

Drastic demographic events triggered the Uralic spread

Riho Grünthal¹, Volker Heyd¹, Sampsa Holopainen^{1,2},
Juha A. Janhunen¹, Olesya Khanina^{1,3}, Matti Miestamo¹,
Johanna Nichols^{1,4,5}, Janne Saarikivi¹ and Kaius Sinnemäki¹
¹ University of Helsinki | ² University of Vienna | ³ Russian Academy of
Sciences | ⁴ University of California | ⁵ HSE University

The widespread Uralic family offers several advantages for tracing prehistory: a firm absolute chronological anchor point in an ancient contact episode with well-dated Indo-Iranian; other points of intersection or diagnostic non-intersection with early Indo-European (the Late Proto-Indo-European-speaking Yamnaya culture of the western steppe, the Afanasievo culture of the upper Yenisei, and the Fatyanovo culture of the middle Volga); lexical and morphological reconstruction sufficient to establish critical absences of sharings and contacts. We add information on climate, linguistic geography, typology, and cognate frequency distributions to reconstruct the Uralic origin and spread. We argue that the Uralic homeland was east of the Urals and initially out of contact with Indo-European. The spread was rapid and without widespread shared substratal effects. We reconstruct its cause as the interconnected reactions of early Uralic and Indo-European populations to a catastrophic climate change episode and interregionalization opportunities which advantaged riverine hunter-fishers over herders.

Keywords: Uralic, Finno-Ugric, Indo-European, Yamnaya, Indo-Iranian, Siberia, Eurasia, Seima-Turbino, 4.2 ka event, linguistic homeland

1. Introduction

The Uralic language family, probably about 4,500 years old,¹ comprises over 30 daughter languages. At one time or another they, or their immediate ancestors, have extended from the Yenisei River basin in western Siberia to Norway and from near the middle Volga to the Arctic Ocean. There was a window in time, from late medieval to early modern times, when this large region was populated almost exclusively by Uralic languages – when the Uralic northward expansion was absorbing the remaining non-Uralic languages in the far north, just as Swedish, Russian, and Siberian Turkic were beginning their expansions in the west and south.² At present, except for the three national languages Finnish, Estonian, and Hungarian, most of the languages are those of minorities scattered through their former ranges. Most are endangered, and some are extinct.³

The Uralic family is divided into nine elementary branches (Table 1), all between about 2,500 and 1,000 years old internally (i.e., since their own divergence into individual languages), while their external ages (since the initial Uralic split) are all about 4,000–4,500 years.⁴ (Here and below, in tables and prose we list languages from east to west, following the standing directionalities of language spread in northern Eurasia.) See Supplement S1 (<https://doi.org/10.5281/zenodo.6345559>) for tree diagrams and a full list of languages.⁵ Any valid family tree must distinguish these elementary branches from each other; what higher-level structure exists is debated (and discussed below and in Supplement S1). For the branch historical homelands see Figure 1.

The traditional view of the family tree posits an initial bifurcation into Samoyedic vs. Finno-Ugric, a clade comprising the rest of the branches. The bifurcation is based primarily on the fact that Samoyedic has far fewer cognates with other branches than any two Finno-Ugric branches have with each other, which suggests a longer separate development of Samoyedic from the rest. The low cognate count is then due to losses over time. After that initial split, succes-

1. See §2.2, 2.4 below. Kallio (2015: 80–81) gives an overview of dates advocated in the literature. The age increases to 6000 years or more if, as traditionally, it is based on lexical retention rates and the traditional family tree (See Supplement S1: <https://doi.org/10.5281/zenodo.6345559>).

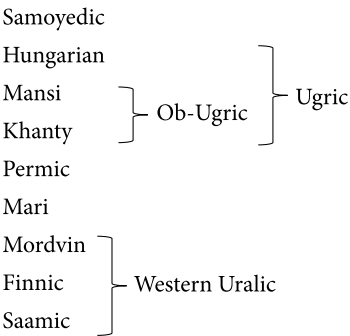
2. Even then, Turkic expansions, chiefly Volga and Siberian Tatar, had absorbed Uralic speakers south of the Volga and in the southern Urals.

3. A map of the modern ranges is: <https://bedlan.net/uralic/> A map of historical ranges is Grünthal and Salminen (1993).

4. This time depth makes them all genera by the definition of Dryer (1989, 2013). The age is supported below.

5. Online supplements contain additional graphs, tables, data, and details (<https://doi.org/10.5281/zenodo.6345559>).

Table 1. The Uralic daughter branches, in approximately east to west order of branch homelands. Brackets = likely or known higher-level groupings. Ugric (and within it Ob-Ugric = Khanty and Mansi) is at least an areal grouping and possibly genealogical (areal: Salminen 2001; Helimski 1982, 2003: 161; J. Häkkinen 2009; genealogical: Honti 1998). Western Uralic is more clearly genealogical, but the internal structure is debated



sive bifurcations within Finno-Ugric produced a west-branching family tree with Finnic and Saamic the last to separate from each other. Since, however, no or almost no shared innovative sound changes define Finno-Ugric or its subsequent bifurcations, a recent proposal sees the family as a star phylogeny or rake consisting of nine coequal branches. (For trees see Supplement S1.) Here and below, we speak of Finno-Ugric as the set of non-Samoyedic branches, without commitment as to whether it is a clade; and likewise for Ugric and Ob-Ugric.

Whether the tree is hierarchical or flat (a rake) has implications for what cognate sets are accepted for Proto-Uralic (henceforth PU). If PU first split into Samoyedic vs. Finno-Ugric, cognates accepted as PU must be attested in both branches, and that necessarily means in Samoyedic and at least one Finno-Ugric branch. This gives Samoyedic veto power over what is considered PU and no doubt produces false negatives (especially given the low number of cognates in Samoyedic). If the tree is flat, however, no one branch has veto power, and any reasonably diverse and dense attestation suffices; then there will be false positives. Recent and ongoing work by Ante Aikio (2013, 2014a,b, 2015a,b, in press) uses a criterion of sufficiently diverse attestation. On the traditional definition there are about 200 firm PU etyma (see Appendix 1); Aikio has from 500 to 700.

The large Uralic range is remarkable for a language family traditionally thought to have expanded from a Neolithic hunter-gatherer population.⁶ Other

6. Ancestral Uralic society is sometimes described as Mesolithic. The technology was Neolithic in that it included pottery, but there was no food production. Consistent with the archaeological evidence, the reconstructable Proto-Uralic vocabulary included a word for ‘pot’ but no terms for domesticates or foods prepared from them.

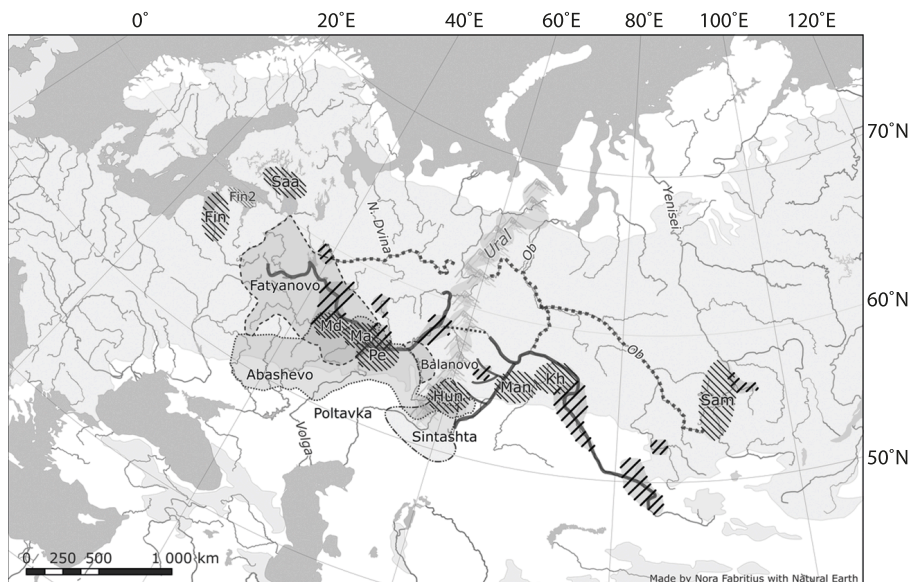


Figure 1. Linguistic and archaeological distributions c. 4,200–4,000 BP. Light shaded band across the entire area: Recent and modern forest zone (steppe to south, tundra to north). Languages and cultures in chronological order: Fatyanovo, Balanovo, Abashevo: Pre-Baltic and Para-Baltic-speaking post-Corded Ware Indo-European cultures. Poltavka, Sintashta: Indo-Iranian-speaking. Shapes with upward coarse parallel hatching: Seima-Turbino major sites. Labeled ovals (finer hatching): core locations of Uralic branch ancestors: Saa(mi), Fin(nic), M(or)d(vin), Ma(ri), Pe(rmic), Hun(garian), Man(si), Kh(anty), Sam(oyedic). Fin2 = later staging ground for Finnic. Heavy line along rivers: southern trade/travel route. Heavy dotted lines: northern route. Light, short dotted line: combined water and overland route across the Urals. (After Saarikivi in press.; Lang 2018; Chernykh 2008.)

large spreads have occurred in northern Eurasia, but they have involved advances in food production (reindeer herding in the north), connections to Chinese imperial expansion (see Barfield 1989; Janhunen 2008, 2012b), or, in the case of Indo-European, the advances in mobility, technology, economy, and network-extension mechanisms that impelled the Yamnaya culture across the western steppe (Anthony 2007: 300–339; Parpola 2012; Anthony & Ringe 2015: 208).

Even more remarkable is the absence of unambiguous early Indo-European (henceforth IE) loanwords in PU. The sweep of the Yamnaya culture, which

spread late Proto-Indo-European (PIE) speech, from c. 5100 BP,^{7,8} was a major economic, cultural, and demographic event that brought wheeled transport, domesticated horses, pastoralism, new forms of social organization, and new forms of exchange and wealth to the entire western steppe and nearby.⁹ In its wake, the PIE language spread to central Europe and from there, as part of the wider Corded Ware culture complex, back eastward along the Volga to the Urals, where the cultural successors of the IE-speaking Fatyanovo culture mined and worked copper.¹⁰ Also remarkable is the tenacity of Uralic-speaking populations along the Volga. Speakers of the ancestral major branches of Uralic succeeded IE to occupy good farming and herding land along the major trade arteries (the Volga, Oka, Kama) and were never dislodged until the Russian expansion,¹¹ while major language spreads and extinctions have taken place both north and south. One would expect to find at least remnants of IE speech along the Volga and to the north, accompanied by aspects of IE social organization, as local hunting, trap-

7. BP = before present, i.e., years ago. These are estimates in calendar years. All are approximate and expressed in rounded numbers. Our sources express prehistoric and protohistoric dates variously as BP, BC/AD, BCE/CE, BZ/AZ, or parts of millennia (e.g., last quarter of third millennium). We convert them all to BP. This is consistent with historical-linguistics practice, where language family ages are estimated in years ago. Dates not accompanied by a before/after abbreviation are in the present era (we use such dates only for historical times).

