Comment: Was Late Proto-Uralic spoken in the Altai Region?

Jaakko Häkkinen, August 14th, 2025

A reply to:

Kommunikationsräume und interkulturelle Netzwerke im bronzezeitlichen Sibirien: Ein neuer Blick auf Sejma-Turbino, den „Flowerpot Complex" und die Ausbreitung der uralischen Sprachen

(Henny Piezonka, Elina Salmela & Outi Vesakoski 2024)

Uralic Archaeolinguistics

(Outi Vesakoski, Elina Salmela & Henny Piezonka 2025)

Contents   **Page**

Figures and tables **Page**

1. Foreword

"...maybe Proto-Uralic was spoken within the eastern settlements of the 'flowerpot complex', located near the 'Altaic homeland' at the time when manufacturing and trading of prestigious bronze items started to become an important sociocultural trait, and while 'Common Uralic' subsequently developed as the lingua franca of the developing wider communication network along the southern fringe of the West Siberian taiga." (Vesakoski, Salmela & Piezonka 2025: 375)

Two recent multidisciplinary articles from the same authors suggest that Proto-Uralic (henceforth Late Proto-Uralic) could be located in the Altai Region. Here I focus on the latter article written in English, but the same views are present in the German article as well. Unfortunately, even though the authors thank me for commenting the draft version of the English article, they chose to ignore all the linguistic argumentation presented in Häkkinen (2023b). Instead, the authors kept relying on Grünthal et al. (2022), argumentation of which was already thoroughly countered in Häkkinen (2023b). In that article I evaluated a wide variety of arguments and pieces of evidence, showing that the Siberian homeland (outside the immediate eastern side of the Central Ural Region) for Late Proto-Uralic is not supported by any valid piece of linguistic evidence.

As scientists, we must acknowledge all the valid arguments and pieces of evidence – cherry-picking is not acceptable. If we disagree with some results, we must counter the presented arguments with stronger arguments and/or disprove the earlier arguments. Due to the decision of the authors, I feel obliged to write this long comment to explain thoroughly why the valid linguistic evidence does not support locating the Late Proto-Uralic homeland in the Altai Region or Southern Siberia.

It is possible that I might later formulate arguments and proposals from this comment to a series of articles to be published in linguistic journals. But because speed is of the essence here, lest the selective and tenuous views take more root in the forthcoming cycle of scholarly references, I decided to publish this comment right away and on my own. This also serves best those who are currently led astray.

For the defense of the authors, I acknowledge that in their search for support they refer to linguistic articles which I had not considered in my article (Häkkinen 2023b). Therefore, as a sequel to my earlier article, I will in this comment evaluate whether these referred articles can give testimony relevant for the location of Late Proto-Uralic. The articles are: Janhunen (2022), Bjørn (2022), Peyrot (2019a), and Warries (2022). I will show that all the arguments presented in these articles are either insufficient or chronologically too vague to support locating Late Proto-Uralic in the Altai Region.

At the same time, I will use this opportunity to explore further several topics related to the divergence and development of the Uralic languages and the methodology of interdisciplinary studies. The reader might want to start from Chapter 6 (Summary) and from there trace the topics and arguments one finds interesting.

1.1. Terminology

Here the label *Southern Siberia* contains the region around the upper streams of Ob and Yenisei (including Middle Irtysh) as well as the Sayan Mountains. The label *Altai Region* seems to partially overlap with this area, although the Altai Mountains are mainly within the borders of Kazakhstan, China, and Mongolia. In the Uralic context, the Altai Region refers to its northern part within Southern Siberia, so these labels can be used synonymously for this overlapping region.

As I will continue to emphasize the importance of using the most accurate possible chronological stages, I will refer to Late Proto-Uralic (LPU) as the most recent common ancestor of all the Uralic languages. Early Proto-Uralic (EPU) is a somewhat earlier stage, connectable to the earliest contacts with the Tocharian language lineage. Pre-Proto-Uralic (PrePU) denotes here chronologically still more distant continuum of development concerning the Uralic language lineage.

Pre-Proto-Samoyed (PrePSy, and similarly with other Uralic branches) means a reconstruction stage after the regional divergence of Uralic speech communities but before any sound changes occurred, while Early Proto-Samoyed (EPSy) includes at least the East Uralic sound changes (Chapter 3.2.2) and continues through the earliest branch-specific sound changes.

Late Proto-Samoyed (LPSy) is the stage immediately preceding the internal divergence of this branch, and Middle Proto-Samoyed (MPSy) could be applied if necessary to distinguish yet another reconstruction stage between the early and the late proto-stage. Language lineage (LL) covers the whole continuum of developments leading to a certain extant or literarily attested language or group of related languages.

Nine extant Uralic branches are Saami (Sa), Finnic (Fi), Mordvin (Md), Mari (Mr), Permic (Pe), Hungarian (Hu), Mansi (Mn), Khanty (Kh), and Samoyed (Sy). At least one extinct branch is known: Meryanic (Me). These abbreviations combine with the aforementioned abbreviations: PrePSa = Pre-Proto-Saami, LPMs = Late Proto-Mansi, etc.

Concepts substrate, superstrate, and adstrate are used here in the sense of historical linguistics (Kaheinen 2023: 21–26). (1) A substrate language is an earlier local language, the speakers of which shift to a newcomer language, leaving traits from their original language into the newcomer language; (2) a superstrate language is a newcomer language, the speakers of which shift to an earlier local language, leaving traits from their original language into the local language; (3) an adstrate language is an adjacent language which leaves traits in another language without language shift ("normal" contacts).

1.2. The Late Proto-Uralic sound system

I use the established Uralic transcription (UPA) which differs from the International Phonetic Alphabet (IPA), so I will describe the differences briefly. Concerning consonants of Uralic reconstructions, I follow Aikio (2022). Familiar consonants are *p, *t, *k, *l, and *r. For sibilants and affricates (all voiceless), Uralic reconstruction distinguishes alveolar *s, post-alveolar *š [ʃ] and *č [t͡ʃ], and alveolopalatal *ć [t͡ɕ] (or sibilant *ś [ɕ]). For spirants, voiced dental spirant *δ [ð] and palatalized voiced dental spirant *δ' are distinguished. For nasals, *m and *n are accompanied by alveolopalatal *ń [ɲ] and velar *ŋ [ŋ]. Glides, semivowels, approximants, or voiced fricatives (depending on interpretation) are bilabial *w [β] and palatal *j. The mysterious *x was most probably the voiced velar spirant [ɣ].

The assumed phonetic nature of certain sounds in Uralic and elsewhere is relevant for some of the topics considered in this comment: (1) the relationship between *a and *e̞ (Chapter

2.5); (2) the relationship between voiceless sibilants and other voiceless fricatives (Chapter 3.2.2.1). Fricatives are considered in the mentioned chapter.

For vowels, I utilize phonetic results from the Estonian vowel system, as it is nearly identical with the vowel system of eight established vowels reconstructed in Late Proto-Uralic – the only difference being that Estonian has a ninth vowel *ö [ø]. The framework for phonetic realizations of the articulatory parameters, height (three distinct classes), backness (two distinct classes), and roundness (two distinct classes), are considered similar to Late Proto-Uralic (Figure 1).

Estonian /õ/ [ɤ] has sometimes been classified as a central vowel, but it seems to be a clear back vowel concerning both its articulation (as inferred from X-ray pictures) and its auditory realization (Eek & Meister 1994). The attribute *central* is often used opposing *peripheral* in the perceptual vowel space (as in Gordon et al. 2012), although central vowels in this meaning are articulatorily a very diverse group: they can include rounded front vowels (*ü* [y], *ö* [ø]), unrounded back vowels (*ï* [ɯ], *e̮* [ɤ] = õ), and vowels between otherwise similar front and back vowels (like [ʉ] between *ü* [y] and *u* [u]). For the sake of clarity, it is better to talk about *inner **vs.** outer* vowels instead of central **vs.** peripheral vowels when denoting location in the two-dimensional perceptional space. The label *central vowel* should be left for true central vowels like [ʉ], representing the third articulatory class between the front and back vowels.
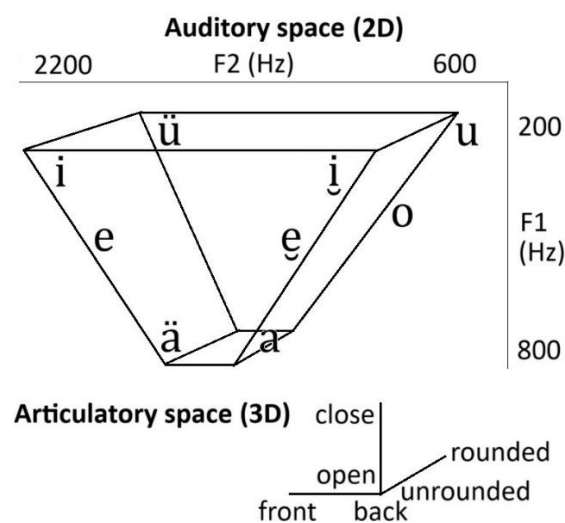
Figure 1: The reconstructed vowel system of Late

Proto-Uralic in auditory and articulatory space.

The reconstructed vowel system of Late Proto-Uralic is shown in Figure 1. The first formant F1 distinguishes the vowel height, while the second formant F2 distinguishes frontness vs. backness. The third formant F3 seems to distinguish slightly between rounded and unrounded vowels but only concerning outer vowels. Here the unrounded back vowel *ẹ/õ* behaves rather weirdly, aligning with rounded vowels (Eek & Meister 1994: 65). Roundedness is also visible in F2: the rounded vowels are more back than the corresponding unrounded vowels.

In the later syllables of Late Proto-Uralic, usually only **a*, **ä*, and **i* (~ **ə* in Häkkinen 2023b following Kallio 2012a) are reconstructed, although there are some words suggesting also **o* (Aikio 2015: 37–38). Janhunen (1981) and Sammallahti (1988) reconstructed four vowels in the later syllables: **a vs. *ä* and **i̬ vs. *i*. Aikio (2022: 8) only accepts **a*, **ä*, and **i*, although also the first two were vowel harmonic allophones.

As pointed out by Korhonen (1988: 11), the opposition **a vs. *ä* would have appeared only after the first syllable **i*, while elsewhere **a* and **ä* were in a complementary distribution. This kind of distribution strongly suggests the secondary nature of the vowel harmony and still an allophonic relationship between **a* and **ä* in the second syllable of Late Proto-Uralic. Here are some established Late Proto-Uralic words which have been traditionally reconstructed as original **i-a*-combinations:

1. LPU **ńi̬la* 'phloem; bast' > PFi **nila* | PSa **ńẹlē* 'the condition in which the bark is loose on the wood; something slippery,' **ńẹlẹ-* 'to peel off' | PMd **nola* | (POU **ńə̂lV-* 'to peel off' >) PMs **ńal-*, PKh **ńi̬l-* (Zhivlov 2023: 168)

2. LPU **wi̬ša* 'green, yellow' > PFi **viha|nta* | PMd **ožə* | PMr **ŭž|ar* | PPe **vež* > PKomi **vež*, PUdm **vož* | (POU **wə̂sV-* >) PKh **wi̬š|ti̬* (Zhivlov 2023: 168)

3. LPU **i̬ša* 'outer skin' > PFi **iho* 'outer skin' | PSa **ẹsē* 'inner side of hide, skin under the hair' | MdE *jožo* 'outer skin, surface' | PPe **ež* > Komi *ež* 'surface, cover, upper or outer side' | PMs **aš* 'surface, face, appearance' | PKh **as* 'skin, surface, outer layer of birch-bark' (Holopainen & Aikio 2023: 11)

The Finnic back combination *i-a* has a back vowel correspondence in many branches. Although Finnic seems to point to original *i*, Saami to original *i* or *ü*, and Permic to original *i*, Zhivlov's (2023: 146–147) Ob-Ugric reconstruction LPU *i-a* > OU *ə̑-V* > Mansi *a-* ~ Khanty *i̮-* is unique. In the other occasions of reconstructed *ə̑-*, Khanty shows *a* (< U *aCCi*, *oKCi*). Likewise, in Mordvin the regular result is unique: the original *u-a* > LPMd *o-a*, but the original *i̮-a* > LPMd *o-ə* with one exception.

Finnic	Proto-Uralic	Mordvin	Mansi
*ü-ä	< *ü-a >	*e-(ə)	*(w)ä-
*i-ä	< *i-a >	*e-(ə)	*ä-
*u-a	< *u-a >	*o-a	*(w)a-
*i-a	< *i̮-a >	*o-(ə)	*a-

Figure 2: The outcome of the four vowel combinations in Finnic, Mordvin, and Mansi points to the presence of *i̮* in the first syllable of Late Proto-Uralic.

Especially Mordvin and Mansi clearly agree with their mergers of the four vowel combinations, as can be seen in Figure 2. Finnic has preserved the distinction between rounded and unrounded vowels, but the two unrounded vowels have merged into *i*. Mordvin and Mansi have preserved the original distinction between front and back vowels, and Mansi has also preserved the distinction of rounded and unrounded vowels next to a velar consonant, and Mordvin in the word-initial rounded front vowel (*ü-* > *ve-*).

Unique reflexes of this vowel combination in Mordvin and Ob-Ugric together with different kinds of mergers in different branches seem to require four distinct vowels in the first syllable of Late Proto-Uralic. Based on "qualitative triangulation" it appears to have been the back pair of *i* – therefore *i̮* is reconstructed here. I have earlier proposed the reconstruction of *ô* (*ə̑*) for the words showing Finnic *i-a* (Häkkinen 2007), while Zhivlov (2023: 120) now seems to agree with Korhonen (1988) and accepts here *i̮*, which means that the

traditional Uralic *ị (since Janhunen 1981 and Sammallahti 1988) must be interpreted as *ẹ, as has been argued earlier on several grounds (Häkkinen 2007: 60–61; Pystynen 2014).

Consequently, I will apply the following modifications in the traditional Late Proto-Uralic reconstructions, in this order:

1. *ị => *ẹ (*mịksa => *mẹksa 'liver')
2. *i-a => *ị-a (*ńila => *ńịla 'phloem, bast')
3. *ä => *a in the later syllables (*ćilmä => *ćilma 'eye')

As an exception, *jilma => *jilmä 'air, weather', because the cognates point to a front vowel everywhere outside Finnic (Zhivlov 2023: 168). Therefore, this is not an original *ị-a-word but Finnic *ilma shows a secondary back vowel in the second syllable.

These changes mean that the vowel harmony in Late Proto-Uralic was just an automatic progressive assimilation: the vowels in the second syllable were pronounced more back after a back vowel in the first syllable and more front after a front vowel in the first syllable. In some Uralic branches (like Finnic and Samoyed) the vowel harmony became phonological, but in others it never did. For example, Saami shows no traces of the Pre-Proto-Saami vowel harmony despite having preserved the second syllable vowels distinct from each other. It is only a consistent habit to reconstruct *ä in the second syllable of the Pre-Proto-Saami words and then assume immediately the merger of *ä and *a there (as in Korhonen 1981: 99; Sammallahti 1998: 181).

In reconstructions, following additional symbols are used: *V = vowel of uncertain quality or any vowel | *V̈ = some or any front vowel | *C = any consonant | *K = any velar consonant | *P = any labial consonant | *V-V, *V-CV, *VC-CV = syllable boundary; vowel combination including the vowels of the first and the second syllable; open first syllable ends with a vowel, closed first syllable ends with a consonant.

To keep the reconstructions readable, I write the Indo-Iranian affricates with digraphs: LPIIr *tś, *dź, *dźʰ > PIr *ts, *dz > Ir *s, *z.

2. Articles supposedly supporting the Late Proto-Uralic homeland in the Altai Region

2.1. Janhunen (2020 ← 2000)

> *"The most uncontroversial information on the pre-historical location and movements of Uralic on the map is, however, provided by the internal taxonomy of the language family. The very fact that the branchings of Uralic seem to become chronologically shallower the farther west we proceed suggests that the main direction of expansion has been systematically from east to west."* (Janhunen 2009: 71)

Vesakoski, Salmela, and Piezonka (2025) refer to Janhunen (2022), in which he refers to Janhunen (2014), in which he refers to Janhunen (2009), in which he refers to Janhunen (2000), in which he drafted the evidence for the Uralic homeland in the Altai Region. His arguments are (1) the shared Ural-Altaic typology, and (2) the westward-branching taxonomic structure of the Uralic language family based on basic numerals 1–10.

The first argument is plainly anachronistic: Even if there was a period of shared areal development between the language families of the so-called Ural-Altaic typology, it would have in any case long preceded the stage of Late Proto-Uralic, most probably by several millennia. The location of so distant a predecessor cannot testify to anything about the location of the Late Proto-Uralic homeland, so it is plainly irrelevant.

Sheer typological properties are in any case weaker evidence than more concrete material correspondences (shared words or morphological material), such as those proposed between Uralic and Indo-European or between Uralic and Eskimo-Aleut (a recent review is Georg 2023). Due to several contradicting pairwise associations between different language families, we do not yet know whether one of these distant comparisons holds the true affinity or are they all mere illusions. We do know that the criteria of the historical comparative method must be loosened in order to even find so distant relatedness, so these all could be false positive results. Consequently, we do not know whether the distant Pre-Proto-Uralic was spoken to the east or to the west from the Ural Mountains.

The second argument by Janhunen is based only on extremely narrow data: an insignificant set of numerals showing random irregularities, which could be countered by several equally small sets of words pointing to contradicting taxonomic hierarchy. For example, we could

find a Uralic word showing the same irregular development in Khanty and Samoyed, on which basis we could then claim that there existed a Khanty-Samoyedic intermediary proto-language after which began the independent development of Samoyed.

Naturally, there is no point in such absurd competition between equally insufficient data sets. It should be obvious to all historical linguists that quantitative data must contain quantity (hundreds of words, not a handful), while qualitative data must contain quality (e.g., regular sound changes, not random irregularities). The minuscule set of basic numerals with random irregularities does not fulfill any criteria for a valid data set concerning the taxonomic structure of the Uralic language family, so its evidential value is zero.

Moreover, even if Samoyed was the first unit to diverge from Late Proto-Uralic, the deepest taxonomic rift alone cannot conclusively testify to the location of the Late Proto-Uralic homeland. There is no linguistic law preventing the deepest taxonomic rift from moving to a new region – just look at the Turkic language family, where the deepest rift is nowadays located in the Middle Volga Region between Chuvash and Tatar, even though the Late Proto-Turkic homeland is usually located in Mongolia.

It is true that we have linguistic evidence for early presence of at least one assumedly primary branch, Samoyed, in the Altai Region, but this alone is not enough to pull also Late Proto-Uralic there. We know that the closest relatives of Samoyed (Khanty and Mansi) have spread to their current regions from the southwest (from the Central Ural Region; see Chapter 3.2), so equating the Late Proto-Uralic homeland with the Samoyed homeland instead of the "Ob-Ugric" or even the assumed "Finno-Ugric" homeland is just an arbitrary decision lacking any supporting evidence (which is acknowledged also by Janhunen himself in 2009: 71–72). Such a decision cannot be deduced from the taxonomic structure of the language family even if Samoyed was the first unit to diverge from Late Proto-Uralic.

Consequently, the arguments by Janhunen cannot offer any valid results for locating Late Proto-Uralic.

2.2. Bjørn (2022)

"The period from the introduction of Afan[a]sievo to the latest possible date of Proto-Uralic at the onset of the Seima–Turbino phenomenon constitutes a 'Goldilocks zone', neither too early for the development of Indo-European derived Bronze Age terminology, neither too late for borrowing at the Proto-Uralic stage. This period stretches from around 3300 to 2100 BC." (Bjørn 2022: 19)

Bjørn (2022) presents a group of words with wide distribution in Eurasian language families, and he derives their origin from Indo-European, mediated into the more eastern language families (mainly Uralic, Turkic, and Sinitic) by the Tocharian language lineage ("language of Afanasievo"). I must give credit to Bjørn, because some of his reconstructed words truly become phonologically closer to the Late Proto-Uralic words through early Tocharian (his "language of Afanasievo") sound changes (about which see Hackstein 2017; Peyrot 2022), so I label the donor language here Early Proto-Tocharian (EPT):

1. PIE *h_3neh_3mn > EPT *ni(:)min (> TochA ñom, B ñem; Bjørn 2022: 11–12)
   → EPU **ni(i)mi > LPU *nimi 'name'

2. PIE *$ĝlh_3(wos)$ > EPT *keliwos (not attested in Tocharian; Bjørn 2022: 12–13)
   → EPU **ke/äliw(o) > LPU *käliw 'brother/sister-in-law'

Earlier the sound correspondences in these word comparisons between Proto-Indo-European and Late Proto-Uralic remained unexplained. As Bjørn writes in the quote above, the presence of the Tocharian language lineage in Southern Siberia seems to have begun already during the late 4th millennium BCE. Therefore, these loanwords do not necessarily concern Late Proto-Uralic – Early Proto-Uralic is also a suitable option. If more of this kind of Early Proto-Tocharian etymologies could be presented in the future, it would increase the probability for the Early Proto-Uralic homeland being situated in Southern Siberia. I will present two new proposals in Chapter 2.5.

However, not all the etymologies of Bjørn seem plausible: Proto-Uralic *ćäjć(ć)imä '7' is still way too different from PIE/EPT *septḿ > LPT *ṣäptä (Bjørn 2022: 9): There are no matching sounds beyond *ä ~ *e and *m ~ *ḿ. To solve this great difference, Bjørn reconstructs sound changes in several evolutionary stages preceding the Late Proto-Uralic form. However, by adding such intermediary stages it becomes too easy to bridge very different words to each other – the method is so effective that it easily produces a multitude of false positive results. With enough additional reconstruction steps for successive sound changes, any pair of words between any languages of the world could be made to look like cognates.

Nevertheless, if enough credible word comparisons with the same sound correspondences (no matter how phonetically distant the sounds are) can be presented, the situation changes. At the moment, there are no further examples for the sound changes reconstructed by Bjørn between the words meaning '7', so chance resemblance remains a valid option.

Even though Proto-Samoyed *wäsa 'iron' (Aikio 2015: 43), when looked as an isolated instance, could be borrowed from Tocharian (LPT *ẇäsa 'gold' < PIE *h₂ues-h₂ 'gold') as proposed by Bjørn (2022: 15), it probably reflects Common Uralic *wäćka – the only irregularity in this comparison concerns the second syllable vowel, as demonstrated by Aikio. Also semantically the Samoyed word agrees better with other Uralic branches (Grünthal et al. 2022: Supplement S2).

Tocharian does not appear a credible source for the word in other Uralic branches, like Proto-Finnic *vaski 'copper, bronze'. This word could come from Pre-Proto-Finnic *waski or *waćki, so it rather corresponds to similar Uralic words with original internal consonant cluster *ćk (which does not match with Tocharian): Proto-Saami *veaškē -(< PrePSa *wećka, *wäćka, or even *waćka) and Proto-Mordvin *uśkə (< PrePMd *oćki, *aćki, or irregularly *waćki). Therefore, *waćkV seems a possible original form for this western Uralic word, and even Proto-Mari *wåž could reflect this, although *k regularly would have been preserved in this cluster. If the Tocharian word is indeed of Indo-European origin, then we are possibly dealing with chance resemblance between Tocharian and Samoyed, although "partial borrowing" (contamination) is also a theoretical possibility.

In any case, the new more credible etymologies for Late Proto-Uralic *nimi 'name' and *käliw 'brother/sister-in-law' by Bjørn seem to point to contacts between earlier stages of Uralic and Tocharian. The Tocharian language lineage is generally connected to the

movement of Yamnaya-related people from the Pontic-Caspian steppe to Southern Siberia around the late 4th millennium BCE (Anthony 2007: 264). This does not contradict the Late Proto-Uralic homeland in the Central Ural Region centuries later, as there is nothing in these loanwords pointing exactly to the stage of Late Proto-Uralic.

Therefore, I can propose a timeframe for these contacts encompassing roughly the early 3rd millennium BCE. We can call the corresponding reconstruction stages Early Proto-Uralic (as in Häkkinen 2023b: 77) and Early Proto-Tocharian, because some branch-specific sound changes must have already occurred in the latter language before the contact. This scenario agrees nicely with the timeframe of Bjørn (2022: 19); more about the chronology of the contact in Chapter 2.6.

Bjørn proposes that Late Proto-Uralic could be associated with the Okunevo Culture, but this seems impossible considering all the valid linguistic arguments concerning the date and location of Late Proto-Uralic (Chapter 3). Apparently, the only basis for this association by Bjørn is that the Okunevo Culture appeared within the former region of the Afanasievo Culture, which is generally associated with the Tocharian language lineage. However, a suitable context for linguistic contacts alone is not a proof, because there can be several suitable contexts available, and linguistic contact can occur even within "unsuitable" (archaeologically less visible) contexts. Possibility does not equal necessity.

We must also remember that all ancient languages most probably had contact with several adjacent languages, but we can identify traces only from those languages which still have extant descendants or relatives. Because many ancient language families in Inner Eurasia must have disappeared under the expansions of Indo-European, Uralic, and many still later highly expansive language families, we might easily forget to acknowledge the large variety of ancient language families. As the linguistic results do not support the Southern Siberian homeland for Late Proto-Uralic, the Okunevo Culture must be associated with some other language lineage; perhaps Para-Uralic could be possible.

To conclude: the results by Bjørn (2022) do not contradict the Late Proto-Uralic homeland in the Central Ural Region, but they can illuminate the contacts between Early Proto-Uralic and Early Proto-Tocharian (see Chapter 2.6).

2.3. Peyrot (2019a) and Warries (2022)

"...the contacts must have taken place, in view of the linguistic evidence presented above, well before Proto-Samoyedic dissolved, at a relatively early Pre-Proto-Samoyedic stage. - - I will in the following assume that the contacts are to be dated to the Afanas'evo period, but I note here explicitly that this is at this point no more than a working hypothesis that is inspired by archaeological, not by linguistic arguments." (Peyrot 2019a: 107, 109)

These two articles consider possible contacts between Uralic/Samoyed and Tocharian; Warries (2022) builds upon the results of Peyrot (2019a). Peyrot found similarities between Tocharian and Samoyed in vowel system, consonant system, and certain inflectional categories. He only considers Uralic/Samoyed substrate in Tocharian, because in such a situation there can be cases where loanwords are few but influence on phonological or grammatical system can be strong (Thomason 2001). However, foreign influence could appear in all levels of language also in the cases of superstrate or adstrate, because language shifters can bring traits from their original language to the target language irrespective of which language was there first. Typically, we would expect to see many loanwords in the latter two cases, but there are also examples of massive loanword borrowing from a substrate language (Aikio 2004; Saarikivi 2006).

Acknowledging the movements of great distance for the Tocharian language lineage both before and after the Afanasievo phase, there could have been several intermediary steps during its advancement. In such a process there could also have been several successive substrate languages. After all, on the way from Southern Siberia to Tarim Basin there existed a variety of different environments (forest, steppe, mountains, desert) probably accompanied by a variety of different local communities speaking different languages.

The importance of loanwords is that they alone enable us to identify and confirm the languages in contact. Without relevant loanword layer, it is in theory possible that similar developments in phonological or grammatical systems (in typological level, without concrete material similarities) in two languages could be due to language-internal factors, general

typological tendencies, or pure chance. Therefore, the Early Proto-Tocharian loanwords into Uralic, as proposed by Bjørn (2022), contribute also to the assumed contact-induced traits considered by Peyrot (2019a) and Warries (2022).

Peyrot demonstrates the necessity for a Finno-Ugric phonological innovation concerning the diverse reflexes of Late Proto-Uralic *u in Samoyed in the same environment (more so in Peyrot 2019b: 192–195). However, these kinds of things always have two possible interpretations: either there was a merger in one branch, or there was a split in another branch. Peyrot ignores the option for a split in Samoyed due to the same phonological context for different reflexes, but an irregular development in a small group of words can occur also randomly, without any defined condition. Therefore, there is no conclusive evidence that the result was caused by a merger shared by all the Uralic branches except Samoyed.

Nevertheless, I believe Peyrot is right when explaining some of my earlier proposed East Uralic (Ugro-Samoyedic) sound changes as parallel independent developments in these branches (Peyrot 2019a: 108); especially the split of Late Proto-Uralic *ę into *i̮ in Khanty and Samoyed under different conditions (see now Zhivlov 2023: 149, 151). However, all the shared East Uralic sound changes cannot be explained as independent developments, even though nowadays it must be acknowledged that these changes indeed occurred between already diverged speech communities and not in the uniform East Uralic proto-language (*contra* Häkkinen 2007).

