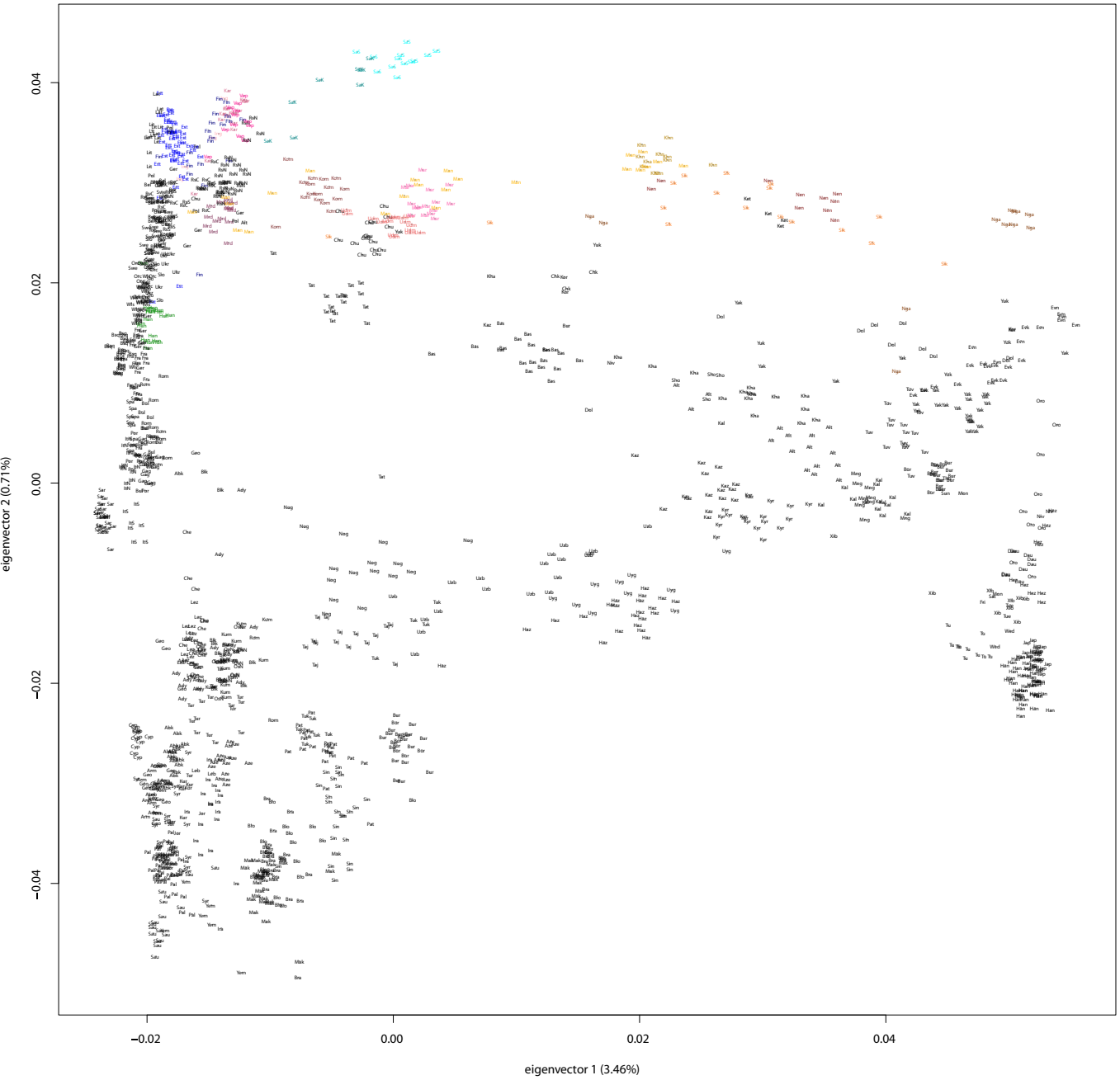


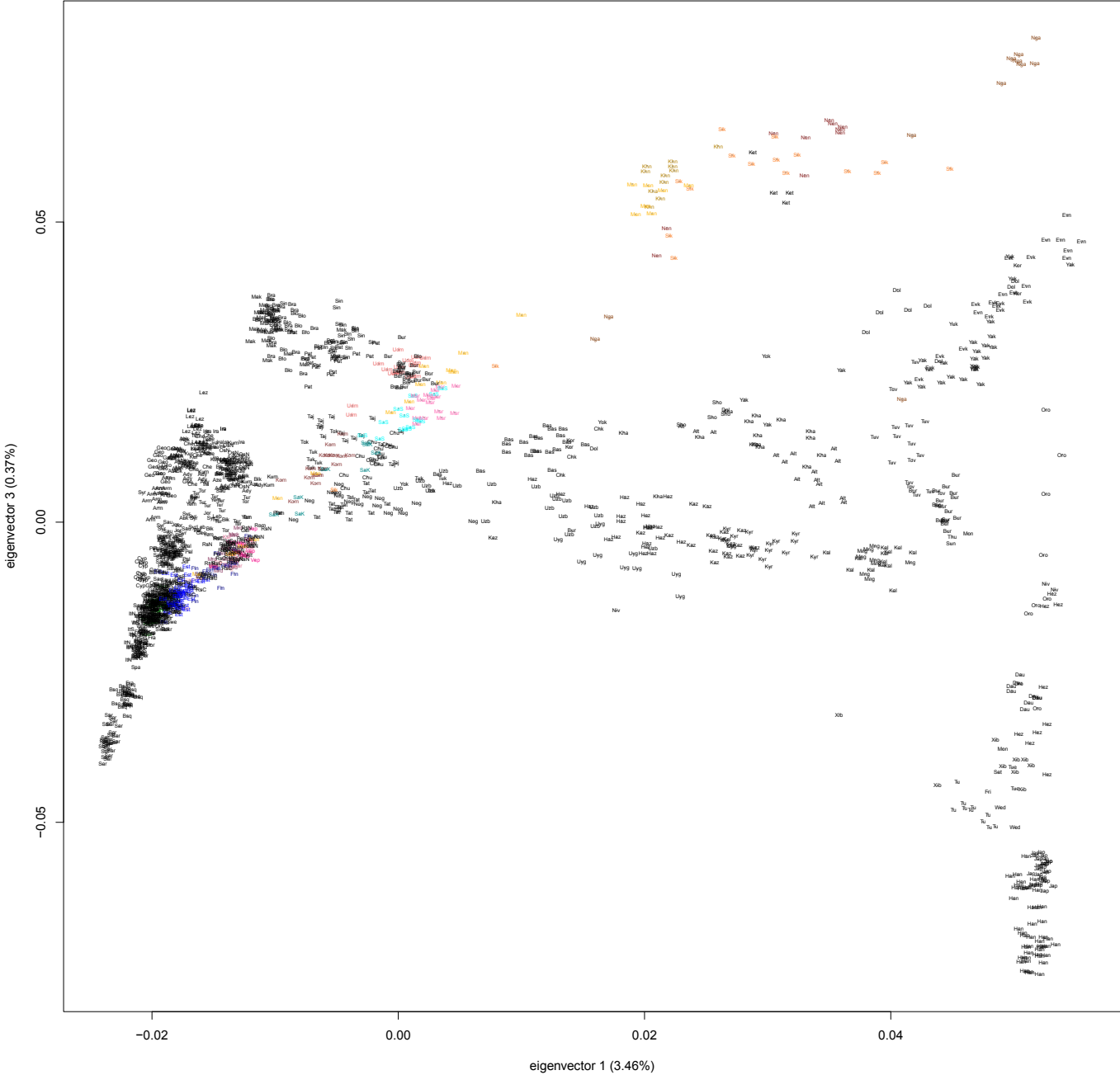
A.

1_2_PCA



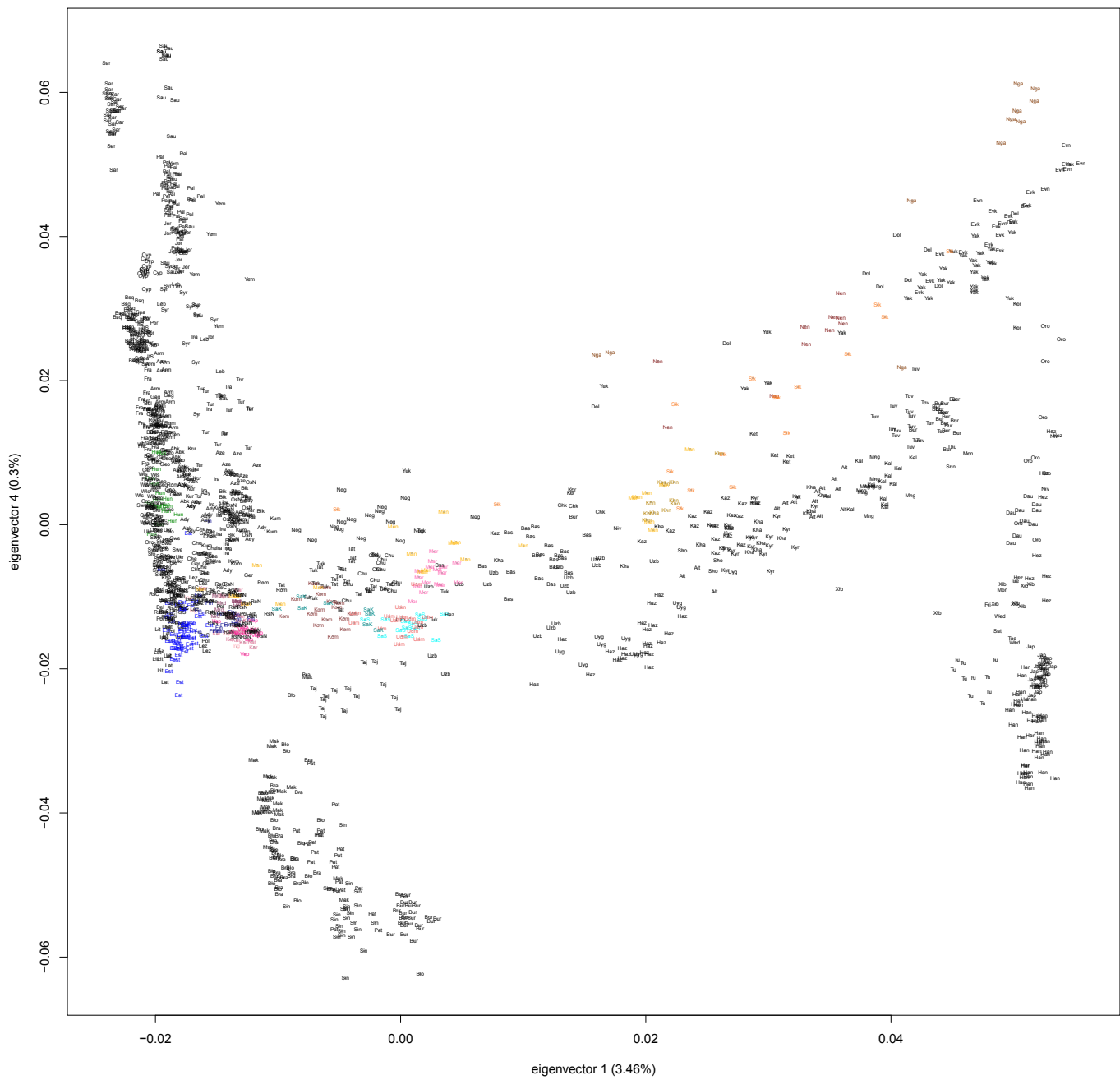
B.

1_3_PCA



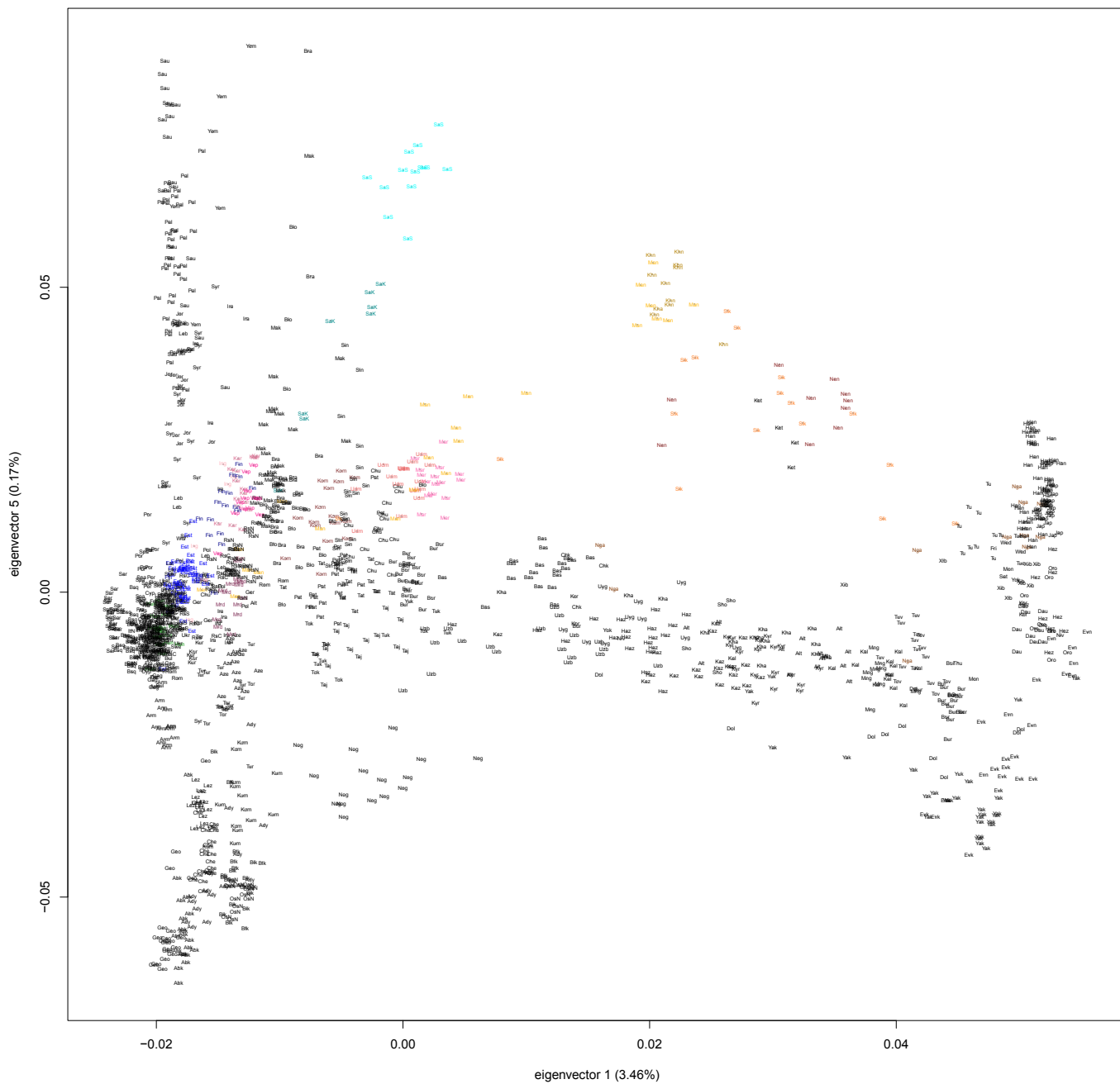
C.

1_4_PCA



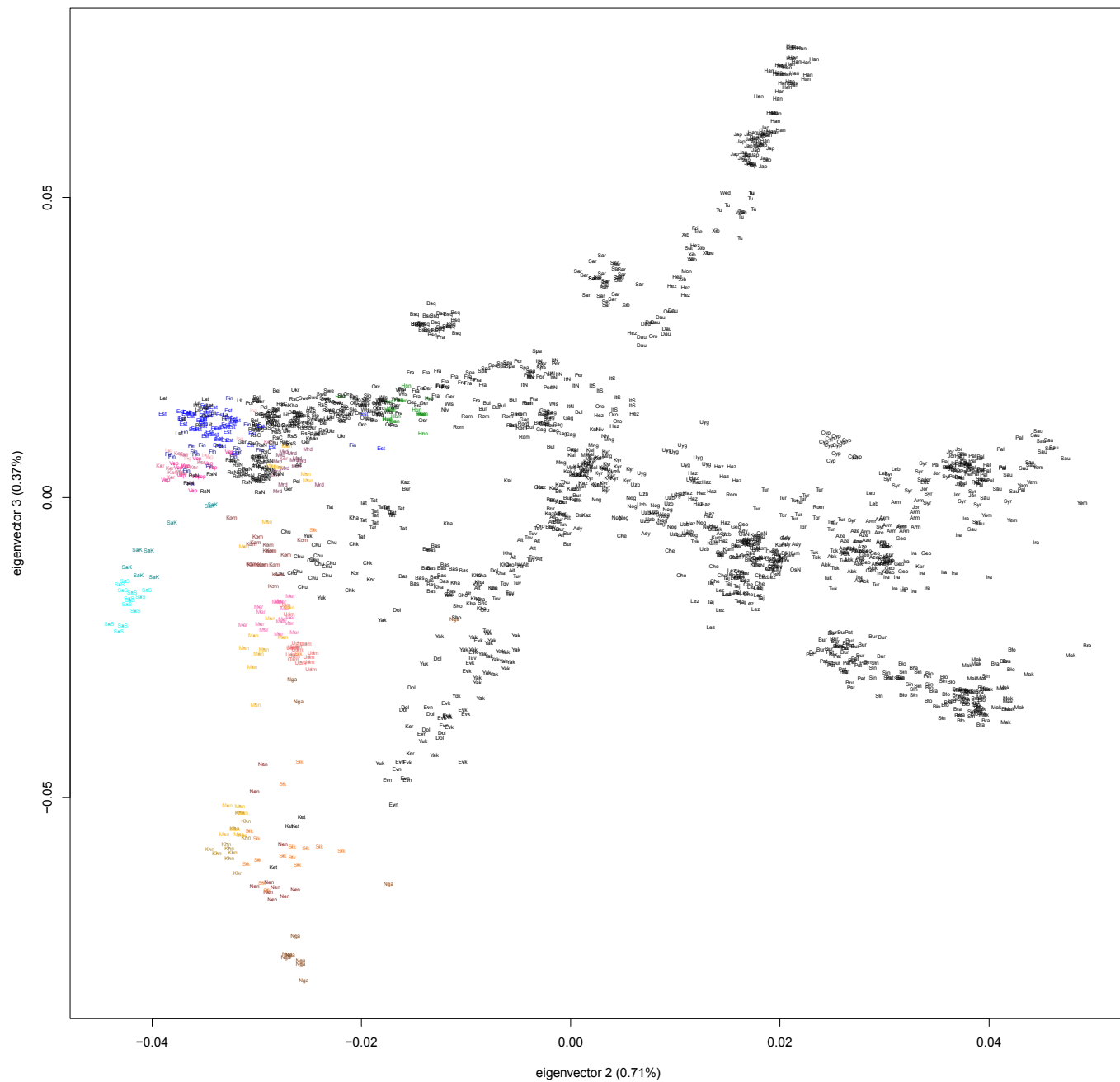
D.

1_5_PCA



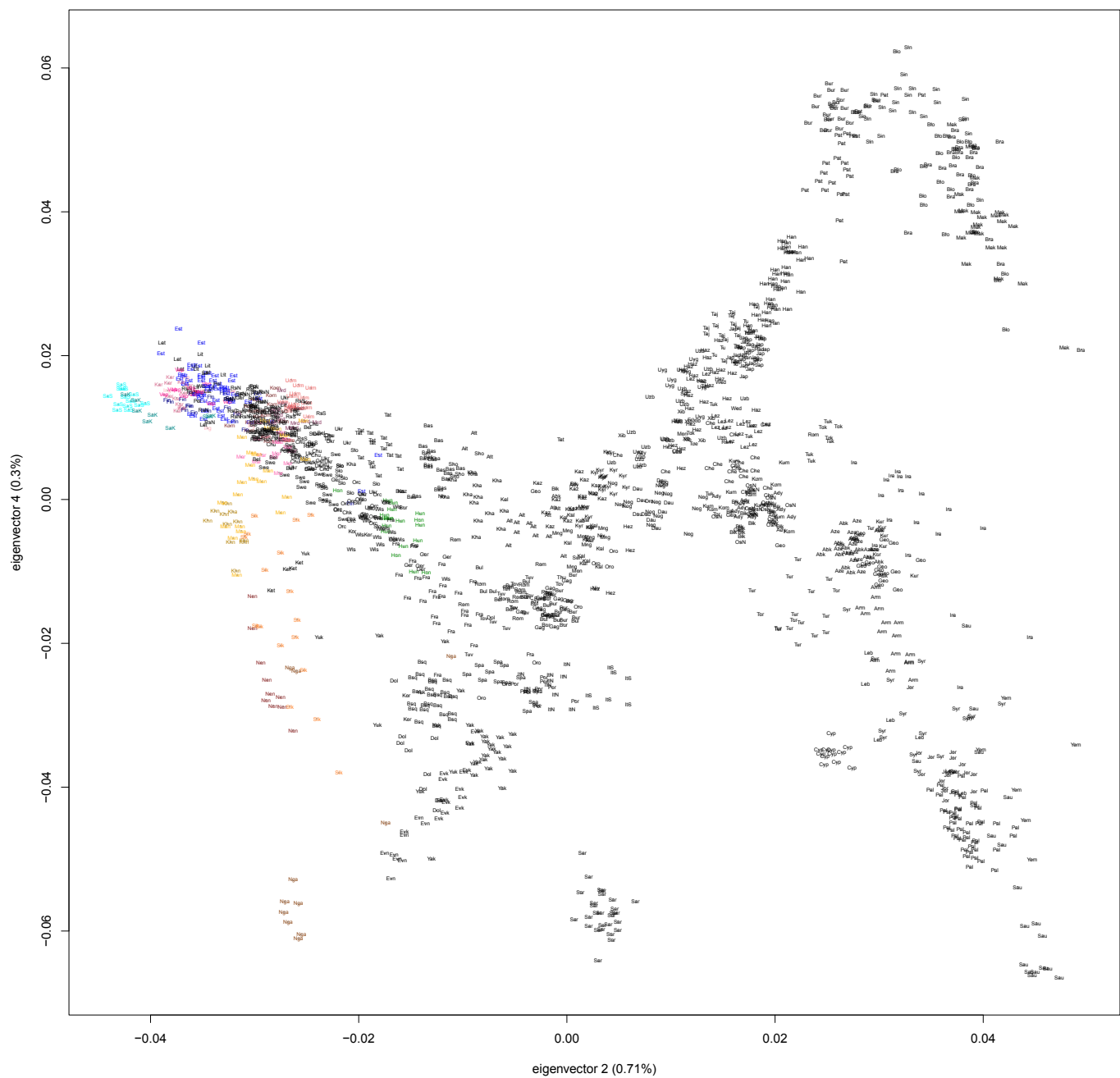
E.

2_3_PCA



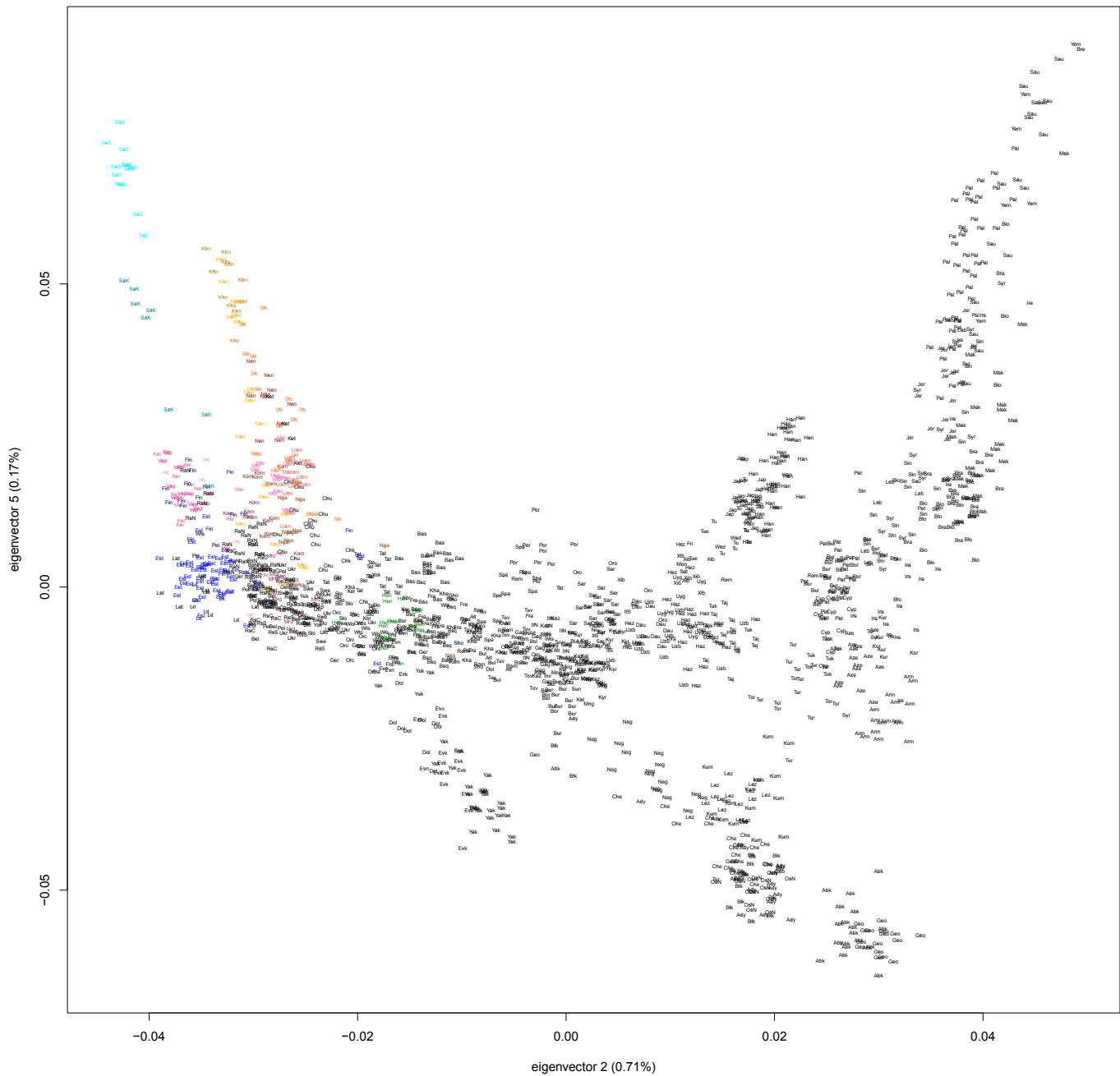
F.

