

Genes and language in the prehistory of the Uralic speaking peoples

Outi Vesakoski



**UNIVERSITY
OF TURKU**

Mervi de Heer, Michael Dunn



UPPSALA
UNIVERSITET



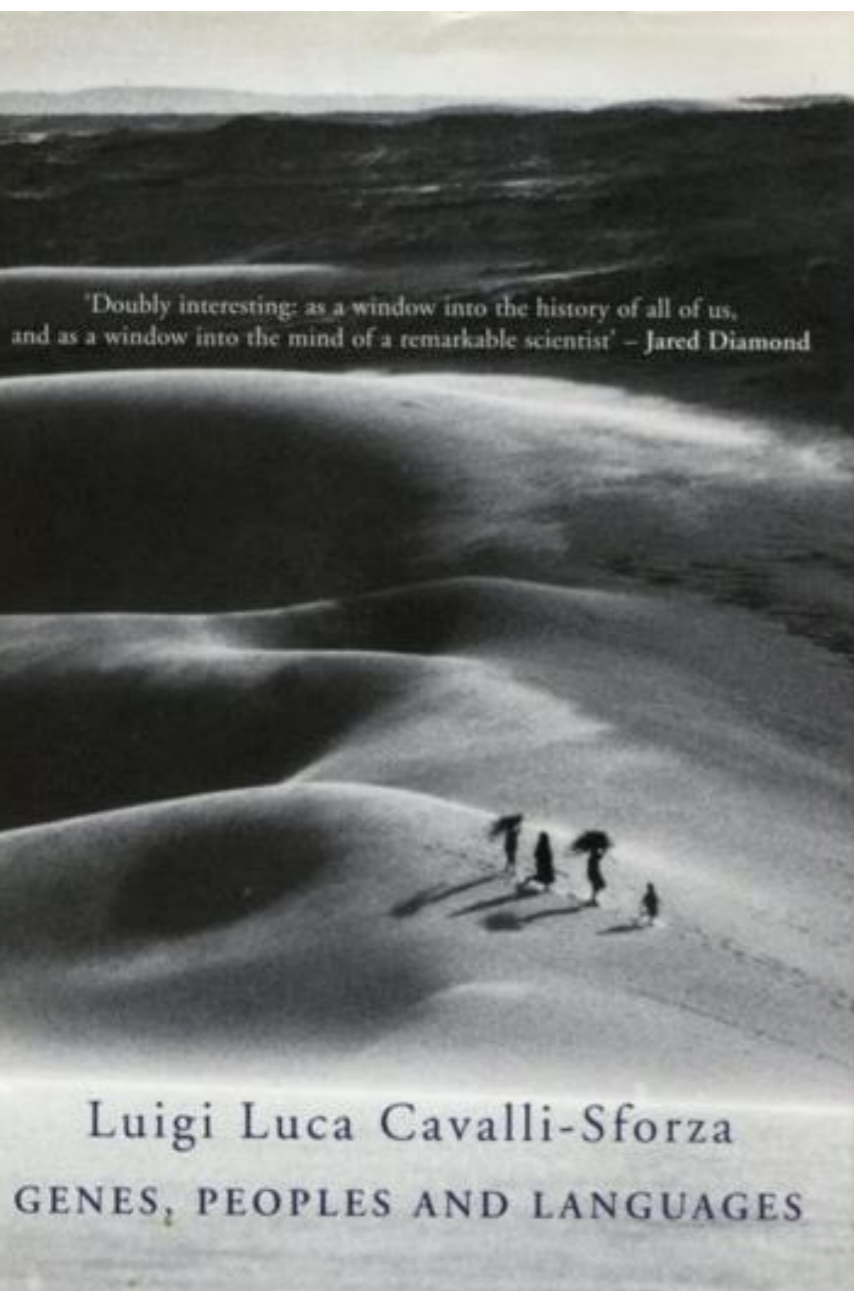
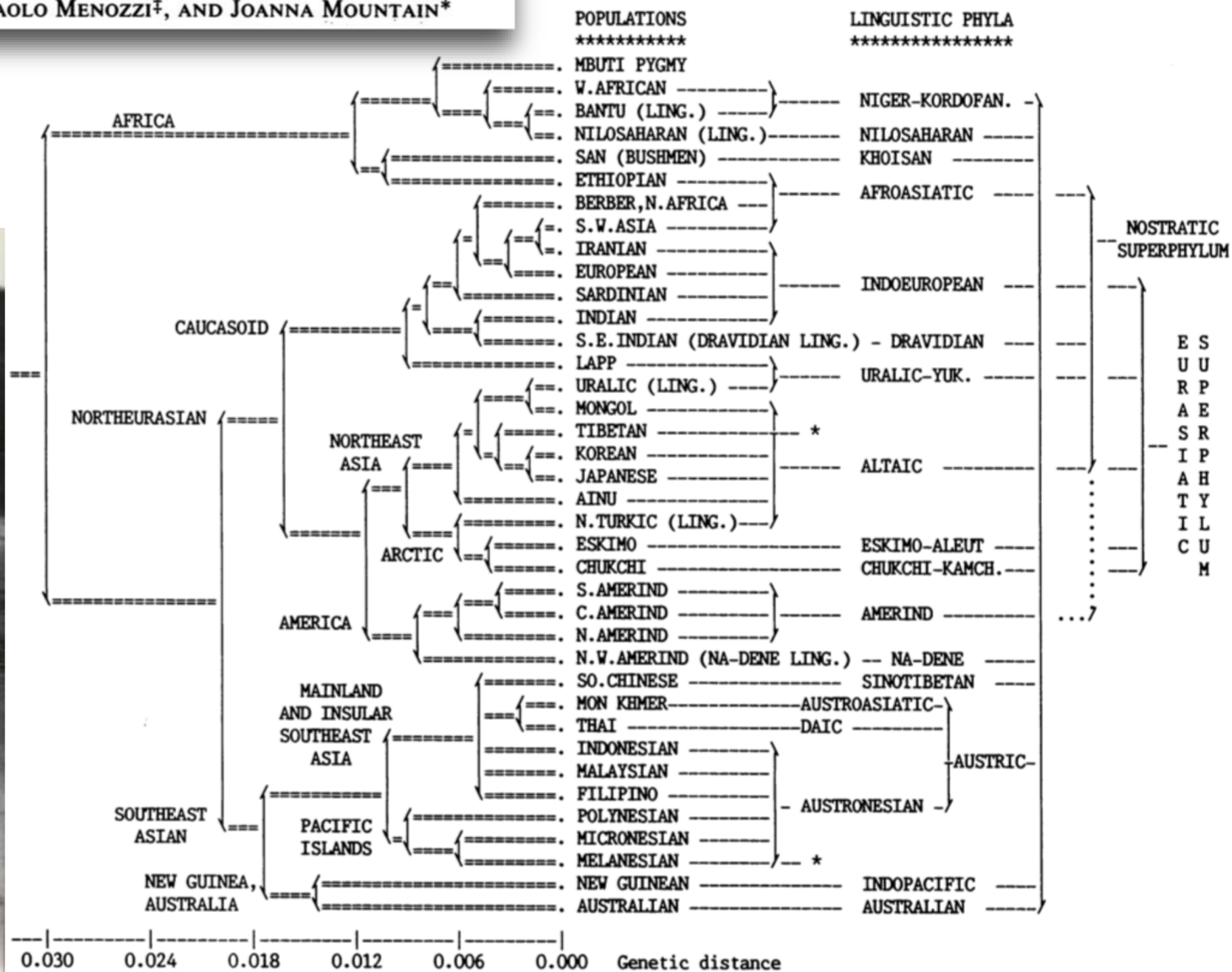
Art by Kerttu Majander



Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data

(origin of modern humans/phylogenetic trees/paleoanthropology)

LUIGI LUCA CAVALLI-SFORZA*, ALBERTO PIAZZA†, PAOLO MENOZZI‡, AND JOANNA MOUNTAIN*



'Doubly interesting: as a window into the history of all of us, and as a window into the mind of a remarkable scientist' - Jared Diamond

Luigi Luca Cavalli-Sforza
 GENES, PEOPLES AND LANGUAGES

Melanesian and Asian Origins of Polynesians: mtDNA and Y Chromosome Gradients Across the Pacific

Manfred Kayser,*† Silke Brauer,*†‡ Richard Cordaux,§ Amanda Casto,* Oscar Lao,†‡ Lev A. Zhivotovsky,|| Claire Moyse-Faurie,¶ Robb B. Rutledge,# Wulf Schiefenhoefel,** David Gil,†† Alice A. Lin,‡‡ Peter A. Underhill,‡‡ Peter J. Oefner,§§ Ronald J. Trent,||| and Mark Stoneking*

