

ASIA OR AFRICA? LOCATING THE AFRASIAN HOMELAND

Keywords: Lexicostatistics, Afrasian languages, Semitic languages, Middle Egyptian, Berber languages, Chadic languages, Cushitic languages, Omotic languages, dental anthropology, archaeology, population genetics.

To test the competing theories of the Afrasian homeland, the matrix of pairwise lexical matches between 58 Afrasian languages based on short (50-item) wordlists from *The Tower of Babel: The Global Lexicostatistical Database*, compiled by G.S. Starostin, A.S. Kasian, and M.A. Zhivlov, was subjected to several multivariate analyses. Rooted and unrooted networks were constructed, and the quasi-spatial model, which had rarely been used in lexicostatistics, was applied. A gradual relaxation of requirements to classification (from a rooted to an unrooted network, and from a hierarchical to a non-hierarchical ordering) reveals a southern cluster consisting of two families – Cushitic and Omotic. Under a more stringent approach there is no such cluster because Omotic proves the earliest branch of Afrasian, strengthening the African homeland theory. However, a comparison with dental, archaeological, and population genetic data tilts the balance somewhat in favor of Alexander Militarev's Natufian scenario. At the next stage, his own materials, based on 100-item lists will be subjected to the same analyses.

This article is a translation of: Козинцев А.Г. Азия или Африка? О локализации афразийской прародины // *Etnograficheskoe Obozrenie*. 2021. No 4. P. 24–41. DOI: 10.31857/S086954150016696-5

Afrasian classification: which model should be preferred?

The discussion about the location of the Afrasian (AA) homeland does not subside, and the balance of forces shifts now in favor of Western Asia, now in favor of Africa. The most influential proponent of the first point of view is A. Y. Militarev. In his opinion, Proto-Afrasian (PAA) was spoken by people associated with the Natufian culture of the Levant¹. After the disintegration of PAA in the 11th–10th millennia BC, the Proto-Cushito-Omotes migrated to Africa. Afraians who had remained in the Levant and created the Pre-Pottery Neolithic A there became the ancestors of the Semites, and those of them who had later migrated to Africa became speakers of languages of the Egyptian-Berber-Chadic branch (Militarev, Shnirelman 1984; Militarev 2002). This point of view was supported by V. Blažek (Blažek 2013a, 2013b) and P. Bellwood (Bellwood 2013: 111, 171–172). One of the important arguments in favor of the Western Asian theory is based on the Eurasian (Nostratic) ties of the AA macrofamily, which some linguists even included in the Nostratic group (Illich-

Alexander Kozintsev | <http://orcid.org/0000-0002-0165-8109> | agkozintsev@gmail.com | Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Sciences (3 University Emb., St. Petersburg, 199034, Russia)

SVitych 1971: 46–52; *Dolgopolsky* 2013: 49–63; *Bomhardt* 2018: 169–210). Other arguments are parallels of Afrasian (non-Semitic) languages with Sumerian (*Militarev* 1995), Elamite (*Blažek* 1999; *Starostin* 2002) and North Caucasian languages (*Militarev, Starostin* 1984).

An active advocate of the African homeland theory is C. Ehret. He ascribes a considerable ideological significance to this idea, since, in his opinion, it forces us to discard the long-standing prejudice that the ancient history of North Africa was only a reflection of the history of Western Asia. Ehret estimates the antiquity of the AA macrofamily at 15 thousand years, and places its homeland in the Horn of Africa. Proto-Semitic, according to Ehret, was its late northern branch (Ehret 2011: 16, 136, 139–140, 155–158). Indeed, as new glottochronological calculations show, it separated no earlier than late 6th millennium BC (*Militarev* 2005; *Kitchen et al.* 2009)², which, however, does not mean that there were no Afrasians in Western Asia before that. The earliest branch of the Semitic family – Akkadian – split off in the early 4th millennium, and other Semitic languages appeared in Arabia and Northeast Africa even later (*Ibid.*). A generally positive attitude to Ehret's theory was expressed by G. S. Starostin (*Starostin* 2017: 226). The position of H. C. Fleming (*Fleming* 2006: 140–142) and R. Blench (*Blench* 2006: 159–160) is rather similar. There are other versions of the African theory, in particular, southeastern Sahara (*Diakonoff* 1965: 99³; *Bender* 1997) and the Maghreb (*McBurney* 1975⁴; *Fellman* 1993).

Not being a linguist, I will nevertheless try to demonstrate the informative potential of some lexicostatistical methods that have not yet been widely used in the study of AA languages, but may be able to help solve the AA problem. In addition, I will point out new archaeological and biological data important in this context.

The work is based on a statistical analysis of a matrix of pairwise lexical correspondences according to short (50-word) lists. The matrix, kindly provided to me by G. S. Starostin, A. S. Kassian and M. A. Zhivlov, derives from their database *The Global Lexicostatistical Database* (GLD-1 n.d.). Lexical data on the following languages were used⁵:

I. Semitic: 1 – Northwest Semitic, 2 – Akkadian, 3 – Arabic, 4 – Modern South Arabian, 5 – Ethiosemitic.

II. Middle Egyptian.

III. Berber: 1 – East Berber, 2 – Tuareg, 3 – Zenaga, 4 – North Berber.

IV. Chadic:

– East: 1 – Tumak-Somrai, 2 – Lai, 3 – Kera-Kwang, 4 – Mokilko, 5 – Ubi-Sokoro, 6 – Mubi, 7 – Dangla-Migama;

– Central: 8 – Tera, 9 – Bura-Marghi, 10 – Higi, 11 – Mandara, 12 – Matakam, 13 – Sukur, 14 – Daba, 15 – Bata, 16 – Kotoko, 17 – Musgu, 18 – Gidar;

– Masa: 19 – Masa

– West: 20 – Hausa, 21 – Bole-Tangale, 22 – Angas-Sura, 23 – Ron, 24 – Bade-Ngizim, 25 – North Bauchi, 26 – South Bauchi.

V. Cushitic:

– North: 1 – Beja;

– Central: 2 – Agaw;

– East: 3 – Saho-Afar, 4 – Somaloid, 5 – Dasenech-Arbore, 6 – Oromoid, 7 – Werizoid, 8 – Highland East Cushitic, 9 – Yaaku;

– South: 10 – West Rift, 11 – Qwadza;

– Other: 12 – Dahalo, 13 – Ma'a.

VI. Omotic:

– South: 1 – South Omotic;

– North: 2 – Maoid, 3 – Dizoid, 4 – Kefoid, 5 – Yemsa, 6 – Bench-She, 7 – Chara, 8 – Omoto.