8. Or 5,300, counting the predecessor Repin culture. The large spread begins with Yamnaya c. 5,100 BP.

9. Geographical terms (all approximate): western steppe = steppe from eastern Romania to the Urals; its subparts: Pontic steppe (west; southeastern Europe, Ukraine, and Russia), Caspian steppe (east; lower Volga to Urals). Western Siberia = Urals to Yenisei. Lower Volga = the north-south extent in the Caspian steppe up to the Kama confluence (Astrakhan to Kazan); middle Volga = the east-west extent in the forest-steppe region (from the big bend at Kazan to the Oka confluence at Nizhnii Novgorod; also the lower Oka and lower Kama); upper Volga = the more tortuous course from Nizhnii Novgorod east (and via portages to the Baltic).

10. Here and below, when we speak of a prehistoric population as “Uralic-speaking”, “IE-speaking”, “pre-Samic-speaking”, etc. we do not imply that the population was monolingual or that we know its speech situation with certainty. Rather, that language was the dominant one as determined by best methods (and usually justified here).

11. Though the rise of Volga Bulgar power (650–900 CE = 1350–1100 BP) and the Tatar Golden Horde (1,200–1,500 AZ = 800–500 BP) pushed the southern Uralic frontier northward along the middle Volga. East of the Urals, the original Ugrić range may have been in the south of the forest zone and not far from the steppe, as suggested by three pieces of evidence: early loans in Ugrić languages dealing with horse culture and nomadic life; the persistence of the Indo-Iranian word for ‘honeybee’ in Hungarian, the only eastern Uralic language whose earliest known location was in honeybee range; non-Ugrić substratal toponyms in much of the Ob-Ugrić territory (§2.1, §3.2.7, and Supplements S2 and S3.2). If that is correct, the earliest Ob-Ugrić core areas have not persisted but have shifted to Turkic, chiefly Siberian Tatar.

ping, and fishing societies shifted to the language of wealth and power. Instead we have Uralic languages from the Volga north, with only recent Russian and Turkic overlays. Here we present a model of the Uralic spread that accounts for these things.

The paper consists of the main article and linked online materials¹²: supplements with further information and references; appendixes of Proto-Uralic, early Indo-Iranian, and early Ugric vocabulary; and the R script and the spreadsheet data used in Supplement S4.

2. Prehistory and timeline

2.1 The PU homeland

Probably the majority view among Uralicists through the 20th century, and still widely taken as received view by archaeologists, is that the PU homeland was on or near the middle Volga and Volga bend. There are four current arguments for a PU homeland in that region.

The first is the region of highest phylogenetic divergence, or center of gravity: the projection on the ground of the highest-level divergence in the family tree is the homeland (Sapir 1916; Dyen 1956). This criterion does not always obtain, but the success rate is good for directional spreads such as those of Austronesian (on the traditional view: Blust 1985, 2009) and Algonquian (Goddard 1996). Highest phylogenetic divergence requires a family tree with an initial split like that of the traditional Uralic tree. For a phylogeny like the Uralic rake proposal there is no single highest divergence.

The second is the region of greatest diversity: the area where the greatest number of branches are clustered together is the homeland. In the Uralic literature that is usually identified as the Volga bend area, where Mordvin, Mari, and Permian are in close proximity.

Third is lexical comparison. About a dozen resemblant proto-roots found in PU and PIE, regarded variously as loans (Koivulehto 2001) or cognates (Helimski 2001b), are taken as evidence for proximity of PIE and PU, on the grounds that, whether they were sisters with a common homeland or neighbors exchanging loans, they were necessarily in the same region.¹³

12. Supplementary materials can be found online at: <https://doi.org/10.5281/zenodo.6345559>

13. Both the possibility of a much earlier PIE-PU genealogical relatedness and the putative loanwords are invoked by Anthony and Ringe (2015:206–207, citing Koivulehto 2001), and somewhat similarly Anthony (2007:96–97), Mallory (1989:149), as evidence for a PIE origin

Fourth is one key item in linguistic paleontology. The word for ‘honeybee’ (*Apis mellifera*), an Indo-Iranian (henceforth I-I) borrowing, points to an origin west of the Urals; in Siberia the natural range of the honeybee does not extend to the forest zone.

The linguistic arguments against the middle Volga homeland are stronger than the arguments for it. Against the first argument, the center of gravity for the traditional tree is in western Siberia where Samoyedic approaches or contacts the nearest Finno-Ugric branch (represented by Khanty), well to the east of the Urals. A relevant consideration even on the rake model is that diversification and geographical distancing arguably began earlier in western Siberia (§3.4).

Second, the criterion equating greatest diversity with greatest number of branches in the area has to our knowledge never been shown to reliably identify homelands of language families; perhaps it comes from a misunderstanding of the technical sense of greatest phylogenetic diversity, which considers only primary branches. On grounds of sheer number of branches the western Middle Volga area where ancestral Finnic and Saamic were in proximity to Mordvin would appear to be an equally good candidate, as would northwestern Siberia where Khanty, Komi, and Forest Nenets are in proximity.¹⁴ In fact, those regions and the middle Volga are what are variously known as residual zones, accretion zones, or language sinks (Nichols 1992, 1997; Nichols & Rhodes 2018): areas where languages move in more readily than out and diversity increases over time.¹⁵ Language sinks are unlikely proto-homelands.

Third, whether based on cognacy or loans the argument from lexical resemblances is flawed. Despite careful attention by both Koivulehto (2001) and Helimski (2001b) to systematicity in correspondences and PIE forms, the set of resemblant roots is not large enough to exceed chance, and the relatively wide

not far from the Volga bend, near where they judged the Uralicist literature to place the PU homeland.

14. An extreme example of this understanding of diversity is the Austronesian family, of which all but one of the primary branches are found on Taiwan while the dozens or more branches of the other primary branch, Malayo-Polynesian, stretch across island Southeast Asia, Melanesia, Micronesia, and Polynesia (Blust 2009; the number of primary branches recognized varies).

15. The Minusinsk basin and middle-upper Yenisei area in south central Siberia, in the likely Samoyedic homeland region, is a high-diversity language sink where different Samoyedic branches have neighbored with Yeniseian, Turkic, and Tungusic languages over time; and it is also the starting point for major northward spreads down the Taz and Yenisei, today illustrated by northern Samoyedic languages (Janhunen 2012a; Khanina et al. 2018). The western Middle Volga area hosts diversity and was the starting point for major northward spreads by Saamic and then northern Finnic. It must be that northward spreads are conditioned by ecological and economic factors that are orthogonal to those that create language sinks.

ranges of both form and meaning reduce the probative value of the individual pieces (Nichols 2010; Nichols & Rhodes 2018; Simon 2020).¹⁶

Fourth, the term for ‘honeybee’, the key item of linguistic paleontology, is a good diagnostic, but not of the homeland. Like all I-I loans, it was borrowed not into PU but into post-dispersal early Uralic; it is irrelevant for the PU homeland. It is borrowed into those branches whose branch homelands are in the honeybee range (Supplement S2; Appendix 2.).

There are additional arguments against the Volga homeland. Probably most telling is the absence of loanwords in PU from the late PIE language of the Yamnaya spread or the pre-Balto-Slavic or Para-Baltic¹⁷ languages of the Fatyanovo culture or its descendants (§1 above, §2.4, and Supplement S3). More generally, the reconstructed wordstock of PU points to a Neolithic technology and economy without food production (Janhunen 2009, 2020), which is unlikely for regions near the middle Volga or Volga bend, as food-producing PIE and early IE-speaking groups had been a major cultural force in the nearby steppe and forest-steppe areas for nearly a thousand years before the PU dispersal.

There is additional linguistic paleontological evidence favoring a western Siberian homeland. A PU term is reconstructed for the Siberian stone pine or cedar pine (*Pinus cembra sibirica*), a food plant found only in Siberia (Supplement S2, on this and other ecological terms; Appendix 1). Importantly, PU had only a single undifferentiated term for ‘metal’, inconsistent with propinquity to the southern Urals, while the early post-dispersal stage we call Common Uralic (§2.3 below) used loans and derivation to create a more elaborate terminology consistent with involvement in the bronze trade (see Supplement S2).

Finally, a Volga homeland makes the long Samoyedic movement to its branch homeland in the area around the middle Ob to middle Yenisei area implausible, and that movement bucks the generally east-to-west trajectory of north Eurasian language spreads (Janhunen 2014, 2012b).

The conclusion is that there is no linguistic evidence in favor of a middle Volga homeland, or any homeland west of the Urals, and strong evidence in favor of a western Siberian homeland.

16. Of what we take to be the two statistically soundest recent quantitative tests, Kessler and Lehtonen (2006), using a 100-item Swadesh-like wordlist, found no evidence for Indo-Uralic, while Kassian et al. (2015), using a shorter wordlist, found evidence (but see the questions on their data and method in Kallio 2015; Kessler 2015; and Ringe 2015).

17. Here and below, “Pre-” labels any unspecified stage prior to “Proto-” (as that is reconstructed from attested daughter languages), and “Para-” labels an immediate sister to an existing branch. Thus, Pre-Baltic is any stage prior to (depending on whether Baltic is regarded as a clade) Proto-Baltic or Proto-Balto-Slavic, and Para-Baltic is a sister to all of Baltic or Balto-Slavic (see also Supplement S3).

2.2 Linguistic and extralinguistic chronology

Early Uralic prehistory needs to be placed in the context of three temporally and geographically overlapping important events. The earliest of these is what is known in the earth sciences as the 4.2 ka event, a global climate development from 4,200 to 3,900 BP that caused drought in the low and mid latitudes of continental interiors, wetter conditions in the northwest of Eurasia, and global cooling. Kingdoms fell and urban centers collapsed in the Near East and Southwest Asia, and the Caspian and Kazakh steppes suffered extreme drought (while the westernmost steppe and eastern Europe may have had more rainfall than usual). Temperatures were cooler overall (Perçoiu et al. 2019; Helama & Oinonen 2019; Dalfes et al. 1997.)

In the same time frame came the Seima-Turbino Transcultural Phenomenon (henceforth ST), an archaeological complex marked by distinctive bronze artifacts, especially symbols of power such as spearheads and axe heads, found across many archaeological cultures and along major waterways from the Altai to Scandinavia. In particular, tin from Altai mines made possible large-scale bronze production, with the forging and casting done there as well as in the southeastern Ural area and Southwest Asia. The most recent radiocarbon dates place ST between c. 4,200 and 3,900 BP, and somewhat later west of the Urals along the Volga (Marchenko et al. 2017; Krause et al. 2019; Chernykh 2008).

ST is evidently the archaeological signature of a waterborne trade network (Barfield 2009; Nichols & Rhodes 2018) that brought metal from Ural and Sayan mines and metal artifacts from Ural forgeries westward to Europe. Its heyday coincides rather closely with the 4.2 ka event. The trade network itself probably existed long before the Bronze Age and demonstrably continued into the Middle Ages as the Bulgar and Viking trade routes.

Barely later came an Indo-Iranian (I-I) contact episode. Approximately 4,000 years ago, the Finno-Ugric branch ancestors absorbed a good deal of vocabulary from the I-I branch of Indo-European (Holopainen 2019). I-I figured prominently in the development and spread of bronze technology from the vicinity of the southern Urals in this time frame. Proto-I-I is well reconstructed and dated on linguistic grounds, including very early stages of daughter languages that are well attested in writing, and the time and general location of its spread are well established archaeologically. The Poltavka archaeological culture and its Potapovka and Sintashta-Petrovka descendants (see Figure 1) were culturally, and almost certainly linguistically, Indo-Iranian (Kuzmina 2007). The Indo-Iranian episode is one of the clearest linguistic signatures known to archaeology, short of inscriptional evidence, and its time frame needs to be stipulated as an absolute date (albeit an approximate one) in any chronology of Uralic.

