

Horizontal transfer and phylogenetic calibration in linguistics: a Bayesian approach

Luke Maurits¹ Mervi de Heer² Michael Dunn² Outi Vesakoski¹
¹University of Turku, Turku, Finland ²University of Uppsala, Uppsala, Sweden

Introduction & Summary

- Dating **language divergence** poses a challenge in linguistics
- Bayesian methods and results of historical-comparative linguistics together help answer the challenge.
- Absolute calibrations** are usually used to obtain divergence times of a linguistic subfamilies in calendar years. However, they are hard to establish[1], especially for families without written records.
- Most linguistic datasets contain results of **horizontal transfer**, e.g. loanwords, due to language contacts.
- Although often thought of as “noise”, **loanwords carry valuable timing information** as contact happens between contemporaneous languages.
- Uralic has long-standing contacts with Indo-European, a family which has been extensively researched and whose timing is better understood.
- We test “**relative calibration**” of linguistic phylogenies using Indo-European loanwords acquired into the Uralic family and the effects on dating Proto-Uralic and intermediate protolanguages.

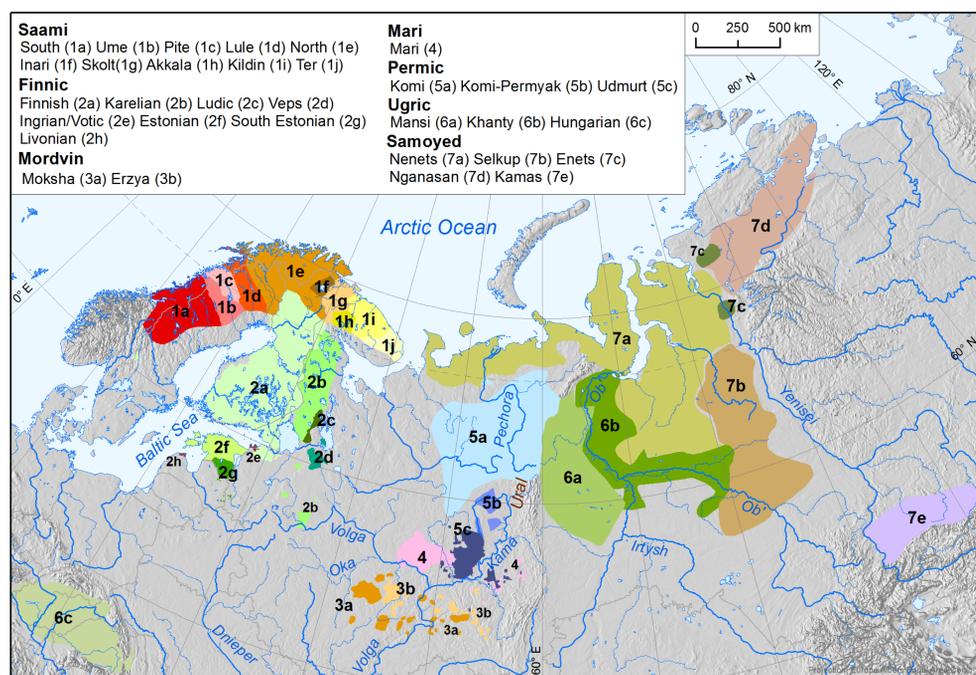


Figure 1: The Uralic language family. Map by T. Rantanen (BEDLAN)

