

Original Research

Simulation-based Benchmarking of Ancient Haplotype Inference for Detecting Population Structure

Jazeps Medina Tretmanis ¹, Flora Jay ^{2,†*}, María C. Ávila-Arcos ^{3,†*}, Emilia Huerta-Sanchez ^{1,†*}

1. Center for Computational Molecular Biology, Brown University, Providence, RI 02912, USA;

Email: jazeps_medina_tretmanis@brown.edu

2. Interdisciplinary Laboratory of Numerical Sciences, Université Paris-Saclay, 91400, Orsay, France

3. International Laboratory for Human Genome Research, UNAM, 76230 Querétaro, Mexico

† These authors contributed equally to this work.

* **Correspondence:** Flora Jay; Email: flora.jay@lri.fr;

María C. Ávila-Arcos; Email: mavila@liigh.unam.mx;

Emilia Huerta-Sanchez; Email: emilia_huerta-sanchez@brown.edu.

Supplementary Materials

Performance of population and reference panel phasing.

Population split

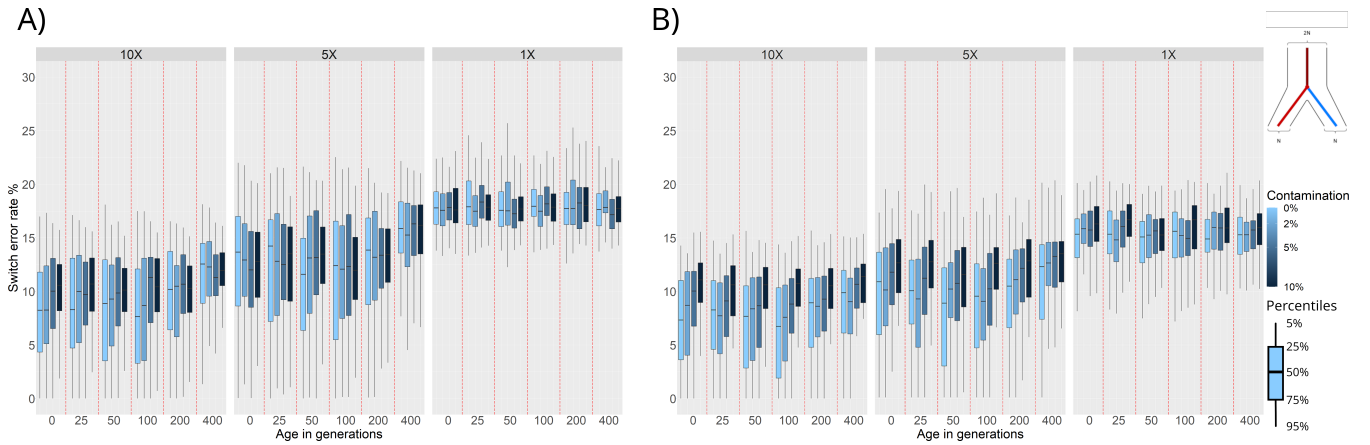


Figure S1. Switch Error Rate (SWE) distributions for phasing of ancient individuals simulated under a population split 100 generations in the past: **A)** Population phasing. **B)** Reference panel phasing. SWE (y-axis) is presented in 3 facets corresponding to 10 \times , 5 \times , and 1 \times average read depth. Within these facets are the SWE distributions for 100 simulated individuals for each combination of age (x-axis) and contamination level (shades of blue).