The inventory of I-I words differs from branch to branch in Uralic, and etyma come from time frames ranging from Pre-Proto-Indo-Iranian to Proto-Iranian to early Iranian (for these loanwords see Appendix 2). This shows that I-I interacted not with a single Proto-Uralic but with an incipiently differentiated early Uralic, over some extent of time and some extent of space. In view of this distribution, the I-I contact episode cannot be regarded as a single clade-defining event and therefore as establishing the reality of a unitary Finno-Ugric branch. What it does establish is the time of the initial Uralic divergence: it occurred before 4,000 BP but not long before. Evidence is the fact that the I-I loans entered at the branch protolanguage level or not long thereafter, that they entered the early Uralic branches separately, and that the internal evolution of the daughter branches began after 4,000 BP as shown by the application of branch-specific sound laws to the I-I material.

The Samoyedic branch lacks the I-I stratum almost entirely. This, together with its low number of cognates, may point to an early and fairly clean separation of Proto-Samoyedic from the rest of the family, as was widely assumed in 20th century Uralic studies. On the other hand, the retention in Samoyedic of much PU inflectional morphology and the regular phonological evolution of its surviving native vocabulary suggest that that separation did not precede the I-I episode by long. The spread of Finno-Ugric could have been simultaneous with the separation of Samoyedic or later; their different histories may be due to different directionalities and geographies as much as to different chronologies. The geography of a reconstructed PU homeland needs to provide for a clean break and exit of Proto-Samoyedic and a rapid spread of Finno-Ugric to bring it into I-I contact in a spatially differentiated but structurally homogeneous form.

The 4.2 ka event and ST are nearly simultaneous and the I-I episode occurred within that time range, and we suggest they are causally connected. During the drought, herders saw their traditional subsistence falter and fail, and the populations of herds and probably herders were reduced by famine. There is evidence of overgrazing on the steppe in late Yamnaya times (Anthony 2007: 330), half a millennium earlier, so it must be assumed that by 4,200 BP the entire grazable Pontic-Caspian steppe was inhabited to carrying capacity, with no leeway except what could be gained by warfare. To the east, the alternative was to seek water sources at the steppe periphery. Herders from the almost certainly Indo-Iranian-speaking Poltavka culture crowded into the river valleys around the Urals, where they established fixed year-round settlements (Anthony 2007: 371–411, especially 389–391, describing permanent settlement as a strategy for claiming access among traditionally mobile societies facing diminishing resources). Competition and warfare were intense, and militarization increased.

Around the southern Urals, IE-speaking societies found economic security in bronze production and bronze trade. The Poltavka- and Abashevo-derived Sintashta culture of the southeastern Urals developed bronze manufacture into a major cottage industry, and c. 3,900 BP invented the chariot and developed chariot warfare, solidifying Iranian-speaking domination of the entire Caspian steppe (Lindner 2020). ST is the visible signature of this strategy: hoards, burials, and other finds along the routes of what had long been a waterborne trade network, which now quickly adapted and began to carry bronze from Ural and Altai mines. ST was a symptom of a broader process of interregionalization that brought expanded trade networks and an expanded inventory of trade goods to a widening range of markets (Frachetti 2008: 47–67). Bronze-producing societies in the southern Ural region established trade connections as far afield as Southwest Asian cities (Anthony 2007: 389–393, 418–427).

A background condition is the presence of endemic bubonic plague (*Yersinia pestis*) and salmonella (*Salmonella enterica*) on and near the steppe (Rascovan et al. 2019; Andrades et al. 2017; Rasmussen et al. 2015; Key et al. 2020). Both bacteria had undergone major genealogical diversifications beginning in the Neolithic period, when the denser village and urban populations and proximity to livestock favored their transmission. Bubonic plague became virulent, that is, transmissible by fleas rather than requiring direct contact, somewhat later but still probably by 5,000 years ago (Spyrou et al. 2018). Both diseases would have presented particular risks in the denser conditions and year-round presence of livestock that accompanied the drought. Mobile and less dense hunter-fisher populations would have been much less affected.

In the model proposed here, early Uralic speech spread rapidly along the waterborne trade network north of the steppe, expanding as part of the same interregionalization as ST. Uralic speakers were the prospectors, miners, boatmen, trade managers, procurers, and first settlers of trading posts at major river confluences; the Indo-Iranian-speaking Sintashta culture and its successors financed prospecting, trade, and markets. Before the pastoral steppe populations recovered from the drought, Uralic-speaking trading post settlements had already become well entrenched and demographically strong along the trade routes, allowing Uralic-speaking populations to dominate the forest-steppe and forest zones thereafter.

2.3 Early Uralic stages

Some previous work has suggested that PU was the language of ST, but it is essential to distinguish PU, dated to about 4,500 BP, from the language of 4,200–3,900 BP when IE retreated and Uralic speech spread. We define PU as the linguistic sys-

tem that can be reconstructed by applying the comparative method to the Uralic daughter languages, and the unified speech community that spoke it. The divergence of PU into incipient daughter branches, and application of changes only branch-internally, brought about the end of PU and the beginning of the separate evolution of the daughter branches. We use the term Common Uralic (CU) to refer to the speech community and the language from the time of the initial divergence to the point, some centuries later, when the daughter dialects ceased to be mutually intelligible. PU began diverging before ST arose, and it had only a minimal terminology for metals, probably only a single term for ‘copper’, the metal that was mined and worked near the Urals prior to the ST phenomenon. CU was involved in ST and appears to have had a richer terminology for the important ST trade items and materials (Supplement S2).

2.4 Timeline

A timeline for Uralic-relevant ethnolinguistic events along the upper and middle Volga, c. 6,000–3,000 BP, is as follows (Anthony 2007; Lang 2018; Nordqvist & Heyd 2020; see Supplement S3).

- c. 6,000–
3,500(?) BP:*

The Volosovo culture of settled hunting-fishing societies is found along the entire Oka-Volga-lower Kama and well to the north. Several of its cultural features persist in ethnographically documented societies of the north European forest zone, but they are not specific enough to suggest any particular language identity. A territory that large must have contained a number of different societies and languages and probably from more than one language family. Much of the Volosovo range has been Finno-Ugric-speaking in historical times, but this is due to parallel northward spreads (§3.1) from different branch centers rather than continuing an ancient ethnolinguistic unity.
- 5,100 BP:*

The Yamnaya culture, almost certainly late PIE-speaking,¹⁸ spread rapidly across the Caspian and Pontic steppes. It adapted wheeled transport and expanded by a mix of language shift, migration, and population growth. Contemporaneous with the Yamnaya spread, the Afanasievo culture, Yamnaya-like and likely speaking an early IE language, appeared in the Minusinsk basin (upper Yenisei) and across the Altai foothills.

18. More precisely, spreading late PIE if not necessarily speaking it as home language. Before the Yamnaya spread, more than one language must have been spoken by the mostly herding cul-

Neither Yamnaya nor Afanasievo has left any detectable loan vocabulary in PU.

Middle Bronze Age, 4,800–3,900 BP: The Fatyanovo, Balanovo, and Abashevo cultures, ultimately Yamnaya descendents via the central European Corded Ware culture, spread east from the Western Dvina/Daugava and middle Dnieper along the forest-steppe belt. Fatyanovo (4800–4200) extended mostly along the upper and middle Volga (and its tributaries) and in the west ultimately far north to Lake Ladoga; Balanovo was a Fatyanovo extension east along the middle Volga and to the Kama confluence; Abashevo, a Fatyanovo descendent (4,200–3,900 BP), extended along the Volga Bend (and south of the Volga to the middle Don catchment) and the lower Kama and the Belaya to the southern Urals. These were farming and herding societies (Abashevo mostly herding) which made and used metal implements. Fatyanovo was very likely Pre-Baltic-speaking (and almost certainly so in its western range). There is no trace of contact with any of these languages in the PU lexicon.

Midway through this phase, PU began to diverge.

4200–3900 BP: The 4.2 ka event and ST; late in this period, the I-I episode; see §2.2. The I-I episode was the first evidence of early Uralic contact with any IE language. By this time the CU varieties had begun to diverge and separate, and they borrowed I-I words separately.

3,900 BP and later: Rebound. With climate amelioration, agriculture recovers across the northern steppe periphery. Indo-Iranian languages come to be spoken across the entire Pontic-Caspian steppe and into the Kazakh steppe. The center of bronze production and trade shifts west and the ST complex comes to an end, perhaps because the new opportunities presented by recolonization of lands abandoned during the drought made warfare and displays of power less necessary. Bronze artifacts exported to eastern Europe and Scandinavia now originate chiefly along the western Urals and near the Volga bend.

tures of the steppe and steppe periphery; the range is too large for a single pedestrian language. The rapid Yamnaya spread indicates language shift, so pre-Yamnaya cultures must have adopted Yamnaya customs and economy and begun to enter the Yamnaya social networks before becoming primarily IE-speaking. However it happened, Yamnaya spread only one language – late PIE – across the steppe and into southeastern and central Europe. The only surviving trace of other languages is terms for crop plants and a few others acquired from the farming cultures of eastern Europe (Iversen & Kroonen 2017).

The ethnolinguistic composition of the middle Volga and southwestern Ural areas has changed: formerly Indo-European and probably Pre-Balto-Slavic- or Para-Baltic-speaking prior to the 4.2 ka event, it becomes durably Finno-Ugric-speaking after that (Supplement S3) and remains so until marginalized in the Russian colonial expansion.

3. The CU spreads

3.1 Northward spreads

The large northern extent of the attested Uralic range is secondary. It was reached after the initial Uralic dispersal, as dialects and daughter languages of Uralic languages spread northward from the branch ancestor¹⁹ homelands located in the southern part of the early Uralic range (Figure 1; see also Saarikivi in press) and eventually reached the Arctic Ocean coast. Such northward spreads are a recurrent pattern in northern Eurasia, though not in North America (Nichols & Rhodes 2018).

These northward spreads are useful for comparison to the primary CU spread, as they give us the geographically closest and most comparable known cases of hunter-gatherer spreads. The process may have begun early but proceeded slowly enough that in several cases the far north was Uralicized only in recent centuries (Aikio 2012; Helimski 2001a). The present-day northernmost languages, the Saamic and Samoyedic branches, exhibit exotic substratal vocabulary and/or grammatical and phonological effects. In most places, local northern toponymy and vocabulary for tundra flora and fauna include words of non-Uralic origin, showing that today's northernmost languages were the frontier languages in the spread (Aikio 2012; Helimski 2001a; Saarikivi 2006, in press). Additionally, the fact that this vocabulary is borrowed shows that not only PU and CU speakers, but also branch ancestor speakers, were unfamiliar with the tundra ecology and needed to borrow terms for it. The rate of northward spread appears to have been accelerated by the emergence of reindeer herding in the last two millennia

19. We use the term *branch ancestor* to refer to the protolanguages of the nine branches (Table 1) and their presumed speech communities (see Saarikivi in press). Branch ancestor homelands, or branch homelands, are the places from which the branches have expanded and spread northward. Sometimes the homelands are supported by archaeological and/or toponymic evidence (strongest for Saamic and Finnic: Aikio 2012; Lang 2018; Saarikivi & Lavento 2012; Saarikivi in press). The branch homelands are likely to have been settled in the initial spread, with the exception of Saamic and Finnic, whose points of dispersal are known to have been secondary.

and (possibly, though debated) the development of large-scale fur trade, in which Uralic speakers were trappers and procurers for Scandinavian and Russian traders (Aikio 2012; Helimski 2001a). Since resources in tundra landscapes are sparse and patchy, survival there requires larger ranges per capita, larger-scale mobility, and spatially more extensive social and economic networks. Correspondingly, the ranges of speakers of the northernmost languages have spread out widely in this environment.

