

"Bayesian reconstruction of the Transeurasian language family tree using grammatical features of modern and extinct languages"
Nataliia Neshcheret & Denise Kühnert

The Transeurasian macrofamily is a unity of 5 language families: Turkic, Mongolic, Tungusic, Koreanic and Japonic that is gradually gaining acceptance in linguistics as a genealogical grouping. Previous studies dating the Transeurasian language family through a range of different methods applied to lexical data are inconclusive. Three different approaches yield dates around 6000 BC (lexicostatistics by Starostin et al. 2003 and Blažek and Schwarz 2014, the algorithm behind Automated Similarity Judgement Program). A recent study suggests a much younger date of 4700 BC based on Bayesian phylogenetic analysis of a 200 cognate coded list of basic vocabulary (Robbeets and Bouckaert 2018). In order to verify the topology and infer the most recent common ancestor of the tree and its subclades, we will apply Bayesian phylogenetic analysis to 228 structural features coded for modern and extinct Transeurasian languages. A prerequisite for the dating (without strong prior information) is that there is correlation between time and root-to-tip divergence, called phylogenetic signal in the data, which we will test prior to the phylogenetic analysis. We expect the inclusion of extinct languages to empower robust estimation of the tree height and topology without needing node calibrations.