Supplemental Data

Inference of Population Structure

from Time-Series Genotype Data

Tyler A. Joseph and Itsik Pe'er
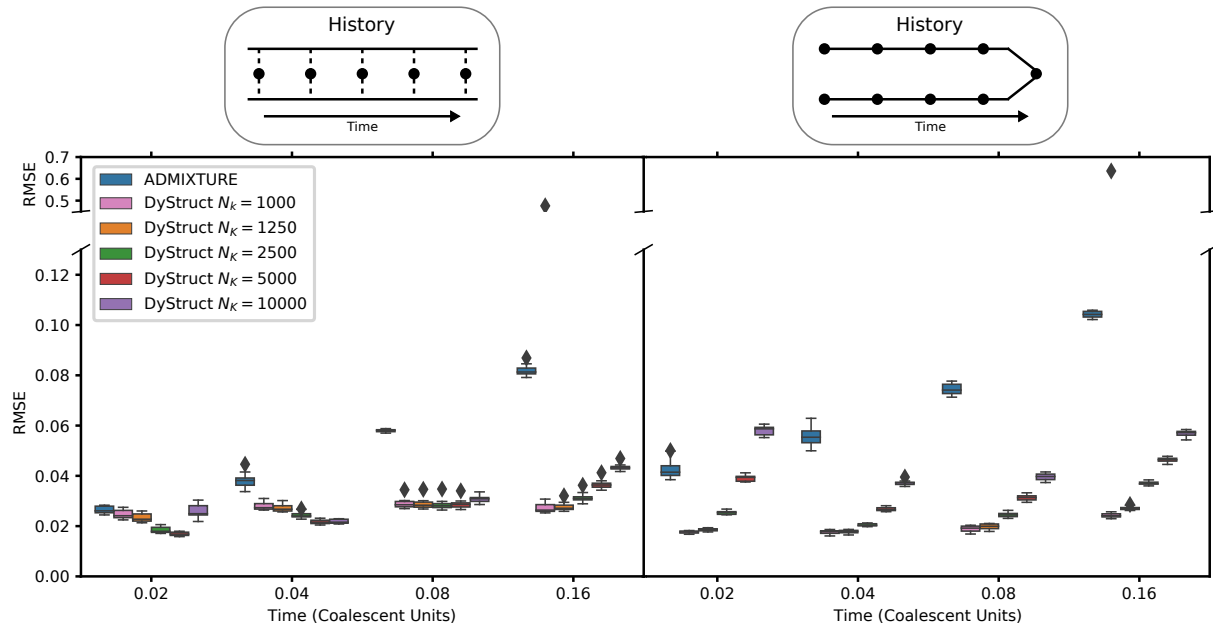
1 Supplementary Figures

Figure S1: **Boxplots of RMSE for DyStruct and ADMIXTURE on two historical scenarios.** The results for ADMIXTURE and DyStruct with $N_k \in \{1250, 2500, 5000, 10000\}$ are the same as presented in Figure 2 (right). Simulations with DyStruct $N_k = 1000$ have been added for comparison. At time = 0.16 for both scenarios DyStruct with $N_k = 1000$ achieves a local optima in 1 of the 10 simulations (outliers at the top of both figures). Note that all simulations depicted start from the same random seed, and thus have the same initial conditions and iterate through the same order of loci. This suggests that larger population sizes are less susceptible to this local optima. Notably, rerunning DyStruct with $N_k = 1000$ and a different random seed on the outliers produces results similar to the better performing runs.
