Chapter 8

Proto-Tungusic in time and space

Martine Robbeets

Max Planck Institute for the Science of Human History & Johannes Gutenberg Universität Mainz

Sofia Oskolskaya

Institute for Linguistic Studies, Russian Academy of Sciences

Although there is a general consensus among historical comparative linguists that the Tungusic languages are genealogically related and descend from a common ancestral language, the internal structure of the family, its age, homeland and prehistoric cultural context remain subject to debate. In addition to four competing concepts of classification, the linguistic literature yields a wide range of time estimations for the family covering more than a millennium as well as four different proposals with regard to the location of the homeland covering Eastern Siberia and Manchuria. Here we will combine the power of traditional comparative historical linguistics and computational phylogenetics to shed light on the prehistory of the Tungusic languages. Our aim is to build on a recent Bayesian verification of the Tungusic family and examine its implications in determining a plausible time depth, location and cultural context of the ancestral proto-Tungusic speech community. We will compare spatial inferences based on two different statistically well-supported Tungusic classifications, namely one in which the break-up of Manchuric constitutes the first split in the family as well as a North-South classification with a northern branch including Even, Evenki, Negidal, Oroqen, Solon, Oroch and Udehe as opposed to a southern branch including Manchuric and Nanaic languages. Situating Proto-Tungusic in time and space, we will estimate the break-up of Proto-Tungusic in the beginning of the first millennium and place its homeland in the area around Lake Khanka. Our study pushes the field forward in answering some tantalizing questions about the prehistory of the Tungusic family, providing a quantitative basis for some conflicting hypotheses and in triangulating linguistics, archaeology and genetics into a holistic approach to the Tungusic past.

1 Introduction

The Tungusic language family is distributed over a vast area in China and Russia, ranging from the Sea of Okhotsk in the east to the Yenisei Basin in the west, and from the Bohai Sea in the south to the Arctic Ocean in the north. Figure 1 shows the distribution of 12 Tungusic languages, notably Oroch, Udehe, Hezhe,[1] Nanai, Orok, Ulch, Xibe, Even, Solon, Evenki, Negidal and Oroqen, as well as dialectal varieties.[2] Whereas the four Nanaic varieties, Hezhe (Heilongjiang), Najkhin Nanai (Middle-lower Amur), Kur-Urmi Nanai (Khabarovsk) and Bikin Nanai (Ussuri), are so diverse that it is not clear whether they should be considered dialects or separate languages, the Momsky and Olsky Even doculects show less internal variation. The map in Figure 1 further shows two historical varieties, notably Jurchen and Manchu. Since written materials in Jurchen, the now extinct language of the Jin dynasty (1115–1234), are only partially deciphered, the earliest well documented stage is Manchu, the official language of the Qing dynasty (1636–1911).

There is a general consensus that the Tungusic languages are genealogically related and descend from a common ancestral language, conventionally called "Proto-Tungusic". However, due to the wide geographical distribution and the considerable internal variation of these languages, the internal family structure along with its root age and homeland are subject to debate. Here we will combine the power of traditional comparative historical linguistics and computational phylogenetics to shed light on the prehistory of the Tungusic languages. Our aim is to build on recent Bayesian analyses of the Tungusic family and examine their implications for determining a plausible time depth and location of the ancestral proto-Tungusic speech community. In addition, we would like to shed light on the factors that drove early Tungusic language spread.

To this end, we will organize our paper as follows. In §2, we will summarize how the recent application of Bayesian inference methods was able to quantify the reliability of previously proposed classifications of the Tungusic family. In §3, we will compare the time range of Tungusic, previously inferred by various linguistic dating techniques, against the quantitative basis provided by Bayesian analysis. In §4, we will test various competing hypotheses concerning the possible homeland of Proto-Tungusic, applying the diversity hotspot principle on the best supported tree models. Finally, we will map our linguistic inferences about Tungusic prehistory on findings from archaeology and genetics in a holistic approach, for which we use the term "triangulation".

---

[1]We use the term *Hezhe* as a cover term for the Kilen and Hezhen dialects and used sources for both dialects.

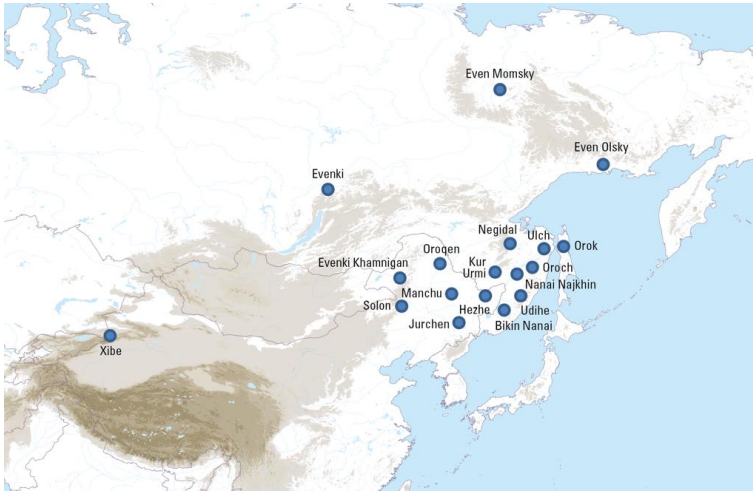[2]We used Natural Earth vector map data for the maps printed in this chapter, which are available in the public domain from https://www.naturalearthdata.com/.

Figure 1: The distribution of the Tungusic languages

2 Family structure

2.1 Previous classifications

Previous studies of the internal taxonomy of the Tungusic family reach different results, particularly with respect to the early separation of the Manchuric (Jurchen, Manchu, Xibe) branch.[3] Cincius (1949), Benzing (1956), Menges (1968: 27) and, more recently, Kormušin (1998), Georg (2004) and Janhunen (2012) proposed a binary north-south classification, in which the separation of Manchuric from the other Tungusic languages does not constitute the earliest split in the family (Figure 2a). The classical approaches by Cincius, Benzing and Menges separated a Northern branch consisting of Evenki, Even, Solon and Negidal from the rest of the Tungusic languages, while the more recent approaches added Oroch and Udehe to the Northern branch instead (Figure 2b). Sunik (1959: 333–335), Vasilevič (1960: 44), Doerfer (1978: 5), Vovin (1993: 102), Whaley et al. (1999: 291), Robbeets (2015), Robbeets & Bouckaert (2018), Dybo & Korovina (2019) and Whaley & Oskolskaya (2020) all argued for an early breakup between Manchuric and the rest of Tungusic, even if their precise configurations do not overlap in each detail (Figure 2c). Moreover, Ikegami (1974) proposed a polytopology, distinguishing as many as four branches, namely Manchuric, Evenic, Nanaic and Udeheic (Figure

---

[3]The term Manchuric is used elsewhere in literature to designate this branch, see among others Alonso de la Fuente (2011) and Robbeets & Savelyev (2020).

2d).[4] The basic configuration of these four topologies is represented in Figure 2.

Except for Dybo & Korovina's (2019) classification, which is based on lexicostatistics, the majority of previous research is based on the traditional Maximum Parsimony method. This approach used by historical comparative linguists, seeks a tree that explains a dataset by minimizing the number of evolutionary changes required to produce the observed data. Lexicostatistics is an early and less reliable form of statistical tree-building, which uses the shared cognate proportion in a basic vocabulary list as a distance metric to estimate linguistic relationships.

The Bayesian method seeks to explain a set of observed data by quantifying how likely it is that they have been produced by a certain model. As this is a statistical approach in which all forms of uncertainty are expressed in terms of probability, it can contribute to the current state of the art by verifying which of the models in Figure 2 is best supported by the data by quantifying the statistical robustness of different proposals and by inferring absolute divergence dates. In this way, a Bayesian approach can provide a quantitative basis for previous classifications based on classical historical linguistic approaches. Here we aim at interpreting the results of a recent Bayesian analysis of the Tungusic family (Oskolskaya et al. 2022) and at inferring spatiotemporal and cultural patterns of Tungusic linguistic dispersal.

2.2 A recent Bayesian approach to the classification of the Tungusic languages

Oskolskaya et al. (2022) took a Bayesian approach to the classification of the Tungusic languages, based on the dataset of 254 basic vocabulary items collected for 21 Tungusic varieties. The maximum clade credibility tree in Figure 3, which is the best supported tree among all trees generated by applying different evolutionary models, summarizes the results of this study.[5]

The trees underlying Figure 3 were generated running the software BEAST 2.4.7 (Bouckaert et al. 2014), which only allows for a binary structure of splits. Jurchen, which is usually considered as a direct ancestor of Manchu is represented as a separate branch because a written standard is never considered directly ancestral to a spoken variety. The model thus assumes that Jurchen and Manchu-Xibe have separated from a common ancestor, spoken at some point in the past.

---

[4]Some authors, such as Ikegami and Janhunen used a different terminology for these groupings, but for reason of accessibility, we use a consistent terminology here.

[5]All datasets and coding details are accessible through the supplementary information in Oskolskaya et al. (2022).