The branch homelands were in forest lands, and the northward spreads eventually moved into tundra ecologies. These northernmost phases are quite likely to have involved primarily language shift, as sparsely distributed northern hunter-gatherers shifted to the languages of the denser and more economically advanced populations to the south. Unpressured shift from the language of a sparse foraging society to the language of a denser food-producing society, especially if the latter is also technologically advanced, is a common though not universal development where foraging and food-producing economies are in contact (Güldemann et al. 2020: 30–32). When tundra populations adopted reindeer herding and/or became procurers in the European fur trade, individual languages became influential and spread widely (e.g., North Saami, and especially Komi and Tundra Nenets).

Judging from the reconstructed and partly attested recent history of Saamic and Samoyedic groups (Aikio 2012; Khanina et al. 2018), northward spread proceeded unevenly, responding to fluctuations in climate, economic and demographic pressures, technological advances, and other factors. The northward spread of the Samoyedic branch proceeded in spurts, most probably driven by advances in reindeer herding (Khanina in press). In the more recent northward spreads of the Finnic branch and Komi of the Permic branch, important factors were the adoption of agriculture and stockbreeding and their concomitant population increases. The scale of the agriculture, which used slash-and-burn methods or flood-retreat sowing, was small, and the livestock often amounted to two or three cows per household for dairy products. Fishing and hunting were important in the diet. This economic scale persisted in rural households in parts of central and northern Finland well into the 20th century.

Northward spread likely involved a mix of small-group migration and language shift, and was chaotic locally and in the short term but northward overall. The process gave rise to a modern stratigraphy in which the more southerly language or languages lack recent substratal effects and have more compact ranges while the northern languages have recent substratal effects, especially terminology for arctic phenomena, and larger ranges. In their overall gradual pace, substratal effects, and expansion primarily from the frontier, these northward spreads are different from what can be reconstructed for the initial Uralic spread.

3.2 The CU east-west dispersal

In contrast to the northward spreads, the initial Uralic spread was almost entirely east-west in direction, with daughter branches taking root along most of the east-west extent of the Volga and probably along the middle and upper Tobol, Irtysh, and Ob and the upper Yenisei (Figure 1). This spread appears to have been rapid, largely without substratum, and with minimal evidence of frontier expansion and isolation by distance. It was westward overall; only the Samoyedic branch probably did not take part in the westward spread and may have moved eastward instead. Evidence in favor of these claims includes the following points.

3.2.1 *Minimal isolation by distance (IBD) effects in vocabulary*

Isolation by distance (IBD; also called autocorrelation) refers to the general phenomenon of finding decreasing numbers of shared traits as geographic distance between related populations increases (Holman et al. 2007; Haynie 2014: 344–345, 349–350). Strong IBD effects suggest a slow and steady movement away from the center of expansion owing to ongoing exchange between neighboring populations. Minimal IBD effects, on the contrary, suggest the spread from the center of expansion was rapid. In historical linguistics, IBD has been applied to modeling language history in several continents. For instance, it has been used to model the Bantu expansion (de Filippo et al. 2011), the linguistic diversity of Japonic languages (Huisman et al. 2019), and the language history in Melanesia (Hunley et al. 2008; see Supplement S4 for more information).

To the extent that ancestors of the Uralic branch spread slowly we should thus find monotonically increasing linguistic distance from language to language as their geographic distance from the center of expansion increases. However, we do not find this kind of regular IBD effect. Figure 2 shows the numbers of Proto-Uralic reconstructed etyma retained in each daughter language.^{20,21} It also shows the numbers of early Indo-Iranian (I-I) loans that entered the early Uralic branch ancestors c. 4,000 BP (§2.2 above; source for the items: Holopainen 2019). The I-I stratum was borrowed early enough that the items reconstruct to the proto-language of each Uralic branch, so that stratum should have undergone IBD attrition at rates similar to those of native Uralic vocabulary. IBD effects should show up as a more or less monotonic dropoff from an origin point; the dropoff should be unidirectional if the origin point was near the present edge of the range, or bidirectional if it was in the center of the range.

20. Used here are PU cognates from the major languages and varieties for which lexical documentation is adequate for meaningful comparison, and I-I loans for those same languages.

21. Holopainen (2019) and Appendixes 1–3 here.

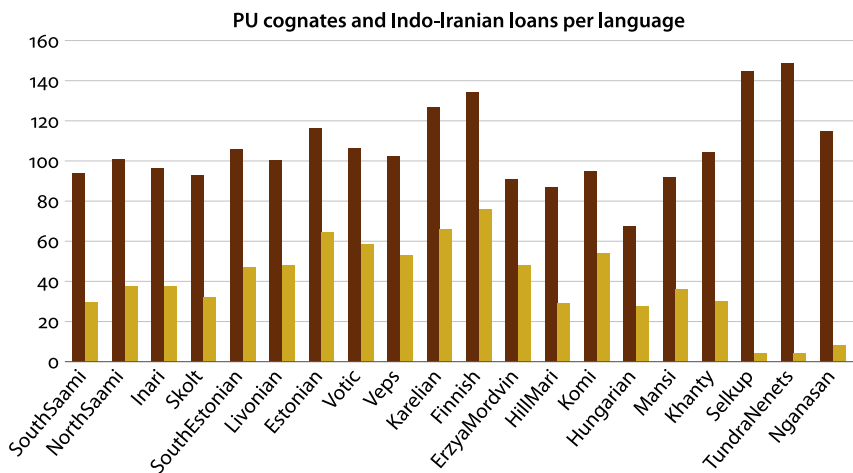


Figure 2. Numbers of Uralic (brown) and Indo-Iranian (yellow) etyma retained per language

Languages in Figure 2 are shown in geographical ordering, west to east (by branch) and then south to north (within branches). There is no evidence of IBD effects; this is also true if, following traditional ordering, Finnic is listed before Saamic and Hungarian before Khanty. For statistical tests of monotonicity on languages and branches, see Supplement S4. If anything, retention rates are highest at the far west of the Uralic range, a highly unlikely origin point for the initial spread; these figures undoubtedly reflect peripheral archaism in the case of PU vocabulary, early isolation of branches in the east (§3.4), and a concentration of I-I contacts west of the Urals (so that the Ugric languages have few I-I words and the Samoyedic ones next to none).²² They may also reflect the larger numbers of daughter languages in the Saamic and Finnic branches, as that raises the probability that a PU or I-I item will be attested in the branch.²³ For other factors see Supplement S4.

PU words are identified, following the traditional phylogeny with its initial split of Samoyedic from the rest (§1 and Supplement S1), as those having a reflex in Samoyedic and one other branch, as well as regular sound correspondences.

22. A regular decrease from west to east, if one exists, should be visible within Finno-Ugric if all of the reconstructable Finno-Ugric vocabulary could be surveyed. A definitive compilation of currently accepted Finno-Ugric etymologies does not yet exist, but a good approximation should be extractable from the ongoing Proto-Uralic project of Ante Aikio when that is complete (see Aikio 2013, 2014a,b, 2015).

23. Samoyedic also has a fairly large number of daughter languages, but several of them became extinct before they could be documented lexically.

There are about 200 such words (Appendix 1), each with a reflex in Samoyedic by definition and a more scattered distribution over the other branches. Since these words are always present in one or more Samoyedic languages, the Samoyedic languages have artifactually high frequencies.

3.2.2 *Homelands*

Branch homelands and attested or reconstructed core areas are in the southern parts of the current branch ranges, along the Volga or nearby (Figure 1). These homelands date to the Middle Bronze Age and are identified on criteria such as lexical comparison, toponymy, archaeological evidence, and early historical evidence. They attempt to represent the general area from which the branch later spread to its larger historical range (Saarikivi in press). Branch homelands are clearest for the Volga languages (Mordvin, Mari, and Permic), though they shifted and retracted somewhat in the Middle Ages in response to Bulgar and then Russian rule. For the Ob-Ugric languages they are uncertain, but probably in the southern parts of the historical ranges, and for Hungarian somewhat south of Ob-Ugric in Western Siberia. For Finnic and Saamic the centers of later spread are secondary locations, reached by westward spreads from nearer to the middle Volga (Supplement S5).

3.2.3 *Innovations*

Each branch has undergone a number of separate innovations since splitting from the rest. In some cases, there are as many as 20 or 30 such branch-specific innovations prior to the internal breakup of a branch, or even more, and at least as many since then in the individual daughter languages (Saarikivi in press). This points to longer periods of independent evolution from the emergence of branch ancestors (when Pre-Permic, etc. emerged as distinct languages) to their internal breakups (e.g., of Proto-Permic into Komi and Udmurt), and from that to the present, as compared to the short CU period from dispersal to emergence of branch ancestors established in §3.2.1. In addition, the general uncertainty about the higher branching structure of Uralic is due to scarcity or lack of diagnostic shared changes, which in itself indicates a rapid separation of the initial branches.

3.2.4 *Peripheral archaisms*

Archaisms are substantial and well preserved at the peripheries of the Uralic range (Janhunen 2020), some in both the far east (Samoyedic) and the far west (Saamic, Finnic) and some only in the far west. These include the well-preserved inflectional paradigms of Samoyedic, Finnic, and Saamic; the preservation of the ancient two-syllable root structure also in Samoyedic, Finnic, and Saamic; the survival of the dual number only in Samoyedic, Ugric, and Saamic; preservation

of the sequence of numerals 1–9 with some modifications in Saamic and Finnic;²⁴ and the high lexical retention rates in Saamic and Finnic. Peripheral retention of archaisms, a common occurrence in dialect geography, is consistent with rapid full separation, as it indicates that inherited material was prone to be replaced by innovations diffusing from post-spread centers of innovation and this was more likely to occur closer to such a center. The central branches of Uralic, especially Permic and Mari, have less PU vocabulary and more changes in the design of inflectional paradigms than Finnic or Saamic. This is because the entire Finno-Ugric range ceased to function as a single dialect-geographical area early on, and these central languages formed a local interactive area whose innovations did not spread to the far peripheries. Besides, unlike the peripheral dialects, the ones in the center could continue to be in contact with each other, which made them more prone to lexical differentiation based on an ideological urge for a separate linguistic identity (see Ellison & Miceli 2017). This is not relevant for phonological or grammatical retentions, but could be so for lexical ones, including the numerals. Some criss-crossing isoglosses in Volga languages based on presumably substrate lexical influence point in this direction.

3.2.5 *Isolation by distance: Loans*

The distribution of Indo-Iranian loans displays no IBD effects, as Figure 2 shows. The number of PU etyma retained per language correlates highly significantly (Kendall's Tau = 0.632; $p = 0.0005$) with the number of I-I etyma if the Samoyedic languages are omitted from the count. If Samoyedic is included, the disparity between its artifactually high PU counts and the low I-I counts singlehandedly destroys the correlation (Kendall's Tau = 0.188; $p = 0.254$). For branches (Supplement S4) there is no significant correlation either with or without Samoyedic, because the number of branches is low. All of this means that the different numbers of I-I etyma in the different languages and branches should not be ascribed to different local intensities of contact between early Uralic and early I-I; the most parsimonious solution is that I-I etyma counts reflect overall vocabulary evolution in the branches and languages, including their different rates of change and loss. This pertains to just the ratios of I-I loans and PU words. The different inventories of I-I etyma in each branch show that borrowing proceeded individually in the branches. Samoyedic is an exception, with very low numbers of I-I words, out of proportion to its artifactually high numbers of PU words, indicating that it was not in the steppe sphere of interaction. (See §3.4 below; also Figure S3b–c.)

