We must also contemplate the hypothesis that kuku ~ koko forms were borrowed from other language families, maybe from Arawakan. Table 4 recapitulates the various KOKO terms found in Arawakan (vocative) and Panoan, suggesting possible mutual borrowings. Southwestern Arawak languages, including Piro, Cuniba, Kanamaré, Cushichineri21 appear as the best candidates in this respect. The problem is that the northern Panoan languages of group IV (Mayoruna, Korubo, Matis and Matses)22, all having koko or kuku forms, are separated from the Southwestern Arawak languages by the main body of Panoan languages, all having koka forms. This is not the case for the Bolivian languages of group IV (Chacobo and Pacahuara) which are neighbors with the Southwestern Arawak language groups displaying kuku forms. These Bolivian Panoans are also neighbors with the Takanan groups which, as we know, also display huhu or kuku forms. The similarity of KOKO forms in all these languages could have resulted from ancient areal diffusion.

Table 4: Comparison of Arawakan and Panoan KOKO forms

<table>
<thead>
<tr>
<th>Panoan koko ~ koka forms</th>
<th>Arawakan koko address forms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panoan kuku ~ koko forms</strong></td>
<td><strong>Southwestern Arawak kuku ~ koko forms</strong></td>
</tr>
<tr>
<td>Cashibo</td>
<td>koko MB, EF</td>
</tr>
<tr>
<td>Mayoruna (Matses)</td>
<td>koko MB, EF</td>
</tr>
<tr>
<td>Matis</td>
<td>kuku MB, EF</td>
</tr>
<tr>
<td>Chaco</td>
<td>koko U</td>
</tr>
<tr>
<td><strong>Panoan koka ~ kuka forms</strong></td>
<td><strong>Southwestern Arawak koka forms</strong></td>
</tr>
<tr>
<td>Amahuaca</td>
<td>kuka MB</td>
</tr>
<tr>
<td>Shipibo-Cunibo</td>
<td>koka MB, EF</td>
</tr>
<tr>
<td>Capanahua</td>
<td>koka MB, EF</td>
</tr>
<tr>
<td>Panobo</td>
<td>koka U</td>
</tr>
<tr>
<td>Isconahua</td>
<td>koka MB</td>
</tr>
<tr>
<td>Cashinahua</td>
<td>kuka MB, FZH, WF</td>
</tr>
<tr>
<td>Catuquina</td>
<td>koka MB</td>
</tr>
<tr>
<td>Marubo</td>
<td>koka MB</td>
</tr>
<tr>
<td>Sharanahua</td>
<td>koka MB</td>
</tr>
<tr>
<td>Shanenahua</td>
<td>kuka U</td>
</tr>
<tr>
<td>Yawanahua</td>
<td>kuka U</td>
</tr>
<tr>
<td>Yaminahua</td>
<td>koka MB, EF</td>
</tr>
<tr>
<td>Poyanawa</td>
<td>kuka U</td>
</tr>
<tr>
<td>Marunahua</td>
<td>koka U</td>
</tr>
<tr>
<td><strong>Southwestern Arawak koko forms</strong></td>
<td><strong>Other Southwestern Arawak forms</strong></td>
</tr>
<tr>
<td>Piro</td>
<td>koko MB, IIF, WFB</td>
</tr>
<tr>
<td>Cushichineri</td>
<td>koko U</td>
</tr>
<tr>
<td>Cuniba</td>
<td>koko MB, EF</td>
</tr>
<tr>
<td>Kanamaré</td>
<td>ghuglu MB</td>
</tr>
<tr>
<td><strong>Campas forms</strong></td>
<td><strong>Upper Rio Negro koko forms</strong></td>
</tr>
<tr>
<td>Mashko</td>
<td>koka U</td>
</tr>
<tr>
<td>Sirineiri</td>
<td>koka U</td>
</tr>
<tr>
<td>Chamicuro</td>
<td>koka MB, EF &lt; Panoan?</td>
</tr>
<tr>
<td>Ashaninka</td>
<td>koki my (🕶) MB, EF</td>
</tr>
<tr>
<td>Ashéninka</td>
<td>koko-ini ad. (flatMap) U, (flatMap) EF</td>
</tr>
<tr>
<td>Matsiguenga</td>
<td>koki MB, EF</td>
</tr>
<tr>
<td>Nomatsiguenga</td>
<td>koki ad. (flatMap) MB, EF</td>
</tr>
<tr>
<td><strong>Resigaro</strong></td>
<td><strong>Piapoco</strong></td>
</tr>
<tr>
<td>grayscale</td>
<td>-khiigi MB</td>
</tr>
<tr>
<td>Achaguine</td>
<td>kúiti ~ kúwi MB</td>
</tr>
<tr>
<td>Piapoco</td>
<td>kui MB, EF</td>
</tr>
</tbody>
</table>

21 Spoken or formerly spoken in the north of the Madre de Dios River, and along the Purús and its tributaries (southeastern Peru and Brazil).
22 Spoken along the Javari River and its tributaries (along the border between northeastern Peru and Brazil).
Girard (1971: 165) reconstructed *kuku/a MB in Proto-Pano-Takanan on the basis of regular sound correspondences between Proto-Panoan and Proto-Takanan (1971: 158-9), leaving open the nature of the second vowel (u or a).

2.3.3. Reconstruction of KOKO in Proto-Arawan: 1,764 BP (ASJP)

The Arawan language family is composed of 5 languages: Sorowahá, Madi, Paumari, spoken in the state of Amazonas (Brazil) along the Purús River and its tributaries and the Jurúá River, Kulina-Deni, spoken along the upper Purús and the Jurúá River, and finally Arawá which is extinct and, for the most part, undocumented. The four living languages and their dialects display KOKO forms, referring to MB and EF (table 5). Note that de Créqui-Montfort and Rivet (1922: 175) mentioned the existence of Paumari kuku EM and Jamamadi koko S. We have not been able to confirm or infirm the existence of these meanings in more recent data.

Dixon (2004b: 48, 61) reconstructs *koko MB, EF in Proto-Arawan (*k > k in all languages and all environments, and *o > o in all languages and all environments), but considers this form as an Amazonian areal feature, and concludes, as we already mentioned (2004b: 17), that it "could have been borrowed from anyone of a number of sources in Proto-Arawan, or else into individual languages." He also claims that Arawan has no genetic link with Arawakan (Maipuran) whatsoever. A likely location for Proto-Arawan could be between the lower Purús and the Jurúá River, where the vast majority of Arawan languages (including Arawan until its extinction) are spoken. No date is suggested for Proto-Arawan, except by ASJP: 1764 BP.

2.3.4. Reconstruction of KOKO in Proto-Cariban: ca. 3,700 BP

Proto-Cariban is estimated back to ca. 3,700 BP (Kaufman 2007: 75); its homeland is posited in Venezuelan Guiana (Villalón 1991; Heckenberger 2002: 103). De Goeje (1909: 30) first published a list of Cariban KOKO forms, on the basis of which he later proposed *koko U (1946: #
The meaning of the reconstructed masculine form surely includes MB, but also possibly EF and GF. This list has now been considerably augmented, and one can remark that there is a basis for postulating also a feminine proto-form kuku ~ koko GM, FZ, EM, reconstructed as *kuku GM in Proto-Taranoan by Meira (1998: 172, #118), on the basis of *k > k in all languages and all environments, and *u > u in all languages and all environments.

The date of the Proto-Taranoan split calculated by Meira (1998: 159-160) is between 500 and 900 years ago. Our “inspectional reconstruction” *kuku ~ koko GM in Cariban is founded on the occurrence of koko or kuku GM in most of the branches composing the Cariban family, as defined in Gildea’s classification (2012: 445), i.e., Venezuelan, Nahukwa, Guianan (*kuku in Proto-Taranoan), as well as in some unclassified languages (see the comparative data below). Similar occurrences of GM forms are also massively found in Chibchan languages, in Mapudungu and in North and Meso-American language families (see Appendix and tables 11, 12). KOKO terms in Cariban are address terms. Reference terms referring to the same kin types can be reconstructed (notes 23 and 24).

Cariban comparative data. Classification of languages adapted from Gildea (2012).

*koko MB, (GF, EF), *kuku GM, (FZ, EM);
Parukotoan Branch: Hixkaryana (Parucoto) owhoko U, EF;
Parukotoan Branch: Hixkaryana (Parucoto) owhoko U, EF;
Parukotoan Branch: Hixkaryana (Parucoto) owhoko U, EF;

Pekodian Branch: Bakairi kogo MB, FZH; Arara (Pará) koko ~ koko MB, FB, FBS, FZS;
Pekodian Branch: Bakairi kogo MB, FZH; Arara (Pará) koko ~ koko MB, FB, FBS, FZS;
Pekodian Branch: Bakairi kogo MB, FZH; Arara (Pará) koko ~ koko MB, FB, FBS, FZS;
Pekodian Branch: Bakairi kogo MB, FZH; Arara (Pará) koko ~ koko MB, FB, FBS, FZS;

Venezuelan Branch: Makushi koko, okoko GM, koko MB; Pemon ko'wai ad. GM; Taurepang (Pemón dial.) ko'way ~ koko (my) GM; Arekuna (Pemón dial.) koki GM; Akawao 'a?ai M; Panarc koko FF, (c)FBS, (c)MJD, (c)eB, kokon yB, young C; Tamanaco a koko form?;
Venezuelan Branch: Makushi koko, okoko GM, koko MB; Pemon ko'wai ad. GM; Taurepang (Pemón dial.) ko'way ~ koko (my) GM; Arekuna (Pemón dial.) koki GM; Akawao 'a?ai M; Panarc koko FF, (c)FBS, (c)MJD, (c)eB, kokon yB, young C; Tamanaco a koko form?;
Venezuelan Branch: Makushi koko, okoko GM, koko MB; Pemon ko'wai ad. GM; Taurepang (Pemón dial.) ko'way ~ koko (my) GM; Arekuna (Pemón dial.) koki GM; Akawao 'a?ai M; Panarc koko FF, (c)FBS, (c)MJD, (c)eB, kokon yB, young C; Tamanaco a koko form?;
Venezuelan Branch: Makushi koko, okoko GM, koko MB; Pemon ko'wai ad. GM; Taurepang (Pemón dial.) ko'way ~ koko (my) GM; Arekuna (Pemón dial.) koki GM; Akawao 'a?ai M; Panarc koko FF, (c)FBS, (c)MJD, (c)eB, kokon yB, young C; Tamanaco a koko form?;

Guianan Branch: Carib (Karinja, Galibi) molekoko boy, kah-tobo MB, WF, (c)FZH, MBS, FZS; Carib (Cachama) kax-topo GF, MB, EF; Carib (Oiapoque) kah-tobó MB, MBS, (c)FZH, MBS, S, SS, etc.; Ye'kwana (Makiritare) kóókó ~ koko ad. GF, FFZS etc.; Wayana (Roucouyenne) ku-ni GM, old woman, kőnko FF, MB, EF, (c)HB; akon eB, kono WB, HZ; Tarananoan group: *kuku GM, (FZ, EM); Pianacoto ku-ni GM; Akuríyo (Triometesem) kúkú-ni-komo GM; Trió kuku ad. & ref. (c)GM, (c)FZ, WM, (c)MBW, (c)GM, (c)FZ, HM, (c)MBW, konóka ref. (c)FZH, WB, ZH, etc., koko ref. & ad. (c)FZD (rarely), (c)MBD, (c)BW, HZ; Caríjona kuku GM, FZ; Hianáculo-Umaúá kúuku GM;
Guianan Branch: Carib (Karinja, Galibi) molekoko boy, kah-tobo MB, WF, (c)FZH, MBS, FZS; Carib (Cachama) kax-topo GF, MB, EF; Carib (Oiapoque) kah-tobó MB, MBS, (c)FZH, MBS, S, SS, etc.; Ye'kwana (Makiritare) kóókó ~ koko ad. GF, FFZS etc.; Wayana (Roucouyenne) ku-ni GM, old woman, kőnko FF, MB, EF, (c)HB; akon eB, kono WB, HZ; Tarananoan group: *kuku GM, (FZ, EM); Pianacoto ku-ni GM; Akuríyo (Triometesem) kúkú-ni-komo GM; Trió kuku ad. & ref. (c)GM, (c)FZ, WM, (c)MBW, (c)GM, (c)FZ, HM, (c)MBW, konóka ref. (c)FZH, WB, ZH, etc., koko ref. & ad. (c)FZD (rarely), (c)MBD, (c)BW, HZ; Caríjona kuku GM, FZ; Hianáculo-Umaúá kúuku GM;

Residue: Apalai e-o-o ‘my MB’; Yukpa group: Japériá yuvan-koko A; Unclassified: Sapara kun-nutú GM; Apiaka koko U; Pimenteira kuekú Oheim (MB), boy; Paravilhana gocko MB, taany gocko GF.

23 Another Cariban (reference) term refers to MB and EF, probably from Proto-Cariban *j-awo- (de Goeje 1946: 58).
24 Another Cariban term refers to GM and FZ, probably from Proto-Cariban *noti (de Goeje 1946: 57).
25 The symbol * is the convention that Blench adopted from the Niger-Congo volume edited by Bendor-Samuel (1989), to distinguish “reconstructions established by regular sound correspondences [marked with an asterisk *]. AM & PB] from those derived by quick inspection of cognates … The effect of this is to translate the starred forms of various writers to hache” (Blench 2008: 204).
2.3.5. **Reconstruction of KOKO in Proto-Guahiboan: ca. 2,300 BP**

Kaufman (2007: 65) gives 23 centuries for Proto-Guahiboan, whose homeland is posited by Wichmann *et al.* (n.d.: 19) on the northeastern side of the Colombian Andes. Christian & Matteson’s article on Proto-Guahiboan (1972: 150-51) is rather scanty and based on a comparison between 3 languages: Guahibo, Cuiva and Guayabero. Unfortunately, we have not been able to access Keels’ more recent study (1986) on Proto-Guahiboan, which is focused on the same languages. Correspondences – existing in almost all environments – given in table 6 are from the first authors. Correspondences in Macaguán are deducted from available comparative material extracted from Buenaventura (1993). We do not have the term for MB in Guayabero but the sound correspondences in this language, deducted from Keels (1985), are probably *x > x, *k > k, *u > u. Thus we just venture a reconstruction of *axu-jo MB, EF, FZH and *akwe FM, MM in Proto-Guahiboan. Note that -jo is a diminutive suffix.

<table>
<thead>
<tr>
<th>Table 6. Guahiboan comparative data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proto-Guahiboan</strong></td>
</tr>
<tr>
<td>Cuiva (Hiwi) *a &gt; a, *x &gt; x, *u &gt; u, *j &gt; j, *o &gt; o, *k &gt; k, *w &gt; w, <em>e &gt; e</em></td>
</tr>
<tr>
<td>Sikuani *a &gt; a, *x &gt; x, *u &gt; u, *j &gt; j, *o &gt; o, *k &gt; k, *w &gt; w, <em>e &gt; e</em></td>
</tr>
<tr>
<td>Guahibo (Rivet 1948)</td>
</tr>
<tr>
<td>Guahibo (Rivet 1948)</td>
</tr>
<tr>
<td>Macaguán *x &gt; k, *k &lt; k, *a &gt; a, *u &gt; u, <em>e &gt; e</em></td>
</tr>
</tbody>
</table>

2.3.6. **Reconstruction of KOKO in Proto-Chibchan: ca. 6,700 BP**

Constenla Umaña (2009: 209) gives 6,550 BP (6682 BP in Constenla Umaña 2012: 419) for the split between Pech (Payá) and Core Chibchan, and suggests a Proto-Chibchan homeland in Costa Rica and Panama (Constenla Umaña 2009: 209). This date matches the 7,000 years obtained through genetic studies of populations from this language group (Hoopes & Fonseca 2003: 61). Constenla Umaña (1981: 381, 399-400) reconstructed the following items in Proto-Chibchan:

1. *káka F* (cognate set numbered 1 to 3 in table 7), with *k having sound correspondences in Cabécar k (all positions), Teraba kʰ (word initial position), k elsewhere, Dorasque g (intervocalic position), k elsewhere, and *a having sound correspondences in Cabécar a (all environments), Terraba o (all environments except in contact with laryngeal or nasalized), Dorasque a (all environments).
2. \(^*\text{gAkA}\) EM\(^{26}\) (cognate set numbered 4 to 7 in table 7), based on sound correspondences: \(^*\text{g}\) > Cabécar \(h\) before \(i\) and \(u, j\) elsewhere, Bribri \(\emptyset\) before \(i\) and \(u, j\) elsewhere, Cuna \(s\) in all positions, Cagaba \(\emptyset\) before \(i\) (one etymology), \(g\) elsewhere; \(^*\text{k}\) > Cabécar \(k\) in all positions, Bribri \(k\) in all positions, Cuna \(k\) in all positions, Cágaba \(g\) in intervocalic position; \(^*\text{a}\) > Cabécar \(a\) in all positions, Bribri \(a\) the same, Cuna \(a\) the same, Cágaba \(a\) the same. As for the final vowel -\(i\) of Cágaba \(\text{gagi}\) EM, it is explained by Constenla Umaña (1981: 362) as a nominal suffix very frequent in this language.

Table 7. Chibchan comparative data (from Constenla Umaña 1981 and additional material)

<table>
<thead>
<tr>
<th>Proto-Chibchan</th>
<th>* k á k a</th>
<th>F</th>
<th>* g a k a</th>
<th>EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pech</td>
<td></td>
<td></td>
<td>u ' a</td>
<td>FZ, EM</td>
</tr>
<tr>
<td>Pech (Paya)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core Chibchan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isthmic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cabécar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cabécar (Estrella)</td>
<td>1</td>
<td>k a k a</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Cabécar (Ciripo)</td>
<td>1</td>
<td>k a k a</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Cabécar (Tukurrike)</td>
<td>1</td>
<td>k a</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Bribri</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teribe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teraba</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dorasque (Chángena)</td>
<td>3</td>
<td>k a g a</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Cuna</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magdalenic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cogué (Cágaba)</td>
<td>7</td>
<td>g a g i</td>
<td>EM</td>
<td></td>
</tr>
<tr>
<td>Guanama (Malayo. Damana)</td>
<td></td>
<td>gw a g i</td>
<td>EM</td>
<td></td>
</tr>
<tr>
<td>Bintukua (Ica)</td>
<td></td>
<td></td>
<td>gw a t i</td>
<td>A</td>
</tr>
<tr>
<td>Atanques (Cancuama)</td>
<td></td>
<td></td>
<td>gw a s i</td>
<td>A</td>
</tr>
<tr>
<td>Central Tunebo (U’wa)</td>
<td></td>
<td></td>
<td>k a g i</td>
<td>FZD, MZD</td>
</tr>
<tr>
<td>Central Tunebo (U’wa)</td>
<td></td>
<td></td>
<td>g* a k a</td>
<td>(2)EF, (2)DH</td>
</tr>
<tr>
<td>Not classified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nutilbé</td>
<td></td>
<td></td>
<td>gu a k u</td>
<td>old woman</td>
</tr>
</tbody>
</table>

The additional list below presents more data than what was used by Constenla Umaña in 1981.

Additional Chibchan comparative data:

a) Proto-Chibchan \(^*\text{KVv}V\) MB, EF, GF:

\(^{26}\) Constenla Umaña (1981: 399) reconstructed \(^*\text{gAkA}\) EM, with \(A\) indicating that the proto-form had either \(^*\text{a}\) or \(^*\text{a}\). The contrast between these vowels is no longer recognized by the author (2012: 404-5). Consequently, we wrote \(^*\text{gaka}\) in table 6.
Pech Pech (Paya) uku MB, EF; oká B;

Core Chibchan

Isthmic Cabécar kēgōlō [kugolo] MB; Cabécar kikiwa ‘anciano(a);’ Cabécar (Estrella) keke ‘señor’; Cabécar (Chirripo) keke ‘señor'; Bribri akēgōla [akigula] ‘anciano’; Boruca kakaat B, sugu ‘anciano’; Teribe k'oki GF, k'ēgē MB, k'ēgē EF; Teraba kega U, kegi EF; Dorasque (Chumulu) wōka GF; Dorasque (Gualaca) ōka GF; Magdalenic Atanques (Cancuama) kuku U, sukwi GM; Bintukua (Ica) tegue U; Damana (Malayo) -kugu U, GF; Cogui (Cágaba) -kukwi A;

b) Proto-Magdalenic: ~5,200 BP (Constenla Umaña 2012) kaka GM; Bintukua (Ica) -zaga GM; Damana (Malayo) -taka GM; Cogui (Cágaba) kaka GM; Muisca (Chibcha) kaka GM; Tunebo (U’wa) kaka GM; Muisca (Chibcha) kaka GM; Tunebo Sinsiaga of Cobugón kaka GM.

2.3.7. Reconstruction of KOKO in Proto-Lenca-Misumalpan: ca. 7,200 BP

The Lenca family is composed of 2 languages. One of them is spoken in Honduras, the other one in El Salvador. The Misumalpan language family includes 4 languages: Miskito (Nicaragua), Matagalpa (Nicaragua), Cacaopera (El Salvador) and Sumo (numerous dialects in Nicaragua). A number of authors have postulated a genetic relationship between these 2 families, notably Constenla Umaña (2002: 189). In an attempt to establish this idea, Constenla Umaña established a list of 92 joint Lenca-Misumalpan sets (2002: 191-193), among which one finds #7: ‘anciano(a),’ thus Cacaopera kikus, Sumo kikuj, Miskito kuka, Lenca-Salvador koko. The author’s list can be extended, adding similar words from other Misumalpan and Lencan languages or dialects (table 8). Constenla Umaña (2002: 193-196) also observed a significant number of phonetic correspondences between Misumalpan languages and Lenca languages, notably Misumalpan *u = Lenca *o, or Misumalpan *k = Lenca *k. Koontz-Garboden and Francez (2009: 9) reconstructed the 1sg possessive suffix *-ki in Proto-Misumalpan.

Constenla Umaña’s glottochronological study, developed in the same article (2002: 197-202), and based on 120 Lenca-Misumalpan sets, furnishes dates for several language splits. The Lenca-Misumalpan split apparently happened ca. 7,200 BP. Within Misumalpan, the split between Proto-Miskito and the other Misumalpan languages occurred ca. 5,800 BP. The split between Matagalpan (Matagalpa and Cacaopera) and Sumo-Ulwa languages occurred ca. 5,300 BP. Other conclusions concerning further subdivisions were also reached, notably the date of the split of the Lenca dialects ca. 2,300 BP.

Table 8. Lenca-Misumalpan comparative data

<table>
<thead>
<tr>
<th>Proto-Lenca-Misumalpan</th>
<th>*k</th>
<th>*V</th>
<th>*k</th>
<th>*V</th>
<th>GF, U, EF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenca</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Guajiquiro - Honduras)</td>
<td>k</td>
<td>o</td>
<td>g</td>
<td>o</td>
<td>EF</td>
</tr>
<tr>
<td>(Chilanga - El Salvador)</td>
<td>k</td>
<td>o</td>
<td>k</td>
<td>o</td>
<td>‘anciano’</td>
</tr>
<tr>
<td>(Chilanga - El Salvador)</td>
<td>k</td>
<td>o</td>
<td>h</td>
<td></td>
<td>U, eB</td>
</tr>
<tr>
<td>Misumalpan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miskito</td>
<td>k</td>
<td>ū</td>
<td>k</td>
<td>a</td>
<td>GM</td>
</tr>
<tr>
<td>Miskito</td>
<td>k</td>
<td>ū</td>
<td>k</td>
<td>i</td>
<td>(my) GM</td>
</tr>
<tr>
<td>Matagalpa</td>
<td>k</td>
<td>u</td>
<td>k</td>
<td>e</td>
<td>(my) U</td>
</tr>
<tr>
<td>Cacaopera</td>
<td>k</td>
<td>ū</td>
<td>i</td>
<td></td>
<td>GF</td>
</tr>
</tbody>
</table>

2.4. Other KOKO cognate sets in South American linguistic families

2.4.1 KOKO in Quechuan and Aymaran: at least 1,500 BP

Quechuan and Aymaran are the two South American indigenous language families most widely spoken today in the middle Andean area. Quechuan “dialects” or “languages” are spoken by some 8 million people along the Andean cordilleras, from Southern Colombia to Northeastern Argentina. The Quechuan language family is divided into 2 branches: Quechua I (or B), henceforth Q I, comprising dialects occupying a continuous area in the central Peruvian Highlands, and Quechua II (or A), henceforth Q II, comprising the remaining Quechuan varieties, situated both to the north and the south of the central Peruvian zone, from Ecuador to Bolivia and Argentina (Adelaar and Muysken 2004: 185-186, Adelaar 2012: 578-580). The Aymaran family also comprises 2 branches: a northern branch with 2 languages, Jaqaru and Kawki, spoken by a few thousand people (Kawki by just a few) in the Peruvian province of Yauyos (southeast of Lima), and a southern branch, with more than 2 million speakers, straddling Bolivia, Chile and Peru (Adelaar 2012: 577-8). The distribution of these two language families results from their pre-Colombian, but also from their colonial expansions. Soon after the Inca Empire was defeated, Quechua and Aymara were promoted, along with Puquina, as “general languages” for colonial administrative and religious purposes (Adelaar and Muysken 2004: 167). So they continued being spoken at the expense of other indigenous Andean languages, many of which were still spoken at the time of European contact (Adelaar and Muyskens 2004: map 3, p. 166).

A consensus has been reached pointing to Central Peru as the homeland of both Proto-Aymaran and Proto-Quechuan, and to date their respective initial linguistic divergence before 500-600 AD (Adelaar and Muysken 2004: 181, Heggarty 2008: 39, 46-47 and 52). The Aymaran language seems to have been the first language to expand, southwards towards Cuzco and Southern Peru, and possibly northwards towards Ancash (Heggarty 2008: 39-41, 48-49), covering a region roughly corresponding to Huari’s territorial influence. The development of Aymaran on the Altiplano seems to have happened much later, probably after 1000 AD (Heggarty 2008: 50). Quechua’s first linguistic divergence was apparently limited to a region from the north of Ancash to Huancayo in the south (Heggarty 2008: 50 and fig. 6). This is approximatively where the Central Quechuan (Q I) dialects are now spoken. The general expansion of Quechuan languages, towards Ecuador and the Cuzco region, was initiated long after this first linguistic spread, but “still a few centuries before the Inca conquest” (Heggarty 2008: 50). It is debated whether the Late

---

Adelaar (2012: 588) says that “More recently a different scenario has emerged as the expansion of Quechuan was attributed to the centralized state of Huari.” According to this scenario, the region of Ayacucho, where the capital of Huari was situated, could have been either the homeland of Quechuan as a whole, or of Quechua II alone (Adelaar 2012: 588, references included).
Intermediate Period (1000-1476) corresponds to this language expansion phase, or if it is the Middle Horizon (600-1000: Huari culture), that was primarily equated with Aymaran. During this expansion phase “it is unclear how significant a population movement was involved” (Heggarty 2008: 51), but it is generally admitted that, notably in the northern region, “this form of Quechua spread principally as a lingua franca for trade” (Heggarty 2008: 51). Apparently, the Incas were responsible for taking Quechua further south into Southern Bolivia, leap-frogging the Aymara-speaking regions of the Altiplano (Heggarty 2008: 52).

Quechuan and Aymaran have much in common linguistically and their genetic relationship has been regularly debated. Their relationship is a matter of millennia, but it is now questioned if these two families have any genetic relationship at all. Most linguistic convergences have been attributed to heavy language contact (Heggarty 2008: 43-46).

A number of Quechuan and Aymaran languages refer or referred in the past to the maternal uncle and the father-in-law, using the word kaka.

➢ Kaka in Quechuan languages

The term kaka has been recorded in Quechuan languages during the 16th and 17th centuries in various regions of Peru.

a. On the central coast around Lima:

In his Lexicon and Grammatica, both published in 1560, the Dominican Santo Tomás reported the forms ‘cacay [kakaj] abuelo de mi muger’ (WGF) (1560a: 2 v.), ‘caca [kaka] suegro, padre de la mujer’ (WF) (1560a: 98), ‘caca [kaka] tio, hermano de madre’ (MB), ‘caca [kaka] tio, hermano de abuelo, hermano de abuela’ (U, GFB, GMB) (1560a: 100), ‘caca [kaka] tio hermano de madre’ (MB), ‘cacay [kakaj] abuelo de mi muger’ (WGF) (1560a: 111 v.), céca [kaka] ‘dize el sobrino al tio’ (U) (1560b: 8). It is alleged that the language described by Santo Tomás was a now extinct variety of Quechua, spoken on the coast around Lima and generally considered as belonging to the Q II branch of Quechuan29. Adelaar and Muysken (2004: 182, 191) and other authors believe that this form of Quechua was likely identical to the language of the Inca’s administration (Inca general language). So the word kaka not only belonged to the lexicon of the coastal dialect but was understood everywhere across the Inca empire.

b. Around Lima and in the Central Peruvian sierra:

Among the native religious practices (“superstitions”) described in his book Extirpación de la Idolatría del Piru, the Jesuit father Arriaga (1621: chap. 6, p. 33-4) mentioned a ceremony where four or five year old children had their hair cut for the first time, and had their names changed. At this occasion, the relatives and specially the “cacas [kakas] y massas” were invited.30 Arriaga is not specific about the places where such ceremonies were held, but he observed that the locks of hair that were cut and considered as sacred objects, were differently called according to the region where the event occurred, either in the sierra or in the llanos (the lowlands near Lima). In his first chapter, Arriaga mentioned

29 Cerrón-Palomino (1990: 340) notes that the form of Quechua found in Santo Tomás’ Lexicon and Grammatica has much to do grammatically with the southern dialects, and lexically with the central and northern dialects.

30 Mas(s)as, in 16th century Quechua from Ayacucho and central Peruvian dialects, denotes the affinal relationships of brother-in-law and DH. About the mas(s)a-caca relationship, see Zuidema (1977: 259, 261-262) and Webster (1977: 39).
the places that had been already visited by the ecclesiastical personnel in charge of the extirpation of idolatry, covering parts of the Central Peruvian Sierra as well as the Coastal Lowlands around Lima. This statement, as well as other information collected from early colonial documents and pertaining to different parts of Peru, make us believe that such haircuts were common practice throughout Peru in the pre-colonial period.

Moreover, although carefully differentiating between the various regional terms designating the hairlocks, Arriaga does not mention any terminological variation for mother’s brother and brother-in-law. Caca [kaka] “tío hermano de madre” and masa “cuidaño, pariente de afinidad” are among the word listed in the index of his book (1621: Indice de algunos vocablos). We also find these words in the “edict against idolatry” published in the same book (1621: 129-133). This edict, covering the Middle Andean and coastal parts of the vice-kingdom of Peru, contains a list of questions that ecclesiastical visitors were supposed to ask the inhabitants of the towns they visited, in order to unveil native superstitions. Under item 14, it was notably asked to the natives to denounce people who they knew were performing haircut ceremonies, to which were invited the “tios” and the “cuidaños” – called “caca” and “masa”, respectively. Put together, these pieces of information allow us to infer that “caca” [kaka] was the general term for MB in the regions already visited, and mentioned by Arriaga in his first chapter, as well as those not yet visited. Arriaga’s “sierra” regions are those where many Quechuan languages of the Q I branch are spoken today.

c. In the Cuzco region:

Numerous colonial documents, glossaries, and grammars, dated from the second half of the 16th century or shortly after, mention the existence of the word caca [kaka] MB in the Cuzco dialect of Quechua (Yaya 2008a, 2008b). Three publications from this period deserve to be mentioned in our paper, because of the invaluable linguistic material they contain, related to kinship and to kinship terms. The first one is an anonymous Arte y Vocabulario printed in 1586 by Antonio Ricardo, whose author is probably Alonso de Barzana (Zuidema 1977: 242), in which we find – notably in the Anotaciones – the word caca [kaka] or cacay [kakaj = 1st person possessive], glossed: ‘(cacay) hijo de mi tío hermano de mi madre’ (MBS), ‘(cacay) tio del varón o de la mujer, hermana de su madre’ (MB), ‘(caca) dize el consuegro padre del hijo a su consuegro, y el a el, caca’ (SWF), caca dize la consuegra madre del hijo a su consuegro (SWF), ‘(catay) dize el suegro al yerno, y el a su suegro caca’ (WF), ‘(cacay) dize el yerno a su cunado hermano de su mujer’ (WB). The second and third documents are the Gramatica and Vocabulario, written by González Holguín and published in 1607 and 1608, respectively, where we find caca [kaka] glossed: ‘tío hermano de madre’ (MB) (1607: 96 v.; 1608: 35), ‘los consuegros

31 Arriaga (1621: Chap. 1. notably p. 7) stresses the fact that the rites described in his book had been directly observed by him, or observed by other visitors of as much authority as himself, in the provinces of Huarochoi, Yauyos, Jauja, Tarma, Huaylas, Chinchaycocha (Junín), in the districts of Andajes (province of Oyón) and Checras (province of Huaura), in the city of Huanuco, etc., all situated in the archbishopric of Lima. In many of the regions cited, dialects of the Q I branch of Quechuan are spoken.

32 Similar ceremonies stressing the role of the maternal uncle were held throughout Peru at the time of the Spanish conquest. This haircut is sometimes performed by the “closest uncle” on one-year-old infants, and called rutuchiku. For Chinchaycocha (department of Junín), see Duviols (1974: 277-8). For Cuzco, see Yaya (2008a: 71, and 2008b: 208).
varones entre sí’ (CEF), ‘caca, el suegro del yerno que es el padre de su muger (WF), que al suegro y a su padre le llaman el hierno y su hermanos y primos’ (WF, WFF) (1607: 98), ‘caca, cuñado del, que el cassado llama a todos hermanos y primos de su muger caca’ (WB, W's (cousins) (1607: 98, 98v), or ‘caca, dizen tambien los hermanos del varò casado como el a sus cuñados varones, hermanos de su cuñada’ (WB, BWB) (1607: 98v). The Quechuan dialects of the Cuzco region belong to the Q lIc branch of Quechua 33.

To our knowledge kaka has been preserved in modern Quechua of the Cuzco region: in Pitumarca (Milicic 2011), and by modern Q'ero who use kakay to refer to MB, MBS, MMBS (Webster 1977: 30). It is also used referring to MB in the Quechua spoken in Ayacucho, and also referring to MB, EF in the Quechua spoken in the Apurimac region. 34

Table 9. Quechuan comparative data.

<table>
<thead>
<tr>
<th>Proto-Quechuan</th>
<th>*k</th>
<th>*a</th>
<th>*k</th>
<th>*a</th>
<th>*y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santo Tomás (1560): Inca general language, Costal Quechua: Q Ihb?</td>
<td>k</td>
<td>a</td>
<td>k</td>
<td>a</td>
<td>y</td>
</tr>
<tr>
<td>Arriaga (1621): Region of Lima and Central Sierra: Q II, Q I</td>
<td>k</td>
<td>a</td>
<td>k</td>
<td>a</td>
<td>MB</td>
</tr>
<tr>
<td>cuzco (1552, 1570, 1586, 1607): Q lIc</td>
<td>k</td>
<td>a</td>
<td>k</td>
<td>a</td>
<td>y</td>
</tr>
<tr>
<td>Huarochiri (ca. 1608): Q lIc?</td>
<td>k</td>
<td>a</td>
<td>k</td>
<td>a</td>
<td>MB</td>
</tr>
<tr>
<td>Q'ero (Modern): Q lIc</td>
<td>k</td>
<td>a</td>
<td>k</td>
<td>a</td>
<td>y</td>
</tr>
<tr>
<td>Pitumarca (Modern): Q lIc</td>
<td>k</td>
<td>a</td>
<td>k</td>
<td>a</td>
<td>MB</td>
</tr>
<tr>
<td>Ayacucho (Modern): Q lIc</td>
<td>k</td>
<td>a</td>
<td>k</td>
<td>a</td>
<td>MB</td>
</tr>
</tbody>
</table>
| Apurimac (Modern): Q lIc | k  | a  | k  | a  | MB, EF 

Proto-Quechua had a 3-vowel system *a, *i, *u preserved in present-day languages with a wide range of allophonic realizations. Consonant *k has been preserved in all languages (Adelaar and Muysken 2004: 194-198). The cognate set kaka reflects Proto-Quechuan *k > k and *a > a. Thus we can propose *kaka MB, WB in Proto-Quechuan with some confidence, all the more so as kaka was in use 500 years ago in all Quechuan branches.

➢ Kaka in Aymaran languages

Hardman (Pyle 1981: 89, editor’s [i.e. Hardman’s] note) says that Aymara also referred in the past to MB as kaka. But this word is mentioned neither in the Arte y gramatica muy copiosa de la lengua Aymara, nor in the Vocabulario de la lengua Aymara both written by Ludovico Bertonio and published in 1603 and 1612, respectively. What we find in the Vocabulario (1612, vol. 2: 191) is “Lari: tio hermano de la madre y casi a todos los varones parientes de parte de madre llamán lari” (“MB and all the masculine parents on the mother’s

The authors want to thank Pierre-Luc Abramson, professeur honoraire des universités (University of Perpignan, France), researcher at the CRHISM (Centre de recherches historiques sur les sociétés méditerranéennes), and César Itier, maître de conférences at INALCO (Institut des langues et civilisations orientales) for their assistance in the translation of the Golden Age Castilian transcriptions of Quechuan and Aymaran kinship terms.

Most Quechuan dialects today refer to both maternal and paternal uncles as tiyu (from Spanish tío). Hardman (1982: 148) states that kaka was replaced by tiyu because of “interlingual taboos and the preoccupations on the part of the priests regarding what they imagined to be incest”, probably the fact that kaka referred to both MB and EF.
side are called ları”). This word is still used today in modern Aymara of Southern Peru at least, with the meaning WB, and rarely, if ever, with the meaning MB (Collins 1981: 218, 234-5).

In contrast, modern Jaqaru and Kawki (Hardman 1983: 204, 206, Belleza Castro 1995, Hardman de Bautista 1969: book 2, 8) call both the paternal and the maternal uncles kāka and kaka, respectively.

- **The origin of kaka in Quechuan and Aymaran**

Quechuan and Aymaran have at least 9 kinship terms in common, and this fact, along with many other vocabulary similarities, is likely to be interpreted as loans from Quechuan into Aymaran. But Hardman’s claim (Pyle 1981: 89, editor’s note) is that, as with warmi ‘woman, wife,’ “in the majority of cases the direction of borrowing was from a Jaqi language to a Quechua language, particularly Cuzco Quechua.” And she continues: “Because Quechua was the imperial language at the time of the conquest, any similarity anywhere has been regularly attributed to Quechua origins. This is not the case; imperial Quechua was a relatively recent imposition. The full complexity of the prehistory of language contact in the Andes is yet to be told.”

So it is not clear yet which language the word kaka was borrowed from, if it was borrowed at all. This term might well have been original in both language families.

### 2.4.2. Koko in Mondé: ca. 2,000 BP

According to Moore (2005: fig. 1), the Mondé language group, a subfamily of Tupián, is composed of 3 languages distributed between 2 branches. The first branch comprises Suruí, while the second one is divided between Salamay and a third language including 4 dialects: Gavião, Zoró, Cinta Larga and Arua. Anonby and Holbrook (2013), for their part, acknowledge the existence of 6 languages in Mondé: Salamay, Aruá, Suruí, Gavião, Zoró, Cinta Larga, but their classification only accounts for the 4 languages still alive: Gavião, Zoró, Cinta Larga and Suruí.35 Their conclusion assumes that Zoró is more closely related to Suruí than it is to Gavião, and thus delineates two branches, with Gavião constituting the first one, and Cinta Larga and Suruí-Zoró constituting the second. Brunelli’s opinion (1987: 157) is that the Suruí separated from the other Mondé groups 1,500 years ago, while Macedo Brito (2005-2006) posits this split by ca. 2,000 BP. For their part, Anonby & Holbrook (2013: 29) assume, without much precision, that Gavião diverged “very early” from Proto-Mondé, whose homeland is posited along the Roosevelt and Aripuana Rivers by Brunelli (1987: 157) or in the region of the Ji-Paraná River’s headwaters by Campbell (1997: 199). The comparative material involving KOKO terms is compelling. Suruí, Gavião, Zoró, Cinta Larga KOKO terms are all address terms; we do not know if it is the same for Salamay.

In the absence of knowledge of the sound correspondences existing in the whole family, but to take in account the recurring sounds and meanings exhibited in the following cognate set (table 10), we will just propose, an “inspectional reconstruction” preceded by the symbol #.

<table>
<thead>
<tr>
<th>Table 10. Mondé comparative data arrayed according to Moore’s internal classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proto-Mondé</strong></td>
</tr>
<tr>
<td><strong>First branch</strong></td>
</tr>
</tbody>
</table>

35 Their classification concerns the languages where field research was possible.
2.4.3. KOKO in Nambikuáran: 2,807 BP (ASJP)

The Nambikuáran language family is composed of 2 major branches: Northern and Southern Nambikuáran, themselves including several languages (dialects), and one separate language, namely Sabanê (Eberhard 2009-1: 21). This family spans a territory covering the northwestern part of Mato Grosso and contiguous parts of the territory of Rondônia (Brazil). Directed by Professor Wetzel, a language study program was launched in 1998, notably involving the phonological and grammatical description of the languages from the 3 branches, and ultimately including a comparative study and a reconstruction of Proto-Nambikuáran (Telles & Wetzels n.d.: 236).

Several doctoral dissertations since then have been completed, notably Fonologia e Gramática Karirian (Telles 2002), A Grammar of Sabanê (Antunes de Araujo 2004), Mamatê Grammar: A Northern Nambikuáran Language and Its Cultural Context (Eberhard 2009). Another study concerning Sarare, a Southern Nambikuáran language, is in preparation (Borella).

The comparative list below displays the kinship terms referring to MB and EF collected from various languages of this family. Most forms display the root first two syllables, then a suffix -nu or -ni present in kinship terms (Eberhard 2009-1: 186), and the final nominal (referential) suffix -su, -te, -tu (Kroeker 2003: 12, 24, etc.; Eberhard 2009: 365-372, etc.). Not knowing what exact sound correspondences exist in the Nambikuáran family, we will again venture a tentative Proto-Nambikuáran form for MB/EF.

**Nambikuáran comparative data**

**Proto-Nambikuáran**\(^{36}\) *-kügkV-nVCV MB, EF; Southern Nambikuáran* Halotesu -küka-nusu MB, EF; Kokóze (Juina Kitáulhú?) -kókó-zu MB; Waklitisu -kųkį-nusu MB; Elotasu -kųkį-nusu MB; **Northern Nambikuáran** Mamatê -kųmûri MB; dialect b 1/2 (Lévi-Strauss 1948) -kïn-de MB, EF; **Sabanê** Sabanê kooka MB, FF, EF.

2.4.4. KOKO in Karirian

Kariri is considered either as one single (extinct) language including several dialects, or as a language family composed of several languages. The dialects (languages) for which some documentation is available, and thus mentioned in our report are Kipeá, Dzubukúá, Sabuá and Pedra Branca (Kamuri). The presence of Karirian-speaking groups is attested ca. 1670 in the states of Paraiba, Pernambuco and along the rio São Francisco, at least upstream of Belém de São Francisco, in a place named Aracapá (Ouracapa) (Martin de Nantes n.d. [published in 1706 or 1707]: 1, 2, 4, 22, etc.). In the preamble of his Portuguese-Kariri Katecismo Indico da língua Kariris, the Capuchin missionary Bernardo de Nantes (1709: aiiij) gives two precious bits of information: first, that the language in which the catechism is translated was called Dzubukúá and

---

\(^{36}\) A subscript tilde marks a creaky voice (laryngealized) vowel.
spoken along the rio de São Francisco, thus confirming what Martin de Nantes previously reported. Second, in the same preamble, he also mentions that the language in which the Jesuit Mamiani wrote another Catecismo (1698), and also an Arte de Gramatica: (1699) was a different language named Kippea, and adds that the two "nations" were geographically separated by "more than a hundred leagues" (1709: aiiij). In 1818, Spix & Martius (1828: 615) found two Karirian-speaking communities living in the south of what is now the little town of Santa Terezinha in the state of Bahia (Costa Neto 2007: 60): the first one was established in 'Villa de Pedra Branca,' (where the present-day village of Pedra Branca is situated), a toponym which became linguistically associated with the ethnic group, the second one was established "a quarter of an hour" south of this first settlement, in a place then named 'Caranguejo,' and formed a group known as the Sabujás. Altogether these groups consisted of some 600 people. Spix & Martius (1828: 615) added that before they settled down in these two places, the Karin's were scattered in the woody mountains nearby – probably in the Serra da Jibóia.

Rodrigues (1986: 48-55, 1999: 164-206) considers that the Karirian language(s) belong(s) to the Macro-Jê trunk; Greenberg (1987: 84, 384), and Greenberg and Ruhlen (2007: 280) hold that Karin is close to Tupian, including both of them in a putative Kariri-Tupi cluster, itself a component of the Equatorial stock. Kaufman (2007: 73) says that Kariri is an unclassified language.

The first mentions of KOKO kinship terms in Karin (Kipeá) are found in Mamiani’s Catecismo (1698: 197): icucit [ikuku] 'os tios,' as well as in his Arte de grammatica (1699: 19): cucit 'tio.' In Bernardo de Nantes’ Katecismo (1709: 91), displaying the Dzubukua dialect (language), we find i-cucu-te ‘os tios.’ Von Martius, for his part (1867a: 359, 1867b: 215, 218) recorded cucu and cucuith 'Oheim, avunculus' (MB) for Sabuja, and cucuinh 'avunculus' (MB) for the Kariris living in Pedra Branca. The term kuk(k)u(h) MB is consequently not a newly acquired or borrowed term in this language family, and can be postulated in Proto-Karirian.