Especially the very specific chain of changes concerning Proto-Uralic sibilants, which is shared by Hungarian, Mansi, Khanty, and Samoyed, requires adjacency of these branches quite late, toward the mid-second millennium BCE in the Central Ural Region (Chapter 3.2.2). Counterarguments presented in order to cut Samoyed off from the adjacency of the three Ugric branches are not valid (see Häkkinen 2023a: 111–113 and Chapter 3.2.2.3).

Building upon Peyrot (2019a), Warries (2022) focuses on the vowel system of Tocharian, and he compares two slightly different stages of Pre-Proto-Tocharian and Pre-Proto-Samoyed. He acknowledges that the Samoyed vowel system of the earlier stage is still almost identical with Late Proto-Uralic. Relatively early date for the foreign influence in Tocharian is required at least concerning sound system, and while Peyrot tried to make the divergence of Samoyed from Late Proto-Uralic earlier, here again Early Proto-Uralic at the first half of the third

millennium BCE would be a better temporal match. Later contact with Early Proto-Samoyed is also a possibility (see Chapter 2.6).

He also points out that in the genetic data, although the steppe ancestry component associated with the people of the Afanasievo Culture was present in the Dzungarian Basin already during the early 3rd millennium BCE, in Tarim Basin (further to the south) this ancestry component was not yet present during the early 2nd millennium BCE. Even though language cannot be seen from DNA, in this case when we have no conclusive linguistic evidence about the time of arrival of the Tocharian language lineage in the Tarim Basin, the genetic results offer a possible hint. Thus, it is possible that Early Proto-Samoyed speakers still met Tocharian speakers, when they arrived in Southern Siberia around the mid-second millennium BCE. And in any case, Para-Tocharian dialects could have survived there for a long time, even if the lineage of Tocharian proper moved to the south.

There is room for different interpretations concerning both the absolute chronology of Tocharian changes and the stage of Uralic or Samoyed which might have caused these changes. This means that the contacts between Uralic/Samoyed and Tocharian are compatible with different homeland propositions for Late Proto-Uralic, so they cannot testify to or against any specific homeland. The Altai homeland for Late Proto-Uralic, as proposed by Vesakoski, Salmela, and Piezonka (2025), does not find support from the contacts chronologically so vague and open to different interpretations.

2.4. Evaluation of the evidence

Search for the Uralic homeland is full of different arguments and alleged pieces of evidence. However, sufficiency, chronological accuracy, and chronological relevance of these vary a lot. Only if a piece of evidence fulfills all these criteria can it be considered as a valid argument for locating the proto-language. The process of hierarchical assessment goes followingly:

1. Is the argument in general sufficient to testify to anything, or is the evidence insufficient (open to different interpretations)?
2. If it is sufficient, is the argument chronologically accurate or inaccurate?

3. If it is accurate, is the argument relevant or irrelevant for
the language stage in question?

4. If it is relevant, it is counted as a valid piece of evidence for locating the
language stage in question.

During the first step, we can exclude pieces of evidence which are insufficient: they cannot testify to anything, as it is easy to find similar pieces of evidence with contradicting testimony. In Häkkinen (2023b) several often seen homeland arguments were deemed as insufficient (although the evaluation process presented above was not explicitly used).

During the second step, we can exclude pieces of evidence which are sufficient but chronologically inaccurate (ambiguous): they cannot designate which reconstruction stage we are exactly dealing with. During the third step, we can exclude pieces of evidence which are chronologically accurate but irrelevant concerning the reconstruction stage in question: they can only testify to some other reconstruction stage of that language lineage.

The pieces of evidence which pass the process to the fourth step can be considered valid arguments. In this comment, I have assessed the evidential value (sufficiency, chronological accuracy, and relevance) of arguments allegedly concerning the Late Proto-Uralic homeland in the articles referred to by Vesakoski, Salmela, and Piezonka (2025) in order to find support for the Late Proto-Uralic homeland in the Altai Region. The results from this assessment are:

1. Arguments by Janhunen (2022 ← 2000) were shown to be chronologically inaccurate (Ural-Altai typology) or insufficient (taxonomic structure based on basic numerals only) concerning the location of Late Proto-Uralic.

2. Arguments by Bjørn (2022) are sufficient but chronologically inaccurate, so they do not contradict the Late Proto-Uralic homeland in the Central Ural Region; instead, they could illuminate thus far poorly known stage of Early Proto-Uralic.

3. Arguments by Peyrot (2019a) and Warries (2022) are uncertain concerning sufficiency, because definite foreign influence on a phonological system or grammar cannot be verified without an accompanying loanword layer. They are chronologically inaccurate, as they do not necessarily concern Late Proto-Uralic itself, because the absolute chronology of the Tocharian sound changes is unknown. Instead, they could be

explained by contacts between Tocharian language lineage and either Early Proto-Uralic, Para-Uralic, or early stages of Samoyed.

	Sufficient	Accurate	Relevant	Valid
Ural-Altaic typology	**?**	**–**	███	Ø
Taxonomic structure from basic numerals	**–**	███	███	Ø
Loanwords from Tocharian into Uralic	**+**	**–**	███	Ø
Systemic influence from Uralic to Tocharian	**?**	**–**	███	Ø

Table 1: Validity of the arguments concerning the homeland of Late Proto-Uralic. A piece of evidence is excluded after the first negative assessment.

Consequently, there exists no valid linguistic evidence requiring us to locate the Late Proto-Uralic homeland in Southern Siberia, while the valid argumentation presented in Häkkinen (2023b) requires us to locate it in the Central Ural Region, as will be demonstrated at length in Chapter 3.

2.5. Possible Early Proto-Tocharian loanwords in Early Proto-Uralic

Inspired by the findings of Bjørn, I have found two more possible Early Proto-Tocharian etymologies for established Late Proto-Uralic words.

1. PIE *pekʷ- 'cook' (Mallory & Adams 2006: 259)
   > EPT *pekʷ- (> LPT *päkʷ- > Tocharian AB päk- 'become ready for eating [i.e. ripen, be cooked]')
   → EPU **pe/äx⁽ʷ⁾i- > LPU *pe/äxi- 'boil, be cooked' (Aikio 2022: 20, 24: *peji-)

Metsäranta (2020: 150) suggests LPU *päji- as the original form, but the *ä-reconstruction also contains uncertainties. However, a parallel case for the substitution EPTo *e → EPU *ä is seen in *käliw 'brother/sister-in-law' (see Chapter 2.2).

Late Proto-Uralic *x has been proposed as a substitute for Proto-Indo-European laryngeals (Koivulehto 1991), but this substitution is not widely accepted (recently Simon 2020; Holopainen 2020b). Here I propose another source for Uralic *x: Pre-Proto-Tocharian *kʷ (phonetic argumentation follows below). In theory, any archaic Indo-European vernacular could be the donor language, but the next word with the same consonant substitution can only be explained from Early Proto-Tocharian.

Aikio (2022: 24) reconstructs PU *peji- to account for Erzya pije- (intr.), pid'e- (trans.): here Mordvin d' would come from earlier *jt (Aikio 2014c: 3: *pej-ta-). However, Hungarian fő- : fővő rather points to *x, because *j between vowels would have been preserved in Hungarian (cf. LPU *ćoji- > HU zaj, szaj 'noise, racket'; Aikio 2020: 132 | LPU *kaji > Hu haj 'hair' | LPU *waji > Hu vaj 'fat' | LPU *koja > Hu háj 'fat' **vs.** LPU *uji- > Hu ú- 'swim'; Zhivlov 2023: 160, 162, 166).

Moreover, the Komi cognate pu- points to *x, because *j would have been preserved (Metsäranta 2020: 90–91). Mansi and Samoyed are indifferent here; the transitive derivation in Samoyed mentioned by Aikio could come from *pexi- or *peji- (Zhivlov 2023: 150–151). There are no obstacles to assume that *x merged with *j already in Early Proto-Mordvin before the transitive derivative was formed, so we can reconstruct Late Proto-Uralic *pexi-, whence EPMd *peji- → *pej-ta- > LPMd *pid'ə-.

Why Early Proto-Tocharian *kʷ would have been substituted by Uralic *x instead of plain *k? A possible explanation is related to the phonetic realization of a labialized velar stop: tightly rounded lips produce blowing-like sound accompanying the velar occlusion. Perhaps this extra pronuncial feature, which was auditorily recognized as bilabial fricative, led the Uralic speakers to substitute *kʷ with velar fricative *x. Because there were neither stop *kʷ nor cluster *kw in Uralic, the phonetic parameters "velar + fricative" could have been combined in velar fricative *x. This would mean that the parameter "rounded" did not make it to TOP 2, unless Late Proto-Uralic *x developed from Early Proto-Uralic *xʷ or *kʷ. Unfortunately, this possibility seems impossible to verify or falsify.

It is interesting that Late Proto-Uralic *x had a very restricted environment: it could only appear between vowels and before the second syllable *i (Aikio 2022: 7–8), although Zhivlov also reconstructs occasions before a consonant (Zhivlov 2023: 119–120). Perhaps it was still a new and young sound, "invented" only in Early Proto-Uralic to substitute the Early Proto-Tocharian labialized velars?

For the next proposal, my Early Proto-Tocharian reconstruction is based on Kim (2000):

2. PIE *(h₁)eh₂gʷʰ- : *(h₁)ēh₂gʷʰ- 'drink' (Kim 2000; Mallory & Adams 2006: 256)

    > EPT *yākʷ- (> *yōkʷ- > LPT *yokʷ- > Tocharian AB yok- 'drink')

    → E/LPU *jexi-/*juxi- 'drink' (Aikio 2022: 23)

Zhivlov reconstructs only Proto-Uralic *jexi- (e.g. Zhivlov 2023: 161), which would be more suitable here. Even the Finnic and Saami cognates could be explained from LPU *e̩, assuming an exceptional development before velar obstruents (Zhivlov 2014: 116; Zhivlov 2023: 124, 127). Uralic close *u would be difficult to explain from an open Early Proto-Tocharian vowel.

Uralic *e̩ was used to substitute for Late Proto-Indo-Iranian *a (Holopainen 2019: 49–50), and the same substitution could have been possible for Early Proto-Tocharian loanwords, as there was only one open vowel *a but not yet non-open inner vowels like <ä> [ɨ] (Hackstein 2017: 1304). In practice, this would require a very early system of five vowel qualities (*a, *e, *i, *o, *u; see Warries 2022: 195), and perhaps even this would not be enough. After all, Late Proto-Indo-Iranian only had three vowel qualities (*a, *i, *u), which allowed wide enough realization for *a to be heard as like Uralic *e̩.

Despite all this uncertainty, it seems possible that Early Proto-Uralic *e̩ was a substitute for Early Proto-Tocharian *a. Figure 3 contains the vowel qualities reconstructed in Late Proto-Uralic in two-dimensional auditory space, modified from the Estonian vowels of Meister and Meister (2019: 133). It also shows the attested or assumed reflexes for Late Proto-Indo-Iranian and Pre-/Early Proto-Tocharian *a. The vowel system of the latter had two mid-close vowels *e and *o, which excluded the Uralic substitutions of *ä and *o for *a (*ä was apparently the substitute for Pre/EPT *e and *o for *o). There was enough room for realizations of PrePT *a to expand toward the middle of the vowel space where Uralic *e̩ (~ Estonian õ) was situated.
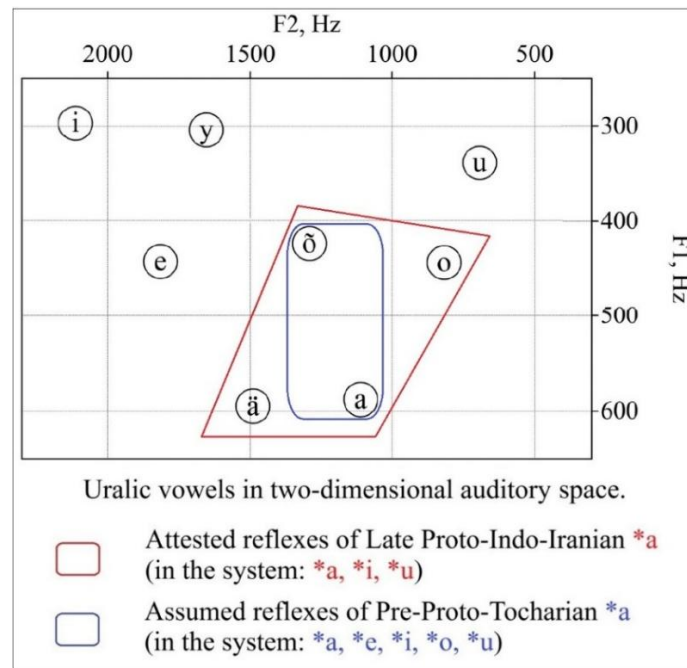
Figure 3: The reconstructed Late Proto-Uralic vowels modified from the Estonian vowels (Meister and Meister 2019) in two-dimensional auditory space compared to *a* in Late Proto-Indo-Iranian and Pre-Proto-Tocharian.

I believe that the sound changes of the Tocharian language lineage could provide more credible etymologies for those Uralic words which have earlier been compared to Proto-Indo-European words, but which are not phonologically satisfactory as such.

2.6. Contact between Early Proto-Tocharian and Early Proto-Uralic

"Similarly, the closely related Afanasievo culture, with its lack of evidence for agriculture, does not provide an evidently suitable context for the Tocharian homeland. - - For Tocharian, it may be necessary to assume an indirect dispersal as well in view of the late spread of agriculture to the eastern steppe. The wooden plows of the Catacomb culture (2500–1950 BCE) offer an archaeological terminus post quem." (Kroonen et al. 2022: 33, 35)

Recently Kroonen et al. (2022) have questioned the traditional association between the Tocharian language lineage and the Afanasievo Culture in Southern Siberia and nearby regions, because agriculture was not known within that culture and because Tocharian has preserved some agricultural vocabulary from Core Indo-European (excluding Anatolian).

However, their own data shows that the meanings of the Tocharian cognates are not strictly restricted to agriculture but can be associated also with other kinds of livelihood contexts, and some are open to different interpretations: 'grain, (sesame, lotus) seeds' (page 6) | 'stone, rock' (8) | 'end, tip (of grass)' (11) | 'head' (11) | 'dust; plow' (12) | 'chaff (of grain), husk' (16). There are different interpretations concerning the form and meaning of Tocharian B *āre*: it could have meant 'plow' or 'dust' (Peyrot 2018: 263), and if it was the latter, then the meaning 'plow' in Tocharian A could be secondary.

These few words still agree with the Afanasievo model: due to the lack of agriculture within that culture, strictly agricultural words would have been forgotten or shifted to new meanings in Tocharian. Therefore, the Tocharian language lineage can still be associated with the earliest Yamnaya-like cultural and genetic expansion to the eastern steppe: the Afanasievo Culture in Southern Siberia and nearby regions. This model is not without problems (Mallory 2015), but it seems to remain the most likely option due to the non-conclusiveness of the agricultural vocabulary argumentation.

New, more critical and precise AMS radiocarbon dates for the Afanasievo Culture set it between 3100–2800 BCE in the Altai Region and 2900–2400 BCE in the Minusinsk Basin along Yenisei. In the former region it was replaced by the Elunino Culture around 2700 BCE and in the latter region by the Okunevo Culture around 2500 BCE (Polyakov, Svyatko & Stepanova 2019). The duration of the Afanasievo Culture therefore agrees with the period of the assumed contact between Early Proto-Tocharian and Early Proto-Uralic, afore tentatively proposed to cover the first half of the 3rd millennium BCE.

However, the proper Uralic language lineage did not remain in Southern Siberia but was relocated in the Central Ural Region (see Chapter 3). Assuming there occurred no complete language replacement during the following centuries, one of the languages in Southern Siberia during the genesis of the Seima-Turbino Network around 2200 BCE could have been Para-Uralic (another descendant of Early Proto-Uralic = an extinct sister lineage of Late Proto-Uralic).

This could also explain both the exact destination and the success of the Samoyed language after leaping over a great distance (the whole wide Western Siberia) and far from its close relatives, from the Central Ural Region directly to the Altai-Sayan Region: the local, still rather closely related Para-Uralic vernaculars could have been updated according to the new-comer dialect, which is an easier process than shifting to a totally foreign language. Still, there must have been also other languages present in the region, because Samoyed has replaced a great portion of its Uralic vocabulary with words from unknown languages.

As the presence of the Tocharian language lineage in or near Southern Siberia was so early and lasted so long, work like that done by Bjørn, Peyrot, and Warries has a potential to in-crease our knowledge about Early Proto-Uralic, Para-Uralic, and/or Samoyed in that region. These newly added reconstruction stages give scholars more room to interpret contacts, and henceforth we all can hopefully be clearer about whether we mean the actual Late Proto-Uralic or some other rather similar language or stage.
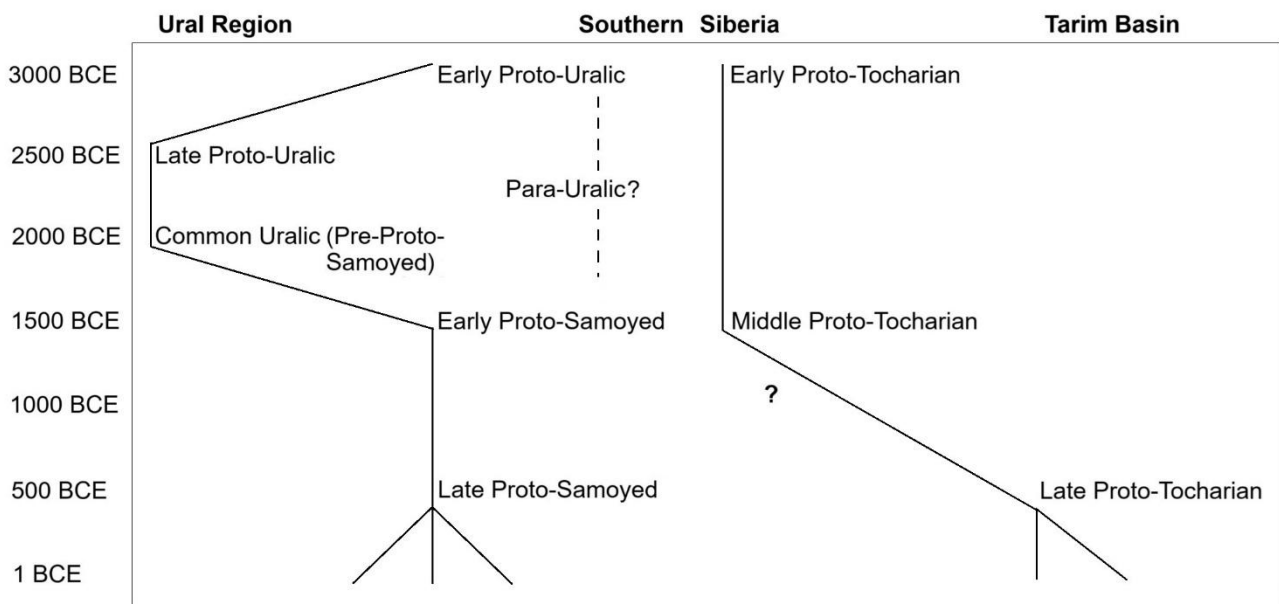
Figure 4: A rough spatiotemporal sketch about possible contact stages between Uralic and Tocharian in Southern Siberia.

It has been assessed by specialists that Late Proto-Tocharian would probably be at least a millennium earlier than the earliest literary attestations at the 5th century CE (Widmer 2017: 1389). Therefore, it would be roughly contemporaneous with Late Proto-Samoyed, which is dated around 500 BCE (Zhivlov 2023: 14). It is uncertain if Late Proto-Tocharian was already spoken in so southern a region that contacts with Late Proto-Samoyed were no longer maintained. However, there are some proposed loanwords from Tocharian to Samoyed at a rather late date (Kallio 2004; Pystynen 2022: 18). Naturally, these could also be borrowed from the Para-Tocharian dialects remaining in Southern Siberia, so they cannot tie the proper Tocharian language lineage in Southern Siberia at so late a stage.

The scenario presented in Figure 4 acknowledges and agrees with all the sufficient linguistic results and therefore should be taken as the starting point when trying to find archaeological or genetic matches for the linguistic results (see Chapter 5). Irrespective of how intriguing the Seima-Turbino Network and the Flowerpot Complex are, they do not match the linguistic results concerning Late Proto-Uralic or Common Uralic (Chapter 3). However, we can assume that during the late 3rd millennium BCE, at least Middle Proto-Tocharian and Para-Uralic could have been spoken in Southern Siberia (among other languages; see Chapter 4).

3. Evidence supporting the Late Proto-Uralic homeland in the Central Ural Region

For the sake of objectivity, I must assess also the arguments presented in support of the Central Ural homeland with the same measure as the arguments above. These arguments are mostly from Häkkinen (2023b), but here I expand many of their details further and respond to some recent interpretations.

3.1. Evidence from early Indo-Iranian loanword layers

It is known that Samoyed shares surprisingly few early Indo-Iranian loanwords with the other Uralic branches (Holopainen 2019: 344; Grünthal et al. 2022: 17). Based on Holopainen (2019: 336–337), Grünthal et al. (2022) accept 14 Pre-Indo-Iranian loanwords, 17 Proto-Indo-

Iranian loanwords, and 55 ambiguous early loanwords (could be Proto-Indo-Iranian or Proto-Iranian). However, they did not compare the coverage of different loanword layers in the Uralic branches, so I will do it here. I accept 11 loanwords representing earlier stages than Late Proto-Indo-Iranian and 29 loanwords representing Late Proto-Indo-Iranian (see the criteria and the list below).

Instead of the traditional division into two distinct layers (e.g. Holopainen 2019: Pre-Indo-Iranian *vs.* Proto-Indo-Iranian), I have recently divided these loanwords into three groups based on the relative chronology of the Indo-Iranian sound changes observed in the Uralic words (Häkkinen 2023b: 56). This is because there occurred many sound changes during a rather long period before the vowel merger which eventually marked the beginning of the Late Proto-Indo-Iranian stage (Lubotsky 2018). Because this stage only lasted for a couple of centuries before the language evolved to Proto-Iranian and Proto-Indic, a more balanced chronological periodization requires more resolution also in the preceding continuum of Indo-Iranian evolution.

However, there is only one loanword in Uralic containing two diagnostic consonants (*kekra*: the initial *k* not yet palatalized, and *r* from earlier *l*), so the layering is based on the relative chronology of the Indo-Iranian sound changes. Consequently, it is difficult to precisely point the layer of individual loanwords, so from now on I will consider the two earlier layers together as Early/Middle Proto-Indo-Iranian loanwords.

Here only Uralic reconstructions are given, followed by the relevant page in Holopainen (2019), possibly accompanied by a more recent reference:

1. Early Proto-Indo-Iranian loanwords (after the inteconsonantal *h > *i, *o > *ō in open syllables, and *l > *r): **A)** *kečra* 'spindle' (116), *kekra* 'ring-like, curved' (118) ‖ **B)** *meti* 'honey, nectar' (146), *pejma* 'milk' (178; Holopainen 2023).
2. Middle Proto-Indo-Iranian loanwords (after the satemization, the Ruki rule, and the palatalization of the velar stops before *e): **A)** *mekši* 'bee' (139), *šera-* 'wake up' (258) ‖ **B)** *erta* 'side' (81), *pejjiš* 'funeral feast' (173 *peijas*; Holopainen 2020a: 640), ?*rećma* 'rope' (207), *tejniš* 'pregnant (of animals)' (276), *wersa* 'sprout' (307 *verso*; Holopainen 2020a: 625).

3. Late Proto-Indo-Iranian loanwords (after the merger *e, *o, *m̥, *n̥ > *a): **A)** *aćnas ‘greedy (for food)’ (60; Holopainen 2020a: 627), *ońća ‘meat, share’ (170), *poći ‘penis’ (185; Häkkinen 2023b: 54), *ćoma- ‘worry, hunger’ (213), *ćarwi ‘horn’ (220), *ć/šarapa (Holopainen & Aikio 2023: 30), *ćaδa- ‘rain’ (224), *ćaxa ‘goat’ (240), *ć/šęta ‘100’ (242), *ć/šasra ‘1000’ (244), *ćišta ‘beeswax’ (249), *ćuka ‘husk of grain’ (254), *waćara ‘hammer, axe’ (292), *woraćV ‘boar’ (313), *ćara- ‘shit’ (324), *moćki- ‘wash’ (Häkkinen 2023b: 55), *ćęra ‘light color’ (Häkkinen 2023b: 59), *ćämi- ‘drink, gulp’ (Aikio 2020: 114) ‖ **B)** *asora ‘lord’ (70), *kęntaw ‘tree stump’ (120), *kota ‘chum, hut’ (126), *ora ‘awl’ (163), *orpa ‘orphan’ (166), *pęŋka ‘psychedelic mushroom’ (186), *pätari- ‘fly’ (198), *repa ‘fox’ (201), ?*sejti ‘bridge’ (226; Holopainen 2023), *sa/oŋka ‘old’ (235), *tora- ‘fight’ (282, 326), *mura ‘song’ (323), *ä(j)sVn-weć(k)V ‘tin/lead’ (Häkkinen 2023a).

The fourth early Indo-Iranian loanword layer would be Proto-Iranian, but it is not included here. As it can be distinguished from the Late Proto-Indo-Iranian layer only in rare favorable situations (words showing LPIIr *tś, *dź **vs.** > PIr *ts, *dz) and as the loanwords distinguishing these affricates (U *ć **vs.** internal *č ~ initial *š; Holopainen 2019: 334) show a clear tendency concerning the distribution of the words in Uralic (wider **vs.** narrower), I have applied the following distribution criteria.

Words in the A-groups show diagnostic consonants matching the respective chronological stage or at least preceding the following stage, while the B-groups lack the consonant criteria. Because (1) there are two layers of *e-words (Early and Middle Proto-Indo-Iranian) and two layers of other words (Late Proto-Indo-Iranian and Proto-Iranian) and (2) it would be realistic to divide these words quite evenly between the layers, I have applied the distribution criterion for the words lacking the consonant criteria.

Those *e-words which have a cognate in at least one European (Finnic, Saami, Mordvin, Mari, Permic) and one Siberian branch (Hungarian, Mansi, Khanty, and Samoyed) are put in the first layer and the rest in the second layer; and likewise the later loanwords which have a cognate in at least one European and one Siberian branch are put in the third layer and the rest in the fourth layer (Proto-Iranian; not listed here). I acknowledge that this practice leads to a situation where the words in layer 3B have wider distribution than the words in layer 2B, but

as the regional divergence of Late Proto-Uralic seems to have begun already before the first layer, this has no practical consequences. This also seems to agree with the spatiotemporal model for the Indo-Iranian contact (Chapter 3.1.3).

Concerning the consonant criteria, the later words with *ć and with only European distribution are included in the third layer, but the words assumedly descending from *ć but having only Siberian distribution are left outside, because the primary outcome of *ć in all the Siberian branches is *s, and therefore such words cannot be reliable distinguished from younger Iranian loanwords due to the changes LPIIr *tś, *dź > PIr *ts, *dz > Ir *s, *z. This asymmetry caused by uncertainty concerning the original sound could lead to a slightly higher number of loanwords concentrating in the European branches.

There is so far only one assumed example of the Ruki rule (*mekši 'bee') and one example of palatalization of the velar stops before the vowel merger (*šera- 'wake up') in the *e-loanwords (U *še ← MPIIr *tše, *dže < EPIIr *ke, *ge). However, examples for these changes are scarce also in the Late Proto-Indo-Iranian layer containing a lot more words: only *ćišta 'beeswax' seems to show the Ruki change. There are only very few words with initial EPIIr *ke, *ge or MPIIr tše, *dže or LPIIr *tša, *dža in all these three loanword layers combined, so this scarcity is rather a statistical character of the data than a qualitative lack. It is also possible that the temporal gap between the palatalization of the velar stops and the vowel merger was very short and therefore we lack further examples from the intermediary stage.

3.1.1. Accepted Early/Middle Proto-Indo-Iranian etymologies

Here I must shortly explain the differences between the sets of *e-loanwords accepted by Grünthal et al. (2022) and by myself here and now (14 *vs.* 11). I apply the principle that if a loanword can equally well be later, there is no reason to assume it is earlier.

