

02: Co-estimating Human Mobility and Language Dispersal with Ancient DNA and Linguistic Data

Lead supervisor:

Dr. Stephan Schiffels
Max Planck Institute for the Science of Human History
Department of Archaeogenetics
Kahlaische Str. 10
07745 Jena, Germany
Email: schiffels@shh.mpg.de
<http://www.shh.mpg.de/person/44297/25522>

Co-Supervisor:

Prof. Dr. Martin Kümmel
Friedrich-Schiller-Universität Jena
Institut für Orientalistik, Indogermanistik, Ur- und Frühgeschichtliche Archäologie
Seminar für Indogermanistik
Zwätzengasse 12a
07743 Jena, Germany
E-Mail: martin-joachim.kuemmel@uni-jena.de
https://www.oriindufa.uni-jena.de/en/k%C3%BCmmel_martin.html

Involved subjects: Population Genetics, Linguistics, Spatial Modeling, Demographic history

Number of positions requested: 1

Abstract:

Human mobility is a key factor for the spread of genes and languages, but it affects them in different ways. While genes spread exclusively by movements of people, cultural traits such as languages can spread independently. The goal of this PhD project is to explore the extent of gene-language co-evolution in Eurasia, with a focus on regions in which Indo-European languages are spoken today, using a novel approach to jointly estimating human mobility and language dispersal from both genetic and linguistic data.

As a starting point, we will build on a recent model by Loog et al. (PNAS, 2017), who used published ancient genomic data from Europe to estimate human mobility through time. Their approach is based on the simple idea that mobile societies will cause lower trait correlations in space relative to time, while less mobile groups will cause stronger trait correlations in space compared to time. While Loog et al. used genetic data alone, we here want to extend it to jointly use genetic and linguistic data. We will focus on Western and Southern Eurasia in the last 3,000 years, a time period and region for which we have extensive genetic data as well as some information on the spoken languages. The genetic dataset will involve hundreds of published and unpublished samples with genome-wide as well as mitochondrial and Y-chromosomal data, and with direct radiocarbon-dating information. The linguistic dataset will include present-day as well as historical language distributions, as inferred from historical and onomastic sources, where uncertainty in historical language affiliations will be expressed probabilistically.

For joint modeling of genetic and linguistic data, we need to define joint trait distances, which besides the standard genetic distance requires also distances between languages, for which two approaches will be explored. First, language affiliations can be used as simple categorical variable (two individuals belong to the same language, with some probability, or not). Second, basic vocabulary and cognate lists available from standard databases for Indo-European (as provided by the MPI-SHH) can be used to quantitatively compare languages. The joint genetic and linguistic data can then be co-analysed in

an extended model based on Loog et al., ultimately providing estimates of mobility as well as language dispersal through time. Importantly, such joint estimates will show to what extent languages spread via human migrations vs. cultural diffusion.

The ideal candidate for this project has a mathematical or computer science background and a sincere interest in human history and languages. We roughly envision in year 1: introductory reading and literature review, and compilation of the datasets; in year 2: implementing the Loog et al. model with the proposed extension to include linguistic information; in year 3: preparation of publications, writing the thesis and identifying new challenges and future work.

Key Reference

Loog, Liisa, Marta Mirazón Lahr, Mirna Kovacevic, Andrea Manica, Anders Eriksson, and Mark G. Thomas. 2017. "Estimating Mobility Using Sparse Data: Application to Human Genetic Variation." *Proceedings of the National Academy of Sciences of the United States of America* 114 (46): 12213–18.