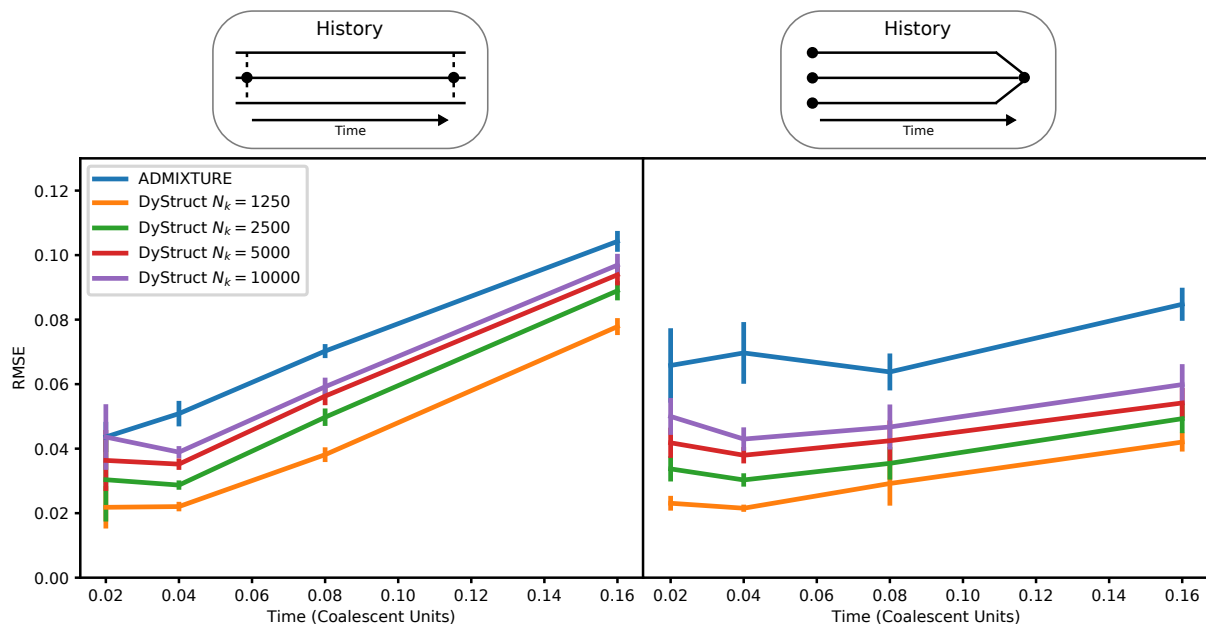
Figure S2: **Performance of DyStruct and ADMIXTURE on simulated data for three populations at two time points.** The root mean square error (RMSE) between ground truth ancestry components and estimated ancestry components (y-axis) is computed across total simulation time (x-axis). Error bars depict 1 standard deviation computed from 10 simulation replicates. In contrast with increasing the number of sample time points, model performance remains similar to the case where two populations are sampled.
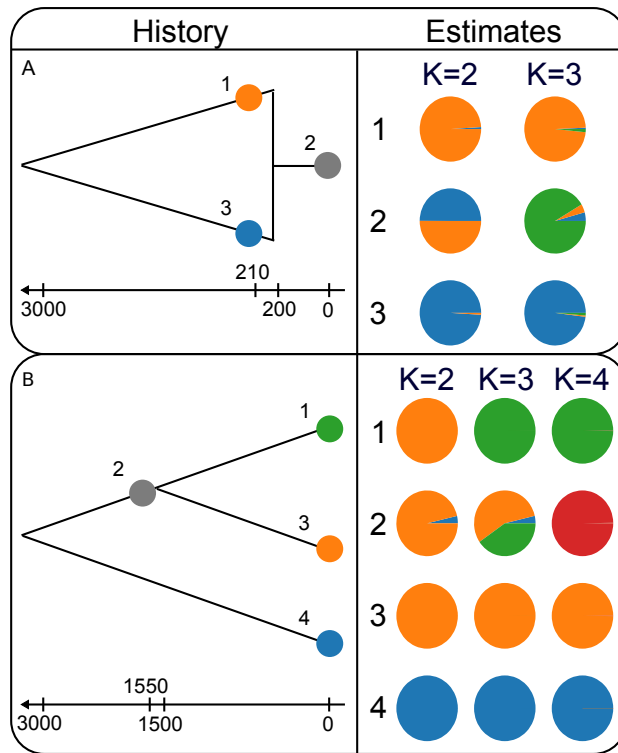
Figure S3: **ADMIXTURE's estimates for out-of-model coalescent simulations.** Mean ancestry estimates are computed by averaging each sample's ancestry components across 10 simulation replicates. ADMIXTURE qualitatively identifies similar patterns as DyStruct, but estimates less shared ancestry between ancient and modern samples for $K = 3$ in the coalescent merger scenario (A) and $K = 4$ for the coalescent split scenario (B).