2.5. KOKO in other South American families and languages isolates

 Families

Barbacoan: Guambiano kásuко U; Colorado Ḵa'ko B; Chocoan: Emberá Catio káu D; Choco (Emberá) kau D; Epena Ḵau D, girl; Noanamá ka D, kái Z, kawá EM; Cholonal: Xibito kók F, keek M; Chonan: Puelche Ḵitči B; Tehuelche kóka F, dueño, kok’an madrina, gôo B, Z, qon GM; Selknam ho’o GF, hoho’nh GM; Jéan: Northwestern Jé: Panará (Kren Akarore) kokri-pia EM; Southern Jé: *káke eB, *kakř MB, EF; Kaingang (Toldo das Lontras, Nonoai, etc.) kakre MF, MB, WF; Kaingang (Rio Grande do Sul) kakrö MF, MB, EF, kêke eB, eZ; Kaingang (Palmas, Parana) kakrá ~ kakř EF, kâke eB, eZ; Kaingang (Serra do Chagu, Parana) ikáke B; Kaingang (Duque de Caxias) = Xokléng kôkla ceremonial F = MB, kakê relative, cousin, but not B, rarely A; Ingáin kau U; Jivaroan: Jivaro Acheh uai (?B). (?)Z; Huambisa kair (?Z); Aguaruna kaig (?Z); Kamakanan: Kamakán gkoong GF, MB; Karajá language area: Karajá hi (?E)B, ixi yB;

37 The prefix i- is the 3rd person possessive marker.
38 Adam (1897) gathered a lexical comparative material based on the reports done about the Kipeá and Dzubukua dialects, by Mamiani and Bernardo de Nantes respectively, and about Sabujá and Pedra Branca dialects, done by von Martius. Adam, who faithfully reported von Martius’ transcriptions, remarked that the final vowels in Pedra Branca and Sabujá were frequently affected by an aspirated h (1897: 6). His comparative glossary also shows the occurrence, notably in the transcriptions of the second syllable in Sabujá and Pedra Branca languages (henceforth S. and P.B.), of geminated consonants c [k]. m. t. b or p. not present in Kipeá and Dzubukua (henceforth K., D.): K., D. cucu MB, S., P.B cucuinh MB; K., D. ami ‘food’, P.B. ammih id.; D. batí star, P.B. batiíh id.
Krenánakan: Botucudo ñ-gike'ñ F, ÊF; Maxakalián: Maxakali xaxíyí MB, EF, FF, MF, FFF, MFF, i-kuax, FZ, EM, GM, MMM; Lule-Vilelan: Lule-Tonocote kue FZ; Makúán: Hupda ʔaw GF, ʔoh GM, ʔe? eB; Nadéb óow GF, hooí GM; Puinave hika A, -uíi EF; Mascoyan: Sanapaná Enlhít dial. haavok eyag eB; Mascay (western) koko MB, havok eB, BS; Mascayo (Lenguá) havok eB, BS; Matacoan: Chuluípí (Nivaclé) k’uca?x old man, k’uca?x old woman, yi-kakt’e WF, yi-kakt’e my WM; Chorote kihíl ~ kihí FF, old man; Macá keiuk B, k’uca?x old man, -kewket EF, -kewket-iP EM; Matacoan: Wichi Lhamtés Vejóz -kó M, kawaká ZH; Muráng: Pirahá kaai D; Paezan: Paez ká?ka MB, MZH; Qawasqaran: Alakuluf (Qawasqar) xoyko-las old woman; Sálíban: Sáliba koko man, tihoho A; Timotean: Timote kúxioy GM; Mucuchi kaak man; Tukanoan: Coreguaje cu’eu GF, cu’eo GM; Siona k’i MB, MZH; Mondean #koo(t)koo(t) MB, (EF) (see 2.4.2); Paurúborá Paurúborá koko MB; Raménaman Káro (Urúkú) íká eZ, hakoiga H; Tupí Tupinambá -ike (c5’) eB; Eastern Tupían: Mundurukú Mundurukú ukípíd (c3’), o-köt-köt yMZS, yFBS; Mawé Mawé (Maué) uy-kiwid (c7’)B; Aweti Aweti i-kywít (c7’)B; Tupí-Guaraní ñ-kíwyra (c7’)B; Guaraní branch, Guayaki (Aché) kyvagí (c7’)B; Xingu branch Araweté ciwi (c7’)B; Kayabi kiwit (c7’)B; Apiaká erarkuíre (c7’)B; Northern Tupí-Guaraní branch Ka’apór ihékywyra (c7’)B; Wayampí kaky eB; Witotoan: Nonuya hókhá man, hóhekíWB, HB hokohókí WZ, HZ; Ocaina hohí man, hóhóhó(hi)š WB, HB, hóhóhó cousin; Witoto uota uakiká GF = F + old man; Zamucoan: Ayoreo axai (c7’)B, axu MZ; Chamacoco -okyok EF; Ebidoso -ohot EF; Turena -ohi ‘my EF’.

> Language isolates

Aikanà (Masaca) kokonai U; Kaliana (Papé) ma-kohai ~ ma-kyhái my GF, kohai old man; Camsá -ki MB; Canichana eu-axa my GF; Cayuvava -kie U; Cofán toto U; Capixán ñ (Kanoe) kúkúi man, keke GM, uká HBD, WBD; Koayá (Kwaza) ha’kái GP; Mapudungu (Mapuche) huiku MB, kúcu FZH, kuku FM, FFZ; Mochica (Chimú) kokoal (c7’)eZ, (c7’)A, ikiš EM; Movima aiku A, ákai eB, eZ; Munichic te’aró U; Oti kooba B; Taushiro ’ikku (c7’)Z; Trumá koko MZ, FBW, aoké EM, FZ; Urarina (Simacú) ka-kum FZ; Warao ku, da -ku MB; Yató (Fulnio) ñ-ki B.

3. The kinship term koko in Meso-America

Among the first records of KOKO kinship terms in Meso-America, we find Tarascan cucu [kuku] GM, by Gilberti (1901 [1559]: 33), Mixe ocataac [ok(taak)] GM, A, by Quintana (1733: 80-81), Totonac coco [koko] U, by Zambrano Bonilla (1752: distintos p.1, nombre de parentesco p. 3).

---

39 We use the term Meso-America to stay in line with Kaufman’s (2007b) terminology.
3.1. The geographic and linguistic distribution of KOKO in Meso-America

We have studied 178 languages/terminologies\(^40\), living or extinct, from this region. KUKU ~ KAKA forms have been found in all of the 11 genetic units acknowledged in Kaufman’s (2007) classification (map 2, table 11). The Lencan and Misumalpan languages mentioned in this classification have been already presented in subsection 2.3.7.

3.2. The reconstruction of KOKO in Meso-American linguistic families

From now on, the linguistic material will be presented by means of tables. These tables display the languages, the language families with their supposed time-depth, the ‘stocks’ when there is some consensus about their existence, and the reconstructed forms when available. These reconstructed forms haven’t been evaluated like their South American counterparts, because of the

\(^{40}\) Lenca-Misumalpan languages not included.
space that such a discussion would entail. We will also venture a few “inspectional reconstructions” when the comparative tables display very similar forms.

<table>
<thead>
<tr>
<th>Stocks, families, languages</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Tequislatec-Jicaque (Tol)</strong></td>
<td>Proto-Jicaque *(ko)koq (♀)GF</td>
<td>Proto-Jicaque *(ku)kus D</td>
<td>R: Campbell &amp; Oltrogge 1980</td>
</tr>
<tr>
<td>Jicaque</td>
<td>kokoy FF, (♀)GC, kik</td>
<td>kukus D, SW, kik (♀)GC</td>
<td>Turner &amp; Olmsted 1966</td>
</tr>
<tr>
<td>Western Jicaque</td>
<td>cui, cohó GF, goain U</td>
<td>gut D</td>
<td>Neuenswander 1977-81</td>
</tr>
<tr>
<td>Montaña de la Flor</td>
<td>kokoy GF, kokamam U, ngokam my U</td>
<td>kukus ~ kuku D</td>
<td>Campbell &amp; Olt. 1980</td>
</tr>
<tr>
<td>Cabeza de Vaca</td>
<td>kokoy old man, kokam U</td>
<td>kukus D</td>
<td>Campbell &amp; Olt. 1980</td>
</tr>
<tr>
<td>Lagunita</td>
<td>cocoy muy old man, uncacom U</td>
<td>kuku D</td>
<td>Campbell &amp; Olt. 1980</td>
</tr>
<tr>
<td>Lean and Mulia</td>
<td>cocoy old man</td>
<td>kuku D</td>
<td>Campbell &amp; Olt. 1980</td>
</tr>
<tr>
<td>Jicaque de Yoro</td>
<td>cocam U</td>
<td>cucusuy D, guías DinL</td>
<td>Campbell &amp; Olt. 1980</td>
</tr>
<tr>
<td><strong>2 Totonac-Tepehuan: 2,500 BP?</strong></td>
<td>*kuku U (MB)</td>
<td></td>
<td>University of Alberta</td>
</tr>
<tr>
<td>Totonac (18th cent.)</td>
<td>coco [koko] U, gag</td>
<td></td>
<td>LDRC, R: AM&amp;PB</td>
</tr>
<tr>
<td>Filomeno Mata</td>
<td>chhi (♀)eB</td>
<td></td>
<td>Zambrano Bonilla</td>
</tr>
<tr>
<td>Totonac</td>
<td>kikbu'U, -kukusta'q</td>
<td></td>
<td>1752</td>
</tr>
<tr>
<td>Totonac of Xicotepec de Juárez</td>
<td>ix-kuk thy U</td>
<td></td>
<td>McFarland 2009</td>
</tr>
<tr>
<td>Upper Necaxa</td>
<td>kuku one's U, pl.</td>
<td></td>
<td>Reid &amp; Bishop 1974</td>
</tr>
<tr>
<td>Totonac</td>
<td>kuku'mu</td>
<td></td>
<td>Beck 2011</td>
</tr>
<tr>
<td>Totonac Papantla</td>
<td>cucu [kuku] U</td>
<td></td>
<td>Aschmann 1973</td>
</tr>
<tr>
<td>Misantla Totonac</td>
<td>kin-kuk my U</td>
<td></td>
<td>MacKay 1999</td>
</tr>
<tr>
<td>Huehuetsla Tepehua</td>
<td>kuuk MB</td>
<td></td>
<td>Smythe Kung 2007</td>
</tr>
<tr>
<td>Tlachichileo Tepehua</td>
<td>kin-kuku my U</td>
<td></td>
<td>Watters 2010</td>
</tr>
<tr>
<td>Tepehua de Pisaflares</td>
<td>kin-kuku [kîn kûkû] my U</td>
<td></td>
<td>MacKay &amp; Trechsel 2013</td>
</tr>
<tr>
<td><strong>3 Mixe-Zoquean: 3,000 BP</strong></td>
<td>*?oko GM, (GC)</td>
<td></td>
<td>Campbell &amp; Kaufman 1976, R:</td>
</tr>
</tbody>
</table>

41 For reason of space, the references – except a few in the general references – have been only reported by author names and publication dates. Complete references can be obtained from the first author: a.matthey@free.fr, or found online at language-kinship.org (click on tab Databases). References of reconstructed terms (in bold) marked * or # can be found in the general references.
### Stocks, families, languages

<table>
<thead>
<tr>
<th>Stocks, families, languages</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixe (Totontepex)</td>
<td>?ahé (♂) older male relative, ?oknì hš- GS</td>
<td>ok GM, GD</td>
<td>Bcales 1945</td>
</tr>
<tr>
<td>Mixe (Coatlan)</td>
<td>ok GS</td>
<td>ok GD</td>
<td>Hoogshagen &amp; Merrifield 1961</td>
</tr>
<tr>
<td>Mixe (Metaltepec)</td>
<td>okunk (♂)GS</td>
<td>ok, oktaak GM, oknox (♂)GD, xoixc (♂)EM</td>
<td>Beals 1945</td>
</tr>
<tr>
<td>Mixe (Juquilla)</td>
<td></td>
<td></td>
<td>Beals 1945</td>
</tr>
<tr>
<td>Zoque</td>
<td>oko MF, MGP, oko unk GS</td>
<td>oko han GD</td>
<td>Radin 1931 (Quintana 1733), Beals 1945</td>
</tr>
</tbody>
</table>

### 4 Mayan: ca. 4,200 BP

<table>
<thead>
<tr>
<th>Languages</th>
<th>Denotata</th>
<th>References</th>
</tr>
</thead>
</table>

### Huastecan

- **Huastec**
  - koko tutor
  - Radin 1931, Schuller 1928

- **Huastec Western (Tancanhuitz)**
  - itxaan U, itxak’ nephew
  - Kaufman & Just. 2003

### Yucatecan-Core Mayan

- **Epigraphic Mayan Maya (16th – 18th centuries)**
  - ?ichaan MB
  - Kaufman & Just. 2003

- **Yukateko (Mopan)**
  - acaen MB, (a)chak male kin
  - Kaufman & Just. 2003

### Lacandon

- **Core Mayan**
  - ikán MB, ZS, WF
  - kik eZ. chiich MM
  - Boremans 1979

### Quiche

- **Q’anjeb’al-Chuj’ecan**
  - ikan MB, ikan-ey U
  - ikan-ey A
  - Kaufman & Just. 2003

- **K’iche’-Mayan**
  - ikan MB
  - Kaufman & Just. 2003

- **Quiche**
  - r-ikan MB
  - Kaufman & Just. 2003

- **Chichicastenango**
  - male cousin.
  - Kaufman & Just. 2003

- **Chicaj**
  - nephew
  - Kaufman & Just. 2003

- **Sipakapense**
  - kyaan U
  - Kaufman & Just. 2003

- **Tz’utujil**
  - ityan U
  - Kaufman & Just. 2003
### MOTHER TONGUE

*Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014*

<table>
<thead>
<tr>
<th>Stocks, families, languages</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Luis Jalotepeque</td>
<td>ikaan U</td>
<td>ikaa A</td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>Usstanteko (Musre)</td>
<td><em>r-ikaan U</em></td>
<td>*Ikaan na? (ikan na') A, ikaq' nephew</td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>Q'eqchí? (Kekchi)</td>
<td>ikan U, ikak'bej nephew</td>
<td>ikak'bej niece</td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>Western Q'eqchí? (Coban, Chamelco)</td>
<td><em>ikan MB, ik'aq nephew</em></td>
<td><em>ik'aq niece</em></td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>Mam /qyool/</td>
<td><em>ikiam U</em></td>
<td>ikia B</td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>San Idlefonso</td>
<td><em>t-kyaan (♂)male cousin</em></td>
<td><em>t-kyaan (♂)female cousin</em></td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>Teko (Cuijqueno)</td>
<td>n-chaan U</td>
<td><em>kwaHn (♀)male</em></td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>ca. 4,000 BP?, 6,400 BP?, 7,400 BP?</td>
<td><em>kuHn Sib</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lenca-Misumalpan:</td>
<td><em>kVbV GF, (MB), (EF)</em></td>
<td></td>
<td>Kaufman &amp; Just. 2003</td>
</tr>
<tr>
<td>7,200 BP (see section 2.3.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isolates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Tarascan</td>
<td>cucu [kuku] GM</td>
<td>Radin 1925 (Gilberti 1559)</td>
<td></td>
</tr>
<tr>
<td>8 Xinca</td>
<td>aguà GM</td>
<td>Breton 1919 (1770)</td>
<td></td>
</tr>
<tr>
<td>8 Cuitlatec</td>
<td>ohci GF</td>
<td>Drucker, Escalante, Weitlaner 1969</td>
<td></td>
</tr>
<tr>
<td>9 Huave</td>
<td>-koh eB, okwaàk WF</td>
<td>-koh eZ, okwaàk WM</td>
<td>Diebold Jr 1966</td>
</tr>
</tbody>
</table>

---

4. The kinship term **KOKO** in North America

Among the first **KOKO** kinship terms recorded during the postcontact period are those from the Algonquian languages: Wampanoag (Natick) kokimmes thy A, kokummes thy GM (Trumbull 1903, from Eliot's Bible, 1663), Montagnais *n8k8nis* [nookoomis] FB (Silvy ca. 1678), Abenaki *n8’k8mes, n8’k8mi* [nookoomes, nookoomi] my GM, my A, *n8’k8m* [nookoom] my UW, my UZ (Rasles 1691 [1833]: 498-9), Miami *n8e8ma* [nookooma] GM (Gravier 1700). Let us also mention caca [kaka] FM in Cohahuiltecan by Garcia (1760), quoted in Romney (1967: 229).

4.1. The geographic and linguistic distribution of **KOKO** in North America

252 languages (or dialects) out of the 360 North American languages and dialects of our sample display **KOKO** kinship terms. As shown on map 3, **KOKO** terms pervade the entire North American territory and are found in 39 out of the 58 genetic units considered in Golla *et al.* (2007).
Map 3. Approximate locations across North America of koko proto-forms (in bold), with additional terms from individual languages showing the extension of the distribution of koko words.

4.2. KOKO in languages and language families of North America
As was done for Meso-America, the linguistic material is presented in a table.
Table 12. North American comparative data

<table>
<thead>
<tr>
<th>Stocks, families</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eskimo-Aleut</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eskimo</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inuit</td>
<td>*akkak FB</td>
<td>*aka eZ, M</td>
<td>R: EHL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*a(a)kkak M, eZ</td>
<td>R: Fortescue et al. 1994, EHL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inuktitut</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iglulik</td>
<td>qak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Simpson Peninsula</td>
<td>ak'aq FB</td>
<td></td>
<td>Birket-Smith 1928</td>
</tr>
<tr>
<td>Melville Peninsula</td>
<td>a'kan FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Pond Inlet</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Chesterfield Inlet</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Upper Kazan River</td>
<td>ak'aq FB</td>
<td></td>
<td>Birket-Smith 1928</td>
</tr>
<tr>
<td>Southampton Island</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Rankin Inlet</td>
<td>acug FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Baker Lake</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Eskimo Point</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Cumberland Inlet</td>
<td>zuk'-iγuih FB</td>
<td></td>
<td>Dall 1877</td>
</tr>
<tr>
<td>Frobisher Bay</td>
<td>akakulu FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Lake Harbour</td>
<td>akakulu FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Sugluk</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Port Harrison</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Labrador</td>
<td>akka FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Great Whale River</td>
<td>aka' FB</td>
<td></td>
<td>Rasmussen 1941</td>
</tr>
<tr>
<td>West Greenland</td>
<td>akak FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Thule (North)</td>
<td>ak'ak FB</td>
<td></td>
<td>Gessain et al. 1982</td>
</tr>
<tr>
<td>Green.)</td>
<td></td>
<td></td>
<td>Birket-Smith 1928</td>
</tr>
<tr>
<td>Northumberland</td>
<td>uk'-kä FB</td>
<td></td>
<td>Morgan 1871</td>
</tr>
<tr>
<td>East Greenland</td>
<td>akka FB</td>
<td></td>
<td>Gessain et al. 1982</td>
</tr>
<tr>
<td><strong>Imiutut</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Alaskan</td>
<td>a'ka M, äxxa eZ.</td>
<td></td>
<td>EHL</td>
</tr>
<tr>
<td>Innupik</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Point Barrow</td>
<td>akaakuk FB</td>
<td></td>
<td>Stevenson 1964</td>
</tr>
<tr>
<td>Point Hope etc.</td>
<td>aqigia ~ okaakayu FB, aakakayu M</td>
<td></td>
<td>Heinrich 1960</td>
</tr>
<tr>
<td>Nunamiut</td>
<td>akanakan male cousin</td>
<td>aaka M, aqakarak stepM</td>
<td>Pospisil &amp; Laughlin 1963, Pospisil 1964, Rasmussen 1941</td>
</tr>
<tr>
<td>Kangianermiuat dialect</td>
<td>ak'ak'ak FB</td>
<td>a'kan M</td>
<td>Heinrich 1960</td>
</tr>
<tr>
<td>Bering Straits</td>
<td>akaakayu MB</td>
<td></td>
<td>EHL</td>
</tr>
<tr>
<td>Seward Peninsula</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Yupik</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alutiiik Yupik</td>
<td>*aka eZ</td>
<td></td>
<td>R: EHL</td>
</tr>
<tr>
<td>Norton Sound</td>
<td>akąq eZ</td>
<td></td>
<td>EHL</td>
</tr>
<tr>
<td>Sirenik</td>
<td>akąq eZ</td>
<td></td>
<td>EHL</td>
</tr>
<tr>
<td>Chaplino</td>
<td>aką eZ</td>
<td></td>
<td>EHL</td>
</tr>
<tr>
<td>Naukan</td>
<td>akąq eZ</td>
<td></td>
<td>EHL</td>
</tr>
<tr>
<td><strong>Aleut</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Aleutian</td>
<td>kukaq GM</td>
<td></td>
<td>Geoghegan 1834</td>
</tr>
<tr>
<td>and Alaskan Islands</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

See also note 25.
<table>
<thead>
<tr>
<th>Stocks, families</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2 Nuclear Na-Dene: 8,532 BP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tlingit</td>
<td>-k 'a'k' MB, k 'i'k' ((\tilde{c}))yB, 'i'k' ((\tilde{c}))B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyak-Athapaskan: 3,500 BP, 4,234 BP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyak</td>
<td>aqaq MeB, aqaq-cta</td>
<td>ak(\tilde{c}) ((\tilde{c}))FeZ, ak(\tilde{c})-cta ((\tilde{c}))FYZ</td>
<td>Mayer-Durlach 1929</td>
</tr>
<tr>
<td>Athapaskan: 2,500 BP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanaina</td>
<td>-o'k(\tilde{a}) MZ</td>
<td></td>
<td>Campbell 1997, ASJP</td>
</tr>
<tr>
<td>Anvik</td>
<td>-t(\tilde{a}) MZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanana</td>
<td>-a(\tilde{a}) MZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kutchin</td>
<td>-(\tilde{a})k(\tilde{a}), ak(\tilde{a}) MZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yellow Knife</td>
<td>ak(\tilde{a})a MZ, stepM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dene (Chipewyan)</td>
<td>-(\tilde{a})(\tilde{a}) MZ</td>
<td>*-ankay? MZ</td>
<td></td>
</tr>
<tr>
<td>Carrier</td>
<td>ak(\tilde{a}) MZ, ZD, akwal my M</td>
<td>Goldman 1941</td>
<td></td>
</tr>
<tr>
<td>Tolowa</td>
<td>kaka M, onkai MZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hupa</td>
<td>-y(\tilde{k})a? MZ</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Wailaki</td>
<td>-(\tilde{k})kait MZ, FBW, stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Lassik</td>
<td>unkai MZ, FBW, stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Sinkyone</td>
<td>unkai MZ, FBW, stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Navajo</td>
<td>-(\tilde{a})k(\tilde{a})? MZ, stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>San Carlos</td>
<td>-(\tilde{k})(\tilde{a})(\tilde{a}) MZ, ((\tilde{a}))SC, stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Chiricahua</td>
<td>-(\tilde{a})(\tilde{a}) stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Mescalero</td>
<td>-(\tilde{k})(\tilde{a}) MZ</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Jicarilla</td>
<td>-(\tilde{k})(\tilde{a}) FBW, stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Lipan</td>
<td>-(\tilde{k})(\tilde{a}) stepM</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Kiowa Apache</td>
<td>-k'(\tilde{a})(\tilde{a}) MZ</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Apache</td>
<td>-k'(\tilde{a})(\tilde{a}) MZ</td>
<td>Donald &amp; Tighe 1987</td>
<td></td>
</tr>
</tbody>
</table>

<p>| <strong>3 Algic: 7,200 BP?, 4,000 BP?, 5,554 BP?</strong> | | | |
| Wiyot | yi-d-oko-tck GM, EGM, hakwi E (endearing) | | Swadesh, Golla 2007, ASJP, R: AM&amp;PB |
| Yurok | ne-kwa EP | | Gifford 1922 |
| Algonquian: 3,000 BP, 3,343 BP | | | |
| Central Algonquian | | | |
| Montagnais | nookoomis my FB | noq(\tilde{a})m my GM, noqum(\tilde{a})my M FZ, ns(\tilde{a})ga(\tilde{a})'wi my M | Silvy ca. 1678, Speek 1918 |
| Cree | no'hkomis my FB | no'hkom my GM, no'koh ad.GM, ni(\tilde{a})(\tilde{a})wi'y my M | Hockett 1964 |
| Prairie Cree | no'komi my stepF | noh'kome' my GM, my EGM, n'g(\tilde{a})we my M | Morgan 1871 |</p>
<table>
<thead>
<tr>
<th>Stocks, families</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plains Cree</td>
<td>n'co'komis my MB</td>
<td>nokūm my GM, n'qawi my M</td>
<td>Skinner 1914</td>
</tr>
<tr>
<td>Woods Cree</td>
<td>no'komish my stepF</td>
<td>nokome' my GM, my EGM, n'gāwe my M</td>
<td>Morgan 1871</td>
</tr>
<tr>
<td>Lowlands Cree</td>
<td>no'komis my stepF</td>
<td>nokome' my GM, my EGM, n'gāwa my M</td>
<td>Morgan 1871</td>
</tr>
<tr>
<td>Ojibwa</td>
<td>nokomiss my GM, no'kko' ad. GM, minga my M</td>
<td>Trautmann &amp; Barnes 1998, Hockett 1964</td>
<td></td>
</tr>
<tr>
<td>Ottawa</td>
<td>nokomis' my GM, my WGM, n'gus'-sheh my M</td>
<td>Morgan 1871</td>
<td></td>
</tr>
<tr>
<td>Miami</td>
<td>nēkoma' my GM, my WGM, (c)EM, ningeah M</td>
<td>Gravier 1700, Morgan 1871, Costa 1999</td>
<td></td>
</tr>
<tr>
<td>Peoria</td>
<td>nokomā' my GM, my WGM, ningeah M</td>
<td>Morgan 1871</td>
<td></td>
</tr>
<tr>
<td>Piankeshaw</td>
<td>nokomā' my GM, my WGM, ningeah M</td>
<td>Morgan 1871</td>
<td></td>
</tr>
<tr>
<td>Kaskaskia</td>
<td>nokomā' my GM, my WGM, negeah M</td>
<td>Morgan 1871</td>
<td></td>
</tr>
<tr>
<td>Weaw</td>
<td>nokomā' my GM, my WGM, negeah M</td>
<td>Morgan 1871</td>
<td></td>
</tr>
<tr>
<td>Kickapoo</td>
<td>no'komeza my GM, nō'ko ad. GM, nō'koma my EM, nega my M</td>
<td>Dyneley Prince 1913</td>
<td></td>
</tr>
<tr>
<td>Potawatomi</td>
<td>nok'mas' my GM, nkye my M</td>
<td>Hockett 1964</td>
<td></td>
</tr>
<tr>
<td>Menominee</td>
<td>no'hkomeh my GM, no'hkoq ad. GM, nekiah my M</td>
<td>Hockett 1964</td>
<td></td>
</tr>
<tr>
<td>Fox</td>
<td>no'gome's'GM, ano'hko ad. GM, no'hkoma my EM, nogum'EM, neg'EM</td>
<td>Eggan 1937, Hockett 1964</td>
<td></td>
</tr>
<tr>
<td>Sauk</td>
<td>no'komis my GM, my WGM, noko'mā' my EM, nākea' my M</td>
<td>Morgan 1871</td>
<td></td>
</tr>
<tr>
<td>Shawnee</td>
<td>nokomqda my GM (form uncertain), nokomeea my GM, WGM, nakeah my M</td>
<td>Morgan 1871, Hockett 1964</td>
<td></td>
</tr>
<tr>
<td>Eastern Algonquian</td>
<td>Micmac</td>
<td>no'yōmi'te my GM, no'yōmi'te my stepM, nik'te my M</td>
<td>Speck 1918</td>
</tr>
<tr>
<td>Malecite</td>
<td>no'komas my GM, no'komas my stepM, no'kom my FZ, my MBW, ni'gawus my M</td>
<td>Speck 1918</td>
<td></td>
</tr>
<tr>
<td>Passamaquoddy</td>
<td>no'komas my GM, no'komas my stepM, no'kom my FZ, my MBW, ni'gawus my M</td>
<td>Speck 1918</td>
<td></td>
</tr>
</tbody>
</table>
### Stocks, families

<table>
<thead>
<tr>
<th>Western Abenaki</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(St Francis)</td>
<td>no'kômes my GM, my A, no'kôm ad. FZ, nokôma's my GM, my FBW, my MZ, nok'ôma's my stepM, nok'o'm ~ nuku'm my FZ, my MBW, nigawas my M</td>
<td>no'kômas my GM, my stepM, no'kôm my FZ, my MBW, n'ga'was my M</td>
<td>Rasles 1691, Speck 1918</td>
</tr>
</tbody>
</table>

### Eastern Abenaki

<table>
<thead>
<tr>
<th>Eastern Abenaki</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Penobscot)</td>
<td></td>
<td></td>
<td>Speck 1918</td>
</tr>
</tbody>
</table>

### Wampanoag (Natick)

<table>
<thead>
<tr>
<th>Wampanoag (Natick)</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
</table>

### Mohegan (and Pequot)

<table>
<thead>
<tr>
<th>Mohegan (and Pequot)</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
</table>

### Mahican (Morgan: Mohegan)

<table>
<thead>
<tr>
<th>Mahican (Morgan: Mohegan)</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Munsee</td>
<td>nano'mome my GM,</td>
<td>n'gahase my M</td>
<td>Morgan 1871</td>
</tr>
</tbody>
</table>

### Delaware

<table>
<thead>
<tr>
<th>Delaware</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
</table>

### Wakashan: 2,781 BP

<table>
<thead>
<tr>
<th>Wakashan: 2,781 BP</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gágas your GF</td>
<td>gágas your GM</td>
<td>ASJP</td>
</tr>
<tr>
<td></td>
<td>k'a'oc GGP(S),</td>
<td>k'a'ock GGP(D)</td>
<td>Grubb 1977</td>
</tr>
<tr>
<td></td>
<td>ña'oc'cu middle B</td>
<td></td>
<td>Fleisher 1984</td>
</tr>
</tbody>
</table>

### Salishan: 3,827 BP

<table>
<thead>
<tr>
<th>Salishan: 3,827 BP</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*q'd- eB, *s-xa'xa EF, DH, SW</td>
<td>*k'tc eZ</td>
<td>ASJP, R: Kinkade 1992, Morgan 1980</td>
</tr>
</tbody>
</table>

### Coast Salish

<table>
<thead>
<tr>
<th>Coast Salish</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squamish</td>
<td>skàk' yB, kò'kpî GF</td>
<td>skàk' yZ</td>
<td>Boas 1890</td>
</tr>
<tr>
<td>Bilqula</td>
<td></td>
<td></td>
<td>Boas 1890</td>
</tr>
<tr>
<td>Twana (Skokomish)</td>
<td>sxaxà'z'' EF, DH</td>
<td></td>
<td>Elmendorf 1946, Morgan 1980</td>
</tr>
</tbody>
</table>

### Snohomish

<table>
<thead>
<tr>
<th>Snohomish</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ska eB</td>
<td></td>
<td>Haeberlin &amp; Gunther 1942</td>
</tr>
</tbody>
</table>

### Puget Sound Salish (Nisqualli)

<table>
<thead>
<tr>
<th>Puget Sound Salish (Nisqualli)</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Chehalis</td>
<td>k'uí- F</td>
<td>k'uí- M</td>
<td>Kinkade 1992</td>
</tr>
<tr>
<td>Lower Chehalis</td>
<td>g'â?l eB</td>
<td>g'â?l eZ, káy? GM</td>
<td>Kinkade 1992</td>
</tr>
</tbody>
</table>

### Shuswap

<table>
<thead>
<tr>
<th>Shuswap</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sqâ'qoa EF, k'a'tsk'a eB</td>
<td>ka'ku eZ</td>
<td>Boas 1890</td>
</tr>
</tbody>
</table>

### Spokane

<table>
<thead>
<tr>
<th>Spokane</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s-xa?-xé? EF</td>
<td></td>
<td>Morgan 1980</td>
</tr>
</tbody>
</table>

### Okanagan-Colville

<table>
<thead>
<tr>
<th>Okanagan-Colville</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stocks, families</td>
<td>Masculine denotata</td>
<td>Feminine denotata</td>
<td>References</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Flathead Salish</td>
<td>sxq 'xe' EF, qāqce' (♂)eB, qéćé (♂)eB</td>
<td>qāqe' ~ qāxa' MZ, skúk'i FZ</td>
<td>Krueger 1961</td>
</tr>
<tr>
<td>Siouan-Catawban:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,000 BP?, 6,523 BP?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*khí GM, EM</td>
<td></td>
<td></td>
<td>Campbell 1997, ASJP</td>
</tr>
<tr>
<td>Siouan: 3,000 BP, 3,169 BP</td>
<td></td>
<td></td>
<td>Campbell 1997, R: Matthews 1959</td>
</tr>
<tr>
<td>Hidatsa</td>
<td>øka (♂)eB, (♂)MB</td>
<td>kú GM, HM etc.</td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Crow</td>
<td>ike (♂)eB, (♂)MB</td>
<td>wa-kí-kúte ad. older female relative</td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Mandan</td>
<td>úka (♂)eB</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Assiniboine</td>
<td></td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Santee</td>
<td>khí EM, khí-ší GM etc.</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Teton</td>
<td>khí EM, khí-ší GM etc.</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Winnebago</td>
<td>kúq-ñík ad. GM, EM, etc.</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Iowa</td>
<td>kó-ni GM, EM etc.</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Oto</td>
<td>kó-ni GM, EM etc.</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Omaha</td>
<td>kú GM, EM, etc.</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Kansa</td>
<td>kú GM, EM, kú id. (respectful)</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Osage</td>
<td>kó GM, WM, kó id. (respectful)</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Kwapa</td>
<td>kú GM</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Biloxi</td>
<td>kákó GM, EM</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Ofo</td>
<td>kó-ñí GM</td>
<td></td>
<td>Matthews 1959</td>
</tr>
<tr>
<td>Tutelo</td>
<td>kúkí'k' ~ kokowa'ñá GF</td>
<td>kú GM</td>
<td>Matthews 1959, Speck &amp; Schaeffer 1942</td>
</tr>
</tbody>
</table>

| Catawban |                    |                   |            |
| Catawba | kokó U | tcutúi GM | Speck & Schaeffer 1942 |

| 7 Caddoan: 3,500 BP?, 4,743 BP |       |                   |            |
| Pawnee (South Band) | atí-ka my GM |                   | Taylor 1963 |
| Arikara | atí-ka? my GM |                   | Taylor 1963 |
| Caddo | i-ká? my GM |                   | Bucca & Lesser 1969 |
| Kitsai | i'kani GM |                   |            |
| Wichita | ðkw GF | ð'kw GM | Spier 1924 |

| Penutian stock: 5,522 BP? |       |                   |            |
| *kaka MB, MF, FZH |                   |                   | ASJP, R: AM&PB |

| 8 Chinookan |       |                   |            |
| Chinook | -gaga MF, qäcqäc FF |                   | Boas 1904 |
| Tfalatik | kaka A |                   | Ruhlen 1994 |

<p>| 9 Coosan |       |                   |            |
| Coos | ačaax MB |                   | Sapir &amp; Swadesh 1953 |</p>
<table>
<thead>
<tr>
<th>Stocks, families</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Plateau</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klamath-Modoc</td>
<td>ka’ê-ip MF, kakas-ip DS</td>
<td></td>
<td>De Angulo &amp; Freeland 1931</td>
</tr>
<tr>
<td>(Lutuami)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sahaptan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern-Sahaptan</td>
<td>*tâqâ? MB</td>
<td></td>
<td>R: Aoki 1966</td>
</tr>
<tr>
<td>Yakama</td>
<td>kâka MB, nakâkas my MB</td>
<td></td>
<td>Beavert &amp; Hargus 2010</td>
</tr>
<tr>
<td>Klikitat</td>
<td>ka’ka ad. MB, naka’kas my MB</td>
<td></td>
<td>Jacobs 1932</td>
</tr>
<tr>
<td>Taidnapam</td>
<td>ka’ka ad. MB, naka’kas my MB</td>
<td></td>
<td>Jacobs 1932</td>
</tr>
<tr>
<td>Palus</td>
<td>ta’xa ad. MB, nato’xas my MB</td>
<td></td>
<td>Jacobs 1932</td>
</tr>
<tr>
<td>Walla-Walla</td>
<td>ta’xa ad. MB, nato’xas my MB</td>
<td></td>
<td>Jacobs 1932</td>
</tr>
<tr>
<td>Nez Perce</td>
<td>na’râx my MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yok-Utian: 6,500 BP?, 4,413 BP?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Utian: 4,500 BP</td>
<td>*kâka MB, MBS, MBSS, etc., *kawu FZH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwokan 3,000 ~ 2,500 BP, 2,141 BP?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southern Sierra</td>
<td>kâka MB, MBS, MBSS, kawu FZH, ZH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwok (Pohonichi)</td>
<td>kâwun, kaney FZH, ZH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Sierra Miwok</td>
<td>ka’ka MB, MBS, kâ’wy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Sierra Big Creek</td>
<td>kâka MB, kawu FZH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Sierra</td>
<td>kâka MB, MBS, FZH, kawu FZH, ZH, GDH, (C)BDH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwok</td>
<td></td>
<td></td>
<td>Gifford 1922, Callaghan 1987</td>
</tr>
<tr>
<td>Plains Miwok</td>
<td>kâka MB, FZH, MZH, GMB, kakatci MBS</td>
<td></td>
<td>Gifford 1922, Callaghan 1984</td>
</tr>
<tr>
<td>Coast Miwok</td>
<td>kâka MB, MBS, gau ZH</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Bodega Miwok</td>
<td>kâkâka MB</td>
<td></td>
<td>Callaghan 1970</td>
</tr>
<tr>
<td>Lake Miwok</td>
<td>kâka MB, MBS, kauko FZH, ZH, ZHB</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Costanoan</td>
<td>howo FF, okko GM</td>
<td></td>
<td>Pinart 1878, Gifford 1922</td>
</tr>
<tr>
<td>Costanoan (Santa Cruz)</td>
<td></td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>12 Yokutsan: 1,500 BP</td>
<td>*?akas MB (Proto-Nim-Yokut)</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Paleuyami</td>
<td>kokwat MB, kohotepe S. HBS</td>
<td></td>
<td>Kroeber 1917, Gifford 1922</td>
</tr>
<tr>
<td>Yaudanchi</td>
<td>akash ~ agash MB, kawaiyis MBW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stocks, families</td>
<td>Masculine denotata</td>
<td>Feminine denotata</td>
<td>References</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Yauelmani</td>
<td>agas ~ akash ~ āagas</td>
<td>MB</td>
<td>Kroeber 1917, Gifford 2012, Sapir &amp; Swadesh 1953</td>
</tr>
<tr>
<td>Tachi</td>
<td>agas MB, MBS</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td>Gashowu</td>
<td>n-u'ho MB, MBS, MBSS</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td>Chukchansi</td>
<td>n-o'ho MB, MBS, MBSS</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td><strong>13 Wintuan: 2,500 BP</strong></td>
<td></td>
<td></td>
<td>Whisler 1980, Shepherd 2006, Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>*k'iye ~ *k'iy old man</td>
<td>(GF, MB, FZH?)</td>
<td>(GF, MB, FZH?)</td>
</tr>
<tr>
<td></td>
<td>kiye MF, MB, FZH</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>kiye MB, MeBS, FcSZ, kiye-tciher P(?) ancestor</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>kiye MB, (c?)MBS</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'iyah U?</td>
<td></td>
<td>Whistler 2010</td>
</tr>
<tr>
<td></td>
<td>k'iya(h) U (FB)</td>
<td></td>
<td>Whistler 2010</td>
</tr>
<tr>
<td></td>
<td>čiya(.k) old man</td>
<td></td>
<td>Whistler 2010</td>
</tr>
<tr>
<td></td>
<td>kaka MB, FZH</td>
<td></td>
<td>Whistler 2010</td>
</tr>
<tr>
<td></td>
<td>kaka MB, FZH</td>
<td></td>
<td>ASJP, R: AM&amp;PB</td>
</tr>
<tr>
<td>14 Maidu: 1,000 BP?, 1,219 BP?</td>
<td></td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>borrowed from Miwok?</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' FB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' FB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>k'ana' MB</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td>15 Yuki-Wappo: P-Yukian 5,000 BP?</td>
<td></td>
<td></td>
<td>Campbell 1997</td>
</tr>
<tr>
<td></td>
<td>Coast Yuki</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>Yuki</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>Huchnom</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>Wappo</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>Pailahnilian</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>Achomawi</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td></td>
<td>Atsugewi</td>
<td></td>
<td>Gifford 2012</td>
</tr>
<tr>
<td>17 Washo</td>
<td>-gu (c)DS, koi F</td>
<td></td>
<td>Kroeber 1907, 1917</td>
</tr>
<tr>
<td></td>
<td>-gu MM, (c)DD</td>
<td></td>
<td>Golla 2007, ASJP</td>
</tr>
<tr>
<td></td>
<td>Washo</td>
<td></td>
<td>Kroeber 1907, 1917</td>
</tr>
</tbody>
</table>
### MOTHER TONGUE

*Journal of the Association for the Study of Language in Prehistory* • Issue XIX • 2014

<table>
<thead>
<tr>
<th>Stocks, families</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>18 Pomoan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Pomo</td>
<td><em>kēha</em> ad. FB, stepF</td>
<td>#kats MM</td>
<td>R: AM&amp;PB</td>
</tr>
<tr>
<td>(Clear Lake)</td>
<td></td>
<td>ghais MM</td>
<td>Kroeber 1917</td>
</tr>
<tr>
<td>Southeastern Pomo</td>
<td></td>
<td>im-ka’ MM</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Northern Pomo</td>
<td><em>kēgu</em> C, FZC, HBC, eku yB</td>
<td>kaci-dai? MM</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Central Pomo</td>
<td></td>
<td>kate MM, eku yZ</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Northern Pomo</td>
<td><em>agī’</em> eB, keghawí ref. S</td>
<td>kadai ad. MM, aka ref. HM</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Southwestern Pomo</td>
<td></td>
<td>kakan MM</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Southern Pomo</td>
<td></td>
<td>akatsen MM</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Salinian</td>
<td><em>kā’i</em> eB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Antonio</td>
<td></td>
<td></td>
<td>Mason 1912</td>
</tr>
<tr>
<td>San Miguel</td>
<td><em>kāyE’</em> eB</td>
<td></td>
<td>Mason 1912</td>
</tr>
<tr>
<td><strong>19 Yuman-Cochimi</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cocopa</td>
<td><em>en’-ikwá</em> MB, in-kwo MF, axga DS, agas (5)SibGS, akas (2)SibGS</td>
<td>ini-ka MM, agas (5)SibGD, si’kas (2)SibGD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southern Diegueño</td>
<td><em>in-ikwá</em> MB, FZH, in-kwau MF, axgüau (5)SibGS</td>
<td>in-kías MM, axgüau (5)DD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Northern Diegueño</td>
<td><em>in-ikwá</em> MB, FZH, in-kwáu MF, ku’kau SinL., akatc WcZS</td>
<td>akatc WcZD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Mohave</td>
<td><em>n-akwi-k</em> MB, na-kweu-k MF, ahko’o-k (5)DC, ahkvo’-k (2)DC</td>
<td>n-akau-k MM</td>
<td>Kroeber 1917</td>
</tr>
<tr>
<td>Yuma</td>
<td><em>n-ikwá</em> MB, n-akwiau MF, axgo DC</td>
<td>n-ekyu MM</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Kamia</td>
<td><em>in-kwäu</em> MB, in-kwau MF, axgüu DS</td>
<td>in-kas MM, axgüau DD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Hualapai</td>
<td><em>ko:</em> DC, ówa SC, gwáwa MF</td>
<td>ko: DC, ówa SC</td>
<td>Watahomigie <em>et al.</em> 2001</td>
</tr>
<tr>
<td><strong>20 Chumashan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chumash (Ynezeno)</td>
<td><em>ma-q̤q̤o</em> my F</td>
<td>k-a-wa my A</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Chumash (Barbareño)</td>
<td><em>kó-ko</em> my F</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Chumash (Island)</td>
<td><em>u-ká’-ká</em> my F</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td><strong>21 Kiowa-Tanoan: 3,434 BP</strong></td>
<td></td>
<td></td>
<td>ASJP</td>
</tr>
<tr>
<td>Taos</td>
<td><em>k’a</em>- M, EM. k’aʔu M (diminutive)</td>
<td></td>
<td>Trager 1943</td>
</tr>
<tr>
<td>Hopi-Tewa</td>
<td><em>kuk’u</em> MF, kóðó’ʔe’ cZS</td>
<td>kóðó’ my Z, kóðó’ʔe’ cZD, ká’káh cZ</td>
<td>Dozier 1954</td>
</tr>
<tr>
<td>Kiowa</td>
<td>*q’o ad.(5)GF, (5)GS</td>
<td></td>
<td>Lowie 1923</td>
</tr>
<tr>
<td>Isleta</td>
<td><em>k’iʔa</em> my F</td>
<td>k’iʔa ad. FZ</td>
<td>Trager 1943</td>
</tr>
<tr>
<td><strong>22 Comecrudan</strong></td>
<td><em>kia’m</em> U</td>
<td>ke’m ~ ken A</td>
<td>Swanton 1940</td>
</tr>
</tbody>
</table>
### MOTHER TONGUE

*Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014*

<table>
<thead>
<tr>
<th>Stocks, families</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keresan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zia</td>
<td>sawá’u ($\gamma$)MB, ano’wa ($\gamma$)MB</td>
<td>sa’ko’ye ($\gamma$)Z</td>
<td>Hawley 1950</td>
</tr>
<tr>
<td>Santa Ana</td>
<td>sawá’u ($\gamma$)MB, sa’nawe ($\gamma$)MB</td>
<td>sakoi’te’ ($\gamma$)Z</td>
<td>Hawley 1950</td>
</tr>
<tr>
<td>Cochiti</td>
<td>anawa ($\gamma$)MB</td>
<td></td>
<td>Hawley 1950</td>
</tr>
<tr>
<td>San Felipe</td>
<td>s’a’namwe ($\gamma$)MB</td>
<td></td>
<td>Hawley 1950</td>
</tr>
<tr>
<td>Laguna</td>
<td>s’a’namwe ($\gamma$)MB</td>
<td>ak’wi ($\gamma$)Z</td>
<td>Hawley 1950</td>
</tr>
<tr>
<td>Acoma</td>
<td>sanawe ($\gamma$)MB</td>
<td>s’akuich ($\gamma$)Z</td>
<td>Hawley 1950</td>
</tr>
<tr>
<td><strong>Shastan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shasta</td>
<td>kwako ($\gamma$) B, akwi S</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td><strong>Muskogean</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koasati</td>
<td>laha eB</td>
<td>ki M</td>
<td>Swanton 1928</td>
</tr>
<tr>
<td>Creek</td>
<td></td>
<td>tcki M</td>
<td>Swanton 1928</td>
</tr>
<tr>
<td><strong>Uto-Aztecan:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8,900 BP?, 5,000 BP?, 4,100 BP?</td>
<td>*ka’ FF, ($\gamma$)SS, *kwa’u MF, ((\gamma))DCS, *pooka old woman</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Numic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaibab Paiute</td>
<td>guno FF, ($\gamma$)SS, gago ($\gamma$)DD, kulu FF, ($\gamma$)SS</td>
<td>*kaku MM, ((\gamma))DC, *kó eZ</td>
<td></td>
</tr>
<tr>
<td>Battle Mountain</td>
<td></td>
<td>kulu MM ($\gamma$)DD, kulu ($\gamma$)SD</td>
<td>Steward 1938</td>
</tr>
<tr>
<td>Shoshone</td>
<td>kaku ($\gamma$)DS, kunu FF, ($\gamma$)SS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elko Shoshone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bannock</td>
<td>igómu’ FF, ($\gamma$)SS</td>
<td>igómu’ ($\gamma$)SD</td>
<td>Lowie 1930</td>
</tr>
<tr>
<td>Plaviotsa</td>
<td>qómu’* FF, ($\gamma$)SS</td>
<td>qómu’* ($\gamma$)SD</td>
<td>Lowie 1930</td>
</tr>
<tr>
<td>Northeastern Mono</td>
<td>gunu’ FF, ($\gamma$)SS</td>
<td>gunu’ ($\gamma$)SD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Southeastern Mono</td>
<td>gunu FF, ($\gamma$)SS</td>
<td>gunu ($\gamma$)SD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Western Mono</td>
<td>gunu FF, ($\gamma$)SS</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>Comanche</td>
<td>ka’ku ($\gamma$)DS, konu FF, ($\gamma$)SS</td>
<td></td>
<td>Gladwin 1948, Ives 1998</td>
</tr>
<tr>
<td>Kawaiisu</td>
<td>kunu-ni FF, kag escape ($\gamma$)DS, ku:ci-ni yBS, kugu-ni FeB</td>
<td>kunu-ni MM, ($\gamma$)DD, ku:ci-ni yBD</td>
<td>Kroeber 1917, Shimkin 1941</td>
</tr>
<tr>
<td>Northern Paiute</td>
<td>kenu’u FF, ($\gamma$)SS</td>
<td>kenu’u ($\gamma$)SD</td>
<td>Kroeber 1917</td>
</tr>
<tr>
<td>Southern Paiute</td>
<td>kenu’u FF, ($\gamma$)SS</td>
<td>kaku’u ($\gamma$)SD</td>
<td>Miller 1967</td>
</tr>
<tr>
<td>Uintah Ute</td>
<td>gunu-ni FeB, kómu FF</td>
<td>kaku’u MM, kenu’u ($\gamma$)SD</td>
<td>Kroeber 1917, Swanton 1913, Shimkin 1941</td>
</tr>
</tbody>
</table>

| Tubatulabal         | aka FF, kunu FeB           | kurič eZ                   | Kroeber 1917, Shimkin 1941 |

---

43 Nichols, who is one of the authors of Hage (2004), has reconstructed the Proto-Numic kinship terms in 2002: *Notes on Proto-Numic kinship reconstructions* (unpubl. ms.).
<table>
<thead>
<tr>
<th>Stocks, families</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Takic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kitanemuk</td>
<td>kūkin FF, kwadi MF, (♂)DS, kwam FeB</td>
<td>kūkin FM, kwadi (♂)DD, kor eZ</td>
<td>Gifford 1922, Shimkin 1941</td>
</tr>
<tr>
<td></td>
<td>kukuri FB, MF</td>
<td>koči eZ</td>
<td>Shimkin 1941, Miller 1967</td>
</tr>
<tr>
<td><strong>Serrano</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kūki? GGF, -ka' FF, SS, k'at MF, (♂)DS, kwa' WF, kumu FeB, aka (♂)eBS</td>
<td>kūki? GGM, -ka' FM, SD, k'at (♂)DD, kakaiyek (♀)PinL (after ego bears a child), kakaiye DinL. (after she bears a child), kēr eZ, aka (♂)eBD</td>
<td>Gifford 1922, Shimkin 1941, Miller 1967</td>
</tr>
<tr>
<td><strong>Luiseño</strong></td>
<td>ka‘ FF, (♀)EF, ka‘-mai SS, kwa’ MF, ka‘-mai (♂)DS, kwa-pa-na HF, (♀)DH</td>
<td>ka‘ FM, (♂)DL, ka‘-mai SD, (♀)ZSD, kwa-itaï (♂)DD</td>
<td>Kroemer 1917, Gifford 1922, Miller 1967</td>
</tr>
<tr>
<td><strong>Cahuilla</strong></td>
<td>-ka‘ FF, (♀)SS, kum FeB</td>
<td>-ka‘ FM, (♀)SD, nē-qa? my GM, kis eZ</td>
<td>Shimkin 1941, Miller 1967</td>
</tr>
<tr>
<td><strong>Desert Cahuilla</strong></td>
<td>ka‘ FF, kala SS, kwa’ MF, kwala (♂)DS, kux GMB, kuxhum (♀)BGS, (♀)ZGS, kum FeB</td>
<td>ka‘ FM, kala SD, kwala (♂)DD, kux GFZ, kuxhum (♀)BGD, (♀)ZGD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td><strong>Cupeño</strong></td>
<td>ka‘ FF, kama SS, kwa’ MF, kwana (♂)DS, kūk GMB, kūkima (♀)BGS, (♀)ZGS, kum FeB</td>
<td>ka‘ FM, kama SD, kwana (♂)DD, kūk GFZ, kūkima (♀)BGD, (♀)ZGD</td>
<td>Gifford 1922</td>
</tr>
<tr>
<td><strong>Hopi</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hopi</td>
<td>k‘ā?áa MF, FF (♂)DS</td>
<td>k‘ā?áa (♂)DD, k‘ā?áa FZ, ɨeqeqa‘ eZ</td>
<td>Shimkin 1941, Miller 1967</td>
</tr>
<tr>
<td><strong>Pimic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pima</td>
<td>ka‘k (♀)SS, kēle FeB</td>
<td>ka‘k FM (♀)SD, ka‘kāks FMZ, hu?ul MM, ḍákāks FeZ</td>
<td>Parsons 1928, Shimkin 1941</td>
</tr>
<tr>
<td>Papago</td>
<td>ka‘amad (♀)SS</td>
<td>ka‘amad - ka‘a FM, ka‘amad (♀)SD</td>
<td>Shimkin 1941, Miller 1967</td>
</tr>
<tr>
<td>Tepecano</td>
<td>kaka-ri (♀)SS</td>
<td>kaka-ri FM, (♀)SD</td>
<td>Shimkin 1941</td>
</tr>
<tr>
<td>Tepehuan</td>
<td>boscica [boski-ka] FF</td>
<td>kasulī FM, (♀)SD, komata id. mat-kasulī MinL</td>
<td>Radin 1931, Shimkin 1941</td>
</tr>
<tr>
<td>Tepehua - Santa Maria Ocotán (Southeastern)</td>
<td>kuxwi’ MeB, (♀)eZS, kuxwi’ (♀)eZD, kuulsi kudlî GF, (♀)GS, o’kix (♀)GD, o’kix (♀)yZD, MeZ (♀)yZS</td>
<td>Willett 2006-2010</td>
<td></td>
</tr>
<tr>
<td><strong>Taracahític</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarahumara</td>
<td>ačikari FF</td>
<td>akácuri FM, cochi [koci] eZ</td>
<td>Radin 1931, Shimkin 1941</td>
</tr>
<tr>
<td>Opata</td>
<td>kuh FeB</td>
<td>coo [koo] ~ ku eZ</td>
<td>Radin 1931, Shimkin 1941</td>
</tr>
<tr>
<td><strong>Corachol</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cora</td>
<td>-yaxu FF, (♀)GS, ne-yē- k’wari (♀) my GS</td>
<td>-yaxu (♀)GD, ne-yē-k’wari my GM, (♀)GD, -kuci eZ</td>
<td>Radin 1931, Shimkin 1941</td>
</tr>
<tr>
<td>Huichol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aztecán</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classical Nahuatli</td>
<td>coli ~ colli ~ cohcolli</td>
<td>ne-k’koci my GM</td>
<td>Miller 1967</td>
</tr>
<tr>
<td></td>
<td>[ko7ko(ōli)] GF, ačka (♂)eB</td>
<td></td>
<td>Molina 1555, 1571, Gardner n.d.</td>
</tr>
</tbody>
</table>
## Stocks, families

- **Nahuatl (Xalitla, Guerrero)**
  - Mecayapan Nahuatl, Tatahuicapan de Juárez
  - **Masculine denotata**: no-coco [no-koko] my U, my eb, my eHB, te-coco [te-koko] masculine elder, i-ćajol [kohkol?] godfather
  - **Feminine denotata**: kuhkul old man, tu kuhkul our GF

- **Tetelcingo Nahuatl**
  - i-culzi his GF

- **Cuismahuat Pipil**
  - kuhkul old man, tu kuhkul our GF

- **Comasagua Pipil**
  - kuhkul old

- **Nicarao**
  - kuhkul old man, evil spirit

### Isolates

<table>
<thead>
<tr>
<th>Number</th>
<th>Language</th>
<th>Masculine denotata</th>
<th>Feminine denotata</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>Tunica</td>
<td>ki MB, ōka S</td>
<td>ōka D</td>
<td>Swanton 1919</td>
</tr>
<tr>
<td>29</td>
<td>Atakapa</td>
<td>waxe MB, FB, hacka eB, hicun EF</td>
<td></td>
<td>Swanton 1919</td>
</tr>
<tr>
<td>30</td>
<td>Alsea</td>
<td>haʔt eB</td>
<td></td>
<td>Ruhlen 1994b</td>
</tr>
<tr>
<td>31</td>
<td>Kootenay</td>
<td>xaʔ FB, xaʔ'sa MB</td>
<td>kukt' MZ, MBW</td>
<td>Boas 1919, Morgan 1969</td>
</tr>
<tr>
<td>33</td>
<td>Haida</td>
<td>qä' MB</td>
<td></td>
<td>Mayer-Durlach 1929</td>
</tr>
<tr>
<td></td>
<td>Haida Massett</td>
<td>q’a ref. MB, ga’gai ad. id.</td>
<td></td>
<td>Murdock 1934</td>
</tr>
<tr>
<td></td>
<td>Haida Hydaburg</td>
<td>q’a ref. MB, ga’ge ad. id.</td>
<td></td>
<td>Murdock 1934</td>
</tr>
<tr>
<td></td>
<td>Haida Skidegate</td>
<td>q’a’-gə ref. MB, ha’gai ad. id.</td>
<td></td>
<td>Murdock 1934</td>
</tr>
<tr>
<td>34</td>
<td>Karok</td>
<td>hogam MB, (c)ZS, aka F, katc S</td>
<td></td>
<td>Gifford 1922</td>
</tr>
<tr>
<td>35</td>
<td>Natchez</td>
<td>gaga (c)eB</td>
<td></td>
<td>Swanton 1928</td>
</tr>
<tr>
<td>36</td>
<td>Seri</td>
<td>axaac MB (little used term)</td>
<td></td>
<td>Moser &amp; Marlett 1997</td>
</tr>
<tr>
<td>37</td>
<td>Takelma</td>
<td>xaga- MZ, MBD, xagā ad. id.</td>
<td></td>
<td>Sapir 1907</td>
</tr>
<tr>
<td>38</td>
<td>Zuni</td>
<td>kaka MB, MMZS, HMB</td>
<td>kuku FZ, FZD, FZDD, HFZ</td>
<td>Schneider &amp; Roberts 1956</td>
</tr>
<tr>
<td>39</td>
<td>Tonkawa</td>
<td>ekak ~ ekac GM</td>
<td></td>
<td>Sapir 1989</td>
</tr>
</tbody>
</table>

### 5. Synthesis of reconstructions

The following table offers a synthesis of the results obtained thus far, including some additional genetic units not developed in the present paper for reasons of space.
<table>
<thead>
<tr>
<th>Stock and Region</th>
<th>Approx. time depth</th>
<th>Term</th>
<th>Meanings</th>
<th>Source of reconstructions</th>
<th>Proposed homeland</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>North America</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-Inuit</td>
<td></td>
<td>*akkak</td>
<td>FB</td>
<td>Fortescue et al. 1994</td>
<td></td>
</tr>
<tr>
<td>Proto-Yupik</td>
<td>2,000 BP</td>
<td>*âka</td>
<td>eZ</td>
<td>EHL</td>
<td></td>
</tr>
<tr>
<td>Proto-Athapaskan</td>
<td></td>
<td>*ankay?</td>
<td>MZ</td>
<td>Hoijer 1956</td>
<td></td>
</tr>
<tr>
<td>Proto-Algic</td>
<td>7,200 BP, 4,000 BP</td>
<td>*-oko</td>
<td>GM</td>
<td>AM &amp; PB</td>
<td>Upper Yukon River (Golla 2007)</td>
</tr>
<tr>
<td>Proto-Algonquian</td>
<td>3,000 BP</td>
<td>*-okko-</td>
<td>GM, (EM)</td>
<td>Sapir 1922</td>
<td>Columbia Plateau (Golla 2007), NY State – CT (Wichmann 2010)</td>
</tr>
<tr>
<td>Proto-Salishan</td>
<td>3,800 BP</td>
<td>*gâ-</td>
<td>eB</td>
<td>Kinkade 1992</td>
<td>Between Skagit &amp; Fraser Rivers (Kinkade 1991)</td>
</tr>
<tr>
<td>Proto-Salishan</td>
<td>3,800 BP</td>
<td>*xaxa</td>
<td>EF, DH, SW</td>
<td>Morgan 1980</td>
<td></td>
</tr>
<tr>
<td>Proto-Siouan</td>
<td>3,000 BP</td>
<td>*khy</td>
<td>GM, EM</td>
<td>Matthews 1959</td>
<td></td>
</tr>
<tr>
<td>Proto-Caddoan</td>
<td>3,500 BP</td>
<td>*ka(?)</td>
<td>GM</td>
<td>Taylor 1963</td>
<td></td>
</tr>
<tr>
<td>Proto-Penutian</td>
<td>5,500 BP</td>
<td>*kaka</td>
<td>MB, FZH</td>
<td>AM &amp; PB</td>
<td>North Great Basin, Columbia Pl. (Golla 2007)</td>
</tr>
<tr>
<td>Proto-Sahaptian</td>
<td></td>
<td>*tâqa?</td>
<td>MB</td>
<td>Aoki 1966</td>
<td>North of San Francisco Bay (Callaghan 1997)</td>
</tr>
<tr>
<td>Proto-Miwokan</td>
<td>2,000 ~ 3,000 BP</td>
<td>*kaka</td>
<td>MB, MBS</td>
<td>Callaghan 1997</td>
<td></td>
</tr>
<tr>
<td>Proto-Nim-Yokuts</td>
<td>1,500 BP</td>
<td>*ta-kas</td>
<td>MB</td>
<td>Callaghan 2001</td>
<td>Pre-Proto-Yokuts in Great Basin? (Golla 2007)</td>
</tr>
<tr>
<td>Proto-Maiduian?</td>
<td>1,000 BP</td>
<td>*kaka</td>
<td>MB, FZH</td>
<td>AM &amp; PB</td>
<td>Pre-Proto-Wintuan: Northern Great Basin (Golla 2007)</td>
</tr>
<tr>
<td><strong>Proto-Yukian-Wappo</strong></td>
<td>5,000 BP</td>
<td>*keka</td>
<td>MB, FZH</td>
<td>AM &amp; PB</td>
<td>Eel River drainage, NW California (Foster 1996: 83)</td>
</tr>
<tr>
<td><strong>Meso-America</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-Totonac-Tepehua</td>
<td>2,400 BP</td>
<td>*koko ~ kuku</td>
<td>U (MB)</td>
<td>AM &amp; PB</td>
<td>Builders of Teotihuacan (AD 200-650; Campbell 1997)</td>
</tr>
<tr>
<td>Proto-Jicaque (Tol)</td>
<td>*kok'cam</td>
<td>U</td>
<td>(c)GF</td>
<td>Campbell &amp; Oltrogge 1980</td>
<td>Spoken by the Olmecs (Campbell 1997), Tuxtla Mountains?</td>
</tr>
<tr>
<td>Proto-Mixe-Zoquean</td>
<td>3,000 BP</td>
<td>*oko</td>
<td>GM, GC</td>
<td>Wichmann 1999</td>
<td></td>
</tr>
<tr>
<td>Proto-Mayan</td>
<td>4,000 BP</td>
<td>*ikaan</td>
<td>(c)MB, (c)FZH, (c)MGF, WF cousin</td>
<td>Wichmann &amp; Brown n.d., Kaufman &amp; Justeson 2003</td>
<td>Cuchumatanes Mountains, Guatemala (Campbell 1997)</td>
</tr>
<tr>
<td>Proto-Uto-Aztecan</td>
<td>8,900 BP ~ 5,000 BP</td>
<td>*ka'a'</td>
<td>FF, FM, (c)SC</td>
<td>Miller 1967, Voegelin.</td>
<td>Great Basin USA (Merrill et al. 2009), Arizona – northern Mexico (Fowler)</td>
</tr>
<tr>
<td>Stock and Region</td>
<td>Approx. time depth</td>
<td>Term</td>
<td>Meanings</td>
<td>Source of reconstructions</td>
<td>Proposed homeland</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------</td>
<td>------</td>
<td>----------</td>
<td>--------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Proto-Otomanguean</td>
<td>6,500 BP–4,000 BP</td>
<td>*hkeh</td>
<td>GF, GS male affinal kin</td>
<td>Voegelin &amp; Hale 1962</td>
<td>1983, Meso-America (Hill 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*kha</td>
<td></td>
<td>Merrifield 1981</td>
<td>Tehuacan Valley (Hopkins 1984, Campbell 1997)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*kuHn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*kwalHn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-Chibchan</td>
<td>6,700 BP</td>
<td>*gaka</td>
<td>F</td>
<td>Constena Umana 1981</td>
<td>South Central America (Constenla Umana 2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*gaka</td>
<td>EM</td>
<td>AM &amp; PB</td>
<td>Northern Colombia?</td>
</tr>
<tr>
<td>Proto-Magdalenic</td>
<td>5,200 BP</td>
<td>*kaca</td>
<td>GM</td>
<td>AM &amp; PB</td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-Cariban</td>
<td>3,700 BP</td>
<td>*koko</td>
<td>U</td>
<td>De Goeje 1946</td>
<td>Venezulan Guiana (Villalón 1991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>idem</td>
</tr>
<tr>
<td>Proto-Cariban</td>
<td>3,700 BP</td>
<td>*kuku</td>
<td>GM, FZ, EM</td>
<td>AM &amp; PB</td>
<td>Central &amp; Southern Surinam (Meira 1998)</td>
</tr>
<tr>
<td>(Cariban)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-Pano-Takanan</td>
<td>4,700 BP</td>
<td>*kaka~</td>
<td>MB, EF</td>
<td>Girard 1971</td>
<td>East Central Peru (Fleck 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*kaka</td>
<td></td>
<td></td>
<td>idem</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-Takanan</td>
<td>2,000 BP?</td>
<td>*kuku</td>
<td>U (MB)</td>
<td>Girard 1971</td>
<td>Northern Bolivia?</td>
</tr>
<tr>
<td>Proto-Arawakan</td>
<td>4,500 BP</td>
<td>*kuko</td>
<td>U, EF</td>
<td>Payne 1991</td>
<td>Middle Orinoco – Upper Amazon (Heckenberger 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*aku-</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-Arawan</td>
<td>1,700 BP</td>
<td>*koko</td>
<td>ad. MB, EF</td>
<td>Dixon 2004b</td>
<td>Middle Juruá River?</td>
</tr>
<tr>
<td>Proto-Guahiboan</td>
<td>2,300 BP</td>
<td>*axu</td>
<td>MB, EF</td>
<td>AM &amp; PB</td>
<td>Northern Colombia (Wichmann n.d.)</td>
</tr>
<tr>
<td>Proto-Quechuan</td>
<td>&gt; 1,500 BP</td>
<td>*akwe</td>
<td>MB, EF</td>
<td>AM &amp; PB</td>
<td>Central Peru (Adelaar 2012)</td>
</tr>
<tr>
<td>Proto-Aymaran?</td>
<td>&gt; 1,500 BP</td>
<td>*kaka-</td>
<td>MB, EF</td>
<td>AM &amp; PB</td>
<td>Central Peru (Adelaar 2012)</td>
</tr>
<tr>
<td>Proto-Nambikuaran</td>
<td>2,800 BP</td>
<td>*kaka-</td>
<td>GF, MB, EF</td>
<td>AM &amp; PB</td>
<td>Upper Juruena River (Wichmann n.d.)</td>
</tr>
<tr>
<td>Proto-Karirian</td>
<td>&gt; 300 BP</td>
<td>*kuk(h)</td>
<td>MB</td>
<td>AM &amp; PB</td>
<td>Lower Rio São Francisco - State of Parába (Brazil)?</td>
</tr>
<tr>
<td>Proto-Mondé</td>
<td>2,000 BP</td>
<td>*koko</td>
<td>ad. MB, EF</td>
<td>AM &amp; PB</td>
<td>Roosevelt &amp; Aracuana Rivers (Bruneau 1987), Jiparan River headwaters (Campbell 1997)</td>
</tr>
<tr>
<td>Proto-Kaingang</td>
<td>?</td>
<td>*kakra</td>
<td>MB, EF</td>
<td>Jolkesky 2010</td>
<td>Parana State (Brazil)?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*kake</td>
<td>eB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 13. **KOKO** forms in various proto-languages of America. Proto-forms preceded by an asterisk * have been reconstructed by specialists of the concerned group, as indicated in column Sources; proto-forms preceded by a hash mark # are not reconstructions, but are postulated by us (AM & PB) according to the generalized presence, in all or nearly all member languages of the concerned group, of forms.
The time depths reported in our paper have been calculated either using the method of glottochronology or the automated dating method based on lexical similarity used by ASJP. Although the time depths obtained by these methods differ somehow, all of them are to be counted in thousands of years. By coupling time depths of language families with reconstructed KOKO terms, table 13 makes apparent that KOKO terms were in use several millennia ago by the ancestor languages of many “major” linguistic families (major in terms of the number of languages they include), whose postulated homelands were, most of the time, far away from each other (see maps 1, 2, 3). Among the most ancient stocks or families, let us mention Lenca-Misumalpan (7,200 BP), Chibchan (6,700 BP ~ 4,484 BP), Algic (> 3,000 BP), Arawakan (Maipuran) (4,500 BP ~ 4,134 BP), Pano-Takanan (4,700 BP), Cariban (3,700 BP ~ 2,362 BP), Uto-Aztecan (8,900 BP ~ 5,000 BP ~ 4,118 BP), Yok-Utian (5,000 BP ~ 4,413 BP), Caddoan (3,500 BP ~ 4,828 BP), Siouan (3,000 BP ~ 3,169 BP), Oto-Manguean (6,500 BP ~ 7,418 BP). A KOKO form has not been posited at the proto-level of some other major stocks like Jéan (4,989 BP), Tupían (5,500 BP ~ 3,585 BP), Tukanan (2,699 BP), Siouan-Catawban (4,000 BP ~ 6,856 BP), although some families or branches belonging to these groups do display KOKO terms, sometimes even at the level of the proto-language, like Mondé (2,000 BP) for the Tupían family, Kaingang (no date) for the Jéan family, Siouan (2,500 BP ~ 3,169 BP) and Catawban for the Siouan-Catawban stock. KOKO forms have been reconstructed for “younger” families like Quechuan (1,500 BP ~ 1,717 BP) or Arawan (1,764 BP), and are also postulated in some of the “smaller” families like Nambikuaran (2,807 BP), or Karirian. Finally, as we already mentioned, they are also present in 60% of the language isolates for which there are substantial data (maps 1 and 2: South and Meso-America).

At this stage of our study, we will just give a general idea of the semantic scope of the form KOKO at the linguistic family level, drawn from table 13. A more fine-grained analysis of the semantic scope will be performed in the anthropological part of our paper. For now, the meanings MB/EF, MB/FZH, MB/MBS are rallied in the category [MB], so much for the meanings GM/EM and GM/FZ which are rallied in the category [GM]. This results in the figures given in Table 14.

6. Discussion: the origin of KOKO in the Americas

A matter of convergence?
Is it possible that the continental distribution of KOKO terms should be the result of chance or of some convergence global process? We will refer the reader to some of our previous publications and notably Bancel & Matthey (2002) in which this question was thoroughly treated at the world level for the etymon KAKA. We will also refer them to Matthey & Bancel (2008).
Matthey, Bancel & Ruhlen (2011), Bancel & Matthey (2013), showing that contrary to Trask’s (2004) allegations, nursery kinship terms are not perpetually recreated or reinvented in the world’s languages, as former nursery terms get linguistically eroded, but are deeply rooted in linguistic families, and transmitted faithfully from one generation to another, through millennia, with very few phonetic transformations. As we stressed in Bancel & Matthey (2002), and the following publications, the daily use of this term as well as the use of other reduplicated terms like PAPA and MAMA, generally called “nursery terms”, the ease of their transmission to young children, their high symbolic significance, have made them extremely resistant to phonetic and semantic change, as is otherwise fully demonstrated by their continuous written transcriptions in the course of the past 5,000 years or so (Matthey de l’Etang & Bancel 2008, Bancel & Matthey de l’Etang 2013). Likewise, KAKA terms, just like their PAPA, MAMA and TATA counterparts, have also left traces in the written records of Indo-Hittite languages enabling the reconstruction of Proto-Indo-Hittite *HawH-os (*xawx-os) MB, GF (Nikolayev 2007), likewise in the written records of Chinese since Preclassic Old Chinese gu? MB some 3,000 years ago (Starostin 2005). But even with no ancient written records left behind, there is little doubt that *ka(a)ka GP in Niger-Congo is extremely ancient if not the proto-form in this language phylum. This very same form *-kaakd GP has been reconstructed in Proto-Bantu by both Meeussen (1969) and Guthrie (1967-1971).

Let us add that, in our opinion, the hypothesis that the transcontinental distribution of the same forms and meanings, let us say KOKO MB, in the languages of the Americas results from sheer coincidence appears utterly improbable. Unless some underlying principle attaching CVCV reduplicated forms, with velar consonant, in occurrence KOKO, to the MB or GM relationships can be demonstrated, the convergence theory would predict forms with more erratic meanings, than just those, found in all regions, and for the most part consistent with Dravidian-type terminology.

Diffusion and borrowing

The second hypothesis that must be debated is the possibility that the general distribution of KOKO terms all over the Americas results from an ancient but still ongoing diffusion-process, making this region as a whole look like a linguistic area.

Upholders of such an hypothesis must explain, and provide some kind of evidence as to how cascade borrowing accounts for the phonetic and semantic parallelisms of the American series KOKO, the huge accumulated evidence of its transcontinental distribution, and the apparent linguistic retention of KOKO forms, as shown in their reconstructions in linguistic families, whose supposed time depths are sometimes as remote as 7,000 BP.

It is worth mentioning at this point that our survey of kin terms diffusion among intermarrying language groups, including those practicing linguistic exogamy, has not provided support for a model of mass diffusion of kinship terms, as the one envisioned by Dixon & Aikhenvald (1999: 8; see Presentation), notably concerning KOKO forms (Matthey de l’Etang & Bancel in preparation). Our survey of the ethnological literature includes the Arawakan-Tukano-Maku cluster in the Vaupés basin, the Arawakan-Tupi-Cariban-Trumai cluster of the Upper Xingu

44 “Linguistic exogamy” refers to marriage prescribed between groups speaking different languages, and implies that one spouse will join the residence of the other (see details in notes 45 and 46).

45 In the Vaupés region (north-east Brazil), since residence is patrilocal, a woman has to leave her paternal group and join her husband’s. Social identity is “established by patrilineal descent and has language group affiliation as its primary marker;” identification is done with “one’s father’s language group” (Stenzel 2005: 3-4). As Sorensen (1967: 677) puts it, “an individual belongs to his (or her) father’s tribe, and to his father linguistic group, which is also his own.”
region, the Arawakan-Cariban cluster of the Western Indies, as well as the relationships between Arawakan and Takanan in Bolivia, between Arawakan and Panoan in Eastern Central Peru, between Quechuan and Aymaran in Peru, and between Arawakan and Quechuan in Peru. The explanation given by these authors does not match the facts correctly, notably at the semantic level, as KOKO terms do substantially refer to kin types different from just EF/MB, and particularly to the GM. This casts a serious doubt on the reality of the borrowing mechanism hypothesized by Dixon & Aikhenvald.

This is not to say that a number of divergent phonetic forms of KOKO within a language family series do not result from the borrowing of forms from other families. But most of the documented cases of linguistic and ethnic contact show that kinship terms are seldom borrowed by one group from another, because language appears determinant in keeping the group’s identity.46

But our key contention regarding the theory of mass diffusion does not basically rest on an empirical demonstration, but on a theoretically compelling argument, recently developed by Martin Haspelmath facing the problem of knowing whether Australia as a whole could actually be defined as a linguistic area. According to Haspelmath (2004: 211), “linguistic areas need not only be internally coherent, but also distinctive with respect to languages outside the area. Thus, one would have to show that the Australianisms are uncommon in the rest of the world, or at least in adjacent areas.” Transposed in the particular case of the American KOKO ~ KAKA MB, GF, FZ, GM, eSib, the second criterion is not met, because this word is in fact one of the most common kinship terms in the world, and is found in languages of New Guinea, Australia, Africa, Eurasia, Oceania, with similar meanings, and consequently does not represent a distinctive feature with respect to the languages outside the Americas.

We can then assume that the pan-American distribution of KOKO ~ KAKA MB, GF, FZ, GM, eSib does not result from the borrowing or diffusion of an American areal feature, but looks consistent with its inheritance from kinship terms present in the language or languages spoken by the group or groups which first colonized the Americas. This better explains why, as far back in time as comparative linguistics can point out, a number of language groups in their ancestral stage, as well as some ancestors of the languages isolates spoken across the Americas in areas very distant from one another, as maps 1, 2 and 3 above do show, had KOKO forms in their kinship lexicon. Thus, if we may venture a paraphrase of a famous comment made by Sapir about the distribution of first person root n- across the Americas: “How in the hell are we gonna explain the general American KOKO except genetically?”47

KOKO and the peopling of the Americas

Of course, the central question that one will ask at this point is: how do the massive geographic and linguistic distribution of KOKO kinship terms, and the fact that they have been preserved through millennia and transmitted with very little or no modifications within families at a high level, fit into the history of the peopling of the Americas?

46 Stenzel (2005) speaking of the Vaupés plurilingual situation. Seki (1999) speaking of the Xingu situation, after Sorensen (1967), have stressed the role of language as a marker of identity, as one individual always associates himself with his (her) parent’s language according to the line of descent.

47 In a personal letter to Franck Speck, dated from August 1, 1918, Sapir wrote: “Getting down to brass tacks, how in the Hell are you going to explain general American n- ‘I’ except genetically? It’s disturbing, I know, but more non-committal conservatism is only dodging, after all, isn’t it? Great simplifications are in store for us.” (Quoted in Darnell and Hymes (1986: 229-230).
Following from Ruhlen (1994a & b), Bancel & Matthey de l’Etang, and Matthey de l’Etang & Bancel (2002, 2011), the massive distribution of KOKO kinship terms in America appears as a local development from the Proto-Sapiens etymon KAKA EF, MB, GP. It is the consequence of the colonization of the Americas by one or several groups, whose kinship terminologies originally comprised KOKO kinship terms. But there is certainly one thing that the distribution pattern does not tell us, in and of itself, and this is the number of language groups that were involved in the migration process. The global distribution of KAKA ~ KOKO words is consistent with both a single-migration and a multi-migration model.

One of the main contentions concerning the initial peopling of the Americas bears on the number of migration waves that occurred during the colonization process: three temporally distant waves according to the well-known and much debated conclusions of Greenberg et al. (1986)48, only one according to many of the recent genetic studies of mtDNA, Y-chromosome, autosomes in American Native populations. These two conclusions may appear contradictory, but they are not, as we shall see.

It is noteworthy that the three language families defining Greenberg’s migration waves, i.e. Amerind, Na-Dene and Eskimo-Aleut, all display KOKO terms to a certain extent, as is indicated in tables 12 and 13, a fact apparently in line with the observations made in the preceding paragraph. Thus, *aka eZ, M has been reconstructed in Proto-Eskimo (EHL), *akkak FB has been reconstructed in Proto-Inuit (Fortescue et al. 1994, EHL)49, while no proto-KOKO form has been reconstructed for either Proto-Athapaskan50, or Nuclear Na-Dené (Eyak-Tlingit-Athapaskan), even if Eyak and Tlingit do display such forms. Let us finally mention that Haida, which is considered a Na-Dené language by Greenberg and Ruhlen, does have q’a ‘ga ‘my MB’. All the other KOKO forms have been attributed to the Amerind macrophyllum by Ruhlen.

➢ The genetic studies of Native Americans

1. A major migration wave along the Pacific coast ca. 16,000 BP

The most recent genetic studies of Native Americans have stressed the fact that virtually all Native American populations trace their ancestry to a limited number of founder mtDNA haplogroups, A2, B2, C1, D1, and D451, with similar coalescence times, and to the Y-chromosome haplogroups Q1a3a, Q1a3*, C3. These studies consequently make the assumption that only one population wave of Asiatic origin52 was responsible for the initial peopling of the Americas, and

48 Zegura et al. (2004: 164) write: “In 1986, Greenberg, Turner & Zegura published a widely cited, synthetic, position paper on the early peopling of the Americas that stressed the apparent congruence of the then available data from linguistics, dental morphology, and traditional biparental nuclear genetic systems within the context of the archaeological record. Their major explanatory hypothesis, the ‘three-wave’ or ‘tripartite’ model, was based on the proposition that all indigenous Native American populations could be allocated to three distinct linguistically defined groups (i.e., Amerind, Na-Dené, and Aleut-Eskimo) that had their origins in three chronologically separate migrations from different geographic areas of Asia (Greenberg, Turner & Zegura 1986).”

49 Aleut refers to the GM as kukaq.

50 Except perhaps ankay? MZ in Proto-Athapaskan, reconstructed by Hoijer in 1956.

51 Perego et al. 2009 identified the mtDNA D4h3 haplogroup in California, Mexico, Peru and Chile and suggest, based on its age estimate of ca. 16,000 BP, that it entered the Americas with the major wave along the Pacific coast.

52 The sequencing of the genome of a ±24,000-year-old anatomically modern human individual from Mal’ta in south-central Siberia by Raghavan et al. (2014) revealed mtDNA and Y-chromosome
for spreading the pan-American haplogroups into the continent (Tarazona-Santos & Santos 2002, Zegura et al. 2004, Tamm et al. 2007, Fagundes et al. 2008, Achilli et al. 2008, Kumar et al. 2011). In line with these conclusions, it has been shown that the very same genes were present in archaic human remains dated from 13,000 to 4,000 years BP, all over the Americas (Smith et al. 2005, Kemp et al. 2007, Manriquez et al. 2011, Chatters et al. 2014). Finally, Schroeder et al. (2007, 2009) have taken the high frequency of a private allele, “the 9-repeat allele at microsatellite D9S1120 in all sampled Native American and Western Beringian populations... as evidence that all modern Native Americans descend primarily from a single founding population” (2009: 995).

This (major) migration wave, generally dated back to ca. 18,000-14,000 BP – thus predating the Clovis culture (Zegura et al. 2004, Tamm et al. 2007, Fagundes et al. 2008, Reich et al. 2012, Bodner et al. 2012). These dates are in accordance with those given in recent archeological studies of Paleo-Indian sites (Goodyear 2005). Besides, Wang et al. (2007: 2049), studying the genetic diversity and population structure in the Americas, observed “gradients both of decreasing genetic diversity as a function of geographic distance from the Bering Straits and of decreasing similarity to Siberians – signals of the southward dispersal of populations from the northwestern tip of the Americas.” They concluded (2007: 2059) that this “genomic continent-wide pattern” is consistent with a model in which, “at each step in the migration, a subset of the population splitting off from a parental group moves deeper into the Americas, taking with it a subset of the genetic variation present in the parental population.”

2. The presence of other lineages

Additionally, a few other minor lineages such as the mtDNA haplogroups X2a, D2, D3, C4c, or the Y-chromosome haplogroups Q1a5 and Q1a6 have been reported in Native American populations (Tamm et al. 2007: 4, Perego et al. 2009, Kumar et al. 2011, Dulik et al. 2012), but they appear to be restricted to North America. The presence of these lineages raises the question

\[ \text{haplogroups, U and R respectively. MtDNA haplogroup U has been found at high frequency among Upper Palaeolithic and Mesolithic European hunters-gatherers, while the Y chromosome haplogroup R is basal to modern-day western Eurasians. The same study revealed autosomal evidence that the Mal'ta individual was basal to modern-day Eurasians, and also genetically closely related to modern-day Native Americans, with no close affinity to East Asians: 14% to 38% of Native American ancestry may originate through gene flow from this prehistoric population. What Raghavan et al. suggest is that this gene flow occurred “after the divergence of Native American ancestors from east Asian ancestors, but before the diversification of Native American populations in the New World.” One of the implications of the study, according to the authors, is that the results “may provide an explanation for the presence of mtDNA haplogroup X in Native Americans, which is related to Western Eurasians but not found in East Asian populations.” Another implication that the study provides is “a possibility that the non-East Asian cranial characteristics of the First Americans derived from the Old World via migration through Beringia, rather than by a trans-Atlantic voyage from Iberia as proposed by the Solutrean hypothesis” (Raghavan et al. 2014: 89, and see note 55).} \]

53 Less than 2,000 years for the entire Pacific coast, according to Bodner (2012: 6).
54 Dulik et al. (2012: 2) report that one Tłı̨chǫ, one Slave and possibly one Alaskan Athapaskan belong to Q1a5.
55 These clades have only been identified in North America. The haplogroup X (X2a) is found in North America at a low frequency. The fact that its coalescence time appears younger than those of the American haplogroups A-D has made some researchers argue that haplogroup X “represents an
as to whether they represent minor founding haplogroups participating in the same (major) population expansion or lineages belonging to more recent gene flows.

3. X2a and C4c: an inland route?

The phylogeographic analyses of the mtDNA haplogroup X2a, identified in Native populations of North America, notably in Algonquian, Wakashan, and Sahaptian speakers (Perego et al. 2009), and of the mtDNA haplogroup C4c detected in Cherokee, Creek, Siouan, Chippewa (Algonquian), Shuswap and a few other individuals from unknown ethnic origin (Kashani et al. 2012), whose coalescence ages appear similar to those of the pan-American mtDNA haplogroups, make the authors (a single research team) suggest that X2a and C4c could possibly have entered the Americas with a second major migration wave using the ice-free corridor between the Laurentide and Cordilleran ice sheets, more or less concomitantly to the first coastal migration-wave. The same research group (Achilli et al. 2013) wonders if this additional migration wave could also concern the Na-Dené speaking groups, because some of them – the southern Athapaskan, notably – do show the presence of X2a. Suggesting that X2a and C4c could actually be present in northern Na-Dené speaking groups as well, the same authors conclude (2013: 5) that “the intermediate migration highlighted by nuclear data in the Chipewyan by Reich et al. would be part of a larger-scale migratory event that did not affect only the ancestor of Modern Na-Dené” (see 7 below).

4. Genes and Na-Dené language

No one-to-one correspondence between genetic data and the Na-Dené-speaking population has been clearly demonstrated. The two mtDNA haplogroups reported at high frequencies in this group are in the first place a sublineage of the pan-American A2, i.e. A2a, and secondly a sublineage of the Siberian-Beringian D2, i.e., D2a1a, whose age is estimated back to 6,900 BP ± 4,100y (Volodko et al. 2008). Both these haplogroups are also reported in Eskimo-Aleut and Chukchi populations. Volodko et al. (2008; 1087) report that the pan-Amerindian “A2 mtDNA coding regions available from GenBank share no mutation” with the Chukchi and Na-Dené A2a lineage, except for the A2 root, and conclude (2008; 1095) that the “geographic specificity of these lineages confined to Chukotka and Alaska is the main argument in favor of the refugial hypothesis which assumes the origin of the founding populations of the Eskimo-Aleut and Na-Dené Indians in the southern Alaska at the terminal Pleistocene or early Holocene.” This conclusion was also reached by Starikovskaya et al. (2005): “The geographic specificity and phylogeny of haplogroup D complete sequences support the refugial hypothesis which proposes that the founding populations of Eskimo-Aleuts and Na-Dené Indians originated in the eastern Beringian/Alaskan refuge area during the early postglacial period.”

Besides, studies of paternally inherited Y-chromosomal DNA has long revealed that Na-Dené populations notably harbor the Y-chromosome haplogroups Q1a3 and C3 that Zegura et al. (2004) also identified in Native American and Eskimo groups, making them endorse a one-wave independent migration from Asia or even Europe” (Fagundes et al. 2008: 1, and references given for the “Solutrean” hypothesis). Perego et al. (2009) reported a coalescence time of ca. 16.7 to 15.5 ky for all the Native American clades, i.e. including X2a. D2 has been identified in Siberia and in North America (Na-Dené and Eskimo-Aleut populations), and D3 has been identified in Siberian and Eskimo populations.

56 The presence of B2 in Athapaskan-Navajo populations is likely the result of admixture with their non-Athapaskan neighbors (Torroni et al. 1992).
migration model, all the more so since both haplogroups yielded rather similar coalescence time estimates at respectively 14,700 BP ± 5,700y and 13,600 BP ± 4,100y (lower bound) (Zegura et al. 2004). But in fact, there are two C subclades which have been identified in the Americas, C3b or C-P39 having the P39 marker, and a more ancient C3* or C-217 without the P39 marker. The first haplogroup has been identified in North America mainly in Athapaskan-speaking populations, whereas the C3-M217 clade has been identified in a Tlingit-speaking individual from Southeast Alaska, in 8 Waorani and 11 Kichwa speakers from Ecuador and in 2 Wayuu-speaking individuals from Colombia (Roewer et al. 2013). The time to the most recent common ancestor (TMRCA) for C3b Y-chromosomes given by Dulik et al. (2012) is ca. 5,000 to 10,500 BP, depending on the statistical programs used, signaling, according to the same authors, a population expansion involving “mostly Athapaskan speakers” (2012: 5), different than that of the first (major) wave. Let us remark that the coalescence age estimates of C3b correspond better to the age of the Na-Dene language family, ca. 8,000 BP according to AJSP, than to the age of Proto-Athapaskan, ca. 2,500 BP (Krauss 1973).

5. Evidence for a Paleo-Eskimo migration ca. 5,500 BP

The multiwave hypothesis has otherwise received support from the genomic sequencing of a permafrost-preserved hair belonging to a Paleo-Eskimo individual found in a Saqqaq culture context, in Qeqertusussuk on the west coast of Greenland. The archeological site, excavated between 1983 and 1993, dates back to 3,900-3100 yBP. The individual was assigned to the mtDNA haplogroup D2a1, which is closely related to the common mtDNA haplogroup reported notably in present-day Aleut populations of the Commander Islands (Gilbert et al. 2008: D2a1a; Volodko et al. 2008: D2a1a1), in some Siberian Sireniki Eskimos (Gilbert et al. 2008: D2a1b; Volodko et al. 2008: D2a1a), and in a number of Inupiaq-speaking individuals from the Alaska North Slope region (Raff et al. 2011: D2)\(^6\). In another paper dedicated to the same Paleo-Eskimo, Rasmussen et al. (2010) compared the high-confidence single-nucleotide polymorphisms (SNPs) of the Saqqaq Eskimo to those of contemporary populations in order to find which were the most closely related to this individual. The conclusions were that the Saqqaq individual was more closely related to present-day Nganasans, Koryaks and Chukchis from north-eastern Siberia than to the Amerinds, the Na-Dené speakers, or the Greenland Inuits. Thus the Saqqaq population apparently shared ancestry with Arctic East Asians and not with the people from the first migration wave. The genetic proximity between the Koryaks and the Saqqaq individual is also apparent in the assignment to the Y-chromosome haplogroup (or paragroup) Q1a*\(^6\) of both the Saqqaq individual and the four Koryaks inhabiting the Sea of Okhotsk coast (Malyarchuk et al. 2011). Bisso-Machado et al. (2011), for their part, assigned one Siberian Yupik Eskimo to the same paragroup Q1a*.

Gilbert et al. (2008) and Rasmussen et al. (2010) concluded that these remains (if representative of a population) constitute evidence for a population expansion along the northern fringes of the American continent, different from that of the first wave, and also different from the Thule expansion that occurred ca. 1,000 years ago. This population wave is dated by Gilbert back to ca. 4,500 BP, and by Rasmussen et al. (2010: 757) to ca. 5,500 BP, based on the estimated

\(^{57}\) Also at moderate frequencies in Siouan and Cheyenne populations (Zegura et al. 2004).

\(^{58}\) Dulik et al. (2012) reactivate a model supported by Lell et al. in 2002, and Bortolini et al. in 2003, but challenged by Tarazona-Santos & Santos (2002) and Zegura et al. (2004).

\(^{59}\) See also Zlojutro 2006-2008, Crawford et al. 2010.

\(^{60}\) The word paragroup refers to a lineage of a haplogroup not defined by any specific additional marker, and written with an asterisk *, here Q1a*. 

136
MOTHER TONGUE

Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

mtDNA genetic divergence date between the Chukchis and the Saqqaq individual of between 4,400 and 6,400 years BP, coupled to the oldest archeological evidence of the Arctic Small Tool tradition dating back to 5,500 BP (Rasmussen et al. 2010: 761). Dulik et al. (2012) argued that this population expansion and that involving Athapaskan-speaking populations which was mentioned two paragraphs above are basically concomitant.\(^6^1\)

Dulik et al. (2012) proposed to equate the Y-chromosome Q1a\(^*\) of the Paleo-Eskimo man to the haplogroup Q1a6, which is defined by the NWT01 marker and whose coalescence age between 7,000 and 5,000 BP\(^6^2\) is consistent with the age of the oldest evidence of the Arctic Small Tool tradition ca. 5,500 BP. Furthermore, given the fact that Q1a6 has been reported in Canadian Inuits (Inuvialuit), in Northern Alaskan Inupiaq-speaking communities, in some Yupik populations, and also inferred in the four Koryaks previously assigned to Q1a\(^*\), Dulik et al. (2012) assume a continuity between the Paleo-Eskimo individual and modern Inuits and Yupiks.