a

North

Evenki
Solon
Negidal
Even

South

Nanai
Olcha
Orok
Oroch
Udehe

Jurchen
Manchu
Sibe

b

North

Evenki
Even
Solon
Negidal
Oroch
Udehe

South

Orok
Olcha
Nanai
Manchu

c

Tungusic

Evenki
Solon
Negidal
Even

Nanai
Olcha
Orok
Oroch
Udehe

Manchuric

Jurchen
Manchu
Sibe

d

Even
Evenki
Solon
Negidal
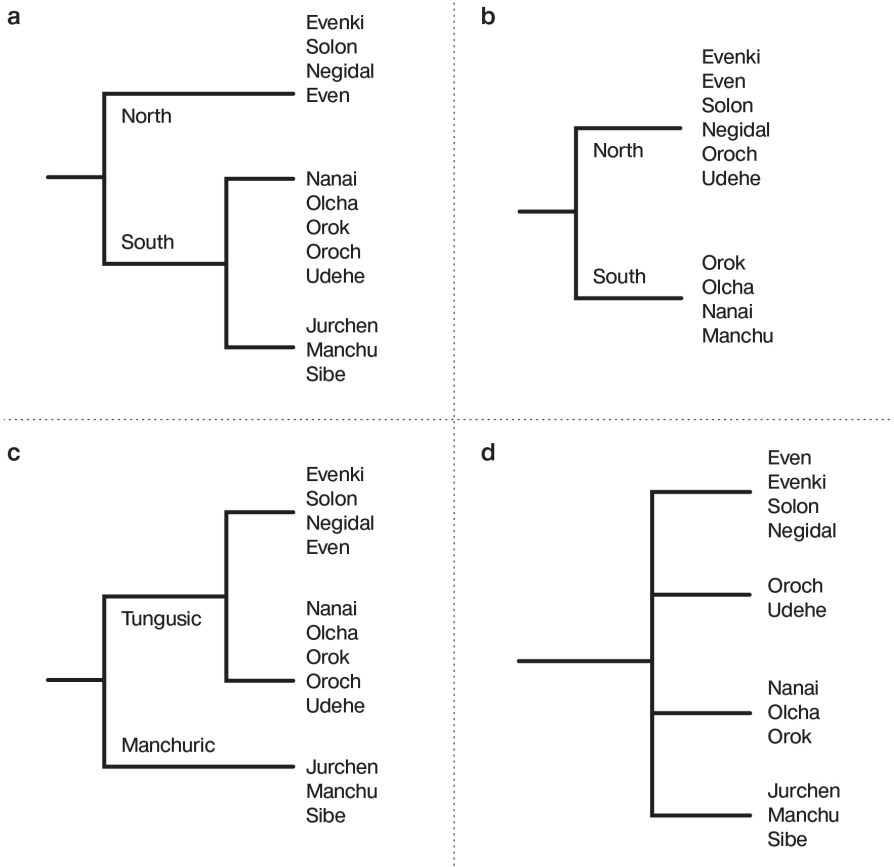
Oroch
Udehe

Nanai
Olcha
Orok

Jurchen
Manchu
Sibe

Figure 2: Basic configuration distinguishing four topologies for the Tungusic family: a. Classical North-South classification; b. Revised North-South classification; c. Manchu-Tungusic classification; d. Quadruple topology
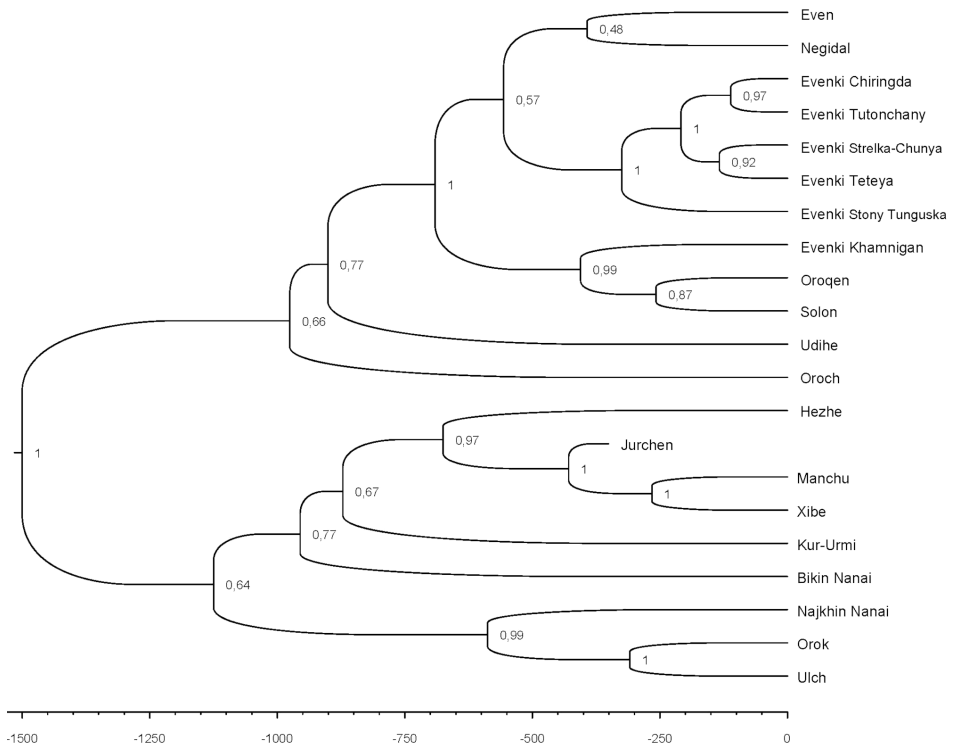
Figure 3: Maximum clade credibility tree (Binary covarion, no gamma variation, relaxed clock) for the Tungusic family (Oskolskaya et al. 2022)

The numbers on the nodes show the posterior probability, which qualifies the statistical robustness of each clade. The higher the number, the more probable the existence of the clade. We can thus safely establish a Northern Tungusic branch (posterior probability = 1), a Nanaic branch with Najkhin Nanai, Orok and Ulch (posterior probability = 0.99) and a Manchuric branch, probably including Hezhe besides Jurchen, Manchu and Xibe (posterior probability = 0.97). With posterior probabilities below 0,80, the exact position in the tree of Udihe, Oroch, Kur-Urmi, Bikin Nanai and the cluster Orok-Ulch-Nakjin Nanai is less secure, but it is interesting to note that the tree in Figure 3 does not support a monophyletic cluster composed by Oroch and Udihe.[6]

Taking into account the probabilities of the branches, the tree in Figure 3 supports two basic classifications previously proposed in the literature, the revised

---

[6]Perekhvalskaya (2022 [this volume]) proposes that Udihe, Oroch, and Kyakala are the ends of a continuum.

North-South classification (Figure 2b) and the Manchu-Tungusic classification (Figure 2c). According to Oskolskaya et al. (2022), the revised North-South classification with a binary split between Manchuric and Nanaic on the one hand and Northern Tungusic, Udehe and Oroch on the other is best supported (i.e. in 48,3% of generated trees), while the Manchu-Tungusic classification (Figure 2c) is also highly probable (31,1% of trees). The other two classifications, namely the classical North-South and the quadruple classification presented in Figures 2a and 2d, are excluded by this analysis.

3 Age

3.1 Previous dating

As shown in Table 1, different linguistic dating principles yield a time range for the primary breakup of Proto-Tungusic between 950 BC and AD 700. Applying lexicostatistic methods, Dybo & Korovina (2019) dated Proto-Tungusic to around 950 BC, while Korovina (2011) dated it to the sixth century BC, but other distance-based methods, such as the Automated Similarity Judgment Program (ASJP) yielded much younger dates, notably AD 681 (Holman et al. 2011: 854). Distance-based methods use a distant metric, such as the lost cognate proportion or the number of operations required to turn one string of phonemes into another, to infer the time depth of language separation. However, as they assume a constant rate of loss over time, their results are not generally accepted. Another loose dating principle that is not entirely foolproof is tracing the primary break-up of Proto-Tungusic back to certain ethnonym shifts. Referring to the name change in Chinese dynastic chronicles of the Tungusic ethnonym "Yilou" to "Wuji", Robbeets (2015: 16–18) situated the break-up of Proto-Tungusic at the end of the Han period (206 BC–AD 220). On the basis of a rough measure of mutual intelligibility, Pevnov (2012: 32) estimated that Proto-Tungusic could not be younger than two thousand years.

Reconstructing the vocabulary of a proto-language, we can examine the cultural and ecological concepts revealed in it. The time when some of these concepts became available to the speakers of the proto-language can also serve as an indication for the language family's time depth. In line with Janhunen's (2012: 8) findings, an Iron Age dating of Proto-Tungusic is supported by the reconstruction of PTg *sele 'iron', reflected in Evenki sele, Even hel, Neg. sele, Solon sele, Xibe selə, Manchu sele, Jur. *sele, Ulch sele, Orok sele, Nanai sele, Oroch sele and Udihe sele. Although a single reconstruction, which is not backed up by other vocabulary items, cannot provide us with a reliable chronology, the Iron Age dating

of Proto-Tungusic is further corroborated by contact linguistics. As words have travelled between languages since prehistoric times, giving voice to new ideas and names to new products and practices, we expect to observe links between cultural diffusion and prehistoric borrowing. One possible candidate for such a link is PTg *murgi*, the reconstructed term for 'barley and similar crops'. There is reason to believe that this word is borrowed from an Old Chinese donor word 來 *mə.rˤək > *mə.rˤə 'a kind of wheat' (Robbeets 2017b: 28–29). The linguistic reconstruction can be correlated to the archaeological evidence for barley being first imported through Chinese contact at the time of the Krounovskaya culture (600 BC–AD 200), situated in the Southern Primorye around Lake Khanka (Sergusheva & Vostretsov 2009: 214–215). This culture also marks the beginning of the Early Iron Age in the Russian Far East with the first uncontested finds of iron. Therefore, on the basis of contact studies, the break-up of Proto-Tungusic has been dated to the period between 600 BC and AD 200.

Table 1: The estimated time of separation of Proto-Tungusic according to different linguistic dating principles

Dating principle	Source	Time depth
Lexicostatistics	Dybo & Korovina 2019	950 BC
Lexicostatistics	Korovina 2011	600 BC
ASJP	Holman et al. 2011	AD 681
Ethnonym shift	Robbeets 2015	206 BC–AD 220
Intelligibility	Pevnov 2012	AD 1
Cultural reconstruction	Janhunen 2012	500 BC–AD 500
Contact linguistics	Robbeets et al. 2020	600 BC–AD 200

3.2 Dating through Bayesian inference

The Bayesian method calibrates the divergence time of the root and the nodes in a language family against known cases of language divergence over attested timespans and quantifies how likely it is that the inferred time depth falls within a certain density interval. For calculating the divergence time of the root and the nodes in Tungusic tree, Oskolskaya et al. (2022) calibrated the tree in three nodes.

First, they estimated a time depth for the transition from Jurchen into Manchu and inserted it as a calibration point. To this end, they used the time of the first known Manchu manuscript which is dated to 1599 (Gorelova 2002: 50). The logic

is that Jurchen is no longer attested by that period and had thus ceased to exist before 1599. Therefore, 351 years before present (conventionally before 1950) was used as an estimation for the separation time between Jurchen and Manchu–Xibe. However, as the model assumes that Jurchen and Manchu–Xibe already separated before the time that Jurchen was first attested, thus before 1185, a separation time that aligns with the model should be at least 765 years before present. This difference of at least 400 years is expected to yield a later date for the calculated time depth of Proto-Tungusic than the real one.

The second calibration point is the time of the split between the Xibe and Manchu languages, 186 years ago. This is based on the dating of the resettlement of Xibe populations to the northwest of China in 1764 (Gorelova 2002: 31).