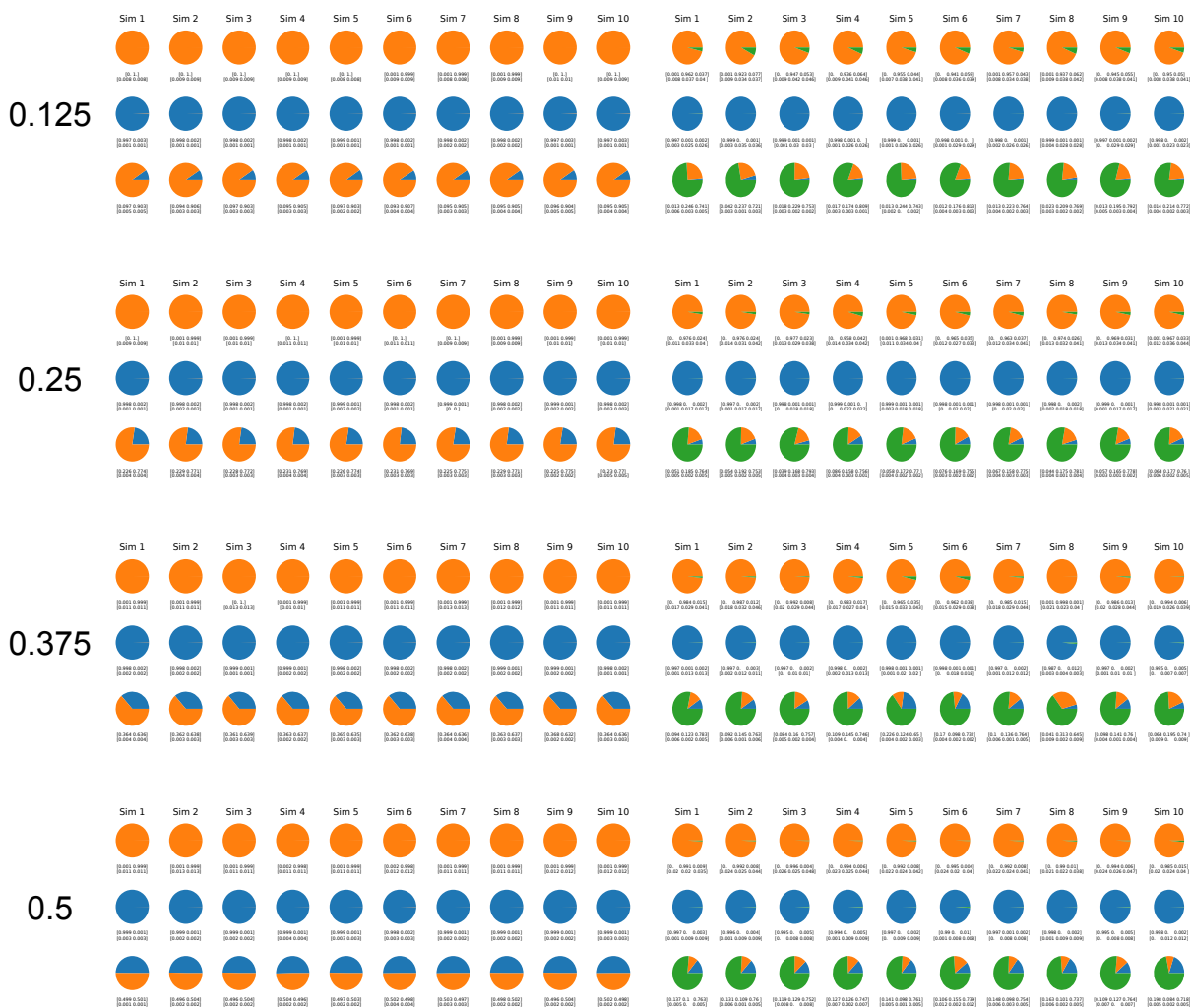
Figure S4: **DyStruct's mean ancestry estimates for 10 simulation replicates under the coalescent merger scenario.** Each pie chart shows the mean ancestry estimates across samples for a particular simulation replicate. The chart is followed by two vectors (enlarge for text). The first gives the mean ancestry estimate across samples, the second gives the standard deviation. Each are rounded to three decimal places.