6. The mtDNA haplogroup D3

Finally, the mt-DNA haplogroup D3 is present in a number of Siberian populations such as the Nganasans, Yukaghirs, Chuvantsis and Chukchis, as well as in Siberian Naukan Eskimos (Volodko et al. 2008), some Aleut individuals (Crawford et al. 2010, Zlojutro 2006-2008) and in Alaskan, Canadian and Greenlandic Inuits (Helgason et al. 2006, Raff et al. 2011). This haplogroup is generally associated with the Thule expansion, dated close to ca. 1,000 AD (Gilbert et al. 2008).

7. New perspectives

Before we put a provisional full-stop to this section, let us mention that a multiple-wave scenario is also supported in a highly comprehensive survey of genetic diversity in Native Americans, carried out at the highest resolution level by Reich et al. (2012), showing that the Saqqaq individual, the Aleutian peoples, and the East and West Greenland Inuits derive 57% of their genetic ancestry from admixture with populations descending from the first migration wave, and 43% from specific Asian lineages, marking a distinct migration from Asia. The same study shows that the Athapaskan (Na-Dene) speaking Chipewyan inherit 90% of their genetic ancestry from likely admixture with populations descending from the first migration wave, and 10% from a third gene stream also marking a distinct migration from Asia. The paper furthermore indicates that the Asian lineages leading to the Eskimo-Aleuts on the one hand and the Na-Dene-speaking Chipewyan on the other hand, are closely related and apparently “descend from a Siberian population that is a sister group to the Han.” Reich et al. (2012: 372) also stress the fact that they have data from just one Na-Dene-speaking group, and that “an important direction for future work will be to test whether the distinct Asian ancestry that we detect in the Chipewyan is a shared signature throughout Na-Dene speakers.”\(^6^3\)

---

\(^6^1\) Achilli et al. (2013) associate the A2a clade of the Pan-American A2 mtDNA haplogroup, which is found among Athapaskan and Eskimo-Aleut speaking populations, and notably the variant found among the latter (A2a2 and A2a3) which “experienced the steepest population expansion” ca. 4,000 BP, to the Paleo-Eskimo wave.

\(^6^2\) Dulik et al. (2012: 4) write: “The Y chromosome of the ancient Paleo-Eskimo man was assigned to paragroup Q1a\(^*\), but the NWT01 locus was not sequenced. Assignment of the Paleo-Eskimo Y chromosome to Q1a6 does not conflict with these data or the TMRCA of Q1a6.”

\(^6^3\) Reich et al. published a corrigendum in Nature, dated November 8, 2012, saying: “At the time of publication of this Letter, the authors were unaware of a manuscript arriving at broadly similar
7. Conclusion of Part I

What emerges from recent genetic research is a consensus on the idea that the prehistoric peopling of the Americas was the result of several migration waves, among which two bear a clear genetic signature: a major Paleo-Indian wave following a Pacific coastal route, generally dated back to ca. 16,000 BP, and a Paleo-Eskimo-Aleut wave in the northern extremes ca. 5,000 BP, to which one should add a more recent Thule Eskimo migration ca. 1,000 AD. There is no consensus, however, concerning the contours of other gene flows, likely involving several populations and language groups, notably Na-Dené and Algonquian, but there is little doubt that more extensive studies of these populations, as well as of populations of the northern Pacific coast, will reinforce or refute a number of pending hypotheses.

Thus, on the basis of current knowledge, it appears that two components of Greenberg's tripartition hypothesis have been given support: a major, initial Amerindian migration wave possibly corresponding, on linguistic grounds, to the Amerind macrophylum, and one Paleo-Eskimo wave possibly corresponding to the Eskimoan linguistic family. The initial genetic pool delineates a unique, homogeneous founding population whose size, at the start of the migration process, has been estimated between 70 to a few hundred individuals (Hey 2005, Fagundes et al. 2007, 2008), a range hardly compatible with several languages. It is thus consistent to transpose these results into the linguistic domain by adopting the most parsimonious hypothesis, i.e. to equate the initial colonizing group with a single language, which one may call Amerind. This assumption provides a genetic and linguistic explanation frame to the distribution of KOKO kinship terms across the Americas, whereby KOKO terms were present in the kinship terminology of the initial Paleo-Indian group, thus confirming the views expressed in Greenberg, Turner & Zegura (1986), Greenberg (1987), Ruhlen (1994a, 1994b) and Greenberg & Ruhlen (2007) about the initial peopling of the Americas and in particular about the Amerind linguistic phylum.

Reconstructing the phonetic form of the Proto-Amerind KOKO kinship terms using the comparative method is still out of reach, but the number of existing or reconstructed reduplicated forms across the Americas, involving the velar consonant k and vowels a and o ~ u, strongly supports the original existence of the phonetic shapes koko ~ kuku ~ kaka.

The statistics of table 14, calculated from the data reported in table 13, unequivocally indicate that these forms were referring to a limited set of kinship relations which altogether display a clear parallelism between the feminine and the masculine sides. Although MB and EF appear as the preeminent relationships referred to by koko terms, at a continental scale, the GM and FZ relationships certainly do not represent random significations, nor do the Sib relationships. The task of the second part of the paper (Matthey de l'Etang and Bancel, to appear) will be precisely to unveil the semantics of koko kinship terms, and ultimately the nature of the kinship terminology of the first Amerinds.

"conclusions based on allotype analysis by Williams et al., which appeared in the American Journal of Physical Anthropology 66, 1985."

64 Gilbert et al. (2008) and Achilli et al. (2013) disagree on the time frame regarding the introduction of the Beringian mtDNA haplogroup A2a into Northern Alaska, Northern Canada and Greenland: during the Thule expansion-wave ca. 1,000 AD according to Gilbert et al., or during the Paleo-Eskimo wave ca. 5,000 BP according to Achilli et al. By contrast, both research teams agree that the mtDNA haplogroup D3 is probably associated with the Thule expansion.
REFERENCES

NB: For references with more than seven authors, only the first seven names are given, followed by the mention “et al.”


BELLEZA CASTRO N. 1995. *Vocabulario Jaconar-Castellano, Castellano-Jaconar (Aimara Tupino)*. Cuzco: Centro de estudios regionales andinos «Bartolomé de las Casas».


GONZÁLEZ HOLGUÍN D. 1607. Gramaticay arte nueva de la lengua general de todo el Peru, llamada lengua Qquichua, o lengua del Inca. Ciudad de los Reyes: Francisco del Canto.

GONZÁLEZ HOLGUÍN D. 1608. Vocabulario de la lengua general de todo el Peru llamada lengua Qquichua, o del Inca. Ciudad de los Reyes: Francisco del Canto.


SANTO TOMAS D. DE. 1560b. Grammatica, o arte de la lengua general de los Indios de los Reynos del Peru. Valladolid: Francisco Fernandez de Cordoua.


MOTHER TONGUE

Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014


ZAMBRANO BONILLA J. 1752. Arte de lengua Totonaca conforme a el arte de Antonio Nebrija. La Puebla: Viuda de Miguel de Ortega.

APENDIX

List of KOKO terms in South American language families. The classification is that of Campbell (2012).

MOTHER TONGUE

Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014


Barbacoan: Guambiano kásuko U [Rivet 1941]; Colorado ?at’kó B [Moore 1968];

Boran: Bora (Miriña) aha DX, SW [Guyot 1977];


Chapacuran: Wari (Pakasnosovos) xa’-yB, yZ [Vilaça 1965, Everett & Kern 1997]; Itene uhuit U [Créqui-Montfort & Rivet 1913];

Nutabe
F,
[Lehmann 1920, Wavrin 1932 in Landaburu 21998];
GM
A,
guaku
-gwati
na-taka
kdke
Unclassified:
[Rivet 1949];
woman
A,
GM
sahga
U, GF,
~
kugu
anzaga
Atanques (Canahuamo) kuku
U, kake
F, gvasi
A, sukui
Arhuaco (Bintukua, Ica) tegwe
U, kakke
~
kake
F, -gwati
A, no-tako
GM [Lehmann 1920, Wavrin 1932 in Landaburu 2 1998];
Unclassified: Nutebê
guaku
old woman [Rivet 1949];

Chipaya-Uru: Chipaya
ouikê
GCh [Snethlage & Snethlage 1932];

Chocoan: Emberá Catio
kau
D [Wavin 1934 in Landaburu 4 1999b]; Choco (Emberá) kau
D [Faron 1981, Reichel-Dolmatoff 1962]; Epena k'au
D, girl [Quiro Dura et al. 2007]; Noanamá ka
D, kii
Z, kawê
EM [Reichel-Dolmatoff 1962];

Cholonan: Hibo
kotk
F, keek
M [Rivet 1949];

Chonan: Tehuelche
koka
F, dueho, kokuču
madrina, kov
U, gow-
B, Z, qon
GM [Fernandez Garay 2004];

Guahiboan: *axu
MB, EF, *akwe
GM [Matthey & Bancel 2014]; Guahibo (Hiwi, Sikuani) axuyu ~axuyu ~(allu
MB, EF, FZH, akiyu
EF, aküê
U, akwe
GM, nokyu my
MB, FZH, EF, akwe
GM [Arcand 1976, Ortiz Gomez 1983]; Guayabero kuwewon
B, kuewew [Tobar Ortiz 1989];

Jéan: Panará (Kren Akarore) kokripia
EM [Giraldin 1994, 1997];

Jivarano: Jivaró Achuar
kai
(?)B, (?)Z [Taylor 1998]; Huamibana
kair
(?)Z [Jakway 1987]; Aguaruna kaig
(?)Z [Bant 1994];

Kamakanan: Kamakan
qköong
GF, MB [Martius 1867b, Ignace 1912, Silva Martins 2007];

Karajá language area: Karajá
hi
(?)eB, ixi
yB [Krause 1911, Pétesch 2000];

Karirian: *kuku
MB [Matthey & Bancel 2014]; Kipeka-kuku
U [Maniani 1698, 1699, Adam 1897, Goeje 1932, Rodrigues 1948]; Dzubukua -kuku
U [Bernardo de Nantes 1709, Correia de Queiroz 2008]; Pedra Branca cuccuh
MB [Martius 1867a & b, Adam 1897, Rodrigues 1948]; Sabuya cuccuh ~ cuccuh
MB [Martius 1867a & b, Adam 1897, Rodrigues 1948];

Lenca-Misumalpan: Lenca: *koko
GF, (EF) [Constena Umaña 2002]; Lenca (Honduras) koko ~ kogo
EF [Lehmann 1920, Schuller 1928]; Lenca (Chilanga - El Salvador) koko ~ kogo
GF, anciano, koh

Lule-Vilelan: Lule-Tonocote
kwe
FZ [Machonri de Cerdeña 1878];

Makúan: Puinave
hika
A, -u'ii
huv
GF, hoh
GM, hov
eB, pit
Sib [Erickson & Erickson 1993]; Nadeb
huv
GF, hoon
GM [Weir 1984];

Mascoyan: Enlhet. dial. of Sanapaná
hawok
easuk
eB [Unruh & Kalisch 1997]; Mascoy (Western) koko
MB, hawok
eB [Braunstein 1983]; Mascoy (Eastern) hawok
eB [Braunstein 1983];
**MOTHER TONGUE**
Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014


**Maxakalian**: Maxakali xiyay MB, EF, FF, MF, GG, xukux, FZ, EM, GM, GGM [Popovich 1980];

**Muran**: Pirahâ D [Everett 2005];


**Paezan**: Paez k’d’ka MB, MZH [Wavrin 1931 in Landaburu 3 1999a, Bernal Villa 1955, Rojas Curieux 1995 in Landaburu 3 1999a];


**Qawasqaran**: Alakaluf (Qawasqar) xoykoy-las old woman [Claris & Viegas Barros 2007];

**Quechuan**: *kaka MB, WF, WB, ((c?)MBS* [Matthey & Bancel 2014]; Cuzco (1586, 1607 etc.: Q2) caca [kaka] U (hononirici); cacay [kakay] MB, WF, WB [Barzana 1856, Holguin 1607, Yaya 2009]; Coastal Peru near Lima (1560: Q1, Q2) cacay [kakaj] MB, WF, WGF [Sanito Tomas 1960]; Central Sierra east of Lima (1618: Q1) cacay [kaka] WF, U [Arliaga 1621]; Modern Q’ero (Q2) kaku my MB, MBS [Webster 1977]; Pitumarca (Q2) kaka MB, HB, WB [Milicic 2011]; Apurimac Quechua (Q2) kaka MB, WF [Camacho et al. 2007]; Ayacucho Quechua (Q2) kaka MB [Zarquiey & Cordoba 2008, Soto Ruiz nd.];

**Sálidan**: Sálita koko man, tihoho A, kaku Z [Martius 1867b, Tastevin 1922 & Estrada 1993 in Landaburu 2 1998];

**Timotean**: Timote-Kuika kuxiy GM [Rivet 1926]; Mucuchi kaak man [Rivet 1926];


**Tupian**: Western Tupian: Arikem: Arikem u-këra (c?)B, u-isja my B (ego ?), my Z (ego ?) [Nimendjaj 1932]; Karitiana syky (c?)B [Landin 2005]; **Mondéan**: *kotikotl* MB, EF, GF [Matthey & Bancel 2014]; Surui

❖ Isolates
Awaké (Sape) ma-kaohai ~ ma-ka’hai my GF, ka’hai old man [Koch-Grüngberg 1923, 1928];
Camah -ki MB [Mateson 1972];
Canichana eu-axa my GF [Créqui-Montfort & Rivet 1913];
Cayuvava -kiche U [Key 1975];
Cofán tōño U [Bormann 1976];
Kapixanà (Kanoë) kūkūn man, keke GM, ñkū WBD, HBD [Nimuendaju 1928, Archives Rivet, Bacelar 2004];
Kwaza ha’ka’i GF [Van der Voort n.d. (2004)];
Mapudungu (Mapuche, Araucano) hweku ~ weku MB, kacú FZH, kuku FM, FFZ [Lacham 1904, Faron 1956];
Masaca (Aikana) kokomai U [Ruhlen 1994];
Mochica (Chimu) kokood (5)eZ, (5)A, ikį EM [Salas 2002];
Movima aiku A, akāl eB, eZ [Créqui-Montfort & Rivet 1914];
Oţi (Chavante?) koaka B [Ruhlen 1994];
Taushiro ‘ukku (5)Z [Ortiz 1975];
Trumai kokó MZ, FBW, aoké EM, FZ [Galvão 1953, Sutherland Louis 1971];
Urarina (Simacu) ka-kaun FZ (dubious) [Walker 2009];
Warao ku, da-kr MB [Williams 1928, Heimen 1997];
Yaté (Fulnio) ţ-xi B [Pinto 1956].
Indo-European and Dravidian: Some Considerations

Stephan Hillyer Levitt

Abstract
This paper examines several considerations with regard to Indo-European and Dravidian from the vantage of Nostratic in the light of one another, and in the light of in certain instances comparable features in other language families grouped under the rubric Nostratic. The paper suggests that Nostratic languages spread with the original spread of anatomically modern humans from Africa, placing the linguistic data in sync with recent genetic and climatological studies. The linguistic features considered are the Indo-European and Kartvelian mobile s-, certain sound correspondences in Dravidian (c: t, r: l, p, r and l: l, l: t, l: f) and their Nostratic parallels and implications, metathesis in suggested cognates between Indo-European and Dravidian and its Nostratic implications, inserted and dropped r in Indo-European etyma from the vantage of Nostratic. In addition, a few individual Indo-European and Dravidian etyma are considered that indicate the antiquity of the Dravidian forms within Nostratic.

1. Introduction

It is becoming clear that the similarities between lexical items in different language families are less a matter of chance than of a relationship as sure as that of different lexical items within the different languages of a single language family, such as Indo-European (IE).

Most recently, such relationship has given rise to such lexicons of suggested correspondences as Bomhard (2008 and 2011) and Dolgopolsky (2008). Neither Bomhard nor Dolgopolsky have to date, though, made use of my articles relating to aspects of the Nostratic hypothesis, nor the work of South Indian linguists, in general, with regard to Nostratics.

Just as when Sir William Jones turned to Persian and Sanskrit, he was able to see the connection between different branches of IE; so, when we turn to Tamil, the most conservative of the Dravidian languages, with a classical literature going back to the early centuries BCE, we are able to see the connection between different families of languages – pointing to a monogenesis of language. It is to this that I attribute the independent

---

1 I would like to thank Bob Scott, currently Co-Director of Columbia University Library’s Digital Humanities Center for his help in accessing information in Bomhard (2008 [ebook]). I would also like to thank the anonymous reader for Mother Tongue for his worthwhile suggestion. An earlier version of this paper, with major printing errors in the Bibliography, appeared in International Journal of Dravidian Linguistics 42.2 (June 2013): 63-99. The customized version of the TransIndic Transliterator font used for some of the diacritics in this paper is available from www.linguistsoftware.com/tintu.htm, +1-425-775-1130.
observations of G. Devaneyan and myself that Dravidian is related to IE; and within IE, that it is related most closely to Germanic.

Devaneyan (1902-1981) was a linguistically savvy Tamil scholar whose Nostratic thesis is often distorted and misunderstood in the West. His arguments go back, though, to such pioneers in Dravidian studies as R. Caldwell (1814-1891) and G. U. Pope (1820-1908), and for instance to D. Savariroyan (fl. 1899-1914) and S. Gnana Prakasar (1875-1947).

In 1998 and 2000 I published two papers in The Journal of Indo-European Studies on the possible relationship between IE and Dravidian. With regard to my filtering of Savariroyan’s, K.C.A. Gnana Giri Nadar’s and Devaneyan’s suggested Nostratic correspondences between Tamil and IE in the 2000 article, Edgar Polomé commented that “I may not agree with some of your etymologies but I find them quite challenging. ... Congratulations on a nice piece of work!” (undated correspondence from July 1999). I note, I now would accept more of Devaneyan’s suggested correspondences than I did at that time.


My method, by and large, is to keep semantic transparency paramount, and to see if there are logical connections between forms that can be argued on this basis, using attested sound correspondences. There appears to be a stability of meaning in a large number of items.

It is my contention that many of the sound changes and alternations that appear in Dravidian are very ancient and can be seen in Nostratic in general, no doubt from pre-Dravidian.

It has been argued by some that there have been untold diffusions of peoples and tongues, mergings, and fresh dispersals. Because of this, we cannot be certain at deep time depths of the actual relation – genetic, borrowing, or simple chance resemblance after attrition and/or innovation. Thus, Hungarian ház ‘house’ looks very like Germanic versions of ‘house’, but it is seen to be cognate with Finnish kota ‘tent, cloth hut’, and to go back to Finno-Ugric *kota ‘tent, hut, house’. So also, devil, diable, diablo etc. in other languages suggest perhaps genetic cognates but documentation says they are all loans and/or devolutes from Gk. diabolos, literally ‘slanderer’. Many suggested cognations at deep time depths may be right, but how do we verify this? In the absence of documentation, how can we be sure?

And, in fact, such indeed is so. Yet, by keeping meaning constant and staying within the parameters of clear sound alternations we can see such developments.

For instance, DEDR 1796 Ta. kurai carries such meanings as ‘to bark, jubilate, shout’, ‘noise, roar, shout’ and ‘dog’. Also compare DEDR 1901 Ta. kūray, which forms mean ‘dog’ and ‘bitch’, and DEDR 2122 Ma. kora ‘asthma’. These forms give rise to Skt. kūkkura, kukkura, kukura ‘dog’, and a variant kukkuṭa (DEDR 1796). Regarding the Sanskrit and related Indo-Aryan words CDIAL 3329 etc. comments “Onomatopoetic”.

Levitt (1998: 151-152 [no. 23]) has argued that these are loan forms in Sanskrit (see also DEDR 1796), related on a deep level to Lat. canis, Eng. hound and, Levitt has argued, through metathesis, Eng. dog. The genetically related Sanskrit cognate is śvan.
With regard to Eng. dog, see the Hindi form kutā ‘dog’, kutā ‘bitch’, listed in CDIAL 3275 and cross-referenced there with Skt. kurkura.

To be kept in mind here is that there is an alternation in Dravidian between r and l (see Section 2.3). And an alternation between l and n in Dravidian forms as against IE forms can be seen in a number of correspondences argued in Levitt (1998). Levitt (2003) greatly simplifies the etymology of Eng. dog by demonstrating that in metathetical forms in Dravidian medial or final l or l becomes initial t- (see Levitt 2003: 179).

In the present context, probably related in English is the English word growl, which is imitative according to Webster’s 508b, and the related Scottish word gurl (see OED 6: 896a, 963c). Such may be a reduplicated onomatopoetic form, or a form showing an echo (as also OE. docga). And as well no doubt related are Eng. cur, ME. curdogge, ME. curren ‘to growl’ and Germ. obs. and dial. kurren ‘to growl, grumble, murmur, coo’ (see Webster’s 278b and OED 4: 135c for these and related forms, and their history).

The explanation of how we can have forms of such different shapes as hound, dog, cur, and growl that stand side-by-side in English all coming from the same protoform *kur-r/l/n/t is that they are from different dialectical sources. The generic name for ‘dog’ in Old English and the Teutonic languages generally, was hund (hound), pre-Teut. *kun-. Dog appears in Old English with its origin and previous history unknown. Cur appears first in Middle English and is of continental origin. Growl, with but two exceptions, appears first in the 17th c. Its continuity from the early usages is questionable, though it may have been preserved in some dialect. Its 13th c. usage is in Anglo-French, and has an ultimately Teutonic source.

Further explaining such variation from a single original source standing side-by-side in English, except for hound, which as a substantive is now used in specialized usage or fashion, the other words seem to appear originally in specialized usages (OED 4: 135c, 921b, 6: 896a, 963c, 7: 432bc).

While I think that such work as Bomhard’s and Dolgopolsky’s, V. Illic-Svityč’s earlier work, and J. Greenberg’s work on Eurasian, which set up clear-cut sound correspondences is premature, and that we ought to be casting a broader net as we are still in the stage of collecting data, I do think that they are very helpful starts. Thus, Illic-Svityč’s examples of correspondences between Nostratic l- and Dravidian t-, which Bomhard has not adopted, however, and which Dolgopolsky seems to have separated from one another (no. 1281) or just discarded as not proved by other etymologies (no. 1309), for example, was expanded by me in Levitt (2003). On this basis I was able to suggest still other correspondences between IE and Dravidian, and support further such formative processes I had given example of in Levitt (1998) as metathesis.

---

2 In this connection, I might point to a recent article by Sayers (2008) in which the English forms dog and cur are explained with what I would judge to be very forced arguments. The proposal here is much simpler and more straightforward, and is therefore preferable.

3 Compare Illic-Svityč’s correspondences as reported in Levitt (2003: 175) with, for instance, Bomhard’s listings under Nostratic *t’uk’-, *t’ok’-, and Nostratic *lak’h-, *lak’h-, accessed by an “All Fields” search for ‘beat’ and ‘leg’ respectively in the “Proto-Nostratic Dictionary” in Bomhard (2008(ebook)); and Dolgopolsky’s discussion of the Dravidian data at the end of his entry no. 1281 for Nostratic ‘push, shove, thrust, fasten to’ and his comment at the end of entry no. 1309 for Nostratic ‘to bend’ in Dolgopolsky (2008). On account of the difficulty in reproducing Dolgopolsky’s reconstructed forms here, I omit them.
Further, neither Bomhard nor Dolgopolsky take into account euphonic combination, which I hope Levitt (2010) has shown is indeed in force in Nostratic.

I propose here to treat several topics, in the course of which I will bring in some of the material I have treated earlier, as well as new material.

First off, though, something ought to be said about the spread of our Nostratic languages and their dating.

There are various “popular” scenarios regarding the origin of Proto-IE (PIE). Two predominate in recent years. One, advanced by Colin Renfrew, argues for the spread of IE languages with the spread of agriculture from southern Anatolia starting in the 8th millennium BCE. The other, advanced by Marija Gimbutas, argues that it was the spread of the Kurgan people from their homeland in the river basins of the Volga and Don in eastern Europe that led to the dispersal of IE languages from the early 5th millennium BCE onward. The International Genomic Project under the direction of Luigi Luca Cavalli-Sforza has combined the two, arguing for primary and secondary diffusion (Haarmann 1998: 391-393).

Still others, such as Paul Thieme and Ward H. Goodenough, have argued for a north Central European homeland, specifically the North European Plain of Germany and Poland, “or at least the eastern end in Poland and the western Ukraine” (Goodenough 1970: 254, 262).

Many Russian archeologists now consider the Kurgan culture to be Indo-Iranian. And, indeed, the contact of this culture with a culture seen by modern Russian archeologists to be the precursor of Baltic and Slavonic culture, would explain the observed linguistic contact between Indo-Iranian and Balto-Slavonic (Kuznetsov 2005: 325; Telegin 2005: 339; Burrow 1973: 18-23). This argument is more in keeping with the spread of Nostratic languages as I see it.

I hold a position which, while it sounds strange when one first hears it – as it did to me, makes perfect sense on reflection, and fits with the genetic, climatological, and geographic considerations.

I would see Nostratic as going back to the original peopling of the world by anatomically modern man (AMH). This opinion is bolstered by linguists of Australian aboriginal languages, who see these as having hived off from Dravidian, or pre-Dravidian, if you would (Dixon 1980: 236-237). So, also, Trombetti (1906-20). Recently, Masica (2001: 255), too, has opined on the basis of parallels with Australian aboriginal languages that a pre-Dravidian substratum in South Asia may well go back to the original peopling of the world by AMH.

Linguistically, the evidence includes phonological features such as both retroflexes and alveolars beside dentals, and typological similarity such as word order, an agglutinative morphology, and an inclusive/exclusive distinction in non-singular first person pronouns, to note just three such similarities.

There is also a connection between Australian aborigines and Dravidians in the kinship system, and in the use of the boomerang.

Ramanathan (1983-84, 2009a; “Introduction” to Devaney 2004: 17-201) has proposed a number of cognates between Australian aboriginal languages, just as Blažek (1988, 2006) has proposed cognates between Australian aboriginal languages and Dravidian. Blažek, though, views these as loan forms in Dravidian from an earlier population, as he accepts Zvelebil’s 1972 model of the descent of Dravidian into South
Asia through the northwest, and later dating for Nostratic than I would argue for (see Morris 2006: 99).

Note that the Y-chromosome genetic marker M130 is found with frequency among Australian aboriginal populations and Dravidian populations in South India, but is rare elsewhere. And recently, an entire village of Kallar near Madurai in Tamil Nadu, with which caste group the boomerang is particularly associated, was found to have the Y-chromosome genetic marker M130 (Wood 2007: 15-16; National Geographic Society and IBM 2005-2006: video and pamphlet, p. 7; Oppenheimer 2007[2004]: 171, 185, 187, and esp. chart on 186).

AMH appears to have reached South Asia by way of a coastal route before the eruption of the Toba volcano in Sumatra in 74,000 BCE. A “furrow” in the genetic evidence in South Asia argues for this (Oppenheimer 2007[2004]: 82). Also, recent archeological research has found similar sets of stone tools in South Indian sites both below and above the layer of ash from the Toba volcano, giving archeological foundation to the genetic data (Patel 2007: 15).

From there, AMH went to Australia perhaps by 65,000-70,000 BCE or earlier, according to Oppenheimer (2007[2004]: 159-163, 192), at least by 40,000 or 50,000 BCE according to others (Dixon 2002: 8-9), being cut off from the rest of mankind by the rising oceans since about 6,000 BCE.

Kivisild et al. (2003: 313) argue that there has been little genetic influx into South Asia from external regions since the original peopling of South Asia by AMH. Further, as Oppenheimer (2007[2004]: 152) has argued with regard to the Y-chromosome mutation M17 (R1a), often referred to as the “Caucasoid” genetic marker, it appears to have developed in South Asia, where it is most diverse (see also Sahoo et al. 2006: 843a).4

Both on climatological and geographical grounds, when AMH left Africa, he could not have gone through the Middle East, but would have gone by the coastal route (Field and Lahr 2005, Oppenheimer 2007[2004]: 68, 78-83).

AMH would have moved out of southern Asia to the region of the Zagros mountains, western Iran and eastern Turkey, and from there he would have moved further into the Eurasian landmass, and into the Levant and North Africa, after 40,000-50,000 BCE with the retreat of the glaciers, etc. (Oppenheimer 2007[2004]: 129-131, 133, 138-141, see map on 132), a second migration into Eurasia from southern Asia coming around 33,500 BCE by one of two possible routes – a trans-Caucasus route, or by way of Kashmir through the Central Asian steppe (Oppenheimer 2007[2004]: 144-146, 149-150, see map on 137).

This data fits with the linguistic scenario of Afroasiatic breaking off from Nostratic first, then Dravidian, then Kartvelian (Bombard 2008[ebook]: Ch. 1, 2011: 4-5).

2.1. Mobile s-

In IE, there is an initial s- that appears inexplicably in front of some forms only without any regularity in a set of etyma, forming in those instances an initial consonant group. Some examples are Skt. tanyati ‘thunders’, Lat. tonāre: Skt. stanayitnu ‘thunder’, see also Gk. steno, OChSl. stenju etc.; Skt. tr-, tārā ‘star’: Skt. str-, Av. star, Gk. aster etc.;

4 For fuller discussion of this data, see Levitt (2009: 140-144, 2010: 53-55, 57-61).

Burrow (1955: 80, 1973: 81) is of the opinion that the phenomenon is due to loss of initial *s-*, and is most probably the result of some kind of external euphonic combination affecting initial *s-* in the IE period.

Vladislav Illič-Svityč was of the opinion that it developed only in Nostratic words with an inherited internal palatal element (Dolgopolsky 2008: 18).

The interpretation of Dolgopolsky (2008: 18, 19) is that Nostratic emphatic consonants in word initial position were pronounced in IE with an initial pre-aspiration. Usually, this pre-aspiration left no trace. But in roots with an inherited palatal element the whole root was palatalized, and the pre-aspiration was transformed into a kind of *h* that later yielded a movable *s-*. Dolgopolsky also mentions a typological parallel in the Turkic language Salar. In Salar, the vowel *i* before voiceless *p, t, and k* (fortes) was transformed into *i* + a pre-consonantal sibilant.


There is a mobile *s-* as well in Kartvelian languages. Here, it is in both consonant clusters and before vowels. Thus, Georg. *ni(s)karti*- ‘beak’: Svan *nikrāt*, *nikrāt* (Common Kartvelian *ni-(s)krt-*); Georg. *sadgis*- ‘awl’: Megr. *odgiš-* (Georgian-Zan *(s)a-dg-is-*); Georg. *savargel*- ‘arm-chair’: Laz *orzm-, orzo-, orzo-‘bench, chair’ (Georgian-Zan *(s)a-warz-3l-‘seat’); Georg. *saxl*- ‘house’: Megr. *oxor-‘dwelling, abode; farmstead’, Laz *oxor-‘house’ (Georgian-Zan *(s)a-xl-‘house’) (Klimov 1998: 125, 143, 172-178, 338; see esp. notes under Common Kartvelian *pu-‘to chop, hack, cut to pieces’ [p. 152] and Common Kartvelian *(s)a-gie-‘milk’ [pp. 177-178]).

Also, Ohno (1983) reports doublets with and without initial *s-* in Japanese.

In this context, we ought perhaps to reconsider Proto-Dravidian *c-*. Proto-Dravidian *c-* can be analyzed as having undergone several different treatments in the various Dravidian languages, which treatments may represent different phonemes in the protolanguage; or rather, it can be analyzed as having undergone several irregular and incomplete sound changes (Krishnamurti 2003: 121; Emeneau 1988: 247a-260b, 265a-266b).

This phoneme is represented in some Dravidian languages by *c-* and in others by *s-. In some languages, there is free variation. Toda represents it always by *t-. Some dialects of Gondi, Kuvi, and sometimes Brahui change it to *h-*(Krishnamurti 1961: 8).

By one of these sound changes, it is by present analysis lost in South Dravidian, Telugu, and some dialects of Gondi (Krishnamurti 2003: 121-124; Emeneau 1988: 248a-250b; Burrow 1947). Thus,

(1) Ta. *il* ‘non-existence, death’, *illai* ‘it is not (in classical Tamil with a complete neg. paradigm); Ma. *illa* ‘does not exist, there is not; no, not’; Ka. *illa*, *illaŋ-, illavu* ‘is or are not; no, not’; Te. *lē-‘ (complete neg. paradigm, 3 sg. non-m. *lēdu*) ‘be not’; Pa. *cila* ‘does not exist’; Go. *sile*, *sille*, *hille*, *hile* ‘not, no’; Konḍa *sil-, Pe. *hil-, (in songs) *sil-, Kui *sidā* ‘not to be’; Kuvi *hill* ‘to be not’ (DEDR 2559) [PDr *cil-]*.

astringent taste’; Tu. ubary, ubāry, ogary ‘brackishness’; Te. ogaru, vagaru ‘astringent taste’; Go. sawwor, sauwar, havar, hovar, ovar, ovor, Konnda soru ‘salt’ (DEDR 2674[a] and [b]) [PDr *cup: *cuv-ar (< *cuy-ar)].

Ramaswami Aiyar (1929a: 152 [sec. vi], 1929b: 7n., 1930: 171, 1932: 29) views some such alternations of initial *c- (Kui s-) with <t> as being a development from an original prothetic front on-glide appearing before original palatal initial vowels. This opinion is consonant with IIIič-Svityč’s analysis of the IE mobile s-. Burrow (1947: 135) dismisses Ramaswami Aiyar’s suggestion, just as he dismisses explanation for the IE mobile s- that would regard the s- as the remains of some kind of prefix.

Devaneyan (1966: 96-99, 2004: 26'-29) takes issue with Burrow’s consideration only of aphetic omission, and not the prothetic addition of initial consonants. He notes that various words in Dravidian add prothetic c-/s~, such as Ta. urul ‘to roll, tumble over and over, become round, revolve, whirl’ (DEDR 664[a]): Ta. curi, curul ‘to be spiral as a conch, whirl round, curl, lie in a circle’ (DEDR 2684) and Ta. ugal ‘to oscillate, swing, be in motion, whirl, revolve’: Ta. cuṟanũ, cuṟal ‘to whirl, spin, rotate, roll’ (both listed under initial *c in DEDR 2698[b] and [a] respectively). In a 1966 article in Tamil in Centamil Celvi he notes uppu = ‘effervescent saline soil, or a lump or pinch of salt’, uman = ‘saline soil, fuller’s earth, salt’ (man = ‘soil, ground, earth’ [Fabricius 1933: 770a]), uvar = ‘salt which rises up and forms a film at the top of saline fields, saline land’. He comments that the Tamil demonstrative u is the ultimate base of thousands of words indicating forward and upward movement, among others, and uppu and uvar are among those words (Devaneyan 2004: 29).

We can also point out that TED 3/2: 55b derives Ta. cimir ‘to bind’, for instance, from Ta. imir, both listed under *c in DEDR 2542; and TED 1/2: 292b gives as cognates for Ta.ippi ‘pearl-oyster, shell’: Ka. cippi, Tu. cippi, Te. cippa (and in Indo-Aryan, Pkt. sippi), all also listed under *c in DEDR 2535.

Given the existence of a mobile s- in Kartvelian as well as IE, Kartvelian having broken off from Nostratic immediately after Dravidian by present opinion (Bomhard 2008[ebook]: Ch. 1, 2011: 4-5), it makes perfect sense that Dravidian, too, would have a mobile s-, and that such forms as in DEDR 2559 and DEDR 2674 ought to be reconstructed as *(c)il- and *(c)up: *(c)uv-ar [<*(c)up-ar]. Note in this regard that in Dravidian, there are no initial consonant clusters in the protolanguage.

That the etyma in DEDR 2559 ought to be reconstructed as *(c)il- is also indicated by a suggested correspondence in Levitt (1998), which correspondence was argued by Devaneyan before as well, between these forms and Eng. no, not (also non-, un-, in-, il- [before l], im- [before m or p], ir- [before r]). Germ. nein, nicht, Skt. na, an- (Pokorny 2005[1959-69]/1: 756-758, PIE *nē, *nē, *nət [*n-, word negation], see also 1: 17, PIE *aiu, *aju and Pei 1962: 154-155; Levitt 1998: 148 [no. 17], 2000: 421). The appearance of n in IE for l in Dravidian occurs in a number of other proposed cognates as, for instance, Eng. inn: Ta. il ‘house, home’ (DEDR 494; Levitt 1998: 147-148 [no. 16], see also entry nos. 13, 18, 23 and the comment toward the end of entry no. 13 with regard to Prakrit sound changes). A form with initial *s- (PDr *c- as presently configured) appears most probably in Lat. sine ‘without’, Sp. sin, Fr. sans, Skt. sanitur ‘besides, without (with prec. acc.)’ (Pokorny 2005[1959-69]/1: 907, PIE *seni-, *senu, [*səni-], *sn-ter-, see also 1: 318, PIE *éneu, *énu; Ernout and Meillet 1985: 627-628; Levitt 2000: 429 [no. 3]).
Another example in which a prothetic *s*- has been added in IE is Lat. *sōlus* ‘alone, sole’, Eng. *solo*. This word at present is without convincing etymology (see Pokorny 2005[1959-69]/1: 884; Ernout and Meillet 1985: 634b, Levitt 2012: 112, 160). Levitt would see it as being related to Dravidian words for ‘one’, single’ Ta. *oru*, or ... Ta. *ogru*, *onru* (*DEDR* 990[a] – [d]). Both Levitt and earlier David (1966: 240, 284-285 [Supplementary Note 4]) see this as related within Dravidian to a root *ol* ‘to unite (often with reality), to be joined, possible, feasible’ (*DEDR* 1006 Ta. *olu*, the latter with an enunciative vowel), which underlies such forms in Tamil as *okka* ‘together’ and *ogru* ‘one’ (*DEDR* 990[b] and [d] respectively; see Levitt 2012: 104-106). It is these forms as a group that Levitt sees as being related to the IE words for ‘one’ such as Eng. *one*, Germ. *ein*, Lat. *īnus* (Pokorny 2005[1959-69]/1: 286, PIE *oi-nos*; Levitt 1998: 144-147 [no. 13], 2012: 108-109, 149, 151; see also Devaneyan 1966: 232).


Also compare Ta. *nākam* ‘snake’ < Ta. *nakar* ‘to creep, crawl’: Eng. *snake*, OE. *snaca* < OE. *snican* ‘to creep, crawl’ (Devaneyan 1966: 250, Fabricius 1933: 594b-595a, 578a). In Finnish we also have *naakia* ‘to hunt crawling on the ground’. And in Hebrew we have *nāḥaš* ‘serpent’, noted in Cohen (2010: 898a) under ḤNS, which latter carries among its meanings ‘to hunt’, ‘to seduce’, and ‘reptile’. Both Dravidian forms are omitted from *DEDR* as well as *DBIA*. Burrow and Emeneau, though, no doubt consider Ta. *nākam* ‘snake’ to be a loanword in Dravidian from Skt. *nāga* ‘snake’. Sanskrit, though, does not show a corresponding verbal form, as we have in Tamil. And within Dravidian, we have other related forms which are reflected in Ta. *nā*, *nakku*, *nāvu* ‘tongue’, *nāvu* ‘to thrust out the tongue and move it sideways, mock by thrusting out the tongue’ (*DEDR* 3633); Ta. *nakku* ‘to lick, lap’ (*DEDR* 3570); Ta. *naku* ‘to laugh, smile’, *nakkal* ‘laughing, mockery’ (*DEDR* 3569). The relationship between the etyma in *DEDR* 3633 and Ta. *nākam* ‘snake’ would be that a snake constantly thrusts its tongue in and out of its mouth sideways. In Levitt (2007b: 20b-21a), it was argued because of these points that Skt. *nāga* was rather a loan form from Dravidian in Sanskrit, the parallel forms in IE displaying a prothetic *s*-.

In light of the IE mobile *s*- phenomenon, though, the Sanskrit form may well be inherited within IE. We just cannot be sure in such situations. In that Sanskrit does not display the matching verbal form or related forms containing any of the related cognitions that can be found in Dravidian (except, perhaps, Skt. *nagara* ‘town, city’ < *DEDR* 3568 Ta. *nakar* ‘house, mansion, temple, palace, town, city’, judged by Burrow and Emeneau to be a loan form in Sanskrit) does however suggest that perhaps Skt. *nāga* is indeed, nevertheless, a loan in Sanskrit.

These examples (and there are more) of prothetic *s*- in proposed IE cognates with Dravidian are, no doubt, related to the phenomenon of a mobile *s*-. With regard to a mobile *s*- being found in both Kartvelian and by the argument here, in Dravidian, and with regard to Kartvelian having separated from Nostratic immediately
after Dravidian by present theorization, note that Fähnrich (1965) discusses extensive similarities between Kartvelian and Dravidian of both phonetic and grammatical nature, to which Fähnrich added extensive lexical similarities. To be kept in mind here is that Fähnrich (1965) assumes an “Ibero-Caucasian” protolanguage that includes Kartvelian and North Caucasian languages. While Nostraticists today include Kartvelian in Nostratic, together with Dravidian and Afroasiatic, the North Caucasian languages are not so included.

As to phonetic similarities, Fähnrich notes (pp. 138-139) that the phonetic systems of the individual Kartvelian and Dravidian languages moved away from one another considerably over time. Nevertheless, certain general characteristics can be laid down. For instance, in comparison with the Kartvelian languages, the Dravidian languages show relatively more vowels and relatively fewer consonants. Further, the Kartvelian languages developed glottalized consonants, which the Dravidian languages lack. The Dravidian languages, on the other hand, possess retroflex consonants, which are foreign to the Kartvelian languages. But against these differences, there exists nevertheless a whole lot of common ground. For instance, all consonants of both language families have points of articulation similarly organized according to the following groups: guttural plosive, labial plosive, dental plosive, dental fricative, m, n, r(l).

The syllable structure, Fähnrich (1965: 139-140) notes, as well shows considerable common ground between the two language families. By far the most frequent initial sound is the guttural plosive. Further, as an initial, r(l) is most frequently replaced.

With regard to word construction, in both language families the word, root, and syllable were originally one and the same. The word was monosyllabic, being the pure root, and as yet no formatives were added. Further, in both language families, the basic models for word construction are similar, the most important structures being CVC (doubled, CVCCVC), CVCV, CVCVC [C = consonant, V = vowel].

As to grammatical similarities, Fähnrich points out (p. 141) that with regard to typology, the primary principle of word derivation is in both language families agglutinative. Further, both language families prefer to represent grammatical references with synthetic types of expression.

As to rules of structure (pp. 141-143), he provides an extensive listing of similarities. These include his noting that the usual sequence of parts of the sentence is subject, object, predicate. The synthetic character of the two families brings with it that this sequence can be changed when emphasis is demanded on a part of the sentence, or on account of some other circumstance. The verb is marked differently in transitive and intransitive in both language families. In both language families, the conjugation system is greatly developed. The infinitive functions in a number of languages as a noun (Georgian, Malayalam, Telugu, Kui, Brahui).

Fähnrich also provides an extensive listing of morphological similarities (pp. 144-146). For instance, a dental element is used to indicate the past tense (Tamil, Malayalam, Parji, Kui, Georgian). The infinitive ends with ‘a’ (Kui, Georgian, Hunzib, Bats). A verbal suffix, formed with a vowel + ‘n’, appears in Kota, Tulu, Georgian, Laz, Lezgi, Avar, Lak, Chechen, Adyghe, and Kabardian. Another verbal suffix, ‘p/b/n’, is found in Georgian and many Dravidian languages. A genitive suffix ‘-n’ is found in Tamil, Kannada, Telugu, Parji, Gondi, Brahui, Chechen, Tabasaran, and Lezgi.
In Fähnrich’s listing of similarities between the two language families in both rules of structure and in morphology, the sheer number of parallels noted presents a very strong argument for genetic relationship. With regard to lexical similarities (pp. 146-158) interestingly he lists correspondences between Kartvelian and Dravidian in words for ‘one’, ‘two’, ‘six’, ‘seven’, ‘eight’, and ‘not to be’, ‘to be’, ‘fire’, and a wide assortment of basic vocabulary. This list, it is noted, was to be continued. It does not cover all possible initials, such as guttural plosives or labial plosives. I am not aware of such a continuation having been published. The lexical correspondences offered, though, provide further compelling argument for genetic connection between the two language families.

Fähnrich (1965: 136-137) places the relationship between Kartvelian and Dravidian in the context of a hypothesized Asian language family most of which languages, such as Sumerian, Elamite, Hurrian, Hattic (the so-called Proto-Hittite), Urartian, etc., are dead. In the case of Dravidian, though, the closest language of which is Brahui, we still have today living languages. As to the geographical distance between Kartvelian and most Dravidian, he points to the distance between Kartvelian and Basque, which he sees to be related to Kartvelian, the two being separated by Indo-European languages as also Dravidian and Kartvelian.

With regard to Hurrian [= Mitanni] and Dravidian, see Brown (1930). With regard to Elamite and Dravidian, see McAlpin (1981). See also Ramanathan (2009b) for a review of Szalek (2006) regarding Lycian and Lydian, and regarding the Pelasgian language. According to Herodotus, the Lycians (Herodotus’ ‘Termilai’) came to Anatolia from Crete. Consonant with this, in a 1981 paper on Kartvelian and Sumerian, Fähnrich argues (pp. 90b-91a) that the Sumerians were not autochthonous in southern Mesopotamia, and that a Dravidian substrate can be found in Sumerian. In this regard, Fähnrich lists 17 lexical items in Sumerian which do not allow Sumerian etymologies, but for which Dravidian etymologies can be found. These include (1) Sum. Uri ‘city in Sumer’: Ta. ār ‘village, city’, Ma. ār ‘village, city’ ... Te. āru ‘village, city’, Naiki ār ‘village’ and (17) Sum. ud ‘storm’: Ta. uta ‘wind’, Ma. utu ‘to blow, to blow about’ ... Te. ādu ‘to blow about’, Gondi udā ‘to blow about’. Fähnrich argues that before the Sumerians arrived in southern Mesopotamia, Dravidian speech was current there.

In a more recent article, Fähnrich (1991) provides 90 regular phonemic correspondences between the protolanguages reconstructed for Kartvelian and Dravidian based on 129 lexical correspondences. In his listing of phonemic correspondences he notes the number of instances of each correspondence in the body of lexical evidence he has pulled together; and in his listing of lexical correspondences he refers back to the number he has assigned to each phonemic correspondence suggested for each phoneme of each lexical item.