2_4_PCA



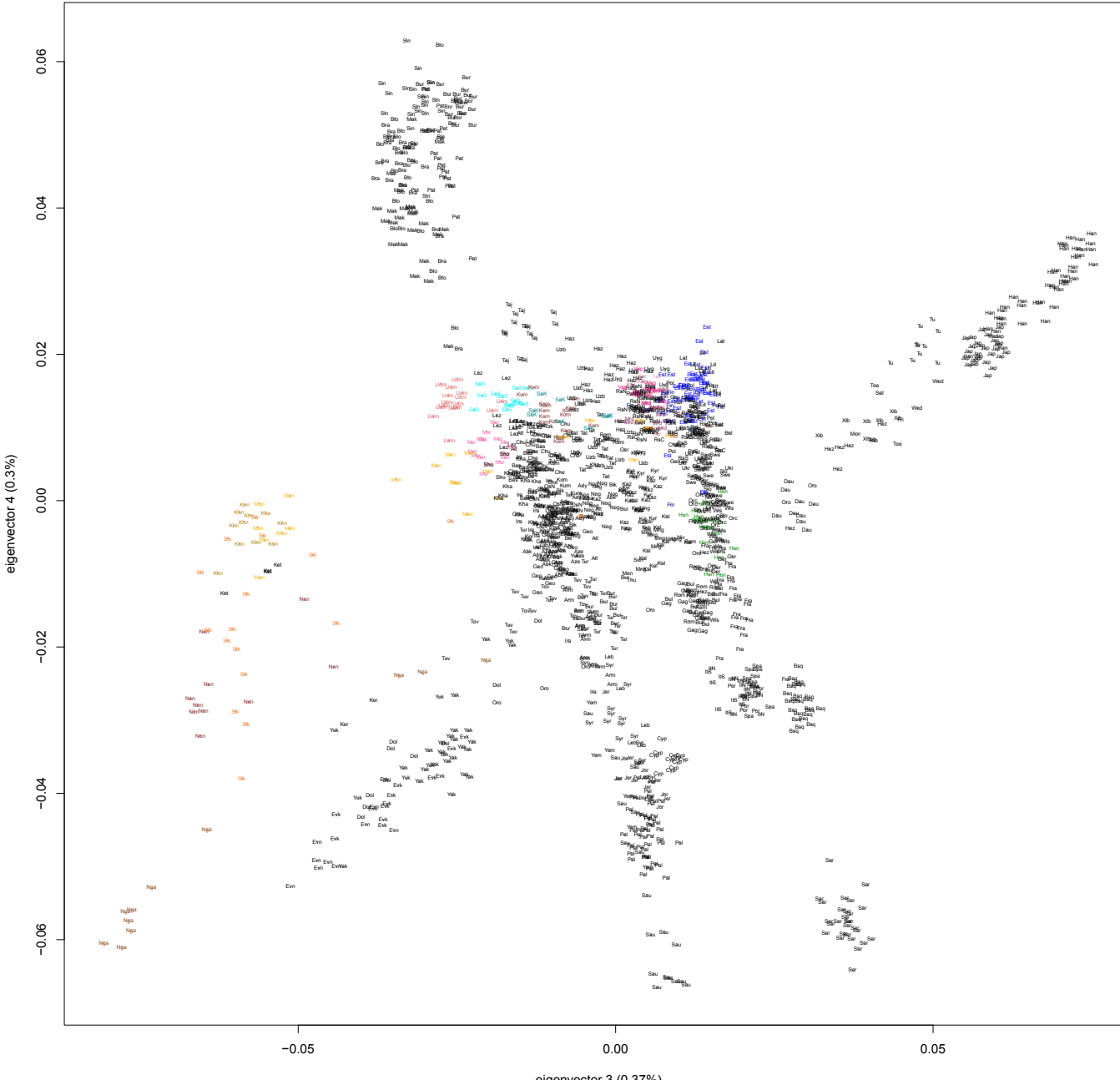
G.

2_5_PCA



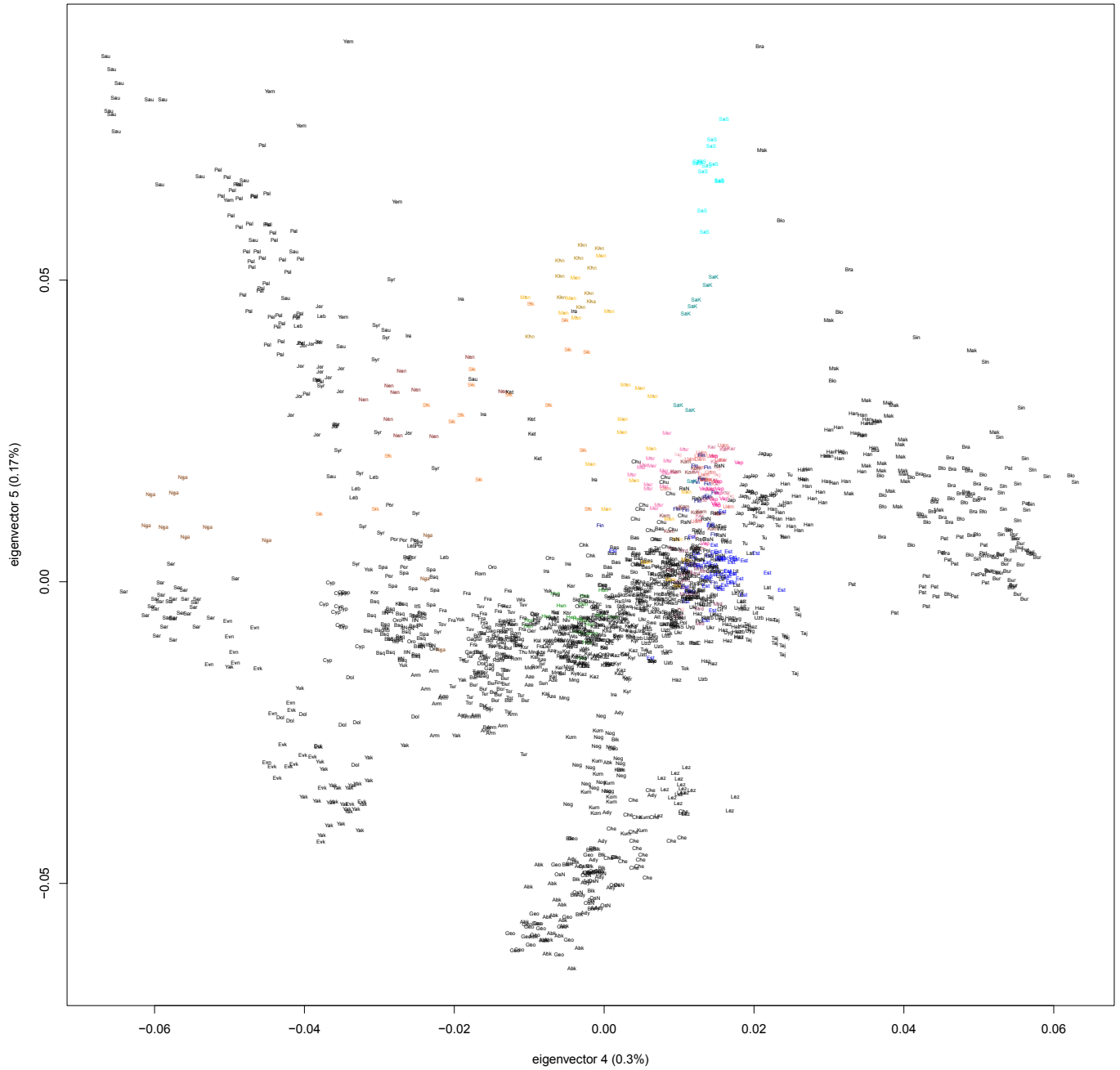
H.

3_4_PCA



I.

4_5_PCA



J.

3_5_PCA

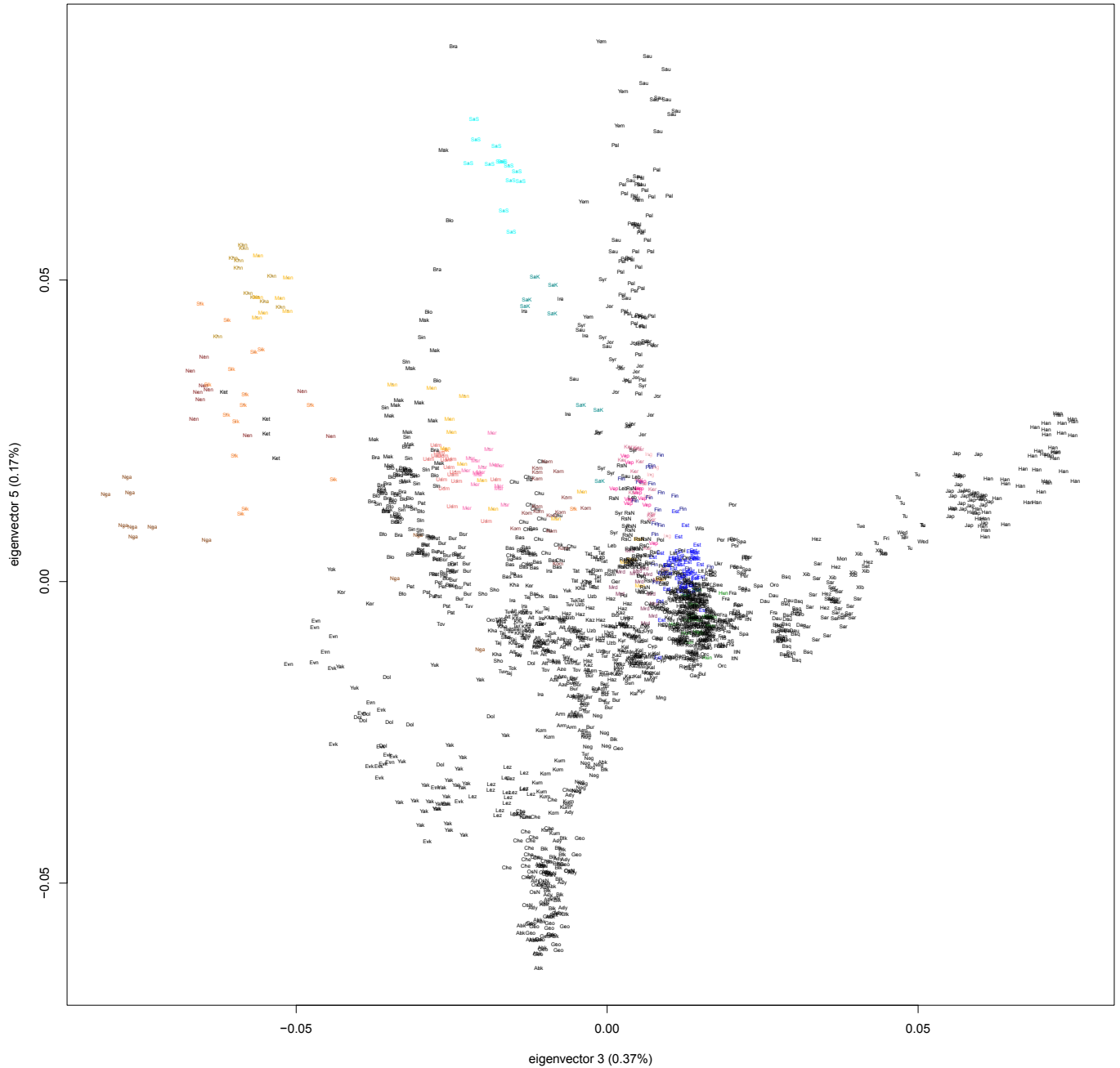


Figure S1. Principal component (PC) analysis of the studied Uralic-speaking populations (highlighted) in Eurasian context (see data and abbreviations of used populations from Additional file 1: Table S1). For each eigenvector the fraction of the variance (%) is specified; **A** – PC1 versus PC2; **B** – PC1 versus PC3; **C** – PC1 versus PC4; **D** – PC1 versus PC5; **E** – PC2 versus PC3; **F** – PC2 versus PC4; **G** – PC2 versus PC5; **H** – PC3 versus PC4; **I** – PC4 versus PC5; **J** – PC3 versus PC5.

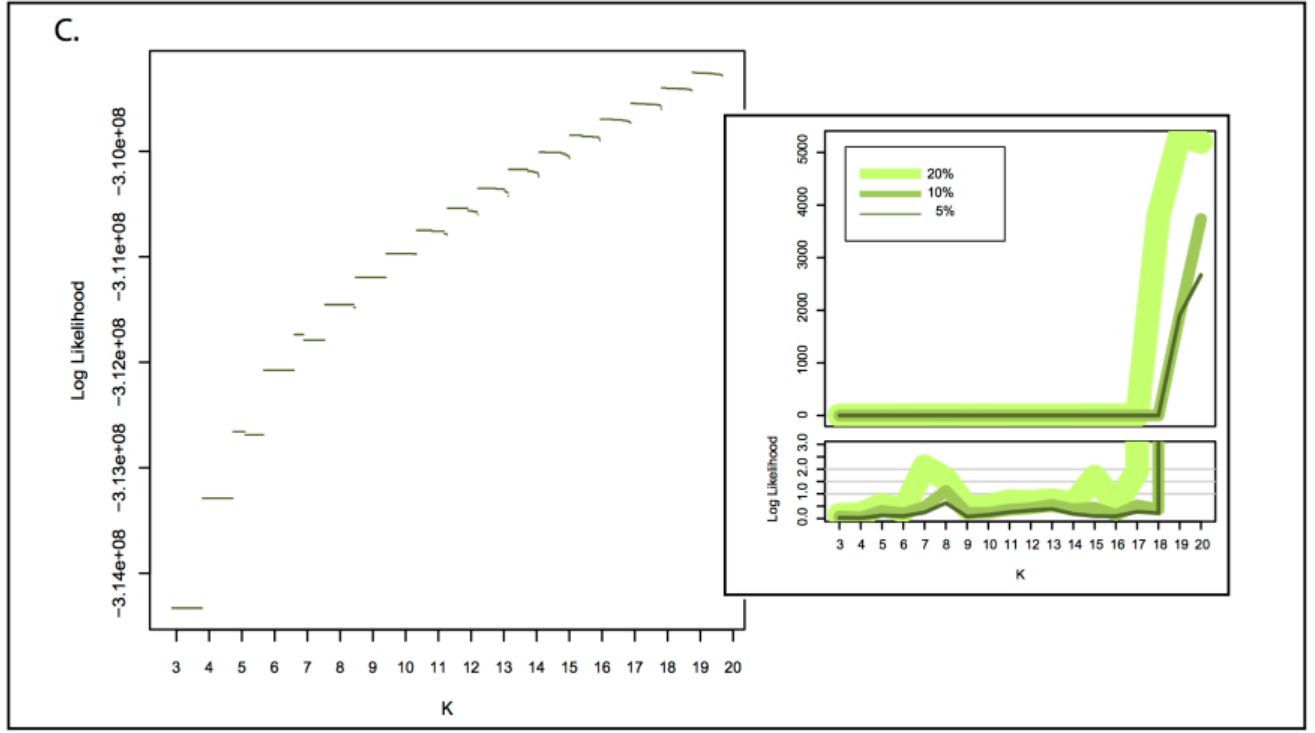
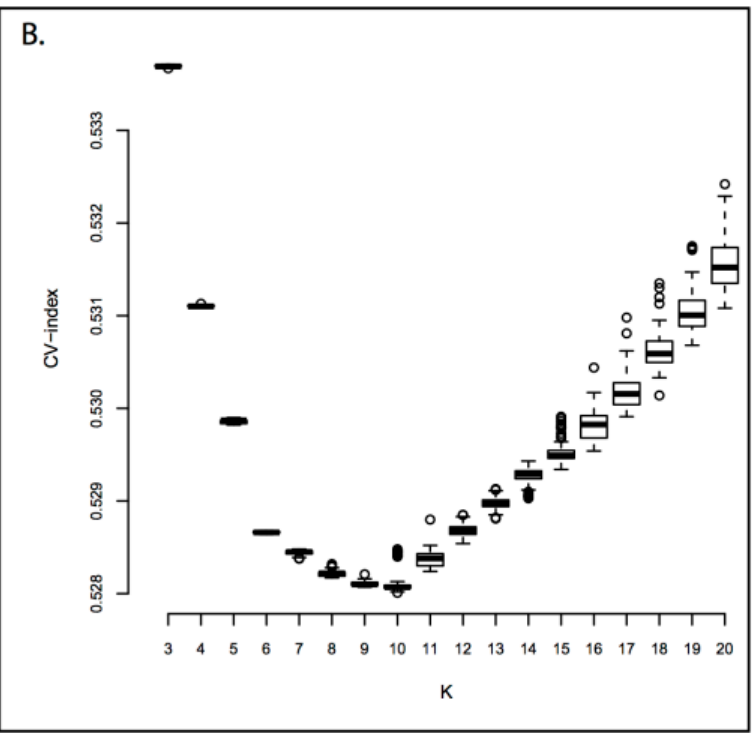
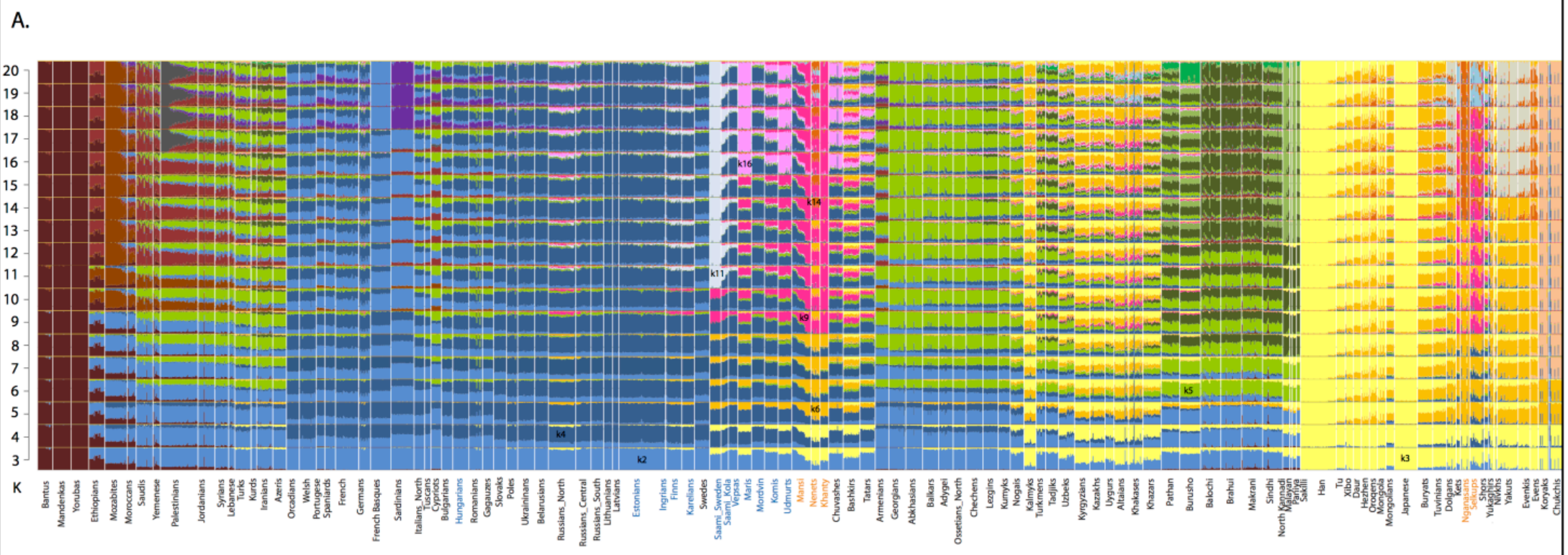
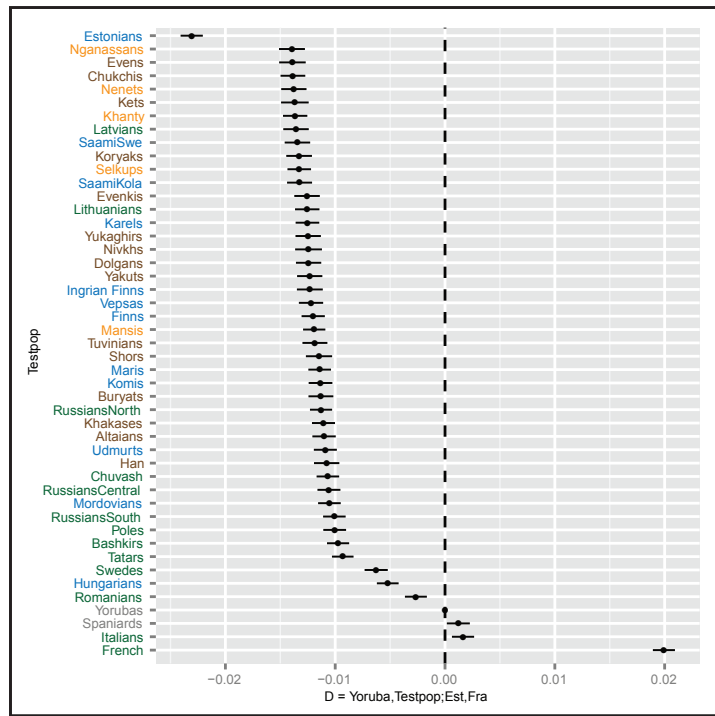
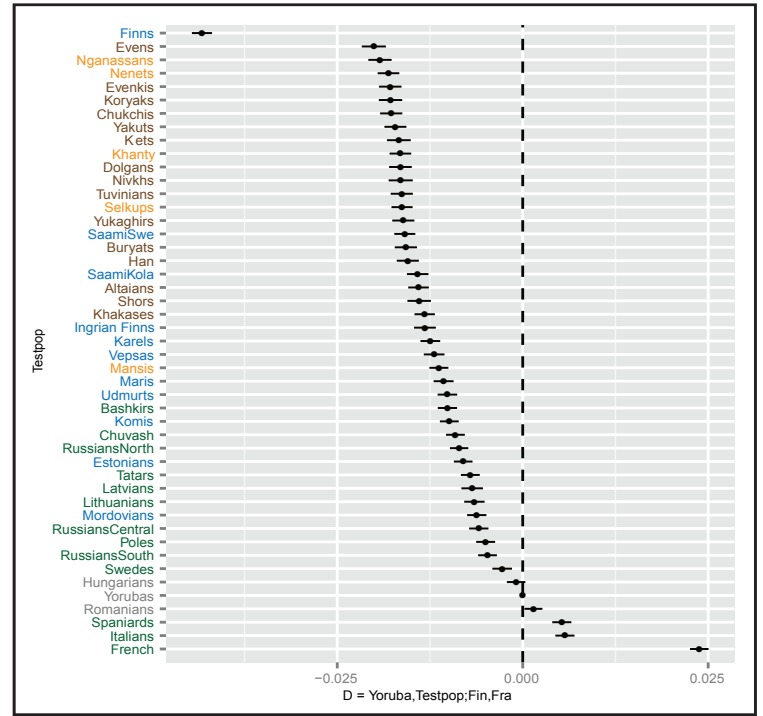


Figure S3. ADMIXTURE plots of the Uralic-speaking populations in a global context based on autosomal SNPs and with the number of assumed ancestral populations (K) ranging from 3 to 20. **A.** Bar plot displaying individual ancestry estimates for studied populations. **B.** Box and whiskers plot of the cross validation (CV) indexes of all K x 100 runs of ADMIXTURE; **C.** Log-likelihood (LL) scores of all K x 100 runs of ADMIXTURE. Inset shows the variation in the fractions (5%, 10% and 20%) of runs that reached the highest LL values.

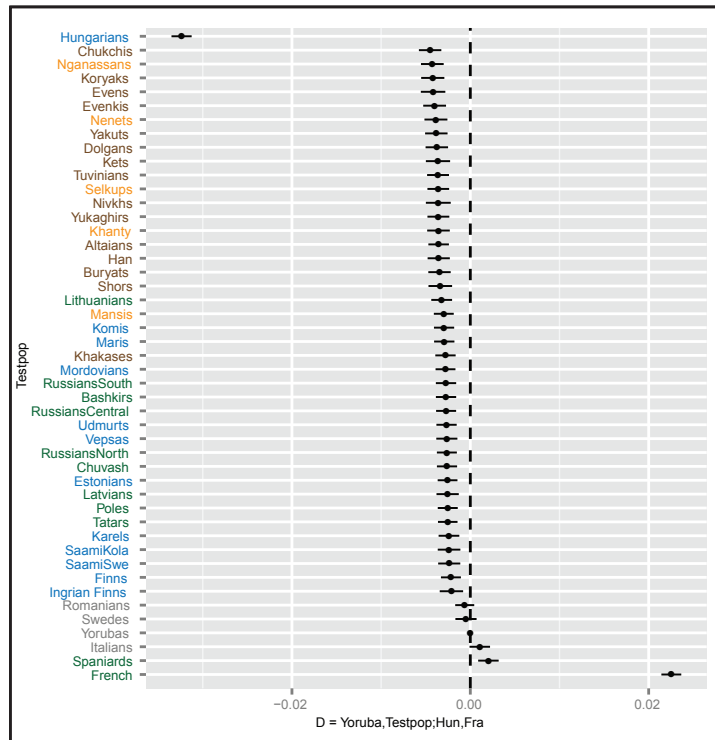
A.



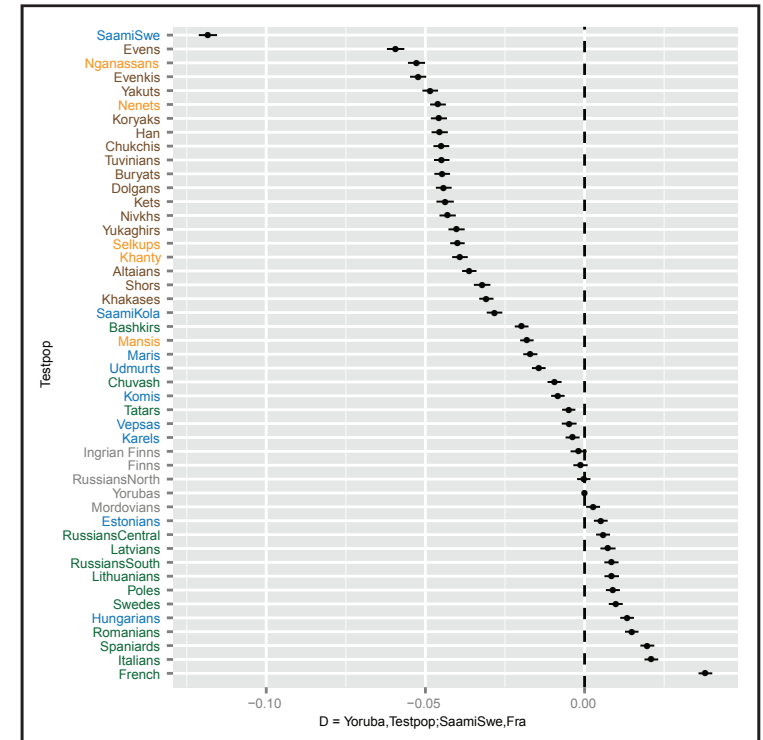
B.