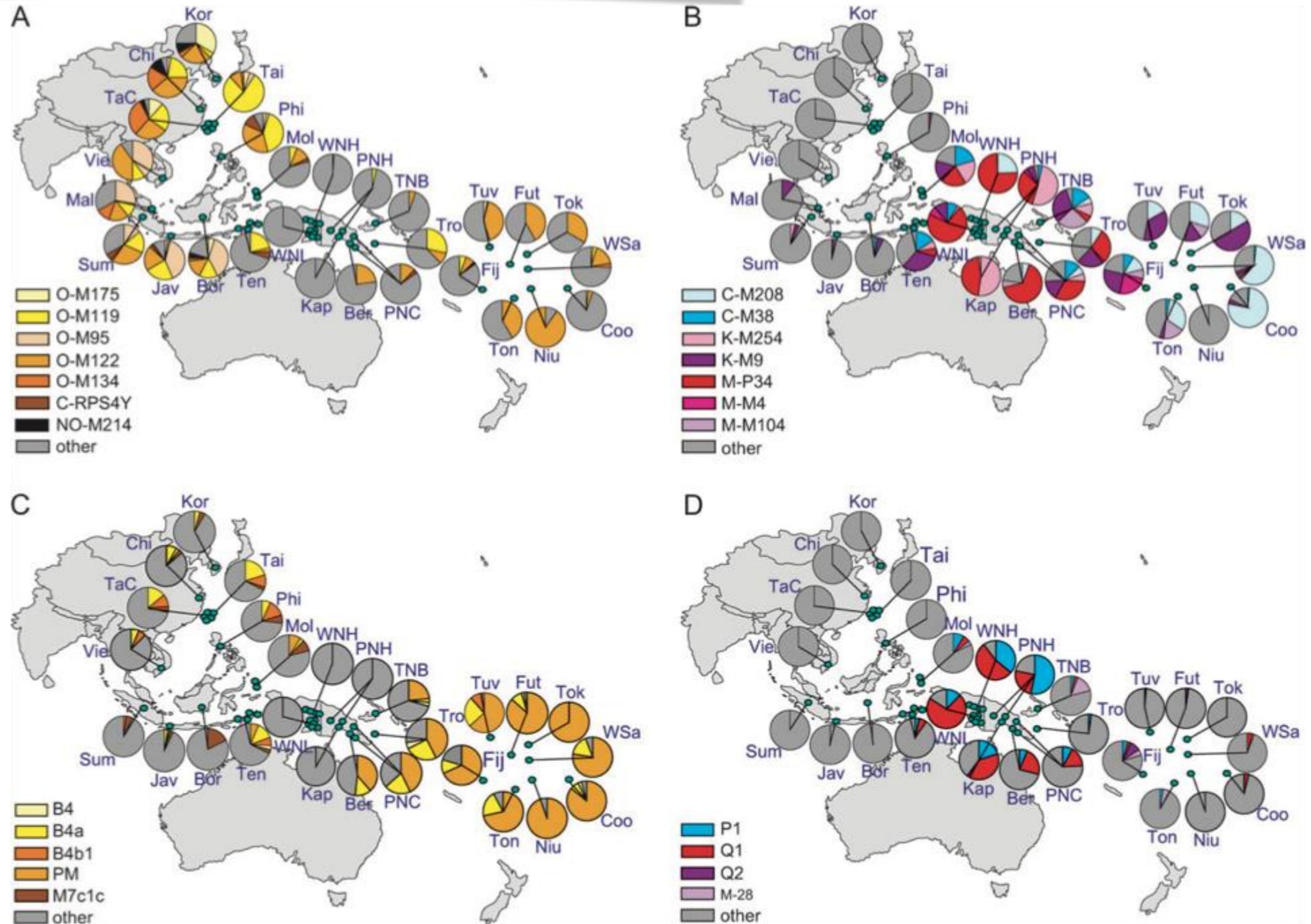
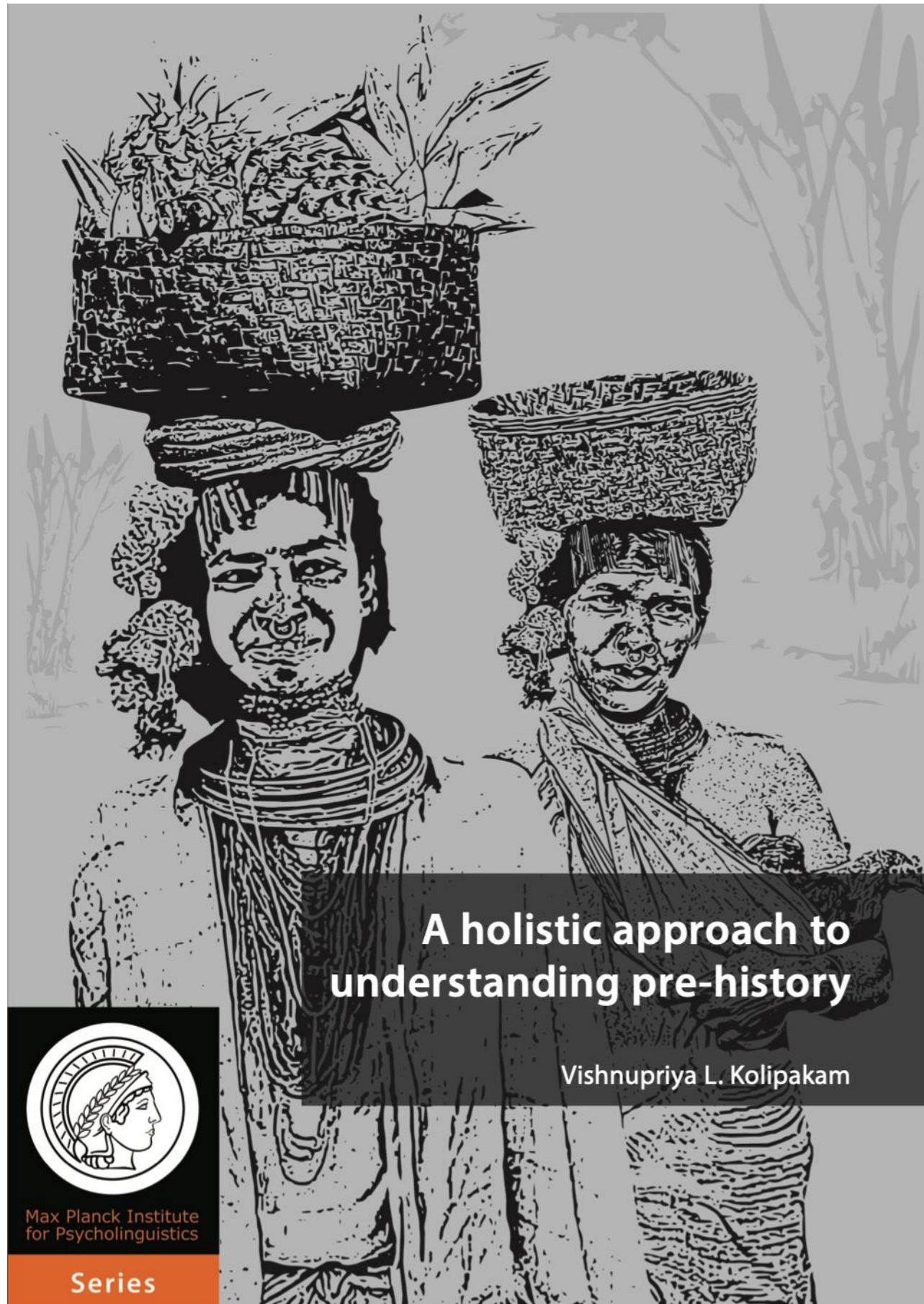


FIG. 1.—Frequency distribution of (A, B) NRY and (C, D) mtDNA haplogroups found in Polynesia with a genetic origin in (A, C) Asia or (B, D) Melanesia.



Kinship terminologies



Postmarital residence



Sex-biased gene flow

2018

Genetic Structure of Human Populations

Noah A. Rosenberg,^{1*} Jonathan K. Pritchard,² James L. Weber,³
Howard M. Cann,⁴ Kenneth K. Kidd,⁵ Lev A. Zhivotovsky,⁶
Marcus W. Feldman⁷

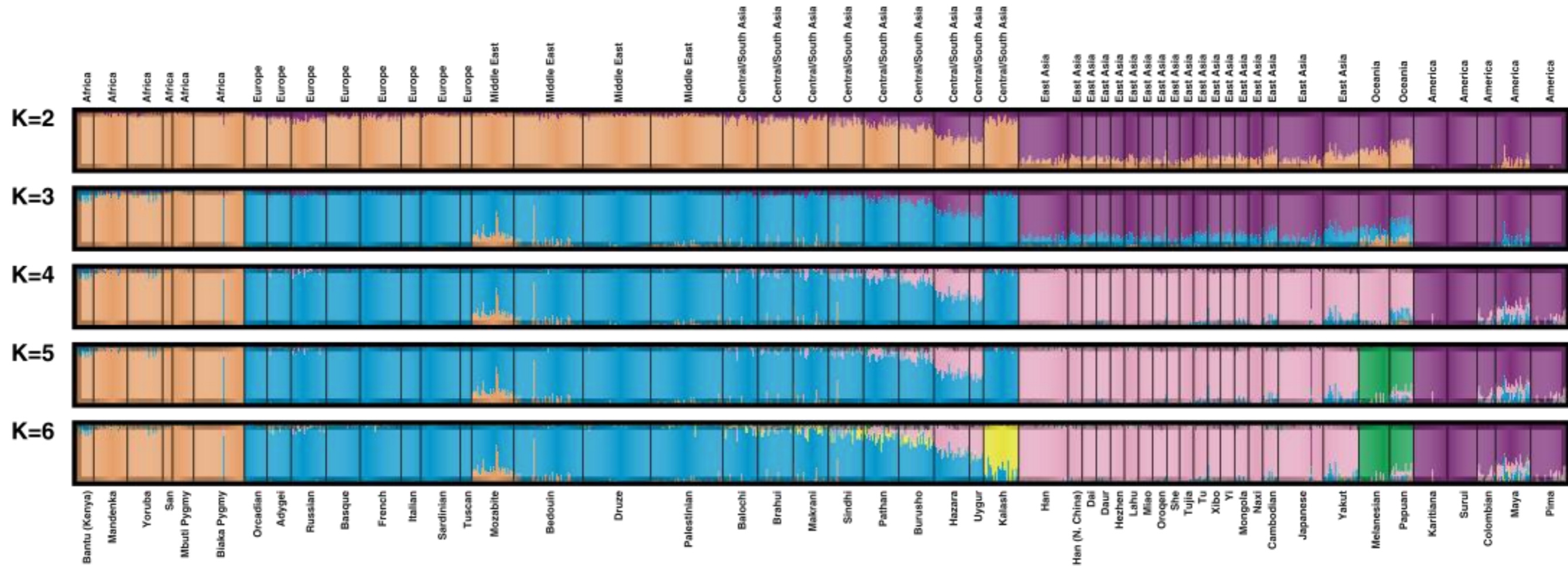


Fig. 1. Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters. Black lines separate individuals of different populations. Populations are labeled below the figure, with their regional affiliations above it. Ten *structure* runs at each

K produced nearly identical individual membership coefficients, having pairwise similarity coefficients above 0.97, with the exceptions of comparisons involving four runs at $K = 3$ that separated East Asia instead of Eurasia, and one run at $K = 6$ that separated Karitiana instead of Kalash. The figure shown for a given K is based on the highest probability run at that K .

aDNA – a new beginning?