He argues that the number of parallels between the Dravidian and Kartvelian protolanguages makes the assumption of an accidental similarity very improbable. He adds that the nature of the parallels – the semantic range of the protolanguages’ lexicons, the elementary area of the vocabulary covered, and above all the root morphemes concerned, also excludes the borrowing of one protolanguage by the other or borrowing from a third source by both protolanguages. Therefore, a genetic connection between the two protolanguages is suggested, if not proven.
It thus ought not be surprising that we would find examples of mobile s- in both Kartvelian and Dravidian, as well as in Indo-European, Japanese, and the Turkic Salar, and a suggestion of same in Hebrew nāhaš 'serpent' (i.e., ‘snake’) and its associated root HNS.

2.2. c: t

One of the treatments of *c- in Dravidian is its sporadic development to t-. This is attributed to the postulated original nature of *c- as an affricate (Emeneau 1988: 250b-151a; Krishnamurti 1961: 12; Burrow 1947: 142-143). And there is, indeed, indication of an affricate pronunciation of c- in the South Asian area, outside of the evidence from Dravidian, in for instance the Sanskrit verbal form ścutf-cyutf-/often in the later language, ścutf- ‘to flow, trickle, ooze’ (Rgveda +; KEWA 3: 380, EWA 2: 658-659 note, “without certain connection” [translation mine]). Emeneau (1988: 252b-255b) lists sets of etyma in Dravidian in which (a) *c- is sporadically replaced by t-, and in which (b) *c- is sporadically replaced by t-, with occurrences of φ in South Dravidian and Telugu.

As noted earlier, it is my contention that many of the sound processes we observe in Dravidian are also applicable to Nostratic (sec, for instance, Levitt 2010). In Levitt (2000: 430 [no. 5]), an example was given of this with regard to this phonological development. This was a proposed genetic relationship between Eng. tell, talk, OHG. zellen (Pokorny 2005[1959-69]/1: 193, PIE *del-; Buck 1949: 18.21; Pei 1962: 45-46) with Ta. col ‘to say, speak, tell, mention, utter, express, recite, repeat; n. word, term, saying, speech, proverb, praise’, collavu ‘saying, telling, proverb’; Ma. colluka, celluka ‘to say, declare, order’; Ka. sol, sollu ‘to say, speak, tell; n. word’; Pa. cul ‘to say’; etc. (DEDR 2855). With regard to the Germanic forms, Buck (1949: 18.21[5]) notes “outside root connections ... dub[ious]”.

Emeneau (1988: 265a-266b) also lists examples in DEDR of the sporadic replacement of *t- by *c- (or as Burrow 1947 referred to it, original c-/s-). These are offered by Emeneau without phonological explanation (p. 265a). To these, we may also add a connection to a Sanskrit form appearing first in Vedic literature, Skt. śūrpa ‘a winnowing basket or fan (i.e., a kind of wicker receptacle which, when shaken about, serves as a fan for winnowing corn)’ and Ta. tūru ‘to scatter, winnow, throw up as dust in the air’; Ka. tūru ‘to winnow, drive off chaff from grain by means of the wind’; Tu. tūpuni ‘to winnow, fan’; Te. tūru ‘husks of grain’, tūru-paṭṭu ‘to winnow’, tūrpidi ‘winnowing’; Kol. tūrpet ‘to winnow’ (DEDR 3402). This connection, first proposed by Burrow (1945: 118 [no. 48]), is noted in DEDR with question; but in this context it appears to me, to be fairly certain. The loan would be from Central Dravidian. Such can be determined by the shape of some of the Central Dravidian forms in comparison with those in South Dravidian.

In the context of the common occurrence of s- for *c- in Dravidian, and my contention that the Dravidian phonological developments are ancient and can be seen as going back to pre-Dravidian, we can also suggest that it is this replacement of *t- by c-/s- that is behind the IE demonstrative pronoun *to-, *tā-, *tio- (Pokorny 2005[1959-69]/1: 1086-1087) appearing in the Nom. sg. m. and f. as *sā, *sā (Pokorny 2005[1959-69]/1: 978-979). Such forms are also listed separately by Dolgopolsky (2008: nos. 2310, 2006) and Bomhard (2011: 68 [no. 114], 95 [no. 201]).
It is often said today that the alternation of \( r \) and \( l \) is not a feature of Dravidian, though such an alternation was observed early by, for instance, Konow (1903: 456), who cites in that place Robert Caldwell, as well.

\( l \) develops to \( r \), of course, in Vedic Sanskrit as against Iranian. And in eastern Prakrits such as Māgadhī, \( r \) becomes \( l \), always in Māgadhī.

It is my contention that \( r \) and \( l \) alternate in forms in Dravidian, but is most often masked by allophones of \( l \) such as \( \mathbf{r} \) and \( n \), which appear for instance in euphonic combination in classical Tamil. Thus kal + pu > karpu, kaṭal + karai > kaṭarkarai, kaṭal + paṭappai > kaṭarpaṭappai, cel + mati > cepmati, kal + mār > kaṃmār, vēl + miḷai > vēṃmiḷai, vel + vēl > veyvēl (Rajam 1992: 105). There is evidence that such rules of euphonic development follow phonological developments within Dravidian (Krishnamurti 2003: 93-94; Zvelebil 1970: 172, 178-180; Ramaswami Aiyar 1935-38/n.s. 28[1]: 20-21; see Levitt 2010: 64-70). It is this which leads Emeneau (1967[1957]: 141-142) to see evidence for an alternation between \( r \) and \( r \), and “morphological doublets” with the alveolar nasal \( b \) and alveolar \( r \).

An alternation between \( r \) and \( l \) as such can be seen in colloquial Tamil, which contains such an alternation probably dialectically as well in some colloquial forms for literary Tamil forms, and in loanwords. See, for instance, Ta. pantar ‘a thatched shed’: pantal (so, also, Ka. pandar, pandara: pandal, Te. pandiri: pandili; DEDR 3922) and Ta. cāmpar ‘ashes’: cāmpal (DEDR 2453). For literary Tamil lāntar: lāntal (< Eng. lantern), colloquial Tamil in Tirunelveli district has rāndal. In literary Tamil such alternation can be seen in, for instance, Ta. mūtari ‘to establish with evidence, confront with proof’: mūtali (see also Te. mudarakincu ‘to remind of something painful or humiliating, ask, question’: mudalakincu; DEDR 5040) and in Ta. tōl ‘to be defeated, lose; n. defeat’, tōli: tōr, tōrvai (DEDR 3558). There is, as well, the standard development in Tulu both medially and finally of \( *l > r \). With regard to these points, see Zvelebil (1970: 144, 143) and, for a fuller discussion, Levitt (1989). 7 Also, see TED 12: 24ab [2.141] which gives 7 examples of \( r > l \) (4) and \( l > r \) (3) in Tamil etymology.

There does not appear to be a geographical pattern in Dravidian with regard to the alternation of \( r \) and \( l \), as we have in Indo-Aryan.

With regard to \( r \) and \( n \), Devaneyan (1966: 204) is of the opinion that these are late developments, and that before the formation of alveolar \( n \), the dental \( n \) was medial and final, as well as initial. It is only Tamil and Malayalam that show evidence for alveolar \( n \) beside dental \( n \). Zvelebil (1970: 130) charts their distribution in Tamil. Zvelebil (1970: 129-130), too, is of the opinion that they represent one single phoneme \( *n \) in Proto-Dravidian.

An alternation between \( l \) and \( n \) in proposed IE cognates with Dravidian has been indicated in Section 2.1.

Hodge (1998) has noted that there also appears to be alternation without obvious pattern in Afroasiatic between \( r \), \( l \), and \( n \), though he tries to develop one from his data.

7 Levitt (1989) contains many typesetting errors. It is hoped that these will be corrected if I am able to publish a collection of my papers on Dravidian topics in the near future.
Pokorny (1960: 161) also pointed to an alternation in Afroasiatic between r, l, n, and d as well.

With regard to an alternation between l and i in Dravidian, Levitt (2003) picked up on an earlier observation of Vladislav Illic-Svityč that Nostratic -l- corresponds to Dravidian -i- in initial position (see Illic-Svityč 1989: 151 and 1990: entry nos. 255, 260, 261). Levitt’s jumping off point was correspondences which this observation led to regarding Nostratic **laka ‘leg’ and Dravidian *ṭāk ‘walk’ (DED 3151) with a correspondence he had made earlier (1998: 139 [no. 5]) with PDr *kāl ‘leg, foot’ (DEDR 1479); and regarding PDr *āl ‘person’ and related forms (DED 399, 400), and the cyma in DEDR 3196 Ta. tāṇ ‘oneself’ and DEDR 3162 Ta. tām ‘they, themselves’. This observation that medial and final *-l/-i alternate within Dravidian with forms with initial *-l- leads to a large number of interesting correspondences, such as Ta. el ‘lustre, splendor, light, sun, daytime’ (DEDR 829): Ta. āl, ēl ... n. ‘fire, lamp, heat’; Ma. āl ‘fire’; Te. āṃḍra ‘light, brightness, heat’, āṃḍramu ‘heat, light, splendor’ (DEDR 3266) and Ta. tegal ‘heat’; Nk. tirup ‘sun’s ray’; Go. ter- ‘to be fierce (heat of the sun)’, tariṭāna ‘to be hot (of sun)’, teṅk- ‘to warm oneself by fire’; Konḍa ṭeṛvel ‘sunshine’ (DEDR 3440); compare Eng. sun (see Section 2.1) and Skt. div, duṣ, deva (see Section 2.6). Also, Ta. kāl ‘air, wind’ (DEDR 1481), Ka. gāli, gāl ‘wind, air’ (DEDR 1499); Kur. tākā ‘air, wind, breeze’ (DEDR 3149); compare Eng. gale etc. And, Ka. mēḷamba ‘the black humble bee’ (DEDR 5098): Ta. tīm ‘sweet’, tīyam ‘sweetness’, tīṃ ‘sweetness, honey, honey-bee’ (DEDR 3268[a]); Ta. tēṅ ‘honey’, tēṅ-i ‘honey-bee’, tīṅ-toṭai ‘beehive’; Tu. tīga ‘beehive, honey’; Te. tēne ‘honey’, tēnt ‘a large black humble bee’ (DEDR 3268[b]); compare PIE *mēdhū (see below, and Section 2.6).

It does not seem possible to tell which is primary, forms with ∗-l/-i or ∗-l-.

The correspondences gathered in Levitt (2003) appear to suggest a split in Dravidian, or pre-Dravidian, between *-l- and *-i-. So also, Ta. kalakali ‘to rustle, tinkle, rattle’: kalakala ‘to rattle, chatter, gurgle’ (DEDR 1302); and Ta. kol ‘to kill, murder’: Br. xalling ‘to strike, kill’ (DEDR 2132): Ta. kol ‘to strike, hurt’, kōl ‘killing, murder’ (DEDR 2152).

---

* P. Ramanathan has communicated to me (correspondence dated July 15, 2013) that the gloss ‘killing, murder’ for Ta. kōl in DEDR comes ultimately from a gloss for the word by Swami Nathaiyar in his edition of the Ĵivavacintamani (2nd ed., 1907; 3rd ed., 1922). This was picked up by Madras University’s Tamil Lexicon (6 vols., 1928-39) and was from that, transferred by Burrow and Emeneau to their Dravidian Etymological Dictionary (1961) and to DEDR. P. Ramanathan notes, though, that in its context, in stanza 264 of the Namakaḷ Ilampakam, the traitor Kattiyāṅkarāṇa enters Ĵivakāṅ’s palace with the objective [kōl] of murdering Ĵivakaṅ, Ta. kōl meaning ‘objective’ and only metaphorically, in this particular context, ‘killing’. P. Ramanathan does not think that the form conveys the meaning ‘killing’ anywhere else in Tamil literature. It would thus fall in with DEDR 2151 Ta. kōl (kōl-, konṭ-) ... kōlai ... ‘determination’; kōl ‘taking, receiving, accepting, holding ...’; the latter also carrying the meanings ‘decision, determination, conclusion’.

Fabricius (1933: 308b-309a [kōl], 320b [kōl = verbal noun]) lists both the transitive meanings and the intransitive meanings, listed separately in DEDR 2151 and 2152 respectively, in the same entries.

Note in this regard Nk. gor- (goṛ-) ‘to strike, beat, KILL’ in DEDR 2152. Also, in DEDR 2152, Ma. kōl ‘hitting, wound, damage’; Te. (B.) konu ‘to be pierced as by an arrow’; Kol. goḷ- (godḍe-) ‘to beat, shoot with bow’; (P.) god- ‘to cut with axe’. It was speculated in Levitt (2003: 188, under entry no. 47) that both DEDR 2152 and DEDR 2132 might be related on a Nostratic level to Uralic *liikk- ‘pierce, insert, push’ and Altaic *liuka- ‘pierce’, as given by Illic-Svityč (1990: entry no. 261). Also note here Eng. kill, OE. cwellan, caus. of OE. cwelan ‘die’, ME. culle, kille, earlier also ‘strike, beat, knock’ (see Levitt 1998: 139-140, entry no. 7). [SHL]
There is as well an alternation of medial and final */l/* and medial and final */t/* in Dravidian, but it does not appear to be as thoroughgoing. Thus, for instance, a connection appears to maintain between Ta. kutirai etc. ‘horse’ (DEDR 1711[a]) and Ta. kul/* to move forward, to go forward, to move the body in a graceful and affected manner’ (as does a horse) (TED 2.2: 506a). The standard traditional explanation of Ta. kutirai is that it is derived from Ta. kuti ‘to jump’ (Devaneyan 1966: 58; Zvelebil 1970: 100; TED 2.2: 414a). Note, though, the etymologies given in TED 2.2: 412a for kuti/* ‘jump’ < kutu, but for kuti/* n. ‘endeavor’ < kutu < kul. Compare Eng. colt ‘the young of a horse’, of obscure origin according to OED 3: 505b.

So also, DEDR 2231 Ta. koru, kolu ‘to bale, draw up (as with an ola basket)’ at present lists with question Ko. ko-t ‘to fill by scooping’.

One reason for the sparsity of such data would appear to be the aforementioned split in Dravidian, or pre-Dravidian, between */l/* and */t/*, and the alternation across the board in Dravidian between /l/ and /t/. Thus,

(1) Ta. kulam ‘tank, reservoir, lake’; Te. kolanu, kolaku ‘tank, pond’; Go. kol ‘tank’; Kui glinju ‘a small pool, puddle’ (DEDR 1828): Ta. Ma. kutam ‘water-pot’; Ta. kutukkai ‘coconut or other hard shell used as vessel, pitcher’; Ma. kutukku ‘shells (as gourds) used as vessels’; Te. kudaka, kuduka ‘cup, bowl, any cup-like thing’ (DEDR 1651).

(2) Ta. il ‘house, home. PLACE’ (DEDR 494): Ta. itai ‘middle in space or time’; Ka. eda, ede ‘place or time between, interval, distance, middle’; Te. eda ‘space, interval, distance’, edamu ‘intervening space, distance’ (DEDR 448). See TED 1.2: 265ab, 260b-261a, which comments il —> il —> itu —< itai. Note, further, in DEDR 494 Ta. il, Kui idu ‘house, dwelling, shed, hut’.

(3) So also, Ta. cutu ‘to be hot, burn’; Te. cādu ‘to burn, brand with a hot iron or the like’; Konda sur- ‘to roast, burn (incense)’; Kur. kurnā ‘to grow warm, become hot, be heated, cook on live embers. bake on an open fire’ (DEDR 2654). TED 3.2: 212ab derives Ta. cutu from cul, which carries a meaning ‘pungency, acrimony’, and which in turn TED 3.2: 271b derives from cul (see also the etyma on 272ab for the semantics here).

TED 12: 24b [2.143, 144] gives 3 examples of / > l/ within Tamil etymology, and one example of / > l/. TED 12: 19a [2.84] gives two examples of / > t/*

In this regard, David (1966: 274) has pointed out that when a word occurs in two forms in Tamil, with both -l/ and -t/ in its basic root, the form with -l/ is always the earlier form.

* The forms cited in TED for l/* are: Ta. affalai ‘apartment on an upper storey; watchtower on a fort; raised covered platform from which one keeps watch on a garden, a field, a village, etc.’ [TED 1/1: 102b] – Ta. affalai ‘upper storey; covered platform for watching a garden’ [TED 1/1: 103a]. Ta. alappatu ‘to chatter, prattle, gossip, talk in vain, to talk nonsense, to talk incoherently’ [TED 1/1: 434b] – Ta. alappatul ‘chatter’ [TED 1/1: 502b]. Ta. alamaral ‘whirling, spinning around; confusion, perturbation, sorrow, fear, dread’ [TED 1/1: 436a] – Ta. alamaral ‘alamaral ‘to be bewildered, to be distressed’ (TED 1/1: 504b)).

The forms cited for / > l/ are: Ta. pollamani (= pol + a + mani), TED 6/3: 216b ‘see pollamani’ [pollal ‘boring a hole; chiseling, as a stone; hole, rent, fissure, puncture; ...’ (TED 6/3: 216ab)) – Ta. pollamani ‘gem which has not been bored; flawless gem; god, as immaculate’ [TED 6/3: 206a]. My focus it must be emphasized is Dravidian, or pre-Dravidian, and not Tamil as such, in which P. Ramanathan has rightly pointed out to me l/ and l/ are not allophones of one another (correspondences dated July 15, 2013 and September 5, 2013). [SHL]

In Vedic Sanskrit on the other hand, but consonant with this latter point, d > l when between two vowels. Thus, īḍe > īle, but īḍya. So also followed by h, for dh. Thus, mīḍhuse > mīḷhuse, but mīḍhvān (Whitney 1960[1889]: 19 [§54]). And so also in Prakrit (Pischel 1965[1957]: 172-173 [§240]). In North Indian Prakrit manuscripts, though, l is written, l having vanished in both writing and spoken classical Sanskrit (Pischel 1965[1957]: 166-167 [§226]; see also Kuiper 1991: 71).

Pischel also notes that sometimes in Prakrit t and d > t; and as dentals are often represented by cerebrals in Prakrit (and sometimes vice versa, depending on dialect), through t and d they become l. But as the North Indian manuscripts write l for l, it cannot always be said with certainty whether l or l is intended (1965[1957]: 174-175 [§244], 160-161 [§218]).

The Prakrits, of course, are commonly understood to reflect Dravidian speech.

Those correspondences that were indicated earlier by Illič-Svityč do not appear to be included by either Bomhard (2008) or Dolgopolsky (2008), as noted earlier (Section 1). But given such alternations in both Afroasiatic and Dravidian, it is likely that we are going to find such in Nostratic in general.

In Levitt (2010), a proposed relationship was spelled out in detail between Eng. bubble, burble and Lat. bulla ‘bubble’, both isolated within their IE sub-families, and Skt. budbuda ‘bubble’ and Te. buḍabuda ‘with a bubbling noise’ (DEDR 4249), which forms Levitt sees to be related on a genetic level (pp. 71-72).

In Levitt (2003), a relationship similarly was suggested between Eng. bone, OHG. bein and Eng. tube < Lat. tubus, akin to Lat. tūba ‘trumpet’, both also of uncertain etymology and isolated within their respective IE sub-families – the n of bone etc. standing for l as noted above, and the etyma in DEDR 5050 Ka. mūle ‘bone’ and DEDR 4528 Te. bomika ‘bone’ (Pa. būla, Go. būlā, bula) (pp. 178-179).

In that place, there also was suggested a genetic relationship with competing forms for ‘honey’ given by Pokorny (2005[1959-69]/1: 707, 723-724), PIE *mēdhu and PIE *mēl-t (Gen. *mēl-nēs), with the etyma in DEDR 3268(a) Ta. ūm and (b) Ta. tēg ‘sweet, honey, bee’ (p. 178; see Section 2.6).

We can also note here a suggested relationship between the locative suffix -il in Tamil, which is related to words for ‘place, house, home’ in DEDR 494 Ta. il (Fabricius 1933: 88b), with the demonstrative pronoun in IE, PIE *tō-, *tā, *tjo- (Pokorny 2005[1959-69]/1: 1086-1087). Again, metathesis would be in force.

---

8 See n. 3 above.
9 Interestingly, note Lakota (= Sioux): North Dakota, South Dakota. In northern Lakota, or Dakota dialects in Alberta, Canada, the form is Nak(h)ota, for which point I must thank John Bengtson.
10 All these forms are taken to be independent onomatopoetic developments, there being argument as to whether or not the Sanskrit word is a loan form from Dravidian or genetically inherited within IE. There is nothing very obvious about a very many onomatopoetic forms, though; and these are a case in point.
2.4. Metathesis

Metathesis has been referred to several times as being a formative process in Dravidian and Nostratic, and several examples have been given above. Neither Bomhard (2008, 2011) nor Dolgopolsky (2008) consider it as a formative process in Nostratic in their etymologies.

In Section 2.3 there was mentioned, among others, Levitt’s earlier suggestion that PDr *kāl ‘leg, foot’ (DEDR 1479) was cognate with Eng. leg etc. Without metathesis, in its meaning ‘column, post’ (Fabricius 1933: 235a; DEDR 1479), it can also be seen as being related to Eng. column < MF. colonne < Lat. columna (Devaney 1966: 236, 2004: 44-45). We can also suggest here a genetic relationship with Eng. calf (of the leg), ON. kālfı, “of unknown origin” (compare also, Ir., Gael. calpa ‘leg, calf of the leg’, OED 2: 781c; see also Devaney 2004: 45 [mng. 7], ‘part of the leg below the knee’).

Levitt (2000) dismissed K. C. A. Gnana Giri Nadar’s connection of Ta. nōkku ‘to look at, consider, reflect; n. sight, look, knowledge’ (DEDR 3794) with Eng. know, OE. cnāwen, Lat. gnoscere, Levitt opting instead for Devaney’s connection with Ta. kāy ‘to see, behold, perceive, understand’, kānkai ‘knowledge’ (DEDR 1443) and connecting Ta. nōkku instead with Eng. look, OE. lōcian, as had Devaney (pp. 413, 420-421, 424, 425). In the context of the clear-cut example of metathesis with regard to Eng. leg, though, and in the context of almost 60 sets of metathetical correspondences within Dravidian in Levitt (2003), many of the sets of correspondences containing more than one example, we ought now accept Gnana Giri Nadar’s connection as also a correspondence. This would be still another example of metathesis between Dravidian and IE forms, though metathesis most likely took place within pre-Dravidian, and the IE forms for ‘know’ are probably related more directly to the etyma in DEDR 1443 Ta. kāy, kānkai etc.

We can also add here Ta. Ma. kulam ‘tank, reservoir, lake’, Tu. kula ‘tank, pond’ (DEDR 1828) and Eng. lake < OF. lac < Lat. locus, OE. lāgu ‘sea’, Gk. lakkos ‘pond’ (Webster’s 644a; OED 8: 595c-596a)

This ought to suffice to indicate in brief the validity of the thesis.

2.5. Inserted and dropped r

It has long been established that the name “Dravidian” [Skt. dravida] is a reflex in Sanskrit of “Tamil” [Ta. tamī; by the standard transcription used for Tamil itself, tamī]. Suggesting this connection are the Sanskrit variants drāmīda and drāmīla and the Pali and Prakrit reflexes dāmīla, dāvīla, dāvīda (Caldwell 2000[1913]: 8-10; CDIAL 3788b [no. 6632]).

Levitt (1998: 142-143 [no. 11]) suggested that Germ. Deutsch ‘people, nation’, and the common language of same, was also a reflex in IE of these names, with reverse vṛddhi-strengthening (au [av] → u) and metathesis of the vowels.††

So also, Levitt argued, the name “Druid” for the ancient inhabitants of Gaul and the British Isles was similarly related. The present etymology that relates the name to ancient magicians, sorcerers, and soothsayers and on an analogy with Irish, analyzes the form as dru-vid ‘very knowing’, would seem to read into the name later attributions, and be forced.

†† Note that Caldwell thinks that the name “D(r)avida” was the original, and “Tamil” a development.
There have been many suggestions as to the meaning of the name “Tamil”, but a meaning ‘people, nation’, used as well to refer to their common language, is supported from a literary aspect by Sivaraja Pillai (1936: 344-346; see Levitt 1998 write-up and Levitt 2009: 145, 2010: 22). Such is supported by the names of such tribal groups in the hills around the Assam Valley in northeast India, such as the Dafla and the Mizo, meaning ‘people’. See also in this regard Bailey (1959: 109-110), who notes that such ethnic names as “Goth”, the name “Evenki” for the Tungus, and the name “Bantu” derive from words meaning ‘man’ or ‘men’. So also, Bailey argues, the people called by the ethnic name Ir. daha-, OPers. dahā. And as well, the name of the Finno-Ugric “Mari” is said to have been borrowed from Indo-Iranian (Anonymous 2009; see Skt. marya ‘man, mortal; pl. people’ < PIE *merjo [MW 791c; Pokorny 2005(1959-69)/1: 738-739]). See also Dolgopolsky (2008: no. 66), who attributes a meaning ‘member of one’s own ethnic group’ (→ ‘freeman’) for the PIE reflex of Skt. ārya, which carries a meaning in Sanskrit of ‘a respectable, honorable man, a man highly esteemed, noble’ (MW 152b; see Pokorny 2005[1959-69]/1: 67).

Levitt (1998) gave as parallel examples of matching forms both with and without r a suggested relationship between Eng. fuck and Germ. pfrecken ‘to prick’ (the latter word listed at Pokorny 2005[1959-69]/1: 167, the origin of the English word generally considered to be obscure), to which we can add colloq. Eng. prick ‘penis’; and between Eng. speak, speech and Germ. sprechen ‘to speak’ (Pokorny 2005[1959-69]/1: 997-998). Buck (1949: 18.21[5]) notes of the r-less forms in this latter set of etyma that their history is obscure.12

We can give here as well a number of other parallel examples:


---

12 With regard to Eng. fuck, see also the colloquial friggin’ and fricken’ as more socially acceptable forms of the word fucking.

13 I bracket this example as dropping an r in a conjunct is normal development in words going from Old Indo-Aryan into Middle and New Indo-Aryan.
Many seemingly possible correspondences have been eliminated here on the basis of standing etymologies in OED, but they may in fact constitute reasonable competing etymologies, such as crackle (in some of its meanings) and cackle. Also note Fr. clé, clef, Lat. clāuis, clāvus: Eng. key (OFris. kei, kay – not in other Teutonic languages; ulterior etymology unknown) (OED 8: 404b).

In the examples given above, time and again it appears that forms with r are more polite, forms without the r are more vulgar or colloquial. This does not seem to apply in all cases, though.

Kuiper (1991: 70-79, 79-85) has written about an “intrusive” r and r in Sanskrit loan forms. Often, the purpose of this is to “Sanskritize” foreign forms, but “often ... the rationale of the insertion is far from clear” (p. 70). In the present context, we can see it as a general phenomenon in IE.

It has been said that such forms as Ta. ētti ‘merchant’, Ta. ēni ‘ladder’, Ta. cāvakar ‘ascetics’, Ta. cunnam ‘powder’, Ta. patimam ‘idol’, Ta. vāñnam ‘color’ come from, respectively, Skt. śreṣṭhin ‘a distinguished man’, Skt. śreni ‘line, or row’, Skt. śrāvaka ‘disciple’, Skt. cūrṇa ‘powder’, Skt. pratimā ‘idol’, and Skt. varṇa ‘color’ through Prakrit, which simplifies Sanskrit conjuncts (Vaidyanathan 1971). The observations here, though, throw such derivations in doubt. Such forms may, rather, be in some cases genetically related. Note that TED gives Tamil derivations for all these forms, whereas for such derivations as Ta. nittam ‘dancing’ < Skt. nṛtta ‘id.’ and Ta. aramiyam ‘palace, terrace of a house’ < Skt. harṣyam ‘palace, large house’ given by Vaidyanathan, TED concurs.

2.6. Miscellaneous

Many of the forms preserved in Dravidian prove to be very ancient from the vantage of IE, and even from the vantage of Afroasiatic. This demonstrates the antiquity of forms preserved in Dravidian within Nostratic.

Etyma as in DEDR 5086 Ta. mē ... mēl, mēlimai, mēlukka, mēlai, mēlor etc. refer in the main to ‘excellence, height, superiority, eminence’. By means of a syncope within Proto-Dravidian, these forms are related to the etyma in DEDR 4841 Ta. micai, which forms have similar cognitions (Krishnamurti 2003: 96). It has been argued elsewhere that these forms are related on a Nostratic level to Heb. melekh ‘king’, Ar. malik ‘king, sovereign, monarch’, Ar. malaka i (malk, milk, milk) ‘to take in possession ... to rule, reign, exercise power or authority’, Ar. mālik ‘ruling, owning; owner, proprietor, master’ etc., the Afroasiatic root showing a common Proto-Dravidian suffix added to a root shown in its more basic form in Dravidian (Levitt 2009: 136-140, 148-150).

Etyma in DEDR 3268(a) Ta. tīm and (b) Ta. tēŋ contain words for ‘sweet, honey, bee’. PIE *mēdhu, Skt. madhu and PIE *mēli-t ‘honey’ appear to be formed by metathesis of a Dravidian form with a common Proto-Dravidian suffix added (Levitt 2003: 178). In that the initial m- of the IE forms corresponds to a suffix in the Dravidian forms speaks to the priority of the Dravidian forms.

The etyma in DEDR 3266 Ta. ti carry such meanings as ‘fire, lamp, heat, light, splendor, shine’. These have a metathetic reflex within Dravidian of the etyma in DEDR 829 Ta. el which carry such meanings as ‘lustre, splendor, light, sun, daytime, shine, spark’. The etyma in DEDR 3266 Ta. ti appear to be cognate within IE to such forms as Skt. div, dyu ‘heaven, sky, day’ (Gk. Zeus), giving rise to the root dyut- ‘to shine, be bright or

Notice Eng. stand, stall < PIE *stä: *stə, Skt. sthā. Levitt (1998) observed that IE forms appear to show common Proto-Dravidian suffixes in a number of instances, here -nt and -l. So also, notice DEDR 4345 Ta. pū ‘flower, blossom, bloom, flourish’: Skt. phulla ‘a full-blown flower, blown (as a flower), abounding in flowers, flowery’, which speaks to the antiquity of the Dravidian forms as they appear without the suffix -l. The Sanskrit form is without clear-cut motivation in IE (see Levitt 2000: 429-430 [no. 4]).

Also, note Ta. taka-tak-enal ‘onomatopoetic expression of dazzling, glowing, glittering’; Te. daga-daga ‘glitter’, Kur. dagnā ‘to light, set fire to, burn’, dagmā ‘to catch fire, be burnt’ (DEDR 2998): Skt. dāh- ‘to burn, scorch, consume by fire’, dagdha ‘burnt, scorched, consumed by fire’ (PIE *dhegʰh, Pokorny 2005[1959-69]/1: 240-241; Bomhard 2011: 61 [no. 90])15; Ta. tī ... (-pp-, -tt-) ... tippī, tippu ‘to be burnt, charred ... to shine; n. fire, lamp, heat, light, splendor’ (DEDR 3266; see above and Section 2.3): Skt. dīp- ‘to blaze, flare, shine, be luminous’, dīpa ‘light, lamp, lantern (Pokorny 2005[1959-69]/1: 183; MW 481a).

To change focus for a moment, in the context of Vasileiadis (2007) regarding mythological connections between ancient Greece and the eastern Mediterranean, and the Coorgs in Karnataka in western India, I could not help but recall the connection made by Legrand (1954: 44) between ancient Greek gala ‘milk’ and the etyma in DEDR 4096 Ta. pāl ‘id.’ Legrand draws attention to a Kannarese form hālu for ‘milk’, not listed in DEDR but it is well-known that p > h in Kannarese, prevalent first in Kannarese prose writing of the 10th c. CE, and shared as well with Tulu, Badaga, and Kurumba dialects (Zvelebil 1970: 86; compare Krishnamurti 2003: 120-121). Legrand notes that Greek g is the same as h in many words.

Webster’s 470a (s.v. galaxy) notes Gk. gala to be akin to Lat. lac, lactis ‘milk’ (< *glact, Pokorny 2005[1959-69]/1: 400-401).

The Greek form, though, may speak to a loanword connection between Dravidian and Greek, and to this areal Dravidian sound change having taken place earlier than assumed at present. See, for instance, a 1989 paper presented by Dr. Uppangala Rama Bhat at the 17th All India Conference of Dravidian Linguists and printed in Bhat (2006) which points out that p > h in Kannarese is already in evidence in the earliest Kannarese inscriptions dating from the mid-4th c. to mid-5th c. CE.

Such loanword possibility ought not be overlooked when etymologizing certain other Greek forms as well, such as Gk. dendron ‘tree’. This Greek form is often given as an IE justification for Skt. danda ‘stick, staff, punishment, army’ – which Sanskrit form Levitt has argued, to the contrary, is rather a loan form from Dravidian in Sanskrit, as the removal of its Dravidian reflexes from the Dravidian lexicon leaves a gaping hole in an otherwise logical semantic spread for Dravidian forms of the same and similar

14 Dolgopolsky (2008), however, lists the IE forms with his no. 2241, and the Dravidian forms under his no. 2250.

15 See also Dolgopolsky (2008: no. 511), which does not correlate the IE forms with Dravidian forms, however, as does Bomhard (2011).

3. Summation

Dravidian has been genetically related by specialists in different language families to languages all over the world, which efforts have generally received mixed reviews. It is hoped that by focusing here on Dravidian and IE, some of the reasons for these connections may emerge, and that a clearer idea of the development of Nostratic languages and of Nostratic morphology may emerge.

Abbreviations


Bibliography


Chennai: The Tirunelveli South India Saiva Siddhanta Works Publishing Society, Ltd.


MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014


Levitt, Stephan Hillyer. 2012. “The Number ‘One’”, together with comments from four
discussants and a ‘Response to Discussants’”. Mother Tongue 17: 103-166 [paper proper,
103-118; response, 149-166].

Masica, Colin P. 2001. “The Definition and Significance of Linguistic Areas: Methods, Pitfalls,
and Possibilities (with Special Reference to the Validity of South Asia as a Linguistic
Area). In Peri Bhaskararao and K. V. Subbarao (Eds.). The Yearbook of South Asian

Transactions of the American Philosophical Society, 71.3. Philadelphia: The American
Philosophical Society.

Tongue 11: 91-121.

National Geographic Society & IBM. 2005-2006. The Genographic Project, A Landmark Study
of the Human Journey [A research partnership of National Geographic & IBM / Global
field sciences supported by the Waitt Family Foundation], Journey of Man, narrated by
Spencer Wells. [Washington, D.C.?]: National Geographic Society & IBM (DVD video
with 19 page pamphlet).


Oppenheimer, Stephen. 2007[2004], The Real Eve. Modern Man’s Journey out of Africa. New
York: Carroll & Graf Publishers.


Pischel, Richard. 1965[1957]. Comparative Grammar of the Prakrit Languages, translated by
Subhadra Jhā. Delhi: Motilal Banarsidass. [Reprint adds index by Martino De Zilva
Wickremasinghe, published earlier in parts as supplements to Indian Antiquary 34-38
(1905-1909).]


Pokorny, Julius. 2005[1959-69], Indogermanisches etymologisches Wörterbuch, 2 volumes.

Possehl, Gregory L. 1996. “Meluhha”. In Julian Reade (Ed.), The Indian Ocean in Antiquity.
   London & New York: Kegan Paul International in association with The British Museum,
   133-208.

American Philosophical Society.

Ramanathan, P. [Irāmanāṭṭa, Pi.] 1983-84. “Āstirēliyap palaṅkuṭi makkal molojikal um tamiḻum”
   [The relationship between the Australian aboriginal languages and Tamil]. Centamίlc
   Celvi 58: 359-362.

Ramanathan, P. 2009a. Tamilum āttirēliyap palaṅkuṭi makkal molojikal um [The relationship
between Tamil and the Australian aboriginal languages]. Chemmozhi, Ceyti Maṭal

Ramanathan, P. 2009b. “Affinities of Tamil with ‘Pre-Indo-European’ Languages”. Chemmozhi:
A Quarterly Newsletter from Central Institute of Classical Tamil 3.2/3: 35-37.
Ramaswami Aiyar, L. V. 1929a. “Notes on Dravidian II (iv-vi)”. *The Indian Historical Quarterly* 5.1: 145-153.


Burushaski and the Western
Dene-Caucasian Language Family:
Genetic and Cultural Linguistic Links

John D. Bengtson
Association for the Study of Language in Prehistory

This paper is divided into three major parts: (a) some of the morphological evidence for the genetic relationship of the Burushaski language to Yeniseian, North Caucasian, and Basque languages; (b) some of the lexical evidence (both basic and cultural) for the same; and (c) phonological correspondences. 1

The hypothesis explored here, that the mysterious and 'isolated' Burushaski language of the high mountains of Pakistan is genetically related to Yeniseian, North Caucasian, and Basque (and more remotely to other Dene-Caucasian languages), has been refined from several earlier proposals.

Apart from a few fleeting observations, one of the earliest forerunners of the hypothesis was outlined by the Austrian Robert Bleichsteiner (1930), who compared Burushaski primarily with Caucasian, and explicitly also with the other “Japhetic” languages (Marr’s term), such as Basque, Iberian, Etruscan and Sumerian. His grammatical comparisons were mainly typological, noting for example (p. 299) that the Dagestani (EC) languages Andi, Tsezi, Lak, Archi, Rutul, Tsakhur, Budukh possess more or less “the same” four nominal classes as Burushaski: (I) male human, (II) female human, (III) other animate beings and various things (inanimates), and (IV) all other inanimates (see further below). He also remarked on the multiplicity of plural suffixes in Burushic and EC (pp. 303-306), and offered a number of material correspondences, e.g., Burushaski har ‘ox’, pi. har-o ~ Bezhta os ‘ox’, pl. os-o; Bur fi ‘soul’, pl. fi-ming ~ Archi anš ‘apple’, pl. anš-um, etc. Bleichsteiner pointed out some typological similarities between Burushaski and EC case markers, but not many of the material comparisons can be confirmed.

Bleichsteiner’s pronoun comparisons (pp. 310-311) were generally quite good, and some conform to present-day DC etymologies, such as Bur 1st pers. sing. źa, ja, Je ~ Udi zu, Lezgi zuн, etc.; Bur 2nd pers. sing. un, um, ung ~ Udi un, Lezgi, Agul wun, etc.; Bur demonstratives ke, ku, gu ~ Botlikh go-, etc.; Bur men ‘who?’ ~ Chechen mi-la, etc. (Some of these comparisons included Kartvelian forms as well.) 2 Bleichsteiner (pp. 319-331)

1 Based on a paper given at the 3rd Harvard Round Table on Ethnogenesis of South and Central Asia, Harvard University, May 12-14, 2001, with the title “Genetic and Cultural Linguistic Links between Burushaski, Caucasian Languages and Basque.” It has been revised (2014) and updated according to the current model of the Dene-Caucasian hypothesis. See also the Postscript at the end of the article. I am thankful to Peter Rowley-Conwy for assistance with the dating of animal domestication, and to Michael Witzel, Václav Blažek, and George Starostin for their discussion, support and encouragement.

2 Since the taxonomic distinction between Kartvelian (“South Caucasian”) and North Caucasian was not well-established until some decades later, the inclusion of Kartvelian comparanda with Burushaski (and Basque) was also frequent in the works of Bouda, Lafon, Uhlenbeck, and others. Most paleolinguists now consider Kartvelian to be a subgroup of the Nostratic (Eurasian) macrofamily, or closely related to it (e.g. J.H. Greenberg).
offered 73 lexical comparisons, but these have not fared so well in the light of present-day DC etymologies. Of these only a handful coincide with our current hypothesis, e.g.:  

§19. Bur (Y) yóqares, (H) yókuras, (N) yókurac ‘raven’ ~ Avar qaqra ‘raven’ [?] – not verified; but cf. Chechen q ‘ig ‘crow’. Lezgi k “ak ‘raven’, Agul q:”aRanaq ‘crow’, etc. < PNC *qHwavyV. However, these words are onomatopoetic, cf. PIE *grāk- ‘rook, daw, crow’; Turkic *KArga ‘crow’, etc.].

§21. Bur (Y) yórkun, (H) yírqm, (N) yírquc ‘frog’ ~ Avar q:”eRg ‘, Khinalug q:”rqa ‘or, etc. (see below) [onomatopoetic, but phonetic forms are quite close].

§32. Bur (Y) =huit- is, (H, N) =üt, =út-is ‘foot’ ~ Avar het’e / het ‘foot’, etc. (see below).


Some others, such as §9a: Bur (Y) gacer, (H, N) gachir ‘vulture’ ~ Andi qačirya ‘vulture’ [Godoberi qar:e:iR, Tsakhur q a:zir, Rutul a:zîr, etc. (SKJa 227)] look interesting at first glance, but a closer examination raises doubts. Berger (1998 III: 141) refers the Bur word to Pali gijjha- ‘greedy; vulture’; Waigali, Dameli grat ‘kite’, etc. (CDIAL 4430), which seems phonetically improbable. The Caucasian words resemble Mongolian qažir ‘vulture’, and may be borrowed from Kalmyk, a Mongolic language in the Caucasus region, or a related source, and the Bur word probably has a similar origin. (The Mongol Empire formerly included the Burushic area.) In sum, these words for ‘vulture’ are not genetic DC cognates, but loanwords from Mongolic sources. Obviously, this kind of winnowing process needs to be performed on all “Dene-Caucasian” comparisons proposed in the past (and present).

In 1950 Karl Bouda’s article “Die Sprache der Buruscho” appeared. On the first page Bouda expounded his view that Burushaski seemed to be a keystone that connected the great archaic language families (today we would say “macrofamilies”), including “Indo-Chinese”, Yeniseian, Euskarco-Caucasian, even Austronesian, Austroasiatic, and Chukchi. The article divides lexical and grammatical comparisons into three parts: I. “Indo-Chinese” connections (including Yeniseian!), II. Caucasian connections (the great majority from [North] Caucasian; a few from Kartvelian), and III. Chukchi connections.

Of Bouda’s 86 “Indo-Chinese” comparisons very few Bur + ST parallels can be verified, possibly only:

§48 Bur gi [Hi yîn, N yîn, Y yên “thief”, yê- “to steal”] ~ Tibetan r-ku “to steal” [< PST *(r-)qho – same etymology as Bouda (1954: §30), discussed below].

§56 Bur ghar [yar] “song”, ghare [yar-] “speak” ~ Tibetan d-gyer “to sing” [< PST *k(t)jâr or *g(j)jâr].

Though we cannot exclude that further study might increase this total. Similarly, very few of Bouda’s Burushaski-Yeniseian comparisons (mixed into the “Indo-Chinese” section) withstand the winnowing process and are retained in the current DC inventory:

---

3 Updated and corrected transcriptions have been added. Note that many of the lexical comparisons cited in this initial section are more fully documented and analyzed in the main Lexical Evidence and Phonological Correspondence sections, later in this paper.
Again, it is possible that further study might enlarge this list. On the other hand, Bouda’s Burushaski-Caucasian comparisons fare much better, and of them the following have been incorporated into the current stock of DC etymologies:

§95 Bur -umus [=i-mus, Y =yi-mus] ‘tongue’ ~ PNC *maz- [Udi muz, etc. < *mêlç].
§103 Bur t’p ‘leaf’ ~ Circassian (Adyge) thâp [PNC *?âpi].
§105 Bur ti [=thi-] ‘to pour’ ~ Avar t’[t’e-] id. [PEC *=VwV] [same etymology as Bouda (1964: §4), see below].
§118 Bur ca ~ Lak ac’ar [=a=ca-] ‘to stand’ [PEC *HêreV-]; but Archi -occ ‘[o=c:i-] < PEC *=VmgVr.
§121 Bur caghum [chãyur] ‘box for grain’ ~ Avar caghum [caýur] ‘corn bin, barn’ [PEC *cVoVr].
§127 Bur su-sum [=súsum, Y =sésen] ‘elbow’ ~ Udi sum id. [PEC *sínô].
§135 Bur tchi ‘mir geben’ [*=chi- ‘to give’ (class IV sing, obj.)] ~ EC *tcha [*=icV] ‘to sell, distribute’ (Chamalal ic- ‘to sell, give’).
§149 Bur nach [nas] ‘blame, reproach’ ~ Lak natch [nač], Avar netch [nič], etc. ‘shame’ [PEC *nač.V].

§28 Bur =s- ‘say’ ~ PY *Pas-/ *Pes- ‘shout, call’ [cf. Bsq *e=ša-n ‘to say’, *hic ‘word’; Tzezi =es- ‘to say, speak’, etc.].
§38 Bur ši ‘eat’ ~ Ket ši id. [cf. Bsq *ausi-ki ‘to bite’; Tzezi =ats- ‘to eat’, etc.].
§63 Bur =ik ‘name’ ~ Ket ī, Kott īx, īx id. [? cf. Tlingit ‘ix’ ‘to call; invite’ (Leer 1993)]
§77 Bur =mâl-/ =mal- ‘to be ashamed’ ~ Kott amalag ‘fright’.

§38 Bur si ‘eat’ ~ Ket si id. [cf. Bsq *ausi-ki ‘to bite’; Tzezi =ats- ‘to eat’, etc.].

§63 Bur =ik ‘name’ ~ Ket ī, Kott īx, īx id. [? cf. Tlingit ‘ix’ ‘to call; invite’ (Leer 1993)]
§77 Bur =mâl-/ =mal- ‘to be ashamed’ ~ Kott amalag ‘fright’.

Again, it is possible that further study might enlarge this list. On the other hand, Bouda’s Burushaski-Caucasian comparisons fare much better, and of them the following have been incorporated into the current stock of DC etymologies:

§95 Bur -umus [=i-mus, Y =yi-mus] ‘tongue’ ~ PNC *maz- [Udi muz, etc. < *mêlç].
§103 Bur t’p ‘leaf’ ~ Circassian (Adyge) thâp [PNC *?âpi].
§105 Bur ti [=thi-] ‘to pour’ ~ Avar t’[t’e-] id. [PEC *=VwV] [same etymology as Bouda (1964: §4), see below].
§118 Bur ca ~ Lak ac’ar [=a=ca-] ‘to stand’ [PEC *HêreV-]; but Archi -occ ‘[o=c:i-] < PEC *=VmgVr.
§121 Bur caghum [chãyur] ‘box for grain’ ~ Avar caghum [caýur] ‘corn bin, barn’ [PEC *cVoVr].
§127 Bur su-sum [=súsum, Y =sésen] ‘elbow’ ~ Udi sum id. [PEC *sínô].
§135 Bur tchi ‘mir geben’ [*=chi- ‘to give’ (class IV sing, obj.)] ~ EC *tcha [*=icV] ‘to sell, distribute’ (Chamalal ic- ‘to sell, give’).