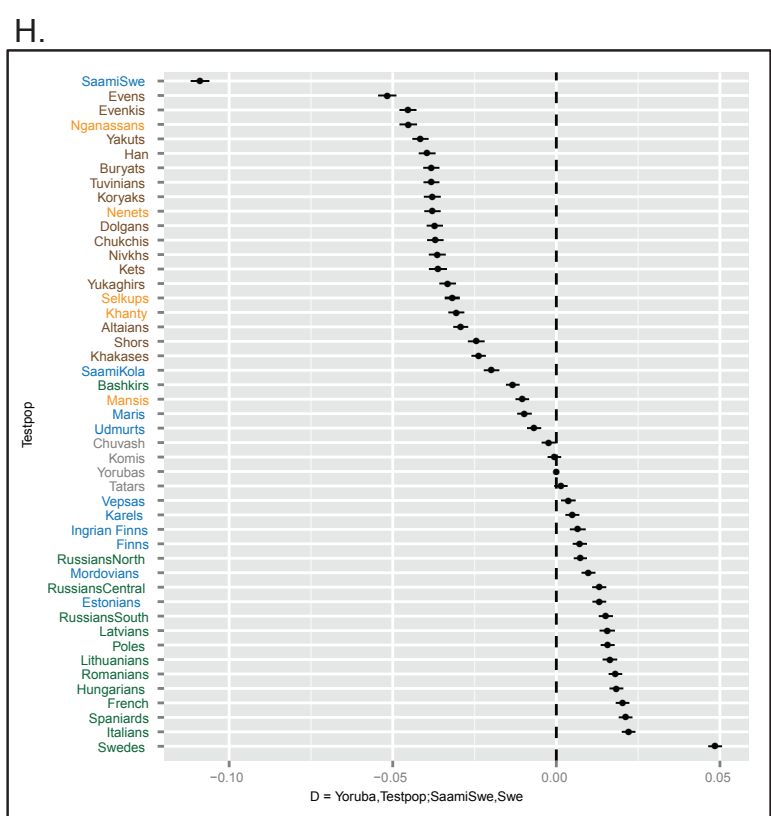
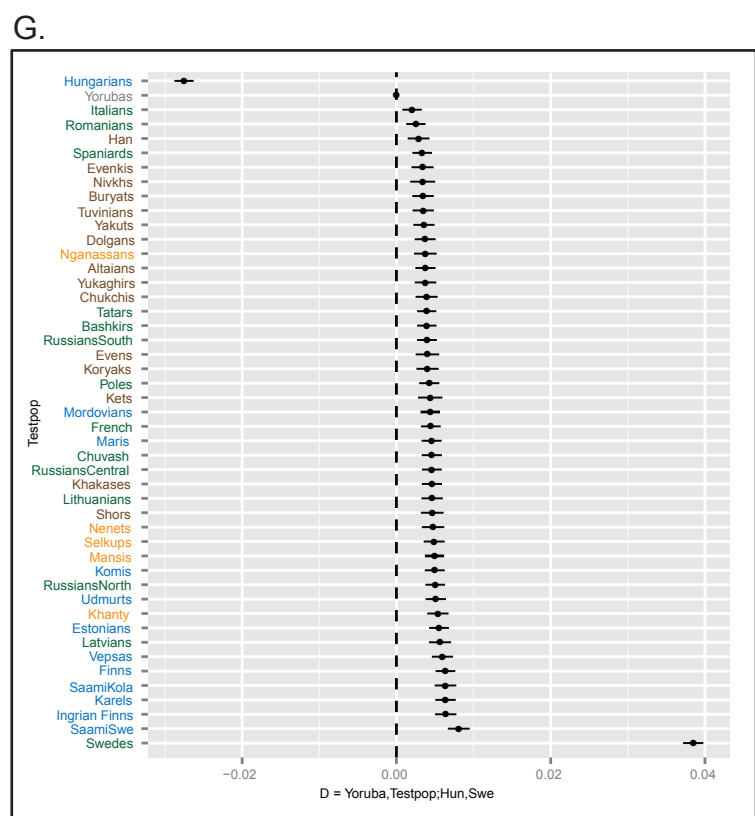
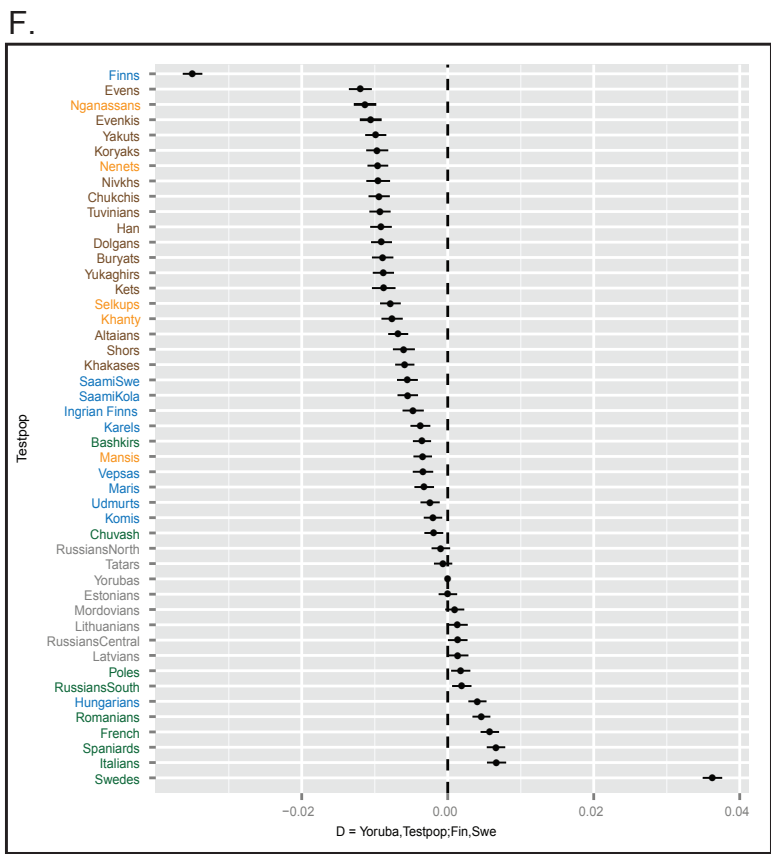
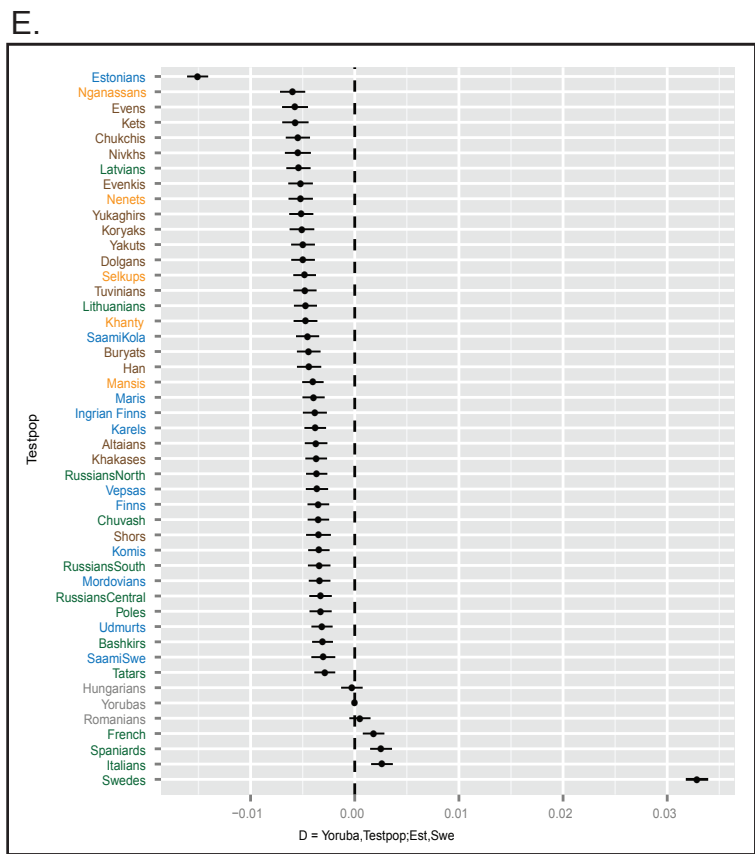


C.

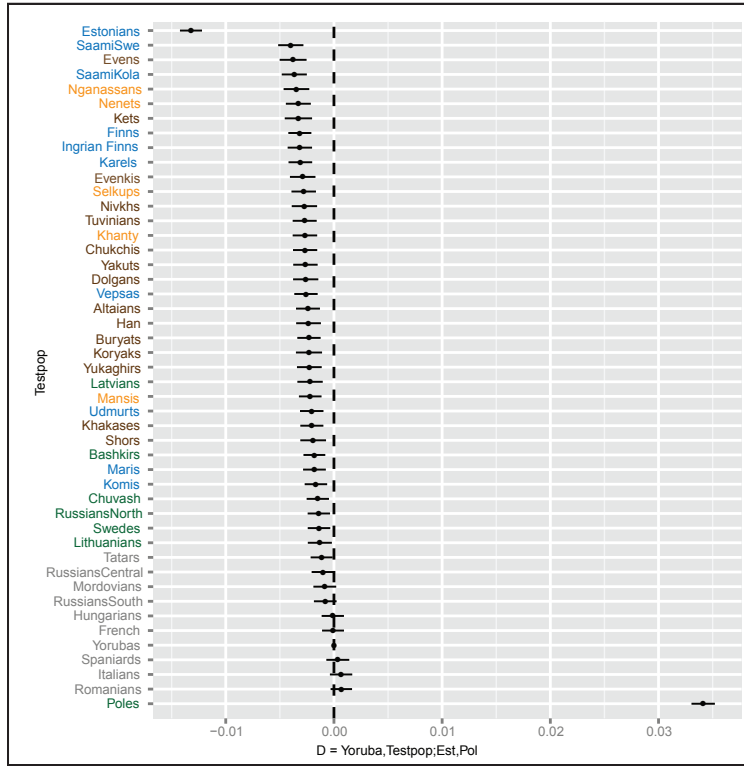


D.

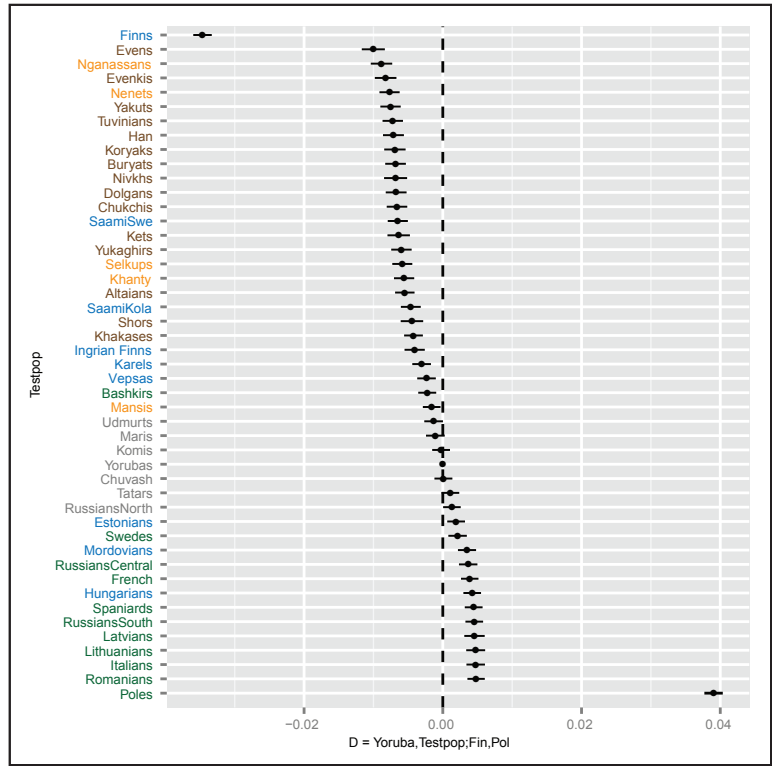




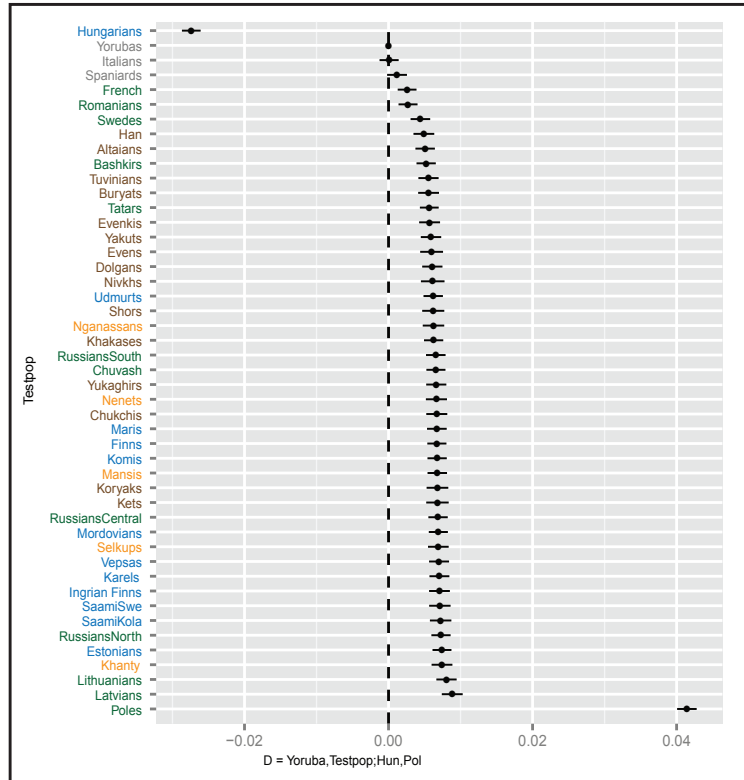
I.



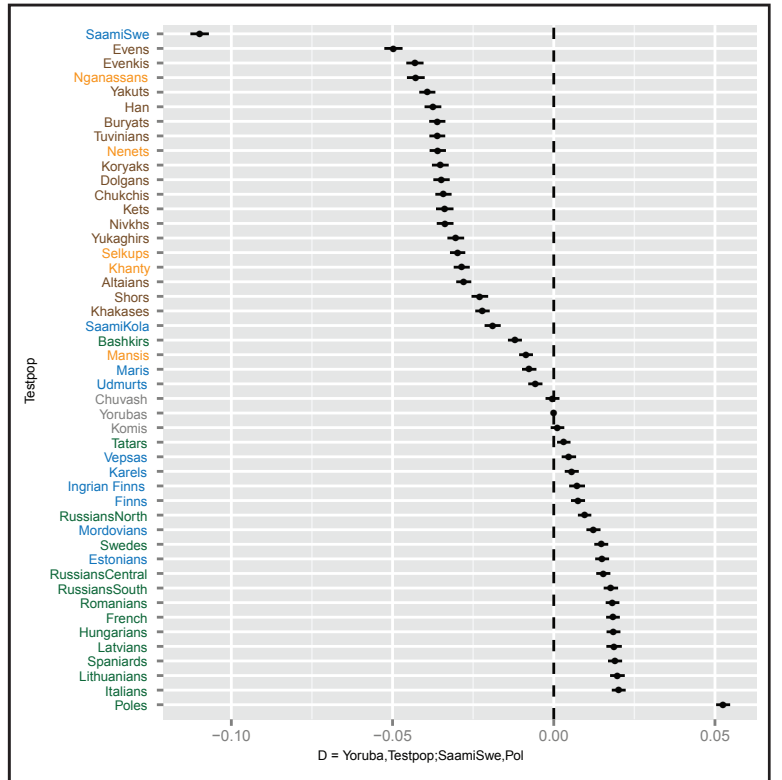
J.



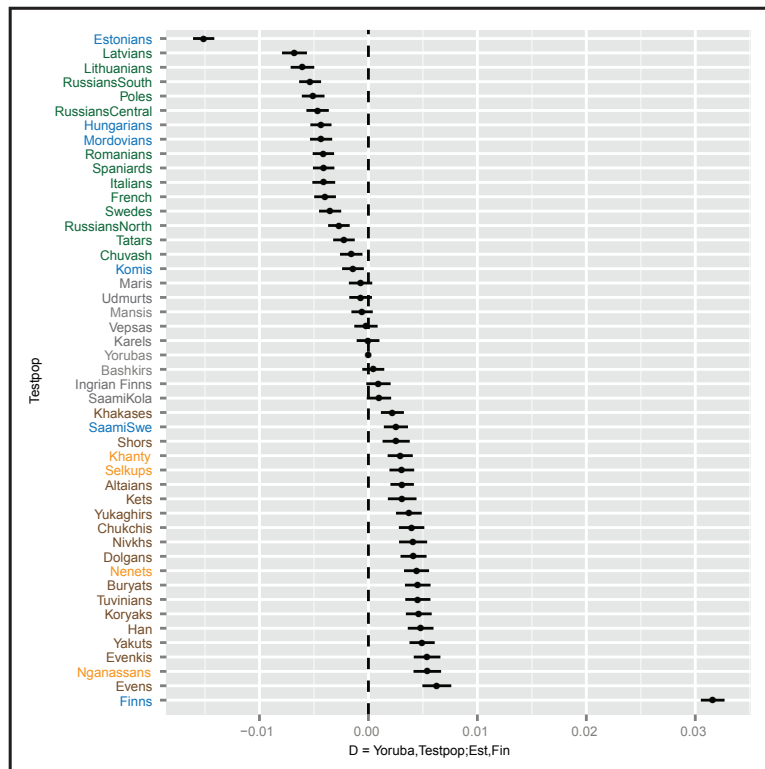
K.



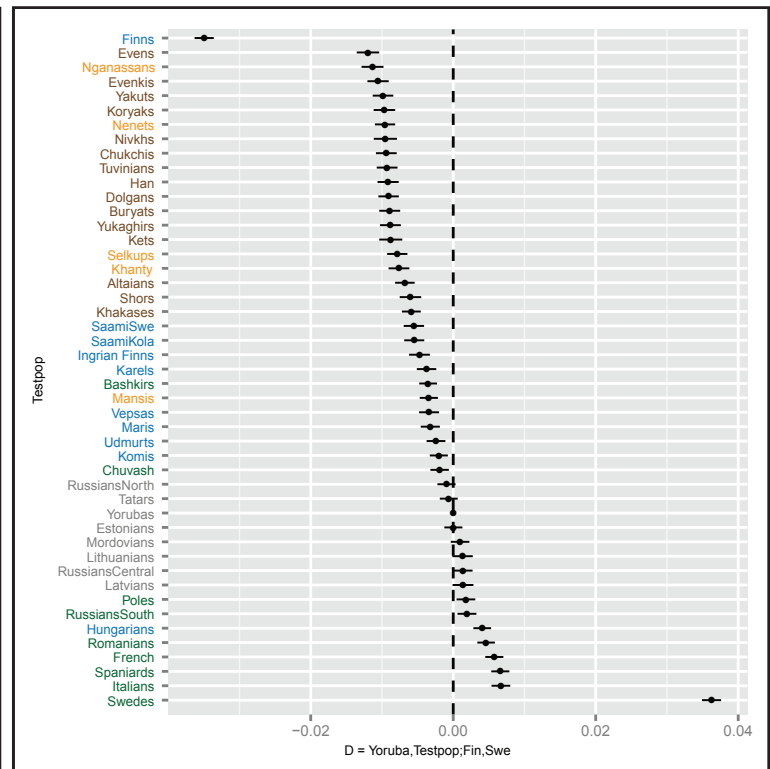
L.



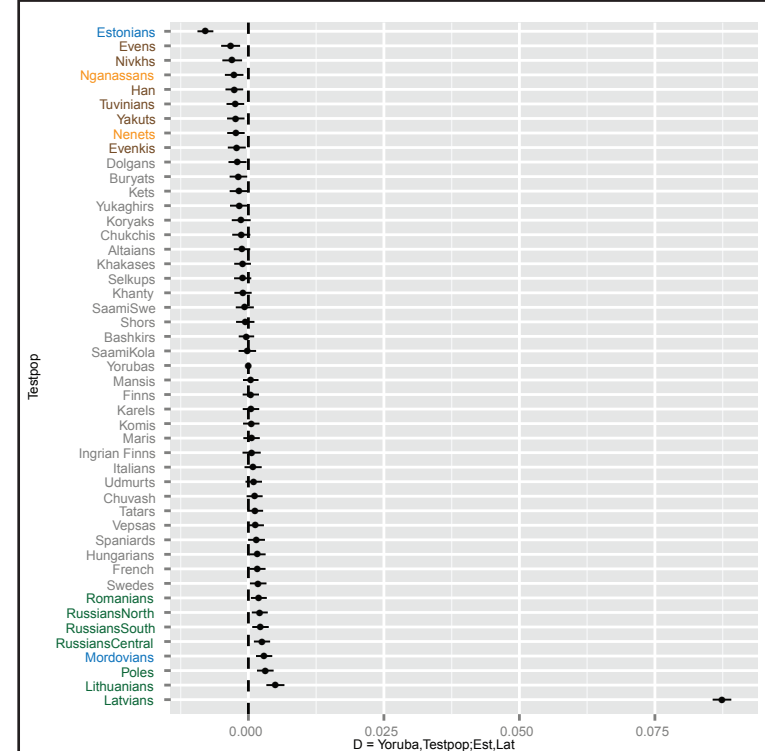
M.



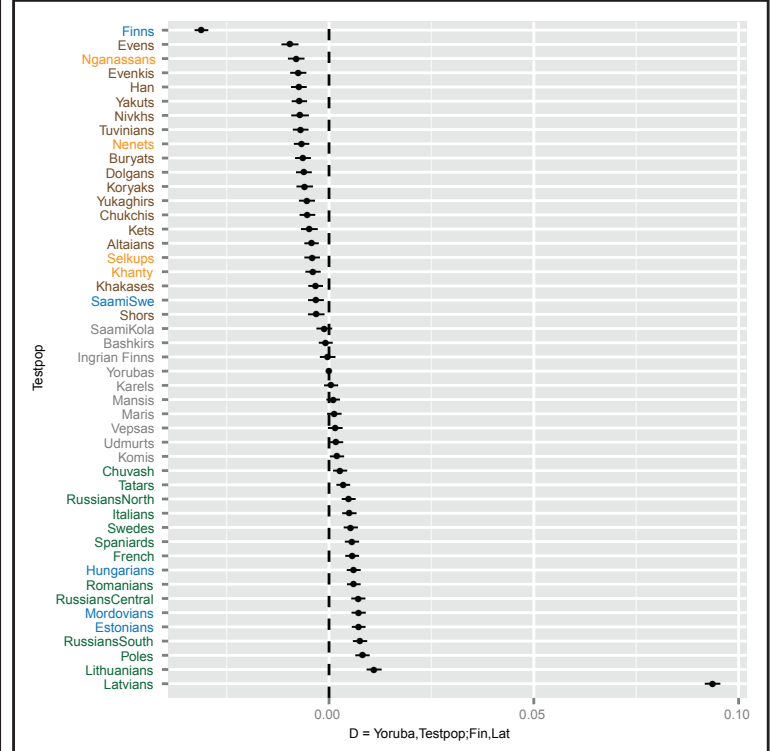
N.



O.



P.



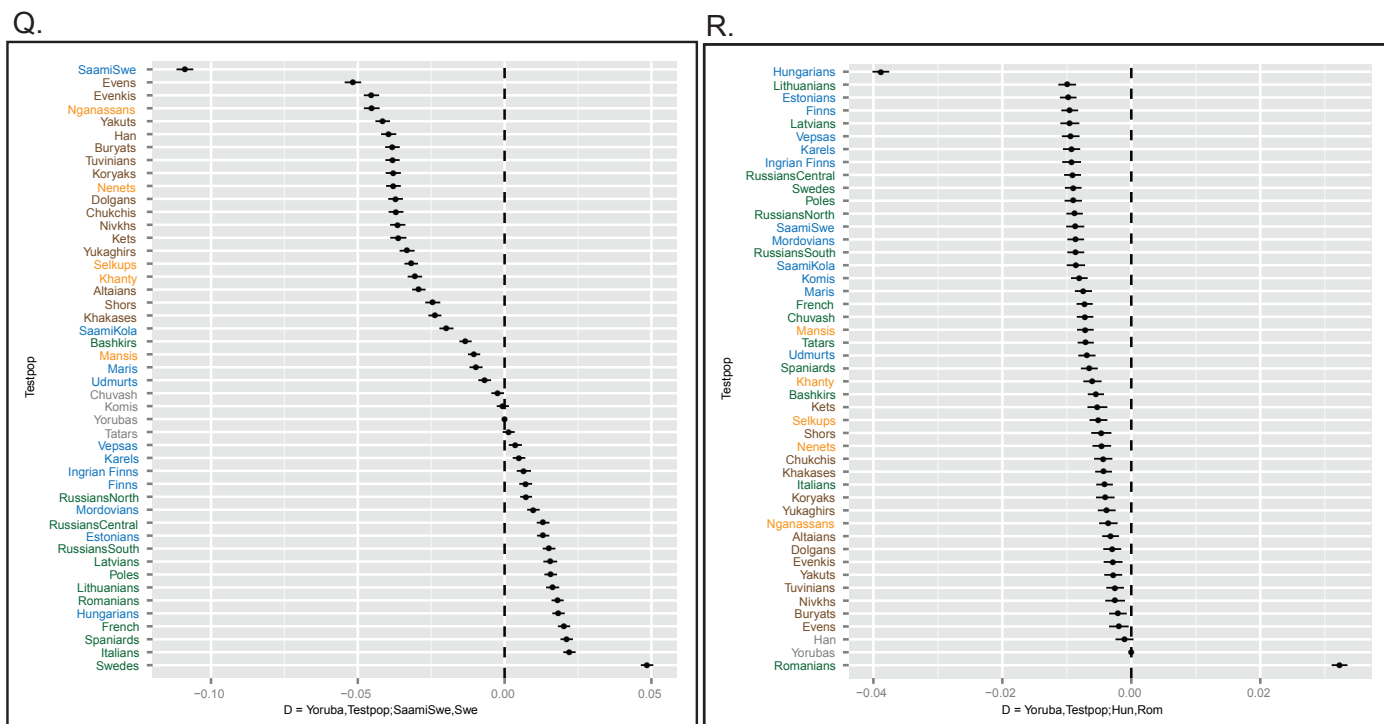
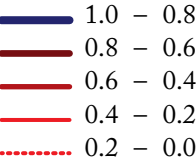


Figure S4. D-statistics calculated for the tree model in the form of $D(\text{outgroup, test population; Uralic speaking population, non-Uralic speaking population})$. Yorubas were used as an outgroup. Uralic speaking populations fixed in the tested model are the Saami from Sweden (SaamiSwe), Finns (Fin), Estonians (Est) and Hungarians (Hun) from Europe. The non-Uralic speaking populations fixed in the tested model are French (Fra, panels **A.-D.**), Swedes (Swe, panels **E.-H., N, Q**) and Poles (Pol, panels **I.-L.**). Panels **M.-R.** depict D-statistics opposing neighbouring population pairs Estonians-Finns (**M.**), Finns-Swedes (**N.**), Estonians-Latvians (Lat, **O.**), Finns-Latvians (**P.**), Saami and Swedes (**Q.**) and Hungarians-Romanians (Rom, **R.**). The values on the Y-axis are sorted by D value. Color codes of populations showing significant deviations from $D=0$ ($Z \text{ score} \geq 3$) correspond to linguistic affinities of tested populations: blue – European Uralic speaking populations; green – European non-Uralic speaking populations; orange – West Siberian Uralic speaking populations; brown – Siberian and East Asian non-Uralic speaking populations. Grey colored labels indicates $D=0$ ($Z \text{ score} < 3$), standard errors to the point estimates are shown with the black bars.

Branch support:



GT group:

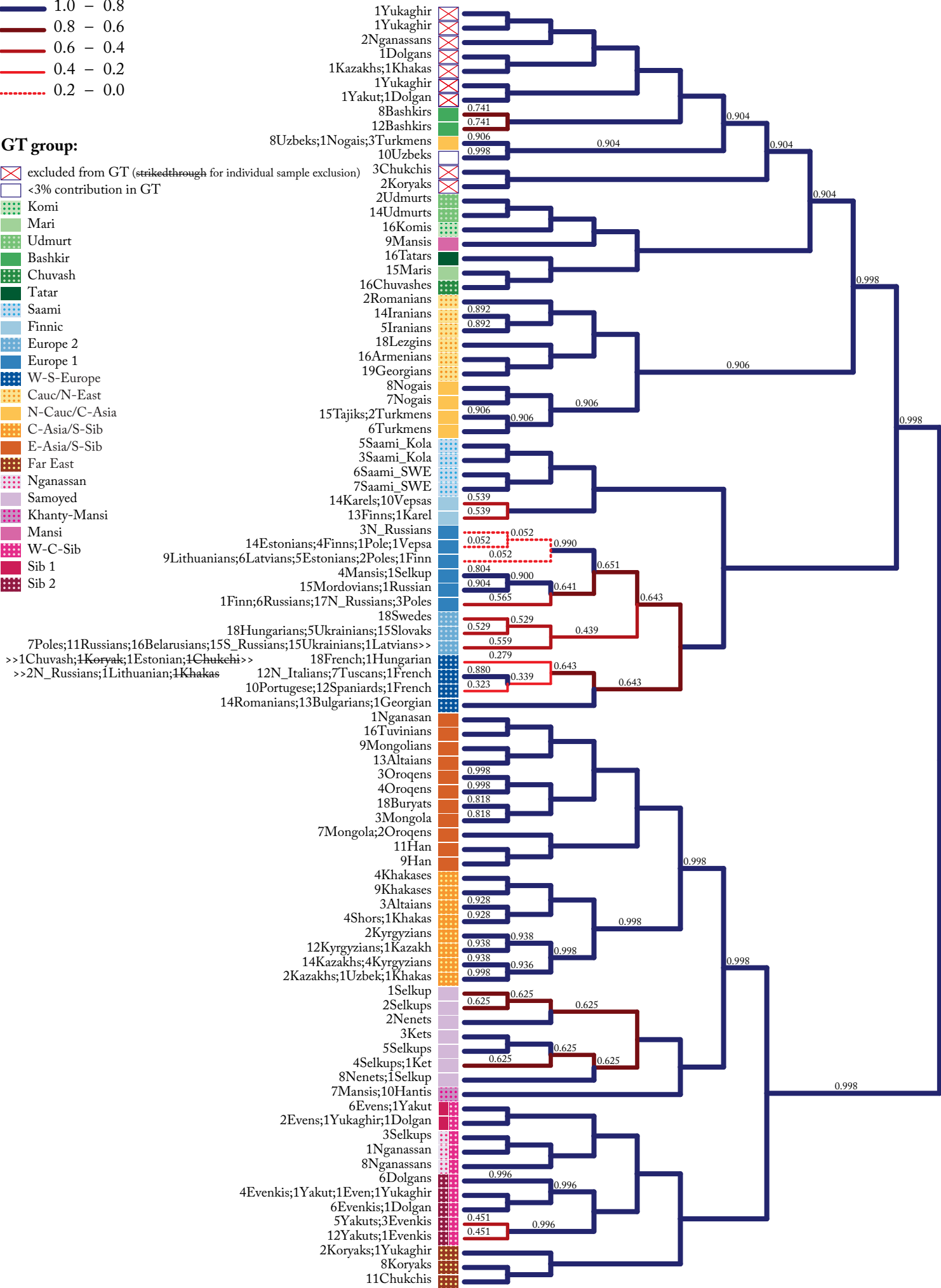
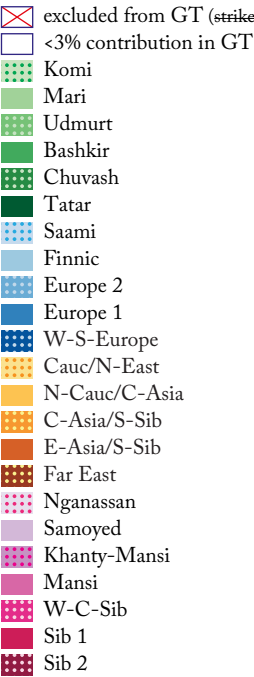


Figure S5. Clustering of individual samples from the comparative dataset, as inferred by fineSTRUCTURE (FS). The tree clusters individuals with similar copying vectors. Labels identify how many samples and which samples are included in each cluster. For example, ‘7Mongola;2Oroqens’ cluster comprises of 7 Mongola and 2 Oroqen individuals. Individual tips were manually inspected and grouped for further admixture analysis in GLOBETROTTER (GT, **Figure 5**). GT groups are color-coded; legend is given on the left-hand side. For all GT runs except of analysis of admixture in the Nganasan cluster, single ‘W-C-Sib’ (■, West and Central Siberia) group was used. For the GT analysis of the Nganasans, ‘W-C-Sib’ was split into three sub-groups: ‘Nganassan’ (■), ‘Sib 1’ (■) and ‘Sib 2’ (■); see Materials and Methods for further details. FS populations (crossed out boxes, ☒) and individual samples (strikethrough font) which show unusually high levels of admixture were excluded from GT. Single donor group, ‘10Uzbeks’, contributed less than 3% of ancestry to all GT target populations (**Figure 5**). Line thickness of individual branches indicates statistical bootstrap support; legend is given in the top left corner.