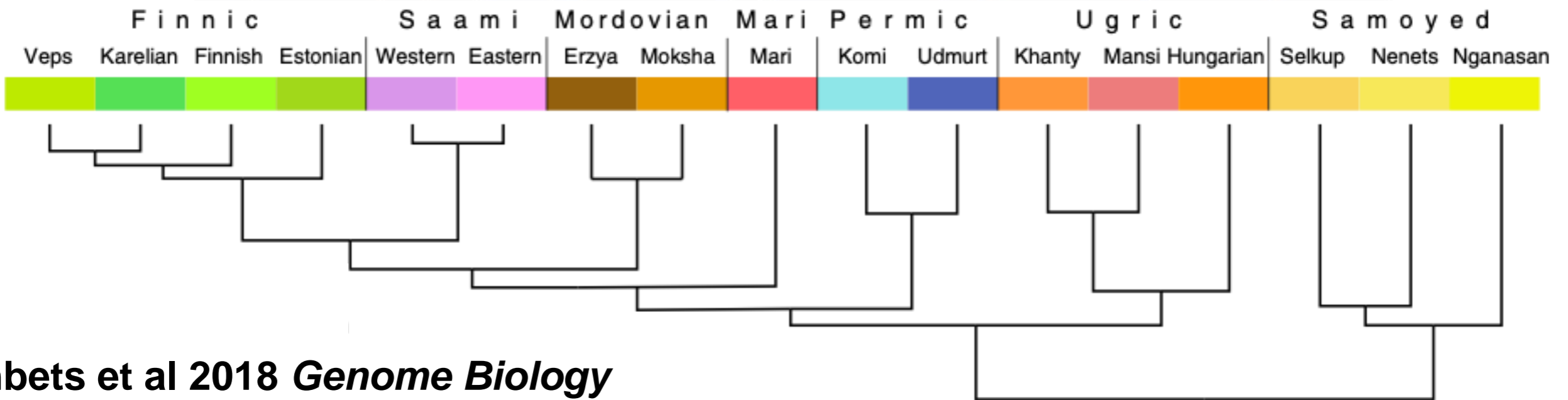
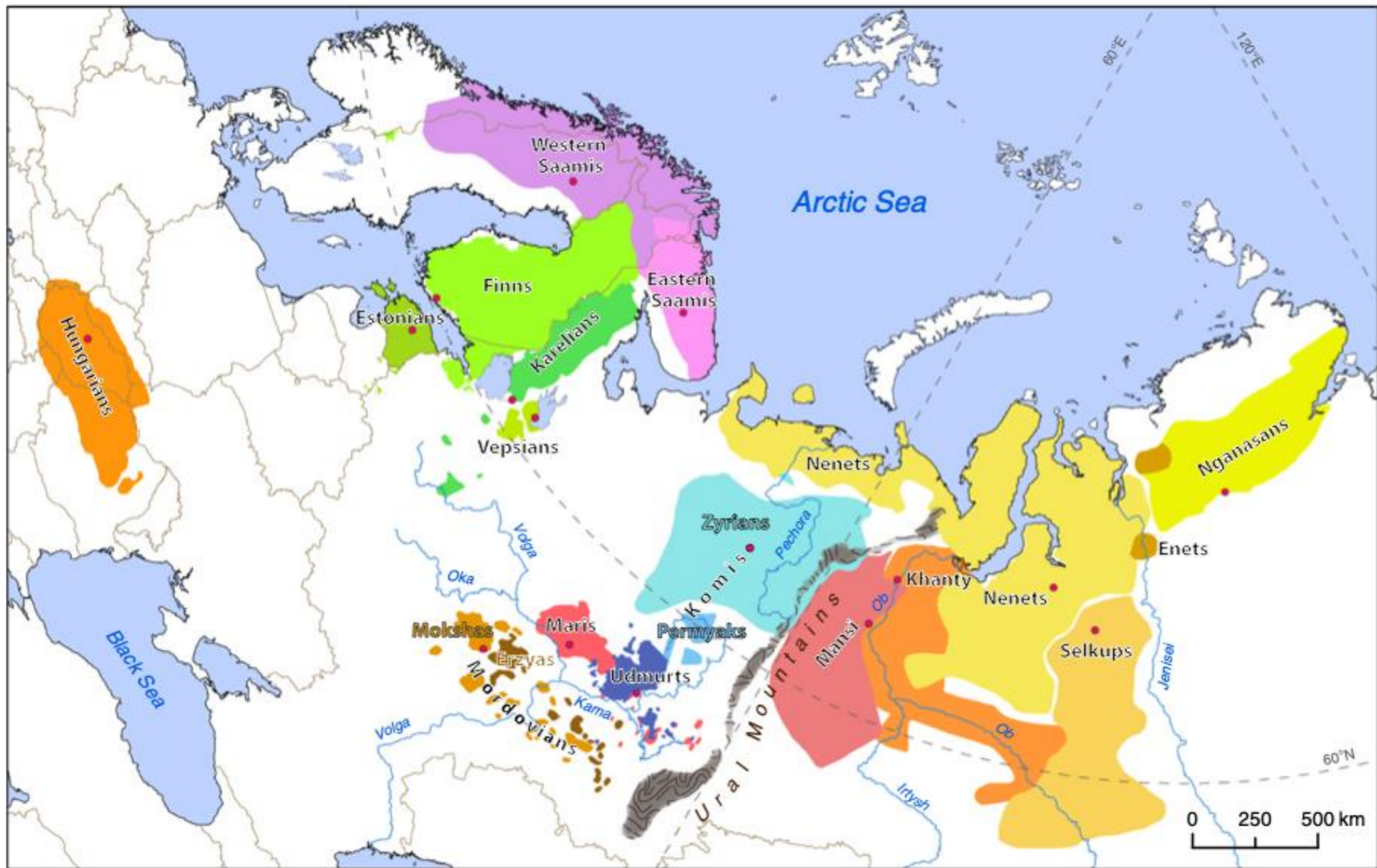


Photo by M. Wallerstedt

Uralic languages and genetics



Art by Kerttu Majander



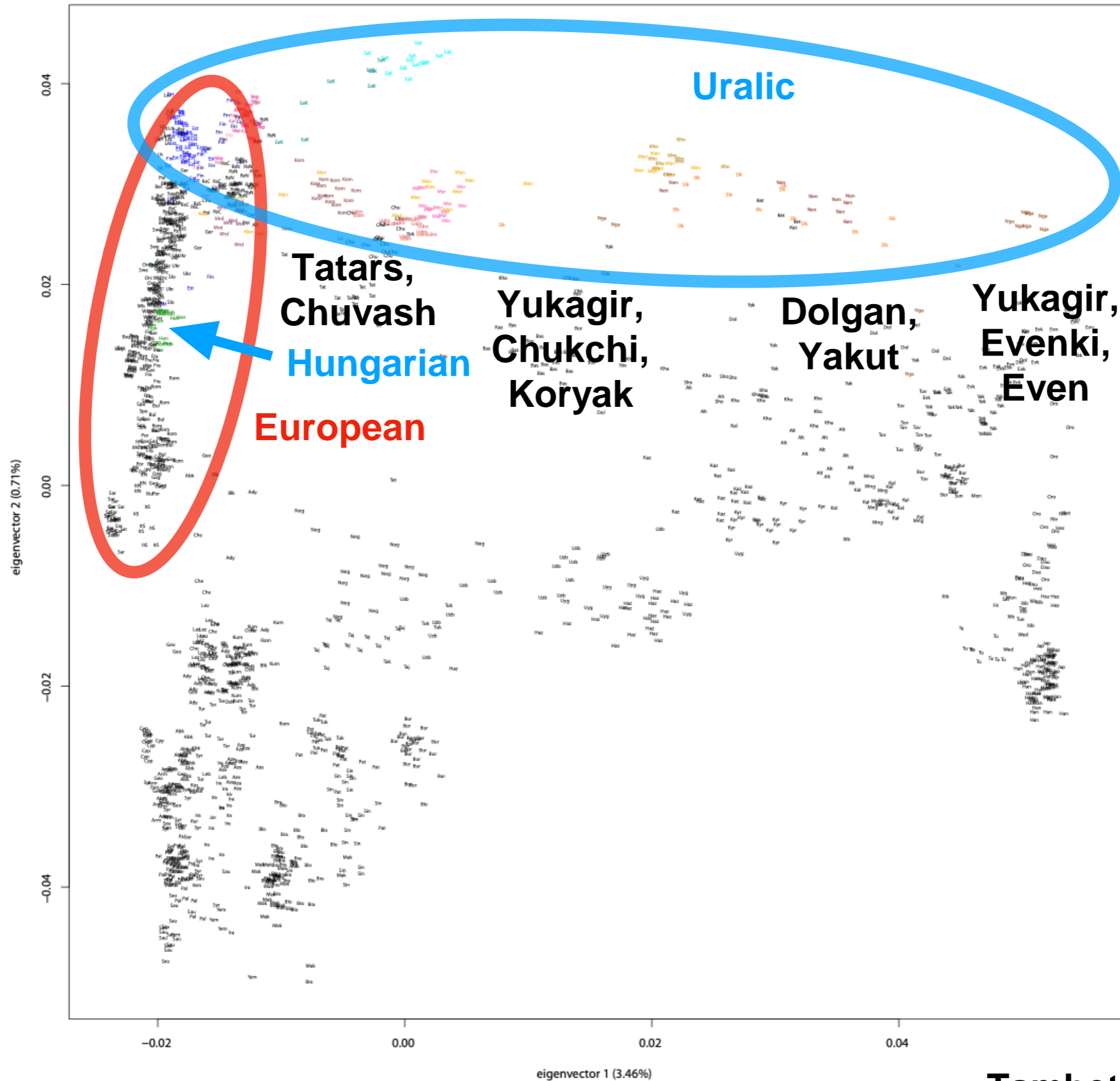
Tambets et al 2018 *Genome Biology*

Map by Timo Rantanen (BEDLAN: Cartographical database of Uralic languages)

Language data: Syrjänen et al. 2018 *lexibank/uralex: UraLex basic vocabulary dataset*

