

Original Research

GRUPS-rs, a high-performance ancient DNA genetic relatedness estimation software relying on pedigree simulations

Maël Lefeuvre 1,*, Michael David Martin 2, Flora Jay 3, Marie-Claude Marsolier 1,4, Céline Bon 1,*

- 1 UMR 7206 Eco-Anthropologie (EA), CNRS, Muséum National d'Histoire Naturelle, Université Paris Cité, 75116 Paris, France; Email: marie-claude.marsolier@mnhn.fr
- 2 Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway; Email: mike.martin@ntnu.no
- 3 UMR9015 - Laboratoire Interdisciplinaire des Sciences du Numérique (LISN), CNRS, INRIA, Université Paris-Saclay, 91400 Orsay, France; Email: flora.jay@lri.fr
- 4 UMR 9198 - Institut de Biologie Intégrative de la Cellule (I2BC), CNRS, CEA, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

* **Correspondence:** Maël Lefeuvre; Email: mael.lefeuvre@mnhn.fr; Céline Bon; Email: celine.bon@mnhn.fr.

Supplementary Materials

Contents

- 1 Figure S1. Alternative template pedigrees on the Koszyce dataset.
- 2 Figure S2. GRUPS-rs accuracy results on simulated dataset.
- 3 Figure S3. Hazleton-North samples pairwise SNP overlap.
- 4 Figure S4. Per-class performance metrics of GRUPS-rs and the Kennett method.
- 5 Figure S5. R-coefficient estimates of GRUPS-rs against the Kennett method.
- 6 Figure S6. Pedigree simulation results for the Koszyce dataset.