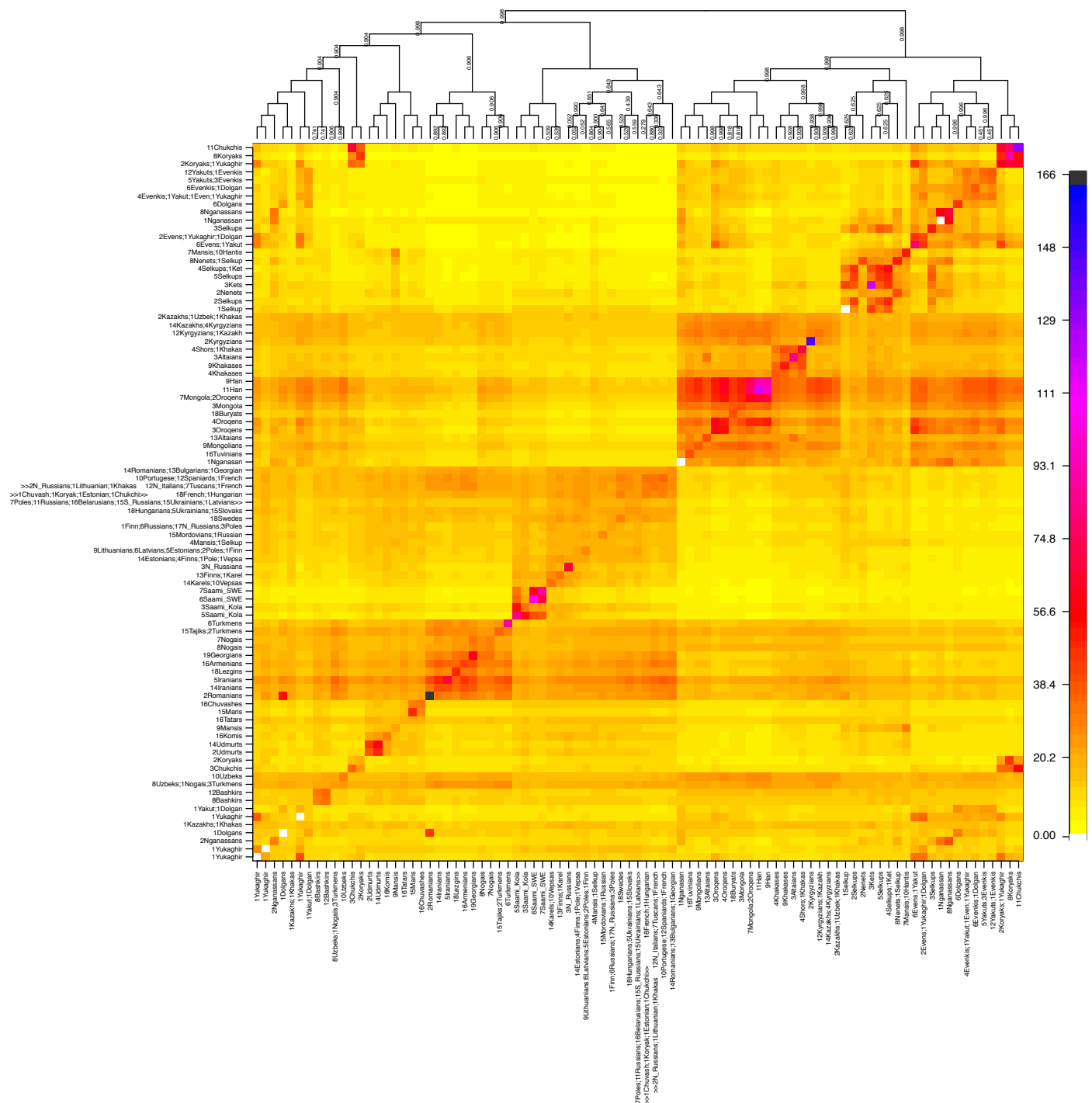


Figure S6. Results of the fineSTRUCTURE clustering analysis using copying vectors generated from chromosome painting. Each row of the heatmap is a recipient copying vector showing the number of chunks shared between the recipient cluster and every donor group (columns). Labels identify how many samples and which samples are included in each cluster. A simplified population clustering dendrogram is shown on top. A detailed dendrogram is given in Figure S5.

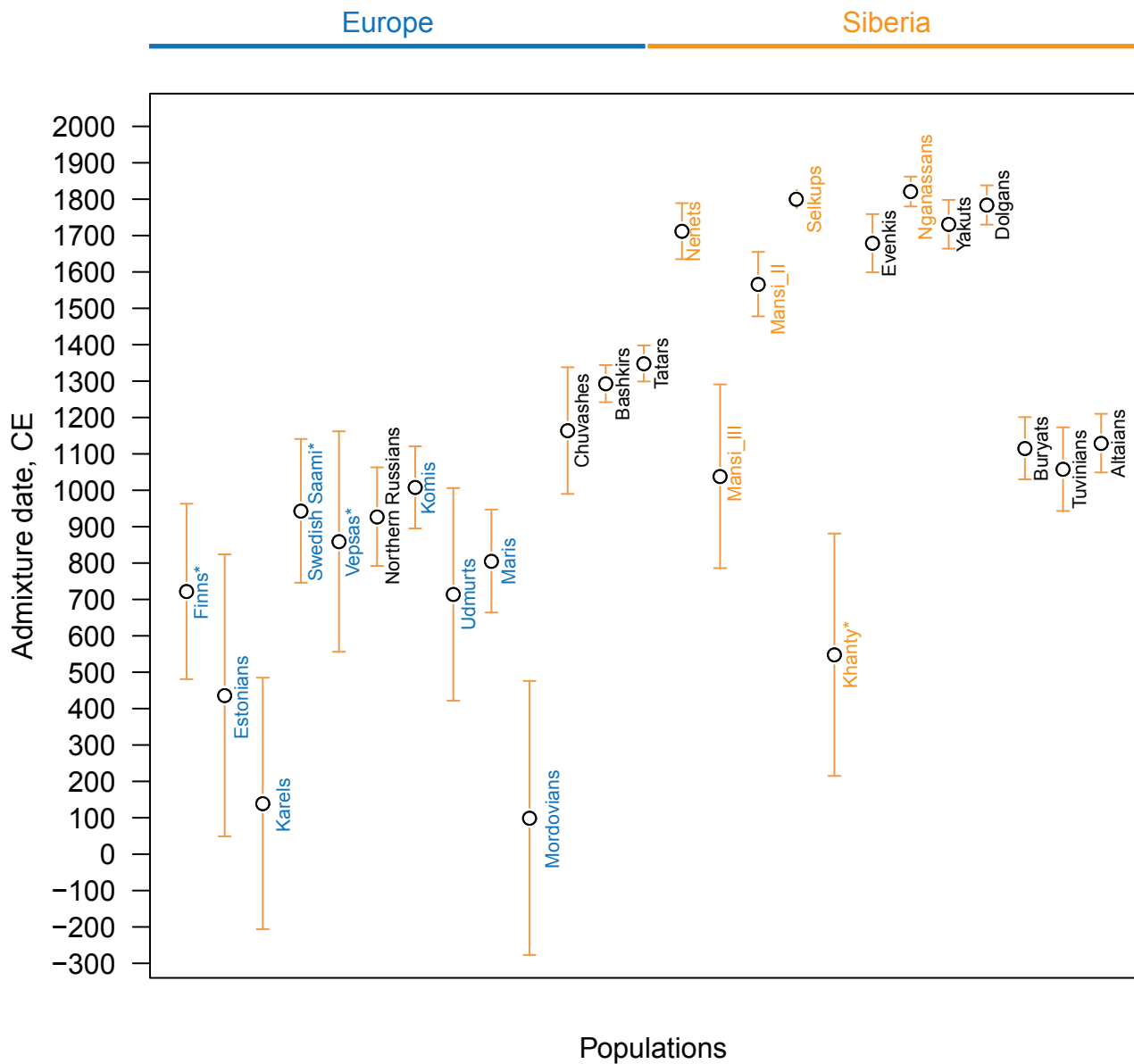
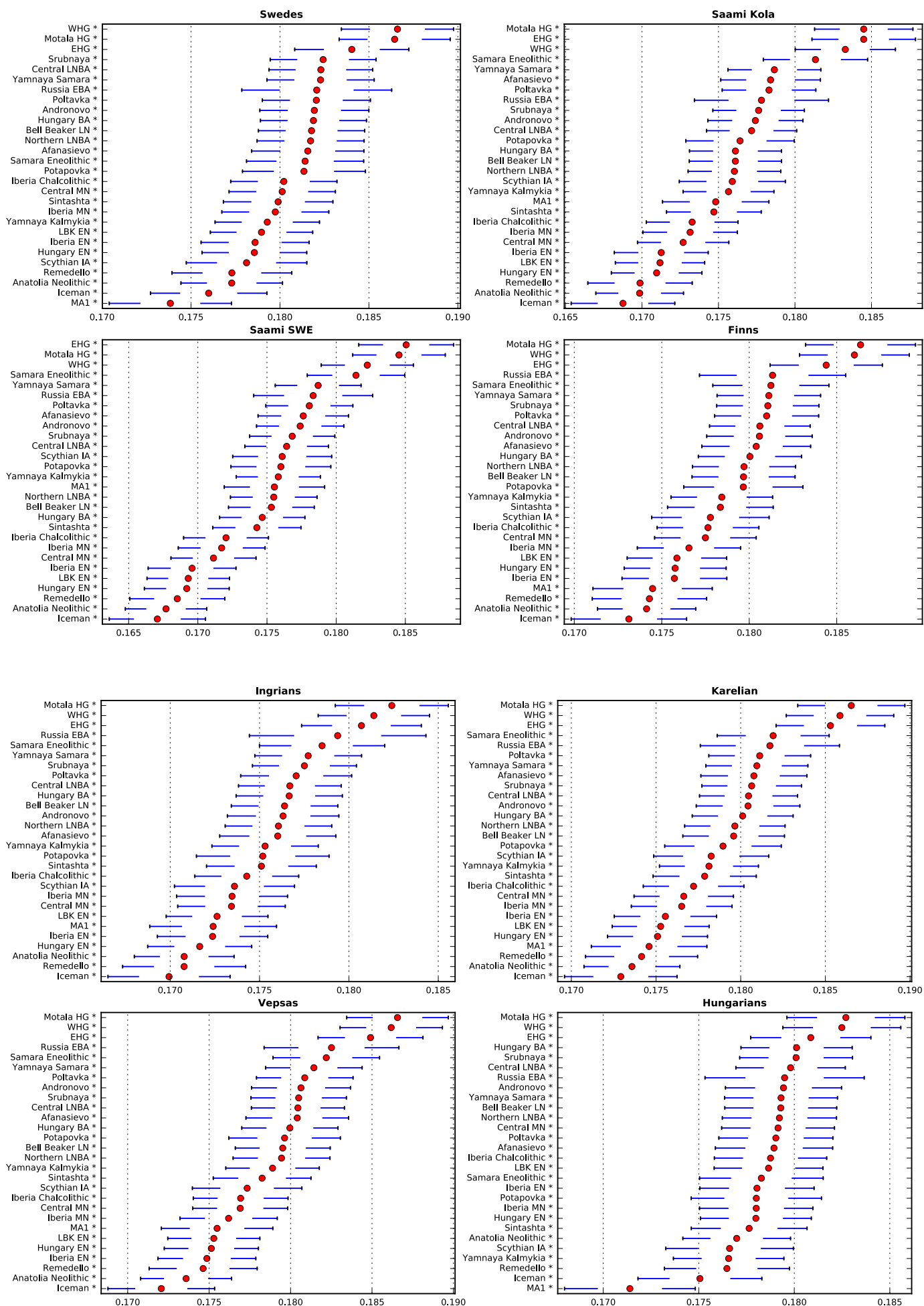
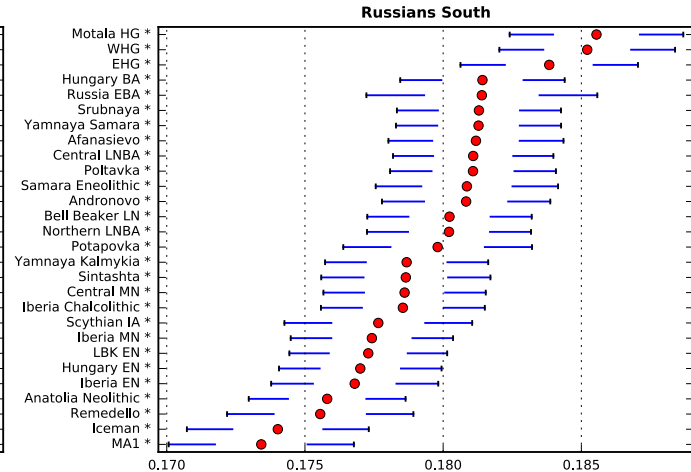
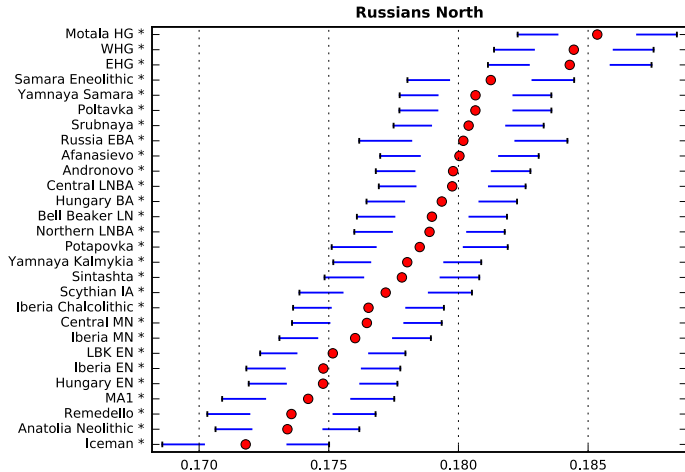
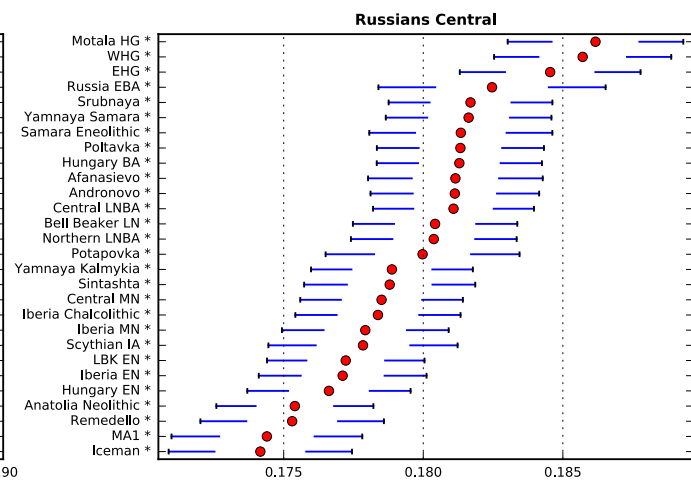
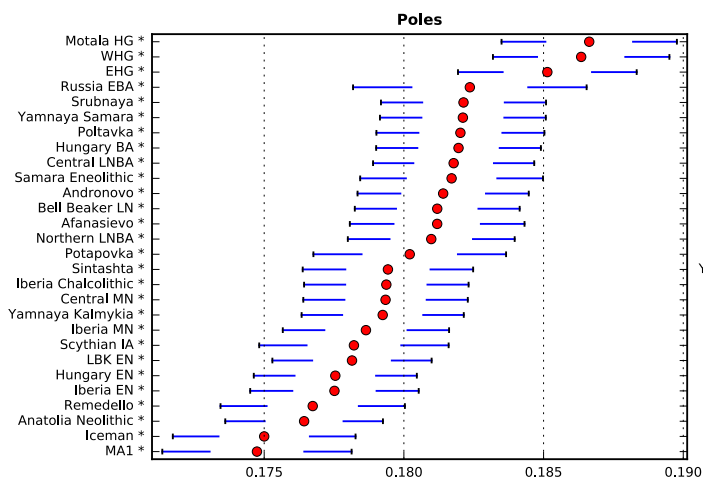
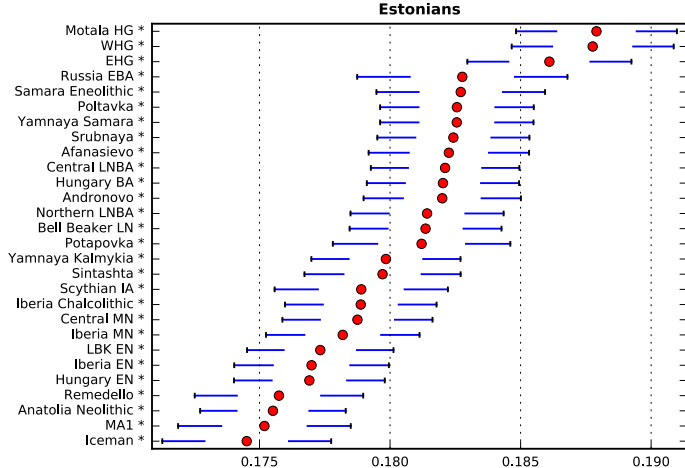
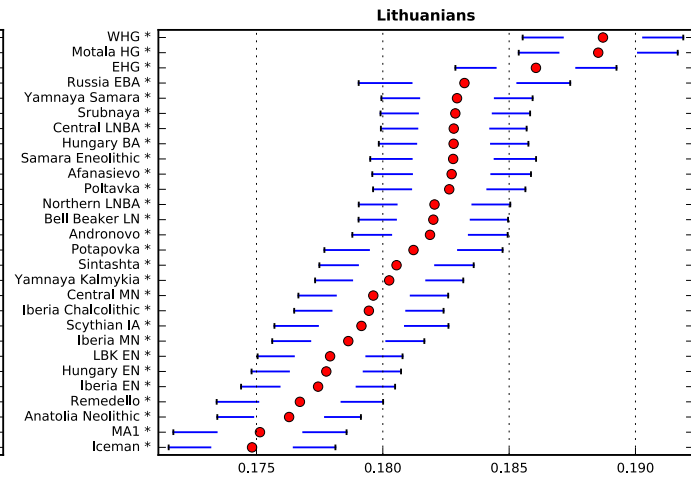
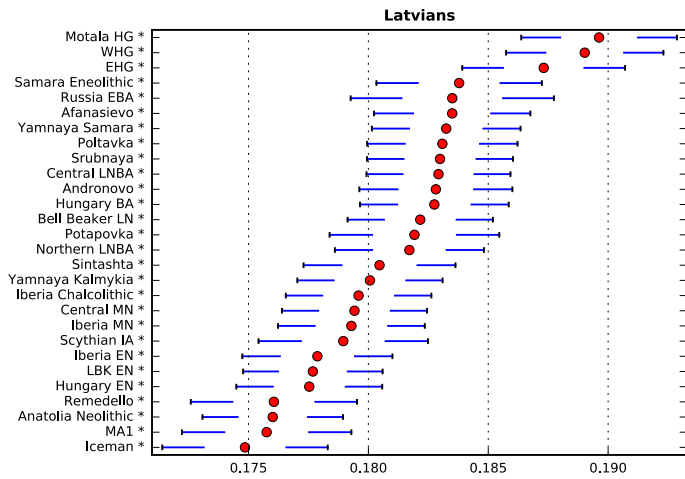
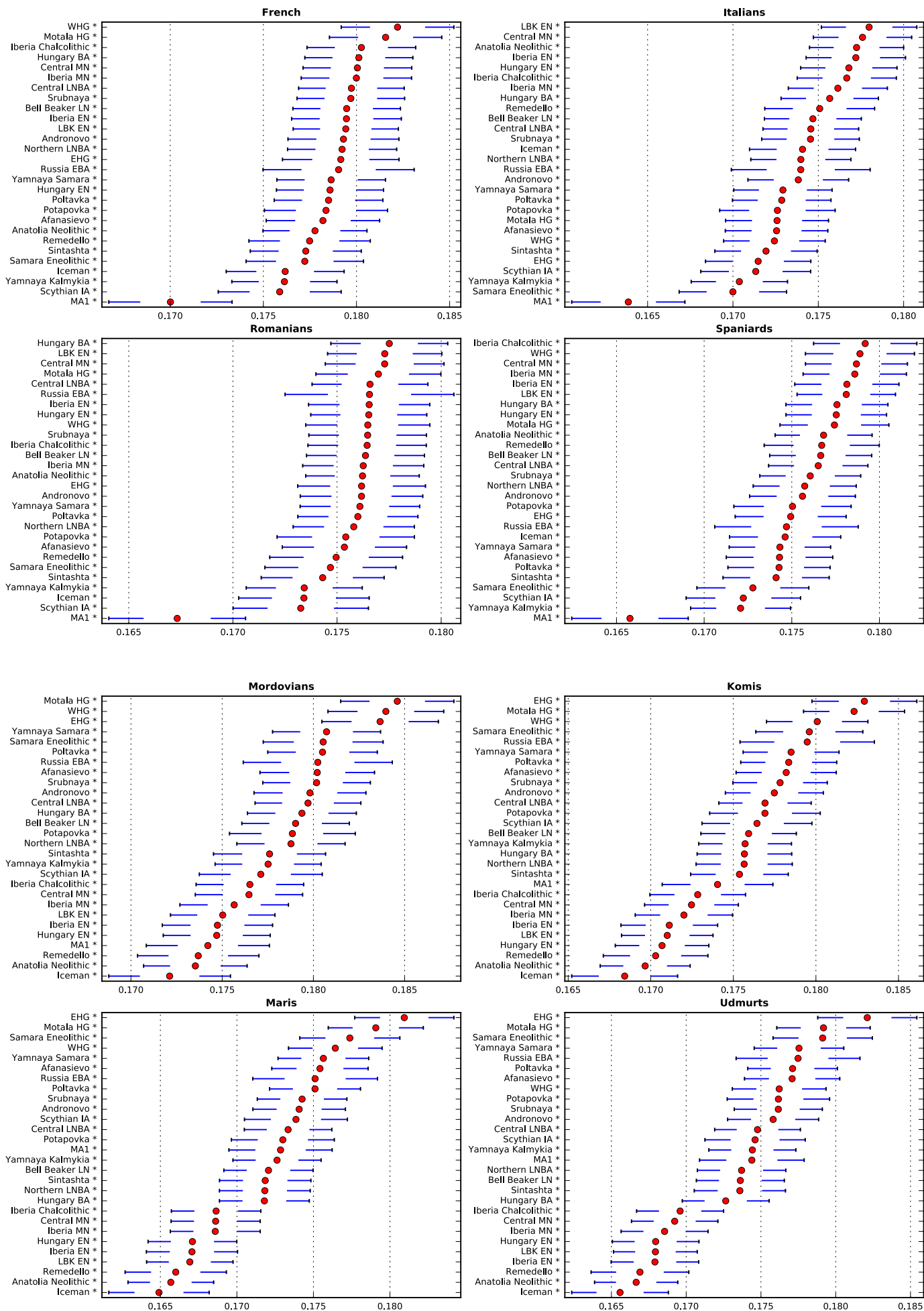