The third calibration point is the time of the first break-up of Evenki, which had taken place already before 1723, i.e. more than 227 years ago. According to Vasilevič (1969), D. G. Messerschmidt in his fieldnotes in 1723 provided vocabulary collected from various Evenki people in different regions. His data show that there were at least two dialects that can be associated with the modern distinction between the Northern versus the Southern and Eastern dialects. Thus, the first break-up of Evenki had already taken place by that time.

Using these three calibration points to calibrate the tree, Bayesian analysis infers a time depth for the root and nodes in the Tungusic tree, as shown in Figure 4.[7]

All dates in Figure 4 indicate a time "before present", conventionally before 1950. Each bar shows a time range, in which a specific split has taken place with a 95% higher posterior density interval (HPDI). The time depth of the primary split is particularly relevant for our present study. The credible interval for this split covers about 1800 years (737 BC–1154 AD), which implies that there are not enough data for more precise results. Nevertheless, the median of this bar is around 1500 BP, i.e. AD 450.

It should be noted that the three calibration points refer to relatively recent events and present upper limits after which the separation cannot have taken place. This could lead to a bias in the estimation of the time-depth of the Tungusic tree, by which the estimated age would be younger than the real age. This effect is increased by the fact that the split between Jurchen and Manchu-Xibe should be calibrated at least 400 years earlier. Therefore, it is probable that the actual break-up of Proto-Tungusic took place several centuries before the inferred date of AD 450, probably in the beginning of the first millennium AD, as implied in the hypotheses provided by Janhunen (2012), Pevnov (2012), Robbeets (2015) and Robbeets et al. (2020).

---

[7]The detailed information of this analysis is described in Oskolskaya et al. (2022).
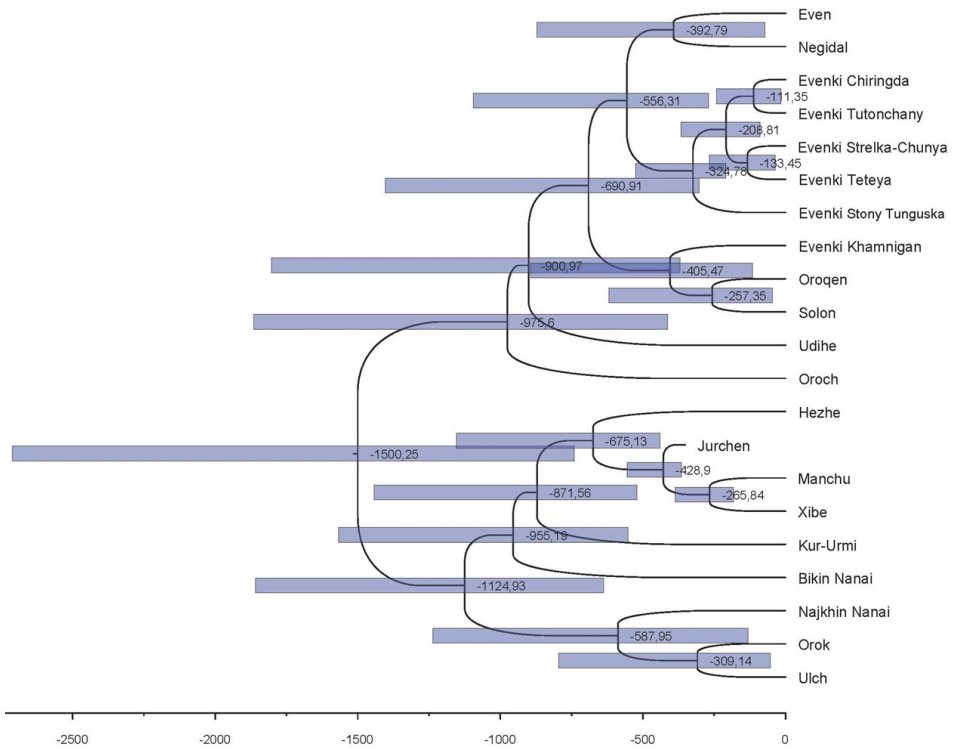
Figure 4: Bayesian age estimation for the nodes in the Tungusic tree (Oskolskaya et al. 2022)

4 Homeland

4.1 Previous proposals

As indicated on the map in Figure 5, there are four competing hypotheses with regard to the possible homeland of Proto-Tungusic, notably (1) the Baikal region (Vasilevič 1960; Menges 1968: 23; Derevyanko 1976; Helimski 1985: 279), (2) the Mid Amur and the lower part of the Upper Amur region (Tugolukov 1980; Janhunen 1996: 169; Korovina 2011; Pevnov 2012; Wichmann p.c., 2019.10.03; Pugach et al. 2016), (3) the region around Lake Khanka (Robbeets 2020; Wang & Robbeets 2020) and (4) the Yalu River region on the border between present-day Liaoning and Northern Korea (Janhunen 2012). The evidence in support of the Baikal region comes mainly from prehistoric contact linguistics assuming ethnolinguistic interaction with ancient speakers of Amuric, Samoyedic, Mongolic and Yeniseic. An original location in the Mid and Upper Amur region is supported by various

approaches: Janhunen (1996: 169) provides ethnolinguistic indications; Korovina (2011) reconstructs names for insects, reptiles, shellfish and fish; Pevnov (2012) combines toponyms with reconstructed river vocabulary and tree names; and Wichmann takes a computer-automated approach to the diversity hotspot principle. The region around Lake Khanka is supported by the diversity hotspot principle and cultural reconstruction, while the Yalu River region is proposed on the basis of a general northward trend of expansion.
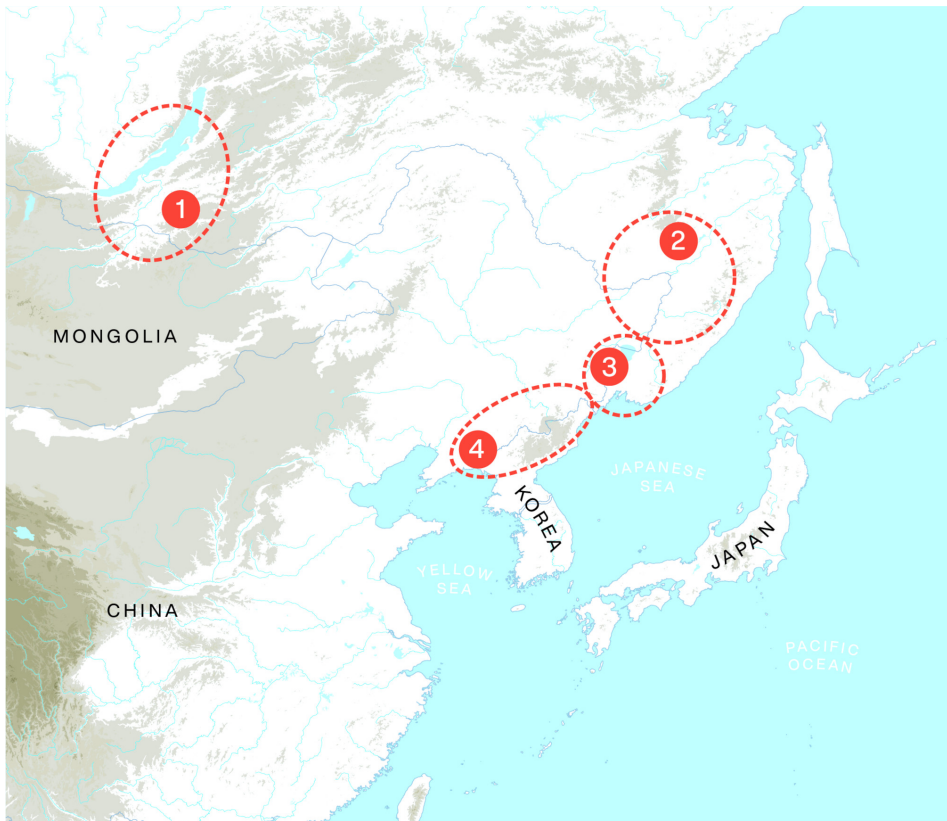

Figure 5: Proposed locations for the homeland of Proto-Tungusic: (1) the Baikal region; (2) the Mid Amur and the lower part of the Upper Amur region; (3) the region around Lake Khanka; (4) the Yalu River region

4.2 Diversity hotspot principle

The "diversity hotspot principle" is based on the assumption that the homeland is closest to where one finds the greatest diversity with regard to the deepest subgroups of the language family. It follows that the primary splits in the family are determinants for the location of the homeland. In the case of the Tungusic family, this implies that depending on which of the four classifications in Figure 2 we favor, the center of primary diversity – and thus also the inferred location of the homeland – will move on the map. The resulting locations are situated on the map in Figure 6. Whereas the classical north-south classification pushes the homeland to the north towards the Upper to Mid Amur region (Figure 6a), the revised north-south, the Manchu-Tungusic and the quadruple classifications pull it more southwards towards the Mid Amur region immediately north of Lake Khanka (Figure 6b), the area around Lake Khanka (Figure 6c) and Manchuria (Figure 6d), respectively.

Although the diversity hotspot principle can provide some clues about the homeland of a language family, it must also contend with several limitations (Wang & Robbeets 2020): the contemporary hotspot of linguistic diversity may diverge from the earlier one or the principle may be upset when a migration was suddenly directed over a long distance rather than representing slow, gradual and random movement into adjacent areas. The Xibe populations, for instance, were suddenly resettled to the west in the 18th century and northern Tungusic populations, such as the Evenki and Even, are extremely mobile. Nevertheless, we can gain more from applying the diversity hotspot principle to Tungusic, than we can lose from ignoring it, because it helps us to set up hypotheses about homelands and it makes us aware of the interconnection between internal family structure and the identification of a homeland.