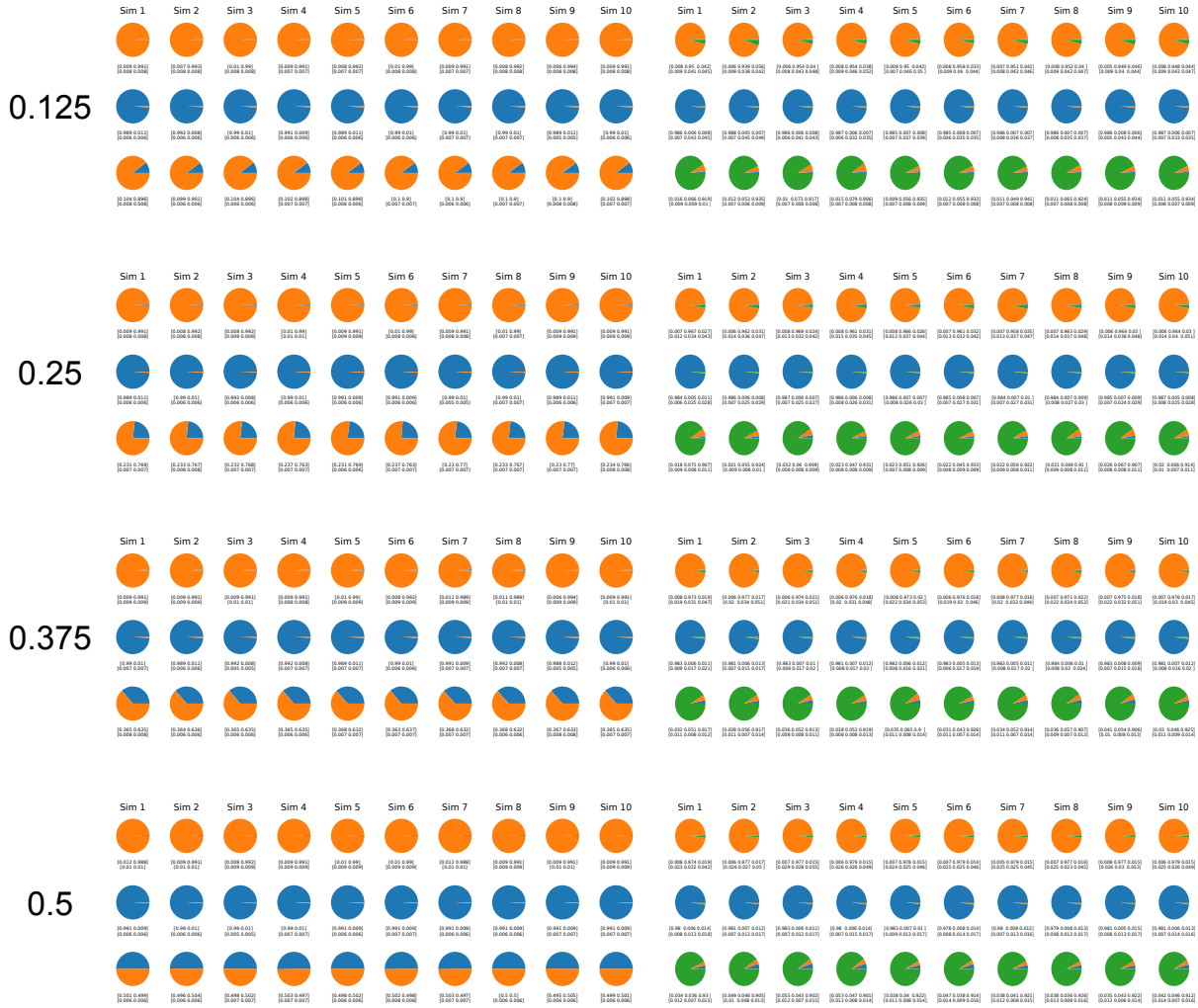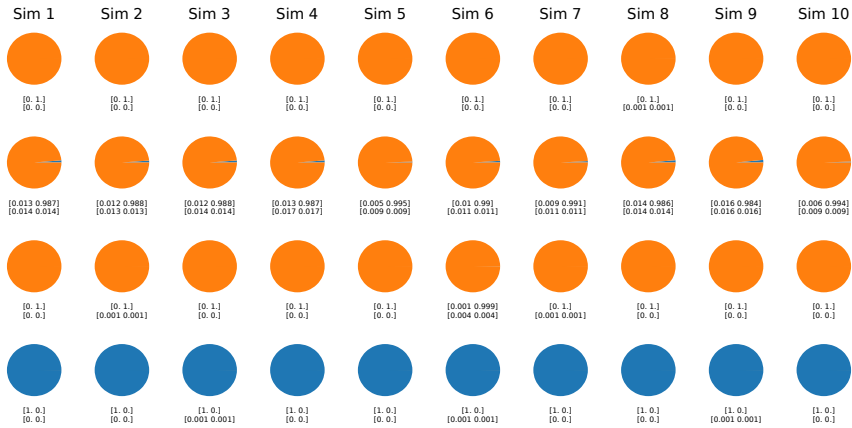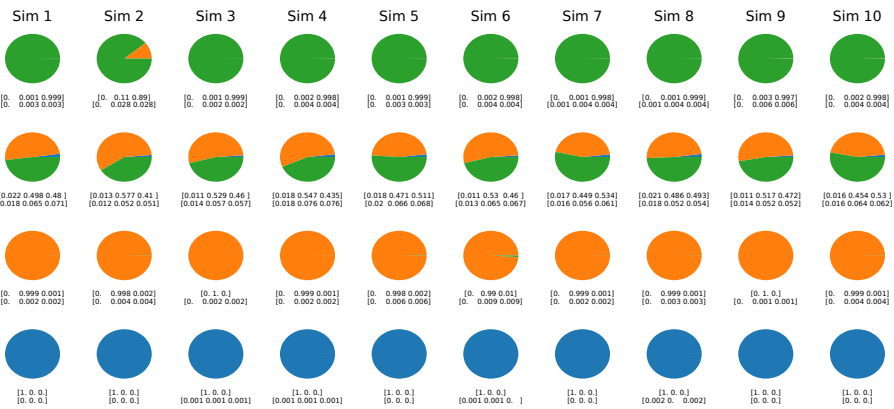
ADMIXTURE Coalescent Merger

Figure S5: **ADMIXTURE's mean ancestry estimates for 10 simulation replicates under the coalescent merger scenario.** Each pie chart shows the mean ancestry estimates across samples for a particular simulation replicate. The chart is followed by two vectors (enlarge for text). The first gives the mean ancestry estimate across samples, the second gives the standard deviation. Each are rounded to three decimal places.