First, I have removed six words from the list of Grünthal et al. (2022). I moved *ońća 'meat, share' and *orpa 'orphan' to the Late Proto-Indo-Iranian layer, because they can represent the well-known substitution *a → *o. I also moved *repa 'fox' and *sejti 'bridge' to the Late Proto-Indo-Iranian layer, because they can represent the original *a just as well as earlier *o (there are no traces of *e in the donor language, see Palmér et al. 2021: 245–246;

see below for *sejti*). I removed *perta* 'wing', because the Permic word can now be explained as a later Iranian borrowing (Metsäranta 2020: 33). Lastly, I removed *peji-* 'to milk a cow' (> Hungarian *fej*), an assumed parallel borrowing for *pejma* 'milk' which can be a later Iranian borrowing, because Hungarian *e* can come from earlier *ä* as noted by Holopainen (2020a: 622).

Second, I added three words: *tejniš* 'pregnant (of animal)', *wersa* 'sprout', and *šera-* 'wake up', which all have cognates in Finnic (Late Proto-Finnic *tiineh*, *verso*, *herä-*), the last two exclusively. For *tejniš* 'pregnant (of animal)', Holopainen (2020a: 624) has recently stated that the Baltic and Indo-Iranian etymologies are equally possible. However, in Baltic loanwords *ei* seems to be reflected regularly by Late Proto-Finnic *ei*: of four possible loanwords I could find, three show *ei* and one shows *ii* (YSUS: *leinä*, *seipi*, *veikeä* **vs.** *liika*). As the Baltic loanwords seem to be later than the Early Proto-Finnic sound change *ej* > *ij* (later > *ii*), the Early/Middle Proto-Indo-Iranian etymology seems acceptable for *tejniš*. This is also supported by the regular cognate in Mari (Aikio 2014b: 90), because Finnic and Mari do not share Baltic loanwords with regular (Late Proto-Uralic-like) sound correspondences.

I have accepted Late Proto-Finnic *verso* 'sprout', recently defended by Holopainen (2020a: 625). Even though the variation North Finnic *e-a* ~ South Finnic *ẹ-a* normally goes back to Late Proto-Finnic *ẹ-a*, it seems that *o* in the second syllable (having no front pair *ö* until North Finnic) has triggered the regressive palatal assimilation *e-o* > *ẹ-o* in Southern Finnic (Pystynen 2019). Therefore, we can reconstruct Early Proto-Finnic front vowel combination *verso* or *versä* (if *o* is a later derivative), which makes the early borrowing possible. I have reconstructed *wersa* in Pre-Proto-Finnic before the appearance of the phonological vowel harmony.

I have added *šera-* 'to wake up' (> Late Proto-Finnic *herä-*), because even though Holopainen (2020a: 620) states that instead of *š*, Uralic *č* would be the assumed substitute for Indo-Iranian *tš* and *dž*, there are no examples of word-initial *č* in the convincing Indo-Iranian loanwords and no further examples of substituting Indo-Iranian *tš* or *dž*.

Moreover, the regular substitute for Proto- and Archaic Iranian *ts* and *dz* is Uralic *š* word-initially and *č* word-internally (Holopainen 2019: 334), and the same pattern seems probable for Indo-Iranian *tš* and *dž*, as well. We have two clear and two possible examples

for the substitution of Proto- or Archaic Iranian *ts and *dz → Uralic *š: *šukta 'burned patch in swidden agriculture' (Holopainen 2019: 264) | *ša/oji- 'take care' (262) | *šatas 'sprout, germ (of seed)' (257; could be also Germanic) | *š/serńa 'gold' (232). After all, Uralic *š seems to be phonetically even more fitting substitute for the latter affricates than for the plain affricates.

A borrowing from archaic Indo-European *ser- 'to keep an eye on, to protect' as proposed by Holopainen is less likely, because we have no evidence of the substitution IE *s → U *š outside the Germanic loanwords.

Third, I have retained eight words in the list of Grünthal: *meti, *mekši, *kekra, *erta, *pejma, *rećma, *pejjiš, and *kečra, but I will comment only the last three. I have earlier rejected *rećma 'rope' (Häkkinen 2022: 138), because Finnic *rihma (< *rićma, *rišma, or *rïsma) is more likely a Balto-Slavic borrowing and because Saami *reaćmē (> South Saami rietjmie ~ North Saami reašmi) shows a secondary affricate. If the word was an early Indo-Iranian borrowing, it should have developed followingly: EPSa **rećma > LPSa **reašmē > South Saami **riesjmie ~ North Saami **reaihmi (cf. EPSa *ćäćna 'woodpecker' > LPSa *ćāšnē | EPSa *ćäćki 'mosquito' > LPSa *ćuoškę | EPSa *kićko- 'tug, drag' > LPSa *kęškō- | EPSa *koćki 'rapids' > LPSa *kuoškę | EPSa *vVćka 'copper/bronze' > LPSa *veaškē; Lehtiranta 2015: 22, 30, 42, 62, 146).

Junttila (2018: 83) has proposed an Old Latgallian etymology for Mordvin *rišmä, but he presents no further examples for the assumed substitution B *s → LPMd *ś. He points out that in Mordvin *sm shows unpalatalized sibilant next to a back vowel and palatalized *śm next to a front vowel. However, in the words containing Late Proto-Uralic *ć, Mordvin shows consistently *ś in all other environments (word-initially, between vowels, and before and after all consonants even after back vowels; Zhivlov 2023: 158–171; Aikio 2020; Aikio 2015: 51–66), and respectively LPU *s shows Mordvin *s in all other environments (but later often > ś before a front vowel in Mokša).

Because *ć and *s in general remained distinctive in Late Proto-Mordvin and because consonant clusters of the type *Sm appeared only later in individual Uralic branches, the merger of these sibilants before *m must be quite recent in Mordvin. Due to first the merger (*śm > *sm) and then the secondary conditioned split of the sibilants before *m (*V̈sm > *V̈śm),

Mordvin cannot testify to the original quality of the sibilant in this context. This means that the substitution could have been *sm, if the word was borrowed from Old Latgallian.

Even though the Old Latgallian etymology seems equally plausible, it cannot explain Saami *reaćmē. Therefore, coveragewise the Early/Middle Proto-Indo-Iranian etymology is stronger, because the Saami and the Mordvin words can be regularly derived from the common proto-form *rećma, if we accept the irregularity concerning the affricate in Saami. No other etymology can explain the affricate, either. Therefore, the E/MPIIr etymology *Hrećmi (Holopainen 2019: 207) appears valid.

I accepted Pre-Proto-Finnic *pejjiš 'feast', because there are two possible reconstructions. If the North Finnic *peijas is regular, this would come from Late Proto-Finnic *pẹijas, which is not a match for the Early/Middle Proto-Indo-Iranian original, because the back-vowel combination *ẹ-a only appeared in Middle Proto-Finnic (Häkkinen 2019) – around 2000 years too late. However, if we base the reconstruction on Estonian and Livonian, Late Proto-Finnic *peijeh is possible, and this would agree better with the Early/Middle Proto-Indo-Iranian original. In that case North Finnic *peijas would reflect some secondary development. Then we would have Pre-Proto-Finnic *pejjiš, and in the Uralic vocabulary *jj has at least one parallel: *äjja 'old man' (Aikio 2020: 33).

A few more words about the age of the Finnic back vowel combinations. Holopainen (2020a) accepts Iranian loanwords showing the Late Proto-Finnic back vowel combination *i-a (opposed to the front vowel combination *i-ä). This goes back to earlier *i̮-a (Kallio 2012a: 171–172), which now can be considered as a direct representative of Late Proto-Uralic *i̮-a as explained earlier (see Chapter 1.2). Within this framework, there can be very old loanwords showing the Late Proto-Finnic back vowel combination *i-a.

Another question then is, why words like Iranian *(H)iš- → Pre-Proto-Finnic *i̮ša 'life force, joy' (Holopainen 2020a: 631) would have been borrowed with *i̮ instead of *i? In any case, the cognates in Mansi and Khanty also point to a back vowel here (Holopainen and Aikio 2023: 11). The explanation could hide in the fact that phonetically *š [ʃ] is very similar to the voiceless retroflex sibilant [ʂ] which makes the following vowel more back by lowering its F2 (Kokkelmans 2021: 29). However, in Late Proto-Uralic we have *šiŋiri 'mouse', so we cannot assume that *ši and *iš were impossible combinations in Uralic.

Instead, we could assume that the phonetic quality of Indo-Iranian *i* was already more back next to *š* than elsewhere, thus triggering the substitution by Uralic back *i̮*. After all, the Uralic vowel space now with 9 distinct vowel qualities against the Late Proto-Indo-Iranian vowel space with only 3 distinct vowel qualities easily leads to a division of each Indo-Iranian vowel into sectors of several Uralic vowels.

The best example is the LPIIr *a*, substituted arbitrarily by Uralic *a*, *o*, *e̮*, or *ä* (Holopainen 2019: 49–50). A division of LPIIr *i* into the auditory realms of Uralic *i* and *i̮* (next to *š*) seems likewise possible, and there is also another example of this substitution: Finnic *ihta* < *i̮šta* 'lust, eagerness' ← LPIIr/PIr *Hišta-* (Holopainen 2019: 95). There are no counterexamples showing the substitution IIr *iš* → U *iš* (with the front *i*).

3.1.2. Suspicions concerning Early/Middle Proto-Indo-Iranian loanwords

Recently there have appeared alternative explanations for some *e*-loanwords in Uralic (Holopainen 2023). For example, the substitution LPIIr *ajC* → U *ejC* was proposed for *pejma* 'milk', *tejniš* 'pregnant (of animal)', and *sejti* 'bridge'. Considering the wide phonetic space for LPIIr *a* (see Chapter 1.1), it is indeed possible that its realizations could have differed according to adjacent sounds.

However, there are also many counterexamples where Indo-Iranian *ajC* was substituted normally with Uralic *ajC* or *ojC*: *ajša* 'shaft' (Holopainen 2019: 62), *tajwas* 'sky' (270), *tojwa-* 'wish' (281), *wa/ojna-* 'see' (296, 312). As all these four words could represent already the Late Proto-Indo-Iranian stage (no diagnostic consonants available), it still seems most likely that the *ejC*-words *pejmä* and *tejniš* represent an earlier layer with *e* in the donor language.

Yet Holopainen (2023: 4) is right when arguing for *sejti* that the coloring effect of *h₂* would have changed the vowel (*e* or *o*; see Derksen 2015: 397–398) to *a* already at a very early stage. But because the laryngeal coloring was a very early change in the Indo-Iranian chronology (Häkkinen 2023b: 56 and references therein), *sejti* could still belong to the Early Proto-Indo-Iranian layer even if it stemmed from this early *a*. The only Indo-Iranian *a* which could testify to the young age of a loanword comes from earlier *e*, *o*, *m̥*, or *n̥*.

Nevertheless, if *sejti* must in any case be derived from the earlier *a*, it could just as well be derived from the later *a*, so it is here grouped within the Late Proto-Indo-Iranian layer. On the other hand, it is also possible that there could be loanwords borrowed from the ancestral stage of Indo-Iranian before the laryngeal coloring.

Holopainen (2023: 4) suggests the possibility for parallel borrowing into Mordvin and Permic, but the cognates LPMd *sed'* 'bridge, floor' ~ LPPe *sǫjt* > Komi *sojt*, *sojd* 'bridge' (? ~ *sod* 'ladder, stairs') come entirely regularly from *sejti*: in Mordvin *jt* > *d'* (Aikio 2014c: 3) and *e-i* > *e-* (Zhivlov 2023: 168), and in Permic *e-i* > *ǫ-* > Komi *o* ~ Jaźva Komi *ú* (Metsäranta 2020: 325–326; no known Jaźva cognate for this word). Unlike Holopainen (2019: 226) writes, there is no *ä* but *e* in Mokša; *ä* is seen only in some Erzya dialects (Paasonen 1990–1999: *sẹd́*). Therefore, Late Proto-Mordvin indeed had *e*, not *ä* in this word (Bartens 1999: 55–56). As there are two different Komi forms with two different meanings, we can focus on the phonologically and semantically best match: *sojt*, *sojd* 'bridge'.

There are two cases of recurring irregularity among the early Indo-European loanwords. The first concerns Uralic *e* from E/MPIIr *o* or LPIIr *a*: *sejti* 'bridge', *repa* 'fox' (Palmér et al. 2021). Note that the latter word is not of the *ejC*-type, so that explanation would not necessarily apply to *sejti*, either (as was already concluded above). As these etymologies seem otherwise plausible and there appear no shared phonetic condition for this vowel representation, it is also possible that we are witnessing traces of some Para- or extinct Indo-Iranian vernacular with a peculiar vowel change *o*/*a* > *e*.

The second occasion of recurring irregularity concerns the irregular Uralic *ć* from Indo-Iranian *s*: *poći* 'penis', *moćki-* 'wash' (Häkkinen 2023b: 54–55). I have earlier proposed hypercorrection as the explanation here, because Uralic *ć* is so common sound in the Indo-Iranian loanwords (in 16/31 = 52 % of the assumed Late Proto-Indo-Iranian loanwords listed above, excluding these two words) that it could have formed a strong mental association with that donor language, resulting in *ć* even in some words where *s* would have been the auditorily correct substitute. Another possibility is that we are dealing with traces from some Para- or extinct Indo-Iranian language – but clearly not the same as in the previous case, judging from the first syllable *o* here instead of *e*.

Whatever is the ultimate reason for these recurring irregularities, the words *sejti*, *repa*, *poći*, and *moćki-* still seem to belong among the early Indo-Iranian loanwords. In the above

listing all these words are included in the Late Proto-Indo-European layer due to the ambiguity of the vowel reflexes.

Another path for questioning some loanwords in the earlier Indo-Iranian layers is based on the fact that there is no convincing Indo-European etymology for all the Indo-Iranian words, and therefore it is not possible to confirm that in these words Late Proto-Indo-Iranian *a comes from earlier *e (Holopainen 2023: *erta 'side', *rećma 'rope', *kečra 'spindle', *mekši 'bee'). Nevertheless, as such non-Indo-European words must then represent (substrate) borrowings from speakers assimilated by Indo-Iranians in the easternmost Europe, these Uralic *e-words can be taken as independent evidence supporting the earlier *e in these words before it changed to *a in Late Proto-Indo-Iranian. It is more plausible to fit the vowel correspondence IIr *a ~ U *e within the established Indo-Iranian sound change *e > *a than to invent some new reason for this vowel correspondence in these apparently non-Indo-European words.

The assumed Ruki rule in *mekši shows that regardless of its ultimate origin, it seems to testify to contact between Indo-Iranian and Uralic. Even if the other three words were borrowed into Uralic parallelly from an unknown language and not through Indo-Iranian, this kind of contact triangle showing mutual borrowings (1) from the unknown language to Indo-Iranian, (2) from the unknown language to Uralic, and (3) from Indo-Iranian to Uralic (most of the *e-borrowings) still requires the presence of all these languages adjacent to each other.

Holopainen (2023: 1) also questions Uralic *kekra 'ring-like, curved' (in Finnic and Saami only) being borrowed from Early/Middle Proto-Indo-Iranian *kekro 'wheel' because "the semantic connection is not perfect." Nevertheless, the meaning in Saami (also reconstructed as the Uralic meaning above) comes very close, even if the meaning in Finnic ('feast at the end of the harvest season') represents more abstract circle. After all, the Finnic *kekri : *kekrin (sg.gen) must be a derivation (< *kekrij < *kekrəj < *kekrä-j; cf. Kallio 2012b: 35), which usually changes the original meaning.

Acknowledging the quality (the phonological and semantic matching) and quantity of the *e-loanwords, we cannot reject the Early/Middle Proto-Indo-Iranian loanwords as mere chance resemblances, and we cannot plausibly explain them all as Late Proto-Indo-Iranian or yet later loanwords. They form a sufficient data set, and even though all the words do not

contain diagnostic consonants, many of them are showing the results of early Indo-Iranian sound changes and/or the lack of later Indo-Iranian sound changes.

3.1.3. The distribution of the early Indo-Iranian loanwords

Early/Middle Proto-Indo-Iranian: 11									
	Finnic	**Saami**	**Mordv.**	**Mari**	**Permic**	**Hung.**	**Mansi**	**Khanty**	**Samoy.**
*erta		+	+	+	+				
*meti	+		+		+	+			
*mekši	+		+		+	+			
*kečra	+	+	+						
*pejma	+		+			+			
*kekra	+	+							
*rećma	+		+						
*tejniš	+			+					
*šera-	+								
*wersa	+								
*pejjiš	+								
11	**10**	**3**	**6**	**2**	**3**	**3**	**0**	**0**	**0**

Table 2: Early/Middle Proto-Indo-Iranian loanwords arranged by their coverage in Uralic.

Table 2 shows the distribution of Early/Middle Proto-Indo-Iranian loanwords in the Uralic branches, arranged by their coverage from the widest to the narrowest. Table 3 shows the Late Proto-Indo-Iranian loanwords. Average number of branches per word is 2.5 for the earlier loanwords (27 / 11) and 3.3 for the later loanwords (102 / 31). Average number of words per branch is 3.0 for the earlier loanwords (27 / 9) and 11.3 for the later loanwords (102 / 9). There are seven later loanwords with greater coverage (5 or more branches) than any of the earlier loanwords (4 branches at the most). For Late Proto-Indo-Iranian *a, the Uralic reflexes in the frequency order are: *a = 8.5, *o = 8.5, *ę = 4, *ä = 3, *e = 2 (5 words had no LPIIr *a: *ćarwi, *tora-, *ćuka, *ćišta, *mura).

The earlier loanwords show surprisingly western emphasis in their distribution, totally lacking cognates in the three easternmost branches (see Chapter 3.7), and the top branches being Finnic and Mordvin. The later loanwords show more even distribution yet still decreasing toward the east, although here Mordvin forms the sole peak. This kind of narrower

distribution has sometimes been seen as suspicious, because the basic principle has been that the older the layer, the wider the distribution should be.

Late Proto-Indo-Iranian: 31									
	Finnic	Saami	Mordv.	Mari	Permic	Hung.	Mansi	Khanty	Samoy.
*ćęta	+	+	+	+	+	+	+	+	
*kota	+	+	+	+	+	+		+	
*moćki-	+		+	+	+	+			+
*ćarwi	+	+	+	+	+	+			
*pęŋka			+	+			+	+	+
*kęntaw	+	+	+				+	+	
*ońća	+	+		+		+	+		
*ćoma-			+	+	+	+			
*orpa	+	+	+					+	
*ora	+	+	+			+			
*repa	+		+	+	+				
*asira			+		+		+		
*ćišta			+	+	+				
*ćuka			+	+	+				
*tora-	+	+							+
*waćara	+	+	+						
*mura			+	+				+	
*ćara-			+	+		+			
*ćasra					+		+		
*woraćV	+		+						
*ćaδa-	+								+
*pätari		+					+		
*poći		+				+			
*sa/oŋka				+		+			
*ä(j)sVn-weć(k)V					+		+		
*ćęra		+							+
*sejti			+		+				
*ćarapa							+		
*ćämi	+		+						
*ćaxa			+						
*aćnas	+								
31	15	12	20	13	12	10	9	6	5
	Finnic	Saami	Mordv.	Mari	Permic	Hung.	Mansi	Khanty	Samoy.

Table 3: Late Proto-Indo-Iranian loanwords arranged by their coverage in Uralic.

However, the spatiotemporal model presented in Häkkinen (2023b: 73–75) plausibly explains this narrower western-central distribution in the earlier loanword layer. The explanation is connected to the advancement of the Indo-Iranians from the west to the east: the successive stages could be connected to the Fatyanovo-Balanovo Culture (Palmér 2025) in the Volga-Kama Region > to the more southern Abashevo Culture (Parpola 2020: 190) in the Volga-Ural Region > to the Sintashta Culture (Lubotsky 2023; Epimakhov and Lubotsky 2023) in the Southern Trans-Urals. Only this geographical expansion/shift toward the east would have led the Indo-Iranians in contact also with the eastern Uralic speech communities, when earlier they were in contact only with the western and central Uralic speech communities.

Furthermore, there are also several other possible reasons for the uneven distribution of the loanwords in the Uralic branches. Possible factors are:

1. Different branches have preserved different amounts of inherited words. Within the Uralic language family, Hungarian and Samoyed are known to be the branches with the smallest retention percentages. As the early Indo-Iranian loanword layers are nearly as old as the Late Proto-Uralic words, this same retention profile must affect them, as well: later lexical replacement cannot discern "original" words from loanwords.

2. The older the loanword, the more there has been time for it to become replaced by younger words. Therefore, older loanwords can have narrower (or wide but patchier) distribution than younger loanwords.

3. The more basic the meaning and the more frequent the usage of the word, the more resistant the word is against being replaced (Pagel et al. 2007). Perhaps the words in the earlier layers were less basic, if the contact was less intensive at the early stage.

4. Some branches have moved a long distance from the original contact region, being subjected to cumulative influence from foreign languages, which often leads to strong lexical replacement. This factor should affect most the lateral branches Finnic, Saami, Khanty, and Samoyed.

5. Some branches have moved to regions where agriculture has not been favored, which naturally leads to the disappearance of such words. For example, Mordvin is the southernmost Uralic language (excluding Hungarian), and it has preserved old Indo-Iranian agricultural loanwords with otherwise narrow distribution: *ćaxa 'goat' (only in

Mordvin) | *woraćV 'boar' (in Mordvin and Finnic) | *ćišta 'beeswax' (in Mordvin, Mari, and Permic).

To conclude: there is nothing suspicious in the result that the Early/Middle Proto-Indo-Iranian loanwords often have lower coverage and narrower distribution in Uralic than the Late Proto-Indo-Iranian loanwords. This can be seen as a natural result of geographical shift concerning the region of the donor language.

3.1.4. Conclusion from the early Indo-Iranian loanwords

Concerning the location of Late Proto-Uralic, these two or three successive early Indo-Iranian loanword layers form a sufficient, chronologically accurate, and relevant argument for locating the Late Proto-Uralic homeland, as they capture the initial stages of the Uralic divergence. These loanword layers point to the contact zone very close to the Indo-Iranian homeland, which was situated in Europe until the appearance of the Sintashta Culture in the Southern Trans-Ural Region (the immediate eastern side of the Ural Mountains) around 2100 BCE or even slightly later. Only around 2000 BCE there appeared Sintashta-related cultures (connected to the Indo-Iranian expansion) further to the east: Petrovka, Alakul, and Fedorovo (Fyodorovo) (Grigoriev 2021).

Concerning the Late Proto-Uralic homeland in the Altai Region, it is highly unlikely that there would have been two or three successive waves of early Indo-Iranian expansions far to the east, when we have no linguistic evidence supporting any such expansions before the disintegration of Late Proto-Indo-Iranian around 2000 BCE and when there are no archaeological traces of such cultural expansion. These waves should have crossed over the wide Western Siberia (and the adjoining regions in present-day Kazakhstan) and repeatedly ended up in the exact same region close to the assumed Late Proto-Uralic homeland in the Altai Region, which sounds unbelievable. Therefore, we must locate all the Uralic speech communities adjacent to the region where the earliest Indo-Iranian stages can be located: in the Central Ural Region close to the Fatyanovo-Balanovo, Abashevo and Sintashta cultures.

Random individuals carrying the steppe ancestry found in Southern Siberia before 2000 BCE do not suffice to explain these distinct early Indo-Iranian loanword layers in Uralic. If random individuals were enough, we should see in every language old loanword layers from very distant languages, because random individuals have always moved over huge distances.

Moreover, if we are dealing with traces from some Para-Indo-Iranian vernaculars (sister dialects to Late Proto-Indo-Iranian) in Uralic, as it now seems possible, they also require the contact zone in the Volga-Ural Region, where the whole continuum of the Indo-Iranian developments up to Late Proto-Indo-Iranian occurred (Figure 5).

Figure 5: Approximate locations of Uralic, Early Proto-Indo-Iranian (EPIIr) and Late Proto-Indo-Iranian (LPIIr) speech communities during the late 3rd millennium BCE.

3.2. Evidence from the Uralic bundle effect

This argument consists of two phenomena from different times, both of which testify to the mutual adjacency of Samoyed with other Uralic or East Uralic branches. The earlier phenomenon is the substitution of Late Proto-Indo-Iranian *a* in the Uralic branches, and the later phenomenon is the shared East Uralic sibilant changes. This argument was first formulated in Häkkinen (2023b: 80).

3.2.1. Evidence from the shared sound substitutions

Late Proto-Indo-Iranian *a* has four possible substitutes in loanwords borrowed into Uralic: *a*, *o*, *ę*, and *ä* (Holopainen 2019: 49–50). Interestingly, the choice of the vowel seems to be arbitrary, even though some tendencies have been observed concerning the *o*-substitution (Holopainen 2019: 327). During the Late Proto-Indo-Iranian layer, all those loanwords which have preserved cognates in Samoyed agree with the exact same substitute in the other Uralic branches:

1. LPIIr *a* → Uralic *o*: U *moćki-* 'wash' > LPFi *moske-* ~ LPSy *måsậ-* (SW 89; Häkkinen 2023b: 55) | U *tora-* 'fight' > LPFi *tora-* ~ LPSy *tåro-* (Aikio 2015: 62; Holopainen 2019: 282)
2. LPIIr *a* → Uralic *ę*: U *pęŋka* 'psychedelic mushroom' > LPMs *pē̮ŋk* ~ LPSy *pęŋkå* (Aikio 2015: 59; Holopainen 2019: 186) | U *ćęra* 'light color' > LPSa *ćuorē* 'light gray' ~ LPSy *sị̆rå* 'snow', *sę̆r* 'white, ice' (Aikio 2020: 125–126; Häkkinen 2023b: 59)
3. LPIIr *a* → Uralic *a*: U *ćaδa-* 'rain' > LPFi *sata-* ~ LPSy *sårå-* (Aikio 2015: 56; Holopainen 2019: 224).

The last word 'rain' shows Uralic *δ* from Indo-Iranian *d*, but this could well be the regular substitution between vowels, even though no further examples are not known. The only other convincing loanword from Indo-Iranian *d* or *dʰ* between vowels is *meti* 'honey, nectar' representing the Early/Middle Proto-Indo-Iranian loanword layer. Late Proto-Permic *gu-* 'steal' comes from Pre-Proto-Permic *kaδa-* or *kata-* (borrowed from Proto-Iranian *gada-* 'thief'; Holopainen 2019: 87), so it does not help here. This cannot be a later Iranian borrowing in Permic due to *u* (see Metsäranta 2020: 172–175). Therefore, Uralic *δ* is just as likely substitute for the Indo-Iranian *d* or *dʰ* between vowels than Uralic *t* (1 **vs.** 1).

Uralic *ä* would yield Late Proto-Samoyed *ä*, but there are no examples of this in the Late Proto-Indo-Iranian loanword layer. When we acknowledge that concerning *tora-*, Samoyed *å* cannot distinguish between the original Uralic *a* and *o* (Zhivlov 2023: 151–152), and that concerning *ćęra* and *ćaδa-*, Saami and Finnic cannot distinguish between the original

Uralic *a-a* and *ẹ-a* (Aikio 2015: 39), the probability for random occurrence of this kind of family-wide agreement between 3 or 4 possible vowel substitutes in five words can be calculated followingly: ⅓ × ¼ × ¼ × ⅓ × ⅓ = 0.00231 = 0.23 %.

To conclude: together these five words, which show the same arbitrary vowel substitution in Samoyed than in the other Uralic branches, are a sufficient argument requiring the adjacency of all the Uralic speech communities during the Late Proto-Indo-Iranian contact.

3.2.2. Evidence from the East Uralic sibilant changes

Samoyed has gone through the same process as Hungarian, Mansi, and Khanty: there occurred a chain of three successive sibilant changes of very specific nature in the same order (Häkkinen 2007: 71–74; Häkkinen 2023a: 111–114). My earlier reconstruction for the main rules was:

1. Uralic *s* merged into *š*.
2. This *š* changed to *L* (voiceless lateral fricative).
3. Uralic *ć* changed to *s* (probably through *ś*).

The result from these main rules is that East Uralic *L* is represented by *L* in Late Proto-Khanty, by *t* in Late Proto-Mansi and Late Proto-Samoyed, and by loss in Hungarian. East Uralic *s* is represented by *s* in all the four branches, except by *š* in Mansi next to a front vowel (as argued in Pystynen 2013).

I will not repeat here my earlier argumentation but instead emphasize the importance of recent observations concerning the exceptions for these rules made by Holopainen and Aikio (2023). Thanks to them, we can now get a more accurate picture about the East Uralic developments, as will be shown in the forthcoming chapters.