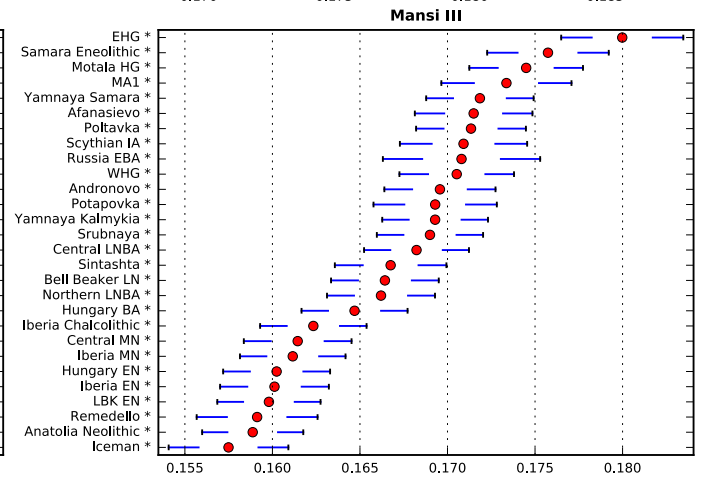
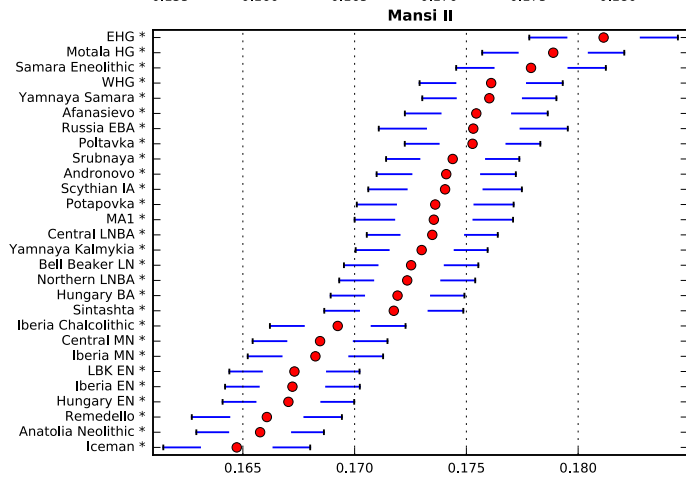
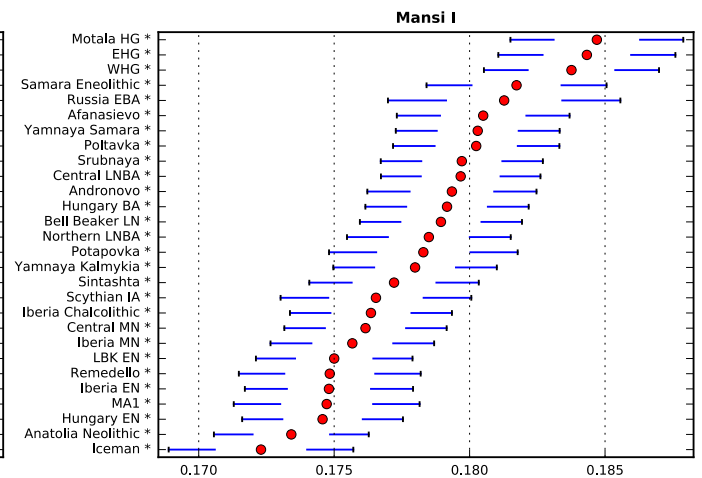
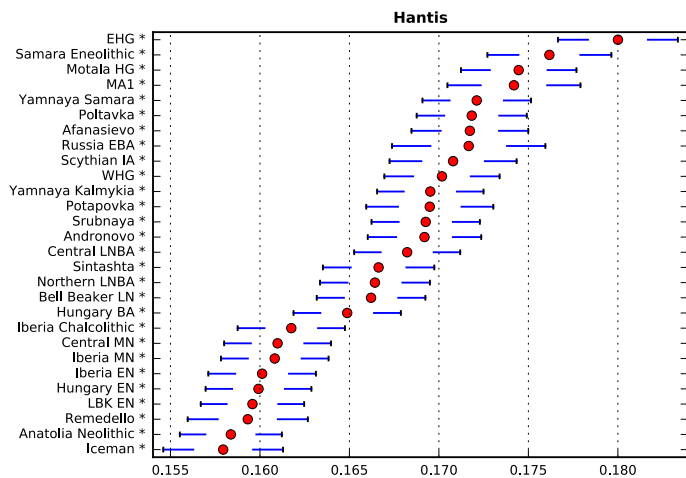
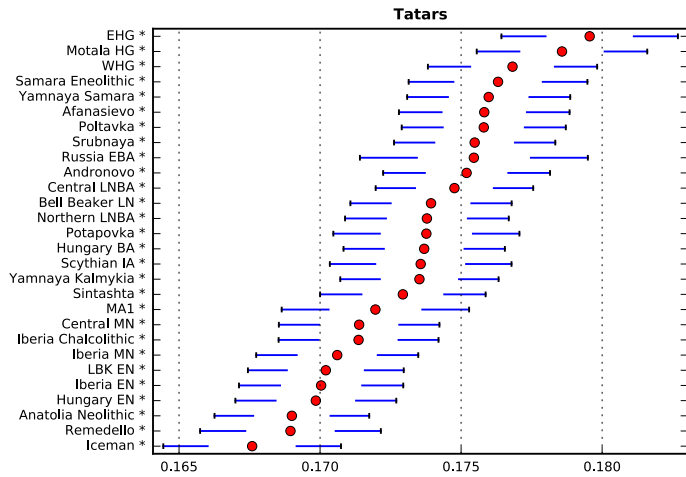
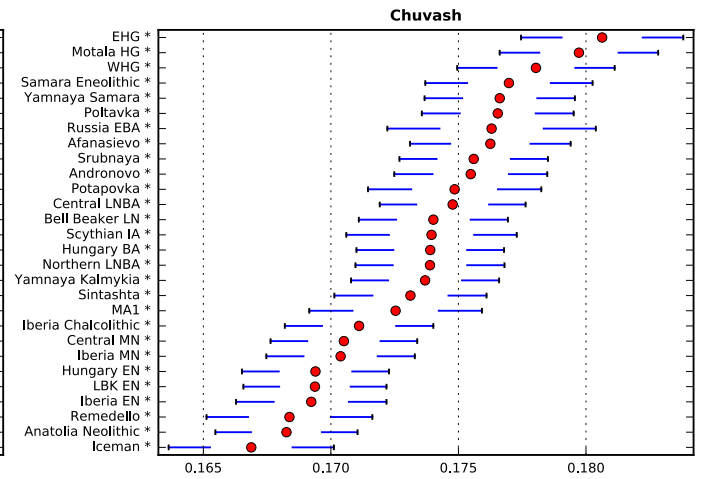
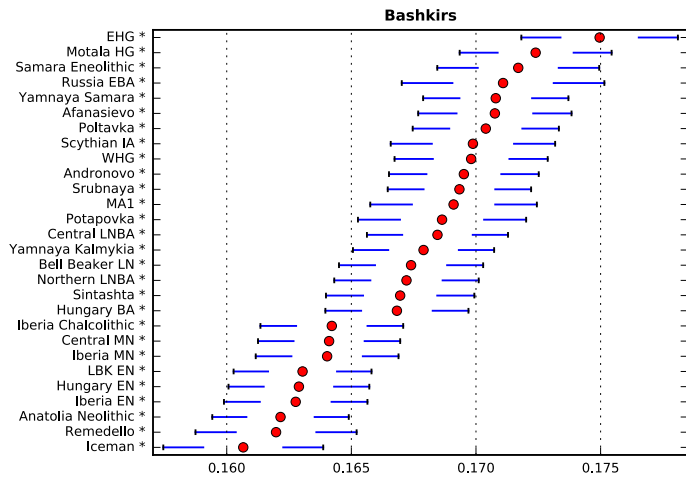
Figure S7. Admixture dates for the eastern and western components of the Uralic-speaking populations (highlighted according to geography blue for Europe and orange for Siberia) in the context of their geographical neighbors on an absolute time scale. Dates are calculated with ALDER according to decay rates of two-reference weighted linkage disequilibrium curve using the generation time of 30 years. Black circles show point estimates and error bars indicate 95% confidence intervals. Admixture dates before Common Era (CE) are shown with a negative sign. (*) indicates admixed populations with inconsistent LD curve decay rates.

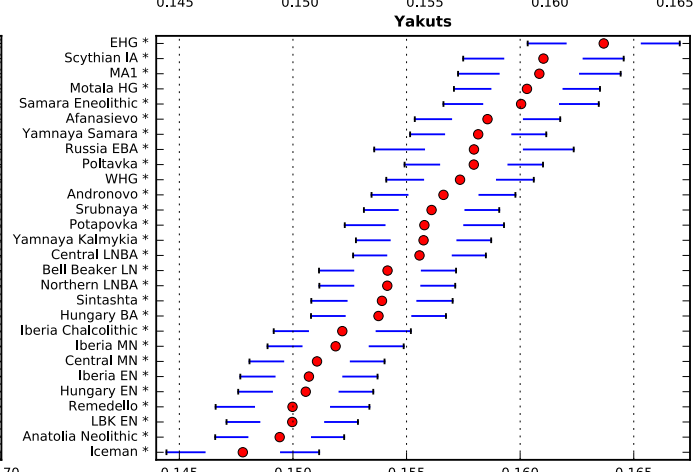
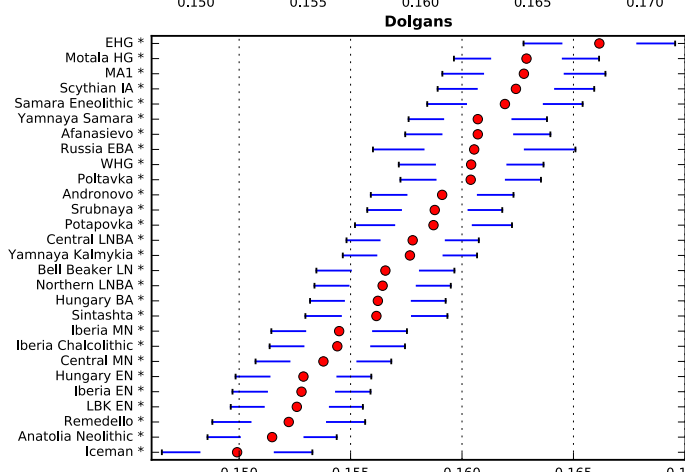
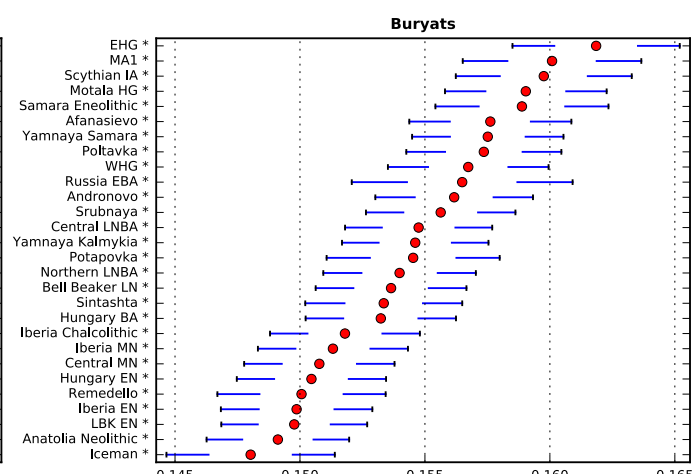
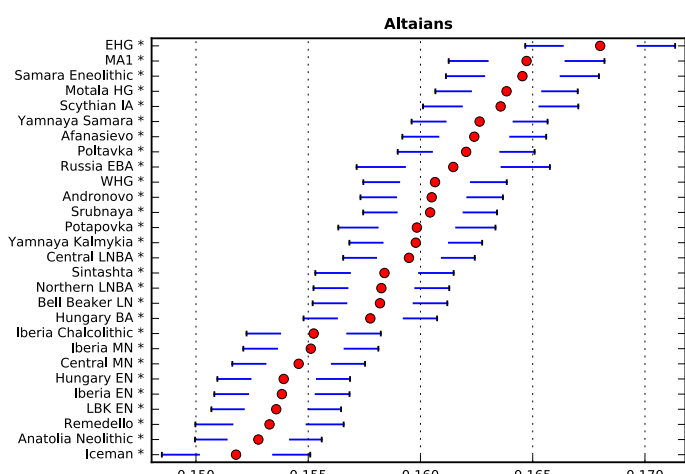
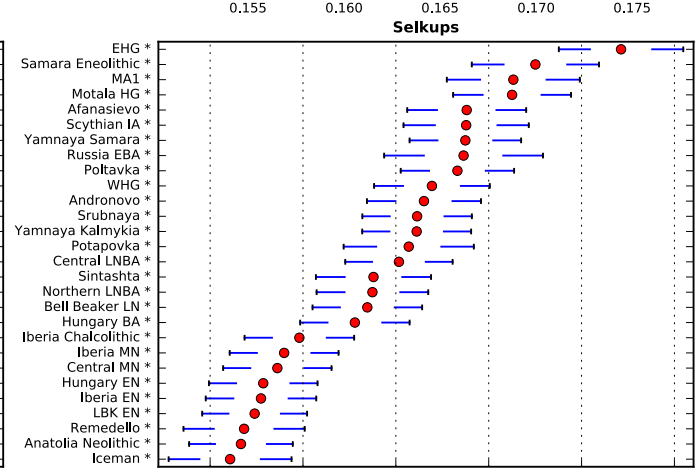
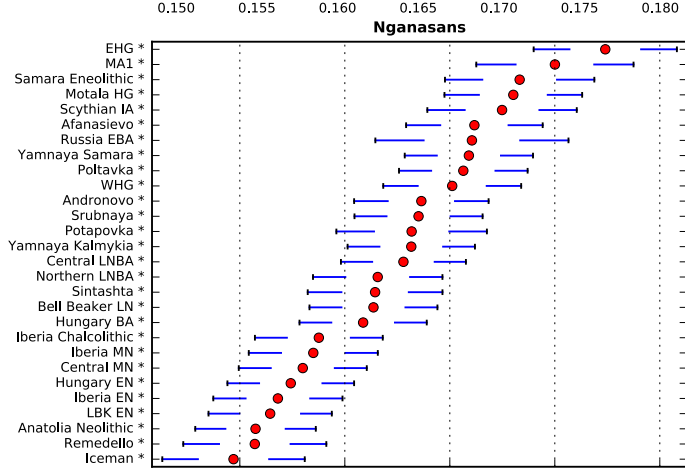
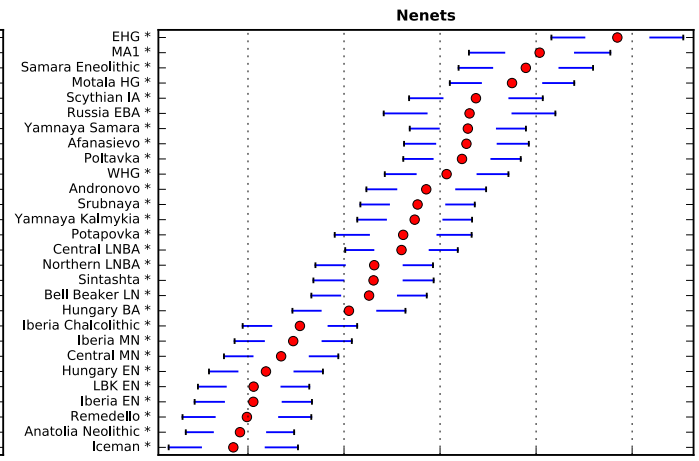
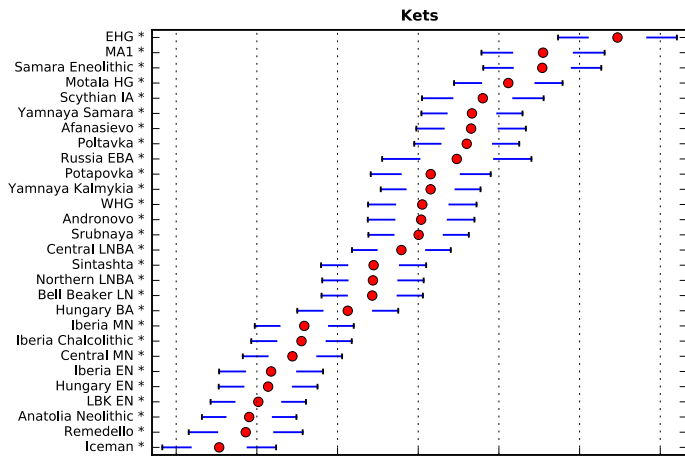
A.

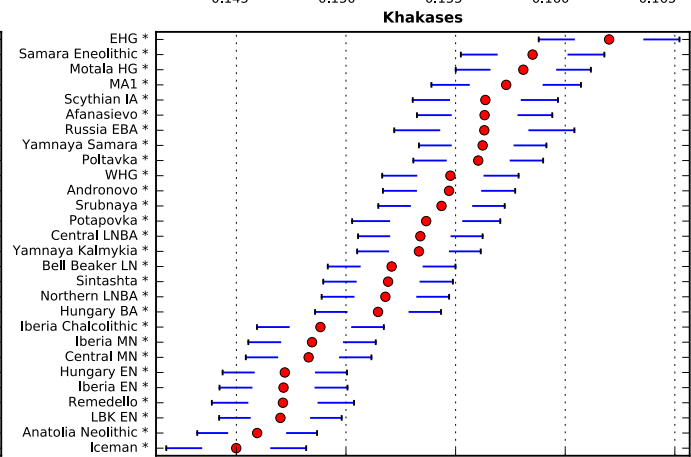
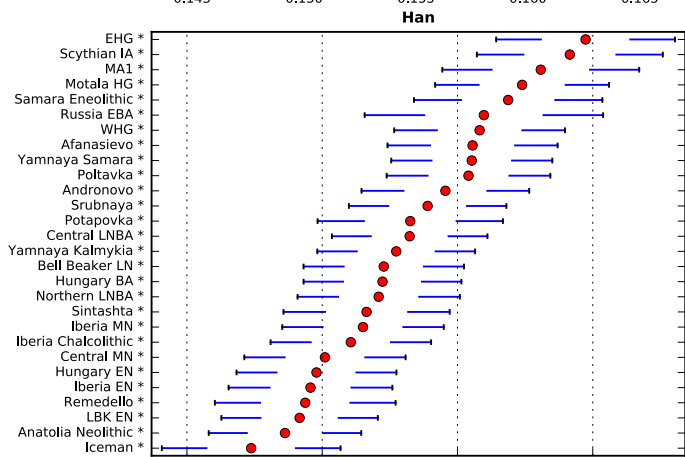
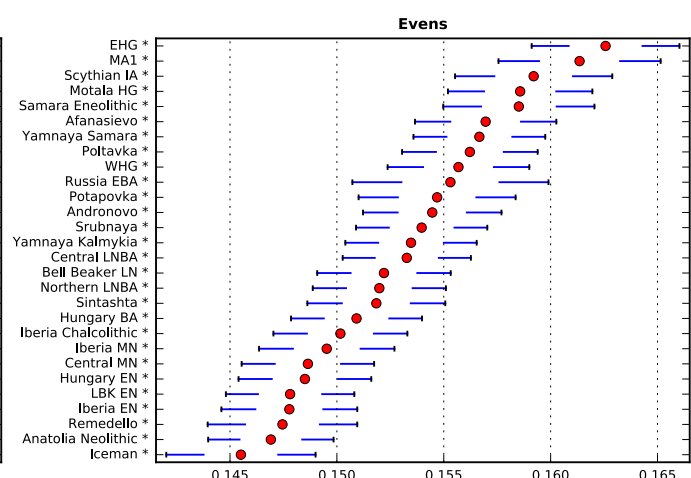
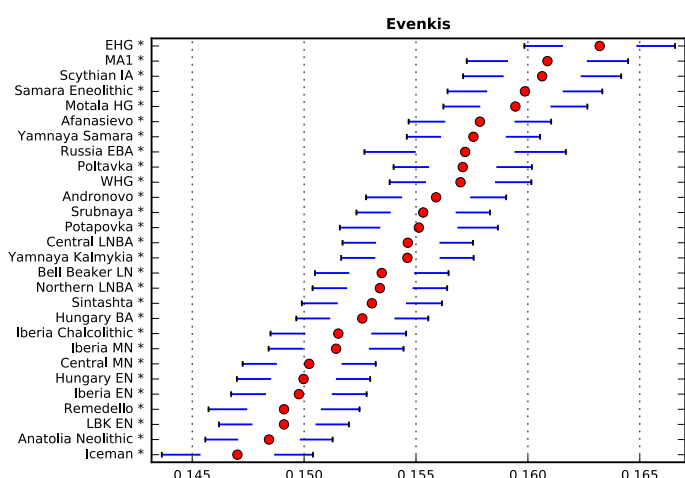
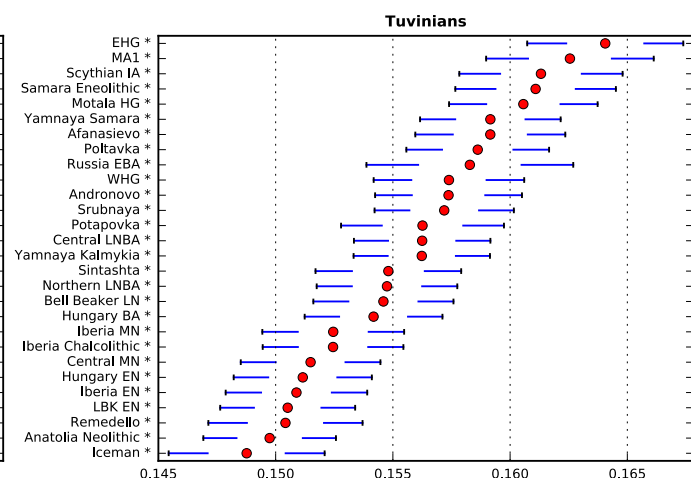
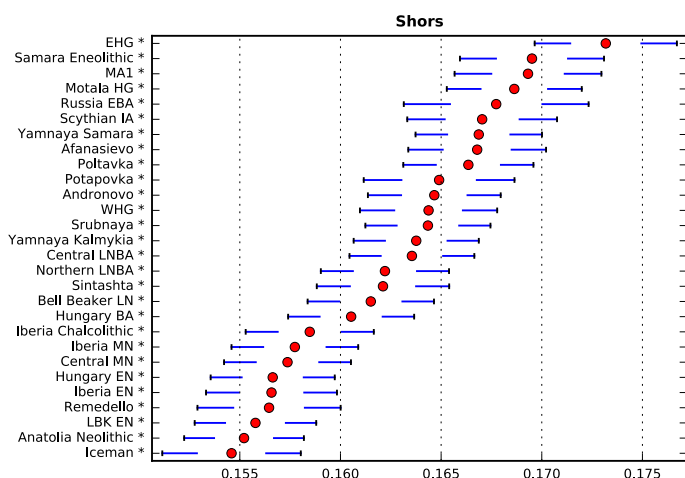


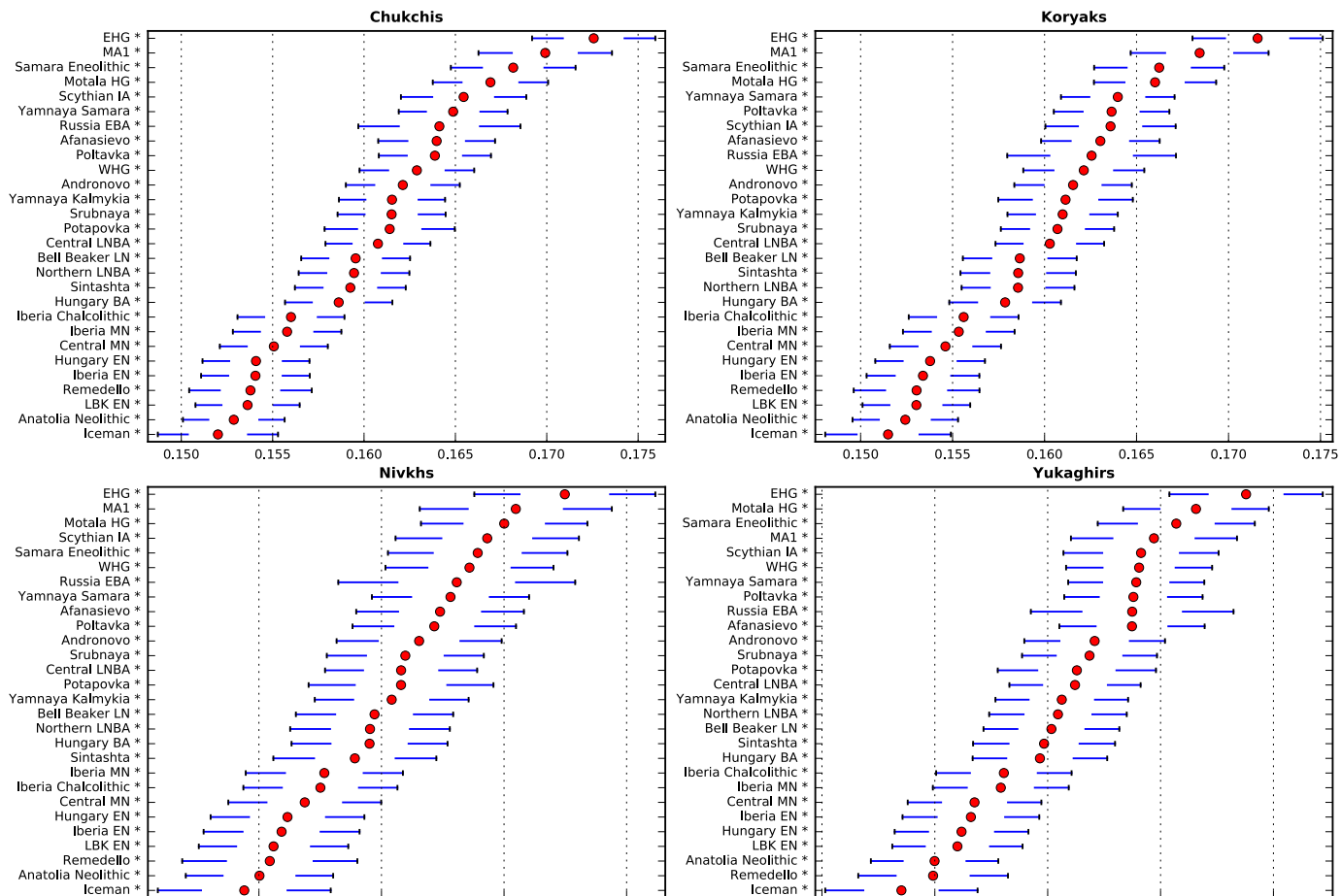




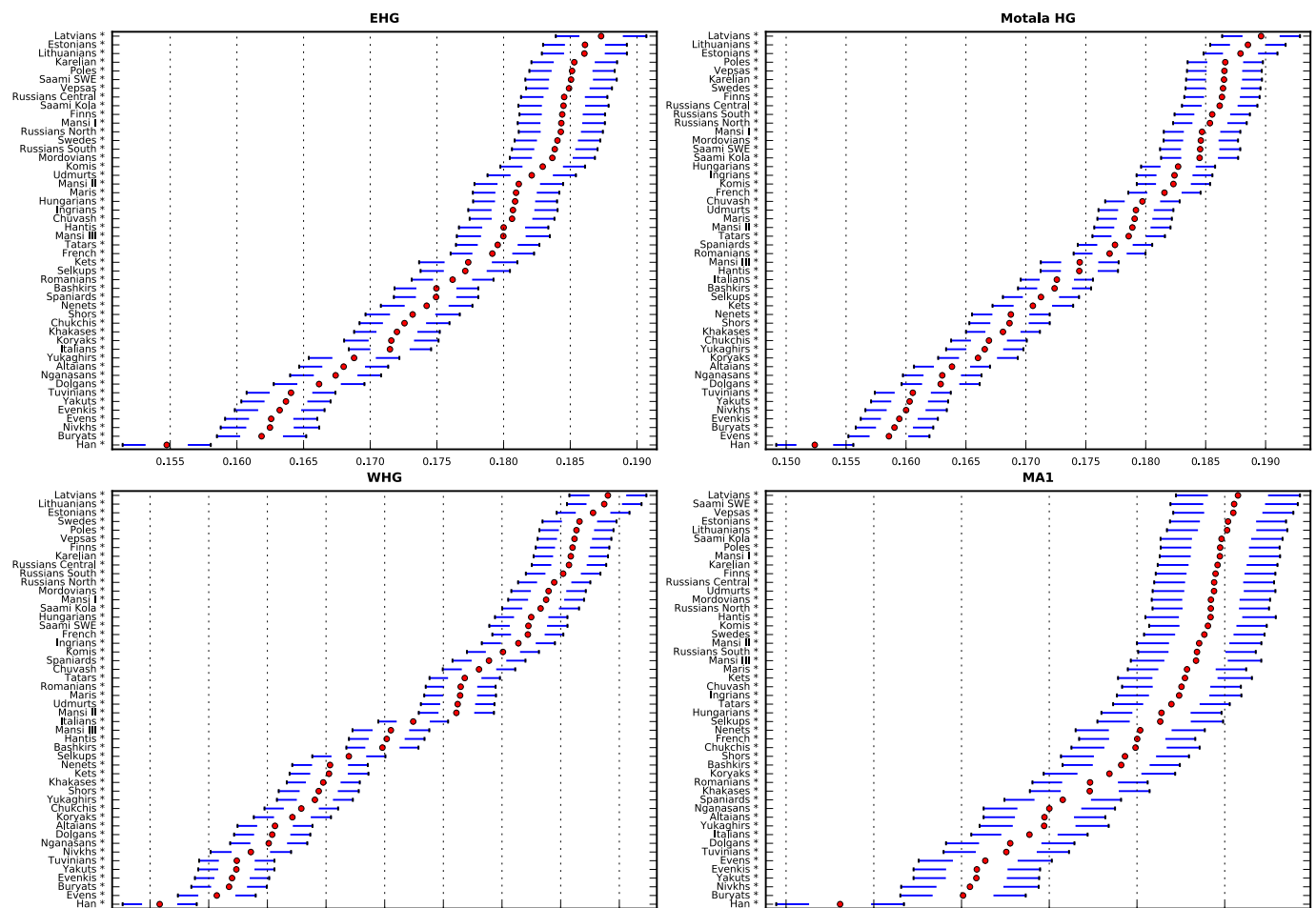




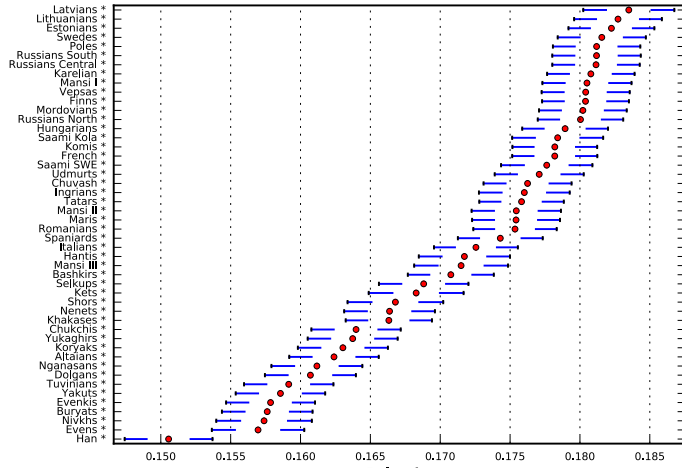




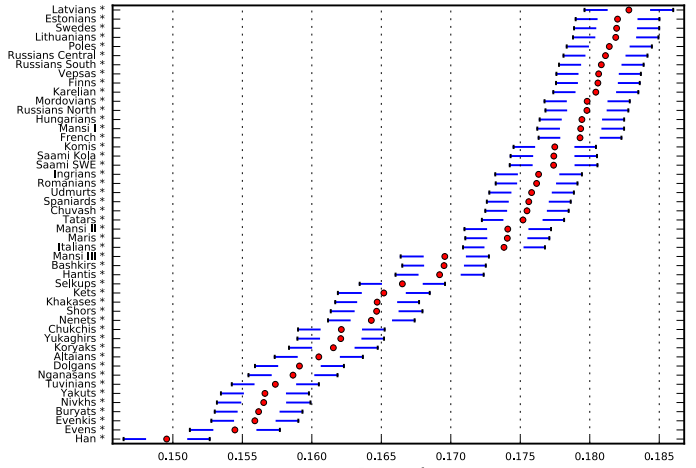
B.