§149 Bur nach [nas] ‘blame, reproach’ ~ Lak natch [nač], Avar netch [nič], etc. ‘shame’ [PEC *nač.V].

§28 Bur =s- ‘say’ ~ PY *Pas-/ *Pes- ‘shout, call’ [cf. Bsq *e=ša-n ‘to say’, *hic ‘word’; Tzezi =es- ‘to say, speak’, etc.].
Bouda continued with articles in *Orbis* “Burushaski Etymologien” (1954) and “Burushaski Etymologien II” (1964). From the first we can cite:

§1 Bur *bal* [causative *s-pal*] ‘to kindle’ ~ WC *bol* ‘burn. shine’.

§4 Bur *dare* ‘threshing floor’ ~ Lak *darac*- [t:araca-lu] id. [< PEC *hr̥ŋy̥*].

§14 Bur *sau* [sa:] ‘oversalted’ ~ Circassian *s* ‘to salt’, *s* ‘salt’ [Adyghe *šəw*, Kabardian *šəw* < PWC *ʃə/ʃəː*; cf. Bsq *i=čə-šo ‘sea’].


§22 Bur *gus* ‘woman’ ~ Dargwa *gaza* [k: ’ac:a], Lak *k̩acca* [k: ’ac:a], Archi *gac:ci~ ’mare’ [< PNC *gwāːjëː*; Lak and Archi words are borrowed from Dargwa].


In “Burushaski Etymologien II” (1964), rather strangely, only one of the 39 North Caucasian etymologies offered seems to pass muster and is included in the current DC collection of etymologies, and it is essentially the same as one of his earlier comparisons: 

§4 Bur Y *hut-in* ‘to fill, pour’ [root is really =thi-] ~ PNC *p* [*=VfyV*] ‘to pour, soak’ = the same comparison as Bouda (1950: §105), see above.

In Bouda’s comparisons with Tibetan we now have to reject many, but note:

§67 Bur *sin* [in Y *sēsen-um, sesēn-um, H,N *sisin-um*] ‘clean, clear’ ~ [West] Tibetan *sigs* ‘clear, thin’, b=*siy* ‘thin, limpid (of fluids)’ [cf. *says* ‘to clean, cleanse’, (b)*sey* ‘clean, white’, etc. < PST *chēn ‘clean, clear’; Bsq *suesen* ‘right, correct, just’, *sin* ‘oath, truth, true’; Chechen. Ingush *c ena ‘clean, pure’, Batsbi *c’ayn, etc. < PNC *HāxEn-.]

Among the Bur-Tibetan comparisons we also find §69 Bur *sat* [šai] ‘power, strength, force’ ~ Tibetan *šed* [šed] id. It seems quite clear that the Bur word is Indo-Aryan, cf. Shina *šat*, Pali, Prakrit *sati*, etc. < OI *šakti- ‘power’ (Berger 1998.III: 392; CDIAL 12250). Since this word is apparently found only in Tibetan and Lolo-Burmese *šatx* ‘strength, force’, according to Peiros & Starostin (1996), it too may be borrowed from Indic. This comparison is the same as Bouda (1950: §17). Some other Bur-ST parallels may indicate areal borrowings in either direction:

- Bouda (1950: §18) Bur *ta* ‘leopard’ ~ Tib. *s=tag ‘tiger’ [Lepcha *s=than*].
- Bouda (1964: §66) Bur *chir*, Y *cer* ‘line, row, turn, chain’ ~ Tib. *chir* ‘order, course, succession’ [PST *chyr*].

---

1 Valid only if there was a semantic shift ‘mare’ > ‘woman’ in Bur. The Yen cognate *kuʔ* means ‘horse’ (‘cow’ in Ket).
Around the same time the Burushaski scholar Hermann Berger (1956, 1959) offered some interesting lexical comparisons between Burushaski and Basque. As usual, some must be rejected, but yet some remain:

- Berger (1956, p. 7) Bur har [har] ‘corn worm’ ~ Bsq har ‘worm’ [~ ār, haar, aar, ar < *ha(m)(a)r], cf. PEC *hābarV ‘worm’.
- Berger (1956, pp. 7, 10) Bur galgi ‘wing’ ~ Bsq hegaj ‘wing’ [also ‘loin, flank (of cow); fin (of fish)’]; cf. PEC *gijj[ə]i ‘elbow, arm, wing’.
- Berger (1956, p. 9, note 16) Bur =l-pur ‘eyelid’ [< *=l-t- ‘eye’ + *bur ‘hair’] ~ Bsq be-puru ‘eyebrow’ [< *be(gi)j-t- ‘eye’ + *buru ‘head’].
- Berger (1956, p. 12) Bur -yon [*yōn] ‘all’ ~ Bsq honi [‘complete, complete number’; *hain-ic ‘many, much’].
- Berger (1956, p. 17) Bur šoq [soq] ‘sole of shoe’ ~ Bsq *oški ‘shoe’ [cf. PEC *šālkV ‘hoof’; or *šwāɡHwV ~ *gHwāšV ‘heel, ankle’?].
- Berger (1956, p. 22) Bur nok ‘bent, crooked (human limb)’ ~ Bsq nokhu [*noku] ‘physical defect’ [if the latter is not < Romance: cf. Lat. nocius ‘hurtful, noxious’].
- Berger (1959, p. 26) Bur bién-um ‘narrow, thin, slender’ ~ Bsq mehe [< *behe-n?] ‘thin’ (inanimate) [cf. PST *pā ‘thin’].
- Berger (1959, p. 26, note 34) Bur tar-in ‘skin bag or bottle’ ~ Bsq larru [*larru] ‘skin, leather’ [PNC *loli ‘colour, to paint’ (‘skin, sheepskin’ in Dargwa)].
- Berger (1959, p. 27, note 35) Bur gisāyas, yašias [root *g=is-/ *y=āš-] ‘to weave, knit’ ~ Bsq josī [josi ‘to sew’] [cf. PNC *iršE ‘to weave’ = Bouda (1950: §147), above].

5 In a note (Berger 1959, p. 17-18, note 3) he joined Burushaski and Basque in a family with – Dravidian(†).
6 This cognate chain exemplifies a common change: ‘fence’ > ‘enclosed space’ > ‘meadow’, ‘field’, etc.
Berger (1959, p. 33, note 57) Bur *multur ‘nostril’ ~ Bsq muthur [*mutur < *muturά] ‘snout, muzzle; end, edge’ [cf. PEC *wěnǐV ‘beak, horn, head’ > Batsbi marú ‘nose’, etc.].

Berger (1959, p. 34, note 57) Bur tur, =ltur ‘horn’ ~ Bsq muthur, =ltur ‘snout, muzzle; end, edge’ [cf. PEC *wenǐV ‘beak, horn, head’ > Batsbi marú ‘nose’, etc.].


Berger (1959, p. 36) Bur halanc ‘moon’ ~ Bsq ilazki [*hila-s-ki?] ‘moon’.


Berger (1959, p. 37) Bur hunatn ‘fine cloth (as wedding gift), bridal ornament’ ~ Bsq *ehun ‘cloth’ [cf. *eihu-n ‘to weave, braid’, PEC *=ir%wV ‘to knit, weave, spin’].


Besides the lexical parallels, Berger noticed some regular and non-trivial correspondences, namely Bur initial *t- = Bsq initial /- (see the three comparisons beginning with Bur tarénum “narrow”), and Bur intervocalic *l- = Bsq intervocalic -rl-/rd- (the comparisons “barefoot”, “nostril”, and “horn”). In current DC phonology these are reckoned as reflexes of original lateral affricates *ɬ/ *ɭ/ *ʎ (see 3. Phonology). After producing what can only be called strong contributions to the evidence for the DC hypothesis, later in life Berger abandoned the connection, or at least considered it unprovable (Berger 1992: 6).

Olivier Guy Tailleur (1958, 1994) included Burushaski and Yeniseian in his “Paleo-Eurasian,” a superstock that also included Basque, Caucasian (incl. Kartvelian), and even Ainu! Nevertheless, he remarked upon some important DC morphological parallels.

---

7 Lezgi č:iy’er ‘pear’, Tabasaran ńiyor, Agul č:iy’ar, Tsakhur ńiyor, Kryz ńhur, Budukh čińur; beside Rutul č:ir, Archi č:er, Udi or ‘pear’; cf. Chechen, Ingush qor ‘pear’, Batsbi qor ‘apple’; other cognates in Lak and Dargwa, all < PEC *gHüre or *qHüre. NCED tries to explain the words with initial sibilant affricates [ʃ, ç, ď, ď]: “As for the strange first part *ju- ..., it is most probable that we deal with a compound *jum-řera ‘quince’ or *pear’ with the first component distorted.” Alternatively, a word such as ēy’er ‘pear’ or ‘apple’ (supported by external comparison with Bur *sugar and Bsq *sugar) could have existed alongside *gHüre or *qHüre, was lost outside of Lezgian, and influence of *jum ‘quince’ was secondary.

8 "...konnte eine Verwandtschaft ihrer Sprache mit irgendeiner anderen bisher nicht nachgewiesen werden: die strukturelle Ähnlichkeit mit dem Baskischen und den Kaukasussprachen ist auffällend, aber es fehlt an überzeugenden Wortgleichungen im Bereich der Körperteile, Verwandtschaftsnamen und niederen Zahlwörter.”

9 Some forms have been updated.
(1994: p. 39) Ket and Yug -d- in oblique cases ~ Basque *-t- (as in su-t-argi ‘firelight’ ['fire' + *-t- + 'light']) ~ various Caucasian and Kartvelian case morphs. [Tailleur interpreted this *d as a genitive morpheme; we now see it as an oblique stem marker. Cf. Burushaski *-t- in some plural forms such as (H, N) guşpûr-t-aro ‘princess’, (Y) =yûhar-t-iy ‘husbands’; and compounds such as =l-t-tûs ‘eyebrow’ (cf. Lak i-t-u-c’ani id.]

(1994: p. 40) Ket -s, -as [instrumental, comitative], Kott -os / -as [comitative] ~ Basque -z [instrumental] ~ Chechen -sa, Ingush, Batsbi -s [ergative animate], Dargwa -s [dative], etc. < PNC *-s: [instrumental animate; general attributive], [Cf. Burushaski -as / -das [verbal infinitive], e.g., her-as ‘to cry, wail’, min-aas ‘to drink, smoke’. The development to [infinitive] is convergent with that in some Lezgian languages, e.g. Lezgi -z [dative].

(1994: p. 41) Yug -ey (in locative -k-ey / -g-ey); fossilized in locutions such as Yug xot-ey ‘(to the) front’, fic-ey ‘down(stairs)’. Northern Ket hit-e6 ‘down(stairs)’, Ket aks-ey5 ‘why?’, Kott peg-ai ‘elsewhere’, etc. ~ Basque -i [partitive], -k [ergative] ~ Northwest Caucasian *-k'V [instrumental, comitative, etc.]. [Cf. Burushaski -e [ergative and genitive], e.g. (H, N) hiles-e ‘boy’ (erg.), hiles-e ‘boy’s’, etc.]

(1994: pp. 41-2) Ket -ya [dative] (in I-class -da-ija / II-II1 -di-ya, etc.) = Yug -y (I-da-y / II -di-y, etc.) ~ Basque -n [locative], as in etxe-an ‘at the house, in the house’, kale-an ‘on the street’, etc.; also -en [genitive], as in gizon-ar-en ‘of the man’ ~ Chechen -n, Lezgi -n, Udi -n [genitive] < PNC *-nV [genitive], also shifted to dative, ablative in other languages. (Note that PSC *y has merged with *n in Basque and Caucasian.) [Cf. Burushaski (Y) -(i)ja [comitative] ‘together with’, e.g. yek-iya ‘with the name ..’; (H) fossilized -aye [instrumental], (N) -aye, -îye, -eje id.]

(1994: pp. 42-3) Kott -çap [ablative, elative], e.g. ül-i-çap ‘out of the water’ ~ Basque*-ca-t, e.g. seme-tzat daukat ‘I consider him my son’, zoro-tzat naukan ‘je le tenais pour fou’; (possibly also *-ç in other agglutinated case forms: *-(r)a-n-c [directional], e.g. mendirantz ‘toward the mountain’; Northern Basque *-(r)a-ko-c [destinative]). [Cf. Burushaski (Y) -êi [general locative], e.g. =rên-êi ‘in the hand’; (H, N) fossilized [locative], e.g. hala-êi ‘in the goal (in polo)’.

(1994: pp. 43-6) Gender (class) systems that distinguish Yeniseian, Burushaski, and (some) Caucasian languages from surrounding Eurasian languages, and some material parallels in class marking among the three families discussed. The Ket vowels -u- (~ -a-, -e-) I-class (masculine) vs. -i- (~ -i-, -o-) II-class (feminine) correspond to some degree with the Northeast Caucasian *y= and *u=, respectively, most clearly in the object person markers -a- ~ -o- I-class / -i(d)- II-class. (Strangely, in Burushaski these are reversed as =i- I-class / =u- II-class.) Tailleur also calls attention to the Ket inanimate (III-class) object marker -b-, corresponding to Northeast Caucasian *w= ~ *b= (inanimate III-class).

10 Except class II (feminine), where the genitive ending is -mu (Berger 1998: 1, 58).
11 This suffix is lexicalized and only occurs with certain nouns and with verbs meaning "strike" or "shoot".
Tailleur relates the Ket Ill-class [inanimate] predicate adjective suffix -am, e.g. sulem-äm 'red', to the Burushaski lexicalized adjective suffix -um/-iäm, as in bur-im 'white', čhajir-iäm 'cold'. In addition he cites the Lak adjectival suffix -ma, mä-, -mi, -mur, as in luhe-mä 'that which is black' (NCED luhi-s:a 'black').

Plural morphemes. Yeniseian plurals -n and -y are compared with n-plurals in Ubykh, Batsbi, Dargwa, Khinalug, Hurrian, Urartian, and Burushaski -y in noun plurals and -n in verbal plurals.

Verbal "ruiniforms": this term, borrowed by Tailleur from geology, aptly describes the tendency in DC (particularly in the 'core' group, i.e. Caucasian, Burushaski, and Yeniseian) for verbal stems to wear down and truncate to very short forms, e.g. Burushaski *t- 'to do, make', Yeniseian *di 'to lie down, put down', PNC *dVr 'to let, leave; to stay'. Since this is also true of Kartvelian (e.g. PK *go- 'to feed') and other languages with extensive prefixation and suffixation, it is more a typological or areal feature than strictly genetic.

Personal markers: notes some of the pronominal parallels already discussed by other scholars, e.g. Ket ad 'I' ~ Avar dí-n, di-, etc.; Ket u 'thou' ~ Abkhaz w-arà, Archi n-n, Burushaski u-n, etc.; Ket ku= 'thou' ~ Basque ã, Burushaski gu= / go=, etc. These and more are discussed in detail in the Morphology section below.

Remarks on various verbal morphemes, such as present marker *-y- in Yeniseian and Burushaski (the latter with verb stems in -n, e.g. =mín- 'to drink': pres. =miy-); past tense markers *-n- and *-r-i- in Yeniseian ~ Basque -n (neki-en 'I knew it'); Burushaski *nu=kú=ći-n 'having given it to thee' [absolutive or converb]; PNC *-nV [past gerund / participle], e.g. Hunzib uhî-n 'having died', Chechen w=a/-ana 'he has left', Lezgi t'i=na 'I have eaten', etc.

More recently, George van Driem (2001) has endorsed the Burusho-Yeniseian link under the name 'Karasuk', mainly based on grammatical homologies.

Beginning in the 1970-80s, some members of the "Moscow School" picked up the threads left by earlier scholars. Vladimir N. Toporov, a prolific and wide-ranging Russian philologist, wrote an article (Toporov 1971) that laid some of the theoretical groundwork for the Sino-Caucasian hypothesis of the next decade.

The existence of genealogical connections between the Yeniseian languages and those of the Tibeto-Burmese linguistic family ... are regarded now already as doubtless ... [and] if the given convergences are taken for reliable, the Basque, North Caucasian and Burushaski languages might be regarded as islets, having remained from a formerly indivisible chain of languages stretched in the latitude direction, from the Atlantic deep into Central Asia ... the Yeniseian languages ... might be regarded as the next remaining link of the chain ...
Table 1: Burusho-Yeniseian Grammatical Comparisons by Toporov (1971)

<table>
<thead>
<tr>
<th>Noun: plural suffixes</th>
<th>Burushaski</th>
<th>Yeniseian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inanimate Plural *-ŋ</td>
<td>-ŋ&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Ket III-class *ŋ&lt;sup&gt;13&lt;/sup&gt;</td>
</tr>
<tr>
<td>Noun: possessive prefixes</td>
<td>(Y) a=rën, (H, N) a=řiŋ ‘my hand’ (Y) gu=rën, (H, N)gu=řiŋ ‘thy hand’ gu=mi ‘thy mother’</td>
<td>Ket ab=ľan ‘my hand’, ug=ľan ‘thy hand’ Yug k=am ‘thy mother’</td>
</tr>
<tr>
<td>Pronoun stems</td>
<td>1sg *(?a)ŋ(a)&lt;sup&gt;14&lt;/sup&gt;</td>
<td>a= *b= / *?ab=</td>
</tr>
<tr>
<td>2sg *(?u)ŋ(u)</td>
<td>gu= / go=</td>
<td>*kV= / *?V)k=</td>
</tr>
<tr>
<td>3sg *mV</td>
<td>mu= (I=class)</td>
<td>*wV &gt; Ket bu</td>
</tr>
<tr>
<td>Deictic *ki-</td>
<td>(Y) kin, khené ‘this’ (I-class) / khamó (II) / guse, khot (III) / guté, khot (IV)</td>
<td>Ket kido ‘this’ (I-class) / kida&lt;sup&gt;6&lt;/sup&gt; (II, III)&lt;sup&gt;15&lt;/sup&gt;</td>
</tr>
<tr>
<td>Verbal subject</td>
<td>Plural *-n</td>
<td>-n *-n</td>
</tr>
</tbody>
</table>

Toporov’s younger Moscow colleague Sergei Starostin, who had worked on the reconstruction of Proto-North Caucasian, Proto-Yeniseian, and Old Chinese, re-examined many of these earlier proposals and concluded that the genetic affinity of Caucasian, Yeniseian, and Sino-Tibetan, at least, could be demonstrated on the basis of shared basic vocabulary and grammar, confirmed by regular phonological correspondences. His first articles describing what he called “Sino-Caucasian” were published in 1982 and 1984. The 1982 article outlined Starostin’s reconstruction of Proto-Yeniseian and proposed etymologies connecting Yeniseian, Caucasian, and Sino-Tibetan, many of which reiterated the earlier comparisons of Trombetti, Bouda, and others, but this time the classic Indo-Europeanist method was applied, proposing regular sound correspondences among the three proto-languages. The 1984 article (1991 in English) expanded on this and included some comparisons not involving Yeniseian as well. It will not be necessary to repeat any examples here, since (with a few exceptions and revisions) they are essentially identical with the most current etymologies and correspondences found in Starostin (2005a, 2005b) and the Tower of Babel databases.


---

12 With various vowel and consonant onsets.
13 Exceptions are usually phonetically conditioned: nouns ending in -n take the ending -y regardless of class, and nouns ending in -y take the ending -n (Werner 1994).
14 These are not Toporov’s reconstructions, but those formulated later by Starostin et al.
15 Pakulikha and Surgutikha dialects.
'oversalted', discussed above, Peiros discovered some new parallels, some of which have been adopted in the current hypothesis:16


- §5. Bur *l-әt(m)- ‘eye’ // PY *de- // PNC *әwәt.[190]

- §10. Bur qәs ‘dwelling house, quarters, lodging’ // PY *qәs ‘tent (made of birch bark)’ > ‘house’ // PEC *qәWAlr ‘house’. [It turns out that “PEC *qәWAlr ‘house’ does not exist, as explained by Starostin (1995a: 305); cf. instead PEC *qәVmV ‘balcony, verandah’?]

- §14. Bur (H, N) asii ‘star’ // PNC *әwәr ‘star’ // PST *әә ‘star’ [As it turns out, the PNC word does not belong here, but goes with Bsq *i=әwә instead. The underlying Bur form seems to be *ә=әm, based on (Y) asımum, and the (H, N) plurals asii-m-uc, and also matches Na-Dene: PA *әә ‘star’.]


- §38. Bur sis ~ ses ‘persons, people’ // PY *әә ‘people’ // PNC *әwә ‘man, male’ (Ubykh әә ‘persons, people’).


- §50. Bur =lәn ‘horn’ // PEC *әwә ‘horn, mane, braid’ [thus expanding Berger’s comparison, discussed above, of Bur =lәn and Bsq *a=әrә < *a=рәrә ‘horn’].

- §63. Bur *әә ‘distant relative’ // PEC *әә ‘relative, kinsman’.

- §65. Bur *lәn ‘to saddle’, *tәn ‘saddle’ // PEC *әwә ‘saddle’ [the comparison is semantically and phonetically perfect, though one wonders if horse riding is early enough to exist at the time of Proto-Western-Dene-Sino-Caucasian (ca. 6000 BCE). See the discussion below].

A paper by Blazek & Bengtson. “Lexica Dene-Caucasica” (1995), was the first to assemble a large number (219) of Dene-Caucasian lexical and grammatical etymologies that included all six families included in the current hypothesis: Basque, North Caucasian, Burushaski, Yeniseian, Sino-Tibetan, and Na-Dene.

For more recent work on Burushaski and Dene-Caucasian see the Postscript, below. In the rest of this paper the lexical and grammatical evidence assembled over a span of more than eight decades is re-examined in the light of the most recent developments.

16 Transcriptions and reconstructions have been updated/normalized.
17 With a seemingly unique (irregular) correspondence of Bur *l = PSC *r. This is not surprising considering the wavering between r ~ l in Old Indic (an areal neighbor of Bur), as in lәn ~ rәn ‘to lick’, lәbә ~ rәbә ‘to seize’, rәmәn ~ lәmәn ‘hair’, etc.
18 This word looks like an old compound, where the first element *beha- (‘liver’) could correspond to PST *beha ‘spleen’ (e.g., Thankur =pay ‘liver’); PY *bәlja ‘kidney’ could be a similar compound < *bәlja ‘intestines’.

190
Morphological evidence: Pronouns

Some of the Dene-Caucasian morphological evidence discussed above, and more, can be summarized as follows:

**Pronouns:** Both Burushaski and the reconstructed Proto-(North) Caucasian have *suppletive* pronoun stems in the first and second person singular; and, in our analysis, the original suppletive stems themselves are inherited, though with some paradigmatic rearrangements. According to Nikolayev & Starostin (NCED pp. 402, 483-84, 855, 1014-15, 1084-85), the original Proto-Caucasian paradigms were very complicated, and difficult to reconstruct with much certainty. For the present purpose, let us compare Hunza Burushaski (Berger 1998, vol. I, p. 80) with two East Caucasian languages, Khinalug and Tsakhur (and PEC):

<table>
<thead>
<tr>
<th>1st person singular ‘I’ / ‘me’</th>
<th>direct</th>
<th>ergative</th>
<th>genitive</th>
<th>dative / oblique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burushaski</td>
<td><em>zc</em></td>
<td>*zá-a</td>
<td>*záa</td>
<td>*áar [dat.]</td>
</tr>
<tr>
<td>Khinalug</td>
<td>zi</td>
<td>yā</td>
<td>i ~ e</td>
<td>as [dat.]</td>
</tr>
<tr>
<td>Tsakhur</td>
<td>zi</td>
<td>za</td>
<td>yiz-in</td>
<td>zā- [obl.]</td>
</tr>
<tr>
<td>PEC</td>
<td>*zō(-n)</td>
<td>*ʔez(V)</td>
<td>*ʔiz(V)</td>
<td>*zā- [obl.]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd person singular ‘thou’ / ‘thee’</th>
<th>direct</th>
<th>ergative</th>
<th>genitive</th>
<th>dative / oblique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burushaski</td>
<td>un ~ um</td>
<td>un-e</td>
<td>góo</td>
<td>góo-r [dat.]</td>
</tr>
<tr>
<td>Khinalug</td>
<td>wi</td>
<td>wa</td>
<td>wi</td>
<td>óy [dat.]</td>
</tr>
<tr>
<td>Tsakhur</td>
<td>wu ~ ku</td>
<td>yīk-na, yīk-n</td>
<td>wa- [obl.]</td>
<td></td>
</tr>
<tr>
<td>PEC</td>
<td>*uō(-n)</td>
<td>*ʔōwV</td>
<td>*ʔiūV</td>
<td>*ʔōwV [obl.]</td>
</tr>
<tr>
<td></td>
<td>~ *kwV</td>
<td>~ *kwV</td>
<td>~ *kwV</td>
<td>~ *kwV [obl.]</td>
</tr>
</tbody>
</table>

Clearly a great deal of rearrangement has taken place in these languages since the postulated original paradigms of thousands of years ago. Nevertheless it is possible to posit the following suppletive pronoun stems:

- **First person singular (1):** PDC *zV ‘I / me’ > Bur *je, ji-; Khinalug, Tsakhur zi, etc. (PNC *zō).
- **First person singular (2):** PDC *yV ‘I / me’ > Bur *āa-r [dat.] ‘to me’, also as pronominal prefixes *a = / *ā = / *āa = ‘I, my’ (i.e. a < *ya, a regular change in Bur: see SCP 48-49). This stem does not figure in the Khinalug and Tsakhur forms above, but only in Lak (na ‘I’)

19 In Indological tradition conventionally written je, jāa, etc.
20 According to NCED Tsakhur wu ‘thou’ (< *uō) and ku ‘thou’ (< *kwV) are in free variation.
21 ṣīn-na for 1- and 2-class, ṣīn-i for 3-class.
and Dargi (m ‘I’, m-ša ‘we’, etc.). PDC *ŋ regularly becomes *n in PNC/PEC *n ‘I’ and Basque *n ‘I’.

- Second person singular (1): PDC *wV ‘thou / thee’ > Bur ō-n; Khinalug wi, Tsakhur ōu, etc. (PNC *yō-n). Some Cauc forms come from the suffixed PEC *yō-n, similar to Bur ō-n, such as Archi. Udi ō-n ‘thou’. Lezgi, Agul ōm, Avar mm, Andi ōn ‘thou’ (assimilated < *hōn < *hōn, etc.

- Second person singular (2): PDC *xōwV ‘thou / thee’ > Bur ōo, ōo-r, also as a pronominal prefix, with many shapes depending on the following stem (gu/ /gu/ = /-kú= /-kó= /-kóo=); Khinalug oy [dat.], Tsakhur ōu [direct], yū-na, yūn-n [gen.]; elsewhere in Cauc: Chechen, Ingush, Batsbi ho ‘thou’, Dargi hu ‘thou’ (dial. gu, uy, u, i), Rutul (dial.) hu ‘thou’; and Basque *h ‘thou’ (restricted intimate use), verbal affixes *h=, *k < *-ga.

In Yeniseian analogous suppletions are evident:

- PDC *zV ‘I / me’ > PY *ʔař ‘I’ > Ket āt, Kott aj (ay), Pumpokol ad, etc.;
- PDC *yV ‘I / me’ > PY *h-/ab/-y > Ket āp ‘my’, ba= /ba= [1st pers. sg. verbal affix], Kott -t [1st pers. verbal affix] (for phonetics see SCP 48-49);
- PDC *wV ‘thou / thee’ PY *ʔaw / *ʔu > Ket ū ‘thou’, Yug u, Kott, Arin au, Pumpokol ùe id.;
- PDC *xawV ‘thou / thee’ > PY *kV- / *ʔVk- > Ket ūk ~ ùk ‘thy’, k =, ku = [2nd pers. sg. verbal affix], etc.

In other Dene-Caucasian subgroups the original suppletive paradigms have been regularized or leveled in various ways. In West Caucasian only the morphs *zō ‘I / me’ and *yō ‘thou / thee’ have survived (PWC *sa and *wa, respectively). In Basque, on the other hand, PDC *zV and *wV have been eliminated, and only the stems *ŋV and *xōwV remain, as *n ‘I’ and *h ‘thou’, respectively. Purely by chance the East Caucasian language Dargi has rearranged the first and second person direct forms to coincide with those of Basque: Dargi (Akusha, Urakhi) m ‘I’ / hu ‘thou’ = Bsq m / *hi.

An interesting discussion of Dene-Caucasian pronouns, including Sino-Tibetan and Na-Dene forms as well, has been provided by George Starostin (2010a).

**Lexical evidence:**

Burushaski-Caucasian-Basque-Yeniseian comparisons

### Body part terms

\[\text{Body part terms}\]

\[^{22}\text{/u/ represents a pharyngealized vowel, caused by the Proto-Dargi “emphatic laryngeal” *h.}\]

\[^{23}\text{The other, less intimate and more frequent Basque word for ‘thou’, zv (‘sv), is a recent development of the original 2nd person plural pronoun (= PNC *zve > Chechen ū, Lak zv, Archi zve-n, etc. ‘you’ [pl.]), analogous to the development in English in which thou has been mostly supplanted by the original plural you.}\]

\[^{24}\text{In each etymological entry, the Burushaski word or words are listed first, with dialectal designations (H = Hunza, N = Nager, Y = Yasin) where appropriate. Next, proposed Caucasian cognates, if any, then Basque and Yeniseian cognates, if any, and finally there may be notes, following the symbol §. The notes may include more remote Sino-Tibetan and Na-Dene cognates.}\]
Bur. *bác-in > bácin (H, N) ‘shank; (animal’s) hind leg above the hock’
~ Cauc: Avar pürč:š ‘ham’, Chamali beču ‘thigh; knee (of animal)’, Tabasaran bac ‘paw’,
Tsezi besi ‘ist’, etc. < PEC *b³ʃe\^\, (NCED 291)
~ Basque *borc,:*bošt- ‘five’ < ‘*hand’ (cf. Tsezi ‘fist’, above).
§ Cf. PST *pùš(-s) ‘knee’ < PDC *b\,v\,st\,v (SCG 19-20).

Bur. *bimbal-, in bimbalten (Y) ‘ankle’ (old compound with *=ltën ‘bone’)
~ Cauc: PEC *b³ml\,v ‘hoof, foot’ (> Tsezi bula ‘hoof’, Chechen ber-g id., Avar mal ‘foot’, etc.: NCED 307) + PEC *\,q\,v\,\,w\,u\,\,\,l\, ‘on foot’, Arin pil ‘feet, legs’, etc.
~ Cauc: PEC *b³ml\,v ‘hoof, foot’ (> Tsezi bula ‘hoof’, Chechen ber-g id., Avar mal ‘foot’, etc.: NCED 307) + PEC *\,q\,v\,\,w\,u\,\,\,l\, ‘on foot’, Arin pil ‘feet, legs’, etc.
~ Basque *buľ ‘head’, *be=puru ‘eyebrow’ (*be(t)= *begi ‘eye’).
§ Bur. and Basque compared by Berger (1956, p. 9, note 16). PDC *burV ‘hair’ (Starostin 2005d, 2007 [TOB]).

Bur. *galgi > galgi (H, N, Y) ‘wing, fin’
~ Cauc: Lak qa ‘wing’, Lezgi qil hand’, Bezhta xaro ‘elbow’, etc. < PEC *gǐl\,i ‘wing, arm, elbow’ (NCED 895)
~ Basque *na=gal ‘wing’, *e=gal ‘wing, fin’ (northern Bsq hegal apparently influenced by *hega- ‘to fly’).

Bur. *yan > =yaan (H, N), =yan (Y) ‘heel’
~ Cauc: Avar erë ‘heel’, Lak niq:a ‘heel’, Chechen haq-olq ‘ankle, ankle-bone’, etc. < PEC *\,h\,g\,u\,l ‘heal; ankle, ankle-bone’ (NCED 248)
~ Basque *hoïn ‘foot’.

---

25 *c or /\,c/ in these and all Dene-Caucasian words henceforth, is consistently the sibilant affricate [\,t\,ʃ\,].
26 /\,=/ at the beginning of a morpheme indicates a bound morpheme that requires a possessive prefix (in Burushaski) or a class prefix (in Caucasian). Postulated fossilized prefixes (as in Basque) are also followed by this symbol, e.g. Bur. (H) a=ltën ‘my bone’, nu=ltën ‘her bone’; Avar w=as ‘son’, y=as ‘daughter’, b=as ‘young (of animal)’; Basque *bi=hoc ‘heart’, etc.
27 =wa is a common plural ending.
28 /\,s/ denotes a retroflex sibilant.
29 The PDC form would have to be *bur\,v, with a palatalized or ‘soft’ rhotic (SCP 60), otherwise the Bsq word would be *burru (*bururu), with a trilled rhotic.
30 Burushaski \,y/, written y and later g by Berger, is really a voiced uvular fricative /\,s/ and is part of the back-velar or uvular series with /\,q/ and /\,qh/ (the latter often realized as [\,k\,\,]) , though in realization the velars and uvulars overlap to some extent (Berger 1998 I, pp. 20-21).
§ Bur + Avar compared by Bouda (1950, §164). Bur + Basque compared by Berger (1956: 40). All < PDC *ʔawNaV 'heel, foot' (SCG 265).31 Bsq development could have been *ʔawNaV > *ʔiV > *hoin.

Bur. *hut > =hútes (Y), =útis, =út (H, N) 'foot'32

~ Cauca: Avar het 'e / het 'foot', Chechen t'a 'front leg (of animal)', Dargi Urakh, Sirgokala t”ah 'foot', etc. < PEC *twiV > *hútwV 'foot, forefoot' (NCED 1007).33

§ Bleichsteiner (1930, §23) compared Bur + Avar. Cf. PST *tiH 'heel, ankle' < PDC *ʔawNaV > *tiV (SCG 207), or possibly *ʔiV would better account for the Bur form.

Bur. *ken > =kin (H, N), =ken (Y) 'liver'34

~ Cauca: Andi k’'unu 'kidney', Chamali k’im, k ü, Beazha koma id., etc. < PEC *kunHV (NCED 728).

§ Cf. PST *qVnH 'kidney' < PDC *(x)kunHV 'kidney' (Starostin 2005d, 2007 [TOB]).

Na-Dene: Eyak ḡ’omar 'kidney, salmon roe'; PA *q’un? 'roe'.35

Bur. *kur- > kúur (H) 'finger-joint, toe-joint'; kurój (H) 'bone'

~ Cauca: Chechen k’urau 'bone (for playing dice)', Lezgi k’ur ‘hoof, leg (of animal)', Archi k’iri 'leg (of animal)', etc. < PEC *kwirV (NCED 736).

§ Cf. PST *k(h)rey 'foot' < PDC *kwirV 'foot, leg' (SCG 123); Na-Dene: PA *qe- 'foot', Eyak =qí- 'knee, salmon roe'; PA *q’un? 'roe'.35

Bur. *ltén > ten (Y) 'bone'; tan-c 'leg'; =ltín, tin (H, N) 'bone', =ltín-c 'leg'

~ Cauca: Avar it’'on 'groin', Archi it’'on-t’ol 'fingernail';36 Agul k’un 'ankle', etc. < PEC *ʔwVnHV (NCED 785).

§ Cf. PST *lay 'shin, ankle' < PDC *ʔwVnHV 'ankle, shin' (SCG 139-140).

Bur. *ltur > tur (Y), =ltúr, tur (H, N) 'horn'

~ Cauca: Avar it’ar, Chechen kur ‘horn', Lezgi firi ‘mane', etc. < PEC *ʔwirV 'horn; braid, mane' (NCED 771)

~ Basque *a=dar 'horn' (< *a=rder).

§ Bur + Bsq compared by Berger (1959, p. 34, note 57); Bur + PEC compared by Peiros (1988, §50). PDC *ʔwirV 'horn' (SCG 134-135).

---

31 *c or /k/ in PDC, PNC, PEC, and PY reconstructions represents a voiced uvular stop, though possibly realized as an affricate [kʰ], and reflexes in Cauca and Bur are often the simple fricative /k/ ~ /y/.32 Confusingly, *t or *l denotes a retroflex stop in Bur, but a glottalized stop /t'/ in PDC, PNC, PEC. Similarly all underdotted consonants in Bur are retroflex, but glottalics in PDC, PNC, PEC.

33 Metathetic variants are quite frequent in Cauca languages.

34 The notation /=/ indicates that the obligatory pronominal prefix, rather than the noun stem, bears the accent, e.g. (H) â=kin ‘my liver', gō=kin ‘thy liver', etc.

35 For typology of ‘kidney ~ liver' cf. Old Indic vrkka 'kidneys' > German Romani pukko 'liver, lungs, spleen, kidney', Oriya buku 'heart, chest, courage', etc. (CDIAL 12064).

36 Archi -t’ol < PEC *t’lV 'finger'.

37 The PNC/PEC sound *ʔ is a glottalized lateral affricate, alternatively written ‘tl’/, as in Navajo.
Bur. *melč > =mēlč (H, N, Y) ‘jaw, jawbone’
   ~ Cauc: Archi muč ‘nose, beak; peak, top’, Lak murč ‘lip, brim (of dish)’, Chechen mēč-ig ‘tip (of something) bent upwards’ < PNC *mēwīlčă ‘edge, tip’ (NCED 816).

Bur. *mes > =mis (H, N), =mes (Y) ‘finger, toe’
   ~ Yeniseian: PY *bes- > Ket bes-taq ‘index finger’ (compound with PY *to?q ‘finger’).
§ PDC *HmičV (SCG 77).

   ~ Dargi biq-ri ‘witness’, Avar muš ‘witness’ (<*nuqV < *nuqV), Chechen boq ‘true’ (adv.), Ingush boq ‘truth, rule’, etc. < PEC *wēmqV ‘witness, true’ (< ‘eye’) (NCED 1050).
   ~ Basque *moko ‘beak; extremity, point; face; façade’, etc.
§ Cf. PST *myVk ‘eye’ < PDC *wēmqV (SCG 216); Na-Dene: Tlingit wāq = wāq ‘eye’, PA *-na-wēg-ɔr ‘eye’; Eyak lāq ‘eye’ < *nāq. Proto-Na-Dene, like Proto-Yeniseian, probably had no initial *m-, which seems to have changed to *w- (see Sino-Dene, p. 225). For Bur & Bsq phonetic development (*wēn- > *mo-) cf. the next set (*wēn- > Bur & Bsq *mu-).

Bur. *mulur > =mulur (H, N) ‘nostril’
   ~ Cauc: Bezhta, Hunzib motl’ ‘beak’, Hinukh milt’u, Khwarshi matl’ u id., etc., Batsbi marlo ‘nose’, Karata butl’a ‘horn’, etc. < PEC *wēnlV (NCED 1041)
   ~ Basque *mutor ‘snout, muzzle’ < *murtu-r.

Bur. *muš ‘nose; end, edge’ > =mūš (H, N) ‘nose; end, edge’, =mūš-puţ ‘beak’; muś ‘end, edge’ (H, N), =mūpus ‘nose’
   ~ Cauc: Chechen, Ingush mus ‘ar ‘snout, muzzle, trunk’, Avar môts ‘u ‘teat, nipple; tip’, Lezgi mus ‘blade; edge, verge; narrow side of an object’, etc. < PEC *mḥērōţ “protruding part, point, edge” (NCED 811)
   ~ Basque *moñis > Gip musu ‘nose’; in other dialects: ‘snout, face, lip, kiss’, etc.
§ A different comparison, with PEC *mḥēqē ‘edge’ (Ingush mēqīz-arg ‘snout’, etc.) in Starostin (2007 = TOB).

38 /c/ represents a retroflex affricate.
39 “PY, however, has no initial *m-: in this position we observe *w- [in pronominal stems only], *p- [the general reflex] or *b- [from initial clusters *Hm- , *mH-]” (SCP 30-31).
40 For semantic typology of ‘witness ~ eye’, cf. Old Indic saksin- ‘witness’ : saksat ‘with the eye, clearly’ < sa- ‘with’ + akṣa- = aksi- ‘eye’ (Buck 21.23). That ‘eye’ is the archaic meaning is shown by the Eastern Dene-Caucasian (= “Sino-Dene”) words, cited above.
Bur. *p(h)at- > *pat (H), *phat (N), *p(h)at (Y) ‘side, flank’
~ Cauc: Lezgi p-ad ‘side’, Kryz badow ‘near’, Budukh bode ‘near’ < PEC *bVdV ‘side’;
  cf. Urartian beda ‘side; (postpos.) on the part of, by’ (NCED 315)
~ Basque *pata-ë ‘slope, rugged slope’ < *side’.
§ A root with two stops, showing probable assimilations and dissimilations: *patV ~
  *padV ~ *badV. Reconstructed as PDC *pVdV (SCG 165-166).

Bur. *phol- > phol-ţó (Y), phul-ţyú (H), phur-ţyú (N) ‘feather’ (compound with *yuy ‘hair’)
~ Cauc: Lak p ‘ihulli ‘feather’, Dargi (Akushi) paheka, Chechen pěla-g id., etc. < PNC
  *pVhV (NCED 879)
~ Basque *bilho ‘hair, mane’ (not derivable from Latin pilum).41
§ PDC *pVhV (SCG 166), or metathesized *pVh(V)hV.

Bur. *=qat > =qat (H), =qhat (N), =qet-arar (Y) ‘armpit’
~ Yeniseian: PY *qot- (~ *qot-) > Ket qota ‘in front, before’, Yug xút ‘(to the) front’.
  Arin un-kut ‘in front of’, etc.: adverbial development of the noun: cf. Eng. abreast, etc.
~ Cauc: Avar me-héd ‘brisket (chest of animal)’, Bezhta nađe ‘brisket’ < PEC *qVdV
  (NCED 897).
  čët ‘chest, breast’ (Boas), xe’-ka ‘beastbone’ (Leer 1993); Haida sq’ut ‘armpit’.

~ Cauc: Udi goš ‘behind’, a development of PNC *-VqV ‘behind’ (NCED 1026).
§ Comparison by Bouda (1954, p. 229. no. 28).

Bur. *=qhát > =qhát (H, N), =xát, =xat (Y) ‘mouth’
~ Cauc: Lak qít ~ q”it ‘Adam’s apple, beak’, Udi qirti ‘Adam’s apple’, Kryz qulut
  ‘larynx’ (< *qul-ut), etc. < PEC *qwiti (NCED 905).
§ PDC *qwiti ‘palate, mouth’ (SCG 172).

Bur. *=qhorpV (H, N), =xorpet (Y) ‘lung’
~ Cauc: Andi qon:ir ‘lung’, Tsezl jot’orí, Archi qur:t:ur-t:i, Dargi Chiragh qusara,
  Akusha qurhalá, qurhari id., etc. < PEC *qwolq(rV) / *qvolq(rV) (NCED 901)43
~ Basque *han spo > (Bzk) aspo, aspo ‘lungs; bellows’, only ‘bellows’ in other dialects.
  The Bsq word seems to be contaminated with *hané ‘ashes, dust’ (as if ‘dust blower’).
§ Bur. and Basque compared by Berger (1959: 21). The words in the three families have
  a certain resemblance to each other, but seem impossible to reduce to a single protoform.

---

41 Latin pilu would become Bsq *biru or *piru, according to the development of other Lat loanwords in
  Bsq, e.g. Bsq goru ‘distaff < Lat colu(m), etc. Further, the laryngeal cluster /lh/ in Bsq is explainable by
  PDC *pVi(V)hV. The IE status of Lat pilum is itself questionable.
42 -iŋ, -aŋ are frequent plural affixes, seen in several of these etymologies.
43 “The root is rather peculiar (although there are no doubts at all in its EC antiquity): it contains a very rare
  phoneme *-o- (which yields t-like reflexes in PTs. Lak. and PL, but s-like reflexes in And. and Darg.)”
  (NCED).
Bur. =sán > =sán (H, N, Y) ‘spleen’
~ Cauc: Avar c:in ‘gall, anger’, Tindi s:imi id., Archi s:am ‘gall’, Lezgi seb, Bezhta simo id., etc. < PNC *cwày̥mê ‘gall; anger’ (NCED 329)
~ Basque *beHa-sun ‘bile’ (apparently an old compound).§

Bur. *=sú[m] > =sú (Y: plural =símu), =súi (H, N: plural =sítmuc) ‘navel, umbilical cord’
~ Cauc: Dargwa Chirag zu ‘navel’, Khinalug c’un, Tindi c:\u0421, Lak c’un, Chechen c’on-ga id., etc. < PEC *ʒ̥n?u (NCED 1096).
§ Bouda (1950, §131) compared Bur + Lak, etc. PDC *ʒ̥n?u (SCG 249). Cf. Na-Dene: Eyak c’ar? ‘umbilical cord’ (< *c’a?); PA *c’e’q ‘navel’ (cf. Chechen c’on-ga < *c’an-k’u).

Bur. *=šVsVn > =šusun (H, N), =šesen (Y) ‘elbow’
~ Cauc: Udi sun ‘elbow’, Rutul sín ‘front part of leg’, Lak s:an ‘foreleg, paw, pad’, Avar san ‘organ, body part’, etc. < PEC *š̥n̥ (NCED 963)
~ Basque *š̥n̥-ko ‘leg, calf, foot’, *šan-ka-r ‘tibia, heel-bone’, etc.
§ Bouda (1948, §100) compared Bsq + Lak; Bouda (1950, §127) compared Bur + Udi. PDC *š̥n̥ (SCG 187-188).

Bur. *tal > tal (H) ‘stomach, belly’
§ Note recurrent correspondences of Bur initial *t- with Caucasian lateral affricates (*ɬ, *ɻ, *L) and Basque *t-: see the next two entries, Bur *hēn ‘bone’ and *ltur ‘horn’ (above), and the Phonology section of this paper. PDC *Hl̥alV ‘liver, belly’ (SCG 76).

~ Cauc: Avar t’imu ‘bottom’, Tindi hinlu, Bezhta ôl’o, Lezgi k’an id., Khinalug k’an-ik’ ‘under’, etc. < PNC *Hl̥ônû (NCED 590).