3.2.2.1. Phonetic evidence for the route of the East Uralic sibilant changes

Zhivlov (2023: 139) assumes that Uralic *š [ʃ] first merged into *s, and later this *s changed into *L [ɫ] or *ϑ [θ]. I have earlier assumed the merger of *s into *š, because phonetically *š resembles *L more than *s does (Gordon, Barthmaier & Sands 2002: 5) and therefore it would be easier to explain *L after the merger of *s into *š than the other way round. However, the current data indeed seems to require the merger of *š into *s.

The key piece of evidence is the early Indo-Iranian loanword '1000' with cognates in Permic and Mansi. LPIIr *ȷ́ʰasra → EPMs *šasra > *šaLəra > LPMs *šātərā (Holopainen 2019: 244; Holopainen & Aikio 2023: 28) shows the secondary *š in Mansi, while the word-internal *s > *L > *t. If the initial merger had been *s > *š, the original *š and the secondary *š would have coexisted, in which case there could not even be secondary *š. The only viable option is to assume that during this borrowing (1) the original *š had already merged into *s; (2) the secondary *š had appeared in Early Proto-Mansi; and (3) the change *s > *L had not yet occurred.

I will in short explain why *s changed to *L and not to *ϑ. The spectrum of ϑ [θ] is very similar to [f], as is seen in Gordon, Barthmaier & Sands (2002: 24). The dental and the labiodental fricative differ clearly from fricatives pronounced with the tip of the tongue more back in the mouth. Because *L* is auditorily very similar to *š* (Gordon, Barthmaier & Sands 2002: 5), these observations make it likely that either sibilant (*s or *š) would have changed more easily to *L than to *ϑ.

Because I could not find information concerning the formantwise calculational similarity of the fricatives, I did a little phonetic home experiment. I pronounced and recorded several different voiceless fricative sounds. I analyzed their first four formants F1–F4 using the Praat software (Boersma & Weenink). Then I created a matrix with the fricatives as rows and the Hertz values for the formants F1–F4 as columns in the Past software (Hammer, Harper & Ryan 2001) and produced graphs based on different clustering methods. I recognize that my homemade results are just a preliminary clue, so hopefully there will appear phonetic studies about the calculational similarity of fricatives.
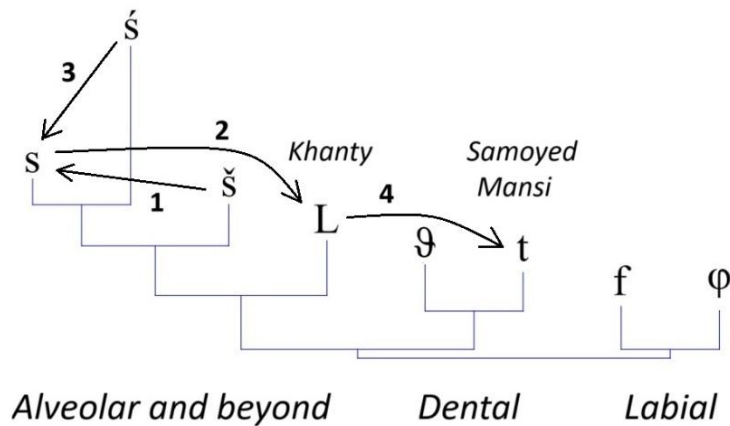
Figure 6: The East Uralic sibilant changes projected on the phonetic background: euclidean neighbor joining graph represents the similarity of voiceless fricative sounds based on the Hertz frequencies of the formants F1–F4.

The sounds included in Figure 6 are: *s* = voiceless alveolar sibilant [s] | *ś* = voiceless alveolopalatal sibilant [ɕ] | *š* = voiceless postalveolar sibilant [ʃ] | *L* = voiceless alveolar lateral fricative [ɬ] | *ϑ* = voiceless dental fricative [θ] | *t* = voiceless alveolar stop [t] | *f* = voiceless labiodental fricative [f] | *φ* = voiceless bilabial fricative [ɸ].

The neighbor joining graph first separates labial sounds from others, then dental sounds from alveolar and further back pronounced sounds. The graph shows *ϑ* similar to *t* and *L* similar to *š*, thus agreeing nicely with the spectral results above.

Furthermore, because there is *L* in Khanty, it is not plausible to assume that the East Uralic outcome was *ϑ* instead of *L*. Consequently, there is zero evidence supporting the reconstruction of *ϑ*. Still, this sound is possible but not necessary intermediary step toward *t* in Mansi, Samoyed, and some Khanty varieties. In Hungarian *L* was lost (which seems to have a parallel case in Yukaghir; see Chapter 4.1), so it is not included in Figure 6.

3.2.2.2. Increased accuracy for the East Uralic sibilant changes

Initially I interpreted the East Uralic sibilant changes proving that East Uralic was a uniform ancestral stage for Hungarian, Mansi, Khanty, and Samoyed (Häkkinen 2007). Since then,

accumulating evidence has made it clear that the East Uralic branches were already diverged as speech communities before these changes occurred (Häkkinen 2023b: 79). Likewise, the exceptions presented by Holopainen and Aikio (2023) show that some of the branch-specific exceptions must be assumed at the very early stage of the East Uralic sibilant changes (Table 4).

Uralic	Hungarian	Mansi	Khanty	Samoyed
*s, *š before *t	?	?	*s	?
*š after *i (and *i̮)	?	*š	*s	?
*s between front vowels	sz [s]	*t'	*L	*t
*s, *š in other contexts	Ø	*t	*L	*t
*ńć	gy [d']	*ńć	*s ~ *ńć	*nć?
*ć next to a front vowel	sz [s]	*š	*s	*s
*ć in other contexts	sz [s]	*s	*s	*s

Table 4: Sibilant reflexes in the East Uralic branches according to Holopainen & Aikio (2023). Samoyed added here.

As can be seen, some of the observed conditions are so rare that there is no data from every branch. Such exceptions could either unite or divide the branches, and it is impossible to guess which option is true. Figure 7 below is the most parsimonious and the most chronologically compact scenario that I could end up with when acknowledging all the current data. Even though there are earlier and later changes concerning individual branches, there are also relatively late changes shared by all the four East Uralic branches.

Several successive stages for the sibilant changes are required (stages 1–6), which underlines the close and long-lasting contact between all the East Uralic branches – even more so than in my earlier reconstructions. This makes it undeniable that there existed the East Uralic linguistic area, containing the ancestral stages of Hungarian, Mansi, Khanty, and Samoyed. Only after these changes could Samoyed have moved far away from the other branches, all the way to Southern Siberia.

TIME

7. East Uralic *s   East Uralic *L

6. *š > *ś Kh Mn / Hu Sy   Mn *V̈L > *V̈L'

5.   *š > *s Kh Mn / Hu Sy

4. Mn *iš < *iś   Kh (Hu? Mn? Sy?) *śt < *št   Kh Mn Hu Sy *š > *s

3.   Hu *V̈ś < *V̈s

2. Kh Mn Sy Hu *ńć ≡   *iś < *iš Kh Mn (Hu? Sy?)

1.   *ć > *š Kh Mn Hu Sy

Common Uralic *ć   Common Uralic *š   Common Uralic *s
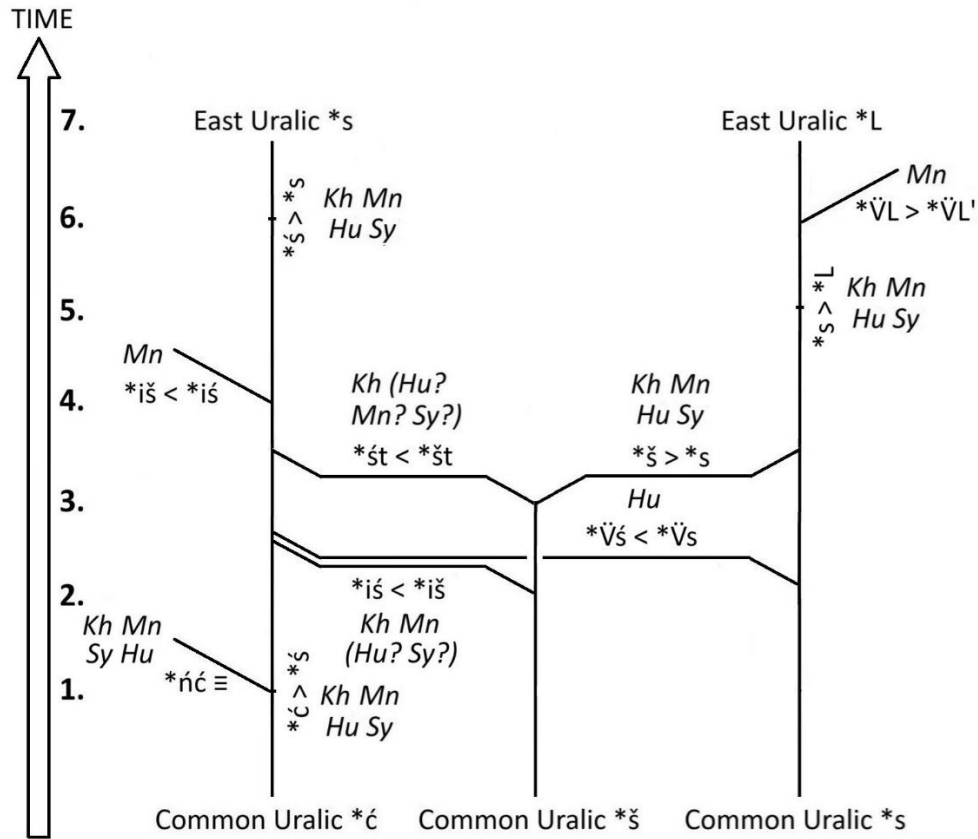
Figure 7: The East Uralic sibilant changes and the exceptions in relative chronological scenario, based on the data of Holopainen & Aikio (2023).

Figure 7 also illustrates the reason why we need the complex back-and-forth change *š > *ś > *š to explain Mansi *iš from original *iš: the sibilant could not have remained as *š the whole time, because (1) the original *iš did not merge into the original *is in Mansi, even though *š in general merged into *s; (2) the original *iš merged with the original *ić in Mansi and Khanty; (3) *š and *s in general changed to *L in Mansi, Khanty, Hungarian, and Samoyed. The reconstructed development in Figure 7 also agrees with and explains the data from early Indo-Iranian loanword layers (see Chapter 3.2.2.4).

To confirm the above reconstruction, in Table 5 there are some Uralic words going through the stages of the East Uralic sibilant changes. Samoyed has been excluded from the table because it shows so few exceptions. Zhivlov (2018b) has proposed that Late Proto-Samoyed *s was still *ś or *ć after *ń, and this seems plausible. In that case, Samoyed would agree with the other East Uralic branches: *ńć would have been preserved in them all.

46

CU	1.	2.	3.	4.	5.	6.	7.→	Branch
Common Uralic >> Late Proto-Mansi								
*pesa					*s > *L	*iL > *iL′	*L′ > *t′	*piťī
*męksa					*s > *L		*L > *t	*majt
*i̯ša		*i̯š > *i̯ś		*i̯ś > *i̯š				*aš
*šiŋiri			*š > *s		*s > *L		*L > *t	*täŋkər
*kuńći-	*ć ≡							*kuńć-
*ćilma	*ć > *ś			*śi > *ši				*šäm
Common Uralic >> Late Proto-Khanty								
*pis-ta-			*st > *śt			*ś > *s		*pestī
*pisi-					*s > *L			*päL-
*šuš-ta-			*št > *śt; *š > *s		*s > *L	*ś > *s		*Li̯stə-
*i̯ša		*i̯š > *i̯ś				*ś > *s		*as
*šiŋiri			*š > *s		*s > *L			*Läŋkər
*kuńći-	*ć ≡			(*ńć > *ś)		*ś > *s		*kus-
*ćilma	*ć > *ś					*ś > *s		*säm
Common Uralic >> Hungarian								
*šiŋiri					*s > *L		*L > Ø	egér
*pesa		*V̈s > *V̈ś				*ś > *s		fészek
*węksa					*s > *L		*L > Ø	máj
*kuńći-	*ć ≡							húgy
*ćilma	*ć > *ś					*ś > *s		szem
	1.	2.	3.	4.	5.	6.	7.→	

Table 5: Example words going through stages 1–7 of the East Uralic

sibilant changes. Branch-specific changes are included in column 7.

However, Zhivlov's assumption that the initial *s was also *ś, because in Nganasan *ę > *e after *s just like it is after *ń, can be interpreted differently. Kaheinen (2023: 60) mentions a group of Samoyed words showing the merger *ę > *e in Nganasan: *ńęr 'cartilage', *ńęrkå 'willow', *sęjə̂- 'sink', *sęr 'ice', *sęr 'salt', *sęrå 'widow', *węńȝ 'soup'. If we assume that a palatalized consonant should be near to explain this fronting, the last word 'soup' only has

*ń following the vowel (while in 'cartilage' and 'willow' *ń precedes the vowel). In that case, also for 'sink' the following *j could be the cause, not the preceding *s. This leaves us three words ('ice', 'salt', and 'widow'), all of which contain the structure *sẹr-. We cannot just decide that *s was the reason for the change *ẹ > *e, when it could just as well have been *r – or neither consonant, for that matter.

Namely, even though change from a back vowel to a front vowel would be natural next to a palatalized consonant, it does not require a palatalized consonant. We see the fronting of *ẹ in North Finnic (> *e; Kallio 2014: 160–161; Häkkinen 2019), in Mari (> *ü; Zhivlov 2023: 133), and in Permic (> *ö; Zhivlov 2023: 136), none of which are caused by palatalized consonant. Therefore, I do not see conclusive evidence supporting the assumption that Uralic *ć was word-initially preserved as *ć or *ś in Late Proto-Samoyed.

Although *L is reconstructed in Late Proto-Khanty, it preserved only in "Central Khanty" (Tromyogan, Yugan, and Kazym), while in the "Peripheral Khanty" (Obdorsk in the north, Vakh and Vasyugan in the east) it changed to *l. In "Southwestern Khanty" (Demyanka, Konda, and Nizyam) it changed to *t, perhaps due to contact with Mansi.

In Mansi *L > *t and *L' > *t'. In Samoyed the change *L > *t is apparently an independent innovation, because the loss of this sound is seen in some Uralic loanwords in Yukaghir, and such a loss can hardly be explained from *t. Instead, it requires a weaker sound, so the fricative *L is reconstructed still in Early Proto-Samoyed (see Chapter 4.1). A parallel case for this Pre-Proto-Yukaghir loss of *L is the loss of *L in Hungarian.

It is interesting to consider whether the shared *t in Mansi and Southwestern Khanty could represent a continuum to the East Uralic linguistic area. However, inferring from the current location of the varieties concerned, this contact zone is situated around the Ob-Irtysh confluence, while the region of the East Uralic linguistic area was some 400 kilometers to the southwest, in the Central Ural Region (map in Häkkinen 2023b: 82). If the change *L > *t occurred already in the Central Trans-Urals, it should be later than the expansion of Proto-Khanty toward the north, since it only concerned the "rearguard" of the Proto-Khanty speech community.

3.2.2.3. Alleged non-contemporaneity of the East Uralic sibilant changes

Zhivlov (2018a; 2023: 144) has repeated his interpretation that the few Indo-Iranian loanwords in which there is *L in Khanty but *s in Mansi or Hungarian could prove that the sibilant changes occurred non-simultaneously in the East Uralic branches. However, considering how many loanwords there are showing the regular sibilant reflexes in all the East Uralic branches **vs.** his few counterexamples, the more plausible interpretation is that these few words represent non-simultaneous borrowings. There are plenty of examples of borrowing the same word in different Uralic branches either parallelly or non-simultaneously.

The counter-argumentation of Zhivlov seems to be based on his presumption that the loanwords in Hungarian, Mansi, and Khanty must be contemporaneous, but the evidence does not support this view. As the goal of Zhivlov is the lateness of change *s > *L in Khanty, only three words allegedly testifying to this lateness are relevant: Late Proto-Khanty (1) *Lįɣər 'body armor', (2) *Lə̄pət '7', and (3) *Līkər 'sledge'.

For Khanty *Lįɣər 'body armor', there are apparently two reasons why Zhivlov considers this a late loanword: (1) the meaning 'chain mail' and (2) preserving of this word only in Ossetic and Pashto. Zhivlov reconstructs the meaning 'chain mail', but we are likely dealing here with later semantic development following technological advancement through time. The original referent of this word was likely bone lamellar armor, such as found in Seima-Turbino grave 33 in Rostovka and dated around 4000 years old (Koryakova & Epimakhov 2006: 107–108; Childebayeva et al. 2024: ROT015 in Rostovka grave 33, date 2133–1919 calBC).

Phonological analysis is primary compared to the meaning for dating the loanword. Even though attested Iranian words are scarce, Sanskrit *jagara* 'armor' is considered a borrowing from Proto-Iranian *dzagara (Holopainen 2019: 267–268), so the word is of considerable age despite its narrow distribution in Iranian. Khanty *į comes regularly from earlier *į-a, but if ablaut was involved, earlier *a-i or *o-i could also be possible (Zhivlov 2023: 146–148). Therefore, we could reconstruct *šakira or *šokira in Early Proto-Khanty, as Uralic *š is a known substitute for Proto-Iranian word-initial *ts, *dz (Holopainen 2019: 334).

In any case, Zhivlov's idea of recent borrowing (cf. Ossetic *æsqær*) is phonologically implausible: (1) we should assume metathesis of the first two sounds during the borrowing, and

(2) we should assume that Khanty *i̯ substituted Ossetic/Alanic *æ which seems far-fetched: how a close back vowel could substitute an open front vowel? Holopainen (2019: 267) gives Ossetic zɣar, borrowing of which would require that Khanty *i̯ was an "extra" vowel breaking the initial consonant cluster of the donor language, but there are no convincing examples of such an epenthetic substitution by this vowel.

For Mansi *tāt '7', there are two possible interpretations. The Mansi word could come from a later Iranian donor language in which *pt > *ft (Cantera 2017: 491), if *f was left unsubstituted. However, Mansi *t could also come from earlier *pt (cf. Aikio 2020: 62). But even if the Mansi word was borrowed from a similar donor form than the Khanty word, the two borrowings need not be contemporaneous. All the way since Late Proto-Indo-Iranian until Old Iranian, the word '7' looked the same: *saptá. As this original *s was still preserved during the Iranian-Elamite contact around the turn to the first millennium BCE (Cantera 2017: 494), there was a full millennium between the earliest possible time and the latest possible time for borrowing this word into Mansi and Khanty. We cannot just decide that the borrowing of '7' into these two branches must have been contemporaneous – especially when there seem to be even Late Proto-Indo-Iranian loanwords requiring separate or parallel borrowing into these branches (see Chapter 3.2.2.4 below).

Proto-Khanty *Līkər 'sledge' has been compared to Hungarian szekér 'wagon'. The main reason why the Indo-Iranian loan etymology for these words is considered uncertain by Holopainen (2019: 256, 269) is that this word is only seen in Indic, not in Iranian, and the final *r requires a hypothetical yet possible Proto-Indic reconstruction *śakarta- (> Old Indic śakaṭa-). According to Zhivlov (2023: 144): "This word shows that the sibilant shift in Khanty must be dated later than the analogous process in Hungarian." However, this is only one possible interpretation.

First, there are two possibilities concerning the Hungarian sz (*s): it could come from earlier *ć borrowed from LPIIr *tś or (through *ts) from a later Iranian *s. Hungarian e in the first syllable could come from earlier *ä, *e, or *i (Zhivlov 2023: 142), and the source vowel could have been the earlier Indo-Iranian *a or later Iranian *æ. When we do not even know the age of this borrowing into Hungarian, we cannot claim that it must be contemporaneous with the borrowing into Khanty.

Second, according to Zhivlov (2007: 285), Khanty *ī in the first syllable can come from earlier Ob-Ugric *ä-i, *e-i, and *i-i (earlier reconstruction: *ā̈-i, *ē-i, and *ī-i), and if ablaut was involved, yet more possible interpretations open. Likely we are dealing with *ä-substitution for earlier *a or for later *æ. The initial *L comes from earlier *s or *š, so Khanty points to an Iranian donor language (although the word is not attested in Iranian). Even here we have two possible interpretations: the *š-substitution from Proto-Iranian **tsakarta- or the *s-substitution from later Iranian **sakarta-. Consequently, it is entirely possible that the Khanty word was borrowed from Proto-Iranian and the Hungarian word from some later Iranian source. Therefore, this word cannot testify to the sibilant changes having occurred later in Khanty than in the other East Uralic branches.

To conclude, there is no evidence supporting the interpretation that these three words in Khanty were so recent borrowings that we should consider the sibilant changes in Khanty later than in the other East Uralic branches. On the contrary, one of them disagrees with the later Alanic/Ossetic origin: Proto-Iranian *pt > Ossetic *vd (Holopainen 2019: 35), but in Khanty we see *saptá → *säpta > *Lā̈pət. There are no obstacles to derive these three Khanty words from Proto-Iranian or Archaic Iranian and consider the sibilant changes in Khanty just as old as in the other East Uralic branches. Moreover, the new more accurate relative chronological scenario (as presented in Figure 7 in Chapter 3.2.2.2 above) convincingly excludes the possibility for non-contemporaneous sibilant changes in the East Uralic branches.

3.2.2.4. The secondary Mansi *š in Indo-Iranian loanwords

Late Proto-Uralic *ć is regularly reflected by Mansi *s next to most vowels: U *ćoδ'ka 'common goldeneye' > LPMs *sē̮l' | LPU *ćara- 'dry' > LPMs *sōr- | LPU *ćęmi 'fish scale' > LPMs *sē̮mə etc. Next to *i and *e in the first syllable, the regular reflex is Mansi *š (Pystynen 2013): LPU *ćilma 'eye' > LPMs *šäm | LPU *ćeδ'ami 'heart' > LPMs *šim | LPU *ćepa 'neck, collar' > LPMs *šip. We can now add the new *i̭ here: LPU *i̭ša 'outer skin' > LPMs *aš. Frontness of the vowel has never been the context here, which is confirmed by the lack of *š next to the original *ä and *ü; rather, the condition could be generalized being related to close unrounded vowels.

Because this condition is violated only by the assumed Late Proto-Indo-Iranian loanwords, Pystynen assumed that these loanwords are so late that the secondary Mansi *š had already appeared through the conditioned change and now spread to new environments. This seems to be the only plausible interpretation. However, this does not mean that these loanwords were borrowed later into Mansi than into other Uralic branches.

There are currently altogether three words in which Late Proto-Indo-Iranian *tś, *dź, or *dźʰ was substituted by the secondary *š in Mansi:

1. LPIIr *dźʰasra → Uralic *ćasra ~ LPMs *šātərā '1000'
   (Holopainen 2019: 244; Holopainen & Aikio 2023: 28)

2. LPIIr *tśatá-m → Uralic *ćęta ~ LPMs *šē̮tə '100'
   (Holopainen 2019: 242; Holopainen & Aikio 2023: 28)

3. LPIIr *tśarabhá- 'a kind of animal' → LPKh *ćerpəj ~ LPMs *šārəp 'male elk'
   (Holopainen 2019: 220; Holopainen & Aikio 2023: 28)

First, it is important to notice that these are the only loanwords in Mansi from Late Proto-Indo-Iranian *tś, *dź, or *dźʰ. Therefore, we should consider Mansi *š as the regular reflex. Second, any other source consonant does not appear even theoretically possible. We know the sound substitutions for the Indo-Iranian sibilants and affricates (Holopainen 2019: 51):

1. LPIIr *tś, *dź, *dźʰ → Uralic *ć
2. LPIIr *tš, *dž, *džʰ → Uralic ?*š/*č (initial/internal)
3. PIr *ts, *dz → Uralic *š/*č (initial/internal)
4. LPIIr/PIr *s → Uralic *s
5. LPIIr/PIr *š → Uralic *š

We know that early Indo-Iranian *s and *š were substituted by Uralic *s and *š, and these both (through merger *š > *s) yield East Uralic *L and further Mansi *t. We do not see the secondary Mansi *š replacing reflexes of LPIIr or PIr *s (no early examples of IIr *š):

4. LPIIr/PIr *saptá '7' → Uralic *säpta > *Läpta > LPMs *tā̆pət '7'
   (Holopainen 2019: 239)

5. LPIIr/PIr *Hásura 'god, master' → Uralic *asora > *aLəra > LPMs *ātər 'lord'
   (Holopainen 2019: 70; Aikio 2015: 37–38)

6. LPIIr/PIr *sáras- 'lake' → Uralic *sara > *Lara > LPMs *tōr 'lake'
   (Holopainen 2019: 217)

As the secondary Mansi *š now seems to be the regular reflex for Late Proto-Indo-Iranian palatalized affricates, the only plausible explanation seems to be that during the assumed Late Proto-Indo-Iranian loanword layer the East Uralic sibilant changes had already begun. How does this scenario agree with the fact that also Proto-Iranian loanwords have gone through the East Uralic sibilant changes? In Proto-Iranian loanwords we see Mansi *š, but here it comes word-internally from earlier *č (as seen in the Uralic vocabulary) which was a substitute for Proto-Iranian *ts, *dz. Word-initially the substitute for these Proto-Iranian affricates was *š (Holopainen 2019: 51, 334).

7. PIr *patsu-ka- 'livestock' (dim.) → *päčəka > LPMs *pā̆šəγ 'reindeer calf'
   (Holopainen 2019: 196)

8. PIr *dz̦ʹHnya- 'gold' → *šẹrńa 'gold/copper' > *Lẹrńa > Mansi *tarəń
   (Holopainen 2019: 232; no certain examples of later Iranian *z → *s > *L)

Now, considering the refined relative chronology of the East Uralic sibilant changes as presented in Chapter 3.2.2.2, the Proto-Iranian words which participated in the change *š > *L must have been borrowed before stage 3, while the secondary Mansi *š appeared at stage 4.

This would indeed require that the two loanword layers were roughly contemporaneous: the assumed Late Proto-Indo-Iranian layer was not the proto-language itself but its conservative daughter dialect – either Archaic Indo-Iranian (if the source sounds were still *tś, *dź, *dźʰ) or Proto-Indic (Proto-Indo-Aryan, if the source sounds were already *ś, *ź, *źʰ).

Unfortunately, most of the Late Proto-Indo-Iranian *ć-borrowings lack the Mansi cognate, so we do not know if some other loanwords are earlier and could be considered borrowings from Late Proto-Indo-Iranian proper. Perhaps those loanwords which have cognates in Samoyed were earlier than the Mansi *š-borrowings? In any case, it looks like the loanwords were still borrowed simultaneously into different Uralic speech communities, but their sibilant substitutions were already different (unlike the vowel substitutions; see Chapter 3.2.1).

An illustrative example is Uralic *ćasra '1000', which contains two different sibilants. It is necessary to assume parallel borrowings here:

A) Early Proto-Permic *ćasra > MPPe *śosra > LPPe *śurs '1000'
B) Early Proto-Mansi *šasra > East Uralic Mansi *šaLəra > LPMs *šātərā '1000'

Permic metathesis *sr > *rs could be connected to the recent loss of the final vowel, as word-final cluster *sr appears to have been impossible. In Mansi, it seems natural that an epenthetic vowel *ə appeared to break the cluster *Lr which feels impossible to pronounce otherwise. Consequently, this now three-syllable word would have later developed like the original Uralic *a-i-combinations (> LPMs *ā-; cf. Zhivlov 2023: 146).

The secondary affricate *ć in Mansi and Khanty sometimes possibly corresponds to Archaic Indo-Iranian *tś, *dź but sometimes to Iranian *tš, *dž (Holopainen & Aikio 2023: 30–33). At this point, not much can be said about the donor languages in such random cases. It seems possible that the secondary affricate *ć appeared in Mansi and Khanty at the stage when there still existed some archaic Indo-Iranian variety with preserved *tś, *dź (like Archaic IIr *tśarabhá- → LPKh *ćerpəj ~ EPMs *šarapa > LPMs *šārəp 'male elk').

To conclude, the refined chronological scenario concerning the East Uralic sibilant changes seems to open new possibilities to interpret the Indo-Iranian loanwords showing diverse sibilants and affricates in Uralic branches.

3.2.3. Conclusion from the Uralic bundle effect

Samoyed could not be situated far from the western and central Uralic branches until after the Late Proto-Indo-Iranian loanword layer, and it could not be situated far from the other East Uralic branches until even later, because the East Uralic sibilant changes occurred only after the Proto-Iranian loanword layer.