VII. Ongota.

The key question on which the solution of the AA homeland problem depends is the following: are the Omotic and Cushite families sisters? A. Y. Militarev answers this in the

positive (Militarev 2005: 398)⁶; he not only attributes the Omotic languages to the AA macrofamily, which some linguists deny (Theil 2012)⁷, but also considers it possible to speak about the former Cushomotic clade that arose in Western Asia. This opinion is shared by V. Blažek (Blažek 2013a, 2013b). According to this point of view, the specificity of the Omotic family was exaggerated by previous authors (in particular, H. C. Fleming); in addition, it is augmented by substratal influences, mainly from the Nilo-Saharan languages.

Proponents of the African homeland hypothesis, who classify the Omotic languages as AA, usually consider the Omotic branch as the earliest in the AA macrofamily (Fleming 2006: 140; Blench 2006: 148; Ehret 2011: 142; see also G. S. Starostin's unpublished data, cit. ex: Blažek 2013b: 126).

The location of the AA homeland also depends on the status of the virtually extinct Ongota language (Southwestern Ethiopia), which A. Y. Militarev attributes to the Omotic family (Militarev 2006: 510–511). Meanwhile, H. C. Fleming considered it a separate branch of AA macrofamily (Fleming 2006: 135–139), and B. Blažek links it with the Nilo-Saharan macrofamily (Blažek 2007). Disputing this, G. S. Starostin calls Ongota a linguistic isolate (Starostin 2013: 226–227). G. Savà and M. Tosco describe it as a Creole language that has Nilotic, Omotic and Cushitic features (Savà, Tosco 2000).

According to *The Global Lexicostatistical Database*, the Omotic family is indeed the most peculiar of all the branches of AA macrofamily. The unweighted average proportion of lexical matches between the Omotic branch and the remaining five AA branches is only 4.9% (2–3 words from a 50-word list), while the respective value is 8.1% for the Cushitic branch (4 words), 13.1% for Semitic (6–7 words), 13.2% for Chadic (also 6–7 words), 13.6% for Middle Egyptian (7 words), and 15.8% for the Berber branch (8 words). As for Ongota, it has 12.3% coincidences with Omotic languages, 7.0% with Cushitic languages, and no more than 2.4% with the other branches of the AA macrofamily.

On the family tree compiled by G. S. Starostin and his colleagues, the Omotic branch is the first to separate from the common AA stem, the Cushitic branch is the second, the Semitic is the third, followed by Middle Egyptian, and the Berber and Chadian branches are the last to split (see: GD-1). If we take into account separate languages, Ongota turns out to be the earliest branch of the Omotic family (see: GD-2).

Thus, in contrast to the tree compiled by A. Y. Militarev, there is no Cushomotic clade here: the similarity between the Cushitic and Omotic branches is not based on the evolutionarily derived features (synapomorphies), but on ancestral ones (symplesiomorphies)⁸. If we follow the requirements of a strictly phylogenetic (cladistic) taxonomy, such a similarity cannot serve as a basis for merging taxa. This contradicts the Western Asian homeland hypothesis because the homeland should be where the earliest branches had separated. There are two such branches here, the Omotic and the Cushitic, and both of them are located in North-East Africa. Can this possibly be the result of two successive migrations from the Levant?

Let's construct a network of AA families based on averaged data. Such a graph is more informative than an ordinary tree, since it takes into account not only "vertical" (genealogical) relations, but also "horizontal" (areal) ones⁹. We will root the network by the most isolated taxon — Ongota (Fig. 1).

Clearly, the result cannot be considered a confirmation of H. C. Fleming's idea of Ongota as an independent branch of the AA macrofamily, because we used this hypothesis as a prerequisite. If we had applied the usual pair-group cluster analysis, Ongota would have joined the Omotic family, but the Cushitic and Omotic families would still not have become sisters.

In the network, areal ties are shown as "collaterals" at the base of the branches. Such a graph, unlike the usual one, shows that the separation order of the Semitic, Berber and Chadic branches, as well as the of Middle Egyptian language, cannot be established as they are connected not only by close kinship, but also by areal contacts. Whereas in an ordinary tree the sequence of taxa within clusters is arbitrary (a cluster of any rank can be rotated 180°

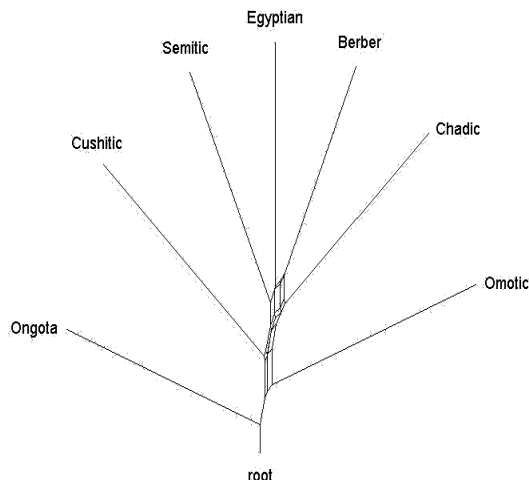


Fig. 1. Network of Afroasiatic families, rooted by Ongota

around the bifurcation point), in a network the branches are arranged in accordance with the geography and population history: the Omotic branch reveals areal connections with the Cushitic, the latter with the Semitic-Egyptian-Berber-Chadic cluster, within which the Semitic branch maintained contacts with the Egyptian, the Egyptian with the Berber, and the latter with the Chadic.

Let's look at the network of individual AA languages, which we will again root by Ongota (Fig. 2). The South Omotic language distinctly deviates towards the root here; this is to a lesser extent true of the North Omotic languages, i.e. the tendency is shown by the entire Omotic family compared to the Cushitic, let alone other AA families¹⁰. In the Cushitic family, a similar deviation is found in the South Cushitic languages, while the central Cushitic language Agaw, and especially the North Cushitic language Beja, are shifted towards the Semitic family. All this is quite consistent with the geographic distribution of languages. The Dahalo language takes an intermediate position between South Cushitic and East Cushitic languages, and the Ma'a language, which is close to it, is lexically adjacent to the South Cushitic languages.

The gradient "Semitic languages – Middle Egyptian – Berber – East-Chadic – Central-Chadic – West-Chadic" suggests that the ancestors of speakers of these languages migrated mainly from east to west¹¹, but the speakers of Chadic dialects could have arrived in Central Africa both from the east, i.e., from the Nile Valley, along the currently dry 1000 kilometer riverbed of Wadi Howar (Blench 1999: 71; Diakonoff 1998), and from the north, that is, from Cyrenaica, across Central Sahara (Ehret 2002: 79). The first scenario is based on the assumption that the Chadic languages originated from the Cushitic languages. But lexically they are much further from the Cushitic than from the northern AA. On the graph (see Fig. 2) there are no direct links between them.