Bur. *tar[f-] ‘*skin’ > tar-'i (H, N, Y) ‘ bag made from animal hide (for containing fluids, or for rafts)’
~ Cauc: Avar t’le ‘color’ (< *‘skin’), Karata t’le ‘id., Dargi Akusha guli ‘skin, hide, sheepskin’, Tabasaran yal ‘bark, shell’, etc. < PNC *Lôlî i’*skin’ (NCED 789)
~ Basque *laru ‘skin, hide, leather’.
§ Bur and Bsq compared by Berger (1959, p. 26, note 34); Bsq and Avar compared by Bouda (1948, §86). PDC *Lûli ‘skin, color’ (SCG 130; I find the PST and PY

---

44 For *beHa-, cf. more distant DC relatives: PST *phe ‘spleen’ (Thankur =pay ‘liver’, etc.); PY *b[a]vI ‘kidney’ could be a similar compound < *b[a]vI ‘intestines’.

197
comparanda here very doubtful). But based on the evidence of outliers (Bsq & Bur) for internal *-r- the Cauc forms may have assimilated the laterals (*Lori > *Ldl), and there may have been vowel metathesis, ie. *Loru > Bsq *laru, *Lori > Bur *tar[i-]. Cf. a similar situation with Bur *ter ‘high pasture’ ~ PEC *Lwēlv ‘fence, yard’ ~ Bsq *lare ‘pasture, grassland’ (below).

Bur. *=úl > =úl (H, N, Y) ‘belly, abdomen, bowels’
~ Cauc: Tindi b=etl ‘stomach, rennet, abomasum’ (b= is a class prefix), Dargi Akusha =arg ‘stomach, inside’ (with changing class prefixes), Agul urav ‘rennet, abomasum’, Rutil yiriy id., etc. < PEC *=íraLV (NCED 670)
~ Basque *urdal ‘stomach’ (Bzk also ‘rennet; womb’).
§ PDC *=hrLV ‘stomach, belly’ (SCG 112-112).

Bur. *=yǎldi- > =yǎldir (H), =yǎldin (N) ‘the part of the ribs under the armpit; middle part of the breast’
~ Cauc: Tindi hatl ‘armful’, Chamali hatl ‘”, Hunzib hethu id., etc. < PEC *HeLV ~ *HeLV (NCED 558).
§ For the regular correspondence of Burushaski -lt- (-ld-) to Caucasian *-L-, *-L-, see the Phonology section of this paper. PDC *Hlv ‘armful, armpit’ (Starostin 2005d, 2007 [TOB]).

Bur. *jal-/*jal- > jalaalimitj (H. N) ‘long hair (of people)’, jalás (Y) ‘hairy’, jaléi, jalii (H) ‘beard (of goat)’, etc.45
§ PDC *žǎlh ‘mane, long hair’ (SCG 251).

Nature

Bur. *bay > bay (N), bangi (Y) ‘resin (of trees), pitch, gum’
~ Basque *muki > (Bzt) muki ‘gum, resin of trees’ (with regular change of *bVnk- > *mVk-).
§ PDC *bhünkV ‘pine tree, resin’ (SCG 13). Cf. Old Indic bhaṅgā ‘hemp, bhang, marijuana’ (a resinous plant) > Hindi bhāg, bhaṅg, etc., possibly one of the Burushic words acquired during the Indo-Aryan sojourn in the Northwest (CDIAL 9354; Witzel 1999, p. 4).

Bur. *bar > bar (H. N, Y) ‘small valley, ravine, gorge’

45 *j’/j’ represent a voiced retroflex affricate, alternatively written /dz/.
~ Basque *i=bar 'valley'.

§ Bur. and Basque compared by Berger (1956: 7; 1959: p. 28, note 39). Apparently an old DC word that was lost in Caucasian but preserved in the outliers.

Bur. *bun[dl]- > bun, (pl.) bundó (H, N) 'mountain pasture, mountain grove; boulder; wild, mountain-', bun, (pl.) bundó ~ bunjó (Y) 'boulder'

~ Caucasian mida ~ mda 'mountain' (cf. Georgian mta 'mountain')
~ Basque *mendi 'mountain'.

§ Bur. and Basque compared by Berger (1959, p. 28, note 41). Possibly another archaic DC word (cf. preceding set) that was lost in Cauc (except in the outlier Khinalug). Georgian mta is also isolated within Kartvelian and may be borrowed from archaic Cauc.

Bur. *dul-düm > dul'düm (H, N) 'rising cloud (of dust, smoke, etc.)'

~ Caucasian Archi dil' 'cloud', Lak?urlu 'cloud', Dargi Akusha dirix 'cloud', etc. < PEC *ñilõwV (NCED 400).

§ PDC *ñilówV 'dust, cloud' (SCG 40).

Bur. *ge > ge ~ gye (H, N, Y) 'snow'

~ Yeniseian: PY *go (~ *go) 'ice' > Ket qō 'ice', Yug xo id.

~ Caucasian Khinalug q:i 'cold' (n.), Lak -q-q:i- 'to grow cold, catch cold', Andi =e+-eš- 'to get cold, freeze', etc. < PNC *ñH sièA 'to freeze, get cold, be benumbed' (NCED 568).

§ PDC *ñH sièA 'ice, to freeze' (Starostin 2005d, 2007 [TOB]).

Bur. *yonder- > ýonderes ~ yondoles (Y) 'water, that runs over many stones'

~ Caucasian Botlikh kadaru 'stream, brook', Godoberi kada 'ravine', Lak atara 'mountain stream', Chechen Sowr-ät 'mountain stream (after rain or thaw)', etc. < PEC *ñHwadṼ (NCED 478).

§ PDC *ñHwadṼ (SCG 185). Bur -n- is unclear.

Bur. *hun > hun (H, N) 'wood, timber, beam, hewn trunk', hun (Y) 'wood, firewood'

~ Caucasian Chechen hun 'forest', Chadakolob xunå-q 'shady side, slope', Khwarshi hun 'mountain', etc. < PNC *ñanV 'mountain, hill' (NCED 425)

~ Basque: *oiham (=*i=ham) 'forest, woods; mountain; desert'.

§ Bouda (1950, §189) compared Bur + Cauc. PDC *ñwemṼ 'height, mountain' (SCG 232). This etymology exemplifies the common shift of 'mountain, hill' > 'forest, woods': cf. Old Slavic gorå 'mountain' ~ Lithuanian giria 'forest', etc. (Buck 1.22, 1.41).

Bur. *ltap > tap (H, N, Y) 'leaf, petal; leaf (page) of book'; =ñápewe- (H, N), =ñápi- (Y) 'to wither'

~ Yeniseian: PY *yäpe 'leaf' > Ket l:à, Yug l:hp, Kott dípi, etc. 'leaf'

~ Caucasian Chamali (dialect) łapä 'leaf', Lak ê'ap'i, Dargi Akusha k'api, Adyge thäp 'leaf', etc. < PNC *ñäpi 'leaf' (NCED 774)

~ Basque *lapa- 'bramble, thorn'.

§ Bouda (1950, §103) compared Bur + Cauc. PDC *ñápi 'leaf' (SCG 136). With the common correspondence of Bur *l-*, *=l-, *=l- ~ PNC *ñ-: see Phonology.
Bur. *mal > mal (H, N, Y) ‘field’
   < PNC *malwxë (NCED 795).
   § PDC *malwxë ‘pasture’ (Starostin 2005d, 2007 [TOB]). Indo-Aryan origin of the Bur
   word (Sanskrit māla- ‘forest near a village’, Prakrit māla- ‘garden’, etc. CDIAL 10088)
   does not seem likely.

   ~ Cauc: Chamali bat’a ‘faeces’, Tabasaran bat ur ‘dirt’, Ingush (dial.) bed ‘faeces’, etc.
   < PEC *phát ‘dirt, faeces’ (NCED 299).
   § PDC *phát ‘dirt’ (Starostin 2005d, 2007 [TOB]). For typology of ‘ashes ~ dirt ~
   excrement’ cf. Old Indic kṣaya ‘loss, waste’ > Panjabi, Hindi kheh ‘ashes, dust, rubbish,
   ordure’, etc. (CDIAL 3661); OI pānś ‘crumbling soil, dust, sand; dung, manure’, etc.
   > Bengali pās ‘ashes’, Awadhi (dialect.) pāsi ‘manure’, Hindi pās ‘dust, dung’, etc.
   (CDIAL 8019); and the set including Bur *ther ‘dirt’, below.

Bur. *phunc > phunc (H, N, Y) ‘dew’
   ~ Yeniseian: PY *pít ‘glue’ > Ket híʔt, Yug fíʔt ‘glue’, Kott fit ōginan ‘to glue’
   etc. < PNC *phunc ‘resin, juice’ (NCED 871)
   ~ Basque *pista > (Bzk) pizta ‘fresh rheum (secretion from eyes)’.
   § PDC *phunc ‘glue, resin’ (SCG 160-161). But surely the original PDC sense was
   ‘secretion, issue of liquid’, whether of human or animal body (> ‘rheum, sweat’), of trees
   (> ‘resin, gum, pitch’) > ‘glue’), or of nature (‘dew’ > ‘water’), etc.

Bur. *ši > ši (H, N) ‘fireplace, hearth’
   ~ Cauc: Ingush c i ‘fire’, Lak c ‘u, Avar c a, Godoberi c aiy, Bezhta c o, Abkhaz a = m ca
   ‘fire’, etc. < PNC *céuy ‘fire’ (NCED 354)
   ~ Basque *šu or *ši = ēu ‘fire’ > (c) su ‘fire’; Araban itsu arri ‘flint’ (‘fire-stone’).
   § Bur + Cauc + Bsq compared by Bouda (1950, §153); Bur + Bsq compared by Berger

Bur. *švre ‘*night’ > gó(i)n-sare (H, N), gón-šere (Y) ‘the whole night, all the night
   through’ (gón = ‘dawn’)
   ‘yesterday’, Ubykh šow ‘night’, etc. < PNC *śvěrV ‘evening’ (NCED 977).
   § PDC *śvěrV ‘evening’, but with a different Bur cognate in SCG (194).

Bur. *tay > tuy (Y) ‘branch, shoot’
   ~ Cauc: Hinukh n’iŋ ‘bough’, Avar t’eg ‘flour (made of young sprouts), oat flour’, Lak
   k’ury ‘sprout’, etc. < PEC *ʔorgv ‘sprout’ (NCED 780).
   § With the common correspondence of Bur *t- ~ PNC *ʔ-: see next three entries, and
   Phonology.
Bur. *tápi > tápi (H, N) ‘stone terrace’
~ Cau: Chadakolob *t’eb ‘millstone, whetstone’, Avar *t’eb ‘stone’, Chechen laba ‘shed; peak (of cap)’, etc. < PEC *t’eb ‘stone plate, shed’ (NCED 777)
~ Basque *tape > (Zub) tape ‘shelter under the eaves of a shed’.
§ PDC *t’eb ‘flat; slab’ (SCG 137). With the common correspondence of Bur *t- ~ PNC *t-: see Phonology.

Bur. *ter > ter (H, N, Y) ‘high pasture, summer mountain pasture’
~ Cau: Archi tol ‘yard, place in front of the house’, Avar lol ‘open enclosure (for sheep)’, Bezhta kalo ‘fence’, etc. < PEC *töl ‘enclosure, fence’ (NCED 791)
~ Basque *tare ‘grassland, pasture’.
§ Bur. and Basque compared by Berger (1959, p. 26, note 34). PDC *töl ‘fence’ (SCG 130, not including the Bur word), but this is virtually the same as the PEC form which, in comparison with the outliers Bur and Bsq, exhibits lateral assimilation (PDC *töl > PEC *töl). Cf. a similar situation with Bur *tar[i] ‘*skin’ ~ PEC *töl ‘*skin’ ~ Bur. and Basque compared by Bur and Basque.

Bur. *tis > tis (H, N, Y) ‘wind’
§ PDC *täč ‘wind’ (SCG 134). With the common correspondence of Bur *t- ~ PNC *t-: see Phonology.

~ Cau: Chechen tum ‘marrow; kernel (of a fruit, nut)’, Archi t'ummul ‘grape’, Abkhaz a-t'ama ‘peach’, etc. < PNC *tum ‘kernel (of fruit), seed’ (NCED 1004).
§ PDC *tum ‘kernel (of fruit), seed’ (SCG 205). Bur. *tumá-y and Lezgian *tumul ‘> Archi t’ummul, etc.) exhibit similar suffixation. Bur. *y is often of lateral origin (see Bengtson & Blazek 2011a).

Bur. *ther > ther (H, Y), ther-k (N) ‘dirt’
~ Cau: Akhwakh tere-ti ‘ashes, dust’, Tindi tira ‘dung (of dog)’, Bezhta tär ‘dung (of sheep)’, etc. < PEC *tär ‘dirt’ (NCED 993)
~ Yeniseian *dor- ‘powder, dirt’ > Kott tara ‘dirt’, etc.

~ Cau: Avar sár ‘branch, bough’, Tsezi atl’ir ‘pod’, Bezhta atl’alo id., etc. < PEC *sul ‘branch, pod’ (NCED 508)
~ Basque *adar ‘branch, knot (of tree), leg (of chair, bed)’ < *ardar: a homonym of *adar ‘horn’.
§ For the regular correspondence of Burushaski *-t- to Caucasian *-t- see the Phonology section of this paper.

Wild animals

Bur. *č(h)argé > čargé (Y) ‘flying squirrel’
~ Yeniseian: PY *sαq’a ‘squirrel’ > Ket saq’a, Yug saq’a / saq’x, Kott šaga, Arin sava, Pumpokol tak ‘squirrel’
~ Cau: Chechen šat’q’a ‘weasel’, Andi sart’u ‘weasel’, Tsakhur sok ‘weasel’, Adyge cəx’a ‘mouse’, etc. < PNC *cərgwV (NCED 322)
~ Basque *sagu ‘mouse’ > (c) sagu; in compounds: sagu-zahar ‘bat’ (‘mouse-old’); sat-or ‘mole’ (< *sag-t-hor ‘mouse-dog’), sat-itsu ‘shrew’ (< *sag-t-icu ‘mouse-blind’).
§ PDC *fu[č]ərgwV ‘squirrel, weasel’ (SCG 21). The original initial consonant is uncertain, probably altered here or there by expressive changes (cf. Bsq sagu / sagu/ ‘mouse’ ~ xagu / sagu/ ‘mousie, wee mouse’). The meaning ‘squirrel’, only in Bur and Yen, may be evidence for their common development in a “Burusho-Yeniseian” subgroup of DC. (Cf. Bur *khen ‘lea’ ~ PY *qəʔi ‘lea’, below.)

Bur. *chen > čhin (H, N), cən (Y) ‘small) bird’
~ Cau: Abaza c′i-s ‘small bird, sparrow’, Avar hinč ‘bird’, Dargi Akusha humuc ‘eagle (poetic)’, Chechen hōza ‘sparrow’, Archi noč ‘(small) bird, sparrow’, etc. < PNC *fimic(w)I ~ *fi[w]im ‘small bird’ (NCED 525)
~ Basque *hunc ‘owl’.
§ The comparison assumes metathesis: PDC *fim[w]I ~ *fi[w]im ‘bird’ (Starostin 2005d, 2007 [TOB]), apparently also *fi[w]im[č] to account for Dargi humuc ‘eagle’ and Bsq *hunc ‘owl’.

Bur. *yórk- > yürqun (H), yürqec (N), yórken (Y) ‘frog’
~ Cau: Tindi q′orq′u: Khinalug q′urq′or, Khwarshi q′urq ač ‘lizard’, Kabardian (handar)-q′: “aq′a ‘frog’, etc. < PNC *qarkVrVqV (NCED 942).
§ Bleichsteiner (1930, § 21) compared Bur + Cau. Obviously of onomatopoeic origin, but Bur and Cau forms exhibit precisely matching forms. PDC *xəq(r)VrV(qV) ‘frog’ (SCG 243).

Bur. *har > har (H, N) ‘corn worm, grain weevil’
~ Cau: Avar hapára ‘worm’, Andi hapara, Tsakhur qbra-wuče id. < PEC *fiabarV (NCED 508)
~ Basque *ha[m]ar ‘worm’ > Rnc ār, Lap har, archaic Bzk haar ‘worm’, etc.

46 I have hypothesized that Bsq *-t- in combinatory forms is a relic of a PDC oblique stem marker = PNC *-dV-.
Bur. *haúlal > haúlal ~ ahúlal (Y), hoólal-as (H, N) ‘butterfly, moth’
~ Cauk: ? Archi hiluku ~ hiliku ~ hikku ‘fly’ (insect)
~ Basque *euti ‘fly’ (insect).
§ Bur + Basq compared by Berger (1956: 16), citing Zarubin’s transcription of Yasin ahúlal. The Archi word hil(i/u)ku ‘fly’ is totally isolated within Cauk, thus dubious.

Bur. *khen > khin (H, N), khen (Y) ‘flea’
~ Yeniseian: PY *qohi ‘flea’ > Ket qohi ‘flea’, Kott imgara-xon ‘flea’ (imgara- ‘little’), xon ‘beetle’
~ Cauk: Chechen sëw-ig ‘louse’, Ingush xon-g id., Dargi q ‘i nít, Lak q’umar q’ay-sáta: ‘worm’ < PEC *qηñ?V ‘louse, nit; worm’ (NCED 911)
~ Basque *a=kain ‘tick’.
§ PDC *xqân?V ‘louse, flea’ (SCG 236). The meaning ‘flea’, only in Bur and Yen, may be evidence for their common development in a “Burusho-Yeniseian” subgroup of DC. (Cf. Bur *č(h)arge ‘[flying] squirrel’ ~ PY *sa?ya ‘squirrel’, above.)

~ Cauk: Chechen polla ‘butterfly’, Andi pera ‘bee’, Khwarshi par ‘bee’, etc. < PEC *pdrV (NCED 875); reduplicated as Andi pirmpa ‘butterfly, Udi pôpô-k id., etc. < PNC *pdrVpôlV (NCED 867)
~ Basque *pinpilin > High Navarrese pinpilin ‘butterfly’, Lapurdian pinpirin id.
§ PDC *pôrV (SCG 162).

Bur. *phen > phin (H, N), phen (Y) ‘fly’ (insect)
~ Cauk: Avar piq:na ‘drone’, Dargi Akusha mirqi ‘bee’, Chechen, Ingush niq ‘beehive’, etc. < PEC *pôrV (NCED 875); reduplicated as Andi pirmpa ‘butterfly, Udi pôpô-k id., etc. < PNC *pôrVpôlV (NCED 867)
§ PDC *pinogwV ‘bee’ (SCG 159). Presumably the Bur development was something like *pôñK > *phen > *phen.

Bur. *Qariuyo > qariuyo (H), yariuyo (N) ‘heron’
~ Yeniseian: PY [*gwirirak] > Kott kurirax, pl. kuriragan, kurirakj ‘crane’
~ Cauk: Andi q’:urru ‘crane’, Karata q’:urru-n, Adyge q:carav id., Dargi Akusha q’angq ‘heron, bustard’, Chechen karul ‘crane’, Lak q’urug id., etc. < PNC *gôrśgjwV (NCED 914)
~ Basque *kurí ~ *kuri ~ *kuri(i)-lo ‘crane’.48
§ PDC *xqôrśxjwV (SCG 238), with reduplication; but some forms point to a simplex such as *xqôrV (Basq *kuru ~ *kuri; Andi q’urru), or to a lateral suffix (Basq *kur(i)-lo;

47 /q’/ represents a pharyngealized + glottalized uvular stop/affecticate.
48 This cannot be of Romance origin, for phonetic reasons; in fact Spanish grulla seems to have been influenced or blended with Basq (Bzk, Gip) kurrillo ‘crane’. (Other Romance languages have no counterpart to the -lla suffix in this word.)
Chechen karsu-li; Bur *Qarǎn-yo). Bur /y/ is of lateral origin (see Bengtson & Blázek 2011a).

Bur. *tal > tal (H, N, Y) ‘dove’
~ Cauc: Avar t’il ‘i’a kind of songbird’, t’il: ‘i-diro ‘a kind of variegated bird’, Budukh kak-il ‘partridge’, Lezgi k’ek ‘cock’, Khinalug k’ak’-id ‘partridge’, etc. < PEC *lele ‘a kind of bird’ (NCED 776).
§ PDC *lele(w)je (SCG 136-137). Note recurrent correspondences of Burushaski *t- with PDC lateral affricates (*l, etc.): see Phonology section of this paper.

Bur. *tur- in tur-cun (Y), tur-sún (H, N) ‘marmot’ (compound of *tur- + obscure second element)
§ PDC *LārHV ~ *lārHV (Starostin 2005d, 2007 [TOB]). Marmot and hare are both rodents. See Phonology section for initial correspondence.

Human relations

Bur. *=cu > =co (N, H), =’cu (Y) ‘brother (of a man) / sister (of a woman); husband of a man’s sister’
~ Yeniseian: PY *h[i]s ‘brother, sister’ > Ket bišeʔp ‘brother, sister’, Yug bišrʔp, Pumpokol bič id., Kott popēš ‘brother’, popēša ‘sister’50
~ Basque *an-his-ba ‘sister (of a woman)’ > (c) ahizpa, aizpa, (Bzk) aizta, (Zub) dhizpa, (Rnc) aizpa.

Bur. *yul > yul (H, N) ‘grudge, enmity, hatred’
~ Yeniseian: PY [*qor- ~ *xor-] > Yug xviř ‘bad tempered, angry’
~ Cauc: Avar ā’el ‘gossip, rumor; abuse’, Khinalug qol ‘offence’, Chechen qel ‘(legal) sentence, fate’, etc. < PEC *gwālho (NCED 465)
~ Basque *bilhau ~ *bilahu ‘curse, blasphemy, hate’ (< *g’ilhau, etc., a regular change).
§ PDC *gwālho ‘bad, angry; anger, quarrel’ (SCG 55-56). An interesting ‘negative emotion’ etymology.

49 Ubykh /h/ is a voiced lateral fricative, shifted to a sibilant /z/ in Abkhazian languages.
50 Here the PDC I-class (masculine) form, corresponding to PNC *u=ici ‘brother’, has been generalized for both ‘brother’ and ‘sister’.
51 ‘Two basic original forms must be reconstructed as *u-ici(jV) ‘brother’ (with frequent further development > *čivijV or *čewijV; exactly this form is reflected, e.g., in PL and PWC), and *j-ici(jV) ‘sister’ (sometimes with a similar contraction > *čivijV) (NCED). Thus, for example, Agul ču ‘brother’ / či ‘sister’, in which the gender element has been transferred from the former prefix to the stem vowel. In Agul the old class system is no longer grammatically productive, so the difference between ču ‘brother’ and či ‘sister’ has become purely lexical.

~ Cau: Andi, Chamali, Tindi *hekʷ* ‘a ‘man, person’, Inkhokwari *hik*‘o ‘man, person’, Hinukh *rekʷ* id., etc. < PNC *HirkwE ‘man, person’ (NCED 579); cf. Chechen *ëra ‘ungelded*, Ingush *äri* id., Akhwakh *b=el:o ‘male*, Lak *b=urx-ni-s:a ‘male* (the latter two with class prefixes), etc. < PEC *PirglV ‘male’ (NCED 210)

~ Basque *ar* ‘male’.

§ PDC *Hir(Y)kwE ‘man, male’ (Starostin 2005d, 2007 [TOB]: only for Bur *hur-ik-* and PNC *Hir(-)kwE). PEC *PirglV* and Basq *ar* may represent an unrelated root.

Bur. *s* > *is* (Y) ‘child; (animal’s) young’, *s-k* (H, N), *s-ko* (H) ‘(animal’s) young; (jokingly) human child’

~ Cau: Avar *w=as ‘son*, y=as ‘daughter*, Bezhta *ožo ‘son, boy*, Lak *ars ‘son*, Dargi *urši ‘son*, Khinalug *ši ‘son*, ri=ši ‘daughter*, Kabardian *ša-wa ‘son*, etc. < PNC *=iswE ‘son* / ‘daughter*, with changing class prefixes (NCED 671)

~ Basque *-sV- [element in kin terms], e.g. (c) *se-me ‘son*; (Bzk) *osa-ba ‘uncle*, gura-so ‘parent*; *osa-ba ‘ancestor*; (Zub) *osâ-ba ‘uncle*, alhaba-so ‘grand-daughter*, sênmê-so ‘grandson*, iloba-so ‘grand-nephew*, bûrû-so ‘ancestor*, etc.

§ PDC *=iswE ‘son, child’ (SCG 113). Cf. Na-Dene: PA *=yaʔzC ‘small, woman’s child’ (Navajo *yaazh ‘baby [woman speaking]*, *yâsh ‘little one*, *yâzhi ‘little, small; young [of animal]*, etc.); Eyak *yâhš ‘child (of a female)*; PST *=šū ‘grandchild*.

Bur. *ses > sis* (H, N), *ses ‘persons, people*


§ Peiros (1988, §38) compared Bur + Cau + Yen. PDC *ćwëjo ‘man’ (SCG 30), reduplicated in Bur and Ubykh.

Bur. *ţâm > jaám* (H, N), *jâm* (Y) ‘(distant) kinsman, relative’


§ Peiros (1988, §63) compared Bur + Cau. PDC *ţâmV / *mâţV ‘relative’ (SCG 251).

**Descriptives**


~ Cau: Andi *=ltl:in- ‘to put on (shoes, footwear, trousers), Akhwakh *itl:e-l ‘stocking*, Tsezi *=ltl- ‘to put on (shoes)*, Archi *=ubtla-s ‘to put on (trousers)*, etc. < PEC *=omLV ‘to put on (trousers, shoes)* (NCED 861)
~ Basque *ortüc ‘1 barefoot; 2 to take off (shoes, stockings)’ > Zub ortüts 1, ortüts-2, BN ortüts 1, orsüts-(tu) 2, Bzk ortoz 1, etc.\(^{52}\)

§ Berger (1959, p. 27, note 35) compared Bur hulás, hollást + Bsq orths, ortoz ‘barefoot’. PDC *lu:nV ‘to put on (shoes)’ (SCG 130), based on a somewhat dubious PST *lu:nH ‘a kind of shoe’,\(^{53}\) altered to the Western DC verb *=ōmLV by a frequent process of metathesis and syllabic reduction, caused by extensive prefix- and suffixation (Starostin 2005a, p. 1). Overall an interesting DC etymology with precise phonetics and semantics.\(^{54}\)

Bur. *Qaqáv- > yaqáv-um (H, N), qaqá-m (Y) ‘bitter; unsweetened; sour’ (-um is a frequent adjectival ending)
~ Yeniseian: PY *qVqVr > Ket qa:l; qo’lin1 / qo’lay5 ‘bitter’. Yug x.xu6; x.xilay6, Kott ogar id., Pumapokol leo-xóxor ‘bladder’ (‘gall-bladder’?)
~ Cauca: Chechen q’āha ‘bitter’, Bezhta n= iq’aro, Archi q’aal id., Khinalug q’al ‘bitter’, q’il-ez ‘salty’, Ubykh q’aq ‘sweet’, etc. < PNC *qêhl光伏发电 ‘bitter’ (NCED 912)
~ Basque *keru ‘stench; rancor’; *kerac ‘bitter; foul-smelling’.

~ Cauca: Chamali s’ik’u- ‘sour’, Khwarshi caqun, Lak c:ixku-, Archi c’eg”-du ‘rank, bitter’, etc. < PEC *câxkwV ‘sour, raw’ (NCED 356)
~ Basque: (with metathesis *cʌxkwV > *xkwâcV) *gasí ‘salty’, (dialed.) ‘acidic; bitter’, *gastana ‘cheese’, *gac ‘salt’.
§ PDC câxkwV ‘sour; bitter’ (SCG 24), with metathetic variant *xkwâcV to account for Bsq *gac / *gas- / *gast-; Na-Dene: Eyak c’i:k’ ‘bitter’; PA *-ɛ’i:k’ ‘to sting, smart; be peppery, bitter’ (Leer 1993).

Bur. *tharén- > tharén-um (H, N) ‘narrow, cramped, tight (of clothes)’
~ Cauca: Avar t’eréna- ‘thin’, Karata =et’ara-, Dargi Akusha b=uk’ula, Khinalug k’ir, id., etc. < PNC *=išlfV ‘thin’ (NCED 639)
~ Basque *hirain ‘slender, svelte, fithe, agile, graceful’.

**Cultural vocabulary:** domestic animals

\(^{52}\) Basque *-rt- is the normal correspondence to Bur *-rt- and PNC *-r-, *-f-, *-l-, when the following vowel in PDC is accented, e.g. Bsq *urte ‘year’ < PDC *VwV ‘last year’ (SCG 259).

\(^{53}\) Based only on Tibetan lham ‘boot, shoe’, Old Chinese 鞋 *loj? ‘shoes for criminals with lopped toes’ (‘shoes [with lopped toes] for criminals’).

\(^{54}\) It is interesting that this verb is also found in Na-Dene: e.g. Navajo =tlé, =tlé, in, isté, =stlé ‘socks, stockings, leggings’.
Bur. *aćás > aćás (H, N, Y) ‘sheep, goat, sheep and/or goat(s) = Kleinvieh, small cattle’
~ Cauc: Adyghe āča ‘buck goat’, Dargi Akusha šeža ‘goat’, Chechen avst ‘goat (about 1 yr. old)’, etc. < PNC *pēyēwē (NCED 245).
§ PDC *pēyēwē’goat’ (SCG 264). Note, however, similarity, even in accent, with Old Indic ādā- ‘goat’ (PIE *aː-g), raising the possibility of borrowing, but this word is known only in eastern (Satem) IE languages. The frequent semantic variation ‘sheep’ ~ ‘goat’ recurs in several of the comparisons below.

Bur. *buč > buc (H, N) ‘(ungelded) male goat, 2 or 3 years old’
§ PDC *?ey[w] ‘goat’ (SCG 264). Note, however, similarity, even in accent, with Old Indie aja- ‘goat’ (PIE *ag-), raising the possibility of borrowing, but this word is known only in eastern (Saments) IE languages. The frequent semantic variation ‘sheep’ ~ ‘goat’ recurs in several of the comparisons below.

Bur. *chigir > cigir (Y), chigir (N), chiir (H) ‘(nanny-)goat’, also ‘female ibex’
~ Basque *siki-ro ‘gelded ram’, *siki-te ‘gelded goat’.
§ PDC *sdi[k]V ‘goat’ (SCG 187). Note the similar suffixes in Bur *chigir-r, Andian *či-ki-r / *čiki-r and Bsq *siki-ro.

culdár (Y) ‘bull’, chindár (H, N) id.
~ Cauc: Andi č’ora ‘heifer’, Tindi č’ara, Agul luč ‘id., Chechen čša ‘calf’, Ingush časa id., etc. < PEC *HcwilaV / *Hlicwiva ‘heifer’ (NCED 556)
~ Basque *čahal > (Zub) xahal /sahal/ ‘calf, heifer’, (Bzk) txaal /caal/ ‘calf’, etc.
§ PDC *HçjwivaV ‘bull, heifer’ (Starostin 2005d, 2007 [TOB]); with metathesis *[ç]VHlaV or *[ç]HIVaV to account for Bsq *čahal.

Bur. *du[m] > du (H, N, Y), diđo (H) ‘kid, young goat up to one year’ (< *diš)
~ Cauc: Andi dam ‘sheep, ewe’, Chechen iš ‘ram’, Lak i’a ‘sheep, ewe’, Kabardian i’o ‘ram’, etc. < PNC *dwanV ~ *dwanH ‘sheep, ram’ (NCED 405)
~ Basque *i=di > (c) idi ‘bull’

---

55 PDC *sd has reflexes similar to PDC *y [d] in Bsq (*s) and PNC (*y), but the reflexes differ in Bur and PY: PDC *sd > Bur *ch, PY *t, but PDC (*y) > Bur *s, PY ‘*y or *s, with distribution yet unclear’ (SCP, pp. 53, 75-76). Since the Yeniseians did not practice animal husbandry (apart from some reindeer herding at a late date) and cereal cultivation, PY lacks cognates for many of the cultural words discussed here.

56 Loss of PDC nasals (via nasalized vowels) is frequent in Bur (cf. convergent cases in Cauc and Bsq), with evidence for the original nasal often preserved in inflected forms. In this case *du[m] is hypothesized based on (H) diđo ‘kid’, pl. diđomuc.

57 For semantic typology, cf. Welsh dafad ‘sheep, ewe’ ~ OIr dam ‘ox’, Gk δομάδ ‘young ox’ (< ‘tamed animal’). Buck 3.25. The loss of PDC nasals (in clusters of the type *-nH-) is also frequent in Bsq.
§ A different comparison of Bur *du[m] (with PNC *sw̃n?V ‘lamb’) was preferred by Starostin (SCG 191).

Bur. *dágar > dágar (N) ‘ram’

~ Cau: Avar de’én ‘billy-goat, buck goat’, Hinukh t’eq “i ‘kid (about 1 year old)’, Inkhokwari t’iq o, Bezhta t’iq ‘i id., etc. < PEC *dVr̥jwV ‘billy-goat’ (NCED 403).

§ PDC *dVr̥jwV ‘male animal’ (SCG 43). Initial retroflex /d/ in Bur apparently conditioned by following /r/.

Bur. *hálgi-t > élgit (H, N), hálkít (Y) ‘(female) goat, over 1 year old, which has not given birth’*

~ Cau: Agul, Tsakhur urg ‘lamb (less than a year old)’, Rutul urg ‘yearling sheep’, Chamali barg “a spring-time lamb’, etc. < PEC *?wilgi (NCED 232).

§ PDC *?wilgi ‘lamb, kid’ (SCG 260). Perhaps *?olgi is better, since the PEC form may incorporate a class prefix: *w=?olgi > *?wilgi, not uncommon in PNC/PEC forms.

Bur. *huyés > huyés (H, N, Y) ‘Kleinvieh, small cattle, sheep and/or goats’

~ Cau: Avar i’ ‘flock (of sheep)’, Lak yar-u ‘flock (of sheep)’ < PEC *hV?V ~ *hVNHV ‘flock’ (NCED 532).

§ PDC *hV?V ‘sheep, small cattle’ (Starostin 2005d, 2007 [TOB]); however there is a note that Bur *huyés “should be moved to [PDC]*hV?V, i.e. *hVNHV ‘a kind of deer’ (PST *y’ ‘doe, antelope’; PY *?y’ ‘elk, deer’), which seems less plausible to me. Cf. also PIE *Howi- ‘sheep’.

Bur. *thugar > thugar (H, N) ‘buck goat, billy-goat’

~ Cau: Karata t’uka ‘buck goat’, Tindi t’uka ~ k’uta (in free variation), Bezhta t’iga, Kabardian da’ga ‘buck goat’, etc. < PNC *f̥g̥̃V (NCED 1003).

§ PDC *f̥g̥̃V ‘goat’ (Starostin 2005d, 2007 [TOB]). The Bur initial *t- (unaspirated) would be expected. The aspirate seems to be the result of a kind of “Verner’s Law” effect, caused by accent on the following syllable (cf. Bur *tharén- ‘narrow’, above).

Cultural vocabulary: dairy


---

58 TOB adds note: “Could be derived from *halk- ‘to bear young’ (not quite plausible semantically, however); or could be secondarily contaminated with this root in Yas[ian].”

59 “An exclusive Avar-Lak isogloss; very unsecure in what concerns PEC reconstruction” (NCED).

60 “Evidence for dairying activities in prehistory can be assessed by the detection of dairy fats associated with archaeological pottery … This has revealed that milk was being processed in the northwestern part of present-day Turkey … as early as 8500 years … BP” (Gerbault et al. 2013). This is well within the postulated time range of the Western DC family (see Postscript).
MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • Issue XIX • 2014

~ Basque: *sen-bera* ‘soft cheese, cottage cheese’ > Salazarese zenbera, etc.
§ PDC *H[∫]ē̂mi* ‘butter, tasty food’ (Starostin 2005d, 2007 [TOB]).

Bur. *čháo > čháo (+ verb ʔ- ‘do, make’) (H, N) ‘to milk’
~ Cau: Lezgi ʔ-ac-a ‘to milk’, Archi =ac-a, Dargi Akusha =iz-es, Lak t-,izi-n id., Ubykh ʔwa ‘to drink’ (< *to drink milk’ < ‘to milk’), etc. < PNC *=āmʔU ‘to milk’ (NCED 262).
~ Basque *e=aici* ‘to milk (a cow)’ > Zub jaitzi, Goizueta (Navarre) jetzí, etc.
§ PDC *=āmsdʔU* ‘to milk’ (SCG 4). There are several nominal derivatives, such as Tindi, Karata zini ‘cow’ = Basque *sesen* ‘bull’.

Bur. *dilta-r > dilta* ‘buttermilk’ (H, N, Y)
§ PDC *rhuʔV* (SCG 183-184). For the regular correspondence of Burushaski -lt- (Id-) to Cau *-l- [tl], see the Phonology section of this paper. PDC *r-* > Bur *d-* is also regular (SCP 41).

§ PDC *[m]hɛʔV* ‘milk, butter’ (SCG 146). The *n- in Cau is still unexplained: possibly by contamination with PNC *rɛŋʁV* ‘butter’ (NCED 948), which has n- initials in several languages (Avar nɛ ‘butter, oil, fat’, Lak nah ‘butter’, Dargi Akusha nerɛ ‘melted butter’, etc.).

~ Yeniseian: PY *deʔn* ‘milk, nipple’ > Pumpokol den ‘milk’, Arin tęŋ-yul ‘milk’, Kott ten ‘nipple’ (with regular change of PDC *š-* > PY *d-: see SCP 63)
~ Basque *e=šene* ‘milk’ > esene, esne, ezne; *seu-(bera) ‘soft cheese’ is a distinct etymon (see Bur *čeau-).
§ PDC *šeğʔW* ‘milk, nipple’ (SCG 195-196).

Cultural vocabulary: the horse

This topic is rather difficult. Archeological evidence indicates that horses were domesticated on the steppes of Central Asia (now Kazakhstan and Russia), certainly by 3500 BCE and possibly as early as 4500 BCE (Outram, et al. 2008; Anthony & Brown 2011). However, according to recent glottochronological results of the Moscow School the “break-up of the North Caucasian–Basque and Yenisseian–Burushaski branches [of Western Dene-Caucasian took place in] the second half of the 9th millennium BC” (Kassian 2010a: 430), thus about four millennia before domestication of the horse. Of course, the horse existed as a hunted wild animal long before that. In the 2001 version of
this paper and others I have suggested the comparison of Bur *hayúr ‘horse’ with PNC *farné ‘horse’ (Adyge fəra, Kabardian xəra ‘thoroughbred horse’, Lezgi č̣ar ‘mare’, Khwarshi xaram ‘foal’, etc.: NCED 425) and Basque *behor ‘mare’, but the phonetics involved in this are quite problematic. Berger (1998: 185) notes the resemblance of Bur *hayúr to Turkish aigir ‘stallion’, which seems to be a plausible source, since the Turks have a long history of mounted nomadism in central Asia.

However, the words for ‘saddle’ in Caucasian and Burushaski are very similar and fit the correspondence of Bur *t-/ *lt- to PNC *l (and other lateral affricates) that recurs in several etymologies discussed here (see Phonology).

Bur. *ltul[i] > =ltul (H, N, Y) ‘to saddle (a horse), prepare mount’, tiliaq (H, N)
tilihan, telêhay (Y) ‘saddle’ (noun, with frequent plural morpheme -aj)
~ Cauc: Avar tl’ilì ‘saddle’. Akhwakh (with assimilation) tl’:etl’:e, Andi tl’:iru,
Lak k’ilì, Dargi Akusha gili. Sirgokala, Tsudakhar guli ‘saddle’, etc. <
PEC *Iwilè ‘saddle’ (NCED 783).
§ Peiros (1988, §65) compared Bur + Cauc.

Of course, the phonetic correspondences do not necessarily prove that the Bur and Cauc words are “genetic” cognates; they could be early loanwords in either direction, or from an unknown source, which I have not been able to track down. Note another interesting comparison involving equids:

Bur. *chardV> charda (H, N), cardé (Y) ‘stallion’
~ Cauc: Abkhaz a-cada, cada ‘ass. donkey’, Adyge šod, Kabardian šod, Ubykh
cado id. (Čirikba 1996: 314)
~ Basque *a=sto or *ar=sto ‘donkey, ass’ > (c) asto, (Zub-archaic) arsto
‘donkey, ass’ (with usual syncopation of *-sto > *.sto).

The domestication of the donkey is also much later than the postulated breakup of Western DC (Marshall & Weissbrod 2011), but the possibility remains of a designation of the wild ass or other equid. Berger (1959: 32, note 55) pointed out the unlikely similarity of Bur jakun ‘ass’ (with retroflex initial = [dz]) and Hausa zaki ‘ass’, pi. zakiına, which, if genuine, would have to be traced to trade contacts.

Cultural vocabulary: cereal cultivation and processing

Bur *bay > bay (H: double plural baçény), bay (N: double plural báyignty), ba (Y) ‘(small-grained) millet’ (Panicum miliaceum)
§ PDC *bolcwi ‘millet, rice’ (SCG 15: also including PST *phr(e)s ‘rice’). Bur *ɣ often derives from PDC laterals or lateral clusters (see Bengtson & Blažek 2011).

Bur. *čha > čha (H, N), ča (Y) ‘millet’ (Setaria italica)
~ Cauc: Bezhta će ‘a species of barley’, Andi ćor ‘rye’, Kryz ćeľ ‘roasted grain’, etc. < PEC *ć(e)hlV ‘a kind of cereal’ (NCED 384).
§ The loss of PDC *-hl- in Bur is difficult to verify. In NCED this cluster is only found in this entry and *kahlV ‘a kind of berry’, the latter with no known Bur cognate. Starostin (2005d, 2007) preferred a different comparison, with PST *Čiď ‘hemp’, but also citing Proto-Nakh *ća ‘straw’ (cf. NCED 978) and Budukh ćeľ ‘roasted grain’ (cf. Kryz ćeľ id., with glottal initial, above).

Bur. *daltán- > daltán- (N) ‘to thresh (millet, buckwheat)’ (< *rVLV-)<
~ Cauc: Ingush ard-, Batsbi arl- ‘to thresh’, Tindi ćeli- ‘to thresh, roli ‘grain ready for threshing’, Bezhta ćol- ‘to thresh’, Archi tlorom ‘threshing board’, etc. < PEC *=VrLV ‘to thresh’, *r=ćelő ‘grain ready for threshing’ (NCED 1031)
~ Basque *larain ‘threshing floor’.
§ PDC *rVLV ‘to thresh’ (SCG 182) ~ *LVrV (> Bsq).

Bur. *darć > darć ‘threshing floor, grain ready for threshing’
~ Cauc: Lak t:arac’a-lu ‘threshing floor’, Dargi darac, Tabasaran rac,; Lezgi rat, Andi hinc :u, Bezhta ćid id., etc. < PEC *hrőřů (NCED 503).
§ Bouda (1950, §4; 1954, p. 228, §4) compared Bur + Lak. PDC *hronsdu (SCG 100). Bur development is similar to those of eastern Dagestan (Lak, Darg, Lezg), with *r > *d (SCP 41).

~ Cauc: Agul q’ir (diaI. q’ur) ‘grain’, Rutul q’ir ‘winter wheat’, Udi ar-um ‘wheat’ (< *?ar- < *q’ar-), Dargi q’ar ‘grass’, etc. < PEC *q[ʃ]rV ‘a kind of weed, (wild) cereal’ (NCED 915)
~ Basque *gar- in *(gara-)gar ‘barley’: a compound with *ga[l]i ‘wheat’. 61
§ Berger (1998: 161) notes the similar Tibetan word, gro ‘wheat’. Starostin (SCG 243) compares instead Tibetan khre ‘millet’, etc. < PST *khriśH ‘a kind of grain’.

Bur. *hars > hars (H, N), hars, hasč (Y) ‘plow’
~ Cauc: Akhwakh šerc-e ‘wooden plow’, Godoberi rec:i, Lak qa-ras id., Chechen āsta ‘hoe, mattock; plane (for woodworking)’, etc. < PNC *Hrőęč ‘wooden plow; mattock’ (NCED 601)
~ Basque *haincu- > Basque (Lab) haintzur ‘hoe’, (Rnc) aintzur ‘(heavy) hoe’, (Zub) háitzur ‘pickax’, (Bzk) atxur ‘spade’, (Gip) aitzur id., etc.

Cultural vocabulary: other artifacts:

61 In the 2001 version of this paper I compared Bur *gur ‘wheat’ with PEC *gölče ‘wheat’ and Bsq gari (*gali) ‘wheat’, but now the SCG comparison seems preferable phonetically.
   notl : ‘house, room’, etc. < PEC *balti ‘house’ (NCED 312).
§ PDC *balti ‘house’ (SCG 15). With the frequent correspondence of Bur *大批 ~ PDC
   *大批- (see Phonology).

Bur. *cháyur > cháyur (H, N) ‘chest or box for grain or meal’
   Cauc: Avar carür ‘corn bin, barn’, Chechen cjar ‘penthouse’, Khinalug cuqa ‘shed,
cattle-shed’ < PEC *cVGVrV ‘shed, penthouse’ (NCED 328).
§ Bouda (1950, §121) compared Bur + Avar, a resemblance also noted by Berger (1998
III: 73).

Bur. *khor > khor (H, N) ‘large wicker basket’
   ~ Cauc: Akhwakh koro ‘trough, gutter’, Andi koru ‘a kind of jar’, Lezgi k:ar ‘a big jar
(for carrying water), Tabasaran, Agul g:ar id., etc. < PEC *kworV (NCED 706).
§ PDC *kworV ‘a kind of vessel, scoop’ (SCG 118).

Bur. *tókur > tókur (H, N) ‘wooden chest (for grain, etc.)’
   ~ Cauc: Ingush t:aq ‘tub (for cheese brine)’, Dargi Akusha t:aq ‘a hive’, Agul t:ak id.,
   Rutul t:ak ‘baskct (for berries), etc. < PNC *t:aqV ‘a kind of vessel’ (NCED 997).
§ SCG (200) has a different comparison, with PEC *tákV ‘a kind of vessel’, which is
more suitable for the second consonant (Bur *k = PEC *k), while the above comparison is
more suitable for the first consonant (Bur *t = PNC *t), with secondary change of PDC
*q > Bur *k (as in Lezgian). Possibly further study will determine which option is best.