Despite the interesting conditional exceptions to the main rules, the totality of the East Uralic sound changes is still fully valid. After acknowledging the new data, the sibilant developments must be seen containing even more stages than was previously assumed, thus testifying to the reality of the East Uralic linguistic area. The reconstructed complexity of the shared sound changes requires long-lasting adjacency of Hungarian, Mansi, Khanty, and Samoyed around the second quarter of the 2nd millennium BCE (after the borrowing of the Proto-Iranian loanwords; see Häkkinen 2023b: 62).

This argument is sufficient, because both the shared vowel substitutions and the East Uralic sound changes are sufficient pieces of evidence for proving the regional adjacency. The argument is also chronologically accurate (moored in the chronology of Indo-Iranian), but as it is slightly later than the previous argument (although overlapping with it), it is temporally slightly more distant from Late Proto-Uralic.

However, the Uralic bundle effect has also a temporal dimension in addition to the spatial dimension. It is highly improbable that the tight bundle of Uralic and later East Uralic speech communities could have remained intact if some members of the bundle moved farther from others. It is even more improbable that the members would all have moved together as a bundle, at the same time to the very same new distant location. Therefore, the location of the bundle of the diverging Uralic speech communities must be the exact same location as the location of Late Proto-Uralic.

Since the western and central branches Saami, Finnic, Meryanic, Mordvin, Mari, and Permic developed on the western side of the Urals, and since Hungarian, Mansi, and Khanty developed immediately on the eastern side of the Urals, the Uralic bundle effect testifies to the Late Proto-Uralic homeland in the Central Ural Region, tying also Samoyed in this region until the mid-2nd millennium BCE.

3.3. Evidence from *sęksa 'Siberian pine'

LPU *sęksa 'Siberian pine' > Permic *sus > Komi *sus ~ Udmurt *sus|i̯- 'juniper'
| Ob-Ugric *sęksᴈ̂ > Mansi *tē̮t : *tajt- ~ Khanty *Li̯ɣəl | Samoyed *ti̯tå̊|jŋ
(Aikio 2015: 60; Zhivlov 2023: 160)

The next argument is the name for 'Siberian pine (*Pinus sibirica*)', *sęksa*. With individual pieces of evidence like the tree names, quality is of paramount importance. Considering the etymology in Late Proto-Uralic, high quality requires regular or recurring sound correspondences, the same or almost the same meaning, and wide enough distribution within the language family.

This word shows the same meaning 'Siberian pine' and regular sound correspondences in Permic, Mansi, Khanty, and Samoyed (Aikio 2015: 60). It testifies to the familiarity of this tree among the Late Proto-Uralic speech community, because Permic was always situated in Europe and was left outside the East Uralic linguistic area. We can assume that the word has disappeared from the more western Uralic branches, which advanced far from the habitat of Siberian pine. In Udmurt (belonging to the Permic branch), which is also spoken outside the natural habitat of Siberian pine, the word has been preserved, but its meaning has shifted to 'juniper'.

Thus, this is a qualitatively sufficient argument. However, its chronological accuracy is not self-evident but depends on other arguments: the further to the east the other valid arguments locate the Late Proto-Uralic homeland, the less chronologically accurate this argument becomes, because Siberian pine has spread from the east to the west. In the east, this word could have appeared in the language thousands of years before Late Proto-Uralic.

In Western Siberia, Siberian pine has been present for a long time (Blyakharchuk 2003), while in the Upper Kama Region it appears only around 1300 BCE (Lapteva et al. 2017: 330). In the Central Ural Region, which is still currently divided by the habitat limit of Siberian pine (Figure 9; AgroAtlas: *Pinus sibirica*) and therefore representing the frontier of its slow advancement, this is chronologically more accurate and relevant argument concerning Late Proto-Uralic. It is also a fully valid argument for excluding all the homelands to the west from

the Central Ural Region (except the northeasternmost corner of European Russia, but nothing suggests that the Late Proto-Uralic homeland could be located there).

A = The north-eastern boundary of the Abashevo Culture (early stage).
B = The southern boundary of natural habitat of Siberian pine.
C = The approximate area of the Sintashta Culture.
D = The approximate area of the Koptyaki Culture.
E = The north-eastern boundary of natural habitat of elm.
F = The northern boundary of forest-steppe.
✦ = The Central Ural Passage.
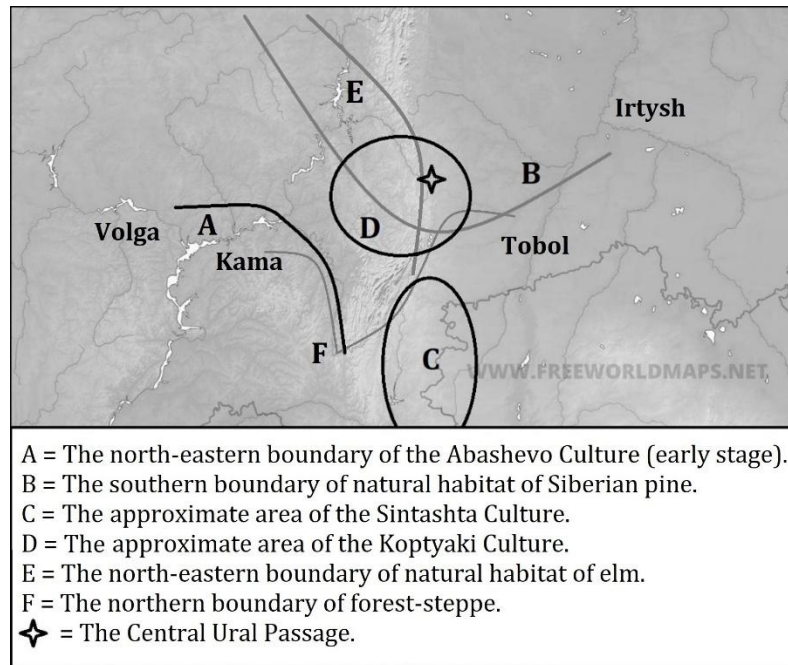
Figure 8: The overlapping of the natural habitats of Siberian pine (B) and elm (E) in the Central Ural Region.

3.4. Evidence from *ćęli 'elm'

LPU *ćęli 'elm' > Mordvin *śäl´əń | Mari *šolə | Hungarian *szil* (Aikio 2014: 67) | Mansi *sēl´t 'linden bast' > northernmost varieties: 'willow bast?' (In compound words only; Kannisto 2013: 741b)

A reverse instance compared to the previous word is seen with *ćęli 'elm' (European elms: *Ulmus laevis*, *Ulmus glabra*), because elm spread from the west to the east. The European elm species still have not advanced beyond the Ural Mountains (AgroAtlas: *Ulmus laevis*, *Ulmus glabra*). However, there is also Siberian elm (*Ulmus pumila*; see below).

This word shows the same meaning 'elm' in Mordvin, Mari, and Hungarian (Aikio 2014a: 67), and regular sound correspondences between Hungarian and Mansi. In Mari, there are three possible reflexes from Uralic *ẹ, although reason for such a division is unknown beyond some observed tendencies (Metsäranta 2020: 80–81, 314–315): there are 15 words with *ü, 5 words with *o, and 4 words with *å. Even though it is less frequent, *o is still a strong recurring minority reflex and agrees with the *ẹ-reconstruction.

For the Mordvin cognate *śäĺəń, further examples (from Aikio 2020) for the vowel development can be presented next to *ć: U *ẹ́ći- 'set' > LPMd *äźəm 'place, bench' (48) | WU *ćẹ́ćki 'mosquito' > LPMd *śäśkə (122). In addition to these three words, there are only three more words with *ẹ next to *ć with cognates in Mordvin (pages from Aikio 2020): U *ćẹlkaw 'pole, rod' > LPMd *śalgə̂ (122) | U *ćẹmi 'scale (of fish)' > ? LPMd *śav 'money' (124; *v irregular) | U *ćẹta '100' > LPMd *śadə̂ (126). Therefore, the game is 3–3, and we cannot claim either Mordvin reflex (*a or *ä) as the regular one – they both are equally regular. And even if the game was 4–3, the status of regularity (equaling majority: 4/7) here would be only slightly higher than the status of recurring irregularity (3/7). Also the latter case would still provide a plausible etymology in such a situation. Therefore, all these four branches can be plausibly derived from Late Proto-Uralic *ćẹli.

For the Mansi cognate *sēĺʹt, the semantic shift to 'bast' is supported by similar shifts in other languages concerning words for 'linden' and 'elm', the most important source trees for bast (Häkkinen 2023b: 66). Unlike elm, linden grows nowadays to the east from the Urals in the southern part of the Mansi region, around the Irtysh-Tobol confluence (AgroAtlas: *Tilia cordata*). The association of this word to different trees has shifted along with the advancement of the Mansi speakers to new regions where earlier trees did not grow: from the original 'elm' to '(elm?) bast' in the Central Ural Region, to 'linden bast' on the eastern side of the Urals, and to 'willow bast' in the northernmost varieties.

Both the extra palatalization of *l and the extra consonant (here *t) are recurring phenomena in Mansi. According to Aikio (2020: 151), the initial *ć could have triggered the palatalization of *l even over a back vowel: U *ćoδka > *ćolka > *ćolʹka > LPMs *sēĺʹ 'goldeneye' | *ćülki > *ćülʹki > LPMs *säĺʹəγ- 'spit'. We can therefore assume the development *ćẹli > *ćẹĺʹi > LPMs *sēĺʹ-t 'bast'.

There are examples showing an unexplained extra consonant in Mansi: U *aδ'i-* > LPMs *āl'āt* 'bed' (Aikio 2020: 7: unknown suffix) | U *ali/*ola* > LPMs *ūləć* 'chin, lower jaw' (Aikio 2020: 11: obscure suffix). In verbs an extra *t* without causing any change in the meaning is more frequent. Therefore, these recurring irregularities do not hinder deriving Mansi *sēl´t* from Late Proto-Uralic *ćęli*.

Khanty and Saami advanced outside the natural habitat of elm, so the disappearance of this word is expected in these branches. Finnic *salaga* 'crack willow (*Salix fragilis*) etc.' has been connected to the Uralic word for 'elm', because it could be derived from it quite regularly, but Germanic origin seems semantically more plausible (Aikio 2014a: 67). The second syllable *a* in Finnic does not agree with either the Uralic or the Germanic word, so it must in any case be assumed irregular.

Because the European elm reached the Kama-Ural Region already around 5000 BCE (Shumilovskikh et al. 2020: 533), this argument is chronologically inaccurate: the word could have appeared in the language long before Late Proto-Uralic. Therefore, it must be considered invalid for giving positive testimony in support of the Late Proto-Uralic homeland in the Central Ural Region. At the same time, it is still a fully valid argument for testifying against all the homelands in Western Siberia.
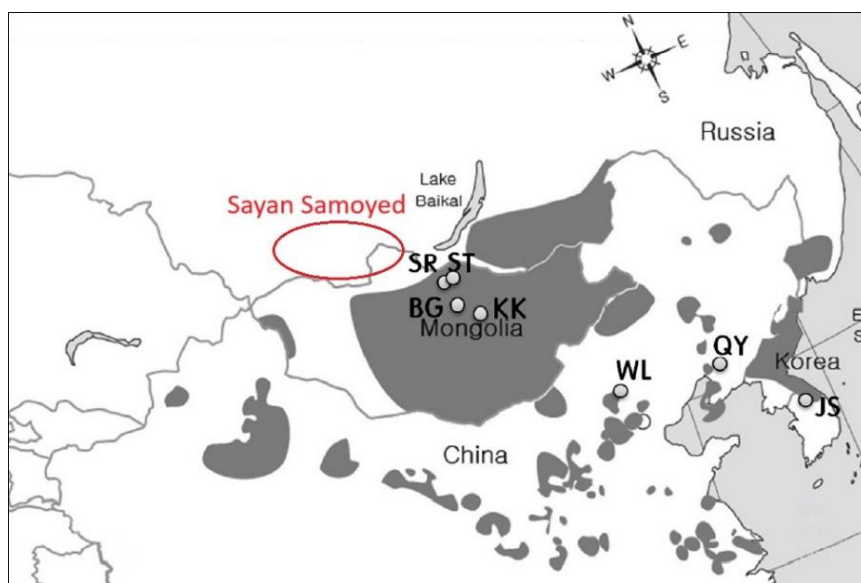

Figure 9: The region of Sayan Samoyed compared to the natural habitat of Siberian elm (*Ulmus pumila*). Original map from Park et al. (2016).

However, there is Siberian elm (*Ulmus pumila*) which despite its name grows mainly in more southern regions: in Siberia it only grows in the East Baikal Region and near Vladivostok (Park et al. 2016: 3; see Figure 9 above). The closest region of Siberian elm to Southern Siberia is in Mongolia. Could the Late Proto-Uralic word for 'elm' be explained from Siberian elm? The answer is negative, for two reasons:

1. If the Late Proto-Uralic homeland was situated in the Altai Region, that would be the same location as the Late Proto-Samoyed homeland. In that case, we would expect to see cognates of *ćęli* in the Samoyed languages spoken in the Sayan Region – but there were none.

2. If the Late Proto-Uralic homeland was situated in the Altai Region, we would expect the word for 'elm' to have disappeared from all the other Uralic branches than (Sayan) Samoyed, because the whole wide Western Siberia was empty of elm (e.g. Blyakharchuk 2003; Lapteva, Korona & Zhilin 2020). After all, the names of trees have generally disappeared from the branches which are not spoken close to the natural habitat of trees concerned.

The distribution of the word *ćęli* 'elm' in the Uralic languages only agrees with the homeland in the Central Ural Region and to the west from there: if the homeland was in Southern Siberia, the distribution of the word for 'elm' should be restricted in the Sayan Samoyed languages. Even though this argument is not chronologically accurate concerning the Central Ural homeland, it is still sufficient, spatially accurate, and relevant to exclude the Altai Region, Southern Siberia, and Western Siberia from the candidate regions for the Late Proto-Uralic homeland.

*3.5. Evidence from *mekši* 'bee' and *meti* 'nectar, honey'*

U *mekši* 'bee' > Finnic *mehi-läinen* | Mordvin *mekš* | Permic *moš*
| Hungarian *méh*, *méj* (dial.) (Holopainen 2019: 139; Metsäranta 2020: 326)

U *meti* 'nectar, honey' > Finnic *meci* | Mordvin *med'* | Permic *må*
| Hungarian *méz* (Holopainen 2019: 146; Metsäranta 2020: 326)

These two words are early Indo-Iranian loanwords, but here relevant is only their paleo-linguistic value: what can they tell about the region where Late Proto-Uralic was spoken? I did not include this argument in Häkkinen (2023b), but due to the temporal dimension of the Uralic bundle effect (see Chapter 3.2.3 above) it is again relevant. Even though these words are lacking in the easternmost Uralic branches Mansi, Khanty, and Samoyed, the Hungarian cognates *méh*, *méj* 'bee' (< *meɣLi* < *mekši*) and *méz* 'honey' (< *meti*) have been adopted before the East Uralic sibilant changes, which requires also the adjacency of the easternmost branches in the region familiar with apiculture or at least honeybee.

From phonological perspective, Uralic *meti* could have been borrowed from any archaic Indo-European dialect, but from semantic perspective its primary meaning in Indo-European was 'mead' (→ 'wine'), while 'honey' as the secondary meaning is only found in Balto-Slavic, Indo-Iranian, and Tocharian (Mallory & Adams 2006: 262). Therefore, only these branches are credible sources for the Uralic word. As (1) *meti* shows the same *e*-vowel, (2) the same distribution, and (3) shows a bee-related meaning together with *mekši*, most likely these words were borrowed at the same time.

Uralic *mekši* has been compared to the Indo-Iranian root *mákš-*, derivations of which mean 'fly' in Iranian and Indic but 'bee' only in Indic (Holopainen 2019: 139). Therefore, 'fly' is the most likely original meaning, and the semantic shift in Indic could have been developed through compounds like *madhu-makṣikā* and *madhu-makṣa* 'honey-fly'; similar compound logic is seen in the Dravidian languages of India (Carpelan & Parpola 2001: 114). Such a compound for 'bee' could have been used in Late Proto-Indo-Iranian and even earlier (< *médhu-mékš-*), increasing the probability that *meti* was borrowed together with *mekši*.

It has been proposed that the Indo-Iranian *mákš-* could come from earlier *moko-* (Mallory & Adams 2006: 149), but because the only assumed cognate would be Lithuanian *mãkatas* 'gnat' and because Indo-Iranian *š* would remain unexplained, there are only three common sounds. Therefore, this seems a possible chance resemblance, and the Indo-Iranian word

still lacks a convincing Indo-European etymology. There are also other probable Paleo-European (substrate) loanwords in Early/Middle Proto-Indo-Iranian (see Chapter 3.1.1).

Carpelan & Parpola (2001) emphasized that the western honeybee (*Apis mellifera*) has not spread beyond the Ural Mountains until by human action during the recent centuries. Another species, *Apis cerana*, is only found in Southeast Asia and is therefore irrelevant for the Uralic homeland. However, since then more eastern endemic subspecies of the western honeybee have been found: the Tien Shan honeybee (*Apis mellifera pomonella*) and the Xinyuan honeybee (*Apis mellifera sinisxinyuan*; Ilyasov et al. 2020: 3617). These newly found subspecies are classified closest to the subspecies in Western Asia, so they do not represent an expansion of the European subspecies. Western Siberia still lacks honeybee except for colonies spread by humans during recent centuries.

Therefore, the value of this argument is similar to 'elm' in the previous chapter: if 'bee' and 'nectar, honey' were borrowed in Southern Siberia, we would expect these words to be preserved in the Sayan Samoyed languages and to have disappeared from all the other branches which would have crossed the wide honeybeeless Western Siberia on their way to the Urals and Europe. Instead, we see cognates in the European branches and Hungarian, and this situation only agrees with the homeland in the Central Ural Region and to the west from there.

Chronologically this argument is not accurate, because utilizing bees and collecting honey have long roots in Europe; **mélit* 'honey' can be reconstructed already in Early Proto-Indo-European (Proto-Indo-Anatolian; Mallory & Adams 2006: 262), which was spoken over 6000 years ago. Nevertheless, just like 'elm', this argument is relevant for excluding the homelands in Western or Southern Siberia.

Figure 10 shows the original regions of the diverging Uralic branches (excluding the three easternmost branches on the eastern side of the Central Ural Region) around the mid-2nd millennium BCE. This is considerably later than the borrowing of 'bee' and 'honey, 'nectar' but illustrates why these words were preserved in Finnic, Mordvin, Permic, and Hungarian.

The northern limit for the natural distribution of the brown bee (*Apis mellifera mellifera*) is near the 60th parallel north, which also roughly aligns with the northern limit of temperate broadleaved trees oak (*Quercus robur*), elm (*Ulmus laevis*, *Ulmus glabra*), maple (*Acer platanoides*), linden (*Tilia cordata*), ash (*Fraxinus excelsior*), and hazel (*Corylus avellana*)

(AgroAtlas). Toward the east, slowest has been ash, reaching only the Great Volga Bend, and the fastest has been linden, having crossed over the Central Urals to Siberia and currently advancing along Lower Irtysh. Oak, elm, maple, and hazel have stopped in the Central and Southern Ural Region already thousands of years ago, just like the brown bee.

Figure 10: The original regions of diverging western and central Uralic branches projected on the natural habitat of *Apis mellifera mellifera* (map from Wainwright 2025)

Interestingly, names for all these other trees except elm are assumed to be borrowings from Paleo-European languages in the western and central Uralic branches: *tammi* 'oak' (Fi Md) | *wakštira* 'maple' (Fi Md Mr) | *päkšna* 'linden' (Fi Md Mr) | *särńa* 'ash' (Fi Mr) | *päški-*/*pekši-* 'hazel, nut' (Fi Md Mr Pe) (Häkkinen 2009: 37–39, 47–48; Aikio 2015: 43–47). As cluster *kš* is also seen in *mekši* 'bee', this word is likely originally a borrowing from the same language family. In that case *š* is probably not the result of the Ruki rule in Indo-Iranian, but the borrowing from Paleo-European to Indo-Iranian still could not precede the appearance of *š* in Indo-Iranian, which is connected to the Ruki rule (but in theory could have occurred already earlier along with loanwords). Another possibility is that the word was parallelly borrowed from Paleo-European to Uralic. Then it could be already Late Proto-Uralic but not earlier, if Early Proto-Uralic arrived from Southern Siberia only slightly before.

3.6. Conclusion from the arguments supporting the Central Ural homeland

	Sufficient	Accurate	Relevant	Valid
Early Indo-Iranian loanword layers in Uralic	+	+	+	✓
The Uralic bundle effect	+	+	+	✓
seksa 'Siberian pine'	+	?/+	?/+	✓
ćẹli 'elm'	+	–	*	Ø / ✓
mekši 'bee' and *meti* 'nectar, honey'	+	–	*	Ø / ✓

Table 6: Assessment of the arguments supporting the Central Ural home-
land for Late Proto-Uralic (* = relevant for excluding eastern homelands).

All the five arguments are sufficient for locating purposes. The early Indo-Iranian loanword layers and the Uralic bundle effect are also chronologically accurate and relevant. 'Siberian pine' is chronologically accurate and relevant, when it concerns the Central Ural Region, but less so, if it concerns Southern Siberia where the presence of this tree is much earlier. 'Elm', 'bee', and 'nectar, honey' are not chronologically accurate and therefore cannot testify for the Late Proto-Uralic homeland in the Central Ural Region, but they are still fully valid arguments testifying against all the homelands to the east from the Ural Region, because the distribution of these words in Uralic does not agree with the spread from Southern Siberia over the elmless and honeybeeless Western Siberia to Europe.

The map in Figure 11 illustrates the incompatibility of the Altai/Southern Siberian homeland with the western-central distribution of the Uralic words *ćẹli* 'elm' (Md Mr Hu Mn), *mekši* 'bee' (Fi Md Pe Hu), and *meti* 'honey, nectar' (Fi Md Pe Hu). Even if these words would have originally denoted the Asian elm and honeybee, we would expect them to have survived in the Sayan Samoyed languages only and to have disappeared elsewhere.

When none of the arguments considered in Chapter 2 could support the location of the Late Proto-Uralic homeland in the Altai Region or Southern Siberia, there is neither contradiction nor competition – the Late Proto-Uralic homeland must be located in the Central Ural Region.
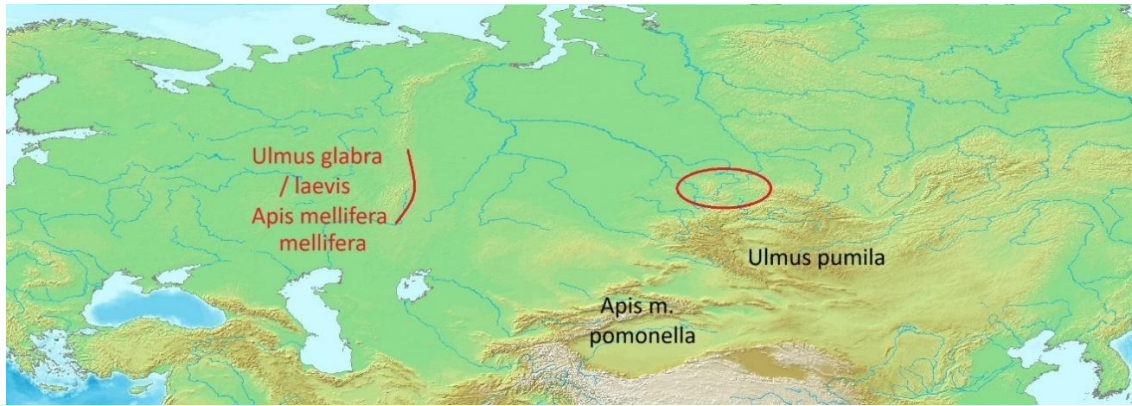
Figure 11: The assumed Southern Siberian homeland for Late Proto-Uralic in comparison to the natural habitat limit of the European elm and the western honeybee (red) and the closest Asian elm and honeybee (black).

3.7. Consequences for the Uralic taxonomy

"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-Samo-yedic from the rest of the family, as was widely assumed in 20th century Uralic studies." (Grünthal et al. 2022: 10)

I have earlier accepted this quoted conclusion. However, an updated selection of the data and more precise layering of the early Indo-Iranian loanwords (as presented in Tables 3 and 4 in Chapter 3.1.3) aligns Mansi and Khanty with Samoyed: all these three branches show the total lack of the Early/Middle Proto-Indo-Iranian loanwords (0/11). This result does not point to an early divergence of Samoyed but instead suggests that perhaps the easternmost Uralic speech community was ancestral to Mansi, Khanty, and Samoyed.

At the same time, this kind of alteration between the results from differently delimited or divided data sets illustrates how vulnerable this kind of argumentation (based on an observed distributional gap) is to chance. Statistical probability calculations would show that the distributional gaps both in the data set of Grünthal et al. (2022: 17) and in the data set presented in Chapter 3.1.3 are highly improbable to have formed through random loss of words. And

yet, the results contradict each other. Perhaps either one of these two results is true, or perhaps the true result has not yet been found.

In any case, concerning the Early/Middle Proto-Indo-Iranian loanword layers, it is no longer plausible to assume that Samoyed alone was the first branch to split off from Late Proto-Uralic. Based on the updated and more high-resolution layering, the more plausible one of the two distributional gaps is the one presented here: Samoyed aligns with Mansi and Khanty, while Hungarian aligns with the western and central Uralic languages. Another argument which aligns Hungarian with the western and central Uralic branches are the words *mekši* 'bee' and *meti* 'honey, nectar' (see Chapter 3.5). Perhaps only later Early Proto-Hungarian became integrated within the East Uralic linguistic area responsible for the spread of the sibilant changes and the shared Ugric developments?

In theory it is also possible that Late Proto-Uralic was still a uniform language during the Early/Middle Proto-Indo-Iranian contact or even during the Late Proto-Indo-Iranian contact (due to the shared vowel substitutions; see Chapter 3.2.1) and that both distributional gaps are illusory. We simply have no conclusive data about the topic.

The often-seen argument for the status of Samoyed as the first branch to split off from Late Proto-Uralic is based on the low number of inherited Uralic words it shares with other Uralic branches. However, it is not a reliable method to directly translate the number of shared words to a family tree, because different languages retain different amounts of the inherited vocabulary. Branches like Samoyed, which have gone through strong lexical replacement, appear illusorily distant from other branches, if we only look at the lexical level (see Häkkinen 2012b).

Still even today the same erroneous presupposition is repeated in phylolinguistic studies all over the world, and family trees are inferred directly from the matrix of pairwise sums of the shared words (mostly retentions from the common proto-language). Such a faulty method can never produce results capable of challenging family trees based on the methods of historical linguistics. Furthermore, when based on an illusory taxonomy, the dates for proto-languages are also illusory.

Concerning the taxonomic structure of the Uralic language family, the valid linguistic evidence forces us to accept both the early divergence of the Uralic branches and the secondary shared traits between branches, as modelled in Häkkinen (2023b). There is no room for too

simplistic divergence models. This kind of overlapping or fluctuation between stages of divergence and convergence can be illustrated in a "family curtain" graph (Figure 12).
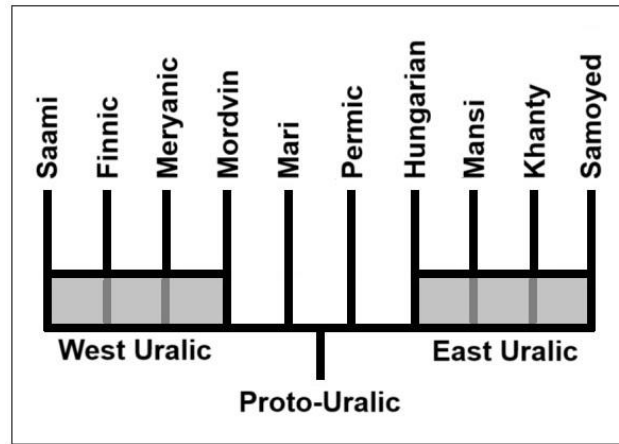
Figure 12: The Uralic family curtain. The grey parts represent areal convergence between already diverged branches.