The main obstacle to an unambiguous solution of the AA homeland problem is that the lexicostatistic classification of A. Y. Militarev is significantly different from that of G. S. Starostin et al. Serious disagreements within the Moscow Comparativist School¹² suggest not only that the methodology is insufficiently developed, but also that the material itself probably does not meet the conditions necessary for building a strict genealogical system of AA languages.

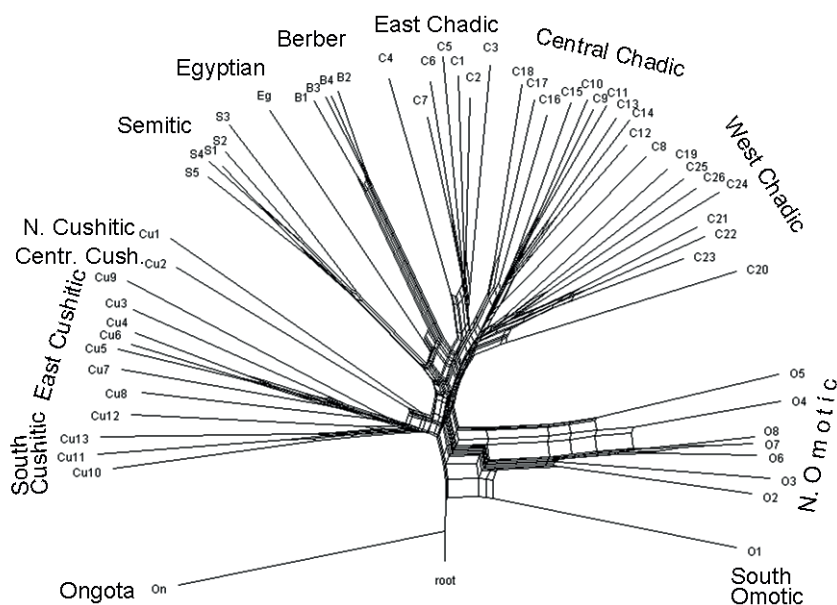


Fig. 2. Network of Afrasian languages, rooted by Ongota. *Semitic*: S1 — Northwest Semitic, S2 — Akkadian, S3 — Arabic, S4 — Modern South Arabian, S5 — Ethiosemitic; *Berber*: B1 — East Berber, B2 — Tuareg, B3 — Zenaga, B4 — North Berber; *East Chadic*: C1 — Tumak-Somrai, C2 — Lai, C3 — Kera-Kwang, C4 — Mokilko, C5 — Ubi-Sokoro, C6 — Mubi, C7 — Dangla-Migama; *Central Chadic*: C8 — Tera, C9 — Bura-Marghi, C10 — Higi, C11 — Mandara, C12 — Matakam, C13 — Sukur, C14 — Daba, C15 — Bata, C16 — Kotoko, C17 — Musgu, C18 — Gidar; *Masa*: C19 — Masa; *West Chadic*: C20 — Hausa, C21 — Bole-Tangale, C22 — Angas-Sura, C23 — Ron, C24 — Bade-Ngizim, C25 — North Bauchi, C26 — South Bauchi; *North Cushitic*: Cu1 — Beja; *Central Cushitic*: Cu2 — Agaw; *East Cushitic*: Cu3 — Saho-Afar, Cu4 — Somaloid, Cu5 — Dasenech-Arbore, Cu6 — Oromoid, Cu7 — Werizoid, Cu8 — Highland East Cushitic, Cu9 — Yaaku; *South Cushitic*: Cu10 — West Rift, Cu11 — Qwadza; *other Cushitic*: Cu12 — Dahalo, Cu13 — Ma'a; *South O. motic*: O1 — South O. motic; *North O. motic*: O2 — Maoid, O3 — Dizoid, O4 — Kefoid, O5 — Yemsa, O6 — Bench-She, O7 — Chara, O8 — Omoto.

Let's try to release the classification even further from the hierarchical principle imposed on it. To do that, we will not merely take into account areal ties, but will also refrain from determining the position of the common ancestor of all AA languages — perhaps the time has not yet come for this. We will now choose the unrooted network model (Fig. 3). Whereas the rooted tree mirrors the evolutionary process in the lateral projection, the unrooted one is a vertical projection, i.e. a cross-section through the crown. The root is hidden here, and the relations between the most ancient branches are not visible either.

The consequences of this change in perspective are obvious. The Ongota language, which we arbitrarily chose as the root language at the previous stage of the analysis, is now connected to the Omotic family, specifically to the South Omotic language, but, like the latter, deviates in the direction of the South Cushitic languages. Since the common ancestor is not indicated in such a model, the ancestral features are no less suitable for classifying groups than derived ones. As a result, the southern (Cushomotic) cluster, which was absent in the strictly genealogical model, is present here, and it opposes the northern (Semitic-Egyptian-Berber-Chadic) cluster, as in Militarev's classification. It would seem that the

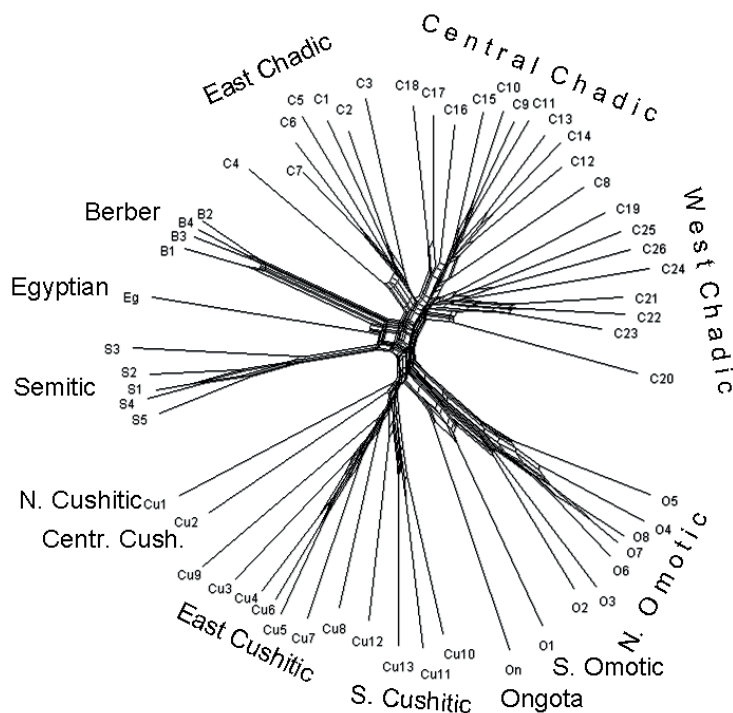


Fig. 3. Unrooted network of Afroasiatic languages (see Fig. 2 for explanations)

main obstacle to the adoption of the Levantine theory of the origin of AA languages disappears.