Bur. *(y)alt- in giyált (H, N) ‘spoon, scoop’ (compound with *giy- ‘to pour, put in, put
on’, etc.)
   ~ Cauc: Hunzib égu ‘wooden shovel’, Khwarshi éku, Lezgi yirf id., etc. < PEC
   *giyáltWV ‘wooden shovel’ (NCED 673)
   ~ Basque *sa-harde ‘rake; (two-tined) fork; (dinner) fork’: compound with obscure first
   element *sa- (perhaps a haplogonic compound with Bsq *sarha-tu ‘to clear land’, i.e.
   *sarha-harde > *saharde?).
§ PDC *ydlWV ‘shovel, ladle’ (SCG 113). With the frequent correspondence of Bur *大批
   ~ PDC *大批- (see Phonology).

Bur. *yeéš > yeéš (H, N) ‘lasting dwelling place, permanent residence’
   ~ Cauc: Tsakhur yie’a ‘sty, cattle shed’, Chechen c’a ‘house, room’, Khinalug c’a ‘house’,
   Ubykh cyá ‘house, room’, etc. < PNC *y:fyW ~ *y:fyWV ~ *y:fy:cu ‘house’
   (NCED 364)
   ~ Basque *eçe ‘house’ (or *e=eče, with fossilized class prefix = PNC *e=) > (Bzk, Rnc)
etse, (c) etxe /ece/ ‘house’. The common form etxe seems to have been the expressive
form, now generalized.
§ Bur + Bsq compared by Berger (1956: pp. 18, 24). PDC c[if]jü ‘house’ (Starostin
Correspondences of stops: In general, Burush aski unaspirated \*p, \*t/t, \*k, \*q correspond to PNC/PEC glottalized \*p, \*t, \*k, \*q (=[p', t', k', q']), respectively; and Burushaski aspirated \*ph, \*th, \*kh, \*qh correspond to PNC/PEC aspirated \*p, \*t, \*k, \*q ([pʰ, tʰ, kʰ, qʰ]), respectively. For voiced stops the correspondences are generally trivial, i.e. \*b = \*b, \*d/d = \*d, \*g = \*g, \*y = \*g.\(^{62}\) See the following examples from the lexical comparisons cited above (some from S.A. Starostin’s later etymologies):

- **Bur.** \*p, \*t/t, \*k, \*q = PNC/PEC \*p, \*t, \*k, \*q
- **Bur.** \*tāpi ‘stony terrace’ ~ PEC \*tāp’ī ‘stone plate or shed’ (SCG 137)
- **Bur.** \*topo, \*tutopo ‘a kind of bread’ ~ PEC \*HārLāpV ‘a kind of food’ (SCG 63)
- **Bur.** \*tókur ‘wooden chest’ ~ PNC \*taqV ‘vessel’: Rutul \*fak ‘basket (for berries)’\(^{63}\)
- **Bur.** \*=huf ‘foot’ ~ PEC \*fiTtwV ~ *twlhV ‘foot’: Avar \*hefe/het ‘foot’
- **Bur.** \*=ken ‘liver’ ~ PEC \*kunHV ‘kidney’: Chamali \*k’u ‘liver’
- **Bur.** \*kur- ‘bone, joint’ ~ PEC \*kwirV ‘bone, leg’: Lezgi \*k’ur ‘leg, hoof’
- **Bur.** \*tuma-y ‘nut, fruit’ ~ PNC \*tumhV ‘kernel, fruit’: Archi \*t’ummu- ‘grape’
- **Bur.** \*tuni ‘small basket’ ~ PEC \*fwdn?e: Dargi \*funi ‘trough’
- **Bur.** \*phen ‘fly’ ~ PEC \*pdngwV ‘bee’: Avar \*puq:na ‘drone’
- **Bur.** \*phiran ‘moth, spider’ ~ PEC \*porV ‘bee’: Tindi \*pera ‘bee’
- **Bur.** \*phunc ‘dew’ ~ PNC \*pincM’A ‘resin, juice’: Lak \*pīc ‘dew; sweat’
- **Bur.** \*ther ‘dirt’ ~ PEC \*tiirV ‘dirt, dung’: Akhvakh \*tere- ‘ashes, dust’
- **Bur.** \*khor ‘large basket’ ~ PEC \*kn’drV ‘vessel’: Akhwakh \*koro ‘trough, gutter’
- **Bur.** \*=qhas- ‘arse, genital’ ~ PNC \*=VgV ‘behind’: Udi \*qos
- **Bur.** \*=qhat ‘mouth’ ~ PEC \*qwiti: Lak \*qit ‘Adam’s apple; beak’
- **Bur.** \*=qhorpVt ‘lung’ ~ PEC \*gwsl6V(rV): Archi \*xurt:ur ‘lung’

**Bur.** \*b, \*d/d, \*g, \*y = PNC/PEC \*b, \*d, \*g, \*g
- **Bur.** \*bac-in ‘shank (of animal)’ ~ PEC \*bh[ə]cV: Chamali \*beč ‘knee (of animal), thigh’
- **Bur.** \*balit ‘front room, veranda’ ~ PEC \*bǔl2V ‘house’: Hunzib \*bulti ‘at home’
- **Bur.** \*bay ‘resin’ ~ PEC \*bhünkV: Chechen \*bağa ‘pine tree’
- **Bur.** \*bay ‘millet’ ~ PNC \*bölćw: Tindi \*boča ‘millet’
- **Bur.** \*buć ‘goat’ ~ PEC \*b[a]ćV ‘goat’: Lak \*buća ‘yearling male goat’
- **Bur.** \*bimbal- ‘ankle’ ~ PEC \*bimłV: Tsezi \*bula ‘hoof’; cf. PY \*bül ‘foot, leg’
- **Bur.** \*dul-dim ‘cloud of dust’, etc. ~ PEC \*dilšwV: Lak \*turku ‘cloud’
- **Bur.** \*du[m] ‘kid’ ~ PNC \*dwān2V ‘sheep, ram’: Lak \*t:a ‘sheep, ewe’
- **Bur.** \*dögār ‘ram’ ~ PEC \*dwrgwV ‘billy-goat’: Avar \*deşēn id.
- **Bur.** \*ehardV ‘stallion’ ~ Abkhaz a-čada, ęda ‘ass, donkey’, etc.

\(^{62}\) Retroflex stops have sometimes developed secondarily in Bur from original dental stops. The conditioning factors are not yet fully understood. See Bengtson & Blazek (2011: 3-4).

\(^{63}\) PEC has another similar root, \*takwV ‘a k. of vessel’ (NCED 990). See the note under \*taq V (NCED 997). The initial \*t- of Bur. \*tōkur fits best with PNC \*taqV, but the internal \*-k- fits \*tākV.
There remain a few seemingly anomalous cases, but there is a growing body of evidence that some reflexes have been conditioned by the position of accent, for example, a tendency for unaspirated stops to become aspirated (or, in the case of uvulars, *q* > *qʰ* > *γ*) when the accent is on a following (or preceding) syllable. (Cf. “Verner’s Law” in Indo-European phonology.) Some dialectal variations in these tendencies (*q* > *qʰ* > *γ*) can be observed:

- Bur. *phol-(yuuy)* ‘feather’ (expected *poI-*) ~ PNC *pVhV* ‘feather’, etc.
- Bur. *tharen-* ‘narrow’ (expected *tar-*) ~ PNC *=iIiV* ‘thin’
- Bur. *thugdr* ‘buck goat’ (expected *tug-*) ~ PNC *tiigV* ‘buck goat’
- Bur. (H) *qaruuyo* - (N) *yariiuyo* ‘heron’ ~ PNC *qpr§qwV* ‘crane’
- Bur. (Y) *qaqd-m* - (H, N) *yaqdy-* ‘bitter; unsweetened; sour’ ~ PNC *qefdV~* *qefdV* ‘bitter’
- Bur. (H) *=qat* ~ (N) *=qhat* ~ (Y) *=get-ar* ‘armpit’ ~ PEC *qVdV* ‘brisket’ (here the accent is on a preceding pronominal prefix)\(^{64}\)

**Correspondences to Caucasian lateral affricates:** As already pointed out in some of the notes to the above lexical comparisons, there are recurrent correspondences between Burushaski initial *t-* and medial *-lt-* and the Proto-Caucasian lateral affricates *ṯ*, *ṯ, *L* = [t̚, t̚', d̚], as shown in the following examples. The corresponding reflexes in Basque are initial *l-* and medial *-rd-*, respectively. (Comparisons already listed above will be cited in abbreviated form):

(a) Examples with Burushaski initial *t-*:

- Bur. *tay* ‘branch, shoot’ ~ PEC *tōrV* ‘sprout’
- Bur. *tal* ‘dove’ ~ PEC *tēlV* ‘a kind of bird’
- Bur. *tal* ‘stomach, belly’ ~ PEC *HlalV ~ *HlaiV* ‘liver’
- Bur. *tano* ‘colon, rectum’ ~ PNC *HX6nu* ‘bottom’
- Bur. *tdpi* ‘stone terrace’ ~ PEC *tēpV* ‘stone plate, shed’ ~ Bsq *iape* ‘shelter’
- Bur. *tar[i]-* ‘skin’ ~ PNC *tēlV* ‘skin’ ~ Bsq *lari* ‘skin, hide, leather’
- Bur. *tis* ‘wind’ ~ PEC *HlaiV* ‘movement of air’
- Bur. *tur-Cún ‘marmot’ ~ PNC *HlaiV ~ *γårV* ‘hare’
- Bur. *tharen-* ‘narrow, tight’ ~ PNC *=iIiV* ‘thin’ ~ Bsq *lirain* ‘lithe, svelte’

(b) Examples with Burushaski alternation of (initial) *t-* / (medial) *-lt-*:

\(^{64}\) Realized as (H) a=qat ‘my armpit, gó=qat ‘thy -’, mó=qat ‘her -’, etc.
• Bur. *ltén (Y) ten ‘bone’ / (H, N) =ltin ‘bone’ (bound form) ~ PEC *ål̥n ‘ankle, shin’
• Bur. *ltur> (Y) tur ‘horn’ / (H, N) =ltür (bound form) ~ PEC *₂wūr ‘horn; braid, mane’ ~ Bsq *ₐrdar ‘horn’ (< *ₐrdar)
• Bur. *ltap > tap ‘leaf’ / du=ltapi- ~ =ltapi- ‘to wither’ ~ PNC *₂ḍi ‘leaf’ ~ Bsq *lapa-r ‘bramble’
• Bur. *lē> (Y) lē- ‘to swear’ ~ PEC *Hl̥LV ‘to say’ (NCED 572)
• Bur. *ltul[i] ‘to saddle’, *tili- ~ *tele- ‘saddle’ ~ PEC *₂wīlē ‘saddle’
• Bur. *ltār> tār ‘bone’ / (H. N) =ltān (bound form) ~ PEC *₂wūn ‘bone’ (bound form) ~ PEC *₂wūn ‘horn’

(c) Examples with Burushaski medial *-lt- (~ -ld-):
• Bur. *balti ‘front room, veranda’ ~ PEC *bālV ‘house’
• Bur. *b̥ltar ‘buttermilk’ ~ PNC *r̥hālV ‘milk’
• Bur. *daltar ‘to thresh’ ~ PEC *V̥r̥LV ‘to thresh’ ~ Bsq *l̥r̥vain ‘threshing floor’
• Bur. *hālt- ‘to wash’ ~ PEC *V̥h̥V ‘to wash, pour; weep’ (NCED 1023)
• Bur. *h̥ltar ‘to show’ ~ PEC *V̥lt̥V ‘to look’ (NCED 209)
• Bur. *n̥lt̥r̥ ‘butter’ ~ PEC *n̥lt̥r̥ ‘milk, dairy product’
• Bur. *n̥lt̥t̥ ‘norris’ ~ PEC *n̥r̥ ‘horn, head, nose’ ~ Bsq *mut̥r̥ ‘snout, muzzle’
• Bur. *ḥ̥l̥t̥l̥ ‘part of the ribs’, etc. ~ PEC *H̥l̥LV ~ *H̥l̥LV ‘armful’
• Bur. *y̥lt̥r̥ ‘spoon, scoop’ ~ PEC *y̥r̥ ‘wooden shovel’ ~ Bsq *s̥yar̥ ‘rake, fork’
• Bur. *y̥lt̥l̥t̥ ‘upper leafy branches’, etc. ~ PEC *h̥l̥l̥V ‘branch, pod’ ~ Bsq *s̥yar̥ ‘rake, fork’

(d) Examples with Burushaski final *-l:*
• Bur. *tal ‘dove’ ~ PEC *l̥l̥ ‘a kind of bird’
• Bur. *tāl ‘belly, abdomen, bowels’ ~ PEC *t̥l̥ ‘stomach, rennet, abomasum’
• Bur. *y̥l̥ ‘to hear’ ~ PNC *e̥l̥ ‘to hear’ (NCED 411, CSG 46)

One East Caucasian language, Avar, partially converges with Burushaski in the development [ˌtʼ] < PDC ʔtʼ.

• Avar tʼalu ‘rock, rocky plateau’ (< PEC ʔl̥l̥) ~ Bur *tali ‘slope (of a mountain)’ (SCG 135)
• Avar (southern) tʼeb ‘millstone, whetstone’ (< PEC ʔl̥p̥V) ~ Bur *t̥p̥ ‘stone terrace’
• Avar tʼerêna-b ‘thin’ (< PNC *u̥l̥V) ~ Bur. *t̥r̥n̥-um ‘narrow, tight’
• Avar tʼim ‘bottom’ (< PNC *H̥l̥m̥) ~ Bur. *t̥n̥ ‘rectum’
• Avar tʼul ‘liver’ (< PEC *H̥l̥l̥V) = Bur. *tal ‘stomach, belly’

Interesting as this may be, the Avar change differs from that of Bur in that it is restricted to PNC/PEC *ʔtʼ(w), and in all positions (NCED 52), while the Bur change involves all lateral affricates, and only in initial position (see above). Some additional examples of Bur *t- / *lt- / *l- corresponding to PDC lateral affricates are given in Bengtson & Blažek (2011).
Conclusions

On the basis of about 100 lexical cognate sets, together with regular phonological patterning of the cognates, and shared irregular morphological patterns, a genetic relationship among Burushaski, the North Caucasian languages, Basque, and Yeniseian languages is the best hypothesis for explaining these similarities. On a competing hypothesis, that Burushaski has a close genetic relationship with Indo-European (Casule 1998, 2003), see Bengtson (2000), Bengtson & Blažek (2011a, 2011b).

Cultural vocabulary shared by the same languages, including words for domestic sheep and goats, dairying, cultivated grain crops (and processes connected with them), and for other artifacts, suggests that the speakers of the proto-language ancestral to these languages (Euskaro-Caucasian or Macro-Caucasian) dispersed as early as 7000 to 9000 years BP in association with the spread of animal domestication and the cultivation of grain.65

Postscript (2014)

Since the original text of this paper was written, for the the Third Harvard Round Table in 2001, many new developments have taken place in Dene-Caucasian studies. The late Sergei A. Starostin, following up on the work of his colleague Ilia Peiros (1988) and myself, worked intensively during the last months of his life on integrating Burushaski into his Sino-Caucasian hypothesis (which up to that point had only included North Caucasian, Yeniseian, and Sino-Tibetan), greatly expanding the number of Burushaski-Dene-Caucasian etymologies, and more fully formulating the comparative phonology briefly sketched above. Before his untimely death in 2005 Starostin was able to integrate Burushaski into his Sino-Caucasian Phonology (2005a), Sino-Caucasian Glossary (2005b), and the Tower of Babel etymological databases (2005c, 2005d, 2007).

Since that time other members of the “Moscow School,” especially George Starostin and Alexei Kassian, have further developed the Dene-Caucasian hypothesis; and credit must also be given to Václav Blažek and Merritt Ruhlen (see References). Using glottochronology, G. Starostin has proposed a new taxonomic model of Dene-Caucasian:

A. ’Sino-Dene’ or ‘Eastern Dene-Sino-Caucasian’:
   A.1. Sino-Tibetan;
   A.2. Na-Dene;
B. ’Western Dene-Sino-Caucasian’:
   B.1. Yeniseian + Burushaski;

Node A (‘Sino-Dene’) essentially confirms the proposal made by Edward Sapir nearly a century ago. Node B.1, ‘Burusho-Yeniseian’, coincides with George van Driem’s (2001) model, which he has called ‘Karasuk’, and has been further developed recently by Bengtson (2010). The approximate dates of these nodes are ~10,660 BCE (division of A from B), ~9000 BCE (division of A.1 from A.2), slightly later (~8,330 BCE) for the split

65 See Postscript (2014) for the glottochronological dates calculated by G. Starostin.
between B.1 and B.2, and about two millennia later (~6,570 BCE) for the breakup of Burusho-Yeniseian.

As to the Burusho-Yeniseian node, some shared semantic innovations were mentioned above:

- Bur. *mes ‘finger/toe’ ~ PY *bes- ‘finger’ (vs. NC ‘hoof, nail, claw’)
- Bur. *é(h)argé ‘flying squirrel’ ~ PY *sa?qa ‘squirrel’ (vs. NC ‘weasel, mouse’, Bsq ‘mouse’)
- Bur. *khen ‘flea’ ~ PY *go?h ‘flea’ (vs. NC ‘louse, nit, worm’, Bsq ‘tick’)

Several other close Burusho-Yeniseian isoglosses can be mentioned:

- Bur. *=yek ‘name’ ~ PY *=ta ‘name’
- Bur. *=reij ‘hand’ ~ PY *=ri ‘hand’
- Bur. *=gau ‘road’ ~ PY *=kaŋ- (~ *gáŋ-) ‘(hunting) path’
- Bur. *=phuŋ ‘nape (of neck); shoulder’ ~ PY *=p[ u]ym- ‘neck’
- Bur. *=toq ‘mud’ ~ PY *=tq- ‘mud, clay’

Among grammatical homologies one of the most obvious is the second person singular verbal affix, in both families inserted before the verb root:

Bur. (2ps verbal affix) gu-/-ku-, gu-, go-/-ko-, goo-/-koo- ~ Ket (2ps verbal affix) k-, ku-

Another grammatical homology is found in the demonstrative pronouns, with a common stem *k(h)i-:

Bur. (Y) khin, khene ‘this’ (hm-class) ~ Ket (Pak, Sur) kído ‘this’ (m.), etc.

Burushaski and Yeniseian have some similar interrogative pronouns made up of the elements *bV + *S:

Burushaski: (Y) besa, bese ‘why?’, (H, N) besau ‘what, which?’, bese ‘why?’
Ket bišëy / bišây ‘where?’, bišše ‘who?’ (masc.), beša ‘who?’ (fem.)

The evidence for Burusho-Yeniseian is discussed in more detail in Bengtson (2010).

What about “Dene-Yeniseian”

Building on a proposal by Ruhlen (1998a) Edward Vajda (see References) has elaborated a “Dene-Yeniseian” hypothesis, attracting attention and acceptance from some ‘mainstream’ linguists, and in the popular press, e.g. Diamond (2011). At a recent Dene/Athabaskan Languages Conference (Berkeley, Calif., 2009) Vajda, George Starostin, and I presented reports about Dene-Yeniseian and Dene-(Sino-)Caucasian. Starostin and I asserted that the binary Dene-Yeniseian model is less adequate than the multilateral Dene-Caucasian one, and that Dene-Yeniseian is taxonomically inaccurate, since, in our view,
Na-Dene is closer to Sino-Tibetan, and Yeniseian is closer to Burushaski, than Na-Dene and Yeniseian are to each other (see G. Starostin’s diagram above). We maintained that the bigger, if somewhat messier, picture of a DC macrofamily is a better and more holistic approximation of the prehistoric reality. The relevant reports have since been published as Vajda (2010), Bengtson (2010) and G. Starostin (2010a). More recently, see further discussions by G. Starostin (2012) and Vajda (2012).

It is plain to see from the cultural lexicon discussed above that the Yeniseian languages share virtually none of the vocabulary pertaining to food production (domestic cattle, dairying, cultivated grain crops and their processing), with the other ‘Western Dene-Caucasian’ languages, Burushaski, Caucasian, and Basque. This is simply because the climate and landscape in the homeland of the Yeniseian peoples, in the Siberian taiga, are not conducive to food production, and presumably the Proto-Yeniseians either did not participate in the Neolithic revolution, or were forced by a new environment to abandon food production. The only exception (until Soviet rule) was the adoption of domestic reindeer by some Ket groups, from Samoyedic neighbors (Vajda 1998). So to some extent, at least, the apparent similarities between Yeniseian and Na-Dene in lexicon associated with sleds, snowshoes, birch-bark, and the like, are to be attributed not to taxonomic closeness but to similar environmental and cultural conditions.

Abbreviations

| BN | bas-navarrais / behe nafarrera / Low Navarrese (Basque dialect) |
| Bzk | Bizkaian = Biscayan (Basque dialect) |
| Bzt | Baztanese (Basque dialect) |
| c | common Basque |
| DC | Dene-Caucasian |
| EC | East Caucasian (= Northeast Caucasian or Nakh-Dagestanian) |
| H | Hunza (Burushaski dialect) |
| Lap | Lapurdian = Labourdin (Basque dialect) |
| N | Nager (Burushaski dialect) |
| PA | Proto-Athabaskan |
| PAE | Proto-Athabaskan-Eyak |
| PDC | Proto-Dene-Caucasian |
| NEC | Proto-East Caucasian |
| PNC | Proto-(North) Caucasian |
| PWC | Proto-West Caucasian |
| PY | Proto-Yeniseian |
| Rnc | Roncalese (Basque dialect) |
| WC | West Caucasian (= Northwest Caucasian or Abkhaz-Adygean) |
| Y | Yasin (Burushaski dialect) = “Werchikwar” |
| Zub | Zuberoan = Souletin (Basque dialect) |

Buck

CDIAL

LACUS

NCFD

SCG

SCP

Buck (1949)

Turner (1962-66)

Linguistic Association of Canada and the United States

Nikolayev & Starostin (1994)

Sino-Caucasian Glossary (Starostin 2005b)

Sino-Caucasian Phonology (Starostin 2005a)
References


219


Leer, Jeff. 1993. “AET Comparative Lexical Database.” Ms. ANLC.


Reviewed by Václav Blažek
Department of Linguistics and Baltic Studies
Faculty of Arts, Masaryk University, Brno
Czech Republic

This long-expected monograph by Michael Witzel, Wales professor of Sanskrit at Harvard University, is unique for several reasons. The author brings forth a revolutionary stratigraphy of myths, determined by two chronological and geographical layers, an earlier southern, called Gondwanan, and a later northern, called Laurasian. Both terms were chosen by the author on the analogy of geological designations of the supercontinents Gondwana and Laurasia, separated 200-180·10⁶ years ago in the mid-Mesozoic era. Naturally, the author does not try to date the oldest myths to this period; his chronological limit is c. 160·10³ years (p. 372), when modern humans appeared in Africa.

M. Witzel summarizes the typical features of the Laurasian mythological traditions as follows (p. 64):

1. Primordial waters / chaos / ‘nonbeing’
2. Primordial egg / giant.
3. Primordial hill or island.
4. (Father) Heaven / (Mother) Earth and their children (4 or 5 generations / ages).
5. Heaven is pushed up (and origin of Milky Way).
6. The hidden sun light revealed.
7. Current gods defeat or kill their predecessors.
8. Killing the ‘dragon’ (and use of heavenly drink), fertilization of the earth.
9. Sun deity is the father of humans (or just of ‘chieftains’).
10. First humans and first evil deeds (often, still by a demi-god), origin of death / the flood.
11. Heroes and nymphs.
12. Bringing of culture: fire / food / culture by a culture hero or shaman; rituals.
14. Final destruction of humans, the world (and) the gods (variant of the Four Ages theme).
15. (A new heaven and a new earth).

The mythological traditions defined as Gondwanan are not as rich, although in several features they agree with the Laurasian ones (pp. 322-323):

1. In the beginning: heaven, earth (and the sea) already exist.
2. A High God lives in heaven, or on earth, or ascends to heaven later. This highest being has been described as the Rainbow Snake.
3. Series of lower gods, often children of High God, act as tricksters and culture heroes.
4. Primordial period ended by some evil deed of son of High God (or by humans).
5. Humans are created from trees and clay (or rock); occasionally, descend directly from the gods / totem ancestors.
6. Humans act haughtily or make a mistake; punishment by a great flood; humans reemerge in various ways. (An end to the world is missing.)

Some concepts seem to be universal, and M. Witzel designates them as Pan-Gaean:
1. High god.
2. Creation of heaven and earth.
3. Creation of humans from clay or from a tree.
4. Cultural hero or trickster.

The following comments concentrate on the linguistic implications of the book under review.

There are fascinating linguistic traces of the Rainbow Snake (cf. the map by Berezkin 2013, 115) and related supreme deities along the coast of the Indian ocean. Witzel (pp. 310, 326, 351, 352) mentions the Andamanese creator deity Puluga or Biliku, often identified with the northeast monsoon. The concrete forms and their sources are as follows:

Little Andaman: Önge Ólugá ‘mythical being identified with the northeast monsoon, name of the Monitor lizard’ (Brown 1914, 40, 51) = õlugé ‘to thunder’ (Portman 1887, 83), õlugé ‘god’ (Portman 1887, 34-35).

This common Andamanese god was apparently specialized in thundering, to judge from the following designations for ‘thunder’:


The closest external cognate may be identified in Australia in one non-Pama-Nyungan family from the eastern Arnhem Land: proto-Gunwinyguan *polong ~ *polung ‘rainbow serpent’ > Dalabon, Jawoyn, Rembarrnga polung, Mangarrayi polokpa (Harvey 2003, 255, #893). Other possible Australian cognates were collected by Riccardo Gatti more than a century ago:

non-Pama-Nyungan: Tangkic: Jakula pargi-gi ‘god’ (Gatti 1906, 6: Jakula + Andamanese) || Pama-Nyungan: Lower Murray: Ngintait poorache; Bagandji: Maraura poorook; Marie: Wadjabangai baringa; Waka-Kabic: Bayali barooongi; Central East Coast: Bigumbil boorinja, booronga, Yugumbal boorongi ‘thunder’ (Gatti 1906, 31: Pama-Nyungan + Andamanese). Related are perhaps also Karadjerri (Pama-Nyungan family Marrngu on the northwest coast) Bulang ‘a tribal creator-goddess’ and Warramunga / Warumungu (Pama-Nyungan isolate in south from the Gulf of Carpentaria) wollungua ‘rainbow serpent’ (Mills 1918, 211).

Gatti (1906, 31) also added Tasmanian (glossary by Peron) bura ‘thunder’.

In his introduction to Gatti 1906, Alfredo Trombetti (pp. ix-x) also added hypothetical parallels from Sub-Saharan Africa:

(i) Oromo of Wellega bulgu ‘ogre, cannibal (with four eyes)’ (Gragg 1982, 67) = bulgu ‘cannibal or man-eater believed to have four eyes’ (Tilahun Gamaa 1989, 95) =

224
bulgu ‘antropofago, cangabile; creazione della fantasia e della superstizione locale’
(Borello 1995, 66) = Boraan bulgu ‘cannibal, bogey man’ (Stroomer 1995, 159).

This term stands isolated within Cushitic (and Afroasiatic in general), and cannot be
explained as an Ethio-Semitic loan, so its origin in some pre-Cushitic substratum seems
quite probable.

(ii) Bantu supreme deity, documented in the zones EFGMNP¹, especially in Eastern
Africa: Nyamwezi mu-lungu ‘tribal creator- & sky-god watching over the earth’.
Kikuyu mu-rungu ‘supreme god’. Chiuta mu-lengi ‘tribal creator’. Kinga pl. emi-
lungu ‘forefathers of the common people’. Swahili mu-unugu, pl. mi-unugu ‘god’.

The missing initial labial syllable, in comparison with the preceding Cushitic examples,
may be ascribed to reinterpretation in the frame of the noun-class system. Rather
peculiar is the semantic determination of the class prefix pair sg. mu- vs. pl. mi-; in
Swahili, for example, where both sg. & pl. operate, this pair is reserved for plants (e.g.,
m(u)-ti ‘tree’ / mi-ti ‘trees’).

Later Trombetti (1918-19, 35/363) abandoned the Bantu examples and added
the following comparanda from the Niger-Congo languages of the Gulf of Guinea:
Lefana (= Lelemi in Togo & Ghana; Kwa family) Burukú ‘name of a deity’ and Sobo
(Edoid; south Nigeria) sono-blúgvé ‘god’. It is possible to add Fon (Gbe group of the
Volta-Niger family; Burkina Faso) Nana Buluku³ ‘androgyne supreme Creator of the
Universe and all that exists in it’, and Yoruba Nana Burukii⁴ ‘great grandmother,
connected with rivers and oceans; she is the ultimate mother of the waters, especially
the sweet river waters’

It is tempting to speculate about any relation to the deity celebrated by Ngbandi
(Adamawa-Ubangi family of the Niger-Congo macro-family), called Mbongo ‘a river-
god and one of the seven deities invoked at sun rise every morning; the creator god of
all tribal people who is said to reside in black waters’ (Jordan 1993, 161).

The tree-diagram shown below [Figure 1] depicts hypothetical genealogical
relations among the world’s linguistic macro-phyla, with a preliminary chronological
scale. The double line cutting across branch lines is a mythological “Wallace Line,”
separating the predominant Laurasian motifs from more archaic Gondwanan ones.

The ‘Gondwanan’ motifs are typical of people speaking Khoisan and Congo-
Saharan languages in Sub-Saharan Africa, and Indo-Pacific and Australian languages
in Australasia, probably descendants of the first out-of-Africa migration. But they
appear more frequently than rarely also in two populations speaking Dravidian and
Austronesian languages, whose proto-languages are affiliated with the Nostratic and
Austro macro-families respectively, and both are connected with the later out-of-Africa
migrations (see Figure 1). This contradiction is explainable on the basis of a strong
influence of the pre-Dravidian and pre-Austronesian / pre-Austro substrata. The
following examples may serve as illustration⁵:

¹ http://goto.glocalnet.net/mahopapers/nuglonline.pdf
² http://www.africamuseum.be/collections/browsecollections/humansciences/blr/results_main
³ http://en.wikipedia.org/wiki/Nana_Buluku
⁴ http://myoruba.tumblr.com/post/78006900721/nana-buruku-is-the-great-grandmother-to-all-of-the
⁵ Bengtson’s (1994) stimulating comparisons between Austro, Indo-Pacific and Australian can be
reinterpreted in this areal sense. Before the settlement of Austronesians in the Indonesian & Philippine
archipelagoes, and other Austro populations on the coast of Southeast Asia, these territories were not
uninhabited. The people living there probably spoke languages related to Indo-Pacific, whose remnants
Figure 1: Hypothetical genealogical relations among the world’s linguistic macro-phyla, with a preliminary chronological scale. The double line [==] is a mythological “Wallace Line,” separating the predominant Laurasian motifs from more archaic Gondwanan ones.

**STAR:** (Trombetti 1918-19, 51/379)

in Timor and North Halmahera exist till the present, and Australian. An analogous situation was probably valid for the Indian peninsula (cf. Blažek 2006).
Kalkatungic: 101 Jalanga booderoo (Australian forms with numbers are quoted after Curr; Gatti 1906, 29: Marie + Andamanese; Trombetti 1918-19, 51: Pama-Nyungan + Austronesian + North Dravidian).

- Tasmanian piterina ‘sun’ & potena ‘star’ (glossary of Lhotsky; Gatti 1906, 27 compared potena with Widugari {Pilbara-Ngayarda subgroup of the Gascoyne R.-Pilbara R. group of South-West Pama-Nyungan} peta ‘star’).
- Austronesian *bi(n)tuun (D 1929) = *bituqen (Dyen 1953a) ‘star’.

WOMAN: (Bengtson 1994, 74)
- Austronesian *binay (Dempwolff) ~ *binsy (Dyen).

STONE: (Blazek 2006, #68)
- Australian: non-Pama-Nyungan: Gunwinyguan *kal- ‘mountain, hill’ (Peiros 428, 441) || Pama-Nyungan: Mabuiag kula; Mbambylmu (‘Princesse Charlotte Bay’) kūla; Gumbaynggiric kullam, etc. ‘stone’ (Schmidt).
- Austronesian *kal ‘rock’ (Dempwolff 1929).

WIND: (Blazek 2006, #62)
- Andamanese: Bea wūl-nga-da ‘wind’ (Portmann 1898, 186).
- Dravidian *val ‘wind’ (DEDR 5312: I, III, V, VI).
- Austronesian *balay ‘wind’ (Dempwolff).

Summing up, The Origins of the World’s Mythologies is not only a unique encyclopedia of comparative mythology, genetics and anthropology, but also some very inspirational reading, likely to stimulate new research in the field of comparative linguistics.
References


Electronic publications

Bantu classification: http://goto.glocalnet.net/mahopapers/nuglonline.pdf

Acknowledgement: The present study was prepared thanks to a grant from the The Czech Science Foundation (GACR). P406/12/0655.

Václav Blažek <blazek@phil.muni.cz>
Department of Linguistics and Baltic Studies
Faculty of Arts. Masaryk University. Brno. Czech Republic

228
Michael Fortescue: Language Relations across Bering Strait: Reappraising the Archaeological and Linguistic Evidence.

Reviewed by Peter A. Michalove
University of Illinois

This book explores the relationships among four language families, Uralic, Yukagir, Chukchi-Kamchatkan, and Eskimo-Aleut (collectively referred to here as the Uralo-Siberian or US languages). While Fortescue is probably most widely known for his work in Eskimo-Aleut, he has worked with all of these languages to varying degrees, and so brings a deep knowledge of the material to this study.

Fortescue appears to be ambivalent about the nature of this relationship, referring to a “mesh, if not an actual genetic stock” (p. 3 and passim), explaining later (p. 28) that this term refers to any degree of historical relatedness between a group of (once) geographically related languages linked by lexical and/or phonological or structural ‘family resemblance’. The possibilities range from Sprachbunds which have drawn unrelated languages into their orbit, through interlocking chains of languages where the ends are unrelated but where there is considerable overlap and even language mixing, to traditional language meshes that exclusively involve languages of a single family. Another, crucial intermediate possibility is the situation where all the ingredient languages are ultimate derived from a single ancestral protolanguage but the time depth is too great to prove it.

Yet elsewhere Fortescue speaks routinely in the terminology of demonstrable genetic language relations. For example he refers (p. 77) to, “elements inherited from Proto-US.” Similarly, when he speaks of the undeniable presence of loans among these languages, he sees the loans, “[a]s opposed to genetically common lexical material apparently shared by the various Uralo-Siberian languages.”

The introductory chapter states Fortescue’s thesis, that there is a demonstrable historical relationship among Uralic, Yukagir, Chukchi-Kamchatkan, and Eskimo-Aleut. Fortescue’s goal is to pin down as much as possible of the historical connections among these languages by supplementing the results of traditional comparative linguistics with typological and archaeological findings.

Chapter Two discusses previous attempts to relate various combinations of these languages, as well as the ongoing debate over Chukchi-Kamchatkan itself. Most scholars today accept a Chukchi-Kamchatkan family, consisting of Chukotian and Kamchadal (the latter represented by only one surviving language, Itelmen), but some writers (e.g. Worth 1962 and Volodin 1976) have not accepted this relationship. Beyond that there have been attempts to relate most of the binary combinations of the four US families: Uralic and Yukagir (Collinder 1940, Angere 1956, Taillleur 1959), Eskimo-Aleut and Chukchi-Kamchatkan (Swadesh 1962, Hamp 1976), Uralic and Eskimo-Aleut (Bergsland 1959), etc. These works are of varying quality, but most have been

---

6 This review by the late Peter A. Michalove (1951-2013) was apparently written several years ago, but never published, as far as we know. Thanks to Allan R. Bomhard for calling our attention to it. Ed.
hampered by the paucity of good descriptive work on these languages (except for that on Uralic). The case for a Uralic-Yukagir connection has apparently been greatly strengthened recently by the appearance of Nikolaeva (1988), but I have been unable to obtain this work.

Fortescue also mentions the possibility of relating all of these languages within a wider Nostratic or Eurasiatic context. He is sympathetic to the idea, but does not pursue it here because he feels that, even if there is a broader Nostratic configuration, the US languages form a more closely related subset. He feels that the Altaic languages, which have been those most frequently compared to Uralic, do not share most of the isoglosses that bind the US group together. However, I find Altaic correspondents to many of the morphological and lexical parallels he considers for the US languages. Further, he states that his view differs from most Nostratic work, "in not seeing any solid grounds for extending these ‘stocks’ further west into Europe or south into India" (p. 52).

Chapter three offers a typological overview of the languages of Siberia. While typological features are not generally considered reliable guides to possible genetic relationship in themselves, Fortescue makes ingenious use of typology in establishing historical connections among languages groups. Although some typological features, such as word-order and ergative or accusative alignment are extremely common, and others, such as postpositions and government relationships, are the result of implicational universals involving other features, Fortescue maintains that rarer features, when geographically clearly defined, can argue for language relationship. For example all of the US languages are characterized by a single (voiceless) set of plosives and a single (voiced) set of non-sibilant fricatives. This is an extremely rare combination among the languages of the world, and it is found nowhere in the languages surrounding US, including those with which they are known to have had extensive contact, such as Yeniseian, Altaic, Na-Dene, and Salishan.

Chapter Four summarizes the morphological evidence for a relationship between Chukchi-Kamchatkan and Eskimo-Aleut, while Chapter Five adds to the picture by bringing in Uralic and Yukagir. Significantly, Fortescue finds that the inclusion of the latter two only deepens the overall picture by confirming many of the commonalities discussed in Chapter Four. Building on the chapter on typology, many of these morphological commonalities are in striking contrast to the material and typological morphologies of surrounding languages.

Chapter Six deals with the lexical correspondences among these language families, and in so doing, sets out a reconstructed phonological system and proposed set of reflexes in the various languages. In this manner, Fortescue is able not only to present a large number of apparently cognate forms, often of quite different phonetic shape, but also to identify many phonetically common forms as loans because they do not conform to these phonological developments. One of the most striking aspects of his list of common US forms (defined as those represented in phonologically regular form in at least three of the four families) is the absence of word-initial */r-//, while initial */l-/ is well represented. This is a phonotactic characteristic of Tungusic as well, and most modern writers who accept the Altaic relationship reconstruct it for Altaic itself (e.g. Starostin 1991, Vovin 1994).

Chapter Seven discusses the archaeological evidence of settlement and migration in the far north, while the final chapter summarizes the cumulative historical picture suggested by the data. Fortescue sees an origin of the US languages in southern and central Siberia c. 8000 to 10,000 years BP.
The greatest weakness of this book is the very poor production quality. There is a huge number of typos, and errors in bibliographic references. The maps, especially those showing the distribution of isoglosses are poorly reproduced and difficult to read, and the endnotes after each chapter are less convenient then footnotes on the page would have been.

But this is a minor quibble. The possibility of a genetic relationship among these languages has long been discussed in mostly impressionistic terms. This book is one of the first efforts to lay out the evidence systematically, in a manner consistent with modern scholarly standards. As such, it is full of stimulating ideas, which are sure to be developed by further research. This book will be an important basis for that ongoing work.

References

Tailleur, Olivier G. 1959. Plaidoyer pour le youkaghir, branche orientale de la famille ouralienne. Lingua 8: 403-423.
Mother Tongue XX (2015)

The next issue of *Mother Tongue* will be dedicated to the memory of its Founder, Harold C. ("Hal") Fleming. All ASLIP members and Mother Tongue readers are invited to contribute. Contributions may take the form of personal tributes or memories, or scholarly articles, or both.

Contributors may send submissions or inquiries to the Editor of *Mother Tongue*, J.D. Bengtson (see contact information on the inside front cover of this volume).

Annual Meeting

The annual meeting of ASLIP was held at 12:38 p.m., 9 November, 2014, at the Department of South Asian Studies, Harvard University, 1 Bow Street, Cambridge, Massachusetts, U.S.A. The meeting was attended by Michael Witzel (President), John D. Bengtson (Vice-President, *Mother Tongue* Editor), Michael T. Lewis (Secretary-Treasurer), Harold C. Fleming (ASLIP Founder, former President, and now on the Board of Directors), Nicholas Davidson (Board of Directors), B.K. Rana (Harvard University), and Caley Smith (Harvard University).

President Witzel reported that the tax-exempt status of ASLIP was now restored. It had lapsed, and its renewal was delayed by internal and external problems. ASLIP is now free to apply for foundation funding, and there was some discussion about conferences ASLIP could sponsor, for example:

- Conference discussing human genetics and/or archeology (e.g., Harvard’s Reich & Meadow)
- Conference on the Nostratic hypothesis (Bomhard, “Moscow School,” et al.)
- Conference on the Dene-Caucasian hypothesis (also “Dene-Yeniseian” & Swadesh’s “Vasco-Dene”)
- Conference on the Khoisan hypothesis: genetic family or areal convergence?

Funding is also needed for expanding the ASLIP homepage (http://aslin.org/) and for electronic publication of Mother Tongue.

ASLIP members are urged to contact the officers with ideas and suggestions about these and other possibilities (see contact information on the inside front cover of this volume).

The ASLIP Council of Fellows: “Council Fellow” is an honorary office, elected by ASLIP members as a whole. At the 2014 annual meeting the following were nominated for the Council:
Back Print Issues of MOTHER TONGUE

Back print issues of Mother Tongue are available for purchase. The following table summarizes some of the topics covered in issues I – XVII:

I (1995) **Inaugural Issue**: Canaanite & Bengali, Austric; Basque & Dene-Caucasian (R.L. Trask & 12 discussants); Proof in Genetic Linguistics (Greenberg)

II (1996): Kusunda, Ainu, Basque, Nihali (Mundlay & 8 discussants); Basque & Dene-Caucasian (S. Starostin, Trask, Ruhlen); Multilateral comparison (Greenberg)

III (1997): Kusunda, Nihali, Sumerian; “Hardware” / Origin of Language Symposium (Zegura, Lieberman, Donald, Fitch, Deacon); Recommendations for Long Rangers (Benedict); S.A. Starostin

IV (1998): Yeniseian; Ainu (Sidwell, Itabashi, Norquest, Bengtson); Deep classifications; Apophony (ablaut)

V (1999): Austric (Hayes, Blažek, Blust, van Driem, Fleming); Basque & Caucasian (Bengtson & 6 discussants); Sumerian (Srinivasan, Witzel, Diakonoff, Bengtson); Climatic influences on language; Bipeds, tools & speech; American prehistory

SPECIAL ISSUE (1999): **South Asian substrate languages** (Witzel, Whitehouse, van Driem, G.D.S. Anderson, Kuiper, Masica, Mundlay); Austronesian taxonomy

VI (2000/2001) **Festschrift for Roger W. Wescott**: Austric; Paleo-linguistics: The State
of the Art and Science (10 discussants); Obituaries: Wescott, Gordon, Greenberg

VII (2002) **In Honor of Joseph H. Greenberg:** Elamite, Dravidian, Ongota, Shabo, Tasmanian, Andamanese, Eurasian; Greenberg’s taxonomic proposals; Proto-Human or Proto-Sapiens

VIII (2003) Linguistic Databases & Taxonomy Workshop (SFI): Nostratic, Salishan & Basque, Khoisan, Negative Evidence (Whitehouse); EHL Project

IX (2004): Australian languages, Ongota, Shabo; Australian languages (O’Grady & Whitehouse); Proto-Sapiens kinship words: (P)APA, (T)ATA; Mario Alinei

X (2005): Kusunda, Basque, Eurasian; Obituaries: Livingstone, S. Starostin, Greenberg; Flores “hobbits”; Great Archeological Debate; Pre-Clovis site; Chinese genome; Trombetti

XI (2006) **Asian Remnant Languages & Year of the Australoid** (Harvard / ASLIP Conference, 2006): Indo-Pacific, South Asian languages, Tibeto-Burman, Austroasiatic, Kusunda, Austro, Australian, Dravidian, Andamanese; Archeology of Southern Route (Harrod); Out of East Africa by 77K BP (Brooks); Population genetics

XII (2007) **In Honor of Harold C. Fleming’s 80th Birthday:** Indo-European, Nostratic, Kartvelian, Bangi Me, Shompen, Dravidian; Nostratic Phonology (Bomhard, Sidwell, G. Starostin); Obituaries (Orel, Helimski, Bender); Glottochronology, Genetics

XIII (2008) **Commemoration of Ann Arbor Language & Prehistory Symposium (1988):** Milyan, Nostratic, Uralic, Chukcho-Kamchatkan, Shompen, Andamanese; Obituaries: Zvelebil, O’Grady; Bio-genetics; Fallacy of time limit; Myth of rapid linguistic change; Linguistic chronology

XIV (2009) **Commemoration of Daniel F. McCall:** Indo-European, Caucasian, Basque, El Molo, Mesames, mystery languages of East Africa; Berber *H (Fournet, Blažek, Kossmann, Prasse); Paleanthropology; Myth of rapid linguistic change II; Numerals (Hurrian, Nilotic); Profiles (Dolgopolsky, Mallory)

XV (2010) **Fifteenth Anniversary Issue 1995-2010:** Areal patterns of myth motifs (Berezin); Holocene etymology of ‘pitch’; Myth of rapid linguistic change III; Venisean numerals; Afrasian etymologies; Review of Campbell & Poser *Language Classification*

XVI (2011): Archeology & Genetics; Indo-European & Fenno-Ugric (Pedersen); Chinese giant Pangu; Minoan; Milyan; Surmic numerals; Dene-Caucasian; Myth of rapid linguistic change IV; Review of Jones & Milicic *Kinship, Language & Prehistory*

XVII (2012) **In Memory of Aharon Dolgopolsky:** Personal memories of Aharon Dolgopolsky; Archeology & Genetics; Trombetti’s “Puluga” and discussion; Kamchukchean and Eskaleutian; Discussion: The Number ‘One’ (Levitt, Blažek, Bomhard, Bürgisser, Janhunen)

XVIII (2013) **Commemorating the 50th Anniversary of J.H. Greenberg’s The Languages of Africa:** Obituaries: C.G. Turner, P.A. Michalove; Paris Linguistic Society Ban on discussion of language origins; The Niger-Congo hypothesis; Greenberg tribute and appraisal; Substrates and isolates of Africa; Proto-Nostratic morphology; The early dispersions of *Homo sapiens sapiens*; Origin of Milyan verbs; Global etymologies and Trombetti; Notes on Moscow conference (2000)

Back print issues can be had for $10 (domestic U.S.) or $20 (foreign). Please contact ASLIP Secretary-Treasurer Michael T. Lewis, 20 Duane Avenue, West Newton, MA 02465, U.S.A. | Tel. 617-964-0978 | lewismtc@rcn.com