1 Figure S1: Alternative template pedigrees on the Koszyce dataset

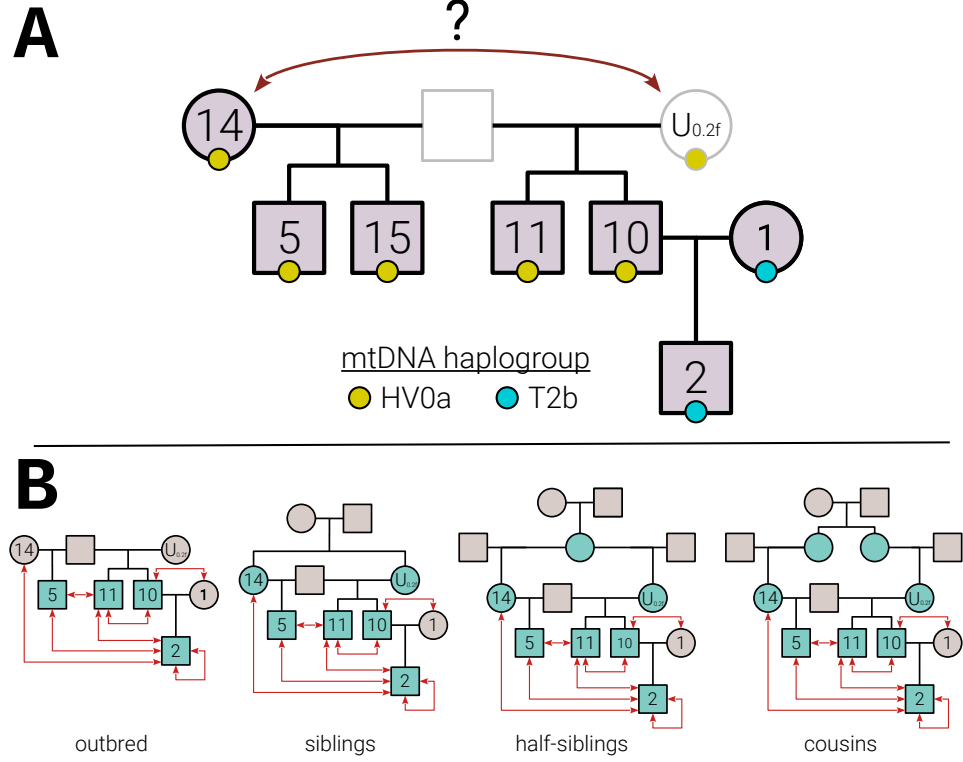


Figure S1: Diagram of the four template pedigrees used as input during the kinship analysis of individuals K[1, 2, 5, 10, 11, 14, 15] from the Koszyce dataset. Each pedigree found in **B** describes an alternative hypothesis regarding the relationship between K14 and U0.2f (from left to right: outbred, siblings, half-siblings, cousins). Grey and blue colored shapes respectively represent founder individuals sampled from the 1000G dataset, and simulated descendants. Red arrows represent the relationships which were investigated by GRUPS-rs within each run. **A**: family tree representing the known genetic ties found between the Koszyce individuals reanalyzed in this study — adapted from (Schroeder *et al.*, 2019). Colored circles represent the determined mitochondrial haplogroup of each individual. Note that every male individual carries the same Y-chromosome haplogroup (I2a-L801). The burgundy arrow highlights the unknown relationship between K14 and U0.2f

2 Figure S2: GRUPS-rs accuracy results on simulated dataset

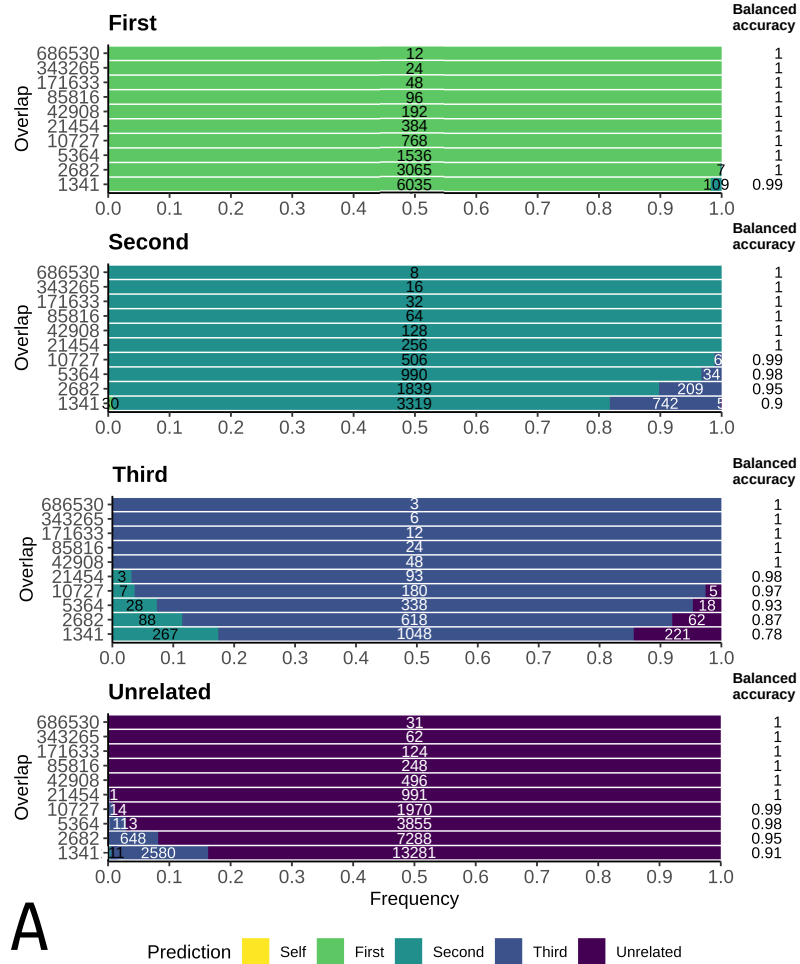


Figure S2: Per class estimates of GRUPS-rs' classification accuracy according to the average pairwise SNP overlap. **A**: Per-class estimates of GRUPS-rs' classification accuracy according to the average number of pairwise overlapping SNPs. Colored bars represent the class predicted by GRUPS-rs. The width of each bar shows the relative frequency of each prediction, while numbers at the centers of each bar amount to the numbers of predictions for that given class. Estimated balanced accuracies for each condition are displayed on the right ($((TP/(TP + FN)) + (TN/(TN + FP)))/2$). **B**: Overall prediction accuracy, for each condition of average pairwise SNP overlap.