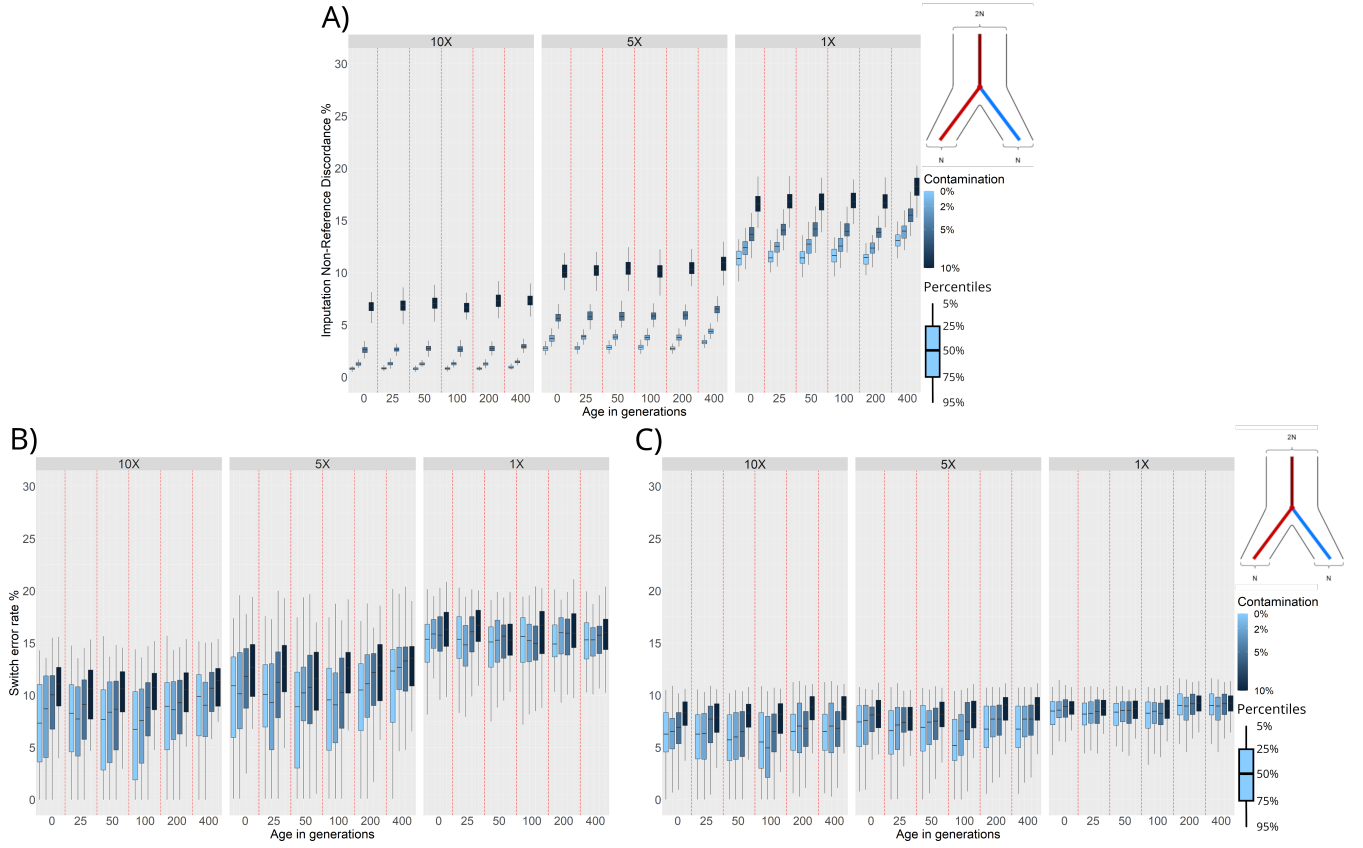


Figure S2. Non-Reference Discordance (NRD) and Switch Error Rate (SWE) distributions for individuals simulated under a demographic model with a population split event 100 generations in the past. NRD or SWE (y-axis) are presented in facets corresponding to 10 \times , 5 \times , and 1 \times average coverage. For each combination of age (x-axis) and contamination level (shades of blue), we plot the distributions of NRD and SWE values computed for each of the 100 simulated individuals. All phasing results were obtained through reference panel phasing. **A)** NRD distributions of imputed variants. **B)** SWE distributions of unimputed individuals. **C)** SWE distributions of imputed individuals.