24. The numerals 'eight' and 'nine' actually reflect a subtractive pattern "less by two" resp. "less by one", presumably an areal phenomenon with parallel etymological structure in Mordvinic, Mari and Permic (Aikio in press; E. Itkonen 1973: 336–339; Parpola 1999: 198–199).

However, a secure determination awaits reconstruction and analysis of a full PU vocabulary without veto power by Samoyedic.

Otherwise, trends are clearest and most meaningful when it is branches that are compared. When individual languages are compared, the high number of Finnic languages in the sample give that branch disproportionate impact; and the ordering in which languages are listed within branches can appreciably change the slope and strength of correlations. For Finnic, ordering south-to-north vs. north-to-south puts Finnish closer to the left edge of the graph vs. near the middle. North-south position within branches is caused primarily by northward spreads (§2.1), which are irrelevant to the initial spread, and there is no consistency between branches as to whether north or south retains the highest numbers of etyma.

3.2.6 *Inflectional paradigms*

Reconstructable inflectional morphology points to a fairly unified picture across the branch ancestors (Aikio in press; Saarikivi in press; Janhunen 1982, 2000, 2009; Kulonen 2001; Majtinskaja 1993). In contrast, developments in individual branches have produced some major reanalyses and reshaping of paradigms. As with sound changes (above), the early similarity and subsequent branch-internal changes point to a short period from initial PU dispersal to branch ancestor formation and longer periods of independent evolution of branches. Thus, a minimally divergent set of future branch ancestors spread out rapidly, at which point branch-internal evolution began (§3.2.3 above and Supplement S6).

3.2.7 *Substrata*

Only the Samoyedic branch gives evidence of substantial substratal effects, arguably dating to the time of the initial Uralic dispersal, and the general lack of initial substrata in the other branches is an unexpected finding given the linguistic geography: the more western languages have spread farthest from the homeland and have had more opportunities for contacts and more need to borrow terms for non-Siberian phenomena. The expanse from east of the Urals to the upper Volga harbors a number of different languages today and should have contained no fewer in early times, so one might expect different effects in every language. In contrast, there is a modest body of words in the Volga and Western Uralic languages for broadleaf trees and items of agriculture and stockbreeding, some containing novel phoneme combinations that were impermissible in PU and brand the words as foreign (Häkkinen 2009: 37–40; Aikio 2015b: 44–47; Zhivlov 2015). All are found in more than one Uralic branch. The set suggests borrowings shared among the sister Uralic languages (and part of a broader Middle Volga areality in grammar and lexicon) and not the pervasive effects we find

from tundra languages in Saamic (see §3.1) or the large and varied non-Uralic vocabulary of Samoyedic (Saarikivi in press). Thus the only evidence for substratum at the shortly post-PU level comes from Samoyedic, where a good deal of PU vocabulary was lost and replaced by non-Uralic loans with un-Uralic canon shapes and phonotactics²⁵ probably beginning shortly post-PU and continuing to the time of Proto-Samoyedic divergence (Aikio in press).²⁶ A counterargument is that the Uralic/non-Uralic contrast may be anachronistic, comparing the young non-Uralic vocabulary (reconstructable only to Proto-Samoyedic, c. 2000–2500 years ago) with PU forms of native vocabulary.

Elsewhere in the family, later substrata are evident in individual languages or subbranches, often connected with northward spreads.²⁷ The Ugric languages have some important exotic vocabulary, such as terms for horses, horse breeding, and nomadic culture, but whether these point to a substratum or are ordinary loans is less clear. The question of substrates in Ugric has had relatively little research.

Just what happened in Samoyedic is debated. On a traditional family-tree model, where the initial split is seen as a bifurcation into Samoyedic vs. Finno-Ugric, the issue can be seen as vocabulary that is present in Finno-Ugric but absent in Samoyedic and replaced there from another language or languages (the traditional view). Alternatively, it can be asked whether it is not Finno-Ugric that has lost and replaced vocabulary, and the putative exotic vocabulary of Samoyedic is in fact native Uralic vocabulary lost in Finno-Ugric. This issue largely evaporates on the rake model (Supplement S1), where Samoyedic is one of nine initial branches and there is no single branch whose lexical differences are crucial to PU

25. That the Samoyedic vocabulary without Uralic cognates is often un-Uralic in form seems to be the received view among Samoyedists, but we do not know who first pointed it out.

26. The Samoyedic replacement is sometimes referred to as relexification, but that term has a more specific meaning in creolization studies, where it refers to taking the phonological forms of words from a lexifier (superstrate) language but the grammar, including semantic structure and lexical classes of words, from the substrate language (e.g., Lefebvre 1998). There is no evidence of anything like creolization in the history of Samoyedic, and no evidence that Samoyedic word grammar or grammar in general is un-Uralic; as noted, Samoyedic inflectional morphology reflects PU morphology well, unlike the usual development in creolization.

27. The Ob-Ugric languages have non-native vocabulary in common, much of it with irregular correspondences indicating separate borrowing by early Khanty and Mansi (Sipos 2002; Saarikivi in press). Saamic and Finnic place names have a number of non-native elements (nearly half of the high-frequency elements in Finnish place names) and the two branches have about 220 unique shared cognate roots of unknown etymology, at least some of which are likely to be substratal; but this vocabulary appears to have been acquired not in the initial dispersal but in the Saamic secondary staging area in southeastern Finland and the Northern Dvina basin of northern Russia. (See Aikio 2012, 2004; Saarikivi 2004a b, 2006.)

reconstruction. Parallel massive vocabulary replacement across the other eight branches is so unlikely that Samoyedic must be seen as different and as having lost vocabulary.²⁸ A view from a different perspective, applying in either case, is that no drastic loss and replacement has occurred in Samoyedic; rather, Samoyedic separated early enough that it has had more time for vocabulary attrition. The good preservation of inflectional morphology in Samoyedic argues against a very early separation. So far the issue remains open.²⁹

3.3 Typology and the Common Uralic spread

The morphosyntactic typology of Uralic is distinctive in western Eurasia. A number of typological properties are eastern-looking overall, fitting comfortably into northeast Asia, Siberia, or the North Pacific Rim. These include (Supplement S7): low finiteness, high inflectional person, high part-of-speech flexibility in the lexicon, possible evidence of non-accusative alignment, and base intransitivity.

28. Atkinson et al. (2008) suggest that languages resulting from more branching retain fewer lexical items from the protolanguage (see also Ellison and Miceli 2017 for a possible explanation based on a sociocultural preference for dissimilar forms in individuals who are bilingual in related languages). This could indeed explain why Samoyedic with more daughters kept fewer Proto-Uralic lexemes than most other Uralic branches did, but crucially it could not explain why Finno-Ugric languages share so many cognates among themselves.

29. Cases where a daughter language preserves morphology well but loses vocabulary massively are not frequent, but they are reported. Comrie (1988, 1989, 2000) shows that Haruai (Piawi family, New Guinea) is a sister of Hagahai (Piawi) and not of Kobon (Kalam family), despite massive vocabulary sharing with Kobon, on the evidence of shared morphological paradigms with Hagahai. Green (2003) shows that Murrinh-Patha, long considered an isolate, is a close sister of Ngangityemerri, making up a Southern Daly family (northern Australia) on the strength of parallel auxiliary verb morphology and despite near-absence of lexical cognates. Comrie suggests that the Piawi discrepancy could have arisen as a result of word taboos compensated by loans from a neighboring language, whose cumulative effect in a small speech community might have been substantial. Green leaves the question open for Southern Daly. Miceli (2015) describes the Pama-Nyungan family as languages which share phonology and grammar, but very little lexicon (and many of the potential cognate forms are nearly identical, so they could as well be loanwords and not inherited from the common protolanguage). She suggests that conscious efforts of bilingual speakers on keeping languages apart (in particular in the case of related languages) may have contributed to extremely high levels of lexical divergence in Pama-Nyungan, attested simultaneously with morphosyntactic convergence (see also Ellison and Miceli 2017 for a more general approach to the same phenomenon). (We note that whether the Pama-Nyungan languages actually have few inherited lexical roots is disputed. See e.g. Alpher & Nash 1999; Bower et al. 2011.) In the case of lexical discrepancy between Samoyedic and the rest of Uralic, the drivers of lexical divergence would have been bilinguals in (Pre-)Proto-Samoyedic and some other Common Uralic variety(/ies).

A number of traits are cross-linguistically infrequent. Despite the evolutionary disadvantage, or fragility in contact situations, suggested by the low frequency, they are persistent in the family, either as morphemes or as types. They include (Supplement S7): a dual number category; negative auxiliary verb; differential object marking with unusual stability in the conditions on object marking; a contrast of subjective vs. objective conjugation; and personless pronoun stems. Taken together, these various traits are consistent with an origin of PU in the eastern part of its range and a rapid initial spread with minimal contact influence, so that they all remained firmly in place as the dispersing CU branch ancestors settled in among new neighbors.

3.4 Early diversification in the east

Though the higher-level branching structure in the Uralic tree is debated, it is clear that if there was any early branching and/or any early geographical separation it began in the east of the PU range. Evidence for an initial binary split of Samoyedic vs. Finno-Ugric includes the substantial vocabulary replacement in Samoyedic and only there (§1, §2.2) and the derivational patterns shared within Finno-Ugric but absent in Samoyedic (Supplement S1). The near-total absence of Indo-Iranian loans in Samoyedic (§2.2.1), though not a clade-defining event, is evidence of early geographical separation of Samoyedic from the rest and its distance from Indo-Iranian-speaking groups. Other possible evidence is a two-step sound change arguably shared by Samoyedic and Ugric, suggesting a very early branching (Supplement S8); numbers of retained PU etyma are generally higher in the west than in the east, as are inter-branch lexical sharings of PU etyma, suggesting that earlier there was less inter-branch integration in the east (Supplement S8). This could be due to the river geography: while the Volga offers direct east-west connections from near the Urals to eastern Europe, the major rivers in western Siberia flow north or northwest and there are no short overwater connections between the Altai-Sayan and the Urals. (Nonetheless, the locations of ST sites indicate that rivers were major transportation routes in Siberia: Figure 1.)

Thus there is evidence for early distancing of Samoyedic and the eastern branches more generally from the rest of the family and early internal distancing in western Siberia. If the early phylogenetic diversification was anything other than a rake, the Samoyedic branch was the first to split off, followed by Ugric.

3.5 Genetic evidence

Rapid spread of a small population and expansion of its subgroups through language shift should have produced exactly the genetic picture we see in Uralic

(Tambets et al. 2018): There is a detectable pan-Uralic component showing that movement of people was involved to at least some extent in the Uralic spread; and Uralic-speaking populations are everywhere similar to their neighbors, so much so that in some of the westernmost groups the original Uralic genetic component is invisible or nearly so, while in Samoyedic populations the Uralic component is the one that is shared with neighbors. (See also Balanovsky 2019; Saag et al. 2019; Ilumäe et al. 2016; Lamnidis et al. 2018.)