3 Figure S3: Hazleton-North samples pairwise SNP overlap

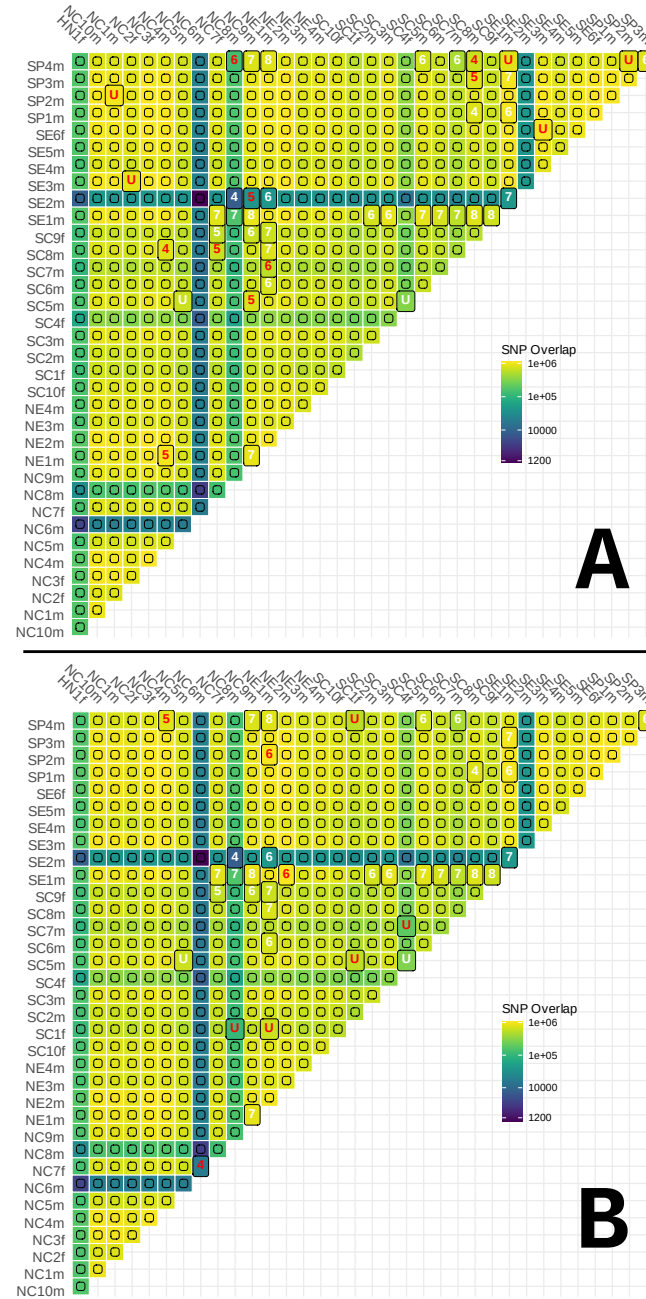


Figure S3: SNP overlap of all Hazleton-North samples against the kinship classification discrepancies of (A) GRUPS-rs and (B) the Kennett method