But with an informal approach to phylogeny, it is impossible to assess kinship with regard to the early stages of evolution. Therefore, the question of whether the early branches, Omotic and Cushitic, can be considered filial remains unresolved. The lesser antiquity of the Semitic and Berber families reconstructed by glottochronology (possibly despite the fact that the required conditions are not met) compared to the Omotic and Cushitic families (GD-1 b.d.; IA 2004; Kitchen *et al.* 2009; Blažek 2013a, 2013b; see also Fig. 1, 2) is also obvious in the unrooted network, and the same holds with regard to their lesser differentiation.

As a final step, we will make an even greater concession to the areal classification principle and apply the spatial model, in which languages are represented not as branches, but as dots (Fig. 4). Along with nonmetric scaling of the matrix of lexical correspondences, we use the minimum spanning tree showing the shortest pathway connecting the dots in a multidimensional space¹³. It was believed that the spatial model is suitable only for studying areal relationships between dialects or, at best, languages of the same family (Dyen *et al.* 1992), but recently it has been applied at a much higher taxonomic level (Kozintsev 2020).

The only thing that remains of the tree model here is the hierarchy: languages cluster into families, but this happens naturally and not because of the intrinsic properties of the model, as in any versions of cluster analysis, even the most undemanding ones, like the unrooted network.

There is no root or even a center in such a graph, the location of taxa is determined only by their lexical proximity. However, there is a correspondence with networks (this is especially evident in Figure 3). Specifically, the “southern” clade, consisting of the Omotic and Cushitic families — they are connected here by a direct link between Bench-She and Agaw,

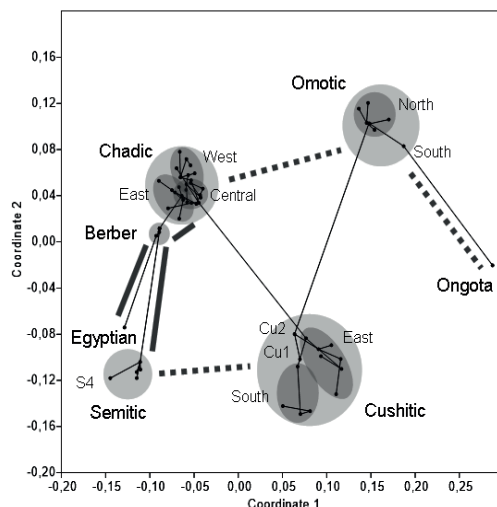


Fig. 4 Arrangement of languages and families based on the results of nonmetric multidimensional scaling of the matrix of lexical correspondences. Dots denote languages, spots denote families, and lines are edges of the minimum spanning tree showing the shortest path connecting taxa in the multidimensional space (thin lines refer to data on individual languages, thick solid lines refer to averaged data on families and isolated languages, connected at the level of 21–24% of lexical coincidences, thick dotted lines show ties with 7–12% of coincidences) (see Fig. 2)

respectively (13% of lexical coincidences) – opposes the “northern” one, which includes the remaining families. However, the average index of lexical proximity between the Omotic and Cushitic languages is only 6.8%.

According to their “spatial” arrangement, the taxa form the same sequence as in networks: North Omotic languages – South Omotic – Ongota – Cushitic family – Semitic family – Middle Egyptian language – Berber family – East Chadic languages – Central Chadic – West Chadic. The latter are separated from the Omotic family by the same hiatus that is also visible on networks. The Middle Egyptian language, being intermediate between the Semitic and Berber families, is closer to the Semitic, but it is nevertheless connected with the Berber family, which is related to the Semitic (the similarity between their individual languages is distributed rather uniformly, only the Modern South Arabian language looks somewhat isolated). The Berber-Chadic connection occurs through a link connecting the East Berber language with the Central Chadic language of Musgu (28%; however, there are many connections of this level between both families).

But the minimal spanning tree, based on the connections between individual languages, connects the families in a different way. For example, there is a direct link between the Cushitic family, in particular the Central Cushitic (Agaw) languages, and the Chadic, in particular the East Chadic languages of the Dangla-Migama group (17%). This hardly confirms Blench’s idea that the Chadic languages originated from the Cushitic ones, since the average number of Cushitic-Chadic lexical correspondences is small, 8.4% in total.

The areal model could give the most interesting results when discussing the issue of substrate and creolization. Thus, the use of Nilo-Saharan data would allow us to establish the direction in which Ongota deviates from the Omotic family. It would also be important to study the role of the West African substratum in Chadic languages. But all these are the tasks for the future.

Similarity indices between individual languages can largely depend on random factors. By averaging these indices across large taxa, we can decrease this dependence (see Fig. 4). The minimum spanning tree arranges the families in a chain with the following links: (1) Cushitic-Semitic (9.1% of correspondences); 2) Semitic – Berber (20.9%); (3) Berber – Chadic (21.8%); (4) Chadic – Omotic (7.1%), 5) Omotic – Ongota (12.3%). The sequence is broken by an offshoot connecting the Berber family with the Middle Egyptian language (23.5% — the strongest of all ties).

So, the four taxa belonging to the northern clade are relatively close (on average, 18.9% matches, and if we take into account only three African taxa, 20.2%). The fact that the Berber-Chadic connection is almost as strong as the Berber-Egyptian one deserves attention, especially given that the distribution areas of Berber-Egyptian and Chadic are separated by the Sahara.

The similarity between the three southern taxa is twice weaker than between the northern ones — on average, 8.7% matches, and if we ignore Ongota and take into account only the Cushitic-Omotic ties, 6.8%. The average similarity index between the northern and southern clusters is quite small — 4.7% (without Ongota, 6.4%), although the land or sea route along the western Red Sea coast, connecting the Horn of Africa with the Nile Delta, is hardly more difficult than the path across the Central Sahara — and still, the average Berber-Chadic connection is more than three times stronger — 21.8%). This could be taken to support Militarev's idea, i.e., that the southern AA languages spread by migration across Arabia and Bab-el-Mandeb rather than across Suez.

But did such a migration really take place? Wasn't the Horn of Africa at the dawn of AA history a kind of core area like Africa as a whole at the dawn of modern humankind? Doesn't the situation with the Omotic and Cushitic branches of the AA macrofamily resemble the situation with the oldest African groups that cannot be considered filial? For instance, Khoisans are opposed not to Pygmies, but to the rest of humankind taken together whereas the Pygmies are opposed to the entire humanity minus the Khoisans, etc. (*Tishkoff et al.* 2009: 1036).