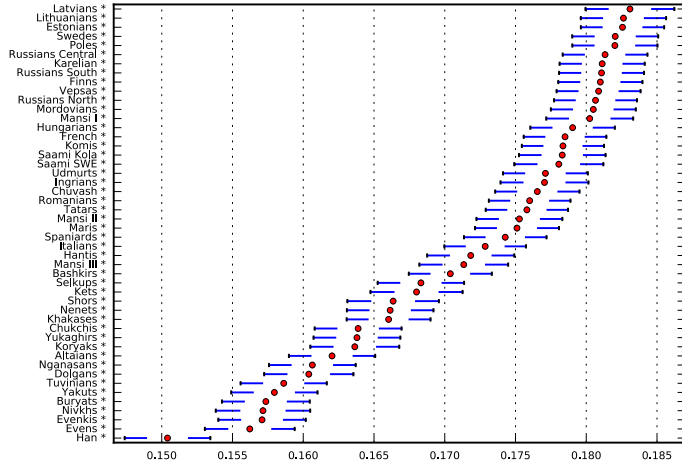
Afanasievo



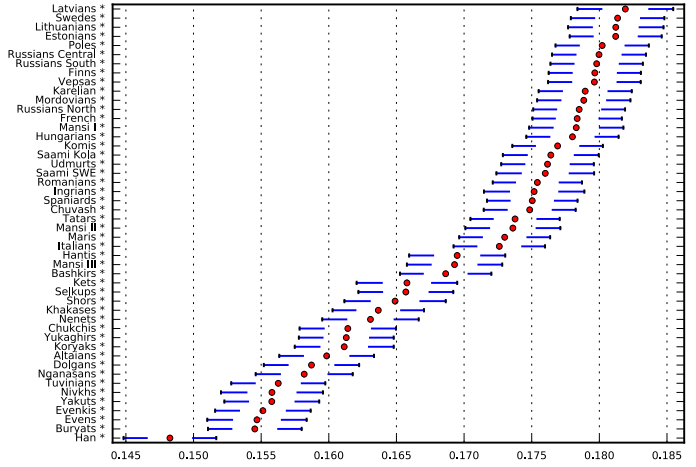
Andronovo



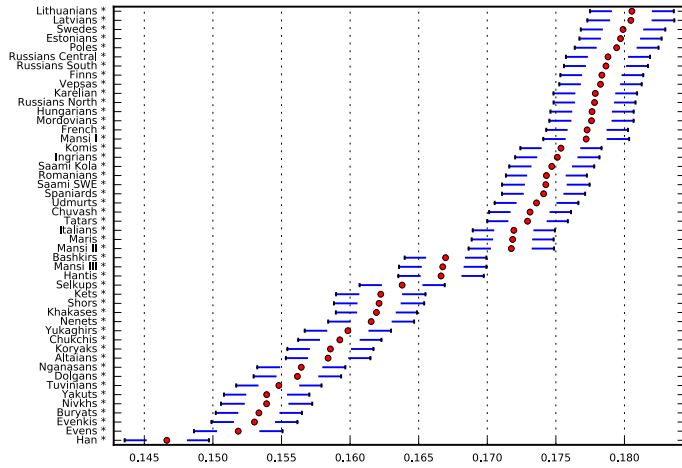
Poltavka



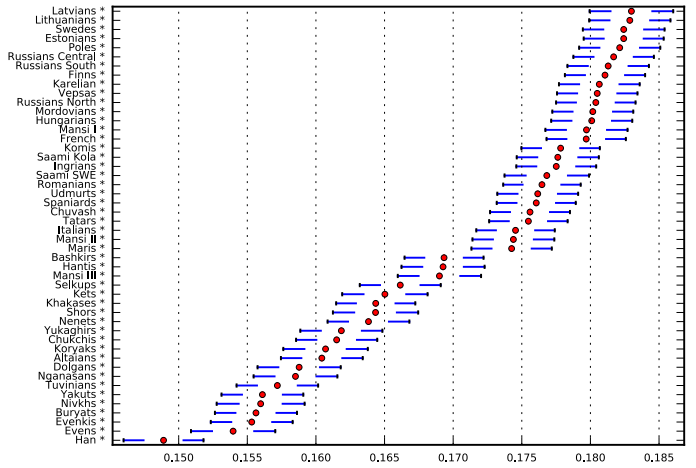
Potapovka



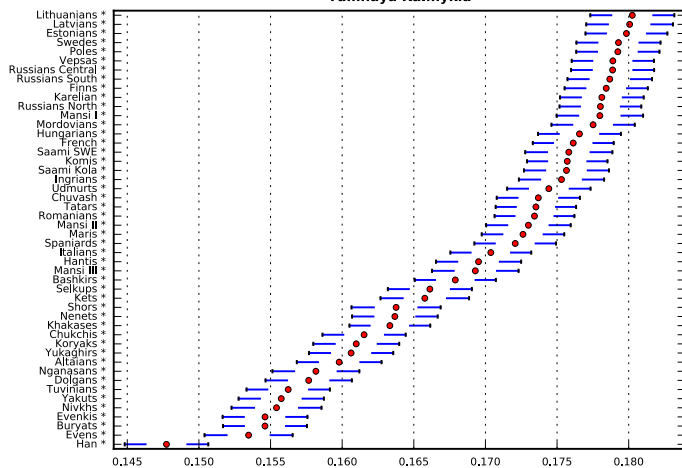
Sintashta



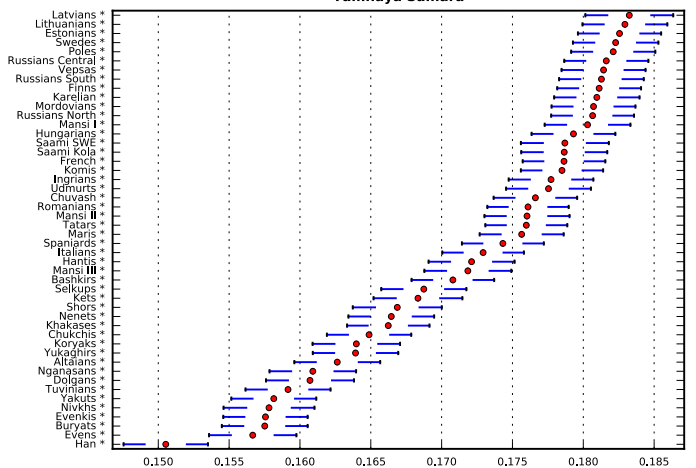
Srubnaya

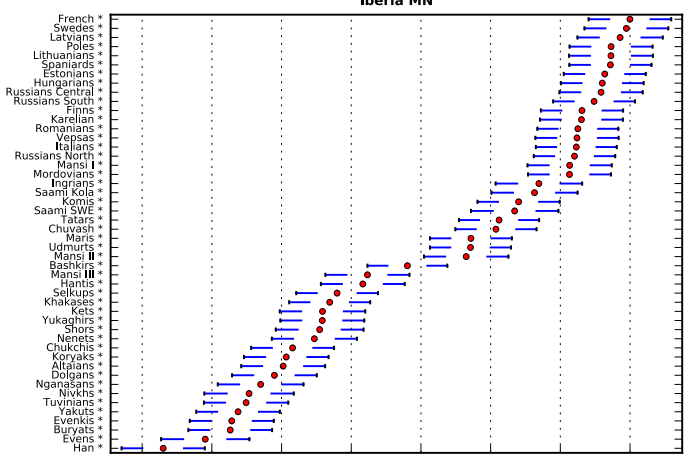
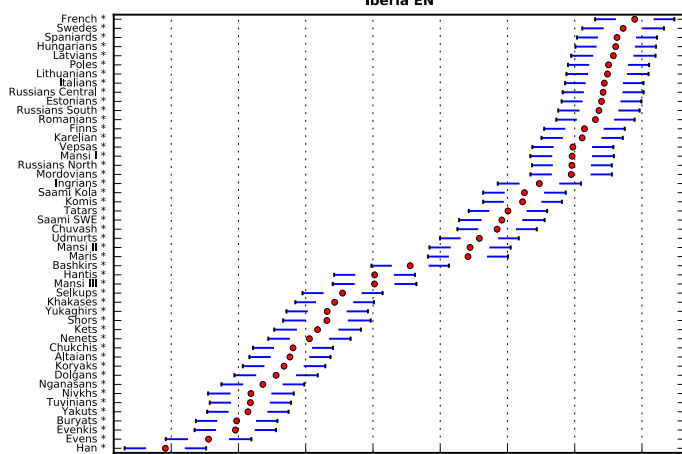
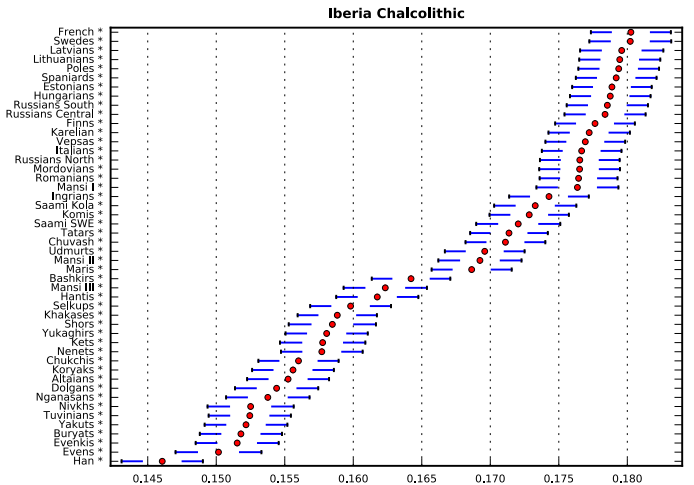
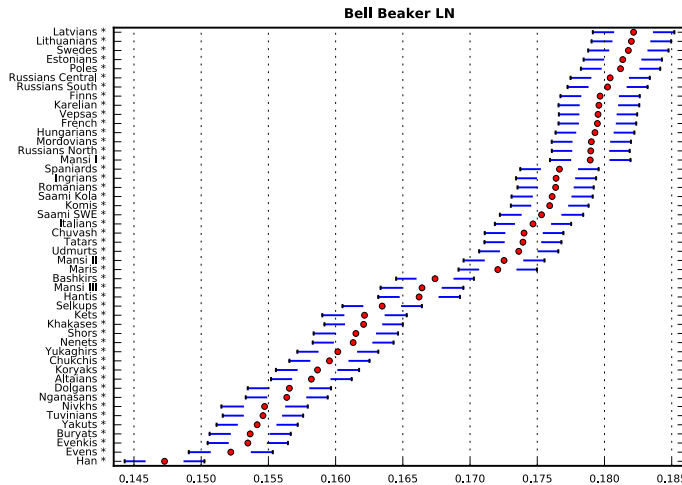
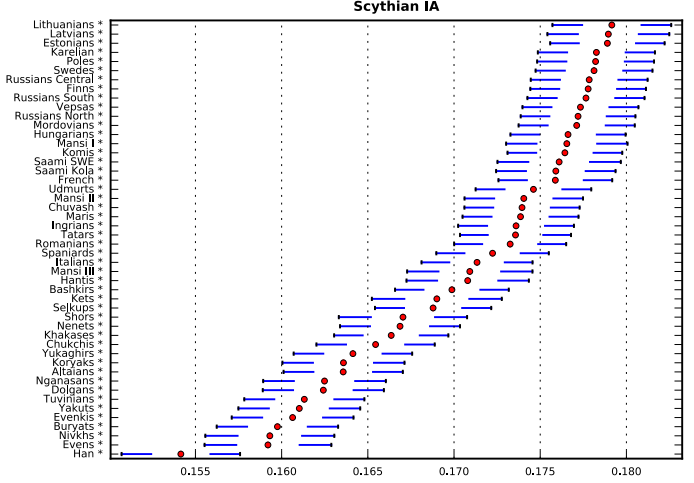
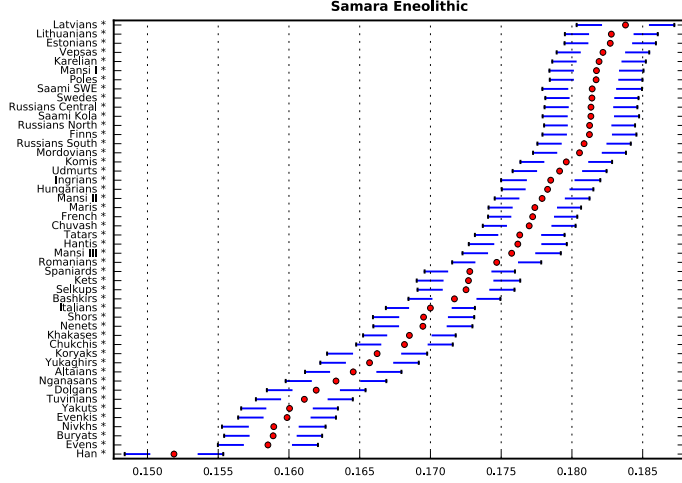
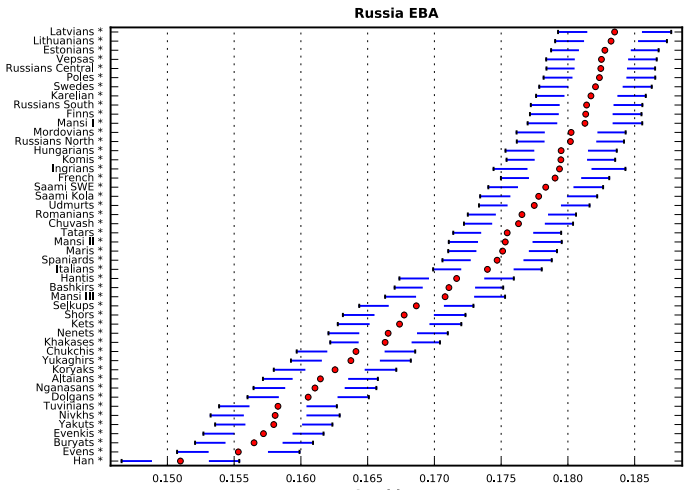
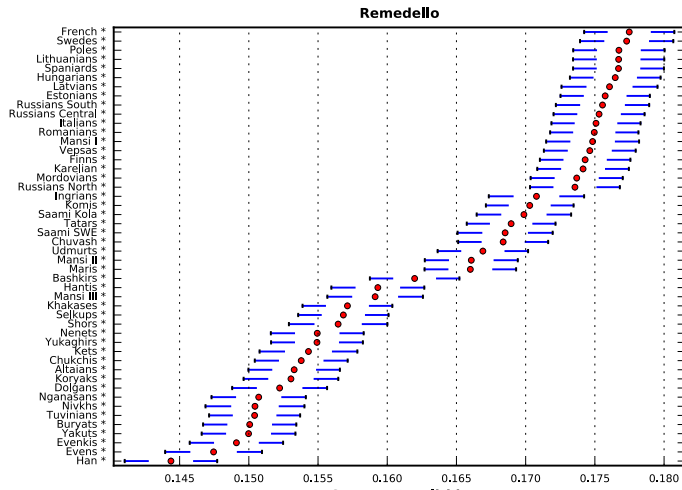


Yamnaya Kalmykia



Yamnaya Samara





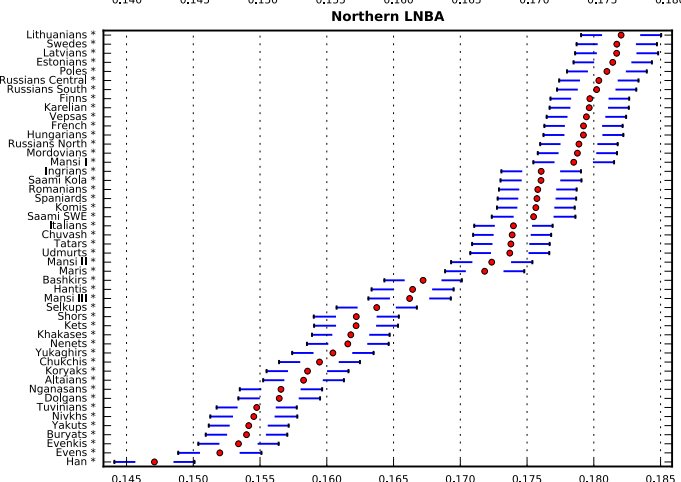
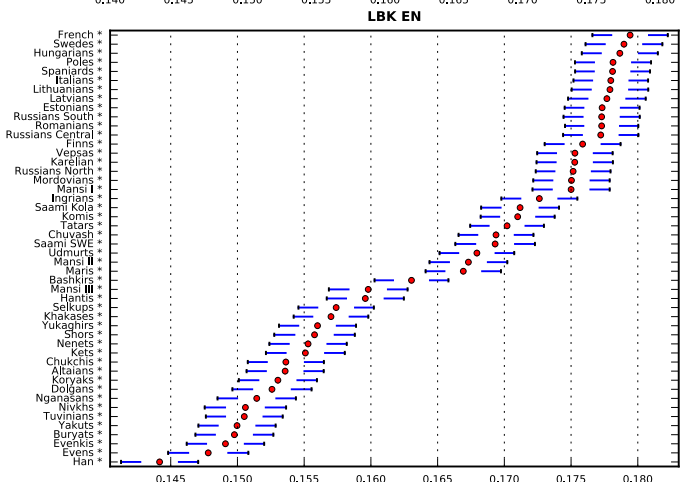
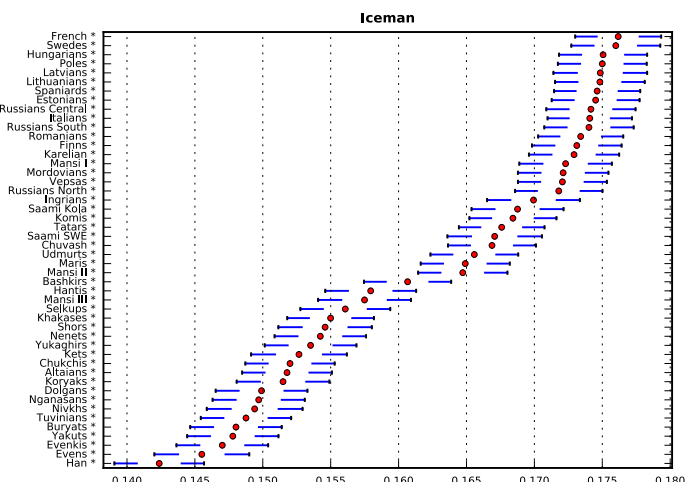
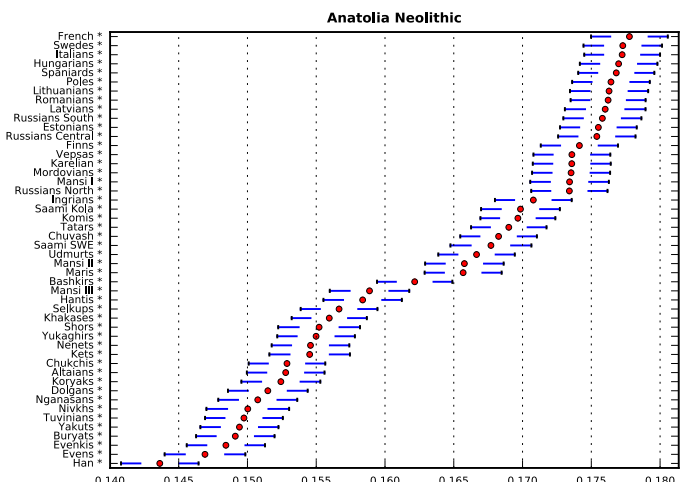
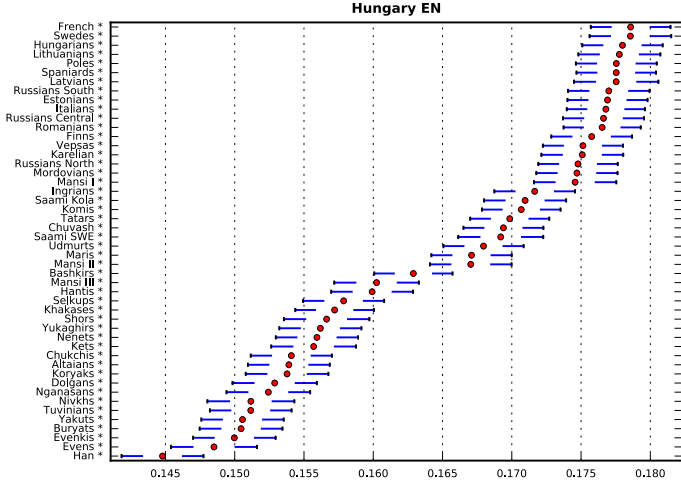
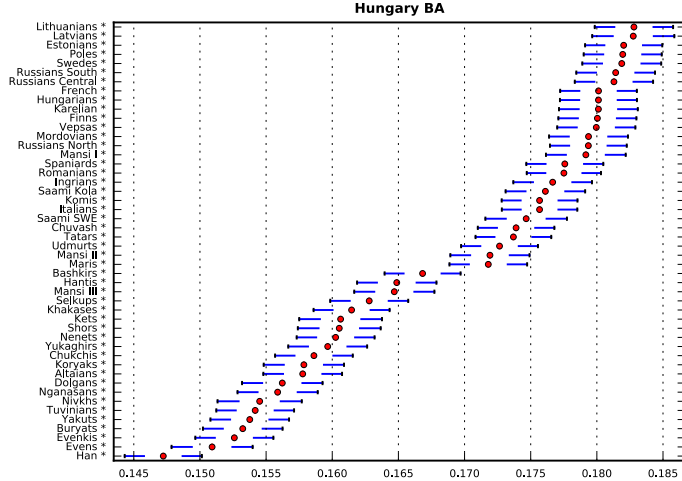
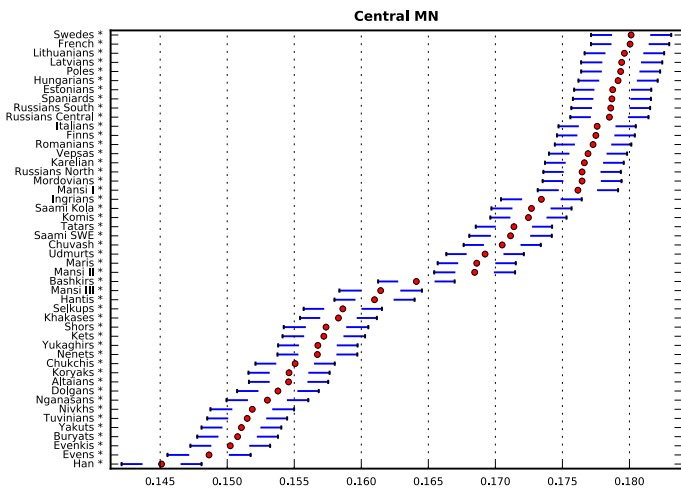
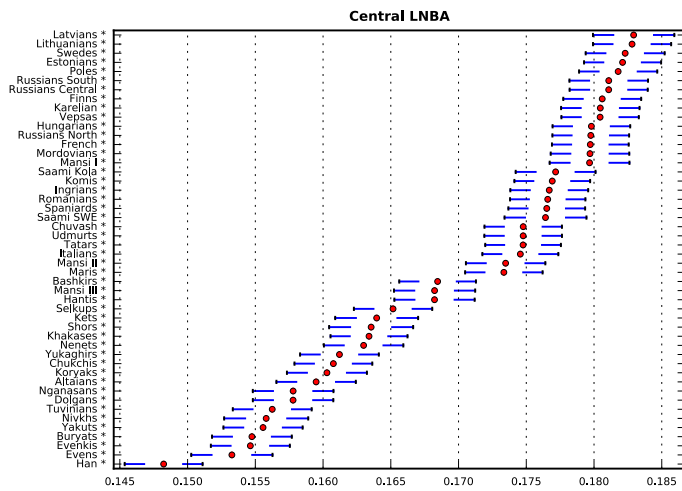
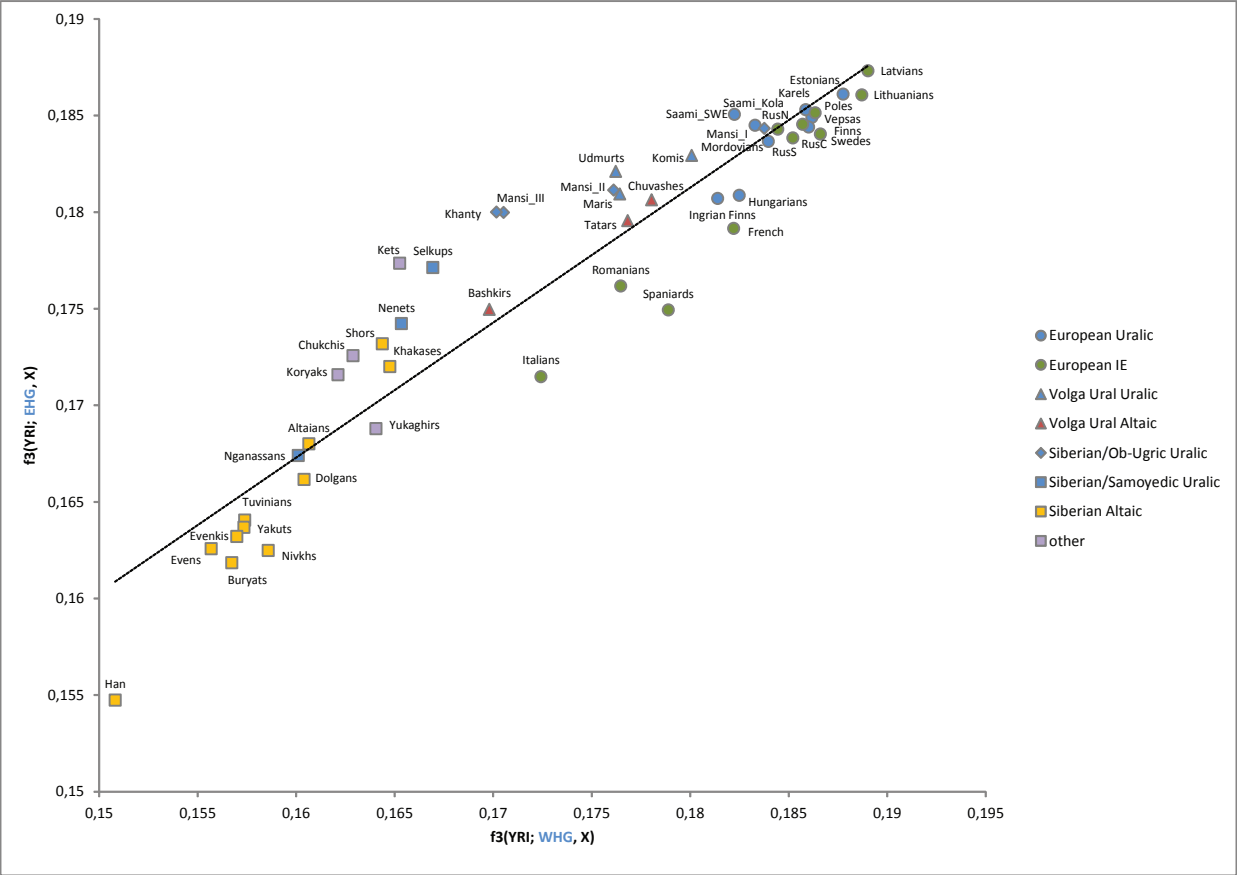
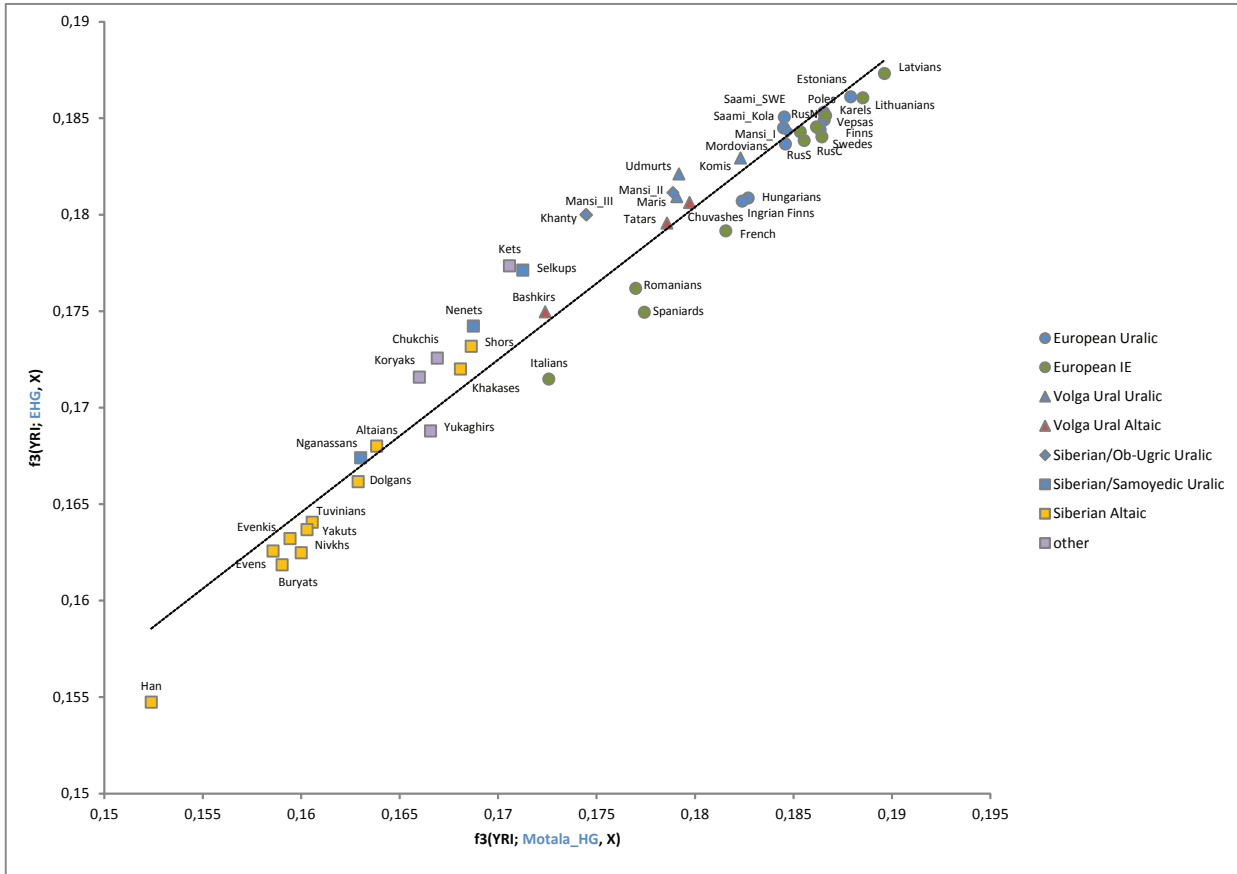