4 Figure S4: Per-class performance metrics of GRUPS-rs and the Kennett method

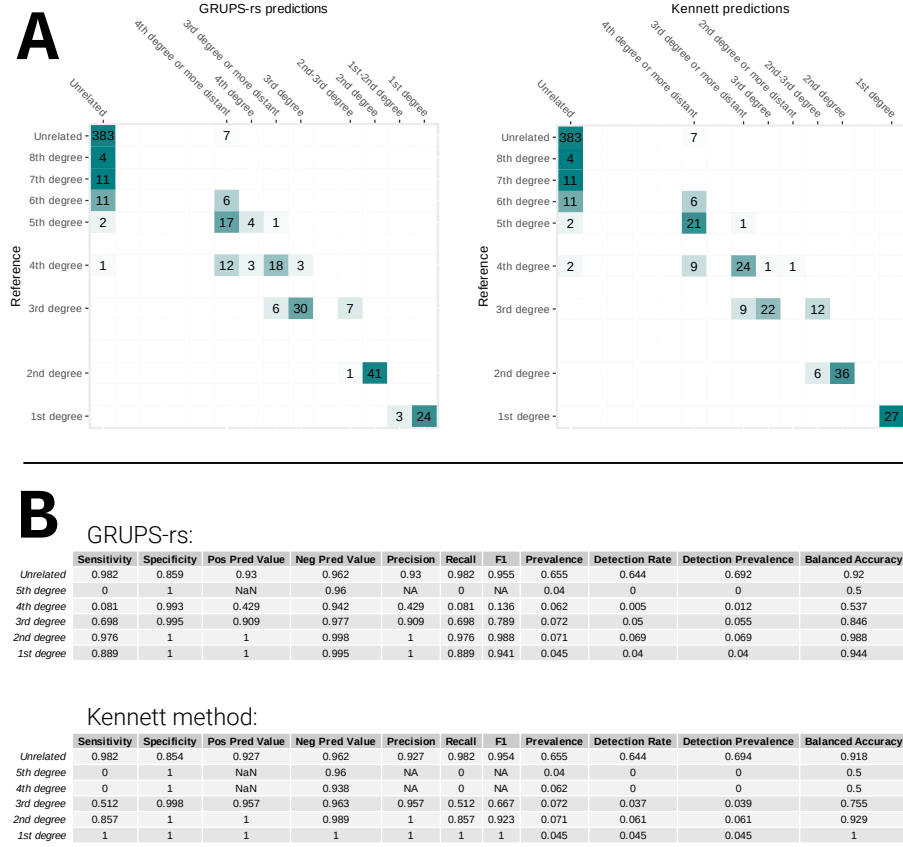


Figure S4: Classification performance of GRUPS-rs and the Kennett method, when applied to the Hazleton-North dataset. **A**: Confusion matrices of both GRUPS-rs (left) and the Kennett method (right). Predictions from the given method are ordered in columns while the validated kinship ties, extracted from the original family tree, are ordered in rows. Values within each cell represent the number of occurrences. The density of each cell's fill color reflects the row-wise relative proportion of that cell. **B**: Classification performance metrics of GRUPS-rs and the Kennett method.

5 Figure S5: R-coefficient estimates of GRUPS-rs against the Kennett method

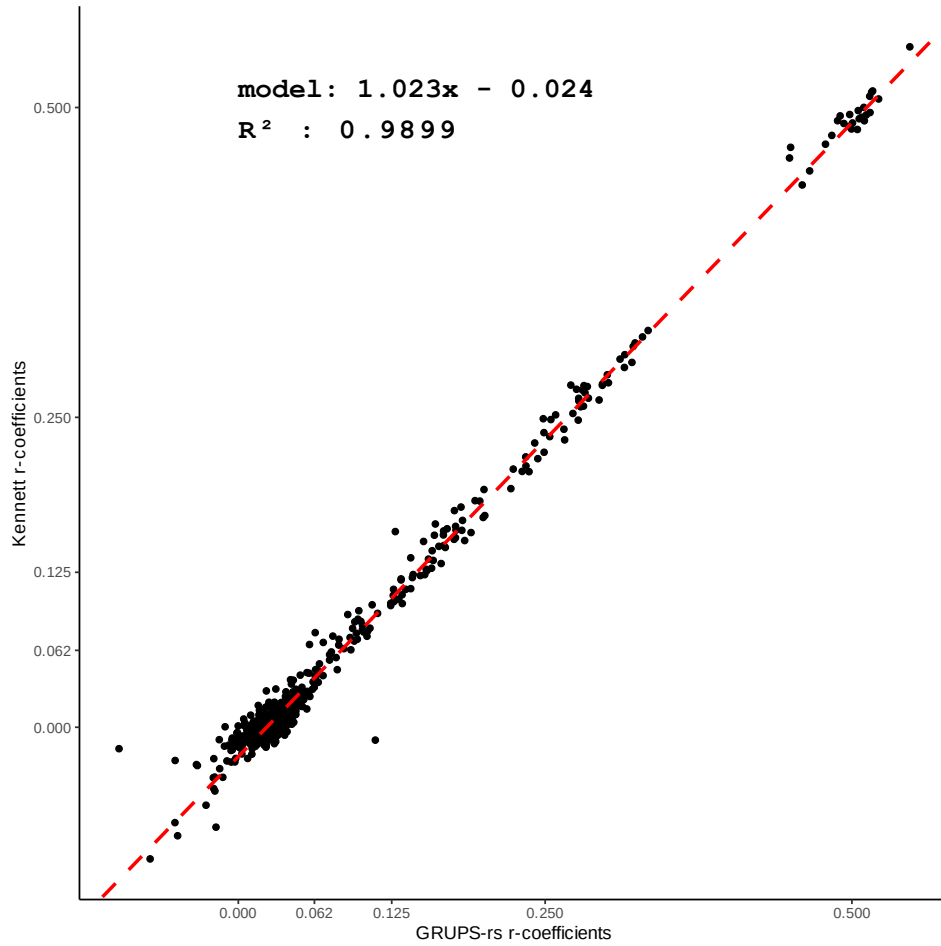


Figure S5: Scatter plot of the r-coefficient estimates obtained through GRUPS-rs (x-axis) and the Kennett method (y-axis). The red dashed line highlights the linear model regression obtained from fitting the data points.