We have come full circle: the reality of the southern (Cushomotic) clade is problematic, and this ambiguity cannot be eliminated by any analytic techniques. The average share of lexical coincidences between both these families (6.8%) is virtually the same as between the northern and southern clusters of AA languages (6.4%) and is even lower than the similarity of the Cushitic family with each of the northern taxa (8.4%; the respective figure for the Omotic family is 4.5%). Both the southern families, therefore, are not merely the most isolated but, moreover, show no specific similarity to one another. In addition, the Cushitic family is the most differentiated of all: the average lexical similarity between the Cushitic languages is only 25% (that between Omotic languages without Ongota is 44%, between Chadic languages, 35%, while the average similarity between Berber languages is 76%, and between Semitic languages, 59%).

So, if the southern (Cushomotic) cluster of AA languages is really a remnant of a former clade, then either this clade is extremely ancient, or the conditions of the region contributed to an accelerated language evolution and creolization. Certain linguists believe that the Cushomotic clade is fictitious not so much because it is based on plesiomorphies as because it is polyphyletic, i.e., does not derive from a single proto-language but represents a Sprachbund (*Orel, Stolbova* 1995: X).

Let me remind that I used the original data of G. S. Starostin and his closest colleagues; calculations by A. Y. Militarev yield different results (see above).

Dental, archaeological, and genetic data

The supporters of the Western Asian theory of AA homeland do not mention their very strong ally — the prominent dental anthropologist C. G. Turner (*Turner* 2008). He found

that the Late Pleistocene (Mesolithic) population of Nubia was dentally similar to the populations of Tropical Africa, while the Early Holocene inhabitants of that region were characterized by a South Caucasoid complex and resembled modern and ancient inhabitants of Western Asia including Natufians, which indicates migration from the Southern Levant to the Nile Valley. According to C. G. Turner, the migrants spoke Afrasian.

There are other facts favoring the Western Asian hypothesis. Common AA words for dog, bow and arrows have been reconstructed (*Militarev, Stolbova* 2007). The dog is known to have been introduced to Africa from Western Asia (*Bergström et al.* 2020). Skeletons of dogs were found in Natufian sites (*Clutton-Brock* 2017: 13–15); there is evidence of dog domestication at an even earlier stage of the Levantine Epipaleolithic, in the Kebaran microlithic culture (*Dayan* 1994). Recently, petroglyphs depicting scenes of people hunting ungulates (horses, antelopes, goats) with bows and arrows were discovered in northern Saudi Arabia¹⁴. Hunters are accompanied by dogs. Although there are no exact dates, certain facts suggest that these representations date to Pre-Neolithic times — 7th and possibly 8th millennia BC (*Guagnin et al.* 2018). The PAA language had already disintegrated by that time whereas the Semitic and the Egyptian-Berber-Chadic branches did not yet exist. This means that the petroglyphs could have been created either by southern Afrasians on their way to Bab-el-Mandeb, or by northern Afrasians, who had not yet separated.

In terms of the AA homeland issue, archaeological parallels between the Natufian culture and the microlithic industries of North Africa such as the Iberomaurusian, Ouchtata, Eastern Oranian, etc., deserve attention. African industries are older than the Natufian — they appeared in Algeria about 25 thousand years ago (*Hogue, Barton* 2016). The paleogenetic analysis of bone samples from Iberomaurusian burials in Grotte de Pigeons near Taforalt, Morocco, dating to 14–15 thousand years confirmed the affinities between those people and the Natufians (*Loosdrecht et al.* 2018).

In particular, the Y-chromosomal E-M78 haplogroup, found in males, is closely related to E-M123 haplogroup detected in Natufians (*Lazaridis et al.* 2016). Both of them are subclades of E-M35 haplogroup, traditionally considered a marker of AA migrations (*Keita* 2008; *Lancaster* 2009). But if its daughter haplogroups already existed in the Upper Paleolithic and Epipaleolithic, then the parental haplogroup, wherever it originated, is older than the AA macrofamily and is not necessarily related to it by origin. The same applies to subclades of the “Afrasian” mitochondrial haplogroups U6 and M1, identified in the Taforalt population (see below).

The archaeological affinity between the Iberomaurusian and Natufian cultures and the greater antiquity of the former as well as the genetic proximity between their carriers would seem to confirm the African theory of the AA homeland. Geneticists initially disagreed with this, arguing that the genome-wide analysis revealed two autosomal components in the Taforalt people — one close to the Natufian (63.5%), the other typical of the population of West Africa, but absent in Natufians (36.5%) (*Loosdrecht et al.* 2018; *Lazaridis et al.* 2016). However, a more detailed analysis of the Natufian genomes still found African components there, North African (21.2%) and even “Omotic” (6.8%) (*Shriner* 2018). Geneticists from David Reich’s team, having re-studied the same materials, also abandoned their original opinion (*Lazaridis et al.* 2016) and decided that the Taforalt population was the source of the North African component in both the natives of West Africa and the Natufians. If so, it is no longer necessary to speak about West African and Natufian admixtures in people from Taforalt (*Lazaridis et al.* 2018). In addition, L. Brace et al. found a slight “Tropical African” tendency in the small Natufian cranial series (*Brace et al.* 2006). It is possible that the genetic affinity between the Iberomaurusian and Natufian people was the heritage of some group that lived in the intermediate territory during the Upper Pleistocene.

According to the results of the genome-wide analysis, most of the gene pool of the modern population of North Africa is inherited from Caucasoids who had migrated to this territory from Western Asia between 38–18 thousand years ago (*Henn et al.* 2012; *Hodgson*

et al. 2014). The analysis of Y-chromosome haplogroups indicates a later time, about 15 thousand years; the only exception is the Berber population of Tunisia, whose ancestors lived there for at least 30 thousand years (*Fadhlaoui-Zid et al.* 2013). Due to the obvious youth of the Berber language family, its weak differentiation and proximity to other North-AA families, it is difficult to consider this as an argument in favor of the Maghreb hypothesis of AA homeland.