Autosomal variation in Eurasian context

1_2_PCA



Tambets et al 2018

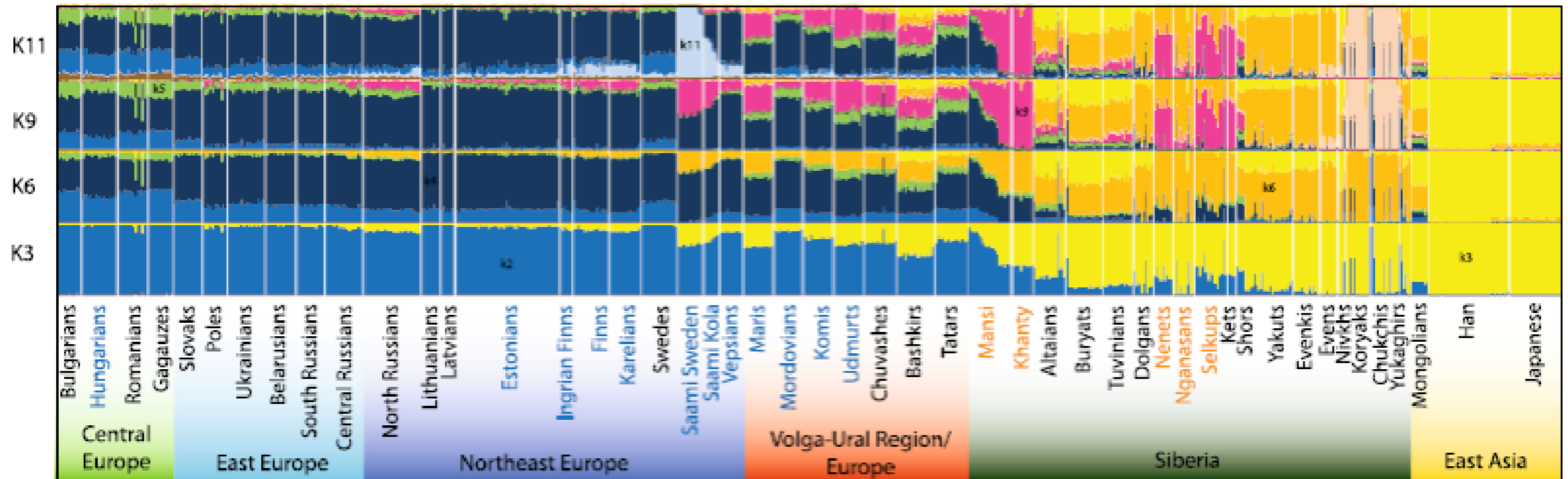
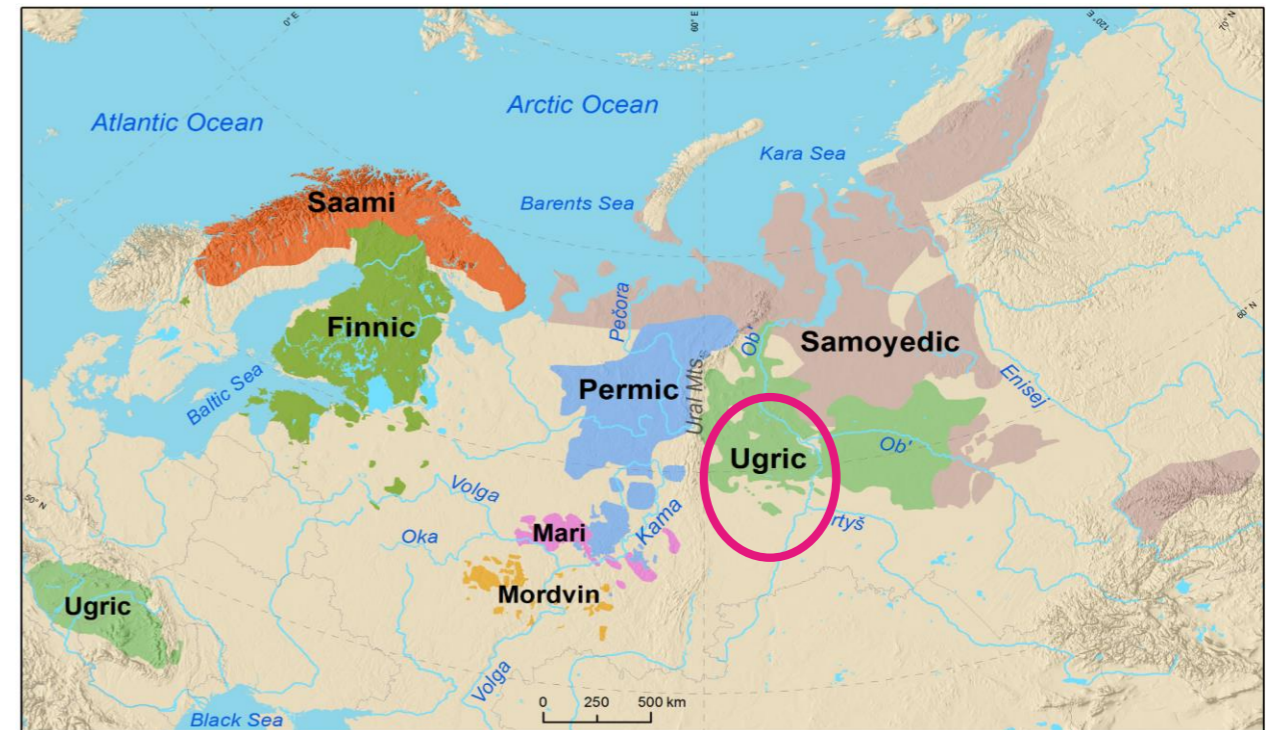
Genetic similarity between the Uralic speaker populations?



Art by Kerttu Majander

Central Siberian “origin” of the “Uralic specific genetics” (autosomes)

Maps: Cartographical Database of Uralic languages by BEDLAN (Timo Rantanen) and Jussi Ylikoski and authors of Oxford Guide for Uralic Languages (forthcoming).



Tambets et al. 2018

North-Siberian “origin” of the “Uralic specific genetics” (autosomal)

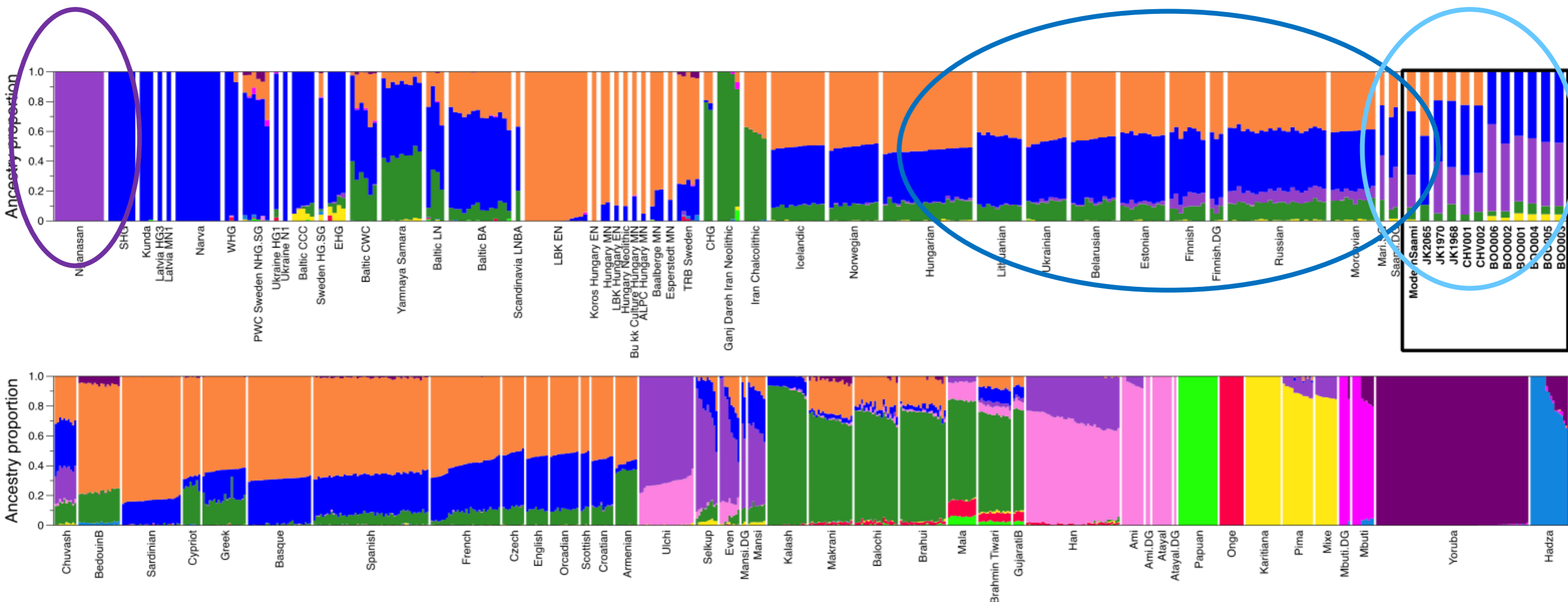
with aDNA populations from Finland (1700-1300 yrs old) and Kola Peninsula (3500 yrs old).



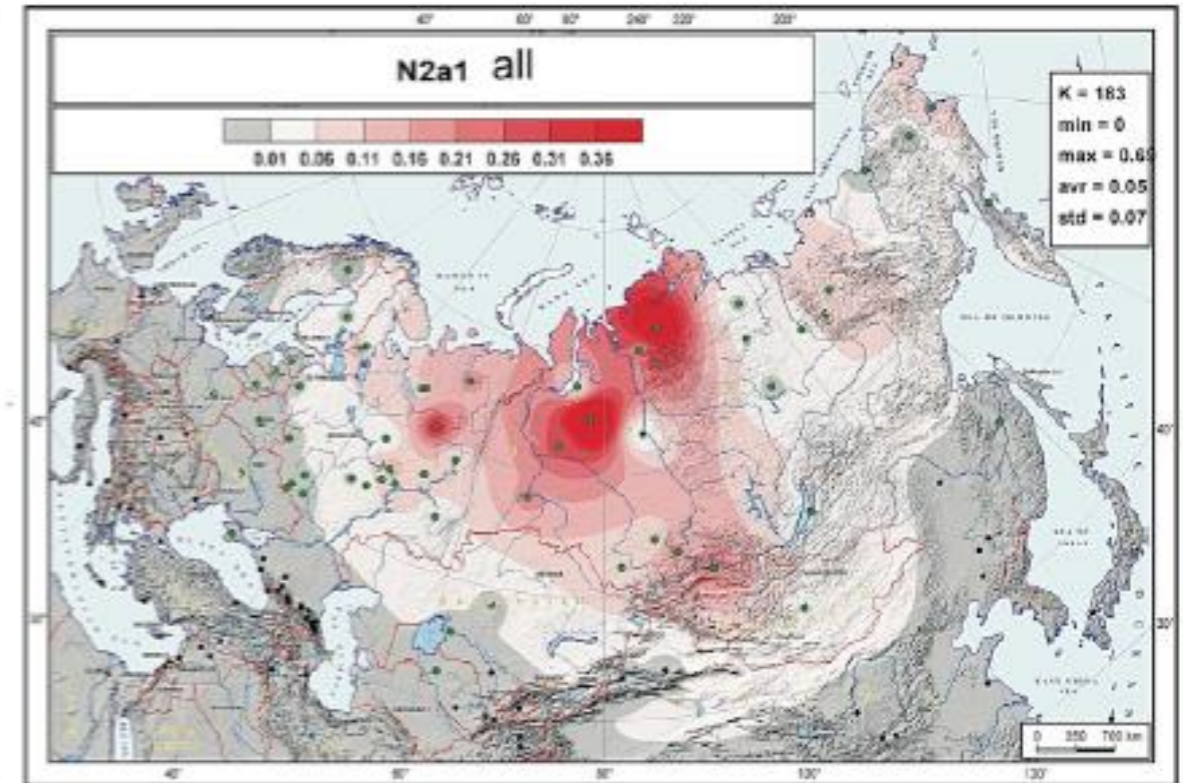
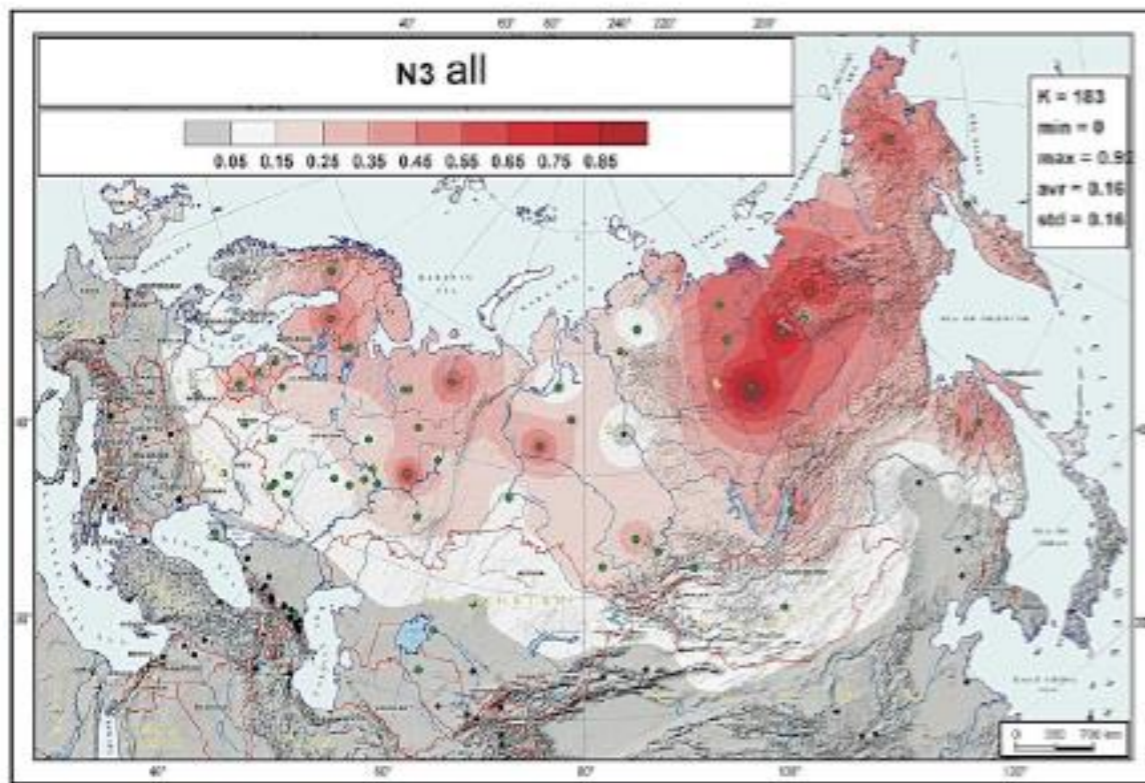
Nganasan