4. Further evidence concerning ancient linguistic diversity in Southern Siberia

4.1. Evidence from Yukaghir

Loanwords between Uralic and Yukaghir can also offer evidence for locating some reconstruction stages of the Uralic language lineage. Based on this contact I have earlier (Häkkinen 2012a) proposed that Pre-Proto-Uralic was spoken near the Altai-Sayan Region. According to Aikio (2014a) there seem to be Uralic loanwords in Yukaghir preceding the Samoyed (and thus East Uralic) sound changes. Three words show Yukaghir *č or *č́ as the reflex for Late Proto-Uralic *ć:

   1. Proto-Yukaghir *ečē 'father' (Aikio 2014a: 69)
     ← Uralic *?*ä(j)ćä* 'father' > Late Proto-Samoyed *äjsä*

   2. Proto-Yukaghir *lanč́in- 'slow, calm' (Aikio 2014a: 71; 2015: 52)
     ← Uralic *lań́ći* 'calm, soft, gentle' (not preserved in Samoyed)

67

3. Proto-Yukaghir *qanč́- 'cold' (Aikio 2014a: 71)

    ← Uralic *kVnćä- (?*känćä-) 'cold; get cold' > Late Proto-Samoyed *kə̆nsä-

As the regular reflex of *ć is Samoyed *s, these loanwords seem to require more archaic donor language than Early Proto-Samoyed, which arrived from the Central Ural Region only after the East Uralic sibilant changes around the middle 2nd millennium BCE. However, it seems possible that Late Proto-Uralic *ńć was preserved as Samoyed *ńć (Zhivlov 2018b). In that case only the first word could testify to the earlier loanword, and its reflexes are very diverse within Uralic. Therefore, it remains uncertain if there was actual contact between Early Proto-Uralic and Pre-Proto-Yukaghir; shared words containing chronologically diagnostic sounds are few. More research is needed on this topic.

    Two words point to the secondary Samoyed *s from Late Proto-Uralic *ć, thus representing a later layer after the East Uralic sibilant changes:

4. Proto-Yukaghir *?*solijə 'intestine, gut' (Aikio 2014a: 74)

    ← Early/Late Proto-Samoyed **sålä (not preserved)

    < Late Proto-Uralic *ćali 'gut' (Aikio 2022: 23) >

5. Proto-Yukaghir *so/alɋ́ə 'loon (Gavia)' (Aikio 2014a: 75)

    ← Early/Late Proto-Samoyed **sålḱå (not preserved)

    < Late Proto-Uralic *ćoδka 'goldeneye'

Of similar age with the previous two words are five words showing the loss of the initial consonant in Yukaghir, explained from Early Proto-Samoyed *L (from Late Proto-Uralic *s or *š):

6. Proto-Yukaghir *al'- 'melt, thaw' (Aikio 2014a: 69)

    ← Early Proto-Samoyed **Ləlä- > Late Proto-Samoyed **təlä- (not preserved)

    < Late Proto-Uralic *sula- 'melt, thaw'

7. Proto-Yukaghir *ere* 'fork' (Aikio 2014a: 70)

    ← Early Proto-Samoyed ***Lärä* 'fork?' > Late Proto-Samoyed ***tärä* (not preserved)

    < Late Proto-Uralic *särä* 'forked object'

8. Proto-Yukaghir *nō-* 'scrape, scratch' (Aikio 2014a: 72)

    ← Early Proto-Samoyed *nəL-* 'scrape' > Late Proto-Samoyed *nət-*

    < Late Proto-Uralic *nusi-* 'scrape, scratch'

9. Proto-Yukaghir *olo-* 'steal' (Aikio 2014a: 72)

    ← Early Proto-Samoyed *Lålä-* 'steal' > Late Proto-Samoyed *tålä-*

    < Late Proto-Uralic *sala-* 'steal'

10. Proto-Yukaghir *oŋ-* 'put on, fit in' (Aikio 2014a: 72)

    ← Early Proto-Samoyed ***Loŋ-* 'enter' > Late Proto-Samoyed ***toŋ-* (not preserved)

    < Late Proto-Uralic *soŋi-* 'penetrate, enter'

It is remarkable that of the seven words representing the post-East Uralic loanword layer in Yukaghir, only two have preserved cognates in Samoyed. This suggests that either (1) strong lexical replacement in Samoyed began only after the contact between Early Proto-Samoyed and Pre-Proto-Yukaghir, or (2) the donor language did not represent the actual Samoyed language lineage but some other (later extinct) East Uralic variant which also spread far to the east (say, Angara Uralic).

Even though the loanwords seem to have gone mainly from Uralic to Yukaghir, there are some indications of possible Yukaghir influence on Samoyed. Within the Uralic language family, exclusively Samoyed logic for the numeral '8' is "2 × 4": *ketä(-jn)-tättə*. This resembles the logic seen in Yukaghir: Tundra Yukaghir *maalajla-* '6', *maalajlakla-* '8' (*maala-* 'both sides'; Nikolaeva 2006: 257) | Kolyma Yukaghir *malγi-jaaloo* '6', *malγ-ilek* '8' ("joint three", "joint four"; Nikolaeva 2006: 273).

Thus, the contact between Early Proto-Samoyed and Pre-Proto-Yukaghir can be considered proven, but Early Proto-Uralic or Para-Uralic contact with the Yukaghir language lineage remain uncertain. However, increasing evidence about contact between these stages and Early

and Middle Proto-Tocharian (see Chapter 2.6) already points to their presence in Southern Siberia. The proposed Early Proto-Tocharian loanwords in Uralic, *nimi* 'name' and *käliw* 'brother/sister-in-law', have matches also in Yukaghir: LPYu *n/ńim* 'name' and *kel'* 'brother-in-law' (Aikio 2014a: 19, 21). Still, there are several possible interpretations for these words: they could be direct borrowings from Early Proto-Tocharian into Pre-Proto-Yukaghir or mediated by Early Proto-Uralic, Para-Uralic, or Samoyed (LPSy *nim* and *kälə*).

It is possible that both Yukaghir and Tocharian language lineages were in contact with Early Proto-Uralic and/or Para-Uralic approximately at equally early stage, during the 3rd millennium BCE. The most probable contact region for Uralic and Yukaghir would be between the water systems of Yenisei and Lena, to the west from Lake Baikal. The Afanasievo Culture, which is generally associated with the Tocharian language lineage, covers also this region. Therefore, it would not be a wonder if direct loanwords (not mediated by Uralic) were found also between Tocharian and Yukaghir.

Recently Zhivlov (2021) has proposed several Yukaghir loanwords in Khanty and Mansi. This would mean that (Para-)Yukaghir languages were earlier spoken also in Western Siberia, all the way to the Ural Mountains. The Seima-Turbino sites in Shaytanka and Satyga are situated on the eastern slopes of the Urals, near the rivers Iset and Konda, both within the region of current or historical Mansi placenames (Matveyev 2011: 455). This hints toward the possibility that Pre-Proto-Yukaghir was one of the languages involved in the Seima-Turbino Network in Western Siberia (see Figure 13 in Chapter 4.2).

4.2. Evidence from Yeniseian

This argument does not concern the location of either Early or Late Proto-Uralic homeland, but it can give us information about the linguistic situation in Western and Southern Siberia before the Uralic expansion.

After Late Proto-Uralic, there appeared a new local case series in the Ugric branches Mansi, Khanty, and Hungarian. In Hungarian this is already the fourth series of local cases (in importance, not chronologically), called the Family local cases. Their usage is very restricted, and therefore these three cases are excluded from many grammars. In Mansi and

Khanty this is the primary local case series (Honti 2006). On the other hand, in Hungarian these local cases and their original functions have been preserved best. This series has a striking material match in the Yeniseian language Ket (Georg 2007: 102), as seen in Table 7. I have not seen this similarity recognized earlier.

Hungarian	Mansi Tavda	Khanty Vakh	Ket
-nott/-nött, -nitt [LOC]	-nāt/-nā̈t [COM]	-nat/-nät [COM]	-ŋta < *-ŋata [ADE]
-nól/-nől, -núl/-nűl [ABL]	-nāl/-nā̈l [ABL]	---	-ŋal [ABL]
-ni/-nyi; -nak/-nek [LAT]; [DAT]	-nā/-nā̈ [LAT]	-na [LAT]	-ŋa [DAT/LAT]

Table 7: Forms and functions of the Ugro-Yeniseian local cases: LOC = 'in' ~ ADE = 'in, on' ~ COM = 'with' | ABL = 'from' | LAT = 'to' ~ DAT = 'for/to someone'.

Intra-Ugric explanations starting from a postposition have also been proposed (Honti 2006 considers these), but borrowing seems a less complicated explanation. The only difference concerns the initial consonant, which is *ŋ in Yeniseian but *n in the Ugric branches. However, a valid explanation can be presented for this difference, if we assume the direction of borrowing being from Yeniseian or a third language to Ugric: The Uralic *ŋ as an independent phoneme disappeared from all the Ugric branches early on by changing to *ŋk (and further to g in Hungarian; Zhivlov 2023: 139–140). After this early change, the natural substitute for foreign *ŋ would have been *n.

As for the other Yeniseian languages, in Yugh there are only ablative and dative, and in Kott there are none of these local cases, so their status in Proto-Yeniseian is uncertain (Werner 2002: Band 3: 73–78); distributionally this seems to concern only Northern Yeniseian (Ket-Yugh). Moreover, there has been no natural contact zone between the known Ugric and Yeniseian speech communities, as the first ones have spread from the Central Ural Region to the northeast and the second ones have spread from or through Southern Siberia to the north.

When we also acknowledge that no credible mutual loanwords between Yeniseian and Ugric have so far been proposed, the most probable explanation for this shared local case series is borrowing from some unknown Paleo-Siberian language family (possibly covering the whole width of the West Siberian Plain on the southern taiga and forest-steppe zones) separately into the Ugric branches (with the substitution *ŋ → *n) and into the Northern Yeniseian languages. Unimaginatively, this language family can be called Paleo-West Siberian. We can now add Yeniseian and Paleo-West Siberian language families among the possible candidates involved in the genesis of the Seima-Turbino Network.

The distribution of undisputed Yeniseian placenames reaches Middle Irtysh in the west (Vajda 2012: 29). Akulov & Efimova (2019) have proposed that certain river names in the Oka Region ending -ul/-ur (Nasmur, Chistur, Shershul, Dandur, Mokshur, Ninur, Pynsur, Syntul, Sesnur) could be related to the Yeniseian river names like Agul and Langur in the Yenisei Region. However, neither of the scholars seem to be linguists, and their method appears quite uncritical.
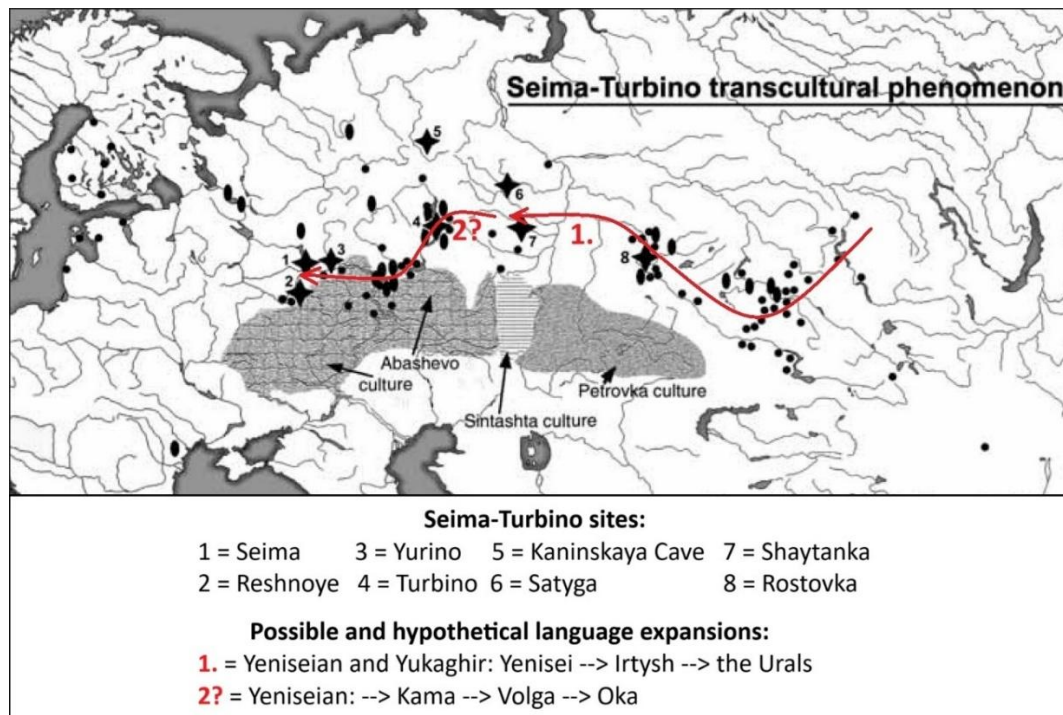

Figure 13: The sites of the Seima-Turbino Network and possible expansions of Yeniseian and Yukaghir languages. Original map from Chernykh (2012).

Nevertheless, it is an interesting coincidence that important Seima-Turbino sites Seima and Reshnoye are situated along Oka. If convincing Yeniseian placenames could be identified there, such Yeniseian speaking colonies could be associated with the expansion of the Seima-Turbino Network from Southern Siberia. River names with the ending *-ul/-ur* are only seen in Assan (Werner 2002: Band 3: 61–64), which was spoken to the east from Yenisei and to the south from Angara. This region was situated in the northeasternmost corner of the Seima-Turbino Network (Figure 13).

However, Yeniseian-looking placename elements have been identified by linguists even in the Kama-Ural Region (see Werner 2002: Band 3: 67–68), and the important Seima-Turbino site of Turbino is situated along Middle Kama. The map in Figure 13 shows the possible and hypothetical steps of expansions of Yeniseian and Yukaghir within the Seima-Turbino Network.

4.3. Conclusion from the extra-Uralic evidence

After a thorough scrutiny of the linguistic evidence, we can conclude that during the late 3rd millennium BCE there seem to have been several unrelated language lineages in or near Southern Siberia:

1. Para-Uralic
2. Middle Proto-Tocharian
3. Pre-Proto-Yukaghir
4. Pre-Proto-Yeniseian
5. Paleo-West Siberian

Any combination of these five language lineages could have been involved in the genesis of the Seima-Turbino Network – but Late Proto-Uralic was not among them (Figure 14). Uralic languages could only be associated with the Seima-Turbino Network since around 2000 BCE when the network spread to Europe through the Central Ural Passage (Häkkinen 2023b: 85–89).

Figure 14: Location of Late Proto-Uralic in comparison to language lineages in or near Southern Siberia during the late 3rd millennium BCE.

5. Interdisciplinary considerations

"In Uralic historical linguistics, a major shift is taking place with regard to timing and locating Proto-Uralic, and its disintegration and spread. This is the result of novel, truly interdisciplinary approaches on how to triangulate between linguistic, archaeological, and genetic evidence of human past in northwestern Eurasia." (Vesakoski, Salmela & Piezonka 2025: 379)

I must say that I have not seen anything fundamentally different in the current interdisciplinary studies compared to the earlier attempts. Knowledge keeps accumulating and methods keep evolving within individual disciplines, just like they have always done. Nevertheless, the basic principles always remain the same: the value of interdisciplinary approach is determined by the quality of the data and the rigor of the method.

This means that (1) we must acknowledge all the valid results and arguments from every discipline concerned, and (2) we must accept the autonomy of all the disciplines involved. It will always be only linguistics which can produce conclusive evidence concerning language. Results from other disciplines can only be used as hints when there exist no conclusive linguistic results.

In this chapter I will consider the Seima-Turbino Network, the location of Indo-Iranian, the paternal haplogroup N, the Yakutia ancestry, and correct a few misconceptions concerning interdisciplinary research.

5.1. The Seima-Turbino Network

Let us take a closer look at the archaeological phenomenon called the Seima-Turbino Network, the expansion of which has been recently associated with the Uralic speakers all the way since its genesis in Southern Siberia around 2200 BCE (Nichols 2021; Grünthal et al. 2022). Vesakoski, Salmela & Piezonka (2025; 2024) take this association as the starting point for their interdisciplinary articles, even though the valid linguistic results clearly exclude this possibility. I will not repeat here the linguistic argumentation, but instead I will focus on general problems in the association of the Seima-Turbino Network and the Uralic language family.

To propose an interdisciplinary correlation for the spread of a certain language family, we should find a precise spatiotemporal match for the linguistic results. Partial spatial match is not enough, because it can only explain the linguistic expansion partially. Seima-Turbino items cover a huge region between Northwestern China and Finland already during the early second millennium BCE, but the Uralic languages did not reach Finland so early – and they never reached Mongolia and China.

In the west, around the mid-2nd millennium BCE the Uralic languages had reached the Upper Volga Region (Häkkinen 2023b: 88–89), and only at the end of that millennium (or even later) did the Uralic languages reach the coast of the Baltic Sea (Lang 2020). Moreover, the presence of unknown Uralic languages in the west before Saami and Finnic seems improbable, because in these branches there are loanwords from non-Uralic Paleo-European languages which they replaced during their advancement. Even the random placenames from the Meryanic *x*-language in Finland (Rahkonen 2013) do not need to precede the Saami expansion – they could just as well represent a later expansion. We have no reason to assume that the Meryanic sound changes occurred earlier than the sound changes in the other Uralic branches or that the Meryanic expansion was earlier than other Uralic expansions.

In the east, the Uralic bundle effect tied Samoyed in the Central Ural Region until the mid-2nd millennium BCE (Chapter 3.2); only after that could Samoyed have moved to Southern Siberia. Consequently, the temporal gap from the beginning of the Seima-Turbino Network to the arrival of the Uralic languages to Southern Siberia (where the STN emerged) approaches a full millennium.

Then we have the northern dimension of this network on both sides of the Urals: Kaninskaya Cave along Upper Pechora and Satyga along Konda (Chernykh, Korochkova & Orlovskaya 2017: 47). These sites cannot be connected to any Uralic branch, as all the branches stem from the southern taiga zone until quite recent northward expansions (Saarikivi 2022).

As a conclusion, the expansion of the Uralic languages to the western and the eastern ends of the known region of the language family is much later than the expansion of the Seima-Turbino Network in these regions, and the southeastern and northern extensions of the network do not match Uralic languages at all. Therefore, this match is spatiotemporally rather poor.

However, a minor early stage of the Uralic expansion, from the Central Ural Region to the Upper Volga Region, could perhaps be connected to the spread of the Seima-Turbino Network through the Central Ural Passage to the Volga-Kama Region in the easternmost Europe, as described in Häkkinen (2023b: 88–89). The radiocarbon dates from the western Seima-Turbino sites along Middle Volga (between the Oka and the Kama confluences) agree with this, fitting within the first third of the 2nd millennium BCE (Chernykh, Korochkova, & Orlovskaya 2017; Marchenko et al. 2017).

5.2. Locating Indo-Iranian

Recently Palmér (2025) has thoroughly considered the development stages from Core Indo-European to Late Proto-Indo-Iranian. He concludes that the lexical evidence is a strong indication for Indo-Slavic dialect continuum. He also concludes that as these words seem to be fully regular cognates, they precede the shared sound changes like satemization and the Ruki rule. Because these sound changes to some extent seem to be shared also by Albanian and

Armenian, the Indo-Slavic stage should be located already somewhere close to the Western Yamnaya region (to the west from Dnieper).

However, another interpretation is also possible: the/some Indo-Slavic words could be mutual borrowings after the satemization and the Ruki rule. It is impossible to reliably distinguish between mutual borrowings and the regular *satem*-reflexes (*$\acute{k}$, *$\acute{g}$ > *$t\acute{s}$, *$d\acute{z}$ in Late Proto-Indo-Iranian and > *$\acute{s}$, *$\acute{z}$ in Late Proto-Balto-Slavic). If words were borrowed from one branch to the other, these very same new sounds would have been the most probable substitutes for the new sounds in the other branch. The same goes with the Ruki rule, as the result was *$\check{s}$ in both branches under the same conditions (although more restricted in Baltic). Therefore, we do not know whether the shared Indo-Slavic words in both branches truly precede these shared sound changes or not. It is possible that these sound changes are after all earlier than most of the shared words, which would agree with the narrower distribution of the words.

Palmér argues convincingly that we cannot exclude the Abashevo Culture, which for some traits offers a better match for the Late Proto-Indo-Iranian vocabulary than the Sintashta Culture (Palmér 2025: 243–251). The Abashevo Culture has recently been associated with Early/Middle Proto-Indo-Iranian (Parpola 2020; Parpola 2022; Häkkinen 2023b), but Palmér considers it now chronologically mostly overlapping with the Sintashta Culture (Figure 13 in Chapter 4.2 shows also these cultures). However, more precise AMS dates and rejecting of the early outlier dates from wood/charcoal have resulted in the Sintashta chronological interval as late as 2010–1770 BCE (Grigoriev & Salugina 2021: 165). The AMS dates have shortened the chronology of cultures and narrowed the margin of error, but they are not yet as numerous as dates produced with the older method.

In any case, yet earlier stages of the Indo-Iranian development can now be associated with the Fatyanovo-Balanovo Culture which represents the easternmost Corded Ware extension. Autosomal and Y-chromosomal DNA have shown that people of the Sintashta Culture are similar to the people of the Fatyanovo Culture; Y-DNA from the Abashevo Culture also agrees here. Balto-Slavic could also have been situated there, because there are apparent loanwords from several early Balto-Slavic varieties into western and central Uralic (Kallio 2022; Häkkinen 2022) and because there were 3–8 distinct regional groups within the Fatyanovo-Balanovo Culture, covering the wide southern forest zone from river Volkhov to the Kama Region (Nordqvist & Heyd 2020: 4). In any case, some Paleo-European languages also survived there

until later times. Archaeological cultures were not necessarily monolingual, and distributions of archaeological cultures do not necessarily agree with distributions of speech communities.

To conclude, the shared words between Indo-Iranian and Balto-Slavic could be also younger, spread together or after the satemization and the Ruki rule. Currently there are still many open possibilities: the phase of adjacency of Indo-Iranian and Balto-Slavic could be situated in (1) Western Ukraine, (2) within the southwestern Corded Ware cultures, or (3) within the northeastern Corded Ware cultures. What seems clear is that earlier stages of the Indo-Iranian language lineage have spread toward the east in rather northern region, offering an explanation for the western-central distribution of the Early/Middle Proto-Indo-Iranian loanwords in Uralic.

5.2. The paternal haplogroup N

I begin the genetic considerations with the paternal haplogroup N, because there is less room for interpretations in uniparental lineages than in the Nth degree interpretations of the auto-somal data. SNP mutations are unique, and the family tree of the human Y-chromosomal DNA tells us accurately when and from where our direct paternal ancestors have arrived.

For many Uralic populations, rather close match is the sample I32545/ROT002 from Ros-tovka (~1940 BCE): N-Z1936 (Zeng, Vyazov & Kim et al. 2025). Mutation Z1936 is dated to 2450 BCE in FamilyTreeDNA Discover, and so far, this Rostovka man is the earliest known representative of this subhaplogroup. So, is it possible that he and his kin could have spread the Uralic language?

Of the tested male Finns, 42 % belong under this subhaplogroup. However, all of them (3708/3708 in FamilyTreeDNA Discover) belong under its later descendant N-Z1925 (850 BCE), which already emerged in Finland based on its distribution. Here we face two obvious complications: (1) Could one man bring a whole new language to the region? (2) The route of this man rather aligns with Saami, not Finnic.

Naturally, many uniparental lineages have gone extinct during the millennia: a paternal lineage does not continue, if a man has only daughters, and a maternal lineage does not con-tinue, if a woman has only sons. There apparently remains no paternal lineages from the Stone

Age in Finland, but all of them have arrived there later. Nevertheless, we cannot speculate outside the data that remains, and the divergence of N-Z1936 in Finland only occurred after N-Z1925, with the gap of almost two millennia between these mutations. This one forefather (even with his relatives) was hardly capable of bringing a whole new language to the region.

And in any case, his language apparently would not be an early stage of Proto-Finnic, which arrived in Finland through Estonia, but an early stage of Proto-Saami, which arrived in Finland from the southeast like N-Z1925. So, even if this one man (perhaps together with his brothers and cousins which have no living paternal descendants) could have brought a new language in the region, it was the wrong language. All his descendants who are not Saami today (this lineage is also frequent among the Saami) must have later adopted a Finnic language, which spread outside the narrow southwestern coastal zone of Finland only during the last 1700 years.

Perhaps the Finnic language carriers could be associated with another lineage: N-L550 (emerged 900 BCE) arrived in Estonia around 600 BCE (Saag et al. 2019). On the other hand, only 4 % of the Finns descend from this mutation, so in any case this could not have been the main language carrier lineage after it reached Southwestern Finland (where the frequency is 11–19 % according to Preussner et al. 2024).

Moreover, this lineage is not under N-Z1936 but under N-Y6058, a brother lineage of N-Z1936 – they diverged already after N-CTS3103 (2700 BCE). Both lineages (respectively VL29 and Z1936 in Ilumäe et al. 2016) are very widespread in Northern Eurasia, from the Baltic Sea Region to Central Siberia. Perhaps these lineages were already involved in the wide Seima-Turbino Network, but connecting them to the spread of the Uralic languages is whole another thing. Much more is needed than just an observation about partial regional overlapping to establish this connection.

Nevertheless, it is interesting to notice the internal division soon after N-Z1936, as the descendants of N-CTS1223 are divided into the western N-CTS9925 (2050 BCE; subhaplogroups frequent in the Finns, the Saami, and the Karelians, but found also in the Estonians) and the eastern N-Y13851 (2100 BCE; subhaplogroups found in the Hungarians, the Mansi, the Khanty, and the Bashkirs). Compared to the valid linguistic results, this division probably developed in the Central Ural Region, and these lineages later perhaps participated in the

expansion of the Uralic languages to the west and east at least to some extent. N-Z1925 (with its problems, see above) is under the western lineage.

However, regional correlation alone is never enough, because these subhaplogroups could have spread also before or after the Uralic languages – we cannot just decide that they spread together with the language. It only leads to circular argumentation, if we first decide that a subhaplogroup must be associated with the Uralic languages, and then we trace the steps of that subhaplogroups in order to trace the spread of the Uralic languages.

So far, all these early N-samples from Southern Siberia have belonged to the haplogroup N1a1-Tat/M46 (ISOGG Y-tree). However, the only lineage shared by all the studied Samoyed populations is under N1a2-L666, especially N-VL63/P63 (Karafet et al. 2018) which emerged around 850 BCE (FamilyTreeDNA Discover). N1a1 and N1a2 diverged already after N-L729 around 15 000 BCE, so this division has nothing to do with the divergence of the Uralic languages. Possibly the latter lineage was involved in the Samoyed expansion toward the north around a millennium later, as it is also found in the populations of the Altai Region. However, other Uralic populations do not have N-VL63, and the Samoyed populations do not have its western brother lineage N-Y3196.

Consequently, it is difficult to find a clear paternal match for the people who brought the Samoyed language from the Central Ural Region to Southern Siberia around 1500 BCE, at least within the haplogroup N. It would be equally difficult if one assumed the movement in the opposite direction, as the common Samoyed paternal lineages differ so clearly from the other Uralic populations. However, the almost two millennia before the expansion of Samoyed toward the north are enough to allow drastic changes in the genetic composition of Samoyed speakers. Genetic drift is especially strong in small populations like hunter-gatherers.

Similar problems are faced concerning the carriers of the Indo-European languages: even though the Corded Ware people are at the autosomal level mainly descended from some Yamnaya-like people, the dominating paternal lineages are totally different.

5.3. The Yakutia ancestry

First, I must emphasize that even though genetics is counted within the natural sciences and nucleotides are concrete at the molecular level, all these autosomal admixture modellings based on qpAdm and other tools are only interpretations of the Nth degree. This becomes obvious when we see that even for the very same DNA samples, every genetic study offers different interpretations. This hypotheticality does not concern uniparental lineages which represent direct observations from the nucleotide data.

When we talk about the Yakutia ancestry in the context of the Uralic populations (like in Childebayeva et al. 2024; Zeng, Vyazov & Kim et al. 2025), we mean the Yakutia ancestry first appearing during the Late Neolithic and Bronze Age within the Ymyyakhtakh Culture. This ancestry emerged from the admixture between the earlier Middle Neolithic Yakutia ancestry and migration from the southwest representing the Neolithic East Baikal ancestry (Gill, Lee & Yeong 2024).

However, the latter study states that neither of the two Yakutia ancestries is the true source for the current Uralic populations, although the actual source population was something quite like these. Consequently, we do not know where and when the true source population existed, so is there any sense to compare the Uralic languages to these only inaccurately matching ancestries? Well, we can always try; but let us not forget how little we still know. Building interpretations upon interpretations tends to lead scholars astray.