6 Figure S6: Pedigree simulation results for the Koszyce dataset

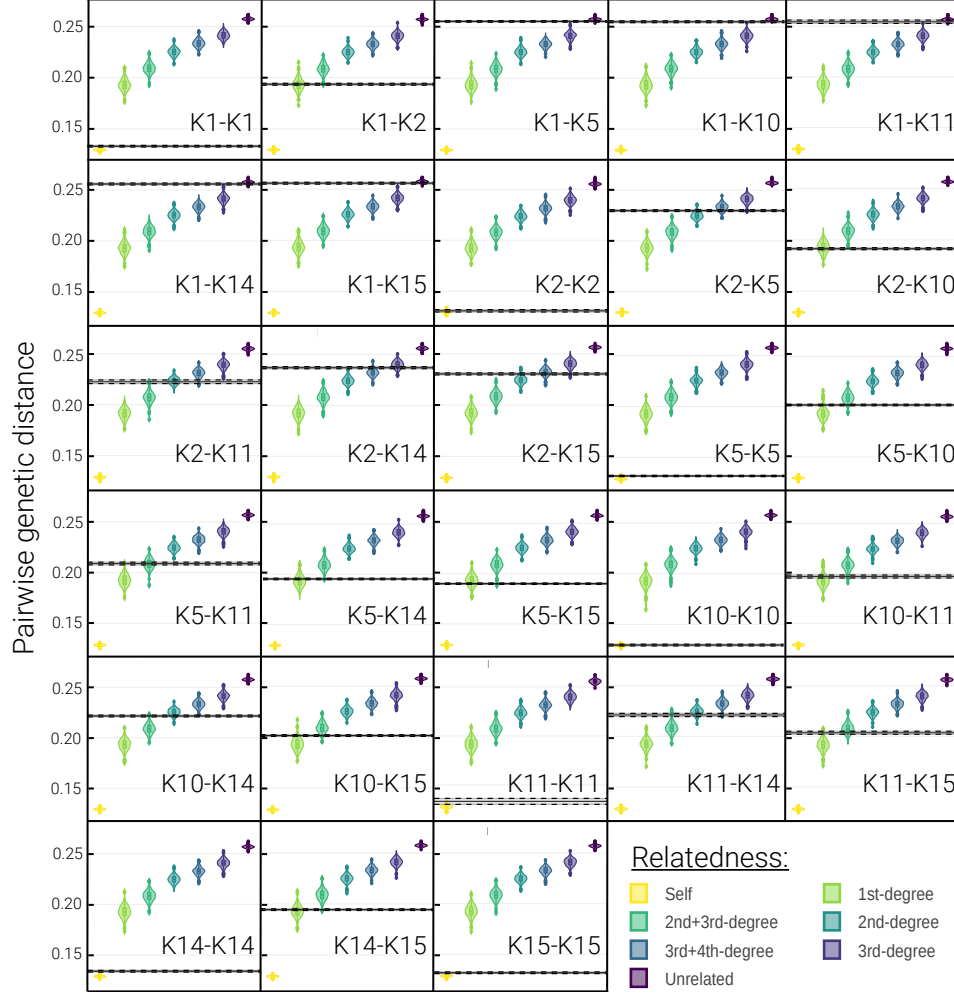


Figure S6: Pairwise pedigree simulation results for the Koszyce dataset using GRUPS-rs along with a template pedigree where K14 and U0.2F are hypothesized to share a first-degree relationship. Colored violin plots represent the distributions of $\widehat{PWD}_{i,j,k}^{sim}$ for the corresponding pair i, j . Solid and dashed black lines represent the average $\widehat{PWD}_{i,j}^{obs}$ and 95% confidence interval, respectively. Relationship labels such as "2nd+3rd-degree" indicate intermediate relationships (e.g.: "2nd+3rd-degree" indicates that the pair shares both a second-degree and a third-degree relationship).