Since our Bayesian analysis presented in Oskolskaya et al. (2022) and summarized in §2.2 supports the classifications represented in Figure 6b/c, it indicates that the original homeland of the Proto-Tungusic speech community was situated in the area around lake Khanka or immediately to the north of it. This location is corroborated by Wichmann et al.'s (2010) attempt to formalize the diversity hotspot principle in a computer-automated approach. Taking into account 11 Tungusic languages, notably Even, Evenki, Negidal, Oroqen, Naykhin Nanai, Oroch, Orok, Ulch, Udehe, Manchu and Xibe, they compute a diversity measure for each language and identify the homeland with the location of Naykhin Nanai, the language with the highest diversity measure.[8] The location of Naykhin Nanai

---

[8]Wichmann et al.'s measure of diversity is derived as the proportion l/g of the linguistic distance l and the geographical distance g between two languages. By way of linguistic distance, they use the *Levenshtein distance*, which is the minimum number of substitutions necessary to transform one string of phonemes into another in a subset of 40 basic vocabulary items.

at latitude 49.28 and longitude 136.47 is equated with that of the Tungusic homeland. As more southern varieties such as Solon on the Nonni River in Inner Mongolia and Bikin Nanai in the southern part of the Primorye province as well as Hezhe in Heilongjiang are lacking from this analysis, we may expect the real diversity hotspot to be slightly more to the south than the inferred one.
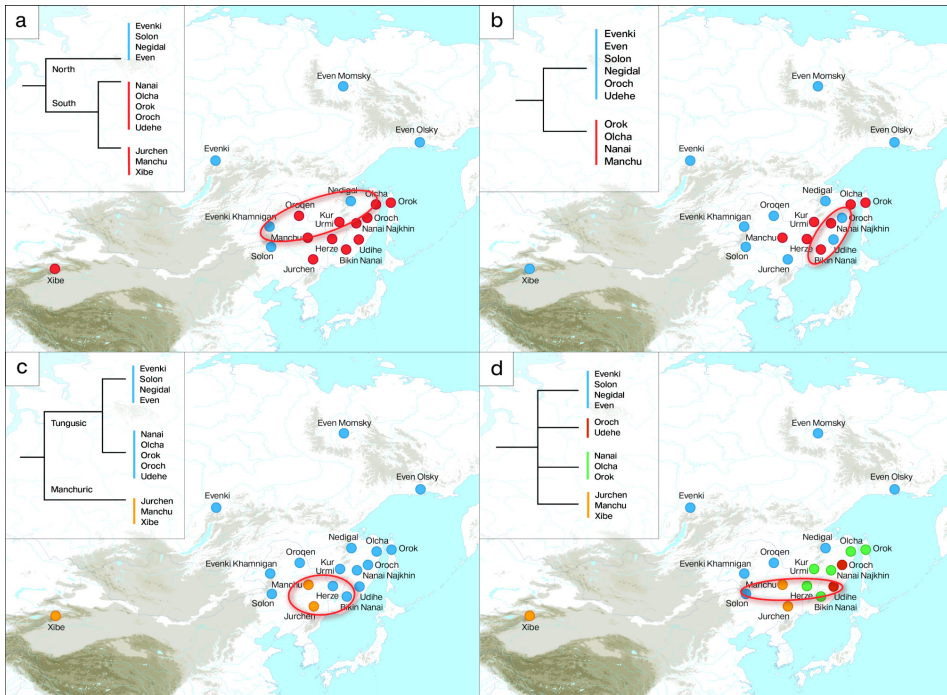

Figure 6: Diversity hotspot of the Tungusic languages under the four proposed classifications

5 Triangulation

5.1 Linguistics

Combining Bayesian inference with other linguistic approaches, we estimate that Proto-Tungusic was spoken in the area around or immediately north of lake Khanka in the beginning of the first millennium AD. As shown in Table 2, cultural reconstruction indicates that the speakers of Proto-Tungusic were familiar with agriculture. The use of words such as *pisike* 'broomcorn millet (*Panicum*

miliaceum)', **jiya-* 'foxtail millet (Setaria italica)',[9] **murgi* 'barley',[10] **üse- ~ üsi-* 'to plant',[11] **üse ~ üsi* 'seed, seedling', **üsin* 'field for cultivation' and **tari-* 'to sow, plant, cultivate' implies that the speakers relied on plant cultivation for subsistence. The derivation of the Proto-Tungusic word **üse ~ üsi* 'seed, seedling' as a deverbal noun from the verb **üse- ~ üsi-* 'to plant' suggests that seeds were not just collected for consumption in the wild but that they were planted as part of a cultivation process. In addition, for some crop names such as 'barley' and 'broomcorn millet', we can argue that the word refers to the domesticated crop rather than to the wild variety of the plant because both crops are not native to the region and have been imported as domesticated crops. It is commonly assumed that broomcorn millet has been imported from the West Liao River region by the people who introduced the Zaisanovskaya culture (3200–1300 BCE) to the Russian far East (Sergusheva & Vostretsov 2009; Leipe et al. 2019; Li et al. 2020).

Whereas several agricultural terms are inherited, maritime vocabulary can often be explained as borrowed from Proto-Amuric or its descendant Nivkh. This

---

[9]The cognates Nanai *jie-kte* and Oroch *jie-kte* suggest front vocalization **jiye-kte* in Proto-Tungusic, while Negidal *ja:kta*, Solon *jakta* and Udehe *jakta* suggest original back vocalization **jiya-kta*. In addition to the vowel alternation in the in Oroch suffix *-kte* and *-kta*, this observation suggests that the alternation between **jiya-kta* and **jiye-kte* was already present at the Proto-Tungusic stage. It was probably due to assimilation with the initial front vowel in **jiya-kta*. Another example of such an alternation is found in Proto-Tungusic **niaya- ~ nieye-*, which is reflected as *yaya-* or *ńaya-* 'to shamanize' in most Tungusic languages but as *leye-* 'to sing' in Manchu. Note that PTg **ye* regularly develops in Manchu *ye*, e.g. PTg **xeye-* 'to sink' and Ma. *eye-* 'to float, flow' (Cincius 1975/77: 440–442; Starostin et al. 2003) or PTg **jeye* 'sharp point; blade' and Manchu *jeyen* 'sharp point; blade' (Cincius 1975/77: 282–283; Starostin et al. 2003), etc. Manchu *je* 'foxtail millet (Setaria italica); grain' is thus considered a contraction from **jiye*.

[10]It is a universal tendency that consonant clusters are susceptible to variation through assimilation, metathesis or consonant loss. Consonant clusters across the Tungusic languages are no exception to this expectation. Although the reflex of the Proto-Tungusic **rg* cluster is relatively stable in most northern Tungusic languages, we find a variety of reflexes, such as *rg*, *gg*, *yg*, *jg* and *g* in Manchuric and other southern Tungusic languages, including an occasional *j* in Nanai, Orok and Ulch, e.g., PTg **burga-kta* 'beard, moustache' reflected in Ulch *buja-qta*, Nanai *boja-qta*; PTg **serge-kte* 'nose bone' in Nanai *sejurẽ*; PTg **irge* 'brain, head' in Ulch *ije*, etc.

[11]Given that the expected reflex of PTg **ü* is Udehe *i*, Udehe *uhi-* 'to sow, to plant a garden', *uhi* 'garden for cultivating plants' may represent cases of intra-Tungusic borrowing and the inherited reflex in Udihe may be *yehu-* 'to grow', given the palatal glide onset. Note however that, even if the majority reflex PTg **ü* is *i* in Oroch, there are a few instances where PTg **ü* is retained as *u* instead, e.g. PTg **erün* 'time' as Oroch *erū(n)* (Cincius 1975/77: 463–464; Starostin et al. 2003), PTg **xulbü-* 'to bind, arrange' as Oroch *ubbuna-* (Cincius 1977: 258; Starostin et al. 2003), PTg **xegün* 'nine' as Oroch *xuju(n)* (Cincius 1975/77: 352–353; Starostin et al. 2003), PTg **tüksa* 'house cover made of birch bark' as Oroch *tuksa* (Cincius 1975/77: 179; Starostin et al. 2003), etc. Therefore, this correspondence is included in the appended list of sound correspondences.

Table 2: The reconstruction of agricultural vocabulary in Proto-Tungusic (adapted from Robbeets et al. 2020: 765)

No.	Proto-Tungusic	Attested Tungusic words
(1)	*pisike 'broomcorn millet (Panicum miliaceum)'	Manchu fisihe ~ fisike 'glutinous millet, broomcorn millet (Panicum miliaceum)'; Ulch pikse; Nanai pikse 'millet'; Kur-Urmi dialect fisxe 'broomcorn millet (Panicum miliaceum)'; Jurchen *fise bele 'yellow rice; coarse rice' (bele 'hulled rice, edible grain')
(2)	*jiya-kta ~ *jiye-kte 'foxtail millet (Setaria italica)'	Negidal ja:kta 'foxtail millet (Setaria italica), small millet,; cleaned grain; oats flour'; Solon jakta 'porridge, food'; Xibe je 'foxtail millet (Setaria italica)'; Manchu je 'foxtail millet (Setaria italica); grain', Jurchen *je bele 'millet'; Nanai jiekte 'foxtail millet (Setaria italica)'; Oroch jiekte, jekte, jiekta 'millet', Udehe jakta 'foxtail millet (Setaria italica), porridge'
(3)	*üse- ~ üsi- 'to plant' → *üse ~ üsi 'seed, seedling' → *üsi-n 'field for cultivation'	Evenki ihəw- 'to grow (of people)'; Even isu:- ~ esu:- 'to sprout, come out (of plants), blossom, grow', Negidal isew- 'to grow, become acclimatized (about plants); mature, grow up (about people)'; Xibe use- 'to sow seeds', use 'seed, grain', usin 'field, farmland'; Manchu use- 'to plant, seed (tr.)', use 'seed; insect egg', usin 'field for cultivation', Jurchen *use 'seedling', usi-in 'field', Ulch use 'seed', usun 'field, garden'; Nanai use 'seed', usī 'arable field for cultivation; private garden for cultivating vegetables', Orok usi 'field (farm), garden', usi- 'to cultivate, till; hunt a bear'; Oroch usi 'seeds, grains', usin 'garden for cultivation', usin- 'to sow, to plant in a garden'; Udehe yehu- 'to grow', uhi- 'to sow, to plant a garden', uhi 'garden for cultivating plants'
(4)	*tari- 'to sow, plant, cultivate'	Evenki tari- ~ tare- ~ tale- 'to sow'; Solon tari- 'to sow seeds; to plant; to cultivate; to grow; to disseminate; to inject; to infect, to catch a disease', tariŋko 'injector'; Manchu tari- 'to cultivate, farm; to plow'; Jurchen *tali- ~ tari- 'to sow, to plant, to cultivate'; Ulch tari- 'to sow, to plant'; Nanai tari- 'to sow seeds', tariko 'sowing machine', tarici- 'to sow (seeds) regularly'; Udehe tali- 'to plant a garden', tali 'garden for cultivating plants'
(5)	*murgi 'barley (Hordeum vulgare)'	Manchu muji 'barley (Hordeum vulgare)', Xibe muji 'barley (Hordeum vulgare)', Ulch muji 'oats', Nanai muji 'oats', Oroch muji 'barley', Jurchen *mirɣei 'product of agriculture', Solon muɾgil 'spring crops, spring-sown field'

is for instance the case for ancient loanwords, such as PTg **laamos* 'wind (from the sea)' from Proto-Amuric **lamos* > Nivkh *lams* 'eastern wind', PTg **kalïmV* 'whale' from Proto-Amuric **kalïmV* 'whale' > Nivkh *kalm* (*qalm*) '(small) whale' (Janhunen 2016) and PTg **laska* 'sea goby' from Proto-Amuric **laskV* 'goby' > Nivkh *lask* 'a goby (of middle size).