Uralic speaker populations

aDNA from Uralic area



Siberian “origin” of the N-haplogroup (Y-chromosome)



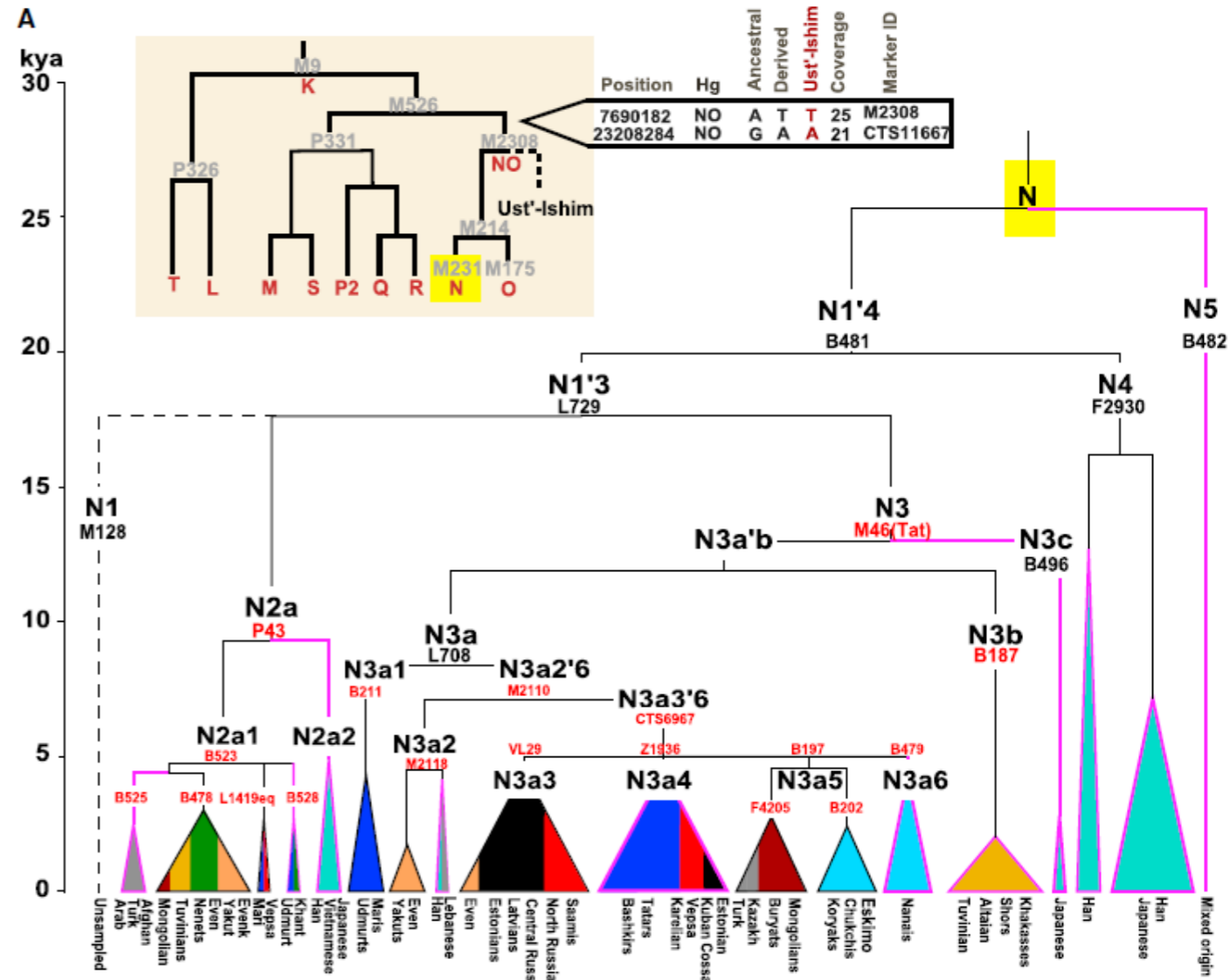
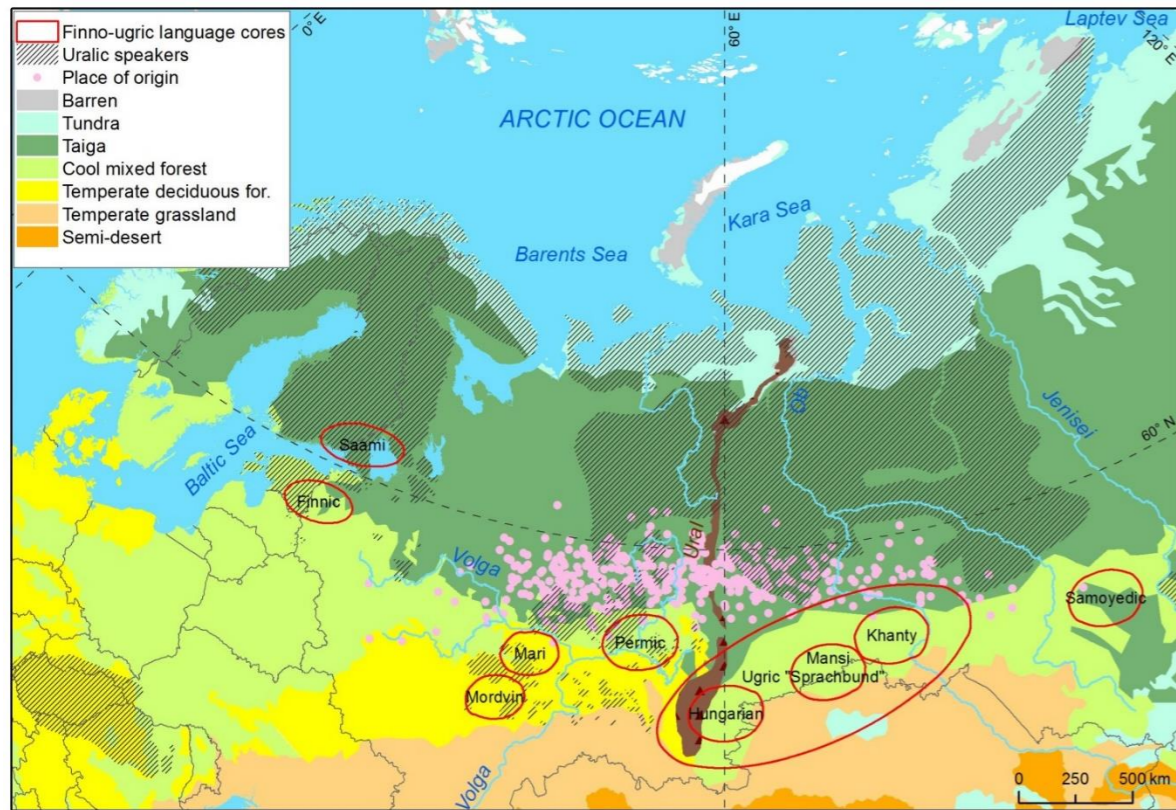
Gene-language co-dispersion?



Art by Kerttu Majander

Males with N haplogroup and Uralic languages moved westwards 5000 yrs ago?