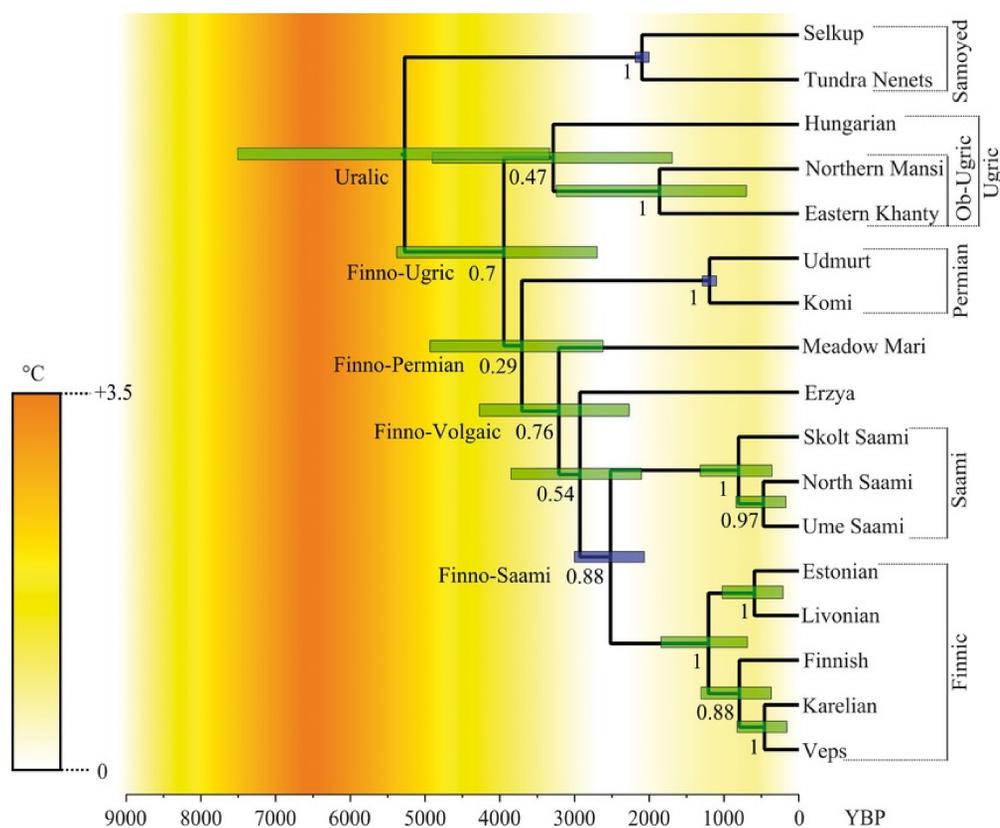


Figure 2: An example of a calibrated phylogeny of the Uralic language family using narrow distribution priors for absolute calibrations only[2]

Materials

Uralex 1.0. and its tagged borrowings

- Basic vocabulary dataset of 26 Uralic languages [3]
- Known borrowings tagged using etymological literature [4]
- Certainty estimate given using evaluative literature

Example: borrowings in Standard Estonian

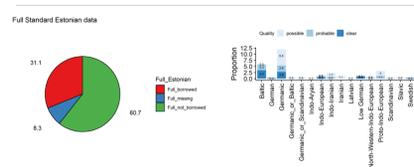


Figure 3: Proportion of borrowings, source languages and certainty estimations in the Uralex Standard Estonian sample

Indo-European timings

- The Indo-European language family is well-studied and has written records, which has enabled thoroughly calibrated phylogenetic analyses to estimate divergence times.[5].

Methods

- Simultaneous sampling of Uralic and IE trees in a single MCMC chain using BEAST 2.
- Uralic tree includes UraLex data, whose evolution is modelled with a relaxed lognormal clock.
- The Indo-European tree has no data, but interior node times are constrained to match published posteriors.
- Sampling is constrained so that trees are only accepted if there is some non-zero overlap in the lifespan of the ancestors of each pair of subfamilies with identified borrowings.

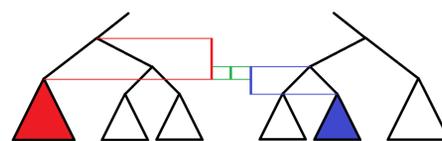


Figure 4: Compatible trees

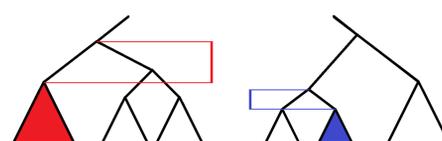


Figure 5: Incompatible trees

Results & Conclusion

- Our relative calibration method estimates Proto-Uralic divergence at approx 5,250 YBP (95% HPD 4,000-6,600)
- Analysing same data with absolute calibrations based on loanwords yields a much older date of approx 6,700 YBP[6].
- Both estimates older than a recent popular hypothesis in Uralic linguistics of ca. 4000 YBP[7, 8].

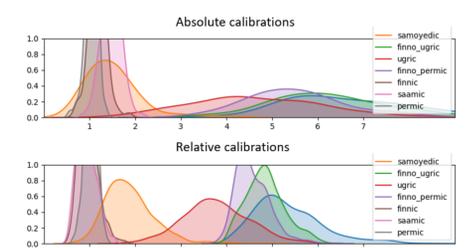


Figure 6: Posterior age estimates for different Uralic subfamilies for relative vs absolute calibrations.

- The histories of language families known to have been in contact should be co-estimated to ensure consistency with linguistic knowledge and make use of all available data.

References

- L. Maurits, M. de Heer, T. Honkola, M. Dunn, and O. Vesakoski. Best practices in justifying calibrations for dating language families. *Journal of Language Evolution*, under review.
- T. Honkola, O. Vesakoski, K. Korhonen, J. Lehtinen, K. Syrjärdnen, and N. Wahlberg. Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology*, 26:1244–1253, 2013.
- K. Syrjänen, J. Lehtinen, O. Vesakoski, M. de Heer, T. Suutari, M. Dunn, U. Määttä, and U. Leino. lexibank/uralex: UraLex basic vocabulary dataset (Version v1.0) [data set]. *Zenodo*, 2018.
- M. de Heer and et al. Borrowings in a Uralic basic vocabulary dataset. in preparation.
- R. Bouckaert, P. Lemey, M. Dunn, S. J. Greenhill, A. V. Alekseyenko, A. J. Drummond, R. D. Gray, M. A. Suchard, and Q. D. Atkinson. Mapping the origins and expansion of the indo-european language family. *Science*, 337(6097):957–960, 2012.
- O. Vesakoski, S. Översti, L. Maurits, and et al. Timing and patterns of proto-uralic and proto-fenno-ugric disintegrations. in preparation.
- P. Kallio. Suomen kantakielten absoluuttista kronologiaa. *Virtittäjä*, 110:2–25, 2006.
- J. P. Häkkinen. Kantauralin ajoitus ja paikannus perustelut puntarissa. *Suomalais-Ugrilaisen Seuran Aikakauskirja*, 92:9–56, 2009.

Acknowledgements



Contact Information

- Web: <https://bedlan.net>
- Email: luke@maurits.id.au, mervi.deheer@moderna.uu.se

BEDLAN
 BIOLOGICAL EVOLUTION AND
 DIVERSIFICATION OF LANGUAGES