The direction of the borrowing is verifiably from Amuric into Tungusic given that Proto-Amuric **lamos* is derived from the simplex root **la* 'wind', while the Tungusic parallel is not segmentable and because the initial liquid phoneme **l-* is atypical for Proto-Tungusic.

Korovina (2011) further finds that fish species that inhabit the Pacific Ocean are not well distributed across the Tungusic languages and are often borrowed from Nivkh into one or more individual daughter languages (e.g. Oroqen *lokko* 'flounder' from Nivkh *lok* 'flounder', Oroqen *la:kka* 'herring' and Orok *la:qqa* 'herring' from Nivkh *laku* 'herring', etc.). This suggests that the speakers of Proto-Tungusic were farmers, who did not acquire maritime vocabulary until they came in contact with indigenous populations on the Pacific coast, some of which might have spoken an ancestral form of Nivkh.

Except for these maritime loanwords, Proto-Tungusic borrowed only few words from Proto-Amuric. By contrast, there are several indications of Proto-Amuric substratum interference in Proto-Tungusic. The evidence comes from atypical structural features in Tungusic that are likely to have developed through imperfect learning from Proto-Amuric. Among others, these features include the development of a word-initial liquid and velar nasal sound in Tungusic, the development of a distinction between 'we (including the addressee)' and 'we (excluding the addressee)' in first-person plural pronouns, the development of a distinction between alienable and inalienable possession and the development of marking possessive relations on the head noun instead of the dependent (Robbeets 2017b).

Whether one prefers to explain the numerous structural similarities between Tungusic and other Transeurasian languages by borrowing or inheritance, it is commonly agreed that Tungusic typology is of the Transeurasian (or "Altaic") type (Robbeets 2017a). Transeurasian languages are typical dependent-marking languages, while Nivkh is – similar to Ainu, Asian North Pacific-Coast languages (e.g., Chukotko-Kamchatkan), wider Paleosiberian languages (e.g.,) and languages of the Northwest Pacific Coast (e.g., Salishan, Wakashan, Chimakuan, Athabaskan) – of the head-marking type. The features above are atypical for Tungusic and more proto-typical of Nivkh in the sense that they represent direct or indirect implicational tendencies of being of the head-marking type or that they more frequently occur in Ainu, other Asian North Pacific-Coast languages,

wider Paleosiberian languages and languages of the Northwest Pacific Coast than in languages of the Transeurasian type.

These linguistic observations thus suggest a situation of language shift whereby some ancestral speakers of Proto-Amuric abandoned their own language and adopted the Proto-Tunguisic target language.

5.2 Archaeology

Is our association of Proto-Tungusic with incoming millet farmers who imposed their language on local fishers speaking Proto-Amuric supported by the archaeological record?

During the Middle to Late Hongshan periods (4000–3000 BC), the cultivation of broomcorn and foxtail millet dispersed from the West Liao River basin in North East China to the Primorye (Maritime) province of the Russian Far East (Sergusheva & Vostretsov 2009; Leipe et al. 2019; Li et al. 2020). The introduction of millet farming in the Primorye was combined with the adoption of Northeast Chinese material culture such as cord-marked pottery, spindle whorls and stone agricultural tools, especially mortars, pestles and constricted-waist hoes (Nelson et al. 2020; Li et al. 2020) and led to the establishment of the Zaisanovskaya culture (3200–1300 BCE). The linguistically inferred time depth in the beginning of the first millennium AD corresponds to the break-up time when Proto-Tungusic separated into its primary branches and thus ceased to exist, but it does not inform us about when the ancestral language arrived in the region or started to exist there. Considering Bayesian inference of the time depth of the split between Tungusic and Mongolo-Turkic at 3300 BC (Robbeets & Bouckaert 2018), it is inviting to associate the arrival of Proto-Tungusic in the Russian Far East with the beginning of the Zaisanovskaya culture (Mallory et al. 2019; Wang & Robbeets 2020; Cui et al. 2020; Li et al. 2020).

Observing the high productivity of rice vis-à-vis millets, archaeobotanists argue that rice tends to be spread more easily through cultural diffusion, while millets are more frequently spread by population migration (Fuller & Qin 2009; Stevens & Fuller 2017). This is explained by the fact that wet rice cultivation can absorb population increase through intensification of land use, while the increased production of millet tends to occur through the agricultural colonisation of new land. The assumption of actual population movements from the West Liao River Basin to the Primorye in the fourth millennium BC is further supported by increases in population density (Peterson et al. 2010; Miyamoto 2014; Drennan et al. 2017; Leipe et al. 2019).

Population migration and cultural diffusion are expected to yield different linguistic outcomes (Thomason & Kaufman 1988). In the first case, when human populations move into new areas along with their language and culture, language shift is frequently observed: local speakers abandon their own language in favour of the incoming target language. Due to imperfect learning, the abandoned language may leave some traces in the structure of the target language, a phenomenon called "substratum interference" (Van Coetsem 2000; Johanson 2002; Winford 2013). Nevertheless, the newly adopted language is genealogically related to the ancestral language of the migrants. By contrast, in the case of cultural diffusion, when certain elements of language and culture move into new areas without the intervention of a migrating population, local speakers frequently maintain their own language but borrow certain words from the model language. The assumption of language shift, whereby a part of the Proto-Amuric speakers abandoned their native language and shifted to the Proto-Tungusic target language is thus in line with a scenario of population migration. Therefore, the archaeological and linguistic observations converge in suggesting that the spread of the Proto-Tungusic farmers was driven by population migration.

At the end of the third century AD, there was a sharp cooling of the climate, which led to a worsening of the conditions for agriculture. This provided the impetus for a gradual migration of millet farmers to coastal regions across most of the Primorye. Based on archaeobotanical data (Yanuševič et al. 1990), it appears that the coastal groups ceased to cultivate millets and wheat and returned to a subsistence strategy of hunting and fishing. If this event can be associated with the separation between Manchuric and Tungusic languages, as suggested by 31% of trees in the Bayesian phylogenetic analysis in Oskolskaya et al. (2022), it would explain why southern Tungusic populations on the Lower Amur such as the Nanai, Oroch and Udehe people were traditionally predominantly fishers and gatherers, rather than farmers.

Hudson (2020) proposed further details of later northern Tungusic expansions. The Evenki are widely distributed hunter-gatherers who also herd domesticated reindeer. According to Anderson (1999: 142) and Zgusta (2015: 166), they first herded wild reindeer around Lake Baikal then moved north ca. AD 1000, reaching the Arctic ocean by the 17th century. The Even probably separated from the Evenki in medieval times (Pakendorf 2007: 15–16), matching the separation estimated at 556 years ago in our Bayesian analysis (Figure 4). They further expanded with reindeer from 17th century onwards, mirroring the separation between Even and Negidal estimated at 393 years ago. Probably due to Russian colonial expansion, the Oroqen and Solon moved south from the Amur in the 17th century, mirroring the estimate of 405 years ago.

5.3 Genetics

The first applications of genetics to the study of human prehistory involved mitochondrial and Y-chromosomal DNA. Whereas mitochondrial DNA is passed down along the maternal line from mother to daughter to granddaughter (and from mother to son but not passed on from sons to their offspring), Y-chromosomal DNA goes along the paternal line from father to son to grandson. Sequencing the chemical building blocks of uniparental DNA from diverse people around the world and comparing the mutations across these sequences, geneticists can reconstruct family trees of maternal and paternal relationship. However, since mitochondrial DNA and Y-chromosomal DNA represent only a tiny proportion of the human genome and provide information on only one out of very numerous ancestors, they shed light on only a limited slice of human prehistory. In fact, our entire genome contains information about many diverse ancestors, not just the two whose lineages can be traced with mitochondrial and Y-chromosomal DNA. The recently acquired ability to sequence the whole genome – meaning, the entire genome analyzed at once instead of just small stretches of it such as mitochondrial and Y-chromosomal DNA – has given us access to richer information recorded into all 23 chromosomes of our genome and representing a multitude of ancestors. Whole genome analysis means a revolution in the study of the human past because it allows us to go beyond the tiny slice of the past sampled by our mtDNA and Y-chromosomal DNA. As recent genome-wide analyses of Tungusic speakers (Pugach et al. 2016; Siska et al. 2017; Wang et al. 2021; Wang & Robbeets 2020) are expected to tell a richer story than previous studies about their mtDNA (Starikovskaya et al. 2005; Sukernik et al. 2012; Duggan et al. 2013) and Y-chromosomal DNA (Malyarchuk et al. 2010; Duggan et al. 2013), we here focus our report on genome-wide analyses.

The Principal Component Analysis in Figure 7 visualizes the genetic distance between contemporary speakers of Tungusic languages and other present-day East Asian populations. In addition, it plots ancient genomes from the Devil's cave in the Southern Primorye dating back to the fifth and sixth millennium BC and from the Ust'-lda site near Lake Baikal dating back to the fourth and third millennium BC onto the contemporary Tungusic-speaking populations.