4. Sociolinguistics of post-catastrophe spreads

Little is known of typical sociolinguistic consequences of post-catastrophe spreads (Supplement S9). Catastrophes such as the Plague of Justinian, the Black Death, and prolonged severe droughts have often resulted in language spreads, shifts, and extinctions, and the shifting and the social turmoil of catastrophes may sometimes have led to decomplexification of the surviving language(s) (as is expected when an expanding language absorbs an appreciable number of adult L2 learners: cf. Trudgill 2011). Most modern Finno-Ugric languages are in fact phonologically and morphologically less complex than the general northern Eurasian level and comparable to the languages that have undergone large spreads (German, Spanish, Turkish, Yakut, Mongolian). The Samoyedic languages, which are notably archaic at least in their morphology, are among the most complex, only partly due to post-Proto-Samoyedic developments. Most Saamic languages are among Eurasia's most complex, due to post-Proto-Saamic phonological developments (Supplement S9; Nichols 2019a). Thus it is possible, but not necessary, that early Finno-Ugric has undergone some decomplexification as might be expected of an inter-ethnic trade language. If real, the effect is subtle, however, and the general lack of substratal effects at the branch ancestor level is a stronger argument and one that speaks against impact of shifting speakers on CU grammar.

5. Conclusions

We have argued that Proto-Uralic originated east of the Urals and out of contact with Proto-Indo-European. Its traceable prehistory begins with a mostly westward spread bringing daughter speech communities to the middle Volga. That spread took place rapidly and for the most part without substratal effects. It occurred in the time frame of the 4.2 ka event, the Seima-Turbino transcultural phenomenon, and the Indo-Iranian contact episode, and taken together these three events explain the Uralic spread and situate it in space and time. Early Uralic

spread with ST trade along the rivers that were the main avenues of communication and transport, and this brought it into position for Indo-Iranian contacts. That contact episode with independently well-dated Indo-Iranian gives a reliable absolute date for the Uralic divergence and dispersal: not long before 4,000 years ago. It took place over some expanse of space and some length of time, as loans come separately into the early Uralic daughters (the ancestors of the modern branches) and reflect time frames from Pre-Proto-Indo-Iranian to early Iranian.

That early spread turned Proto-Uralic as reconstructed by the comparative method into Common Uralic, the set of still mutually intelligible but separate and separately evolving varieties that emerged from the initial diversification of Proto-Uralic. Proto-Uralic has a number of eastern typological features suggesting an eastern origin. It also has some rare features that have remained stable in the family, indicating that, while early Uralic must have expanded via language shift, the shifting population had minimal impact on Proto-Uralic grammar and vocabulary. It was Common Uralic that was involved in Seima-Turbino trade and Indo-Iranian contacts. The non-pastoral, non-agricultural, sparsely distributed Common Uralic populations suffered less from the 4.2 ka event and accompanying plague, and were able to recover from it faster than the denser stockbreeding Indo-European populations. As a result, early Uralic replaced IE-speaking populations along the Volga and near the Urals, acquiring no Indo-European substratum and no evidence of IE lexical contact prior to the Indo-Iranian episode.

The early history of the Samoyedic branch is mysterious. If Pre-Samoyedic spread with Seima-Turbino, it spread eastward and not into the area of strong Indo-Iranian contact. Samoyedic retains a low absolute number of Proto-Uralic etyma, but until more Proto-Uralic reconstruction and statistical analysis are done it is not known whether the number of etyma is significantly low.

Northward spreads, a standing pattern in northern Eurasia, gave the Uralic language family its large northern extent. An important contributor to these spreads was language shift. Major northward spreads of Uralic branches proceeded from the upper Yenisei, the middle Volga, and (later, involving Saamic and then Finnic) from the northeast Baltic region. As a result of these spreads, apart from what are now national languages, Uralic languages became known to linguistics and ethnography primarily as languages of hunter-gatherers and northern people.

Funding

Grünthal: Kone foundation project “Grammatical characters in computational phylogeny: The causative alternation in Uralic”. Heyd: Helsinki University Humanities visiting professorship; European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program under grant agreement no. 788616 – YMPACT. Holopainen: Kone Foundation and Fellowship holder of the Austrian Academy of Sciences (APART-GSK) at the Finno-Ugrian department of the University of Vienna. Khanina: Manuscripta Castréniana project funded by the Kordelin Foundation; Helsinki Collegium for Advanced Studies fellowship funded by EURIAS Fellowship Program (cofunded through the European Commission, Marie-Sklodowska-Curie Actions-COFUND Program FP7). Miestamo: Academy of Finland project 332529 “Negation in Clause Combining: Typological and usage-based perspectives (NiCC)”. Nichols: Helsinki University Humanities visiting professorship; Basic Research Program, National Research University Higher School of Economics, Moscow. Sinnemäki: European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement no. 805371).

Acknowledgments

We thank Valentin Gusev, Petri Kallio, Niklas Metsäranta, Christopher Miller, Tapani Salminen, Zsolt Simon, and Nikolay Vakhtin for comments, references, and unpublished material; Nora Fabritius for cartography; Viljami Haakana, Leena Manninen, and Rebecca Unger for research assistance; Partow Imani and Feicheng Qi for statistical consultation; and, above all, two anonymous reviewers.

Abbreviations

BP	(Years) before present. Present is calculated as the year 2000.
CNG	Connegative (the form of the lexical stem following the negative auxiliary verb)
CU	Common Uralic
IE	Indo-European
I-I	Indo-Iranian
PIE	Proto-Indo-European
PU	Proto-Uralic
PX	Possessive suffix (followed by person-number abbreviation)
ST	Seino-Turbino transcultural phenomenon
Others follow the Leipzig glossing conventions.	

References

- Aikio, Ante. 2004. An essay on substrate studies and the origin of Saami. In Irma Hyvärinen, Petri Kallio, and Jarmo Korhonen (eds.), *Etymologie, Entlehnungen und Entwicklungen*, 187–214. Helsinki: Societe Neophilologique.

- Aikio, Ante. 2012. An essay on Saami ethnolinguistic prehistory. In Grünthal, Riho and Petri Kallio (eds.), *A linguistic map of prehistoric northern Europe* (SUST 266), 63–117. Helsinki: Suomalais-Ugrilainen Seura.
- Aikio, Ante. 2013. Studies in Uralic etymology I: Saami etymologies. *Linguistica Uralica* 49. 161–174. <https://doi.org/10.3176/lu.2013.3.01>
- Aikio, Ante. 2014a. Studies in Uralic etymology II: Finnic etymologies. *Linguistica Uralica* 50. 1–19. <https://doi.org/10.3176/lu.2014.1.01>
- Aikio, Ante. 2014b. Studies in Uralic etymology III: Mari etymologies. *Linguistica Uralica* 50. 81–94. <https://doi.org/10.3176/lu.2014.2.01>
- Aikio, Ante. 2015a. Studies in Uralic etymology IV: Ob-Ugric etymologies. *Linguistica Uralica* 51. 1–20. <https://doi.org/10.3176/lu.2015.1.01>
- Aikio, Ante. 2015b. The Finnic ‘secondary e-stems’ and Proto-Uralic vocalism. *SUSA/JSFOu* 95. 25–66. <https://doi.org/10.33340/susa.82642>
- Aikio, Ante. In press. Proto-Uralic. In Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik, (eds.), In press. *The Oxford guide to the Uralic languages*. Oxford: Oxford University Press.
- Alpher, Barry and David Nash. 1999. Lexical replacement and cognate equilibrium in Australia. *Australian Journal of Linguistics* 19(1). 4–56. <https://doi.org/10.1080/07268609908599573>
- Andrades Valtueña, Aida, Alissa Mittnik, Felix M. Key, Wolfgang Haak et al. 2017. The Stone Age plague and its persistence in Eurasia. *Current Biology* 27. 3683–3691. <https://doi.org/10.1016/j.cub.2017.10.025>
- Anthony, David W. 2007. *The horse, the wheel, and language: How Bronze Age riders from the Eurasian steppes shaped the modern world*. Princeton, NJ: Princeton University Press.
- Anthony, David W. & Don Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1. 199–219. <https://doi.org/10.1146/annurev-linguist-030514-124812>
- Atkinson, Quentin D., Andrew Meade, Chris Venditti, Simon J. Greenhill, and Mark Pagel. 2008. Languages evolve in punctuational bursts. *Science* 319. 588. <https://doi.org/10.1126/science.1149683>
- Balanovsky, Oleg. 2019. Evolution of the Y-chromosomal pool in East Europe. Paper presented at *Contacts: Archaeology, genetics, languages. Joining forces to shed light on early contacts (4000 BC – 1000 AD) between Indo-European and Uralic speakers*. Suomenlinna/Helsinki.
- Barfield, Thomas. 1989. *The perilous frontier: Nomadic empires and China, 221 BC to AD 1757*. London: Blackwell.
- Barfield, Thomas. 2009. Introduction: Frontiers and border dynamics. In Bryan K. Hanks & Katheryn M. Linduff (eds.), *Social complexity in prehistoric Eurasia: Monuments, metals, and mobility*, 235–240. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511605376.014>
- Blust, Robert A. 1985. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26(1). 45–67.
- Blust, Robert A. 2009. *The Austronesian languages*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Bownern, Claire, Patience Epps, Russell D. Gray, Jane H. Hill, Keith Hunley, Patrick McConvell, and Jason Zentz. 2011. Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLOS One* 6(9). e25195. <https://doi.org/10.1371/journal.pone.0025195>