Ilumäe et al. 2016

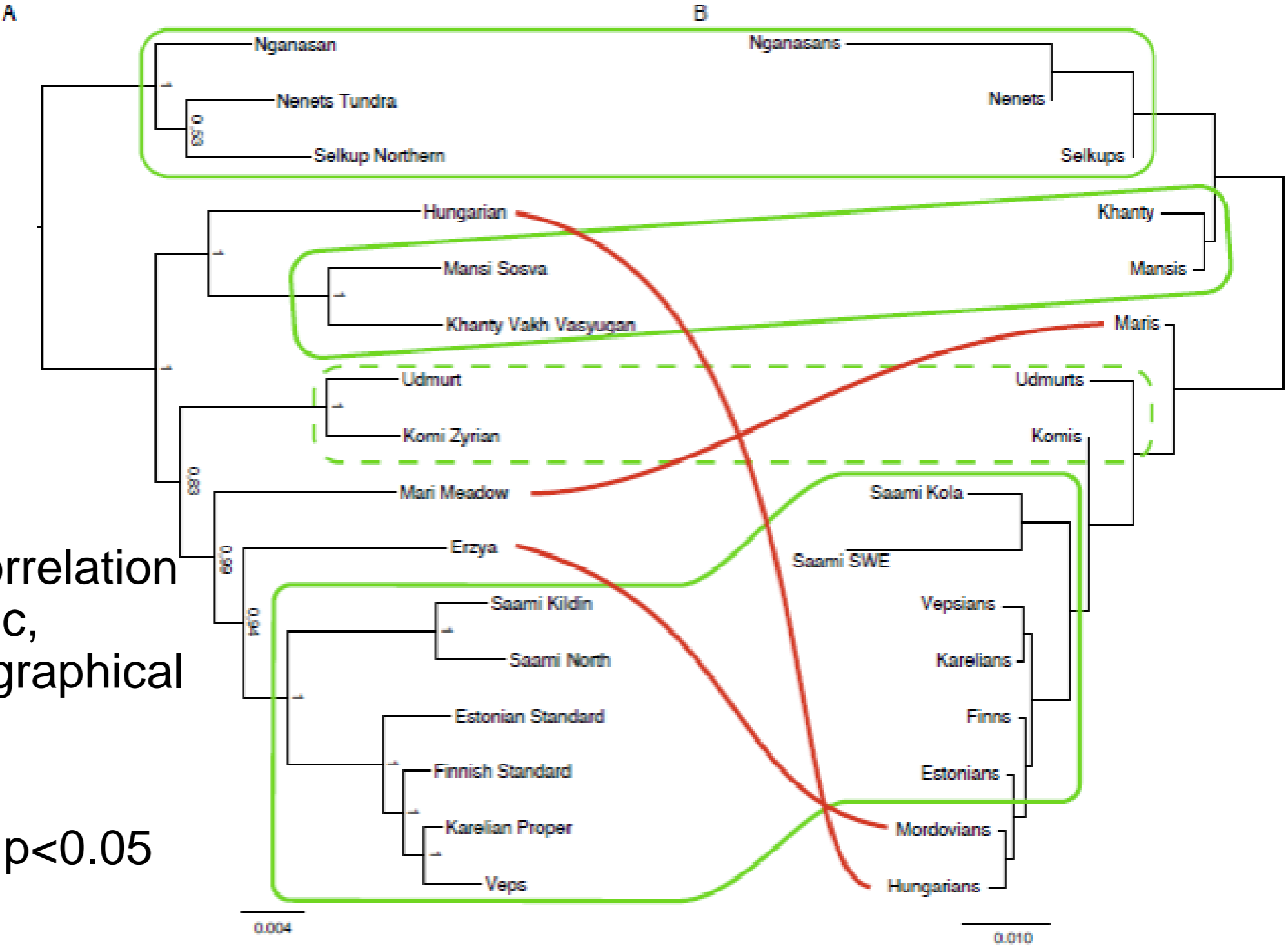


Historical linguistics:

Uralic family 6000 or 5000 or 4000 years old – Temporal match?

Our preliminary phylogeographical modelling of PU homeland (pink) ranges over an area covering all the homeland theories. Suggested areas for proto-stages (circles) were even more south (?) – Spatial mismatch?

Correlations between language trees/distances and genetic distances of speaker populations



Partial mantel correlation between linguistic, genetic and geographical distances:

With autosomes $p < 0.05$

Y and mtDNA *ns*

But when did Siberian influence appear? Case Estonia.



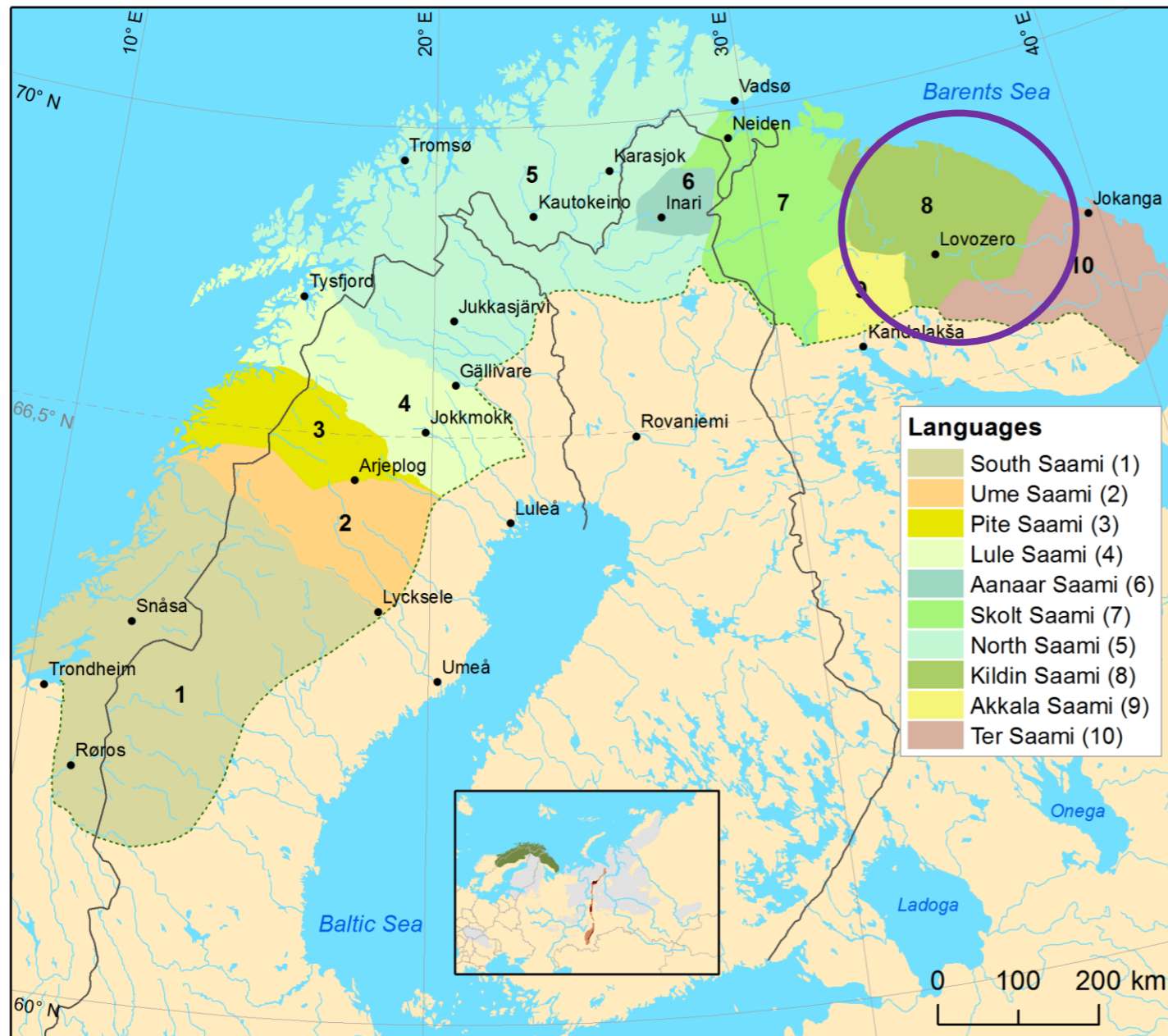
Saag et al. 2019:

N haplogroup appears to Estonia in the beginning of Iron Age; some 2500 YBP

Historical linguistics:
Arrival of Proto-Finnic to Baltic Sea coast (St Petersburg area?) from Volga-Oka tributary some 2500 YBP

Match

But when did Siberian influence appear? Case Lapland.



Lamnidis et al. 2018:

N subgroup and “North Siberian” autosomal influence in **Kola Peninsula** already 3500 yrs ago.

But when did Uralic languages appear? Case Lapland.

Aikio 2012:

Linguistic landscape some 2500 YBP.
Saami languages reached current
speaker area only 1800-1200 YBP.