The contemporary Tungusic speakers in the Amur River Basin, such as the Hezhen, Nanai, Negidal, Oroqen and Ulch are genetically most similar to ancient genomes from the Southern Primorye dating back to the fifth and sixth millennium BC (Siska et al. 2017; Wang et al. 2021; Wang & Robbeets 2020). This is also true for the Nivkh people on Sakhalin island, even if their language is not of a Tungusic descent. The Xibe people in Xinjiang are shifted towards Han Chi-

nese populations due to Chinese influence but they are still very similar to the Amur Tungusic populations and close to the Devil's Gate genome. Whereas some Eastern Evenki are similar to the Amur Tungusic populations, Baikal Evenki and Even populations are shifted towards West Eurasians, such as the Uyghur Turkic populations on the PCA. Wang & Robbeets (2020) estimate that they have about 14% to 35% West Eurasian related ancestry, but that their admixture is a very recent event, going back less than 200 years in time. The ancient genomes from the Ust'-lda site near Lake Baikal dating back to the fourth and third millennium BC show that they derive a large amount of Devil's Gate related Amur-like ancestry and also have some admixture from West Eurasians. Their genetic profile is similar to Even people.
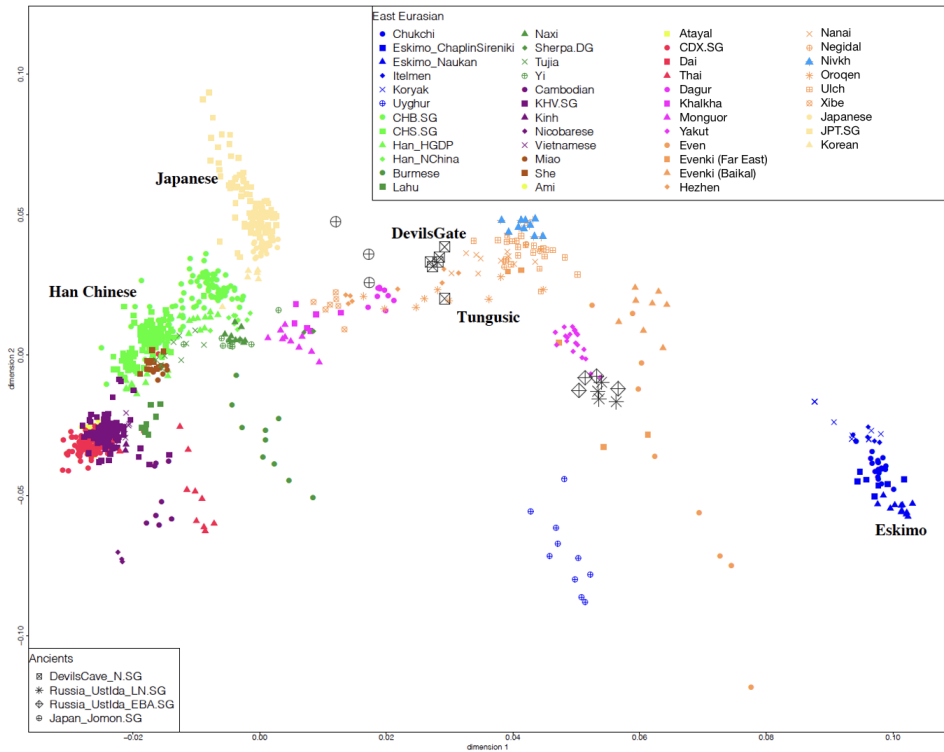

Figure 7: Principal Component Analysis of East Asian populations, projecting ancient Devil's Gate and Ust'-lda genomes onto the present-day speakers of Tungusic languages (adapted from Wang & Robbeets 2020)

This genome-wide perspective is corroborated by analyses of mitochondial and Y-chromosomal DNA. In the maternal line, there are only faint traces of a

genetic relationship between Tungusic-speaking populations in the Amur region, such as Negidal, Ulch and Udeghe and northern Tungusic populations, such as Even and Evenki, due to drift and admixture (Duggan et al. 2013). Nevertheless, the shared haplotypes found in these populations might be retentions from an earlier shared ancestral Tungusic population. Mitochondrial haplogroup frequencies show a cluster of Tungusic-speaking populations in the Amur region with Nivkh populations. The clustering of Even speakers with speakers of Yukaghir is seen as an implication of recent northward expansions of northern Tungusic speakers (Sukernik et al. 2012).

Tungusic speakers are further associated with the Y chromosomal haplogroup C3-M217, which is prevalent in Evenki and Even, as well as in other Tungusic speaking populations in the Amur River Basin including Oroqen, Ulch, Negidal, Udehe and Nanai (Malyarchuk et al. 2010; Duggan et al. 2013). This haplogroup is further well distributed among contemporary Mongolic and Nivkh-speaking populations and has been recovered in human remains of the Boisman culture (4825–2470 BC) in the Russian Far East (Yan et al. 2014; Wang et al. 2021).

It thus appears that both northern and southern Tungusic speaking populations share a proportion of their ancestry, which we refer to as the "Amur" genome. Mongolic and Turkic-speaking populations share a part of this "Amur" ancestry in spite of their increasing admixture with people of Western Eurasian ancestry from the first millennium BC onwards. (Jeong et al. 2018, 2019). Combined with recent analyses of ancient genomes from the West Liao River Basin (Ning et al. 2020), these results suggest that the Amur gene pool has long occupied the region from the Baikal to the West Liao River to the Russian Far East, at least for the last 10 000 years.

Since the Nivkh and the Tungusic-speaking populations share the same Amur ancestry, there are no traces of genetic admixture indicating population migration at the time of the agricultural dispersals. The long-term genetic continuity in the Amur basin is commonly used to argue against population migration and to support demic diffusion of agriculture into the Amur area (Siska et al. 2017). However, this conclusion does not take into account the increase in population density at the time of the agricultural expansions discussed in §5.2, which supports an alternative possibility that the incoming farmers may have shared an Amur-like genetic profile with the local populations (Cui et al. 2020; Wang & Robbeets 2020; Jeong et al. 2020; Li et al. 2020). Given the vast geographical reach of the Amur genetic profile, including the West Liao River region as well as the Russian Far East, a genetic admixture between Proto-Tungusic incoming farmers and Proto-Amuric local fishers would have led to an admixture of two

similar Amur genomes, like mixing two white paints together. Therefore, population migration and admixture are not expected to be visible in the genome. Bringing the archaeological, linguistic and genetic evidence together thus leaves room for agriculture-driven population migration and language shift spreading Proto-Tungusic to the Russian Far East in the Neolithic.

6 Conclusion

Quantitative methods, such as the Bayesian approach adopted in Oskolskaya et al. (2022), have much to offer: they can infer an internal family structure, calculate the statistical robustness of the proposed branches, estimate an absolute time depth within credible intervals without assuming a constant rate of change and help us to determine the location of the original homeland. Nevertheless, Bayesian results should be interpreted with caution, as they are dependent on the quality of the data input and the plausibility of the calibrations. Besides, they are limited in their abilities because even if they can infer information about the time and space of linguistic dispersals, they do not inform us about the natural and cultural environment of the ancient speakers: they can tell us where and when ancestral speech communities were located, but not why these people moved. In order to provide a better understanding of causalities in linguistic prehistory, we need to reinstate comparative historical linguistic tradition and, together with archaeology and genetics, integrate it into a holistic approach. In this paper, we attempted to take such an approach for the dispersal of Proto-Tungusic in time and space.

Combining the power of traditional comparative historical linguistics and computational phylogenetics, we used the recent Bayesian analysis provided by Oskolskaya et al. (2022) to quantify the likelihood of previously proposed classifications. We found that two classifications, namely the revised North-South classification (Figure 2b) and the Manchu-Tungusic classification (Figure 2b) were statistically robust, while other proposals could be excluded. Since we expressed the disagreement among different authors with regard to the exact configuration of the Tungusic tree in terms of probability, we were able to provide a quantitative basis to the ongoing discussions.

Chronologically, we estimated the break-up of Proto-Tungusic in the beginning of the first millennium AD and situated the homeland geographically in the area around or to the north of Lake Khanka.

Triangulating the linguistic evidence for Proto-Tungusic with evidence from archaeology and genetics, we argued for a language shift around 3300 BC, whereby some ancestral speakers of Proto-Amuric in the Russian Far East abandoned

their own language and adopted the Proto-Tungusic target language. The dispersal of Proto-Tungusic from the Liao River basin to the area around lake Khanka was probably caused by the expansion of millet agriculture and driven by population migration. The separation of the Manchuric branch, which may represent the first split in the family, can be associated with a return to hunting and fishing, in part because the conditions for agriculture worsened through climate change.

Abbreviations

PTg   Proto-Tungusic

Acknowledgements

Appendix A Reconstruction of basic sound inventories for Proto-Tungusic

Table 3: Reconstruction of the basic consonant inventory of Proto-Tungusic (Robbeets 2020, supplementary files)

PTg	J.	Ma.	X.	Or.	Ud.	N	B.N.	Ork.	Ul	Evn	Sol.	Evk.	Neg.
*p-	f-	f-	f-	x-	x-	p-	f- x-	p-	p-	h-	ø-	h-	x-
*-p-	f	f b ø	v ø	p w ø	p f w	p	f	p	p	b w ø	p w g ø	p b w ø	p w
*b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-
*-b-	b w ø	b f w ø	v ø	b w ø	b w ø	b w ø	w ø	b w ø	b w ø	b w ø	b p w ø	w ø	w ø
*t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-

PTg	J.	Ma.	X.	Or.	Ud.	N	B.N.	Ork.	Ul	Evn	Sol.	Evk.	Neg.
*-t-	t	t	t s	t	t	t	t	t	t	t	t	t	t
*d- *ji-	d- dʑi-	d- dʑi-	d- dʑi-	d-	d-	d- dʑi-	d- dʑi-	d- dʑi-	d-	d-	d-	d-	d-
*-d- *-ji-	d dʑi	d dʑi	d dʑi	d	d	d dʑi	d dʑi	d dʑi	d	d	d	d	d
*k-	x-	x-	x-	k- ø-	k- ø-	k- ø-	k- ø-	k- ø-	k- ø-	k- ø-	x- ø-	k- ø-	k- ø-
*-k-	k x ø	k x	k ɣ	k ø	k x ɣ ø	k ɣ ø	k ɣ ø	k ø	k ɣ ø	k	k x	k x	k x
*g-	g-	g-	g-	g- ŋ-	g- ŋ-	g- ø-	g- ø-	g- ŋ-	g- ŋ-	g- ŋ-	g- n-	g- ŋ-	g- ŋ
*-g-	ɣ w ø	ɣ w y ø	ø	ɣ w y ø	ɣ w y ø	ɣ w y ø	y ø	ɣ w y ø	ɣ w y ø	ɣ y	ɣ ø	ɣ	ɣ y w
*č-	č-	č-	č-	č-	č-	č-	č-	č- > t-	č- > t-	č-	s-	č-	č-
*-č-	č	č	č	č	s	č	č s	č > t	č	č	š	č	č
*x-	w- ø-	w- ø-	v- ø-	x- ø-	w- ø-	x- s-	x- s-	x- s-	x- s-	ø-	ø-	ø-	ø-
*-x-	x	x	x k ɣ	k	ø	x ø	x k	x ø	x ø	k	x	k	k x
*s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-
*-s-	s	s	s	s	s h ø	s	s	s	s	s	s	x	s
*m-	m-	m-	m-	m-	m-	m- ŋ-	m-	m-	m- ŋ-	m-	m-	m-	m-
*-m-	m	m	m	m	m	m	m	m	m	m	m	m	m
*n-	n-	n-	n-	n-	n-	n- l-	n- l-	n- l-	n- l-	n-	n-	n- l-	n-
*-n-	n	n	n	n	n	n	n	n	n	n	n	n	n
*-r-	r	r	r	y ø	y ø	r	r	r	r	r	r	r	y ø
*-l-	l	l	l	l	l	l	l n	l	l	l	l	l	l
*-y-	y	y	y	y	y	y	y	y	y	y	y	y	y