- Chernykh, E. N. 2008. The "Steppe Belt" of stockbreeding cultures in Eurasia during the early metal age. *Trabajos de prehistoria* 65(2). 73–93. <https://doi.org/10.3989/tp.2008.08004>
- Comrie, Bernard. 1988. Haruai verb structure and language classification in the Upper Yuat. *Language and Linguistics in Melanesia* 17. 140–160.
- Comrie, Bernard. 1989. Genetic classification, contact, and variation. In Thomas J. Walsh (ed.), *Synchronic and diachronic approaches to linguistic variation and change*, 81–93. Washington, D.C.: Georgetown University Press.
- Comrie, Bernard. 2000. Language contact, lexical borrowing, and semantic fields. In Gilbers, Dicky G., John Nerbonne & Jos Schaeken (eds.), *Languages in contact*, 73–86. Amsterdam: Rodopi.
- Dalfes, H. Nüzhet, George Kukla & Harvey Weiss (eds.). 1997. *Third Millennium BC climate change and Old World collapse*. Berlin/New York: Springer. <https://doi.org/10.1007/978-3-642-60616-8>
- DeFilippo, Cesare, Chiara Barbieri, Mark Whitten, Sununguko Wata Mpoloka, Ellen Drofn Gunnarsdóttir, Koen Bostoen, and Brigitte Pakendorf. 2011. Y-chromosomal variation in sub-Saharan Africa: Insights into the history of Niger-Congo groups. *Molecular Biology and Evolution* 28:3.1255–1269. <https://doi.org/10.1093/molbev/msq312>
- Dryer, Matthew. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292. <https://doi.org/10.1075/sl.13.2.03dry>
- Dryer, Matthew. 2013. Genealogical language list. In Dryer, Matthew S. & Martin Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library, 2013. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: <http://wals.info>. Accessed 2-28-20.
- Dyen, Isidore. 1956. Language distribution and migration theory. *Language* 32. 611–626. <https://doi.org/10.2307/411084>
- Ellison, T. Mark & Luisa Miceli. 2017. Language monitoring in bilinguals as a mechanism for rapid lexical divergence. *Language* 93(2). 255–287. <https://doi.org/10.1353/lan.2017.0014>
- Frachetti, Michael D. 2008. *Pastoralist landscapes and social interaction in Bronze Age Eurasia*. Berkeley-Los Angeles: University of California Press. <https://doi.org/10.1525/9780520942691>
- Goddard, Ives. 1996. The west-to-east cline in Algonquian dialectology. In William Cowan (ed.), *Actes du vingt-cinquieme Congrès des algonquinistes*, 187–211. Ottawa: Carleton University.
- Green, Ian. 2003. The genetic status of Murrinh-Patha. In Nicholas Evans (ed.), *The Non Pama-Nyungan languages of northern Australia*, 125–158. Canberra: Pacific Linguistics.
- Grünthal, Riho & Tapani Salminen. 1993. *Geographical distribution of the Uralic languages*. Helsinki: Suomalais-Ugrilainen Seura.
- Güldemann, Tom, Patrick McConvell & Richard A. Rhodes. 2020. Hunter-gatherer anthropology and language. In Güldemann, Tom, Patrick McConvell & Richard A. Rhodes (eds.), *The language of hunter-gatherers*, 3–48. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781139026208.002>
- Häkkinen, Jaakko. 2009. Kantauralin ajoitus ja paikkannus: perustelut puntarissa [The dating and placing of Proto-Uralic: Weighing the evidence]. *Journal de la Société Finno-Ougrienne* 92. 9–56. <https://doi.org/10.33340/susa.82020>
- Haynie, Hannah J. 2014. Geography and spatial analysis in historical linguistics. *Language and Linguistics Compass* 8(8). 344–357. <https://doi.org/10.1111/lnc3.12087>
- Helama, Samuli & Markku Oinonen. 2019. Exact dating of the Meghalayan lower boundary based on high-latitude tree-ring isotope chronology. *Quaternary Science Reviews*, online: <https://doi.org/10.1016/j.quascirev.2019.04.013>

- Helimski, E. A. 1982. *Drevnejšie vengersko-samodijskie jazykovye paralleli* [The oldest Hungarian-Samoyed language parallels]. Moscow: Nauka.
- Helimski, Evgenij. 2001a. Ural'cy i ix predšestvenniki: Belye pjatna na ètnoistoričeskoj karte severnoj evrazii i ural'skie jazyki [Urals and IX predecessors: White spots on the historical map of Northern Eurasia and Uralic languages]. *Congressus nonus internationalis fenno-ugristarum IV, Linguistica* 1. 332–336.
- Helimski, Eugene. 2001b. Early Indo-Uralic linguistic relationship: Real kinship and imagined contacts. Christian Carpelan, Asko Parpola, and Petteri Koskikallio, eds., *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*, 187–205. Helsinki: Suomalais-Ugrilainen seura.
- Helimski, Eugene. 2003. Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: A survey. *Nyelvtudományi Közlemények* 100. 156–167.
- Holman, Eric W., Christian Schulze, Dietrich Stauffer & Søren Wichmann. 2007. On the relation between structural diversity and geographical distance among languages: Observations and computer simulations. *Linguistic Typology* 11(2). 395–423. <https://doi.org/10.1515/LINGTY.2007.027>
- Holopainen, Sampsa. 2019. Indo-Iranian borrowings in Uralic: Critical overview of the sound substitutions and distribution criterion. PhD dissertation, University of Helsinki.
- Honti, László. 1998. Obugrian. In Daniel Abondolo (ed.), *The Uralic languages*, 327–357. London-New York: Routledge.
- Huisman, John L. A., Asifa Majid & Roeland van Hout. 2019. The geographical configuration of a language area influences linguistic diversity. *PLOS One* 14(6). e0217363. <https://doi.org/10.1371/journal.pone.0217363>
- Hunley, Keith, Michael Dunn, Eva Lindström, Ger P. Reesink, Angela Terrill, Meghan Healy, George Koki, Françoise R. Friedlaender & Jonathan S. Friedlaender. 2008. Genetic and linguistic coevolution in northern Island Melanesia. *PLOS Genetics* 4(10). e1000239. <https://doi.org/10.1371/journal.pgen.1000239>
- Illumäe, Anne-Mai, Maere Reidla, Marina Chukhryaeva, Mari Järve, Helen Post, Monika Karmin, ... & Siiri Roots. 2016. Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *American Journal of Human Genetics* 99. 163–173. <https://doi.org/10.1016/j.ajhg.2016.05.025>
- Itkonen, Erkki. 1973. Zur Geschichte des Partitivs. *Finnisch-Ugrische Forschungen* 40. 279–339.
- Iversen, Rune & Guus Kroonen. 2017. Talking Neolithic: Linguistic and archaeological perspectives on how Indo-European was implemented in southern Scandinavia. *American Journal of Archaeology* 121(4). 511–525. <https://doi.org/10.3764/aja.121.4.0511>
- Janhunen, Juha. 1982. On the structure of Proto-Uralic. *Finnisch-Ugrische Forschungen* 44(1–3). 23–42.
- Janhunen, Juha. 2000. Reconstructing Pre-Proto-Uralic typology: Spanning the millennia of linguistic evolution. In Anu Nurk, Tõnu Seilenthal, & Triinu Palo (eds.), *Congressus Nonus Internationalis Fenno-Ugristarum, Pars I: Orationes plenariae & orationes publicae*, 59–76. Tartu: Eesti Fennougristide Komitee, Tartu Ülikool.
- Janhunen, Juha. 2008. Mongolic as an expansive language family. In Tokusu Kurebito (ed.), *Past and present dynamics: The Great Mongolian state*, 127–137. Tokyo: Tokyo University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa.
- Janhunen, Juha. 2009. Proto-Uralic – what, where, and when? *The Quasiquintennial of the Finno-Ugrian Society*, 57–78. Helsinki: Suomalais-Ugrilainen Seura.
- Janhunen, Juha. 2012a. Etymological and ethnohistorical aspects of the Yenisei. *Studia Etymologica Cracoviensia* 17. 67–87.

- Janhunen, Juha. 2012b. The expansion of Tungusic as an ethnic and linguistic process. In Malchukov, Andrej L. & Lindsay J. Whaley (eds.), *Recent advances in Tungusic linguistics*, 5–16. Harrassowitz: Wiesbaden.
- Janhunen, Juha. 2014. A legkeletibb uráliak. [The easternmost Uralians.] *Nyelvtudományi Közlemények* 110. 7–30.
- Janhunen, Juha. 2020. Grammaticalization in Uralic as viewed from a general Eurasian perspective. In Andrej Malchukov & Walter Bisang (eds.), *Handbook of grammaticalization*. Berlin: de Gruyter Mouton. <https://doi.org/10.1515/9783110563146-007>
- Kallio, Petri. 2015. Nugae indo-uralicae. *Journal of Indo-European Studies* 43(3–4). 368–375.
- Kassian, Alexei, Mikhail Zhivlov & George Starostin. 2015. Proto-Indo-European-Uralic comparison from the probabilistic point of view. *Journal of Indo-European Studies* 43(3–4). 301–347.
- Kessler, Brett. 2015. Response to Kassian et al., Proto-Indo-European-Uralic comparison from the probabilistic point of view. *Journal of Indo-European Studies* 43(3–4). 357–367.
- Kessler, Brett & Annukka Lehtonen. 2006. Multilateral comparison and significance testing of the Indo-Uralic question. In Clackson, James, Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 33–42. Cambridge: McDonald Institute for Archaeological Research.
- Key, Felix M., Cosimo Posth, Luis R. Esquivel-Gomez, Ron Hübler et al. 2020. Emergence of human-adapted *salmonella enterica* is linked to the Neolithization process. *Nature Ecology and Evolution* 4. 324–333. <https://doi.org/10.1038/s41559-020-1106-9>
- Khanina, Olesya. In press. A history of Northern Samoyedic: Adding details to the dialect continuum. In Anikin, Alexander, Valentin Gusev & Anna Urmanchieva (eds.), *Siberica et Uralica: To the memory of Eugen Helimski*.
- Khanina, Olesya, Yuri Koryakov & Andrey Shluinsky. 2018. Enets in space and time: A case study in linguistic geography. *Finnisch-Ugrische Mitteilungen* 42. 109–135.
- Koivulehto, Jorma. 2001. The earliest contacts between Indo-European and Uralic speakers in the light of lexical loans. Christian Carpelan, Asko Parpola, and Petteri Koskikallio, eds., *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*, 235–263. Helsinki: Suomalais-Ugrilainen seura.
- Krause, Rüdiger, Andrei V. Epimakhov, E. B. Kupriyanova, Igor A. Novikov, Eliza Stolarczyk. 2019. Petrovskie pamjatniki bronzovogo veka: Problemy taksonomii i khronologii. [The Petrovka Bronze Age sites: Issues in taxonomy and chronology.] *Arxeologija, ètnografija i antropologija Evrazii* 47(1). 54–63. <https://doi.org/10.17746/1563-0110.2019.47.1.054-063>
- Kulonen, Ulla-Maija. 2001. Über die Deklination der Personalpronomina in der finnisch-ugrischen Grundsprache. In Anu Turk, Triinu Palo & Tõnu Seilenthal (eds.), *Congressus Nonus Internationalis Fenno-Ugristarum*, 6. 178–182. Tartu: Eesti Fennougristide Komitee, Tartu Ülikool.
- Kuzmina, Elena E. 2007. In Elena E. Kuzmina and J. P. Mallory (eds.), *Origin of the Indo-Iranians*. Leiden: Brill. <https://doi.org/10.1163/ej.9789004160545.i-763>
- Lamnidis, Thiseas, Kerttu Majander, Choongwon Jeong, Elina Salmela, Anna Wessman, Vyacheslav Moiseyev, ... & Stephan Schiffels. 2018. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature Communications* 9. 5018. <https://doi.org/10.1038/s41467-018-07483-5>
- Lang, Valter. 2018. *Läänemeresoome tulemised* [Finnic be-comings]. Tartu: Tartu Ülikooli Kirjastus.
- Lefebvre, Claire. 1998. *Creole genesis and the acquisition of grammar: The case of Haitian Creole*. Cambridge: Cambridge University Press.