For example, the Yakutia ancestry is not restricted to the Uralic populations (unlike stated in a news article: The Harvard Gazette, July 16th, 2025), but it is found widely in Northern Eurasia, covering the whole width from the Baltic Sea Region to Northeastern Siberia (1.–4. from Zeng, Vyazov & Kim et al. 2025 | 5. from Gill, Lee & Yeong 2024 | 6. from Peltola et al. 2023):

1. The Tundra- and Forest Yukaghir (the Yukaghir language family).
2. The Yakut/Sakha, the Tuvinians, the Altaians, the Tofalar, the Todzin, the Khakass, the Tubalar, the Shor, the Nogai, the Karakalpak, the Tatar, the Chuvash, the Bashkir, the Kyrgyz, the Kazakh (the Turkic language family)

3. The Buryat, the Khamnigan, the Kalmyk, the Mongol, the Daur
   (the Mongolic language family)

4. The Oroqen, the Hezhen/Nanai, the Even, the Evenk
   (the Tungusic language family)

5. The Ket (the Yeniseian language family)

6. The Russians, the Belarussians, the Poles, the Sorbs, the Lithuanians
   (the Indo-European language family)

The Uralic language family (except the modern Hungarians) just happens to be situated within this wide region – just like both extant Yukaghir languages and the only extant Yeniseian language. Do these latter two language families not have even more perfect regional coverage with the Yakutia ancestry? Moreover, according to Zeng, Vyazov & Kim et al. (2025), the population with most of the Yakutia ancestry (98.5 %) is the non-Uralic Tundra Yukaghir, although for some reason they do not consider the Yukaghir languages in their article.

To conclude, Uralic populations make up only one part within the continuum of populations carrying the Yakutia ancestry, and there must have been several language shifts involved during the expansion of this ancestry; this will be considered in the forthcoming chapters. There is no ground to interpret that the Yakutia ancestry was originally associated with the Uralic population, because no Uralic language has ever been spoken in Yakutia.

This final point is not directed to the geneticists mentioned above but to journalists and laymen, who struggle to understand the scientific validity of interdisciplinary comparisons. Genetic methods can never reach language, because language is not inherited in DNA. Therefore, genetic results can never reliably tell anything about the past of a language – only linguistic results can. Interdisciplinary research is scientific only when this premise is accepted. Using results of a scientific discipline in an unscientific way is just pseudoscience.

5.4. Misinterpretations concerning genetic results

In the following subchapters I will shortly consider a few misinterpretations made from genetic results, stemming from insufficient knowledge about language and/or linguistics.

5.4.1. Mistaking genetic continuity as linguistic continuity

"Interestingly, evidence of the second wave of Siberian ancestry expansion, associated with Ymyyakhtakh culture, is discernible in Nganasan individuals but not in another Samoyedic-speaking population, Selkup. This suggests that the divergence within Samoyedic-speaking populations may go back to the Neolithic period." (Gill, Lee & Yeong 2024: 9)

A common misconception is to equate linguistic continuity with cultural or genetic continuity; an example of this is seen in the quote above. Such a method is entirely unreliable, because it leads to contradicting results: one scholar ties certain language to one archaeological or genetic root, while another scholar ties it to another root (examples in Häkkinen 2006). The observation quoted above indeed suggests something about the divergence of populations, but nothing about Samoyed speakers.

Unfortunately, Gill, Lee & Yeong (2024) seem to have erroneously assumed that all the genetic layers in the past of the populations which nowadays speak Samoyed languages would be associated with the Samoyed language. This is not true. Every population has a deep history of accumulating layers of ancestry components and admixture events, one after another. Even though the current population speaks a certain language, this language cannot be projected back in time as if it was tied to a certain ancestry component or an admixture of ancestry components. The current language lineage has been adopted only once at a certain moment, while ancestry components have accumulated during many millennia to a multilayered whole; see Figure 15 (from Piha, Heikkilä & Häkkinen 2024 [2022]).

A common perspective deficiency is to see only the current language of some population. If one does not acknowledge that the "same" population (although all populations constantly change by their genetic composition) possibly has spoken several different languages during its prehistory, every layer of genetic or cultural influx erroneously appears to be associated with the current language lineage. In reality, only one genetic or cultural influx (if it even is perceivable) truly was associated with the arrival of the current language, while all the other influxes were associated with some other languages or different stages of the same language

lineage. The newer genetic layers after the arrival of the current language could be primary associated with some neighboring populations, and they probably left some linguistic traces (like loanwords), even though they could not cause language shift.
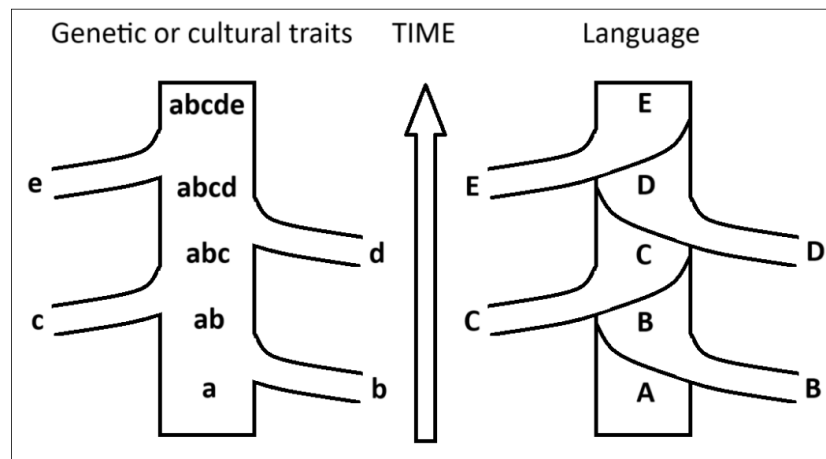

Figure 15: Different nature of language compared to autosomal genetic or cultural traits is important for understanding the restrictions of multidisciplinary comparison.

For every genetic or cultural influx there are three possibilities: (1) it brought a new language to the region; (2) it brought loanwords or another influence from some non-local language; (3) it left no visible linguistic traces. One cannot just decide that certain influx must be associated with the current language. The only scientific way to find a match for the linguistic results is to accept the linguistic results as the starting point and then see if there is an archaeological or a genetic phenomenon appearing in the right place at the right time and spreading to the right direction.

There can be several events of genetic divergence in the past of two adjacent populations, and at the same time several events of genetic convergence. Divergence occurs when only one of the populations receives new genetic material, and convergence occurs when both populations receive the same genetic material (Figure 16). It is probable that all these different influxes of new genetic material were associated with different languages/dialects. A probability to randomly guess right which of the many layers of genetic material was associated with the current language lineage spoken by these populations is very low.
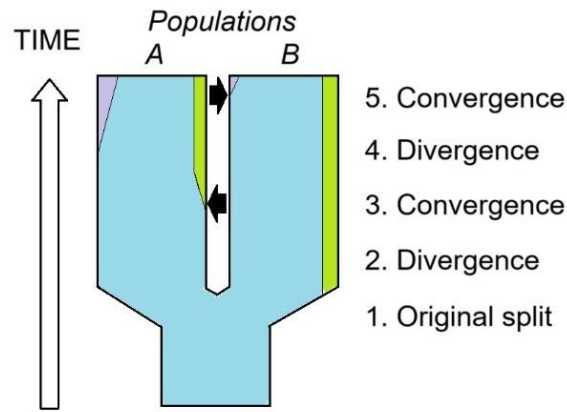
Figure 16: Several events of divergence and convergence could have occurred in the past of neighboring populations. Colors represent different ancestry components.

This is a good time to illuminate the past of the Samoyed languages based on linguistic results. Concerning the evolutionary steps of Nganasan, the northeasternmost Samoyed (and Uralic) language spoken in the Taimyr Peninsula in northern Central Siberia, the following spatio-temporal scenario can be presented:

1. Late Proto-Uralic (Central Ural Region ~2500 BCE)
2. East Uralic linguistic area (Central Ural Region ~1500 BCE)
3. Late Proto-Samoyed (Southern Siberia ~500 BCE)
4. North Samoyed (Lower Yenisei Region ~500 CE)
5. Nganasan (Taimyr Peninsula ~1500 CE)

When the genetic results show that the Nganasan have more of the Yakutia ancestry component than any other Uralic population, this result must be viewed against the framework of the linguistic results: the Nganasan only arrived in Taimyr around 500 years ago (Vesakoski, Salmela & Piezonka 2025) after many steps of advancement over vast distances. Therefore there is no basis to consider the Nganasan as a valid proxy for the Late Proto-Uralic speakers. Just the opposite – the Nganasan are an aberrant population even among the Samoyed populations, resembling also the non-Uralic populations nearby, like the Tundra Yukaghir (Zeng, Vyazov & Kim et al. 2025: Extended Data Figure 11).

With qpAdm, the Nganasan can be modelled e.g. with 59 % Yakutia ancestry + 41 % East Baikal ancestry (Gill, Lee & Yeong 2024) or with 73 % Yakutia ancestry + 21 % Mongolia ancestry + 5 % Tyumen ancestry + 0,5 % Srubnaya ancestry (Zeng, Vyazov & Kim et al. 2025). Gill et al. (2024) model the Nenets, the Enets, and the Selkup without any Baikal ancestry, and although Zeng, Vyazov & Kim et al. (2025) give also to the Enets some Mongolia ancestry, they give to the Enets and the Selkup no Tyumen ancestry but instead Eastern European Hunter-Gatherer ancestry (EHG) and more Steppe (Srubnaya) ancestry.

This suffices to demonstrate the outlier genetic composition of the Nganasan among the Samoyed and Uralic populations. Clearly the Samoyed language carriers, who arrived in the Taimyr Peninsula from the south along Yenisei only recently, have massively assimilated the local inhabitants during the recent centuries, thus acquiring a considerable portion of ancestry typical to northern Central–Eastern Siberia. At this point it should be obvious to everyone what a grave methodological error it would be to tie the Nganasan language lineage to the Yakutia ancestry and then to imagine that by following that ancestry to its region of origin one could find the homeland of Late Proto-Uralic.

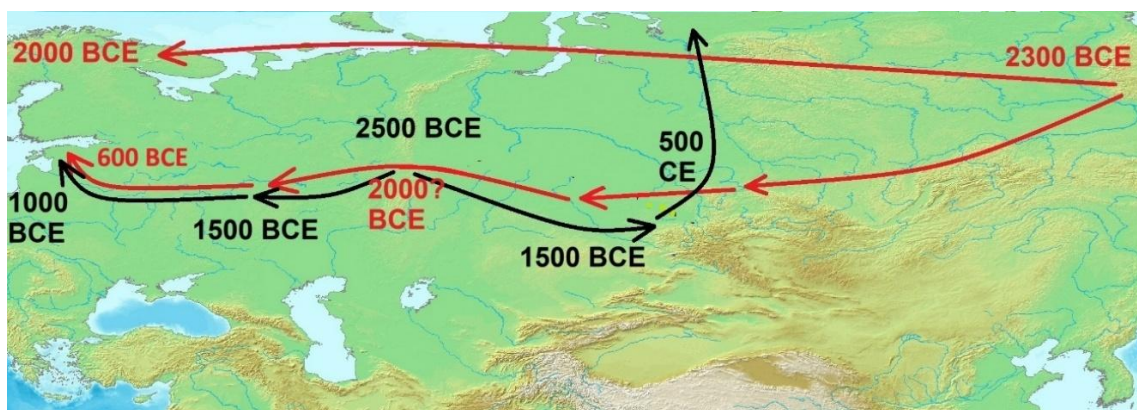Figure 17: The advancement of the Uralic language toward Estonian and Nganasan (black) compared to the spread of the Yakutia ancestry to the west (red).

We should accept the genetic result that the Taimyr population apparently has been distinct from its neighbors for several millennia. At the same time, we should also accept the linguistic results: this genetic separateness had nothing to do with the speakers of the Samoyed

languages until the carriers of the Nganasan language arrived there around 500 years ago. Figure 17 shows the difference between the expansions of the Uralic languages (focusing on Estonian in the west and Nganasan in the east) and the Yakutia ancestry. As it currently seems that the Yakutia ancestry crossed over the Urals around 2000 BCE, it probably was present (among other ancestries) during the Uralic expansion somewhat later. But before that it could not have anything to do with the Uralic languages, because there is no match concerning the time, the place, and the direction of expansion.

5.4.2. Anachronistic and confusing use of linguistic labels

"The [Hungarian] Conquerors, who arrived in the Carpathian Basin after the Avars, had a distinct genomic background with elevated levels of western Eurasian admixture. Their core population carried very similar genomes to modern Bashkirs and Tatars, in agreement with our previous results from uniparental markers. Their genomes were shaped by several admixture events, of which the most fundamental was the Mezhovskaya-Nganasan admixture around the late Bronze Age, leading to the formation of a "proto-Ugric" gene pool." (Maróti et al. 2022: 2867)

"Nganasan" is a linguistic label denoting people speaking Nganasan, the northeasternmost Samoyed (and Uralic) language. As the independent evolution of the Nganasan language only covers ~2000 years at most, it is anachronistic to talk about (the) Nganasan at the Late Bronze Age context. Especially when we are talking about genetic admixture, the labels should be distinct from the labels of non-genetic levels. "Nganasan-like ancestry/population" would already be less misleading.

This terminological miss aside, the results by Maróti et al. (2022) are very interesting, and their reconstruction of genetic events seems plausible, even though it remains hypothetical due to the lack of ancient DNA from all the relevant regions and/or chronological stages. A great portion of the Nganasan-like ancestry (nowadays the Yakutia ancestry) could have been acquired when Ugric speakers spread from the Southern Ural Region (assumedly from the

Mezhovskaya Culture) to the east. Still, the people of the Mezhovskaya Culture already had some Yakutia ancestry (Török 2023: 3).

5.4.3. Genetic methods cannot reach language

"Our paper helps show that the latter scenario [Southern Siberian homeland for Late Proto-Uralic] is more likely. - - We can see this genetic pulse coming from the east just as Uralic languages were expanding." (Tian Chen Zeng to the Harvard Gazette, 16th July 2025)

First, genetics only has methods for reaching DNA, and language is not inherited in DNA: it is learnt from people around a child, irrespective of whether they are genetically related. There is no scientific basis to believe that genetic results could ever make any linguistic homeland solution more probable than another, when there are conclusive linguistic results available.

Genetic results can only agree or disagree with linguistic results, and even if they disagree, they can never question or disprove the linguistic results. In a disagreement situation we just must admit that language has spread without a currently identifiable genetic trace. We can take any old and widespread language family (like Indo-European or Uralic) and see how difficult it is to find clear genetic traces explaining all the linguistic expansions within the language family. During every step of linguistic expansion, the genetic composition of the language carriers has changed (uninhabited islands being the only exception).

In this case, when Zeng sees a match between the spread of the Yakutia ancestry and the Uralic languages, that is a mere result of tunnel vision toward Uralic. As I have explained in previous chapters: (1) this genetic pulse began from Yakutia, where Uralic languages were never spoken; (2) Zeng ignores the Yukaghir languages and all possible ancient languages spoken in Yakutia at the relevant time; (3) this genetic pulse can be connected to the Uralic languages only where the valid linguistic results allow it – in the Central Ural Region around 2000 BCE.

After all, Zeng, Vyasov & Kim et al. (2025) do not assume that the Yakutia ancestry spread together with the Uralic language lineage since the beginning (even though the news article in the Harvard Gazette implies this). This means that they assume a language shift on the way toward the west: in Southern Siberia. When a language shift in any case is inevitable, then it should be set within the coordinates agreeing with the valid linguistic results.

In their defense, Zeng, Vyazov & Kim et al. (2025) accepted the linguistic results as their starting point, and they did not claim (at least in their article) that the Uralic languages spread to the west from Yakutia. Unfortunately, the linguistic results they followed were invalid, and therefore their results are just a blind alley.

To conclude: the only valid assumption we could make is to accept the valid linguistic results and associate the Yakutia ancestry with the Uralic languages in the Central Ural Region around 2000 BCE. What was the language of the Yakutia ancestry carriers before that in Southern Siberia or in Northeastern Siberia, that is another story. But this we know: it was not Late Proto-Uralic or Common Uralic.

5.5. Conclusion from interdisciplinary considerations

If the aim of interdisciplinary research is to find archaeological or genetic matches for language, then the valid linguistic results must be taken as the starting point; otherwise, the method only leads to illusory associations. In this comment I have shown that the valid linguistic evidence supports only the Late Proto-Uralic homeland in the Central Ural Region and that there probably were at least five language lineages in or near Southern Siberia and the Altai Region during the late 3rd millennium BCE: Para-Uralic, Middle Proto-Tocharian, Pre-Proto-Yukaghir, Pre-Proto-Yeniseian, and Paleo-West Siberian.

Whichever combination of languages was involved in the emergence of the Seima-Turbino Network in that region, it could not include Late Proto-Uralic or Common Uralic. This is the starting point for any further interdisciplinary considerations. Instead, the crossing of the STN over the Urals to Europe around 2000 BCE could have involved Uralic speech communities. Around this time also the paternal haplogroup N and the Yakutia ancestry appear to have crossed over the Urals, possibly connected to the STN.

6. Summary

In Chapter 1.2, I argued for the presence of the ninth vowel *i̯ in Late Proto-Uralic, concerning the Finnic back vowel combination *i-a (vs. *i-ä). This is based on different mergers in different Uralic branches and unique vowel combinations in Mordvin and Ob-Ugric. Consequently, this removes the only context for the opposition between the second syllable *a and *ä, thus removing the vowel harmony as a phonological feature from Late Proto-Uralic. This change requires that the traditional Uralic *i̯ must now be reinterpreted as *e̮, like some scholars have already consistently done.

In Chapter 2, I evaluated the arguments interpreted to support the Late Proto-Uralic homeland in the Altai Region. The arguments were:

1. The Ural-Altaic typological properties.
2. The taxonomy of the Uralic language family, based on the basic numerals.
3. The proposed Pre-Proto-Tocharian loanwords in Uralic.
4. The proposed Uralic influence on the phonological and morphological system of Tocharian.

None of these arguments necessarily concern Late Proto-Uralic, because they are chronologically vague and can be interpreted within a wide chronological range. Therefore, no valid support was found for the Late Proto-Uralic homeland in the Altai Region or in Southern Siberia. Instead, two loanwords proposed by Bjørn (2022) could be interpreted as Early Proto-Tocharian loanwords into Early Proto-Uralic: LPU *käliw 'brother/sister-in-law' and *nimi 'name'.

Additionally, I proposed two more Early Proto-Tocharian loanwords borrowed into Early Proto-Uralic: LPU *pexi- 'cook' and *je̮xi- 'drink', reflecting a novel sound substitution EPT *k^w → EPU *$x^{(w)}$ > LPU *x. Increasing evidence concerning the contact between Early Proto-Tocharian and Early Proto-Uralic supports locating Early Proto-Uralic in Southern Siberia during the first half of the 3rd millennium BCE.

In Chapter 3, I evaluated the arguments interpreted as supporting the Late Proto-Uralic homeland in the Central Ural Region applying the same measure as in Chapter 2. The arguments were:

1. The early Indo-Iranian loanwords layers, divided into Early/Middle and Late Proto-Indo-Iranian

2. The Uralic bundle effect, consisting of two parts: A. The arbitrary sound substitutions for Late Proto-Indo-Iranian *a shared by all the Uralic branches, including Samoyed | B. The East Uralic sibilant changes

3. Late Proto-Uralic *sẹksa 'Siberian pine

4. Late Proto-Uralic *ćẹli 'elm'

5. Late Proto-Uralic *mekši 'bee' and *meti 'honey, nectar'

All these arguments agree with the Late Proto-Uralic homeland in the Central Ural Region. Although the last two arguments ('elm', 'bee', and 'honey, nectar') are not chronologically accurate and therefore cannot directly support the homeland in the Central Ural Region, they are still valid arguments to exclude the homeland in Western or Southern Siberia, because the distribution of these words in Uralic does not agree with the spread from Southern Siberia over the elmless and honeybeeless Western Siberia to Europe.

Additionally, several topics were considered in subchapters. In Chapter 3.1, I considered inclusion or exclusion of several Early/Middle Proto-Indo-Iranian loanwords, resulting in total lack of loanwords in this layer in Mansi, Khanty, and Samoyed. Therefore, conclusion from this updated dataset replaces the earlier interpretation that Early/Middle Proto-Indo-Iranian loanwords were lacking only in Samoyed (Chapter 3.7).
 I also argued for a new vowel substitution on phonetic ground: LPIIr *iš → U *i̯š.

In Chapter 3.2 concerning the East Uralic sibilant changes, I argued that on the phonetic ground *L is more plausible intermediary sound than *ϑ and accepted that the updated loanword data indeed requires the merger *š > *s instead of the other way round, as I earlier thought. I also updated the stages of the East Uralic sibilant changes according to the exceptions from the recent data, resulting in a reconstruction of six successive stages.

This new scenario also allows new interpretations concerning early Indo-Iranian loanwords: at least some of the loanwords borrowed from Indo-Iranian *tś*, *dź* and showing the secondary Mansi *š* seem to be contemporaneous with the Proto-Iranian loanwords. Thus, the donor language could not be Late Proto-Indo-Iranian proper but Archaic Indo-Iranian. I also disproved the interpretation that the sibilant changes were later in Khanty than in the other East Uralic branches.

In Chapter 4, I considered two additional phenomena regarding the linguistic situation in Southern Siberia around the time of the early Uralic divergence. These were:

1. The Uralic loanwords in Yukaghir.
2. The Ugro-Yeniseian local case series.

The local case series apparently was adopted both into the Ugric branches and Northern Yeniseian (Ket-Yugh) from Paleo-West Siberian, because no mutual loanwords between Ugric and Yeniseian have been identified. Yukaghir loanwords have been proposed in Mansi and Khanty, and Yeniseian placenames have been identified in the Kama Region, to the west from the Central Ural Region. These linguistic expansions could have been spread together with the Seima-Turbino Network.

Consequently, there were at least five different language lineages in or near Southern Siberia around 2000 BCE: Para-Uralic, Middle Proto-Tocharian, Pre-Proto-Yukaghir, Pre-Proto-Yeniseian, and Paleo-West Siberian. Late Proto-Uralic was not anywhere near that region, so it cannot have anything to do with the early stage of the Seima-Turbino Network.

In Chapter 5, I underlined common problems in interdisciplinary research. We can never reliably see language from DNA or linguistic continuity from genetic continuity, because language is not inherited in DNA. The only scientific way to do interdisciplinary research is to accept the autonomy of all disciplines concerned: only linguistics can study language, only archaeology can study material culture, and only genetics can study DNA. Archaeological or genetic results can never disprove or replace linguistic results.

7. Literature

AgroAtlas.ru = Afonin, A.N. & Greene, S.L. & Dzyubenko, N.I. & Frolov, A.N. (eds.). 2008. *Interactive Agricultural Ecological Atlas of Russia and Neighboring Countries*. Economic Plants and their Diseases, Pests and Weeds [Online]. Available at: http://www.agroatlas.ru

Aikio, Ante. 2004. An essay on substrate studies and the origin of Saami. In Hyvärinen, Irma & Kallio, Petri & Korhonen, Jarmo (eds.), *Etymologie, Entlehnungen und Entwicklungen. Festschrift für Jorma Koivulehto zum 70. Geburtstag*, 5–34. Mémoires de la Société Néophilologique de Helsinki LXIII. https://www.academia.edu/4811770/An_Essay_on_Substrate_Studies_and_the_Origin_of_Saami

Aikio, Ante. 2014a. The Uralic-Yukaghir lexical correspondences: genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62, 7–76. https://journal.fi/fuf/article/download/86078/44958/131067

Aikio, Ante. 2014b. Studies in Uralic etymology III: Mari etymologies. *Linguistica Uralica* 2 / 2014, 81–93. https://kirj.ee/linguistica-uralica-publications/?filter%5Byear%5D=2014&filter%5Bissue%5D=235&filter%5Bpublication%5D=1778&v=f0aa03aaca95

Aikio, Ante. 2014c. Studies in Uralic etymology II: Finnic etymologies. *Linguistica Uralica* 1 / 2014, 1–19. https://kirj.ee/linguistica-uralica-publications/?filter%5Byear%5D=2014&filter%5Bissue%5D=230&filter%5Bpublication%5D=1729&v=f0aa03aaca95

Aikio, Ante. 2015. The Finnic 'secondary e-stems' and Proto-Uralic vocalism. *Journal de la Société Finno-Ougrienne* 95, 25–66. https://journal.fi/susa/article/download/82642/41883/

Aikio, Ante. 2020. *Uralic Etymological Dictionary* (draft version of entries A–Ć). https://www.academia.edu/41659514/URALIC_ETYMOLOGICAL_DICTIONARY_draft_version_of_entries_A_%C4%86_

Aikio, Ante. 2022. Proto-Uralic. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford University Press. https://www.academia.edu/40193033/Proto_Uralic

Akulov, Alexander & Efimova, Nadya. 2019. Hydronymy of Yeniseian origin in the basin of Oka river. *Cultural Anthropology and Ethnosemiotics* Vol. 5, № 4 (December 2019). https://www.academia.edu/41441720/Hydronymy_of_Yeniseian_origin_in_the_basin_of_Oka_river

Anthony, David W. 2007. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton University Press.

Bartens, Raija. 1999. *Mordvalaiskielten rakenne ja kehitys*. Suomalais-Ugrilaisen Seuran toimituksia 232. https://www.sgr.fi/sust/sust232/Bartens-1999-Mordvalaiskielten-rakenne-ja-kehitys.pdf

Bjørn, Rasmus G. 2022. Indo-European loanwords and exchange in Bronze Age Central and East Asia. *Evolutionary Human Sciences* 4, e23, 1–24. https://doi.org/10.1017/ehs.2022.16

Blyakharchuk, Tatyana A. 2003. Four new pollen sections tracing the Holocene vegetational development of the southern part of the West Siberian Lowland. *The Holocene* 13,5, 715–731. https://www.researchgate.net/publication/249868881_Four_new_pollen_sections_tracing_the_Holocene_vegetational_development_of_the_southern_part_of_the_West_Siberan_Lowland

Cantera, Alberto. 2017. The phonology of Iranian. Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.). *Handbook of Comparative and Historical Indo-European Linguistics*. Volume 1. 481–503. Berlin/Boston: Walter De Gruyter.

Chernykh, E. N. 2012. The Northern Hunters, Fishers and Gatherers and their Southern Neighbours in Eurasia in the Early Metal Age. *Networks, Interaction and Emerging Identities in Fennoscandia and Beyond*. Tromsø, Norway, October 13–16th, 2009. Mémoires de la Société Finno-Ougrienne 265, 105–122. https://www.sgr.fi/sust/sust265/sust265_chernykh.pdf

Chernykh, E. N. & Korochkova, O. N. & Orlovskaya, L. B. 2017. Issues in the Calendar Chronology of the Seima-Turbino Transcultural Phenomenon. *Archaeology, Ethnology & Anthropology of Eurasia* 45/2, 45–55. https://journal.archaeology.nsc.ru/jour/article/view/266/351

Childebayeva, Ainash & Fricke, Fabian & Rohrlach, Adam Benjamin & Huang, Lei & Schiffels, Stephan & Vesakoski, Outi & Mannermaa, Kristiina & Semerau, Lena & Aron, Franziska & Solodovnikov, Konstantin & Rykun, Marina & Moiseyev, Vyacheslav & Khartanovich, Valery & Kovtun, Igor & Krause, Johannes & Kuzminykh, Sergey & Haak, Wolfgang. 2024. Bronze age Northern Eurasian genetics in the context of development of metallurgy and Siberian ancestry. *Nature Communications Biology* (2024) 7:723. https://www.nature.com/articles/s42003-024-06343-x

Derksen, Rick. 2015. *Etymological Dictionary of Baltic Inherited Lexicon*. Leiden Indo-European Etymological Series, Volume 13. Leiden/Boston: Brill.

Eek, Arvo & Meister, Einar. 1994. Acoustics and perception of Estonian vowel types. *PERILUS* No. XVIII, 55–90. https://www.academia.edu/18755123/Acoustics_and_perception_of_Estonian_vowel_types

Epimakhov, Andrey & Lubotsky, Alexander. 2023. Fire and water: The Bronze Age of the Southern Urals and the Rigveda. *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics,*

and Linguistics*, 263–274. Cambridge University Press. https://scholarlypublications.universi-teitleiden.nl/access/item%3A3754227/download

FamilyTreeDNA Discover. https://discover.familytreedna.com/

Georg, Stefan. 2007. *A Descriptive Grammar of Ket (Yenisei-Ostyak). Volume I Introduction, Phonology, Morphology*. Languages of Asia, volume 1. Leiden: Brill.