Genetic data relating to the Cushomotic cluster are contradictory. On the one hand, the Cushitic populations show some Nilo-Saharan tendency, and the Omotic ones are even more inclined in this direction, which adds to the doubts of linguists about their AA affiliation (*Baker et al.* 2017). Other AA groups, according to J. L. Baker and her colleagues, are of Western Asian origin (*Ibid.*). On the other hand, the genome-wide analysis revealed in almost all populations of the Horn of Africa including Cushitic and Omotic ones a very ancient Ethio-Somali component, which the geneticists associate with a Late Pleistocene population that had migrated from the Levant to Africa 20–30 thousand years ago. The Ethio-Somali marker is the M1 mitochondrial haplogroup. A related population marked by the U6 haplogroup settled in Maghreb (*Pennarun et al.* 2012; *Hodgson et al.* 2014).

According to J. E. Hodgson and his colleagues, it is most likely that a single proto-AA group, having migrated from the Levant to Africa, split into two subgroups, which peopled the Horn of Africa and the Maghreb, respectively, and since then had essentially lost contact with each other. It could be considered a confirmation of the theory of A. Y. Militarev, if it were not for the time of the alleged events — even C. Ehret does not claim that the disintegration of the AA macrofamily occurred so early. The dog (for which there is a common AA term) was domesticated much later. Another contradiction with the Levantine theory is that the proto-Semitic population, according to the same authors, arose due to a reverse migration to the Levant from Northeast Africa (*Hodgson et al.* 2014).

There are also more specific contradictions between the lexicostatistical and genetic data. There is no explanation for the considerable genetic similarity between the Tuareg and the Beja (*Cavalli-Sforza et al.* 1994: 173). On the other hand, the lexical proximity of Berber languages to Chadic ones is not confirmed by genetics. It was believed that the R-V88 Y-chromosome haplogroup indicates the southward proto-Chadic migration across the Sahara, according to Ehret's hypothesis (*Cruciani et al.* 2010). It turned out, however, that this is a marker of the Arab migration whereas the Chadic migration is marked by the L3f3 mitochondrial haplogroup, which originated in East Africa about 8 thousand years ago (*Černý et al.* 2009; *Shriner, Rotimi* 2018). Contrary to the lexical data, some Chadic peoples are genetically closest to the Cushites, which is consistent with the hypothesis of R. Blench (*Ibid.*). However, they do not show any genetic unity, which indicates intense assimilation processes (*Tishkoff et al.* 2009; *Shriner, Rotimi* 2018).

So, the question of the AA homeland is extremely intricate. Lexicostatistical data allow for various interpretations, but the totality of facts seems to tip the balance somewhat in favor of the Levantine hypothesis.

Notes

¹ The archaeological part of this hypothesis was elaborated under the participation of V.A. Schnirelmann.

² See: *Diakonoff* 1965: 101 for a similar assessment.

³ Subsequently, the views of I. M. Diakonoff changed and became closer to those of A.Y. Militarev. He began to associate Afrasians with the Natufian culture as well as the Sebilian and some other Upper Paleolithic and Mesolithic industries of Upper Egypt (*Diakonoff* 1998). Diakonoff regarded the culture of the Pre-Pottery Neolithic of Jericho as Semitic and even West Semitic. Similar views were expressed by J. Zarins (*Zarins* 1990). Such an early separation of the Semitic branch is not confirmed by glottochronological data (*Militarev* 2005; *Kitchen et al.* 2009).

⁴ C.B.M. McBurney, based on archaeological parallels, believed that Proto-Afrasians had migrated to Maghreb from Southern Europe, in particular from Sicily, during the Upper Paleolithic.

This idea was initially perceived positively (McBurney 1975: 506–511), but it has no linguistic confirmation, and was later refuted by genetic facts (Loosdrecht et al. 2018).

⁵ Classifications that I relied on were taken mainly from the works by A. Y. Militarev (Militarev 2005), R. Blench (Blench, Dendo 2006) and P. Newman (Newman 2013).

⁶ See the family tree of AA languages compiled by him at: https://web.archive.org/web/20090104234232/http://community.livejournal.com/terra_linguarum/95880.html (last accessed 12.01.2021).

⁷ G. S. Starostin, without directly denying the AA-affiliation of the Omotic languages, considers it far from obvious (Starostin 2013: 38; cf.: Ibid.: 148, 447, 449).

⁸ Such a terminology, strictly speaking, is suitable only in relation to character-based (cladistic) methods, and not to distance-based (phenetic) ones, as in this case. However, to the extent that generalized similarity measures indicate kinship, the position of branches that are not filial, at least in the phenetic sense, near the root of the dendrogram means that their proximity is based on symplesiomorphies.

⁹ Networks were built using the SplitsTree4 package of D. Huson and D. Bryant (<https://software-ab.informatik.uni-tuebingen.de/download/splitstree4/welcome.html>; last accessed 12.01.2021).

¹⁰ As I said, networks are much more informative than usual trees, in relation to which it is impossible to speak about the directionality of changes because the position of branches within clusters is arbitrary. It may be objected that the network is unsuitable for a glottochronological analysis. Indeed, it reveals precisely what limits this analysis but is usually ignored — “horizontal” (areal) ties, which in our case are very dense.

¹¹ It would be interesting to trace these links even further west, towards the Canary Islands, but work with the Guanche material has just begun (Militarev 2020).

¹² Disagreements are found even between co-authors of the same article. Thus, S. L. Nikolaev, unlike A. Y. Militarev, adheres to the African hypothesis (Militarev, Nikolaev 2020: 199, footnotes 1, 2).

¹³ The model was implemented using the PAST package of Ø. Hammer, version 4.05 (<https://www.nhm.uio.no/english/research/infrastructure/past> (last accessed 18.01.2021)). One should use either this or newer versions since previous ones calculate the shortest path not in the original multidimensional space, but on the plane, which leads to a significant loss of information (I thank Ø. Hammer for the clarification). The meaning of the term “tree” in the spatial model is completely different from that in the genealogical model.

¹⁴ See the reconstructed PAA terms for ungulates: Militarev, Nikolaev 2020.

Acknowledgments

I am grateful to V. Blažek, A. Y. Militarev, G. S. Starostin, A. S. Kassian and M. A. Zhivlov for providing the unpublished materials. Special thanks are due to A. Y. Militarev for valuable considerations.