- Lindner, Stephan. 2020. Chariots in the Eurasian steppe: A Bayesian approach to the emergence of horse-drawn transport in the early second millennium BC. *Antiquity* 94(374). 361–380. <https://doi.org/10.15184/aqy.2020.37> (accessed July 1, 2020)
- Majtinskaja, K. E. 1993. *Istoriko-sopostavitel'naja morfologija finno-ugorskix jazykov* [Historical-comparative morphology of the Finno-Ugric languages], 2nd edn. Moscow: URSS.
- Mallory, James P. 1989. *In search of the Indo-Europeans: Language, archaeology, and myth*. New York: Thames & Hudson.
- Marchenko, Zhanna V., Svetlana V. Svyatko, V. I. Molodin, A. E. Grishin & M. P. Rykun. 2017. Radiocarbon chronology of complexes with Seima-Turbino type objects (Bronze Age) in southwestern Siberia. *Radiocarbon* 59(5). 1381–1397. <https://doi.org/10.1017/RDC.2017.24>
- Miceli, Luisa. 2015. Pama-Nyungan. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 704–725. Oxon & New York, Abingdon: Routledge.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226580593.001.0001>
- Nichols, Johanna. 1997. Modeling ancient population structures and movement in linguistics. *Annual Review of Anthropology* 26. 359–384. <https://doi.org/10.1146/annurev.anthro.26.1.359>
- Nichols, Johanna. 2010. Proof of Dene-Yeniseian relatedness. In James Kari & Ben A. Potter, (eds.), *The Dene-Yeniseian Connection*, 266–278. Fairbanks: Alaska Native Language Center..
- Nichols, Johanna. 2019. Grammatical complexity in Saami: A historical-typological view. Presented at SAALS 4, Uppsala.
- Nichols, Johanna & Richard A. Rhodes. 2018. Vectors of language spread at the central steppe periphery: Finno-Ugric as catalyst language. In Guus Kroonen & Rune Iversen (eds.), *Digging for words*, 58–68. Oxford: BAR Publishing.
- Nordqvist, Kerkko & Volker Heyd. 2020. The forgotten child of the wider Corded Ware family: The Russian Fatyanovo culture in context. *Proceedings of the Prehistoric Society* 86. 65–93. <https://doi.org/10.1017/ppr.2020.9>
- Parpola, Asko. 1999. Varhaisten indoeurooppalaiskontaktien ajoitus ja paikannus kielellisen ja arkeologisen aineiston perusteella [The dating and placing of early Indo-European contacts on the basis of linguistic and archaeological material]. In Paul Fogelberg (ed.), *Pohjan poluilla: Suomalaisten juuret nykytutkimuksen mukaan* [Tracing the North: The roots of the Finns in the light of recent research], 180–206. Helsinki: Societas Scientiarum Fennica.
- Parpola, Asko. 2012. Formation of the Indo-European and Uralic (Finno-Ugric) language families in the light of archaeology: Revised and integrated “total” correlations. In Riho Grünthal and Petri Kallio (eds.), *A linguistic map of prehistoric northern Europe*, 119–184. (SUST 266.) Helsinki: Suomalais-Ugrilainen Seura.
- Perşoiu, Aurel, Monica Ionita & Harvey Weiss. 2019. Atmospheric blocking induced by the strengthened Siberian high led to drying in West Asia during the 4.2 ka BP event – a hypothesis. *Climate of the Past* 15. 781–793. <https://doi.org/10.5194/cp-15-781-2019>
- Rasmussen, Simon, Allentoft, Morten E., Nielsen, Kasper, Orlando, Ludovic, Sikora, Martin, Sjögren, Karl-Göran, ... & Willerslev, Eske. 2015. Early divergent strains of *Yersinia pestis* in Eurasia 5000 years ago. *Cell* 163. 571–572. <https://doi.org/10.1016/j.cell.2015.10.009>
- Rascovan, Nicolás, Sjögren, Karl-Göran, Kristiansen, Kristian, Nielsen, Rasmus, Willerslev, Eske, Desnues, Christelle, & Rasmussen, Simon. 2019. Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic decline. *Cell* 176. 1–11. <https://doi.org/10.1016/j.cell.2018.11.005>

- Ringe, Don. 2015. Response to Kassian et al., Proto-Indo-European-Uralic comparison from the probabilistic point of view. *Journal of Indo-European Studies* 43(3–4). 348–356.
- Saag, Lehti, Margot Laneman, Liivi Varul et al. 2019. The arrival of Siberian ancestry connecting the Eastern Baltic to Uralic speakers further east. *Current Biology* 29. 1–11. <https://doi.org/10.1016/j.cub.2019.04.026>
- Saarikivi, Janne. 2004a. Palaeo-European substratum interferences in the western branches of Uralic? *Suomalais-Ugrilaisen Seuran Aikakauskirja / Journal de la Société Finno-Ougrienne* 90. 187–214.
- Saarikivi, Janne. 2004b. Über das saamische Substratnamengut in Nordrussland und Finnland. *Finnisch-Ugrische Forschungen* 58(1–3). 187–214.
- Saarikivi, Janne. 2006. Substrata uralica: Studies on Finno-Ugrian substrate in northern Russian dialects. PhD Dissertation, University of Helsinki.
- Saarikivi, Janne. In press. The divergence of Proto-Uralic and its offspring: A descendant reconstruction. In Marianne Bakro-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages* Oxford: Oxford University Press.
- Saarikivi, Janne and Mika Lavento. 2012. Linguistics and archaeology: A critical view of an interdisciplinary approach with reference to the prehistory of northern Scandinavia. In Charlotte Damm and Janne Saarikivi (eds.), *Networks, interaction, and emerging identities in Fennoscandia and beyond*, 177–216. Helsinki: Suomalais-Ugrilainen Seura.
- Salminen, Tapani. 2001. The rise of the Finno-Ugric language family. In Christian Carpelan, Asko Parpola, and Petteri Koskikallio (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*, 385–396. Helsinki: Suomalais-Ugrilainen Seura.
- Sapir, Edward. 1916. Time perspective in aboriginal American culture: A study in method. In David G. Mandelbaum (ed.), *Selected writings of Edward Sapir in language, culture, and personality*, 389–467. Berkeley-Los Angeles: University of California Press.
- Simon, Zsolt. 2020. Urindogermanische Lehnwörter in den uralischen und finno-ugrischen Grundsprachen: Eine Fata Morgana? *Indogermanische Forschungen* 125. 239–266. <https://doi.org/10.1515/if-2020-011>
- Sipos, Mária. 2002. Az obi-ugor alapnyelv lexikális innovációi I. [Lexical innovations in Proto-Ob-Ugric, part 1.] *Nyelvtudományi Közlemények* 99. 7–56.
- Spyrou, Maria A., Tukhbatova, Rezeda I., Wang, Chuan-Chao, Andrades Valtueña, Aida, Lankapalli, Aditya K., Tsybin, Victor A., ... & Johannes Krause. 2018. Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for Bubonic Plague. *Nature Communications* 9. 2234. <https://doi.org/10.1038/s41467-018-04550-9>
- Tambets, Kristiina, Bayazit Yunusbayev, Georgi Hudjashov, Anne-Mai Ilumäe, Siiri Rootsi, Terhi Honkola, ... & Mait Metspalu. 2018. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biology* 19. 139. <https://doi.org/10.1186/s13059-018-1522-1>
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic structure and complexity*. Oxford: Oxford University Press.
- Zhivlov, Mikhail A. 2015. Neindoevropskij substrat v finno-volžskix jazykax [The non-Indo-European substratum in the Finno-Volgaic languages]. *X Čteníja pamjati S. A. Starostina* [Readings in honor of S. A. Starostin, 10].

Appendix 1. Proto-Uralic cognates

<<https://doi.org/10.5281/zenodo.6035956>>

Appendix 2. Indo-Iranian loans into early Uralic

<<https://doi.org/10.5281/zenodo.6035900>>

Appendix 3. Ugric vocabulary table

<<https://doi.org/10.5281/zenodo.6035881>>

Résumé

La famille ouralienne très répandue offre plusieurs avantages pour retracer la préhistoire: un solide point d'ancrage dans une chronologie absolue comprenant un ancien épisode de contact avec une langue (l'indo-iranien) bien datée; d'autres points d'intersection ou bien de non-intersection diagnostique avec les débuts indo-européens (la défunte culture Yamna de la steppe occidentale parlant un proto-indo-européen, la culture Afanasievo du haut Iénisseï, et la culture Fatianovo de la Moyenne Volga); de la reconstruction lexicale et morphologique suffisante pour établir des absences critiques de partages et de contacts. Nous ajoutons des informations sur le climat, la géographie linguistique, la typologie et des distributions de fréquences de mots apparentés pour reconstituer l'origine ainsi que l'expansion ouraliennes. Nous faisons valoir que le territoire ouralien était à l'est de l'Oural et initialement hors de contact avec l'indo-européen. L'expansion a été rapide et sans effets substratiques communs largement répandus. Nous reconstruisons sa cause comme les réactions reliées entre elles des premières populations ouraliennes et indo-européennes à un épisode catastrophique de changement climatique et aux possibilités provoquées par une croissance des contacts qui ont favorisé les chasseurs-pêcheurs riverains par rapport aux éleveurs.

Zusammenfassung


Die weitverbreitete Familie der uralischen Sprachen bietet mehrere Vorteile zur Nachverfolgung ihrer Frühzeit: ein fester absolutchronologischer Ankerpunkt durch eine Episode früher Kontakte mit gut datierten indo-iranischen Sprachen; andere Punkte der Überschneidung oder diagnostischer 'Nicht-Überschneidung' mit frühen indoeuropäischen Sprachen (die spätproto-indoeuropäisch sprechende Jamnaja-Kultur der westlichen Steppe, die Afanasievo-Kultur des oberen Jenissei und die Fatjanovo-Kultur der mittleren Wolga); ausreichende lexikalische und morphologische Rekonstruktion, um kritische Abwesenheiten von Teilhabe und Kontakten herzustellen. Wir fügen Informationen zu Klima, Sprachgeographie, Typologie und Häufigkeitsverteilungen von Kognaten hinzu, um den Ursprung und die Ausbreitung des


Uralischen zu rekonstruieren. Wir argumentieren, dass sich die Urheimat des Uralischen östlich des Uralgebirges und anfangs auch jenseits von Kontaktzonen mit den indoeuropäischen Sprachen befand. Die Ausbreitung war schnell und ohne weit verbreitete, gemeinsame Substrateffekte. Wir rekonstruieren ihre Ursache in verknüpften Reaktionen früher uralischer und indoeuropäischer Bevölkerungen auf einen katastrophalen Klimawechsel sowie den Chancen einer Interregionalisierung, die flussnahe Jäger und Fischer gegenüber Viehzüchtern begünstigte.

Address for correspondence


Riho Grünthal
 Finno-Ugrian Studies
 University of Helsinki
 P.O.Box 24
 Unioninkatu 40
 FIN-00014 UNIVERSITY OF HELSINKI
 Finland
 riho.grunthal@helsinki.fi
 <https://orcid.org/0000-0002-7648-5687>


Co-author information


Volker Heyd
 University of Helsinki
 volker.heyd@helsinki.fi
 <https://orcid.org/0000-0002-0889-6513>

Sampsa Holopainen
 University of Helsinki
 University of Vienna
 sampsa.holopainen@helsinki.fi
 <https://orcid.org/0000-0001-6796-1648>


Juha A. Janhunen
 University of Helsinki
 asiemajeure@yahoo.com

Olesya Khanina
 Institute of Linguistics
 Russian Academy of Sciences
 University of Helsinki
 olesya.khanina@gmail.com
 <https://orcid.org/0000-0001-5930-4656>

Matti Miestamo
 University of Helsinki
 matti.miestamo@helsinki.fi
 <https://orcid.org/0000-0001-9801-9030>

Johanna Nichols
 University of California, Berkeley
 Higher School of Economics, Moscow &
 (during writing) University of Helsinki
 johanna@berkeley.edu
 <https://orcid.org/0000-0002-3709-1861>

Janne Saarikivi
 University of Helsinki
 janne.saarikivi@helsinki.fi

Kaius Sinnemäki
 University of Helsinki
 kaius.sinnemaki@helsinki.fi
 <https://orcid.org/0000-0002-6972-5216>

Publication history

Date received: 20 July 2020

Date accepted: 14 May 2021

Published online: 8 April 2022