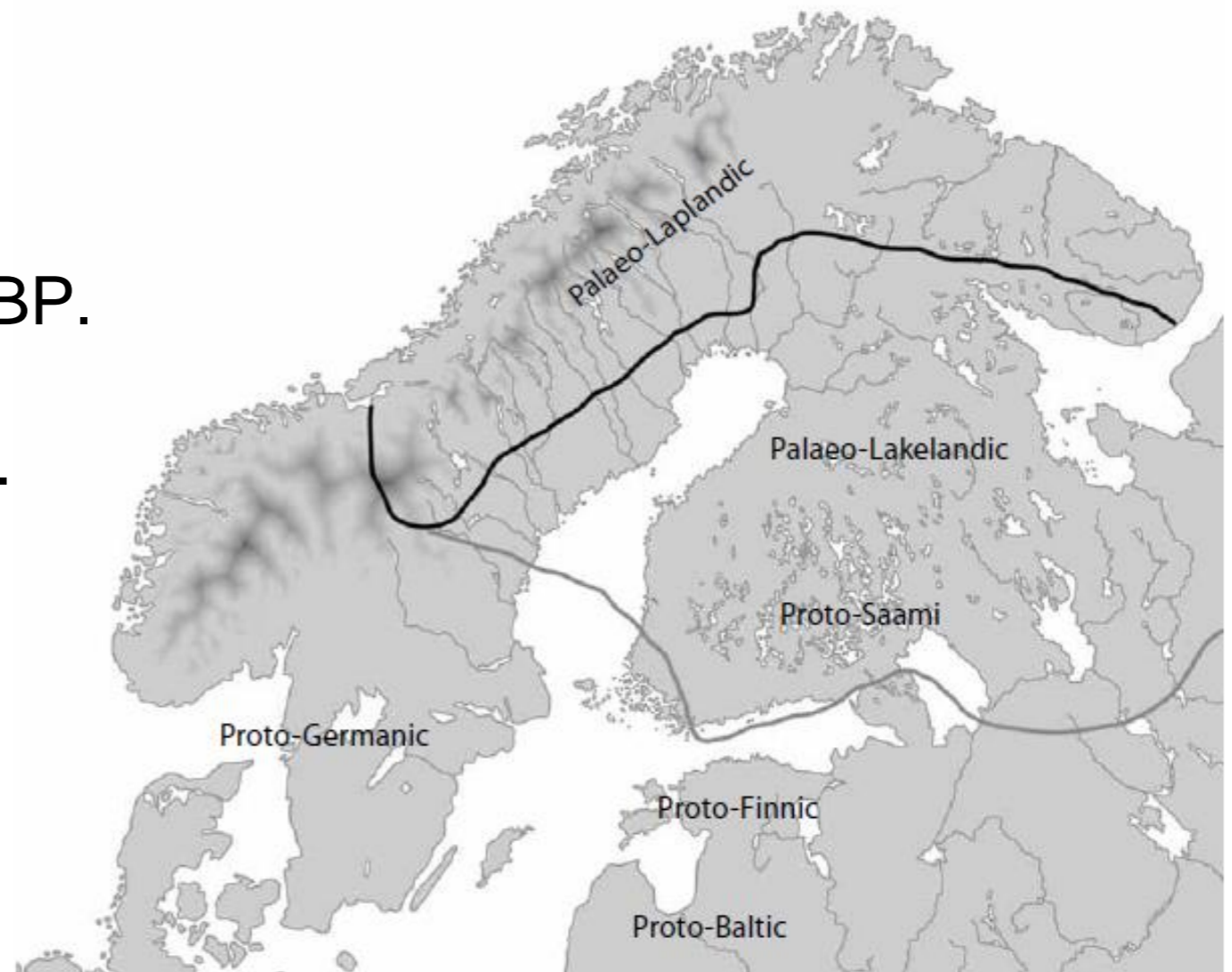
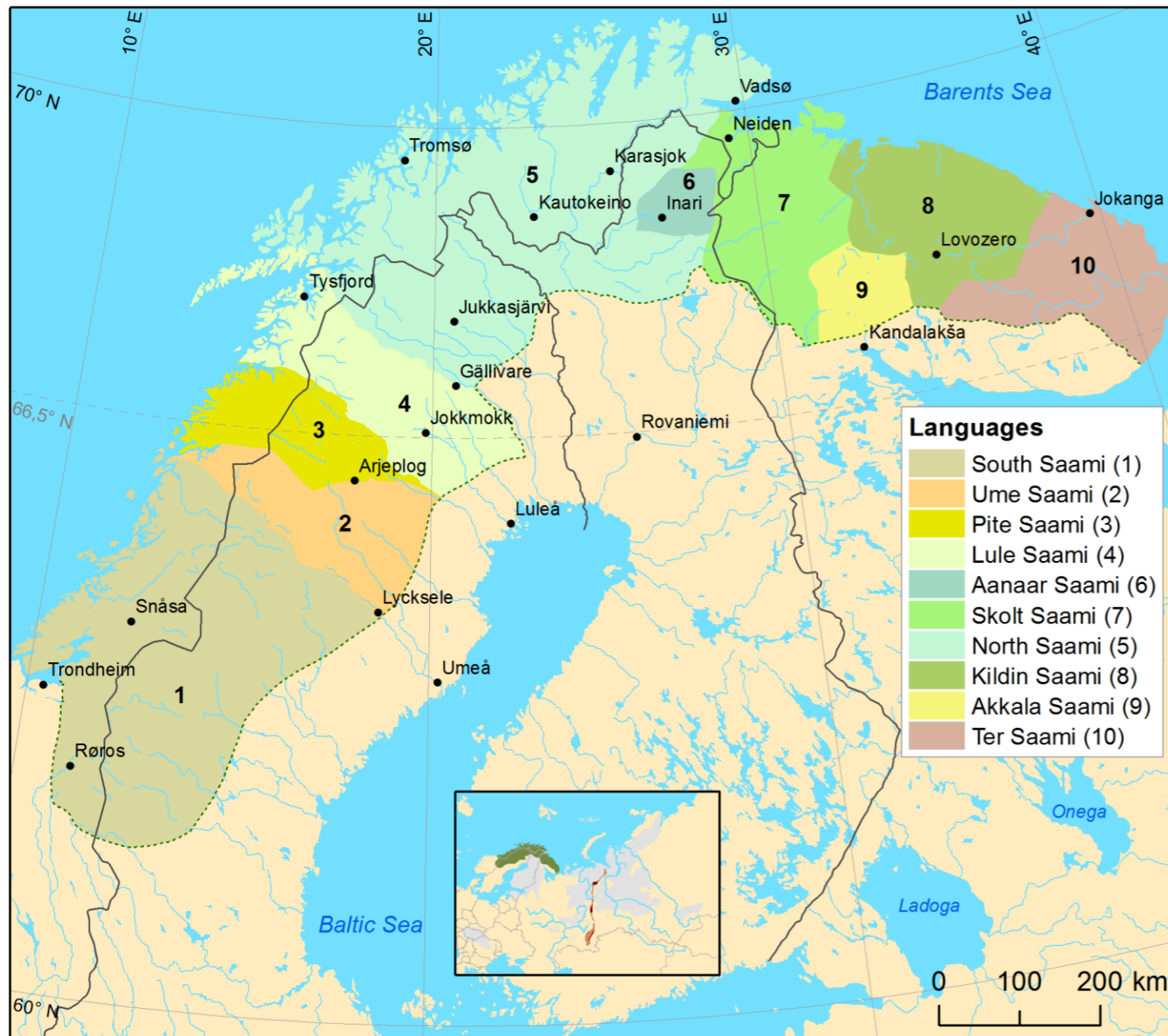


Figure 1. The linguistic situation in Lapland and the northern Baltic Sea Area in the Early Iron Age prior to the expansion of Saami languages; the locations of the language groups are schematic. The black line indicates the distribution of Saami languages in the 19th century, and the gray line their approximate maximal distribution before the expansion of Finnic.

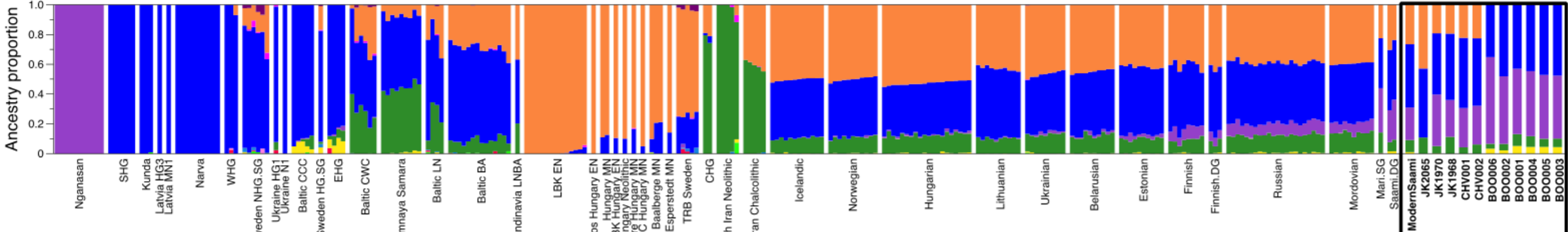
But when did Siberian influence appear? Case Lapland.



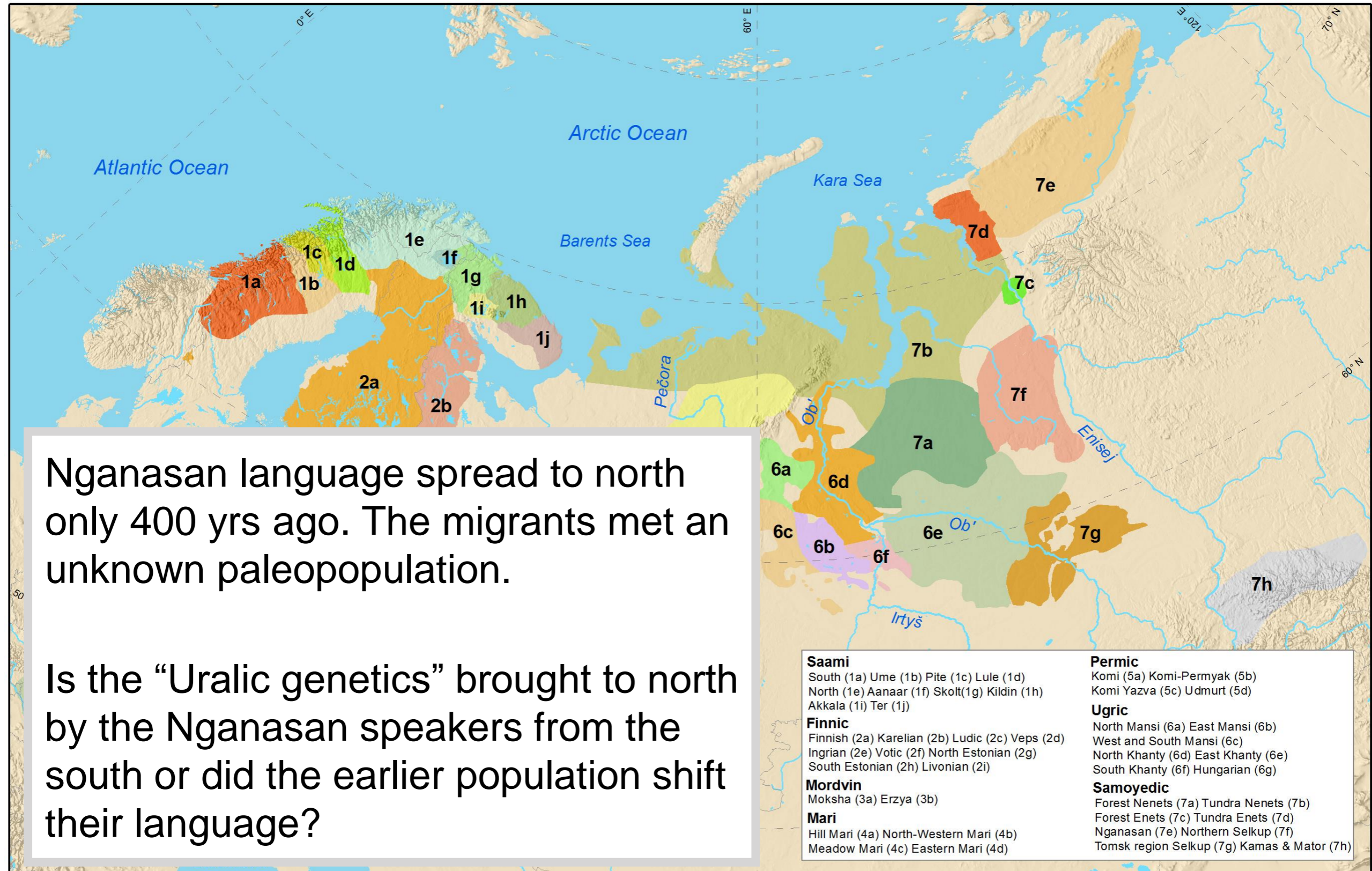
Mismatch:

“Uralic” genetic landscape
earlier

But are when did Siberian influence appear? Case Nganasans.



But are when did Siberian influence appear? Case Nganasans.