Figure S8. Results of outgroup f_3 test between modern samples and ancient genomes. The (modern reference, ancient reference; outgroup) test configuration was used, and Yoruba from Africa (YRI) was used and an outgroup. Each individual plot shows tests results between a reference given in the title of that plot, and all other reference populations (y-axis). For convenience, results for plotting were grouped using modern reference population (**A.**) or ancient reference cluster or genome (**B.**). Detailed results, including the number of SNPs used for each particular test are given in the Additional file 14: Table S13.

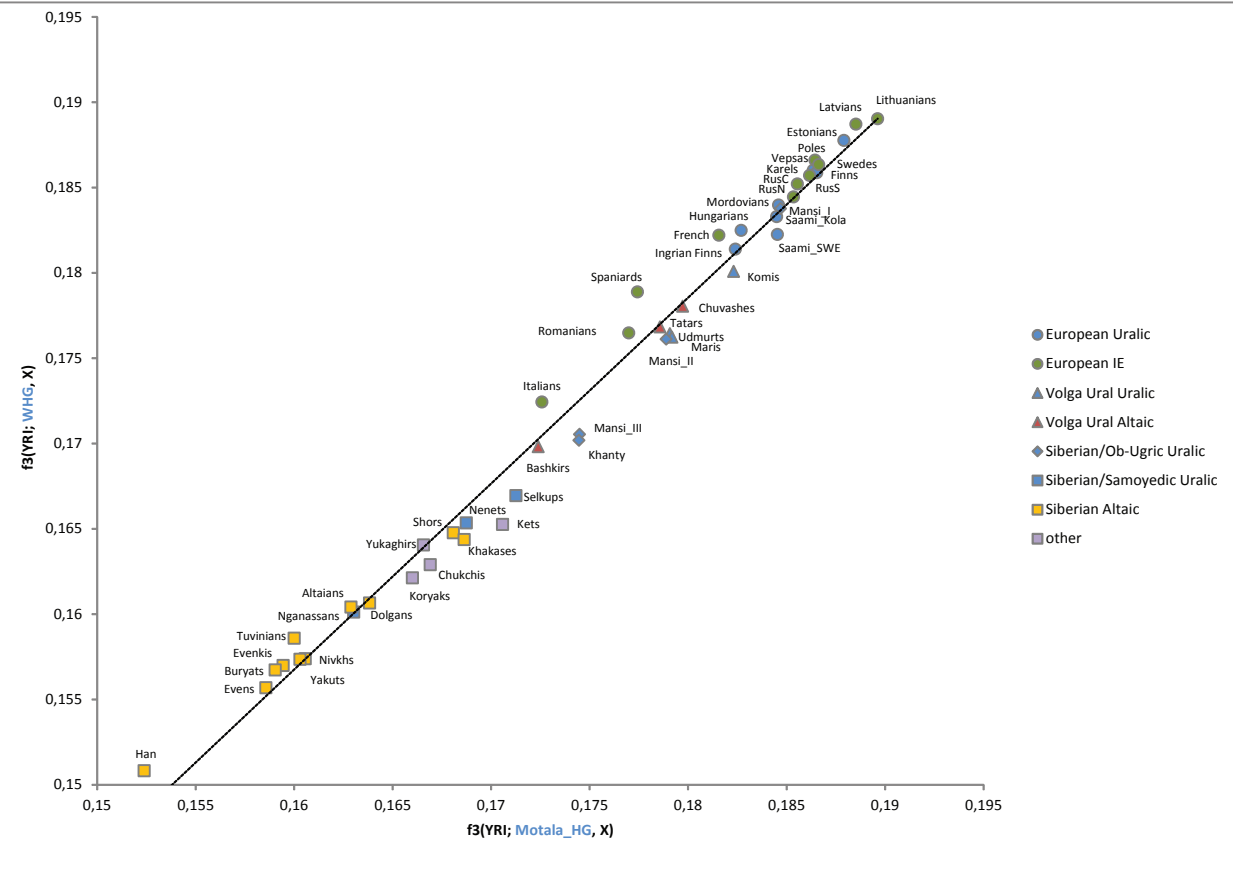
A.



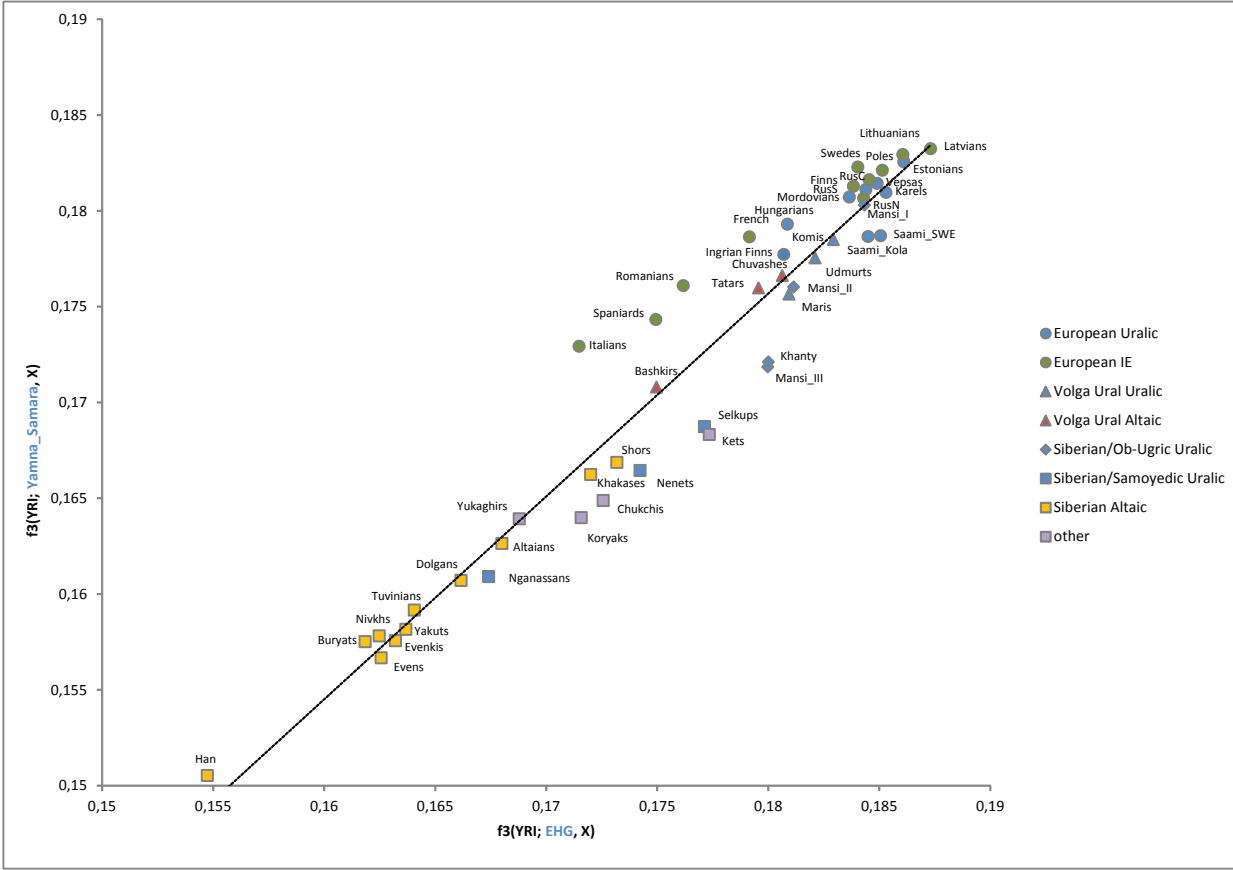
B.



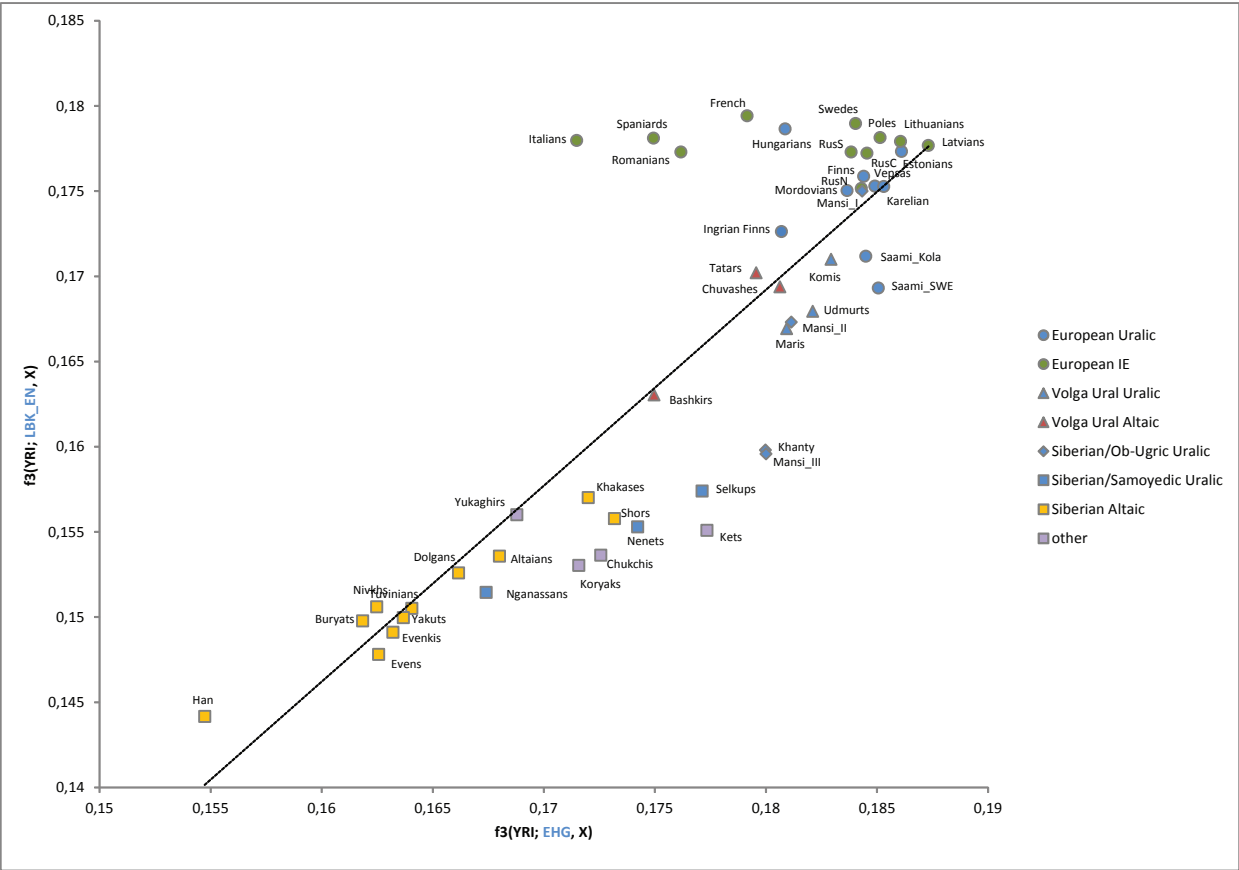
C.



D.



E.



F.

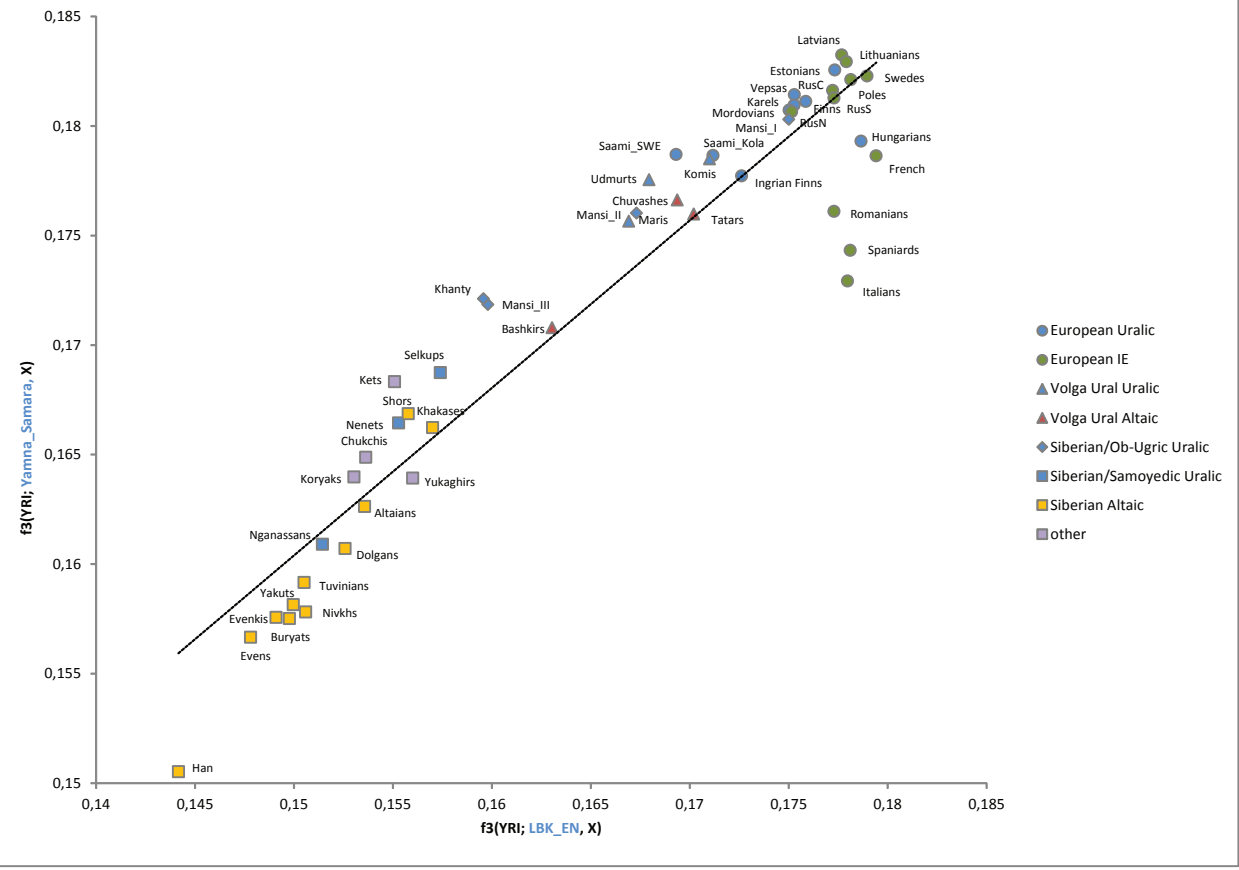


Figure S9. Outgroup f3-statistics' results in the form $f_3(\text{Yorubas}; \text{ancient Eurasian population, modern Eurasian population})$ plotted pairwise against each other. The information about the modern and ancient individuals is listed in Additional file 1: Table S1, underlying data of f3 statistics is from Additional file 14: Table S13. F3 of Eastern hunter gatherer (EHG) is plotted against f3 of **A.** western hunter gatherer (WHG), **B.** Scandinavian hunter gatherer (Motala_HG), **D.** Bronze Age steppe people (Yamna_Samara) and **E.** Central European early farmers (LBK_EN). **C.** F3 of WHG is plotted against f3 of Motala_HG and **F.** f3 of Yamna_Samara against f3 of LBK_EN.

A.

B.

Phylogenetic tree showing relationships between various taxa and their bootstrap values. The tree is rooted at R (62) and branches into Yoruus and A (62). A (36) branches into O and AA (41). O (13) branches into E and CH (41). E (40) branches into EU and WH (49). WH (47) branches into WHG. EU (84%) branches into EH (16%) and CH (62%). EH (38%) branches into EHI (5) and AAA (62%). CH (38%) branches into CH1 (117) and CHG (120). CH1 (38%) branches into CH2 (10%) and CH3 (90%). CH2 (7%) branches into LB (10%) and YML (22%). LB (37) branches into LBK_EN (37) and YML (81). YML (22%) branches into CW (78%) and EstCWC (162). CW (87%) branches into CWC (29%) and CCE (92%). CWC (87%) branches into CCE (29%) and CCA (92%). CCA (92%) branches into Finns (7).

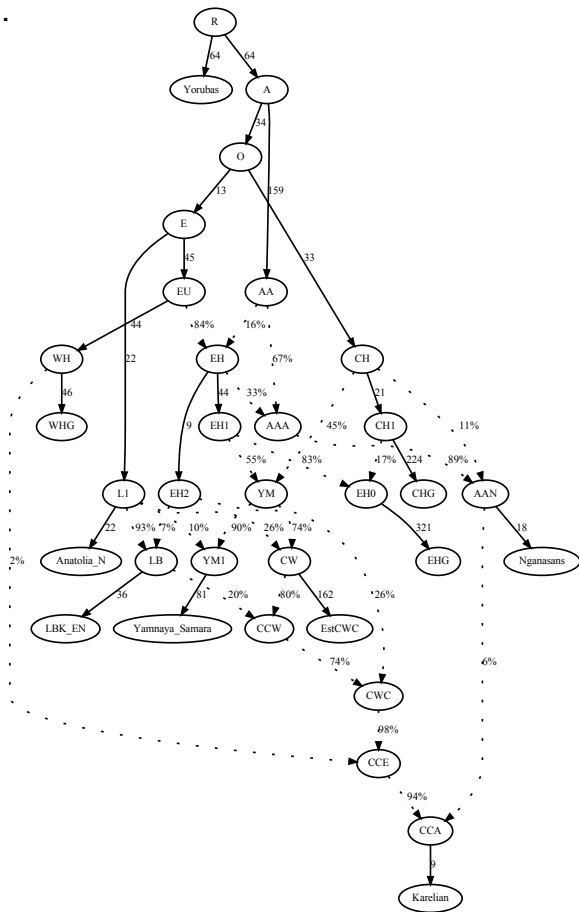
C.

Phylogenetic tree showing relationships between various populations. The tree is rooted at 'R' and branches out to include populations like Yonubas, A, O, E, EU, AA, WH, WHG, EH, EH1, AAA, CH, CH1, AAN, Ngnasans, LI, LBK, LBK_EN, Ymmya_Samara, CCW, EstCWC, CWC, CCE, CCA, and Lithuanians. Bootstrap values are indicated at the nodes.

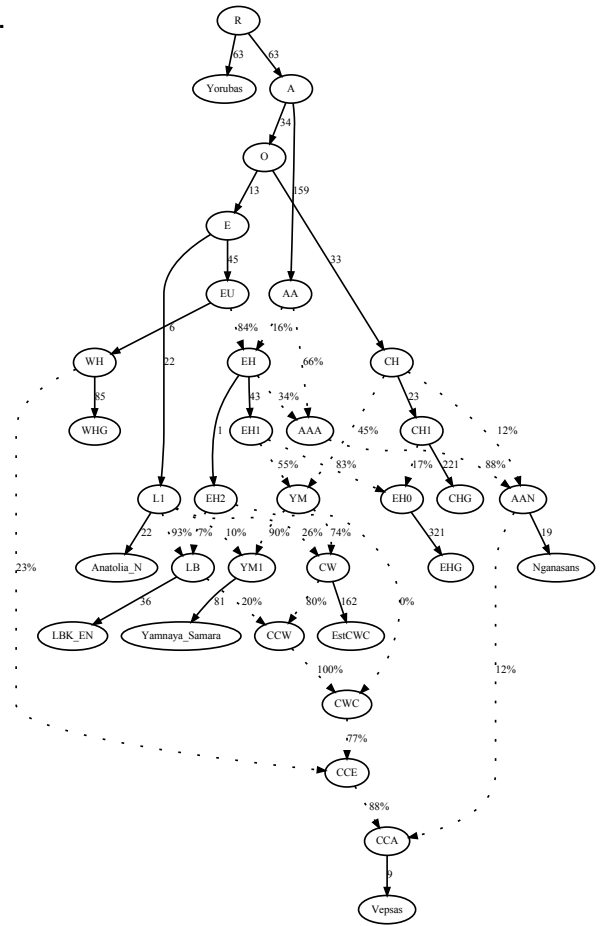
D.

Phylogenetic tree of the R gene in the genus *A.* The tree shows relationships between various accessions, with bootstrap values indicated at the nodes. The tree is rooted at R, which branches into Vorubas and A. A branches into O and AA. O branches into E and CH. E branches into EU and WH. EU branches into L1 and EH2. WH branches into WHG. L1 branches into Anatolia_N and LB. LB branches into LBK_EN and Yamnaya_Samara. EH2 branches into YM1 and CW. YM1 branches into CCW and EstCWC. CW branches into CWC and CCE. CWC branches into CCE. CCE branches into CCA. CCA branches into Latvians. CH branches into CH1 and EH0. CH1 branches into CHG and AAN. EH0 branches into EH1 and EH2. EH1 branches into YM. YM branches into CW. CW branches into CWC. CWC branches into CCE. CCE branches into CCA. CCA branches into Latvians.

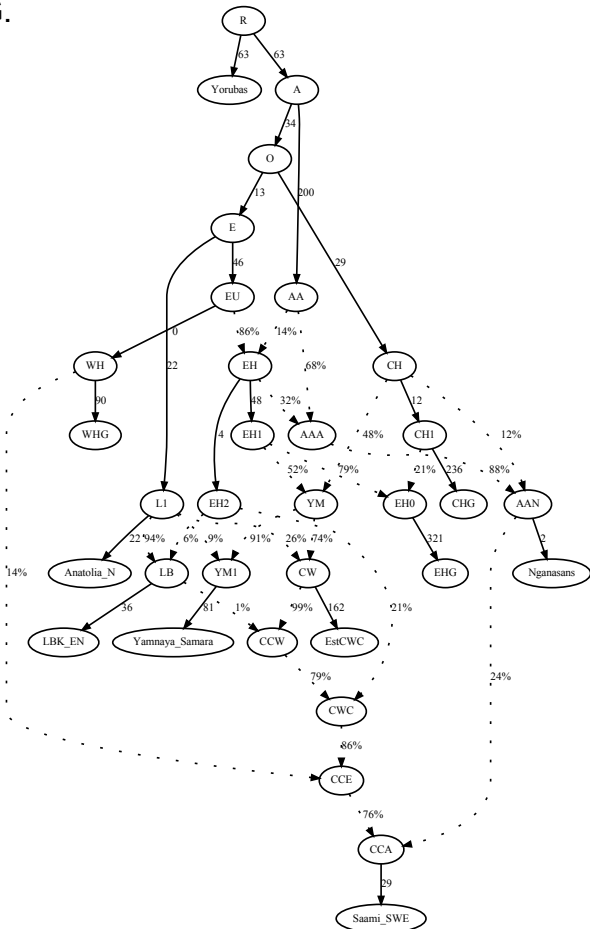
E.



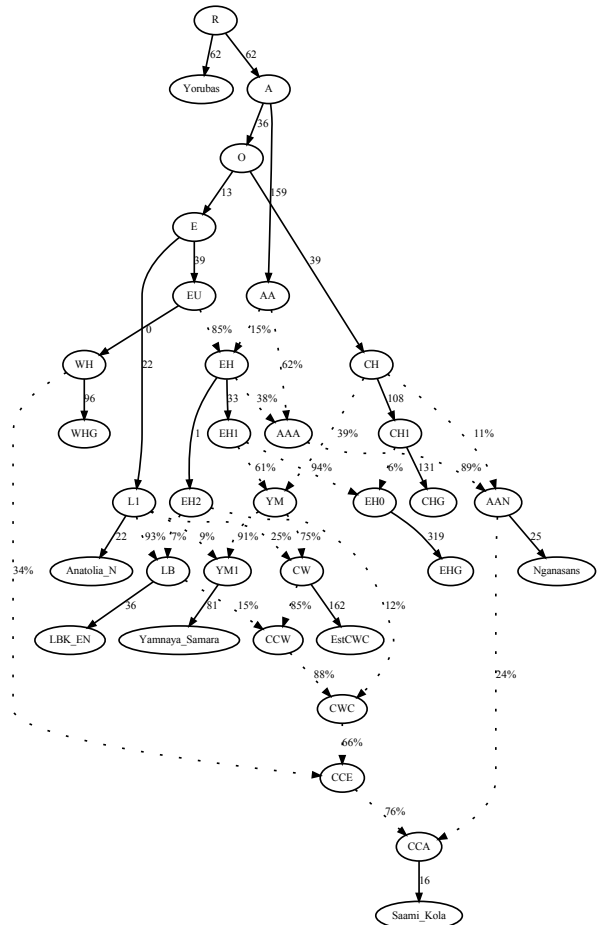
F.