Georg, Stefan. 2023. Connections between Uralic and other language families. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic languages*. New York: Routledge. https://www.aca-demia.edu/122393200/Connections_between_Uralic_and_other_language_families

Gill, Haechan & Lee, Juhyeon & Yeong, Choongwon. 2024. Reconstructing the Genetic Relationship between Ancient and Present-Day Siberian Populations. *Genome Biology and Evolution* 16(4), Volume 16, Issue 4, April 2024, evae063. https://academic.oup.com/gbe/arti-cle/16/4/evae063/7634480

Gordon, Matthew & Barthmaier, Paul & Sands, Kathy. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, Vol. 32, No. 2 (December 2002), 141–174.

Gordon, Matthew & Ghushchyan, Edita & McDonnell, Bradley & Rosenblum, Daisy & Shaw, Patricia A. 2012. Sonority and central vowels: A cross-linguistic phonetic study. In Steve Parker (ed.), *The Sonority Controversy*. De Gruyter Brill.

Grigoriev, Stanislav. 2021. Andronovo Problem: Studies of Cultural Genesis in the Eurasian Bronze Age. *Open Archaeology* 2021. 7, 3–36. https://doi.org/10.1515/opar-2020-0123

Grigoriev, Stanislav & Salugina, Natalia. 2021. Formation of Ceramic Traditions of the Late Bronze Age in the Trans-Urals. *Geoarchaeology and Archaeological Mineralogy* – 2021, 163–170.

Grünthal, Riho & Heyd, Volker & Holopainen, Sampsa & Janhunen, Juha A. & Khanina, Olesya & Miestamo, Matti & Nichols, Johanna & Saarikivi, Janne & Sinnemäki, Kaius. 2022. Drastic demographic events triggered the Uralic spread. Diachronica. Volume 39. Issue 4. 490–524. https://doi.org/10.1075/dia.20038.gru

Hackstein, Olav. 2017. The phonology of Tocharian. In Wiegand, Herbert Ernst (ed.), *Handbook of Comparative and Historical Indo-European Linguistics*, Volume 2, 1304–1334. Berlin/Boston: Walter De Gruyter.

Holopainen, Sampsa. 2019. *Indo-Iranian Borrowings in Uralic. Critical Overview of the Sound Substitutions and Distribution Criterion*. Doctoral dissertation, University of Helsinki. <https://helda.helsinki.fi/handle/10138/307582>

Holopainen, Sampsa. 2020a. Indo-Iranian loanwords in Finnic: A critical overview. *Acta Linguistica Petropolitana* 2020, Vol 16.3, 613–668. https://alp.iling.spb.ru/static/alp_XVI_3/19.pdf

Holopainen, Sampsa. 2020b. Reflexes of Indo-Iranian laryngeals in loanwords to Proto-Uralic. Presentation.

Holopainen, Sampsa. 2023. The earliest Indo-Iranian loans in Uralic: problems with sound substitutions. Handout, LMU, 03.07.2023.

Holopainen, Sampsa & Aikio, Ante. 2023. The development of sibilants and the divergence of Ugric. Presentation. Language Diversification within Uralic, University of Helsinki (Oct 26, 2023).

Honti, László. 2006. Eräästä ugrilaisten kielten postpositioperäisestä kaasussuffiksien perheestä. Journal de la Société Finno-Ougrienne 91, 81–91.

Häkkinen, Jaakko. 2006. Studying the Uralic proto-language. (Translated from: Uralilaisen kantakielen tutkiminen. *Tieteessä tapahtuu* 1 / 2006.) https://www.alkuperasivusto.fi/Studying_the_Uralic_proto-language.pdf

Häkkinen, Jaakko. 2007. Kantauralin murteutuminen vokaalivastaavuuksien valossa. Pro gradu -työ, Helsingin yliopiston Suomalais-ugrilainen laitos. https://helda.helsinki.fi/handle/10138/19684

Häkkinen, Jaakko. 2009. Kantauralin ajoitus ja paikannus: perustelut puntarissa. *Journal de la Société Finno-Ougrienne* 92, 9–56. http://www.sgr.fi/susa/92/hakkinen.pdf

Häkkinen, Jaakko. 2012a. Early contacts between Uralic and Yukaghir. In Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (eds.), *Per Urales ad Orientem : Iter polyphonicum multilingue*. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012. *Mémoires de la Société Finno-Ougrienne* 264, 91–101. http://www.sgr.fi/sust/sust264/sust264_hakkinenj.pdf

Häkkinen, Jaakko. 2012b. After the protolanguage: Invisible convergence, false divergence and boundary shift. Finnisch-Ugrische Forschungen 61, 7–28. https://journal.fi/fuf/article/view/85674/44638

Häkkinen, Jaakko. 2019. Kantasuomen keskivokaalit: paluu. In Santeri Junttila & Juha Kuokkala (eds.), Petri Kallio rocks : Liber semisaecularis 7.2.2019. Kallion etymologiseura. https://helda.helsinki.fi/handle/10138/305150

Häkkinen, Jaakko. 2022. Recurring irregularities in West Uralic 1: Para-Slavic loanwords. In Kaheinen, Kaisla & Leisiö, Larisa & Erkkilä, Riku & Qiu, Toivo E. H. (eds), *Hämeenmaalta Jamalille: kirja Tapani Salmiselle 07.04.2022*, 127–144. HELDA Open Books 2022. https://helda.helsinki.fi/handle/10138/342413

Häkkinen, Jaakko. 2023. An old Uralic compound metal name. In Holopainen, Sampsa & Kim, Jeongdo & Metsäranta, Niklas (eds.), Elämä ja etymologia. Janne Saarikiven 50-vuotisjuhlakirja, 102–121. HELDA Open Books 2023. https://helda.helsinki.fi/handle/10138/358405

Häkkinen, Jaakko. 2023. On locating Proto-Uralic. *Finnisch-Ugrische Forschungen* 68, 43–100. https://journal.fi/fuf/article/view/120910

Ilumäe, Anne-Mai & Reidla, Maere & Chukhryaeva, Marina & Järve, Mari & Post, Helen & Karmin, Monika & Saag, Lauri & Agdzhoyan, Anastasiya & Kushniarevich, Alena & Litvinov, Sergey & Ekomasova, Natalya & Tambets, Kristiina & Metspalu, Ene & Khusainova, Rita & Yunusbayev, Bayazit & Khusnutdinova, Elza K. & Osipova, Ludmila P. & Fedorova, Sardana & Utevska, Olga & Koshel, Sergey & Balanovska, Elena & Behar, Doron M. & Balanovsky, Oleg & Kivisild, Toomas & Underhill, Peter A. & Villems, Richard & Rootsi, Siiri. 2016. Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. The American Journal of Human Genetics 99, 163–173.

Ilyasov, Rustem A. & Lee, Myeong-lyeol & Takahashi, Jun-ichi & Kwon, Hyung Wook & Nikolenko, Alexey G. 2020. A revision of subspecies structure of western honey bee Apis mellifera. *Saudi Journal of Biological Sciences* 27 (2020), 3615–3621.

ISOGG. Y-DNA Haplogroup Tree 2019–2020. Version: 15.73. July 11th, 2020. https://isogg.org/tree/

Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77, 219–271.

Janhunen, Juha. 2000. Reconstructing Pre-Proto-Uralic typology spanning the millennia of linguistic evolution. Congressus Nonus Internationalis Fenno-Ugristarum, 7.–13.8.2000 Tartu, 59–76.

Janhunen, Juha. 2009. Proto-Uralic – what, where, and when? *The Quasquicentennial of the Finno-Ugrian Society.* Mémoires de la Société Finno-Ougrienne 258, 57–78.

Janhunen, Juha. 2015. A legkeletibb uráliak. Székfoglalók a Magyar Tudományos Akadémián.

Janhunen, Juha. 2022. ВЕЛИКОЕ ПРОШЛОЕ МАЛЫХ НАРОДОВ (НА ПРИМЕРЕ САМОДИЙЦЕВ). Археология Евразийских степей, (2), 283–289. https://doi.org/10.24852/2587-6112.2022.2.283.289

Kaheinen, Kaisla. 2023. *Etymologia ex silentio : Nganasanin äännehistoria ja kielikontaktit*. Dissertationes Universitatis Helsingiensis 111/2023.

Kallio, Petri. 2004. Tocharian loanwords in Samoyed? In Hyvärinen, Irma & Kallio, Petri & Korhonen, Jarmo (eds.), Etymologie, Entlehnungen und Entwicklungen. Festschrift für Jorma Koivulehto zum 70. Geburtstag, 129–137. Mémoires de la Société Néophilologique de Helsinki, LXIII.

Kallio, Petri. 2012a. The Non-initial-syllable Vowel Reductions from Proto-Uralic to Proto-Finnic. In Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (eds.), *Per Urales ad Orientem : Iter polyphonicum multilingue*. Festskrift tillägnad Juha Janhunen på hans sextioårsdag

den 12 februari 2012. Mémoires de la Société Finno-Ougrienne 264, 163–175. https://www.sgr.fi/sust/sust264/sust264_kallio.pdf

Kallio, Petri. 2012b. Jälkitavujen diftongit kantasuomessa. *Fenno-Ugrica Suecana Nova Series* 14, 31–40. Stockholm 2012.
http://www.bafity.su.se/polopoly_fs/1.159534.1387186906%21/menu/standard/file/FUS2012-14_07b-Kallio-article_corrected.pdf

Kallio, Petri. 2014. The Diversification of Proto-Finnic. In Ahola, Joonas & Frog (eds.), *Fibula, Fabula, Fact : The Viking Age in Finland*, 155–168. Studia Fennica Historica 18. https://www.academia.edu/10172149/The_Diversification_of_Proto_Finnic

Kallio, Petri. 2022. The Proto-Balto-Slavic Loanword Stratum in the Uralic Languages. (Presentation at the workshop Areal Effects in Prehistoric Contacts between Uralic and Indo-European, Congressus XIII Internationalis Fenno-Ugristarum, August 22–23, 2022, University of Vienna.)

Kannisto, Artturi. 2013. *Wogulisches Wörterbuch*. Bearbeitet von Vuokko Eiras, herausgegen von Arto Moisio. Lexica Societatis Fenno-Ugricae XXXV.

Karafet, Tatiana M. & Osipova, Ludmila P. & Savina, Olga V. & Hallmark, Brian & Hammer, Michael F. 2018. Siberian genetic diversity reveals complex origins of the Samoyedic-speaking populations. *American Journal of Human Biology* 2018, Volume 30, Issue 6. https://doi.org/10.1002/ajhb.23194

Kim, Ronald. 2000. 'To drink' in Anatolian, Tocharian, and Proto-Indo-European. Historische Sprachforschung / Historical Linguistics, 113. Bd., 1./2. H., 151–170.

Koivulehto, Jorma. 1988. Lapin ja itämerensuomen suhteesta. Ieur. -Tr- yhtymän korvautuminen lainoissa. *Virittäjä* 92 Nro 1.

Koivulehto, Jorma. 1991. *Uralische Evidenz für die Laryngaltheorie*. Österreichische Akademie der Wissenschaften. Philosophisch-Historische Klasse. Sitzungsberichte, 566. Band. Wien 1991.

Kokkelmans, Joachim. 2021. *The Phonetics and Phonology of Sibilants : A Synchronic and Diachronic OT Typology of Sibilant Inventories*. Doctoral dissertation. University of Verona, Department of Foreign Languages and Literatures.

Korhonen, Mikko. 1981. Johdatus lapin kielen historiaan. Suomalaisen Kirjallisuuden Seuran toimituksia 370.

Korhonen, Mikko. 1988. Uralilaisten kielten jälkitavujen vokaaliston historiaa. *Virittäjä* Vol 92, Nro 1 (1988), 8–24. https://journal.fi/virittaja/article/download/38186/46502/

Koryakova, Ludmila & Epimakhov, Andrej. 2006. *The Urals and Western Siberia in the Bronze and Iron Ages*. Cambridge World Archaeology. Cambridge University Press.

Kroonen, Guus & Jakob, Anthony & Palmér, Axel I.& van Sluis, Paulus & Wigman, Andrew. 2022. Indo-European cereal terminology suggests a Northwest Pontic homeland for the core Indo-European languages. *PLoS ONE* 17(10): e0275744. https://doi.org/10.1371/journal.pone.0275744

Lang, Valter. 2020. *Homo Fennicus : Itämerensuomalaisten etnohistoria*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lapteva, E. G. & Zaretskaya, N. E. & Kosintsev, P. A. & Lychagina, E. L. & Chernov, A. V. 2017. First Data on the Middle to Late Holocene Dynamics of Vegetation in the Upper Kama Region. *Russian Journal of Ecology* 2017. Vol. 48, No. 4, 326–334.

Lapteva, E. G. & Korona, O. M. & Zhilin, M. G. 2020. The Early Holocene vegetation changes in the vicinity of the Gorbunovo peat bog in the Middle Urals (Russia). *IOP Conference Series: Earth and Environmental Science* 438 012013.

Lehtiranta, Juhani. 2015 [1989]. Yhteissaamelainen sanasto. Mémoires de la Société Finno-Ougrienne 200. https://www.sgr.fi/sust/sust200.pdf

Lubotsky, Alexander. 2018. The phonology of Proto-Indo-Iranian. Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), Handbook of Comparative and Historical Indo-European Linguistics Volume 3, 1875–1888. Berlin/Boston: Walter De Gruyter.

Lubotsky, Alexander. 2023. Indo-European and Indo-Iranian wagon terminology and the date of the Indo-Iranian split. *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics, and Linguistics*, 257–262. Cambridge University Press.

Mallory, J.P. 2015. The Problem of Tocharian Origins: An Archaeological Perspective. *Sino-Platonic Papers* 259.

Mallory, J.P. & Adams, D.Q. 2006. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford / New York: Oxford University Press.

Maróti, Zoltán & Neparáczki, Endre & Schütz, Oszkár & Maár, Kitti & Varga, Gergely I.B. & Kovács, Bence & Kalmár, Tibor & Nyerki, Emil & Nagy, István & Latinovics, Dóra & Tihanyi, Balázs & Marcsik, Antónia & Pálfi, György & Bernert, Zsolt & Gallina, Zsolt & Horváth, Ciprián & Varga, Sándor & Költő, László & Raskó, István & Nagy, Péter L. & Balogh, Csilla & Zink, Albert & Maixner, Frank & Götherström, Anders & George, Robert & Szalontai, Csaba & Szenthe, Gergely & Gáll, Erwin & Kiss, Attila P. & Gulyás, Bence & Kovacsóczy, Bernadett Ny. & Gál, Szilárd Sándor & Tomka, Péter & Török, Tibor. 2022. The genetic origin of Huns, Avars, and conquering Hungarians. *Current Biology*, Volume 32, Issue 13, 2858–2870.e7.

Matveyev, A. V. 2011. Материалы по мансийской топонимии горной части Северного Урала. [Materials on Mansi toponymy in the mountainous part of the Northern Urals.] Министерство

образования и науки Российской Федерации Государственное образовательное учреждение высшего профессионального образования.

Meister, Einar & Meister, Lya. 2019. Production of Estonian vowels by Finnish speakers. *ESUKA – JEFUL* 2019, 10–1: 129–143.

Nichols, Johanna. 2021. The origin and dispersal of Uralic: Distributional typological view. *Annual Review of Linguistics 2021* 7, 351–369.

Nikolaeva Nikolaeva, Irina 2006: *A historical dictionary of Yukaghir*. Trends in linguistics. Documentation 25. Berlin: Mouton de Gruyter.

Nordqvist, Kerkko & Heyd, Volker. 2020. The Forgotten Child of the Wider Corded Ware Family: Russian Fatyanovo Culture in Context. *Proceedings of the Prehistoric Society*. Published by Cambridge University Press on behalf of The Prehistoric Society.

Paasonen, Heikki. 1990–1999. *Mordwinisches Wörterbuch*. Zusammengestellt von Kaino Heikkilä. Bearbeitet und herausgegeben von Martti Kahla. Lexica Societatis Fenno-Ugricae XXIII: 1–6 / Kotimaisten kielten tutkimuskeskuksen julkaisuja 59. https://www.mv.helsinki.fi/home/rueter/PaasonenMW.shtml

Pagel, Mark & Atkinson, Quentin D. & Meade, Andrew. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449, 717–720.

Palmér, Axel I. 2025. *Indo-Slavic Lexical Isoglosses and the Prehistoric Dispersal of Indo-Iranian*. Leiden Studies in Indo-European 26.

Park, Go Eun & Lee, Don Koo & Kim, Ki Woo & Batkhuu, Nyam-Osor & Tsogtbaatar, Jamsran & Zhu, Jiao-Jun & Jin, Yonghuan & Park, Pil Sun & Hyun, Jung Oh & Kim, Hyun Seok. 2016. Morphological Characteristics and Water-Use Efficiency of Siberian Elm Trees (Ulmus pumila L.) within Arid Regions of Northeast Asia. *Forests* 2016, 7, 280.

Parpola, Asko. 2020. Royal "chariot" burials of Sanauli near Delhi and archaeological correlates of prehistoric Indo-Iranian languages. *Studia Orientalia Electronica* Volume 8(1) (2020), 175–198. https://journal.fi/store/article/view/98032

PAST = Hammer, Ø., Harper, D.A.T., Ryan, P.D. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4(1): 9pp.

Peyrot, Michaël. 2018. Tocharian agricultural terminology: Between inheritance and language contact. In Kroonen, Guus & Mallory, J.P. & Comrie, Bernard (eds.), *Talking Neolithic: Proceedings of the workshop on Indo-European origins held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, December 2–3, 2013*, 242–277. Journal of Indo-European Studies Monograph Series no. 65. Washington DC: Institute for the Study of Man.

Peyrot, Michaël. 2019a. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-European Linguistics* 7 (2019), 72–121. https://brill.com/view/journals/ieul/7/1/article-p72_3.xml?srsltid=AfmBOoqcaDwUrIWD-pUCfRv1aFLDCeeolweyAJkph1LwNxfYYeljbL9nG

Peyrot, Michaël. 2019b. Chapter 13 Indo-Uralic, Indo-Anatolian, Indo-Tocharian. *The Precursors of Proto-Indo-European*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789004409354_014

Peyrot, Michaël. 2022. 6 Tocharian. In: Olander, T. (ed.), *The Indo-European Language Family*, 83–101. Cambridge University Press.

Piezonka, Henny & Salmela, Elina & Vesakoski, Outi. 2024. Kommunikationsräume und interkulturelle Netzwerke im bronzezeitlichen Sibirien: Ein neuer Blick auf Sejma-Turbino, den „Flowerpot Complex" und die Ausbreitung der uralischen Sprachen. *Prähistorische archäologie in Südosteuropa*, Band 33, 509–424. Begründet von Bernhard Hänsel. Herausgegeben von Elke Kaiser, Henny Piezonka, und Wolfram Schier.

Piha, Minerva & Heikkilä, Mikko & Häkkinen, Jaakko. 2024 [2022]. Comment on the article Archaeology, Language, and the Question of Sámi Ethnogenesis. *Acta Archaeologica* 93.2 (2022), 456–470. https://www.academia.edu/114870366/Comment_on_the_article_Archaeology_Language_and_the_Question_of_S%C3%A1mi_Ethnogenesis

Poliakov, A. V. & Svyatko, S. V. & Stepanova, N. F. 2019. A Review of the Radiocarbon Dates for the Afanasyevo Culture (Central Asia): Shifting Towards the "Shorter" Chronology. *Radiocarbon* 61(1), 243–263.

Praat = Boersma, Paul & Weenink, David. 1992–2025. Praat, version 6.4.39, July 13th 2025.

Preussner, Annina & Leinonen, Jaakko & Riikonen, Juha & Pirinen, Matti & Tukiainen, Taru. 2024. Y chromosome sequencing data suggest dual paths of haplogroup N1a1 into Finland. European Journal of Human Genetics 2024. Published online 28 October 2024. https://doi.org/10.1038/s41431-024-01707-7

Pystynen, Juho. 2013. The case of Mansi *ś > *š, part 1. *Freelance reconstruction : Thinking out loud on historical linguistics*. Blog post, September 6th, 2013.

Pystynen, Juho. 2014. Two Lemmata: PU *ë, PMs *ee *ëë *oo. *Freelance reconstruction : Thinking out loud on historical linguistics*. Blog post, May 31st, 2014.

Pystynen, Juho. 2019. Kantasuomen keskivokaalit: vahvempia ja heikompia tuloksia. In Santeri Junttila & Juha Kuokkala (eds.), *Petri Kallio rocks : Liber semisaecularis 7.2.2019*. Kallion etymologiseura. https://helda.helsinki.fi/handle/10138/305150

Pystynen, Juho. 2022. The Sources of Proto-Samoyedic *ə. Presentation, Universität Hamburg, September 8th, 2022. https://d1wqtxts1xzle7.cloudfront.net/93780259/PSmy_schwa_online-li-bre.pdf?1667775710=&response-content-disposition=inline%3B+file-name%3DThe_Sources_of_Proto_Samoyedic.pdf

Rahkonen, Pauli. 2013. *Southeastern Contact Area of Finnic Languages in the Light of Onomastics*. Doctoral dissertation, University of Helsinki. https://helda.helsinki.fi/handle/10138/38908

Saag, Lehti & Laneman, Margot & Varul, Liivi & Malve, Martin & Valk, Heiki & Razzak, Maria A. & Shirobokov, Ivan G. & Khartanovich, Valeri I. & Mikhaylova, Elena R. & Kushniarevich, Alena & Scheib, Christiana Lyn & Solnik, Anu & Reisberg, Tuuli & Parik, Jüri & Saag, Lauri & Metspalu, Ene & Rootsi, Siiri & Montinaro, Francesco & Remm, Maido & Mägi, Reedik & D'Atanasio, Eugenia & Crema, Enrico Ryunosuke & Díez-del-Molino, David & Thomas, Mark G. & Kriiska, Aivar & Kivisild, Toomas & Villems, Richard & Lang, Valter & Metspalu, Mait & Tambets, Kristiina. 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. 2006. *Substrata Uralica : Studies on Finno-Ugrian substrate in northern Russian dialects*. Väitöskirja, Tartu. http://ethesis.helsinki.fi/julkaisut/hum/suoma/vk/saarikivi/sub-strat.pdf

Saarikivi, Janne. 2022. The divergence of Proto-Uralic and its offspring: A descendent reconstruction. Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford Guide to the Uralic Languages*, 28–58. Oxford University Press.

Sammallahti, Pekka. 1988. Historical phonology of the Uralic Languages. In Sinor, Denis (ed.), The Uralic languages, 478–554. Leiden: Brill.

Sammallahti, Pekka. 1998. *The Saami Languages : An Introduction*. Kárášjohka: Davvi Girji OS.

Shumilovskikh, Lyudmila S. & Schmidt, Monika & Pereskokov, Mikhail & Sannikov, Pavel. 2020. Postglacial history of East European boreal forests in the mid-Kama region, pre-Urals, Russia. *Boreas* 49(3), 526–543. https://onlinelibrary.wiley.com/doi/full/10.1111/bor.12436

Simon, Zsolt. 2020. Urindogermanische Lehnwörter in den uralischen und finno-ugrischen Grund-sprachen: Eine Fata Morgana? *Indogermanische Forschungen* 125 issue 1, 239–265.

SW = Janhunen, Juha. 1977. *Samojedischer Wortschatz : Gemeinsamojedische Etymologien*. Castre-nianumin toimitteita 17.

The Harvard Gazette, July 16th, 2025: Ancient DNA solves mystery of Hungarian, Finnish language family's origins : Parent emerged over 4,000 years ago in Siberia, farther east than many thought,

then rapidly spread west. https://news.harvard.edu/gazette/story/2025/07/ancient-dna-solves-mystery-of-hungarian-finnish-language-familys-origins/

Thomason, Sarah Grey. 2001. Language Contact. An Introduction. Edinburgh University Press.

Vajda, Edward. 2012. *Yeniseian Peoples and Languages : A History of Yeniseian Studies with an Annotated Bibliography and a Source Guide*. London and New York: Routledge.

Vesakoski, Outi & Salmela, Elina & Piezonka, Henny. 2025. Uralic archaeolinguistics. In Robbeets, Martine & Hudson, Mark (eds.), *The Oxford Handbook of Archaeology and Language*, 353–380.

Wainwright, Michael. 2025. Evolution of Apis mellifera. In *Black bees at Killowen and Stonyhurst* (webpage). https://www.killowen.com/genetics3G.html

Warries, Abel Radu. 2022. Towards a new comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems. *Indo-European Linguistics* 10, 169–213.

Werner, Heinrich. 2002. Vergleichendes Wörterbuch der Jenissej-Sprachen, 1–3. Wiesbaden: Harrassowitz Verlag.

Widmer, Paul. 2017. The dialectology of Tocharian. In Wiegand, Herbert Ernst (ed.), *Handbook of Comparative and Historical Indo-European Linguistics*, Volume 2, 1389–1395. Berlin/Boston: Walter De Gruyter.

YSUS = *Yhteissuomalainen sanasto : Kantasuomen rekonstruktiot*. Kielipankki. https://sanat.csc.fi/wiki/Luokka:Kantasuomen_rekonstruktiot

Zeng, Tian Chen & Vyazov, Leonid A. & Kim, Alexander & Flegontov, Pavel & Sirak, Kendra & Maier, Robert & Lazaridis, Iosif & Akbari, Ali & Frachetti, Michael & Tishkin, Alexey A. & Ryabogina, Natalia E. & Agapov, Sergey A. & Agapov, Danila S. & Alekseev, Anatoliy N. & Boeskorov, Gennady G. & Derevianko, Anatoly P. & Dyakonov, Viktor M. & Enshin, Dmitry N. & Fribus, Alexey V. & Frolov, Yaroslav V. & Grushin, Sergey P. & Khokhlov, Alexander A. & Kiryushin, Kirill Yu. & Kiryushin, Yurii F. & Kitov, Egor P. & Kosintsev, Pavel & Kovtun, Igor V. & Makarov, Nikolai P. & Morozov, Viktor V. & Nikolaev, Egor N. & Rykun, Marina P. & Savenkova, Tatyana M. & Shchelchkova, Marina V. & Shirokov, Vladimir & Skochina, Svetlana N. & Sherstobitova, Olga S. & Slepchenko, Sergey M. & Solodovnikov, Konstantin N. & Solovyova, Elena N. & Stepanov, Aleksandr D. & Timoshchenko, Aleksei A. & Vdovin, Aleksandr S. &Vybornov, Anton V. & Balanovska, Elena V. & Dryomov, Stanislav & Hellenthal, Garrett & Kidd, Kenneth & Krause, Johannes & Starikovskaya, Elena & Sukernik, Rem & Tatarinova, Tatiana & Thomas, Mark G. & Zhabagin, Maxat & Callan, Kim & Cheronet, Olivia & Fernandes, Daniel & Keating, Denise & Candilio, Francesca & Iliev, Lora & Kearns, Aisling & Özdoğan, Kadir Toykan & Mah, Matthew & Micco, Adam & Michel, Megan & Olalde, Iñigo & Zalzala, Fatma & Mallick,

Swapan & Rohland, Nadin & Pinhasi, Ron & Narasimhan, Vagheesh M. & Reich, David. 2025. Ancient DNA reveals the prehistory of the Uralic and Yeniseian peoples. Nature 644, 122–132.

Zhivlov, Mikhail. 2014. Studies in Uralic vocalism III. Journal of Language Relationship • Вопросы языкового родства 12 (2014), 113–148. https://www.jolr.ru/files/(151)jlr2014-12(113-148).pdf

Zhivlov, Mikhail. 2015. Неиндоевропейский субстрат в финно-волжских языках. Manuscript. https://www.academia.edu/11689161/

Zhivlov, Mikhail. 2018a. Историческая фонетика и внутренняя классификация уральских языков. Presentation. https://www.academia.edu/38048089/

Zhivlov, Mikhail. 2018b. Прасамодийский *s – фонология и фонетика. Presentation. https://www.academia.edu/38048065/

Zhivlov, Mikhail. 2021. Юкагирская субстратная лексика в хантыйском? Presentation. https://www.academia.edu/63866105/

Zhivlov, Mikhail. 2022. Ob-Ugric subgroup: shared innovations in historical phonology, morphology and lexicon. Presentation. https://www.academia.edu/88469864/Ob_Ugric_subgroup_shared_innovations_in_historical_phonology_morphology_and_lexicon

Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic languages*. Second edition. New York: Routledge.