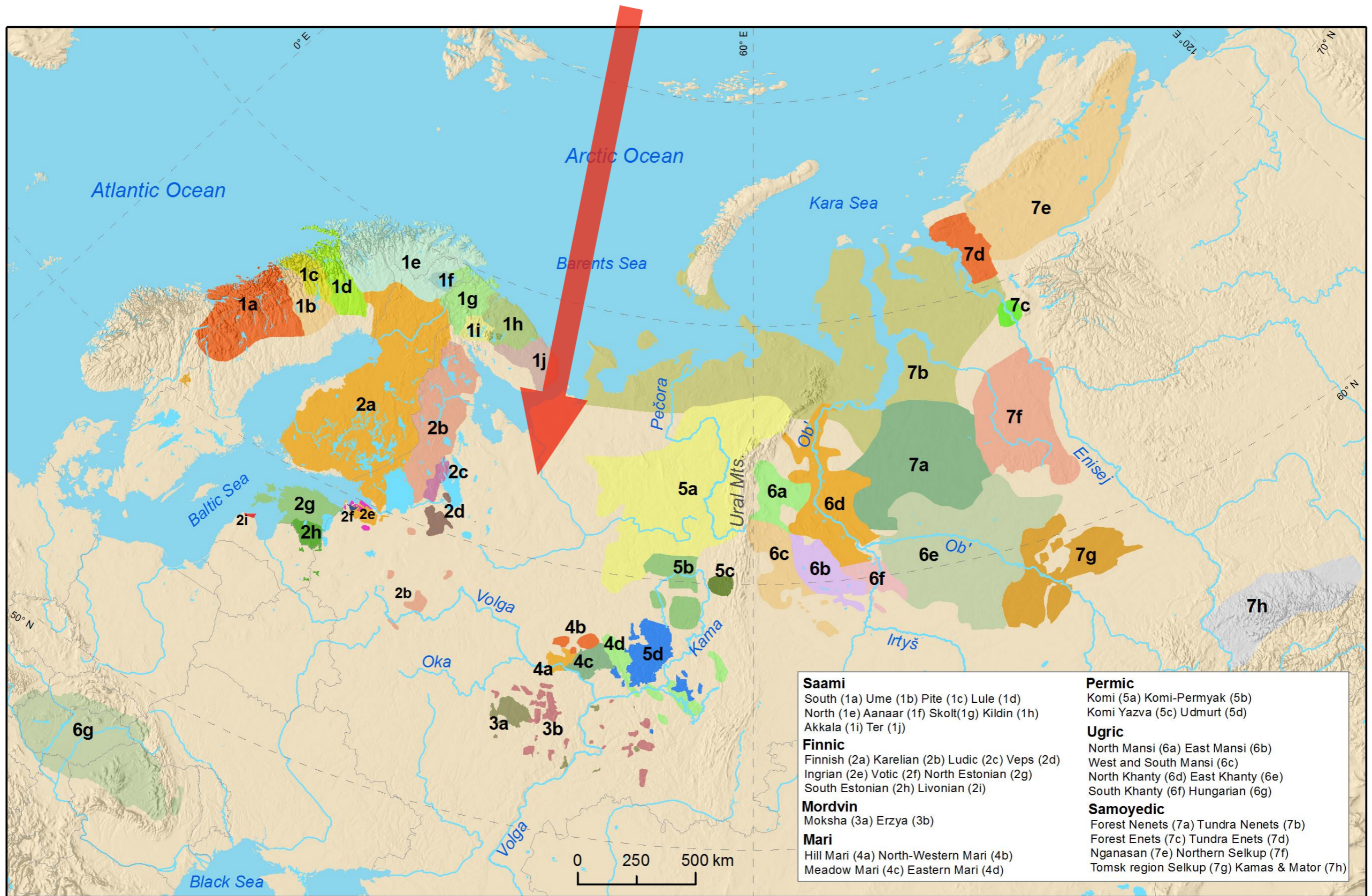


Nganasan language spread to north only 400 yrs ago. The migrants met an unknown paleopopulation.

Is the “Uralic genetics” brought to north by the Nganasan speakers from the south or did the earlier population shift their language?

- | | |
|---|---|
| <p>Saami
 South (1a) Ume (1b) Pite (1c) Lule (1d)
 North (1e) Aanaar (1f) Skolt(1g) Kildin (1h)
 Akkala (1i) Ter (1j)</p> <p>Finnic
 Finnish (2a) Karelian (2b) Ludic (2c) Veps (2d)
 Ingrian (2e) Votic (2f) North Estonian (2g)
 South Estonian (2h) Livonian (2i)</p> <p>Mordvin
 Moksha (3a) Erzya (3b)</p> <p>Mari
 Hill Mari (4a) North-Western Mari (4b)
 Meadow Mari (4c) Eastern Mari (4d)</p> | <p>Permian
 Komi (5a) Komi-Permyak (5b)
 Komi Yazva (5c) Udmurt (5d)</p> <p>Ugric
 North Mansi (6a) East Mansi (6b)
 West and South Mansi (6c)
 North Khanty (6d) East Khanty (6e)
 South Khanty (6f) Hungarian (6g)</p> <p>Samoyedic
 Forest Nenets (7a) Tundra Nenets (7b)
 Forest Enets (7c) Tundra Enets (7d)
 Nganasan (7e) Northern Selkup (7f)
 Tomsk region Selkup (7g) Kamas & Mator (7h)</p> |
|---|---|

A known language shift: the North Russian case



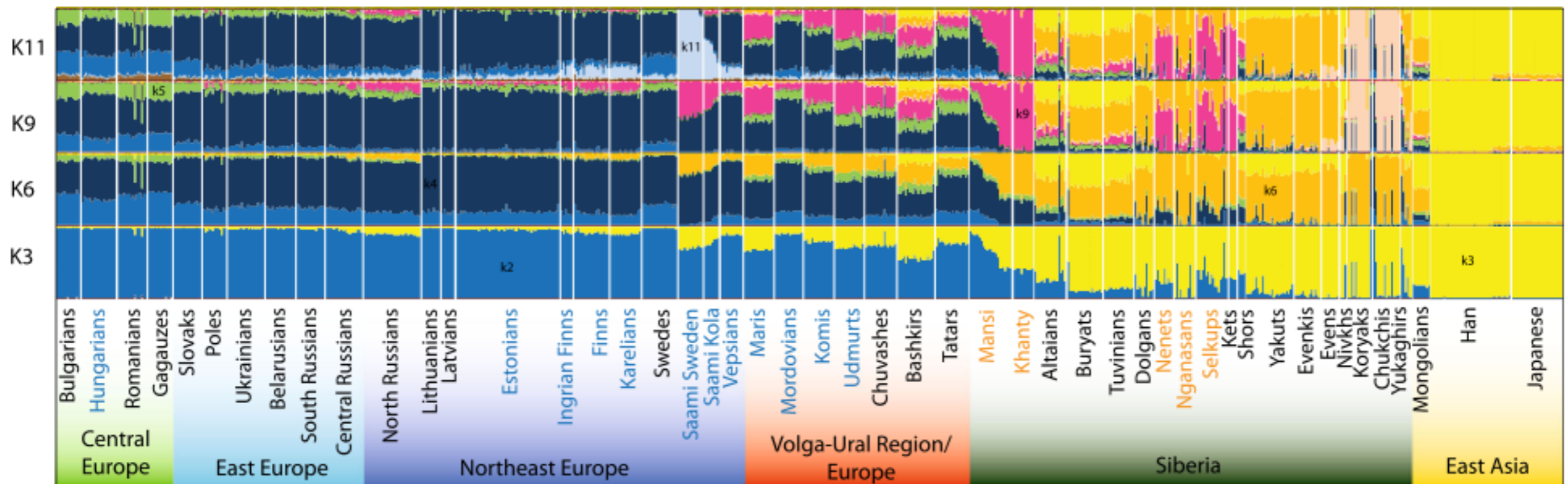
Language shift: the North Russian gap

Hypothesis 1

Slavic speakers immigrated, acquired place names and language substrate from the local population

Hypothesis 2

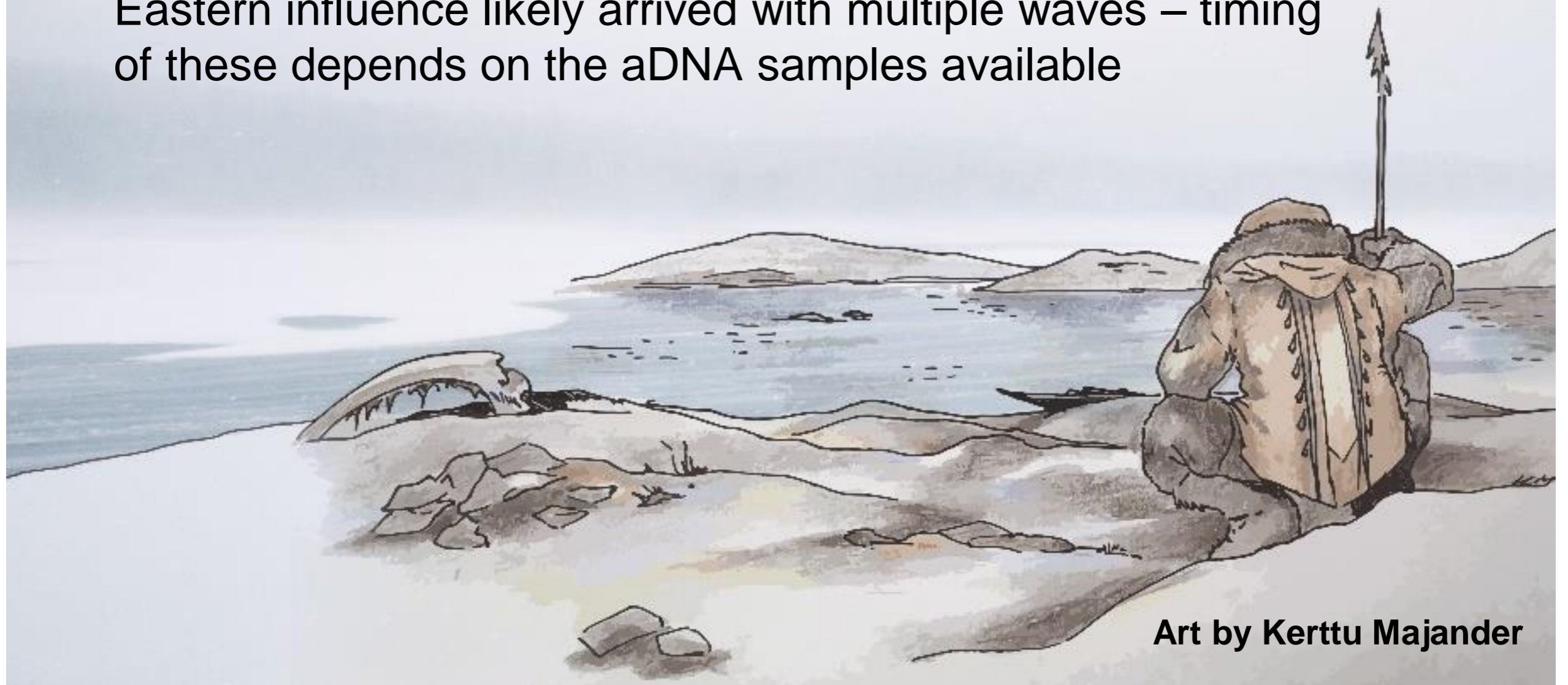
Uralic speakers shift language in contact with Slavic speakers



Sum up:

In the Uralic area there is partial match of gene-language dispersion (Estonia and N), but in the north the genetic landscape is likely older than the linguistic one

Eastern influence likely arrived with multiple waves – timing of these depends on the aDNA samples available



Art by Kerttu Majander

Kiitos



Thanks to the UraHoli team / Salmela et al. (in prep.) review of Uralic past from linguistic, genetic, archaeological and ecological perspectives