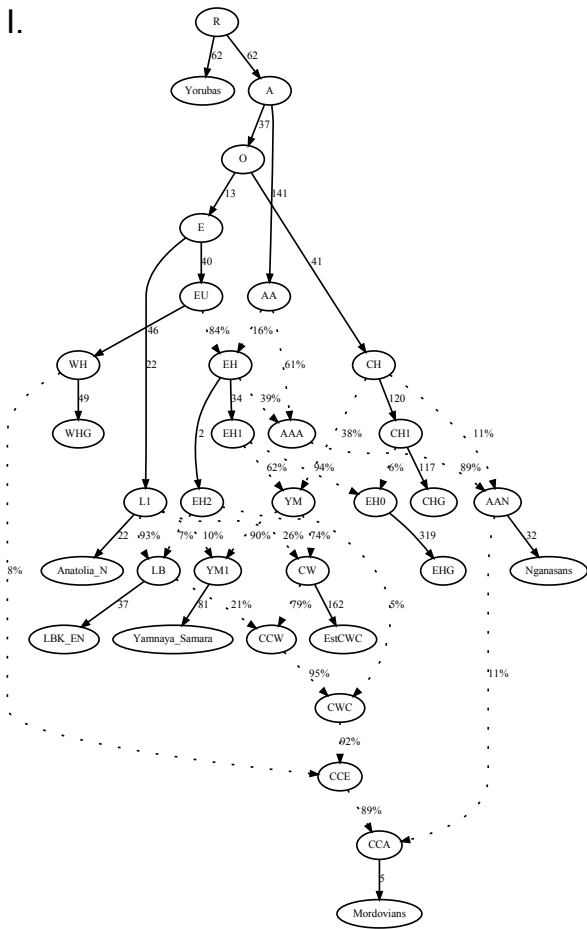
G.



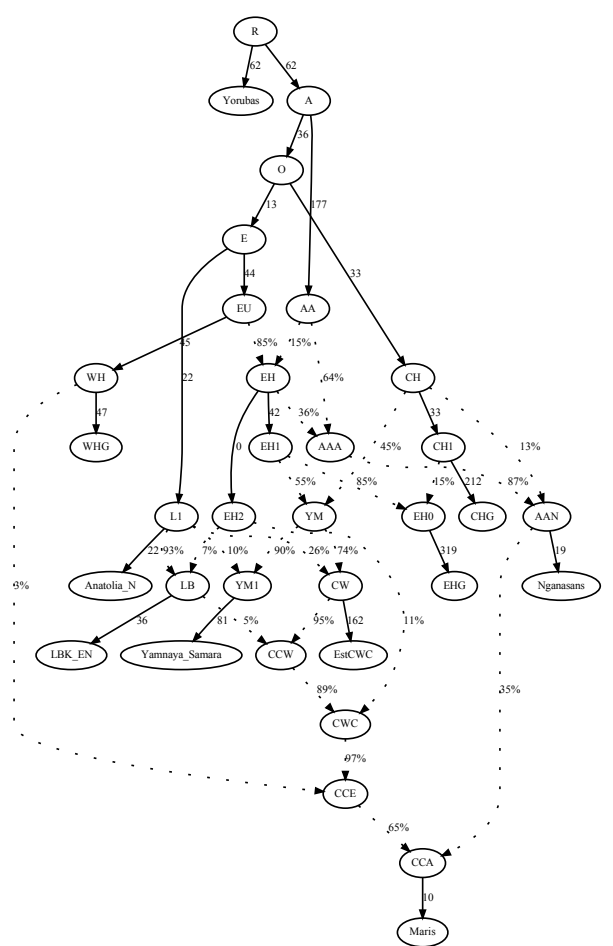
H.



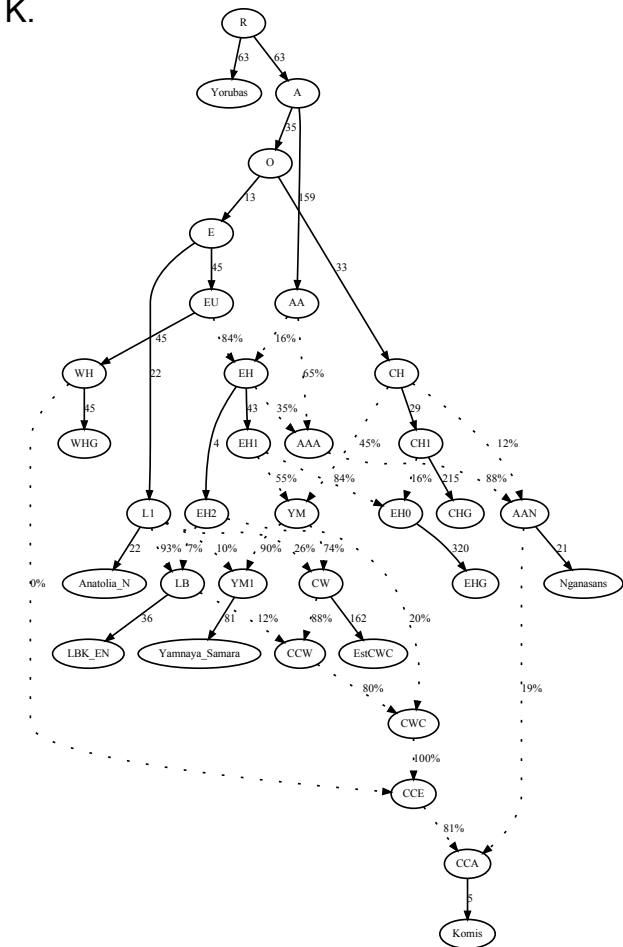
I.



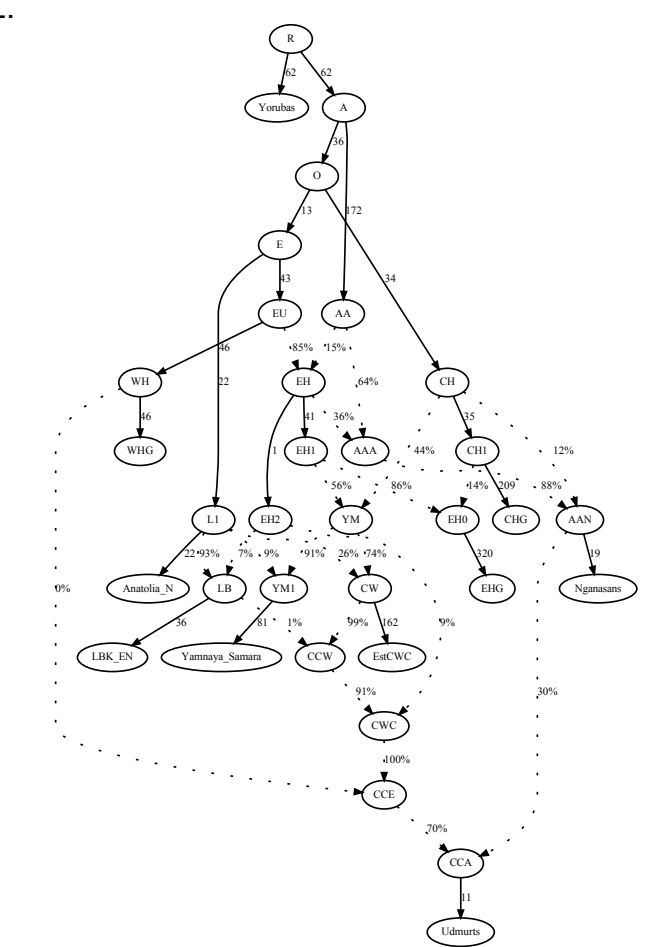
J.



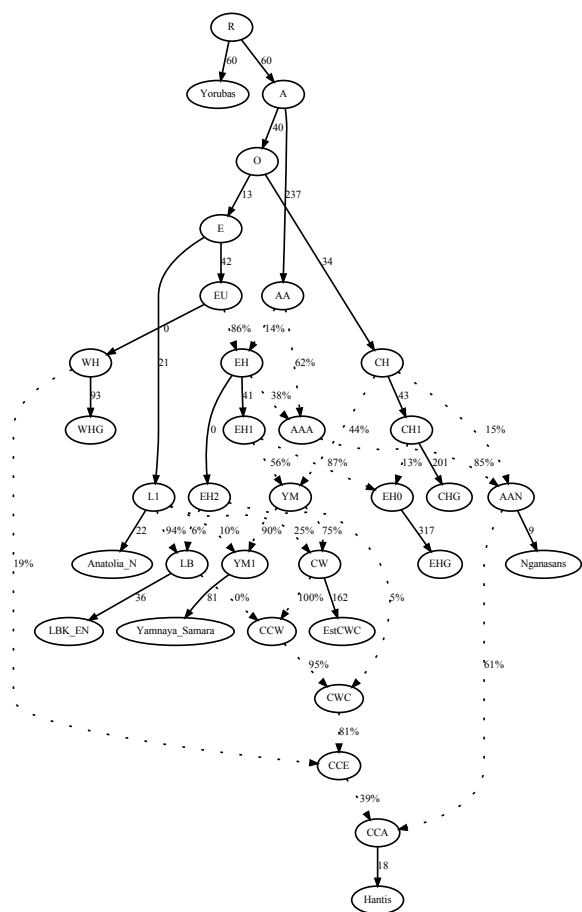
K.



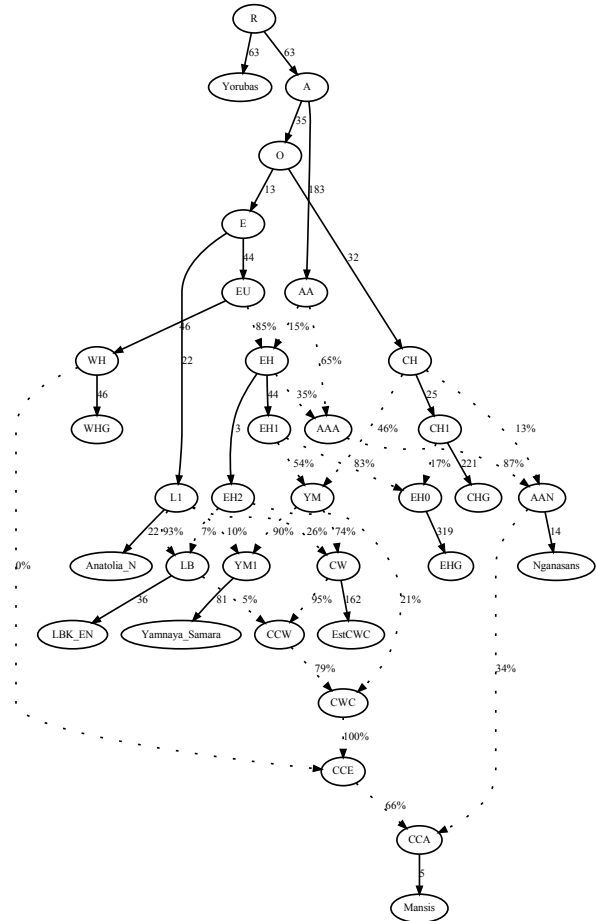
L.



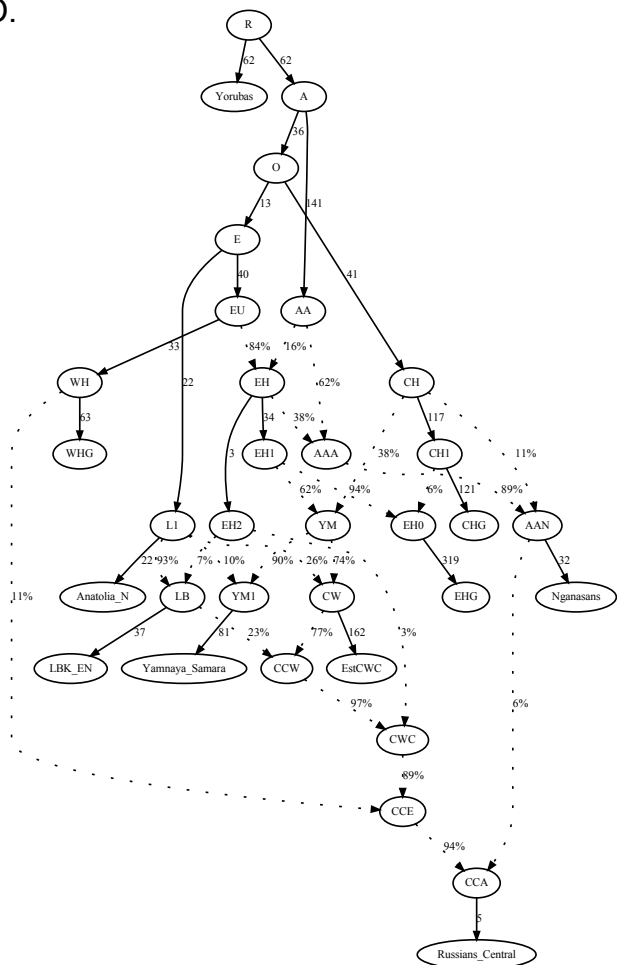
M.



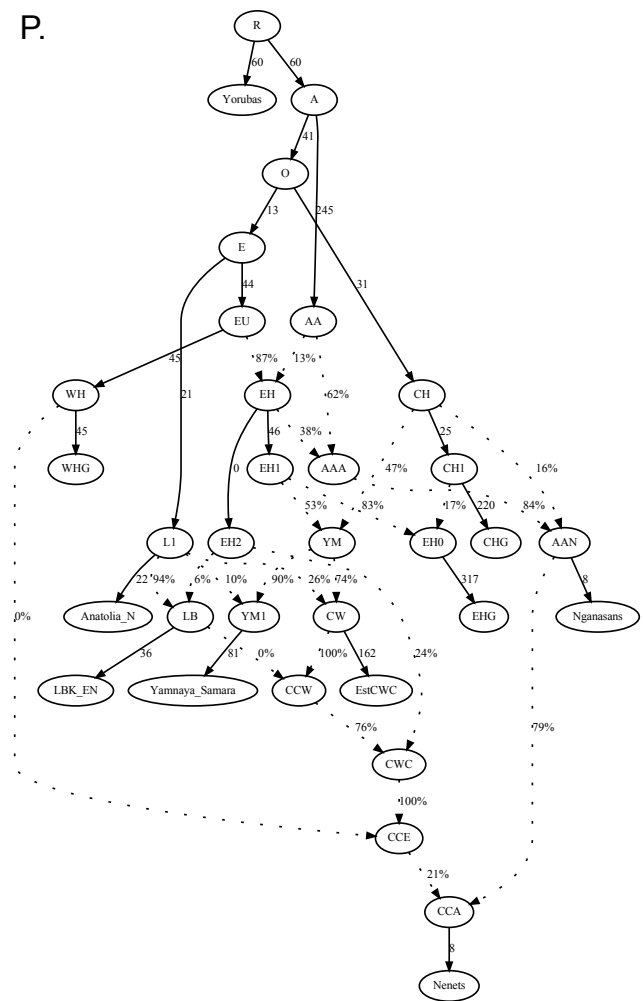
N.



O.



P.



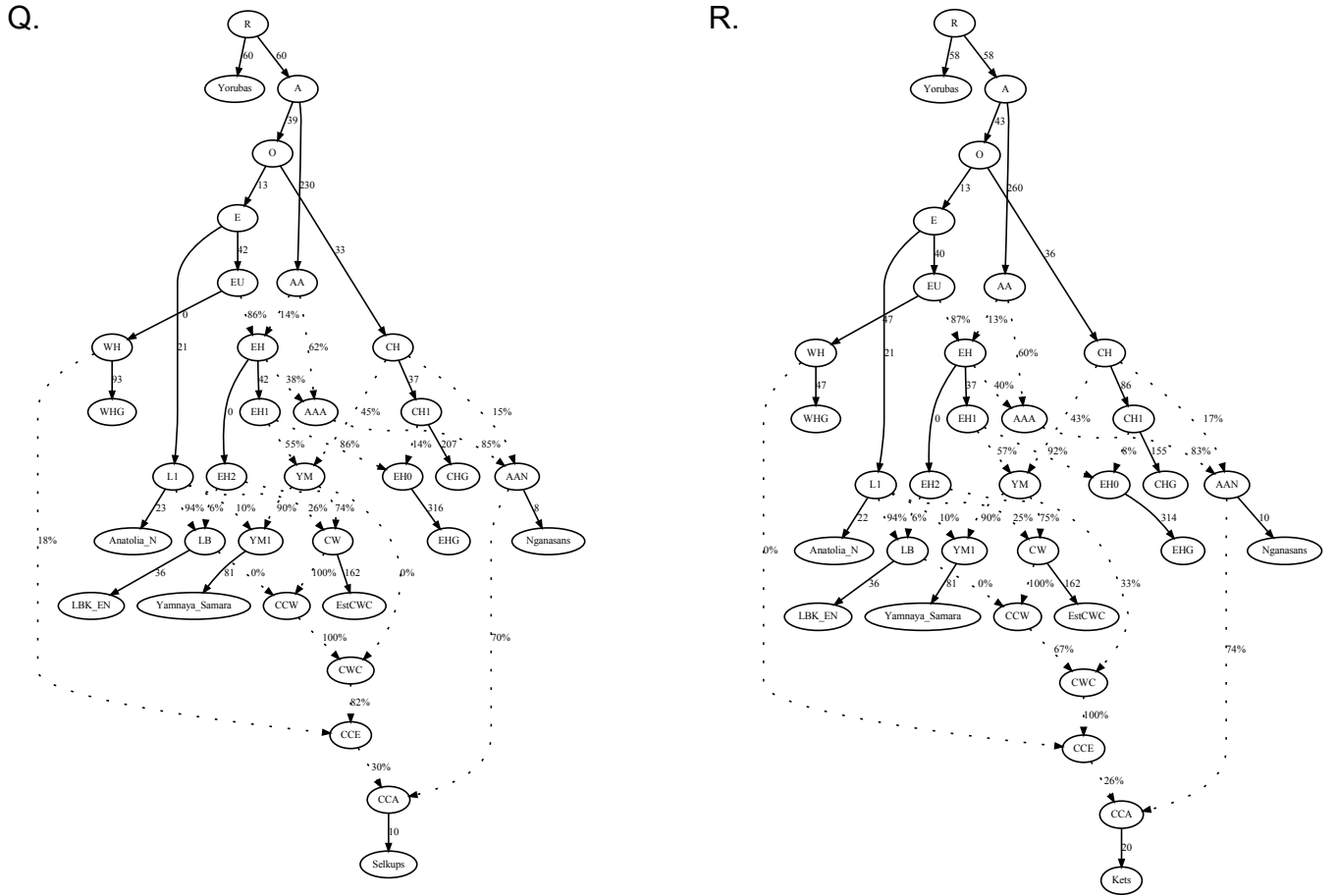


Figure S10. Admixture graphs of the demographic model with several admixture events (with shown % proportions), that fits tested data. **A.-B., E.-N.** and **P.-Q.** – Uralic speaking populations; **C.-D., O.** and **R.** – non-Uralic speaking populations.

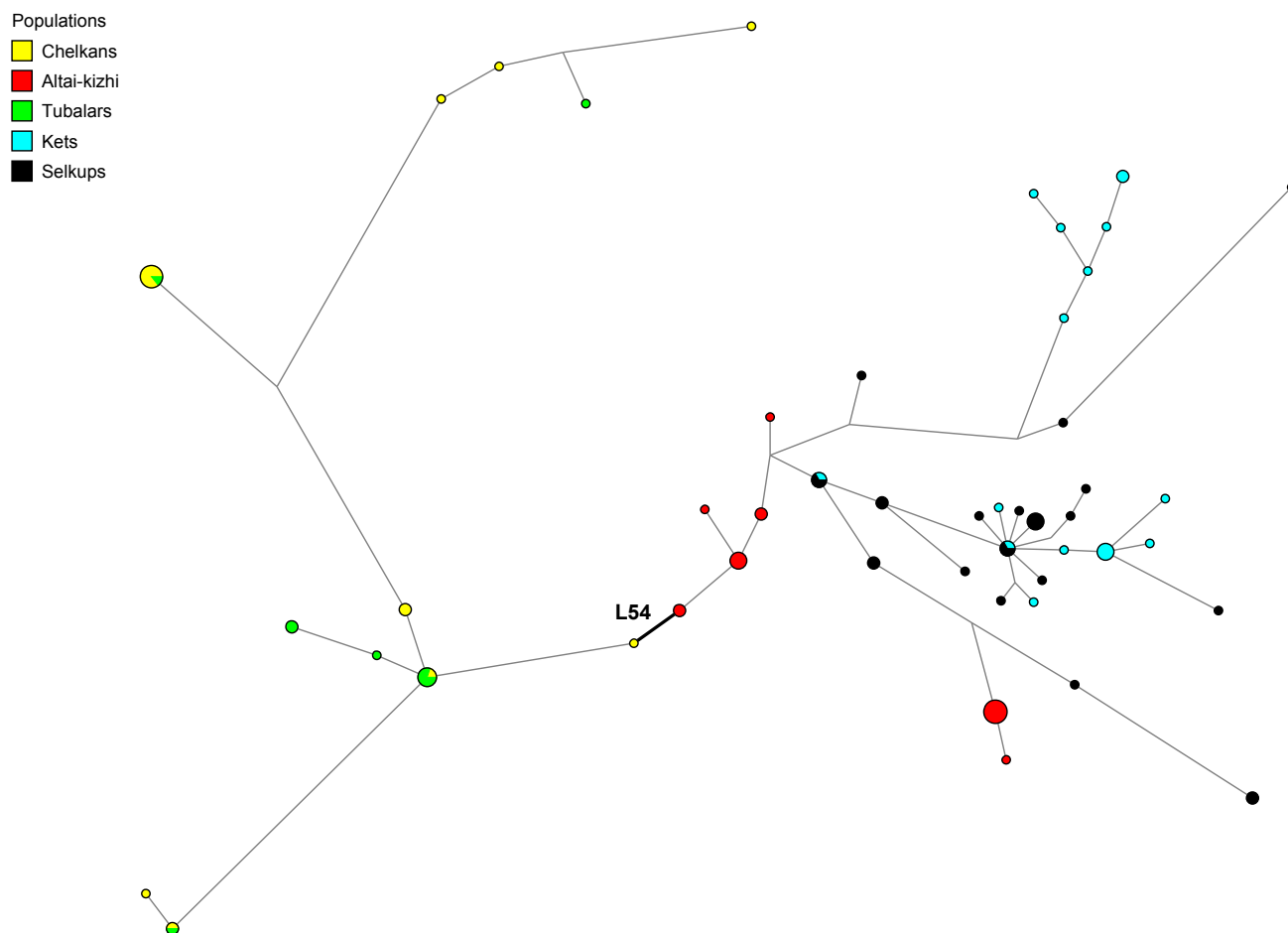


Figure S11. Phylogenetic network of Y chromosome haplogroup Q constructed from 17 Y chromosomal STRs with Network software version 2.0.0.1 (Fluxus-Engineering), applying median joining algorithm with a weight value of 90 for sub-clade defining SNP marker and a weight value of 10 for STR markers. A total of 88 samples (see legend for population data and Additional file 17: Table S16 for STR haplotypes) belonging to subhaplogroup Q1a3a were analyzed using the Y-Filer Kit and run on the ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems). We tested all hg Q Y chromosomes of our sample of Selkups (n=17) and Kets (n=14) and found that all of these belong to the L54 defined branch of Q1a3a1, shown to be very common among Altai-kizhi populations of South Siberia.

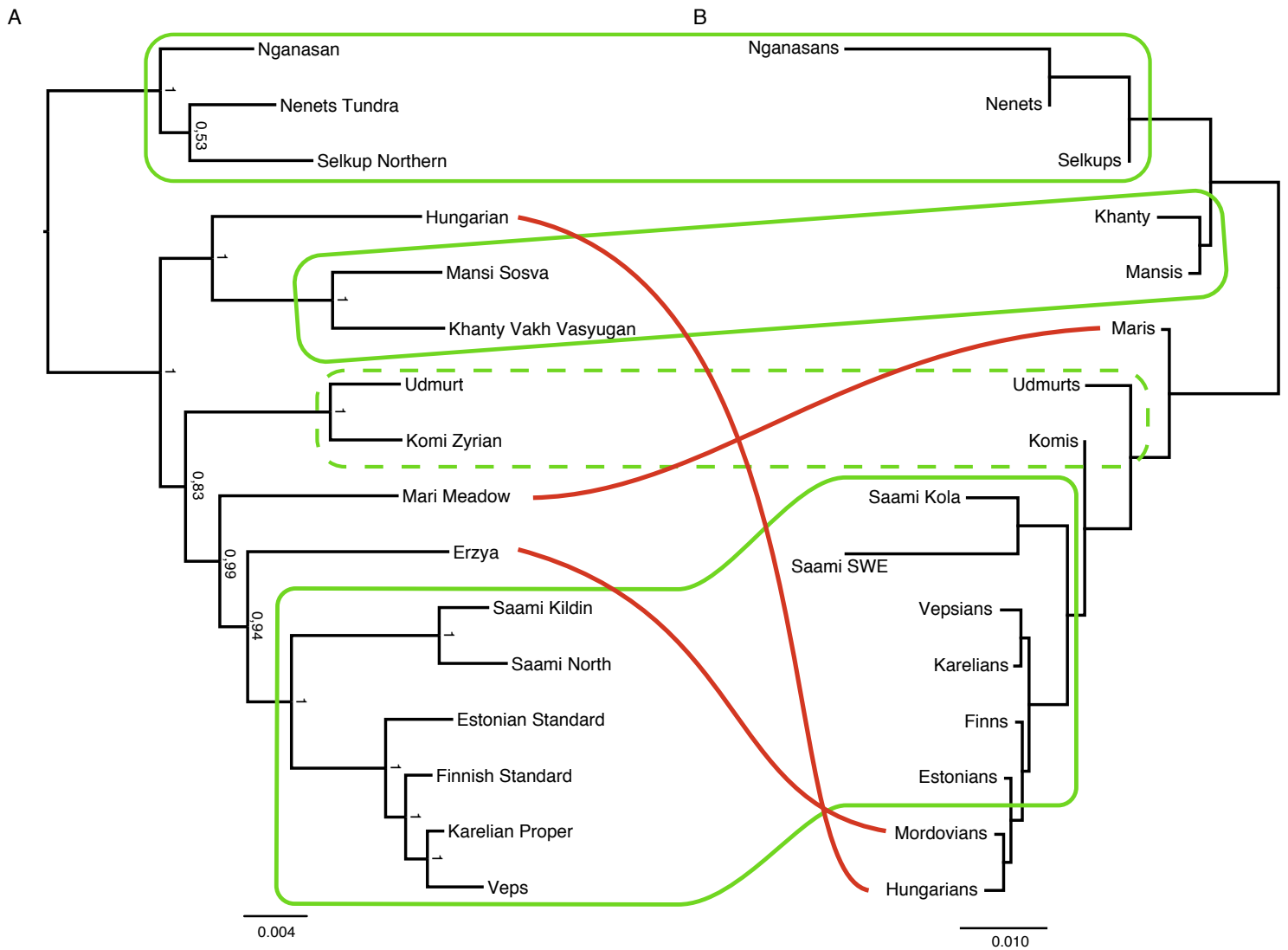


Figure S12. A. Quantitative phylogeny of 16 Uralic languages based on 226 basic vocabulary meanings coded by their cognacy relationships. Phylogeny is made with MrBayes by following the settings in Syrjänen et al. (2013). Shedding more light on language classification using basic vocabularies and phylogenetic methods. A case study of Uralic. *Diachronica* 30, 323–352. **B.** The evolutionary history of the Uralic speaking populations as inferred from pairwise F_{st} distance matrix using the Neighbor-Joining method [1]. The optimal tree with the sum of branch length = 0.13079639 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary analyses were conducted in MEGA7 [2]. Green and red cartoon overlaid on panels A and B highlights the concordances and discordances, respectively, of the two trees.

1. Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
2. Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874.