Table 4: Reconstruction of the basic vowel inventory of Proto-Tungusic (Robbeets 2020, supplementary files)

PTg	J.	Ma.	X.	Or.	Ud.	N	B.N.	Ork.	Ul	Evn	Sol.	Evk.	Neg.
*a	a	a	a	a	a	a	a	a	a	a	a	a	a
*e	e	e	e	e	e	e	e	e	e	e	e	e	e
*o	o	o	o	o	o	o	o	o	o	o	o	o	o
*ö	u	u	u	o u	o	u	u	o u	o u	o	u	u	u
*u	u	u	u	u	u	u	u	u	u	u ö	u ö	u ö	u ö
*ü	u ei	u ei	u i	i u	i y ø	u o	i o	i u	i u o	i	i	i u	i o
*i	i	i	i	i	i	i ị	i ị	i	i ị	i	i	i ị	i ị
*i(y)a	ie ⁱa	iya ai ⁱa	ia a	ⁱæ ei	ⁱæ æ a	ⁱa ea	ⁱæ	ị:	ị:	ia	ị:	ị:	ị:

References

Alonso de la Fuente, José A. 2011. *Tense, voice and Aktionsart in Tungusic. Another case of analysis to synthesis?* Wiesbaden: Harrassowitz.

Anderson, Gregory D. S. 1999. The Evenki of the lower Enisei valley. In Richard B. Lee & Richard Daly (eds.), *The Cambridge encyclopedia of hunters and gatherers*, 142–146. Cambridge: Cambridge University Press.

Benzing, Johannes. 1956. *Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik* (Akademie der Wissenschaften und der Literatur. Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse 11). Mainz: Akademie der Wissenschaften und der Literatur.

Bouckaert, Remco, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A. Suchard, Andrew Rambaut & Alexei J. Drummond. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10(4). e1003537.

Cincius, Vera I. 1949. *Sravnitel'naja fonetika tunguso-man'čžurskich jazykov [Comparative phonetics of Manchu-Tungus languages]*. Leningrad: Učpedgiz.

Cincius, Vera I. 1975/77. *Sravnitel'nyj slovar' tunguso-man'čžurskix jazykov [Comparative dictionary of the Tungus-Manchu languages]*. 2 vols. Leningrad: Nauka.

Cui, Yinqiu, Fan Zhang, Pengcheng Ma, Linyuan Fan, Chao Ning, Quanchao Zhang, Wei Zhang, Lixin Wang & Martine Robbeets. 2020. Bioarchaeological perspective on the expansion of the Transeurasian languages in the Amur River Basin. *Evolutionary Human Sciences* 2. e15.

Derevyanko, Anatoliy P. 1976. *Priamur'ye v drevnosti (do nachala nashej ery) [Priamur'e in antiquity before the beginning of our era]*. Novosibirsk: Nauka.

Doerfer, Gerhard. 1978. Classification problems of Tungus. In Gerhard Doerfer & Michael Weiers (eds.), *Tungusica, vol.1: Beiträge zur nordasiatischen Kulturgeschichte*, 1–26. Wiesbaden: Harrassowitz.

Drennan, Robert D., Xueming Lu & Christian E. Peterson. 2017. A place of pilgrimage? Niuheliang and its role in Hongshan society. *Antiquity* 91. 43–56.

Duggan, Ana T., Mark Whitten, Victor Wiebe, Michael Crawford, Anne Butthof, Victor Spitsyn, Sergey Makarov, Innokentiy Novgorodov, Vladimir Osakovsky & Brigitte Pakendorf. 2013. Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One* 8(12). e83570.

Dybo, Anna V. & Evgeniya Korovina. 2019. *Altaic linguistics and mythology: geography or relationship?* Paper presented at the conference "Centenary of human population genetics", Biology Department of Moscow State University May 29-June 1.

Fuller, Dorian Q. & Ling Qin. 2009. Water management and labour in the origins and dispersal of Asian rice. *World Archaeology* 41. 88–111.

Georg, Stefan. 2004. Unreclassifying Tungusic. In Carsten Naeher (ed.), *Proceedings of the First International Conference on Manchu-Tungus studies*, vol. 2 (Tunguso Sibirica 9), 45–57. Wiesbaden: Harrassowitz.

Gorelova, Liliya M. 2002. *Manchu grammar*. Leiden: Brill.

Helimski, Eugene. 1985. Samodiisko-tungusskie leksicheskie sviazi i ikh etno-istoricheskie implikatsii [Samoyedic-Tungusic lexical correspondences and their ethno-historical implications]. In Elizaveta I. Ubriatova (ed.), *Uralo-altaistika*, 206–212. Novosibirsk: Nauka.

Holman, Eric W., Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Johann-Mattis List & Dmitry Egorov. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology* 52(6). 841–875.

Hudson, Mark J. 2020. Language dispersals and the "Secondary Peoples' Revolution": A historical anthropology of the Transeurasian unity. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 806–813. Oxford: Oxford University Press.

Ikegami, Jirō. 1974. Versuch einer Klassifikation der tungusischen Sprachen. In Gyorgy Hazai & Peter Zieme (eds.), *Sprache, Geschichte und Kultur der altaischen Völker* (Schriften zur Geschichte und Kultur des Alten Orients 5), 271–272. Berlin: Akademie Verlag.

Janhunen, Juha. 1996. *Manchuria. An ethnic history* (Mémoires de la Société Finno-Ougrienne 222). Helsinki: Suomalais-Ugrilainen Seura.

Janhunen, Juha. 2012. The expansion of Tungusic as an ethnic and linguistic process. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 5–16. Wiesbaden: Harrassowitz.

Janhunen, Juha. 2016. Reconstructio externa linguae Ghiliacorum. *Studia Orientalia* 117. 3–27.

Jeong, Choongwon, Oleg Balanovsky, Elena Lukianova, Nurzhibek Kahbatkyzy, Pavel Flegontov, Valery Zaporozhchenko, Alexander Immel, Chuan-Chao Wang, Olzhas Ixan, Elmira Khussainova, Bakhytzhan Bekmanov, Victor Zaibert, Maria Lavryashina, Elvira Pocheshkhova, Yuldash Yusupov, Anastasiya Agdzhoyan, Sergey Koshel, Andrei Bukin, Pagbajabyn Nymadawa, Shahlo Turdikulova, Dilbar Dalimova, Mikhail Churnosov, Roza Skhalyakho, Denis Daragan, Yuri Bogunov, Anna Bogunova, Alexandr Shtrunov, Nadezhda Dubova, Maxat Zhabagin, Levon Yepiskoposyan, Vladimir Churakov, Nikolay Pislegin, Larissa Damba, Ludmila Saroyants, Khadizhat Dibirova, Lubov Atramentova, Olga Utevska, Eldar Idrisov, Evgeniya Kamenshchikova, Irina Evseeva, Mait Metspalu, Alan K. Outram, Martine Robbeets, Leyla Djansugurova, Elena Balanovska, Stephan Schiffels, Wolfgang Haak, David Reich & Johannes Krause. 2019. The genetic history of admixture across inner Eurasia. *Nature Ecology & Evolution* 3. 966–976.

Jeong, Choongwon, Chao Ning & Chuan-Chao Wang. 2020. The Oxford guide to the Transeurasian languages. In Martine Robbeets & Alexander Savelyev (eds.), *Transeurasian unity from a population genetic perspective*, 784–793. Oxford: Oxford University Press.

Jeong, Choongwon, Shevan Wilkin, Tsend Amgalantugs, Abigail S. Bouwman, William Taylor, Richard W. Hagan, Sabri Bromage, Soninkhishig Tsolmon, Christian Trachsel, Jonas Grossmann, Judith Littleton, Cheryl A. Makarewicz, John Krigbaum, Marta Burri, Ashley Scott, Ganmaa Davaasambuu, Joshua Wright, Franziska Irmer, Erdene Myagmar, Nicole Boivin, Martine Robbeets, Frank J. Rühli, Johannes Krause, Bruno Frohlich, Jessica Hendy & Christina Warinner. 2018. Bronze age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proceedings of the National Academy of Sciences* 115(48). E11248–E11255.

Johanson, Lars. 2002. Contact-induced change in a code-copying framework. In Mari C. Jones & Edith Esch (eds.), *Language change. The interplay of internal, external and extra-linguistic factors*, 285–313. Berlin: Mouton de Gruyter.

Kormušin, Igor V. 1998. *Udyhejskij jazyk [The Udihe language]*. Moscow: Nauka.

Korovina, Evgeniya. 2011. *Leksika prirodnogo i kul'turnogo okruženiya v Tunguso-Man'čžurskich jazykach (v istoriko-tipologičeskom osveščenii) [Tungus-Manchu vocabulary related to natural environment and cultural activities (in historical and typological perspective)]*. Moscow: Russian State University of Humanities. BA dissertation.

Leipe, Christian, T. Long, Elena A. Sergusheva, Maike Wagner & P.E. Tarasov. 2019. Discontinuous spread of millet agriculture in eastern Asia and prehistoric population dynamics. *Science Advances* 5. eaax6225.

Li, Tao, Martine Robbeets, Chao Ning, Irina Zhushchikhovskaya & Mark Hudson. 2020. Millet agriculture dispersed from Northeast China to the Russian Far East: Integrating archaeology, genetics, and linguistics. *Archaeological Research in Asia* 22. 100177.

Mallory, James, Anna Dybo & Oleg Balanovsky. 2019. The impact of genetics research on archaeology and linguistics in Eurasia. *Russian Journal of Genetics* 55. 1472–1487.

Malyarchuk, Boris, Miroslava Derenko, Galina Denisova, Marcin Wozniak, Tomasz Grzybowski, Irina Dambueva & Ilia Zakharov. 2010. Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Annals of Human Genetics* 74. 539–546.