References

- Baker, J.L., C.N. Rotimi, and D. Shriner. 2017. Human Ancestry Correlates with Language and Reveals that Race is not an Objective Genomic Classifier. *Scientific Reports* 7: 1572. <https://doi.org/10.1038/s41598-017-01837-7>
- Bellwood, P. 2013. *First Migrants: Ancient Migration in Global Perspective*. Chichester: Wiley – Blackwell.
- Bender, M.L. 1997. Upside-down Afrasian. *Afrikanistische Arbeitspapiere* 50: 19–34.
- Bergström, A., et al. 2020. Origins and Genetic Legacy of Prehistoric Dogs. *Science* 370 (6516): 557–564.
- Blažek, V. 1999. Elam: A Bridge between Ancient Near East and Dravidian India? In *Archaeology and Language IV. Language Change and Cultural Transformation*, edited by R. Blench and M. Spriggs, 48–78. London: Routledge.
- Blažek, V. 2007. Nilo-Saharan Stratum of Ongota. In *Advances in Nilo-Saharan Linguistics*, edited by M. Reh and D. L. Payne, 1–10. Köln: Köppe.
- Blažek, V. 2013. Indo-European Zoonyms in Afroasiatic Perspective. *Journal of Language Relationship* 9: 37–54.
- Blažek, V. 2013. Levant and North Africa: Afroasiatic Linguistic History. In *The Encyclopedia of Global Human Migration*. Vol. 1, Prehistory, edited by P. Bellwood, 125–132. London: Blackwell.

- Blench, R. 1999. The Westward Wanderings of Cushitic Pastoralists: Explorations in the Prehistory of Central Africa. In *L'homme et l'animale dans le bassin du Lac Tchad* [Humans and Animals in the Lake Chad Basin], edited by C. Baroin and J. Boutrais, 39–80. Paris: Éditions IRD.
- Blench, R. 2006. *Archaeology, Language, and the African Past*. Lanham: Altamira Press.
- Bomhardt, A. 2018. *A Comprehensive Introduction to Nostratic Comparative Linguistics with Special Reference to Indo-European*. Vol. 1. Florence, SC.
- Brace, C.L., et al. 2006. The Questionable Contribution of the Neolithic and Bronze Age to European Craniofacial Form. *Proceedings of the National Academy of Sciences of USA* 103 (1): 242–247.
- Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Černý, V., et al. 2009. Migration of Chadic Speaking Pastoralists within Africa Based on Population Structure of Chad Basin and Phylogeography of Mitochondrial L3f Haplogroup. *BMC Evolutionary Biology* 9: 63. <https://doi.org/10.1186/1471-2148-9-63>
- Clutton-Brock, J. 2017. Origins of the Dog: The Archaeological Evidence. In *The Domestic Dog: Its Evolution, Behavior and Interactions with People*, edited by J. Serpell, 7–21. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781139161800.002>
- Cruciani, F., et al. 2010. Human Y-Chromosome Haplogroup R-V88: A Paternal Genetic Record of Early Mid-Holocene Trans-Saharan Connections and the Spread of Chadic Languages. *European Journal of Human Genetics* 18: 800–807. <https://doi.org/10.1038/ejhg.2009.231>
- Dayan, T. 1994. Early Domesticated Dogs of the Near East. *Journal of Archaeological Science* 21 (5): 633–640.
- Diakonoff, I. 1998. The Earliest Semitic Society: Linguistic Data. *Journal of Semitic Studies* 43 (2): 209–219.
- Diakonov, I.M. 1965. *Semito-khamitskie yazyki. Opyt klassifikatsii* [Semito-Hamitic Languages. An Attempt at a Classification]. Moscow: Nauka.
- Dolgopolskii, A.B. 2013. *Indoeuropeiskii slovar' s nostraticheskimi etimologiyami* [An Indo-European Dictionary with Nostratic Etymologies]. Vol. I. Moscow: Rukopisnye pamiatniki Drevnei Rusi.
- Dyen, I., J.B. Kruskal, and P. Black. 1992. An Indo-European Classification: A Lexicostatistical Experiment. *Transactions of the American Philosophical Society* 82 (5): 1–132.
- Ehret, C. 2002. *Civilizations of Africa: A History to 1800*. Oxford: James Currey.
- Ehret, C. 2011. *History and the Testimony of Language*. Berkeley: University of California Press.
- Fadhlaoui-Zid, K., et al. 2013. Genome-Wide and Paternal Diversity Reveal a Recent Origin of Human Populations in North Africa. *PLoS ONE* 8 (11): e80293.
- Fellman, J. 1993. Linguistics as an Instrument of Prehistory. *Orbis* 36: 56–58.
- Fleming, H.C. 2006. *Ongota: A Decisive Language in African Prehistory*. Wiesbaden: Harrassowitz Verlag.
- Guagnin, M., A.R. Perry, and M.D. Petraglia. 2018. Pre-Neolithic Evidence for Dog-Assisted Hunting Strategies in Arabia. *Journal of Anthropological Archaeology* 49: 225–236. <https://doi.org/10.1016/j.jaa.2017.10.003>
- Henn, B.M., et al. 2012. Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genetics* 8 (1): e1002397. <https://doi.org/10.1371/journal.pgen.1002397>
- Hodgson, J.E., C. J. Mulligan, A. Al-Meer, and R. L. Raau. 2014. Early Back-to-Africa Migration into the Horn of Africa. *PLoS Genetics* 10 (6): e1004393. <https://doi.org/10.1371/journal.pgen.1004393>
- Hogue, J.T., and R.N.E. Barton. 2016. New Radiocarbon Dates for the Earliest Later Stone Age Microlithic Technology in Northwest Africa. *Quaternary International* 413: 62–75.
- Illich-Svitych, V.M. 1971. *Opyt sravneniia nostraticheskikh yazykov (semitokhamitskii, kartvel'skii, indoeuropeiskii, ural'skii, dravidiiskii, altaiskii). Vvedenie. Sravnitel'nyi slovar' (b–k)* [An Attempt at a Comparison of Nostratic Languages (Hamito-Semitic, Kartvelian, Indo-European, Uralic, Dravidian, Altaic)]. Moscow: Nauka.
- Keita, S.O. 2008. Geography, Selected Afro-Asiatic Families, and Y Chromosome Lineage Variation: An Exploration in Linguistics and Phylogeography. In *Hot Pursuit of Language in Prehistory: Essays in the Four Fields of Anthropology in Honor of Harold Crane Fleming*, edited by J. D. Bengtson, 3–16. Amsterdam: John Benjamins.
- Kitchen, A., C. Ehret, S. Assefa, and C. J. Mulligan. 2009. Bayesian Phylogenetic Analysis of Semitic Languages Identifies an Early Bronze Age Origin of Semitic in the Near East. *Proceedings of the Royal Society* 276 (1668): 2703–2710.