Menges, Karl Heinrich. 1968. Tungusologie. In Walter Fuchs (ed.), *Die tungusischen Sprachen* (Handbuch der Orientalistik I.5,3), 21–256. Leiden: Brill.

Miyamoto, Kazuo. 2014. The initial spread of early agriculture into Northeast Asia. *Asian Archaeology* 3. 1–12.

Nelson, Sarah, Irina Zhushchikhovskaya, Tao Li, Chao Ning, Mark Hudson & Martine Robbeets. 2020. Tracing population movements in ancient East Asia through the linguistics and archaeology of textile production. *Evolutionary Human Sciences* 2. e5.

Ning, Chao, Tianjiao Li, Ke Wang, Tao Li, Fan Zhang, Xiyan Wu, Shizhu Gao, Quanchao Zhang, Hai Zhang, Mark J. Hudson, Sihao Wu, Yanming Fang, Chen Liu, Chunyan Feng, Wei Li, Tao Han, Ruo Li, Jian Wei, Yonggang Zhu, Yawei Zhou, Chuan-Chao Wang, Shengying Fan, Zenglong Xiong, Guanghui Dong, Zhouyong Sun, Maolin Ye, Lei Sun, Xiaohong Wu, Fawei Liang, Yanpeng Cao, Xingtao Wei, Hong Zhu, Hui Zhou, Johannes Krause, Martine Robbeets, Choongwon Jeong & Yinqui Cui. 2020. Ancient genomes from northern China

suggest links between subsistence changes and human migration. *Nature Communications* 11(1). 1–9.

Oskolskaya, Sofia, Ezequiel Koile & Martine Robbeets. 2022. A Bayesian approach to the classification of Tungusic languages. *Diachronica* 39(1). 128–158.

Pakendorf, Brigitte. 2007. *Contact in the prehistory of the Sakha (Yakuts): Linguistic and genetic perspectives*. Utrecht: LOT.

Perekhvalskaya, Elena. 2022. From consonant to tone: Laryngealized and pharyngealized vowels in Udihe. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 227–262. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053371.

Peterson, Christian E., Xueming Lu, Robert D. Drennen & Da Zhu. 2010. Hongshan chiefly communities in Neolithic northeastern China. *Proceedings of the National Academy of Sciences USA* 107. 5756–5761.

Pevnov, Aleksandr M. 2012. The problem of the localization of the Manchu-Tungusic homeland. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 17–40. Wiesbaden: Harrassowitz.

Pugach, Irina, Rostislav Matveev, Viktor Spitsyn, Sergey Makarov, Innokentiy Novgorodov, Vladimir Osakovsky, Mark Stoneking & Brigitte Pakendorf. 2016. The complex admixture history and recent southern origins of Siberian populations. *Molecular Biology and Evolution* 33(7). 1777–1795.

Robbeets, Martine. 2015. *Diachrony of verb morphology. Japanese and the Transeurasian languages* (Trends in linguistics 291). Berlin: De Gruyter Mouton.

Robbeets, Martine. 2017a. Japanese, Korean and the Transeurasian languages. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics* (Cambridge Handbooks in Language and Linguistics), 586–626. Cambridge: Cambridge University Press.

Robbeets, Martine. 2017b. Proto-Trans-Eurasian: Where and when? *Man in India* 97. 19–46.

Robbeets, Martine. 2020. Basic vocabulary in the Transeurasian languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 645–659. Oxford: Oxford University Press.

Robbeets, Martine & Remco Bouckaert. 2018. Bayesian phylolinguistics reveals the internal structure of the Transeurasian family. *Journal of Linguistic Evolution* 3. 145–162.

Robbeets, Martine, Juha Janhunen, Alexander Savelyev & Evgeniya Korovina. 2020. The homelands of the individual proto-languages: Where, what and

when? In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 753–771. Oxford: Oxford University Press.

Robbeets, Martine & Alexander Savelyev (eds.). 2020. *The Oxford guide to the Transeurasian languages*. Oxford: Oxford University Press.

Sergusheva, Elena A. & Yury E. Vostretsov. 2009. The advance of agriculture in the coastal zone of East Asia. In Andrew Fairbairn & Ehud Weiss (eds.), *From foragers to farmers: Papers in honour of Gordon C. Hillman*, 205–219. Oxford: Oxbow books.

Siska, Veronika, Eppie R. Jones, Jeon Sungwon, Bhak Youngjune, Kim Hak-Min, Cho Yun Sung, Kim Hyunho, Lee Kyusang, Elizaveta Veselovskaya, Tatiana Balueva, Marcos Gallego-Llorente, Michael Hofreiter, Daniel G. Bradley, Anders Eriksson, Ron Pinhasi, Bhak Jong & Andrea Manica. 2017. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Science Advances* 3. 1–10.

Starikovskaya, Elena B., Rem I. Sukernik, Olga A. Derbeneva, Natalia V. Volodko, Eduardo Ruiz-Pesini, Antonio Torroni, Michael D. Brown, Marie T. Lott, Seyed H. Hosseini, Kirsi Huoponen & Douglas C. Wallace. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Annals of Human Genetics* 69(1). 67–89.

Starostin, Sergei, Anna Dybo & Oleg Mudrak. 2003. *Etymological dictionary of the Altaic languages*. Leiden: Brill.

Stevens, Chris J. & Dorian Q. Fuller. 2017. The spread of agriculture in eastern Asia: Archaeological bases for hypothetical farmer/language dispersals. *Language Dynamics and Change* 7. 152–186.

Sukernik, Rem I., Natalia V. Volodko, Ilya O. Mazunin, Nikolai P. Eltsov, Stanislav V. Dryomov & Elena B. Starikovskaya. 2012. Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: Contribution to prehistory of native Siberians and their affinities to Native Americans. *American Journal of Physical Anthropology* 148. 123–138.

Sunik, Orest P. 1959. *Tunguso-man'čžurskije jazyki [Manchu-Tungus languages]* (Mladopi'mennyje jazyki narodov SSSR). Moscow: Nauka.

Thomason, Sarah G. & Terence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tugolukov, V. A. 1980. Etniceskie korni tungusov [Ethnic roots of the Tungus people]. In I. S Gurvich (ed.), *Etnogenez narodov severa [Ethnogenesis of the peoples of the North]*, 152–176. Moscow: Nauka.

Van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Universitätsverlag C. Winter.

Vasilevič, Glafira M. 1960. *K voprosu o klassifikacii tunguso-man'čžurskih jazykov [About the question of classification of Manchu-Tungus languages]* (Voprosy jazykoznanija 2). Moscow: Nauka.

Vasilevič, Glafira M. 1969. *Évenki: Istoriko-etnografičeskie očerki (XVIII – načalo XX v.)* Leningrad: Nauka.

Vovin, Alexander. 1993. Towards a new classification of Tungusic languages. *Eurasian Studies Yearbook* 65. 99–113.

Wang, Chuan-Chao & Martine Robbeets. 2020. The homeland of Proto-Tungusic inferred from contemporary words and ancient genomes. *Evolutionary Human Sciences* 2. E8.

Wang, Chuan-Chao, Hui-Yuan Yeh, Alexander N Popov, Hu-Qin Zhang, Hirofumi Matsumura, Kendra Sirak, Olivia Cheronet, Alexey Kovalev, Nadin Rohland, Alexander M. Kim, Rebecca Bernardos, Dashtseveg Tumen, Jing Zhao, Yi-Chang Liu, Jiun-Yu Liu, Matthew Mah, Swapan Mallick, Ke Wang, Zhao Zhang, Nicole Adamski, Nasreen Broomandkhoshbacht, Kimberly Callan, Brendan J. Culleton, Laurie Eccles, Ann Marie Lawson, Megan Michel, Jonas Oppenheimer, Kristin Stewardson, Shaoqing Wen, Shi Yan, Fatma Zalzala, Richard Chuang, Ching-Jung Huang, Chung-Ching Shiung, Yuri G. Nikitin, Andrei V. Tabarev, Alexey A. Tishkin, Song Lin, Zhou-Yong Sun, Xiao-Ming Wu, Tie-Lin Yang, Xi Hu, Liang Chen, Hua Du, Jamsranjav Bayarsaikhan, Enkhbayar Mijiddorj, Diimaajav Erdenebaatar, Tumur-Ochir Iderkhangai, Erdene Myagmar, Hideaki Kanzawa-Kiriyama, Msato Nishino, Ken-ichi Shinoda, Olga A. Shubina, Jianxin Guo, Qiongying Deng, Longli Kang, Dawei Li, Dongna Li, Rong Lin, Wangwei Cai, Rukesh Shrestha, Ling-Xiang Wang, Lanhai Wei, Guangmao Xie, Hongbing Yao, Manfei Zhang, Guanglin He, Xiaomin Yang, Rong Hu, Martine Robbeets, Stephan Schiffels, Douglas J. Kennett, Li Jin, Hui Li, Johannes Krause, Ron Pinhasi & David Reich. 2021. The genomic formation of human populations in East Asia. *Nature* 591. 413–419.

Whaley, Lindsay, Lenore A. Grenoble & Fengxiang Li. 1999. Revisiting Tungusic classification from the bottom up: A comparison of Evenki and Oroqen. *Language* 75. 286–321.

Whaley, Lindsay & Sofia Oskolskaya. 2020. The classification of the Tungusic languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 81–91. Oxford: Oxford University Press.

Wichmann, Søren, André Müller & Veveka Velupillai. 2010. Homelands of the world's language families. A quantitative approach. *Diachronica* 27(2). 247–276.

Winford, Donald. 2013. On the unity of contact phenomena: The case for imposition. In Carole de Feral (ed.), *In and out of Africa: Languages in question. In honour of Robert Nicolai*, 43–71. Leuven: Peeters.

Yan, Shi, Chuan-Chao Wang, Hong-Xiang Zheng, Wei Wang, Zhen-Dong Qin, Lan-Hai Wei, Yi Wang, Xue-Dong Pan, Wen-Qing Fu, Yun-Gang He, Li-Jun Xiong, Wen-Fei Jin, Shi-Lin Li, andYu An, Hui Li & Li Jin. 2014. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLOS One* 9(8). e105691.

Yanuševič, Zoya, Yuri Vostretsov & Sofiya V. Makarova. 1990. *The palaeoethnobotanical finds in Primorye*. Vladivostok: Nauka.

Zgusta, Richard. 2015. *The peoples of Northeast Asia through time: Precolonial ethnic and cultural processes along the coast between Hokkaido and the Bering Strait*. Leiden: Brill.