- Kozintsev, A. 2020. On the Homelands of Indo-European and Eurasiatic: Geographic Aspects of a Lexicostatistical Classification. *Journal of Indo-European Studies* 48 (1–2): 121–150.
- Lancaster, A. 2009. Y Haplogroups, Archaeological Cultures, and Language Families: A Review of the Possibility of Multidisciplinary Comparisons Using the Case of E-M35. *Journal of Genetic Genealogy* 5 (1): 35–65.
- Lazaridis, I., et al. 2016. Genomic Insights in the Origin of Farming in the Ancient Near East. *Nature* 536 (7617): 419–424.
- Lazaridis, I., et al. 2018. Paleolithic DNA from the Caucasus Reveals Core of West Eurasian Ancestry. *bioRxiv*. <https://doi.org/10.1101/423079>
- Loosdrecht, M., et al. 2018. Pleistocene North African Genomes Link near Eastern and Sub-Saharan African Human Populations. *Science* 360 (6388): 548–552.
- McBurney, C.B.M. 1975. The Archaeological Context of the Hamitic Languages in Northern Africa. In *Hamito-Semitic*, edited by J. Bynon and T. Bynon, 495–511. Paris: Mouton.
- Militarev, A.Y. 1995. Shumery i afraziitsy [Sumerians and Afrasians]. *Vestnik drevnei istorii* 2: 113–127.
- Militarev, A. 2002. The Prehistory of a Dispersal: The Proto-Afrasian (Afro-Asiatic) Farming Lexicon. In *Examining the Farming – Language Hypothesis*, edited by P. Bellwood and C. Renfrew, 135–150. Cambridge: McDonald Institute for Archaeological Research.
- Militarev, A. 2005. Once More about Glottochronology and the Comparative Method: The Omotic-Afrasian Case. In *Orientalia et Classica. Trudy Instituta vostochnykh kul'tur i antichnosti* [Transactions of the Institute of Oriental Cultures and Classical Antiquity], Vol. VI, *Aspekty komparativistiki* [Aspects of Comparative Linguistics], 339–408. Moscow: RGGU.
- Militarev, A. 2006. Towards the Genetic Affiliation of Ongota, a Nearly Extinct Language of Ethiopia (II). In *Orientalia et Classica. Trudy Instituta vostochnykh kul'tur i antichnosti* [Transactions of the Institute of Oriental Cultures and Classical Antiquity]. Vol. XIV, *Vavilon i Bibliia* 3 [Babylon and the Bible 3], 489–512. Moscow: RGGU.
- Militarev, A. 2020. Libyo-Berbers – Tuaregs – Canarians: Linguistic Evidence. *Études et Documents Berbères* 43: 125–152.
- Militarev, A., and S. Nikolaev. 2020. Proto-Afrasian Names of Ungulates in View of the Proto-Afrasian Homeland Issue. *Journal of Language Relationship* 18 (3): 199–226.
- Militarev A.Y., and V.A. Shnirelman. 1984. K probleme lokalizatsii drevneishikh afraziitsev (opyt lingvo-arkheologicheskoi rekonstruktsii) [On Localizing the Earliest Afrasians (An Attempt at a Linguo-Archaeological Reconstruction)]. In *Lingvisticheskaya rekonstruktsiya i drevneishaia istoriia Vostoka*. [Linguistic Reconstruction and the Earliest History of the East], 2: 35–53. Moscow: Nauka.
- Militarev, A.Y., and S. A. Starostin. 1984. Obshchaia afraziisko-severnokavkazskaia kul'turnaia leksika [Common Afrasiatic – North Caucasian Cultural Terms]. In *Lingvisticheskaya rekonstruktsiya i drevneishaia istoriia Vostoka* [Linguistic Reconstruction and the Earliest History of the East]. Pt. 3, *Yazykovaia situatsiia v perednei Azii v X–IV tys. do n. e.* [Linguistic Situation in West Asia in the 10th–4th Millennia BC], 34–43. Moscow: Nauka.
- Militarev, A., and O. Stolbova, eds. 2007. Afroasiatic etymology. In *The Tower of Babel: Evolution of Human Language Project*. 12.04.2007. <https://starling.rinet.ru/cgi-bin/response.cgi?root=config&morpho=0&basename=/data/semham/afaset&first=1>
- Newman, P. 2013. The Chadic Language Family: *Classification and Name Index* (Electronic publication). <https://core.ac.uk/download/pdf/213847723.pdf>
- Orel, V.E., and O.E. Stolbova. 1995. *Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction*. Leiden: E. J. Brill.
- Pennarun, E., et al. 2012. Divorcing the Late Upper Paleolithic Demographic Histories of mtDNA Haplogroups M1 and U6 in Africa. *BMC Evolutionary Biology* 12: 234. <https://doi.org/10.1186/1471-2148-12-234>
- Savà, G., and M. Tosco. 2000. A Sketch of Ongota, a Dying Language of Southwest Ethiopia. *Studies in African Linguistics* 29 (2): 59–135.

- Shriner, D. 2018. Re-analysis of Whole-Genome Sequence Data from 279 Ancient Eurasians Reveals Substantial Ancestral Heterogeneity. *Frontiers in Genetics* 9 (268). <https://doi.org/10.3389/FGENE.2018.00268>
- Shriner, D., and C. N. Rotimi. 2018. Genetic History of Chad. *American Journal of Physical Anthropology* 167 (4). <https://doi.org/10.1002/ajpa.23711>
- Starostin, G. 2002. On the Genetic Affiliation of the Elamite Language. *Mother Tongue* 7: 147–170.
- Starostin, G. 2017. Macrofamilies and Agricultural Lexicon: Problems and Perspectives. In *Language Dispersal Beyond Farming*, edited by M. Robbeets and A. Savelyev, 215–233. Amsterdam: John Benjamins.
- Starostin, G.S. 2013. Yazyki Afriki. Opyt postroeniia leksikostatisticheskoi klassifikatsii [An Attempt at a Lexicostatistical Classification]. Vol. 1, *Metodologiya. Koisanskie yazyki* [Methodology. Khoisan Languages]. Moscow: Yazyki slavianskoi kul'tury.
- Theil, R. 2012. Omotic. In *Semitic and Afroasiatic: Challenges and Opportunities*, edited by L. Edzard, 369–384. Wiesbaden: Harrassowitz Verlag.
- Tishkoff, S.A., et al. 2009. The Genetic Structure and History of Africans and African Americans. *Science* 324 (5930): 1035–1044.
- Turner, C.G. 2008. A Dental Anthropological Hypothesis Relating to the Ethnogenesis, Origin, and Antiquity of the Afro-Asiatic Language Family: Peopling of the Eurafrican – South Asian Triangle IV. In *In Hot Pursuit of Language in Prehistory. Essays in the Four Fields of Anthropology in Honor of Harold Crane Fleming*, edited by J. D. Bengtson, 17–23. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.145.06tur>
- Zarins, J. 1990. Early Pastoral Nomadism and the Settlement of Lower Mesopotamia. *Bulletin of the American Schools of Oriental Research* 280: 31–65.