DyStruct Coalescent Split
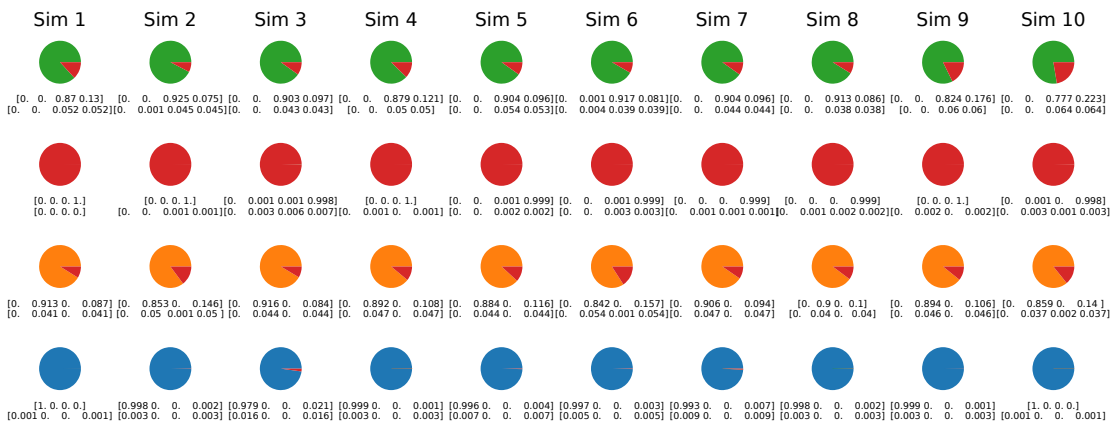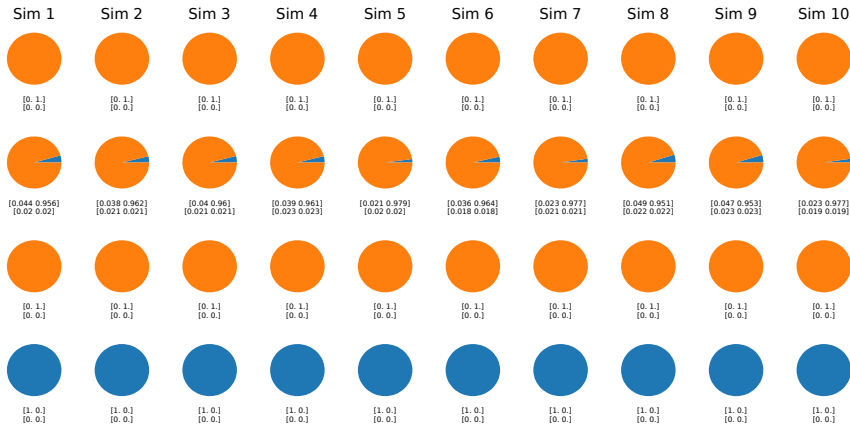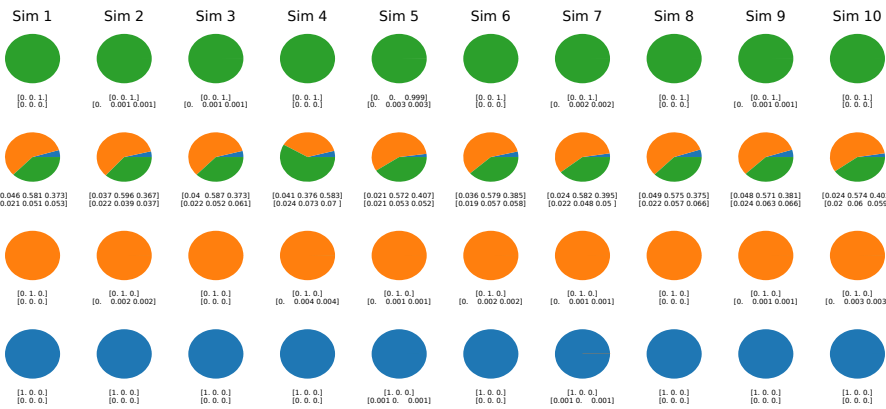
K=2

K=3

K=4

Figure S6: **DyStruct's mean ancestry estimates for 10 simulation replicates under the coalescent split scenario.** Each pie chart shows the mean ancestry estimates across samples for a particular simulation replicate. The chart is followed by two vectors (enlarge for text). The first gives the mean ancestry estimate across samples, the second gives the standard deviation. Each are rounded to three decimal places.

ADMIXTURE Coalescent Split
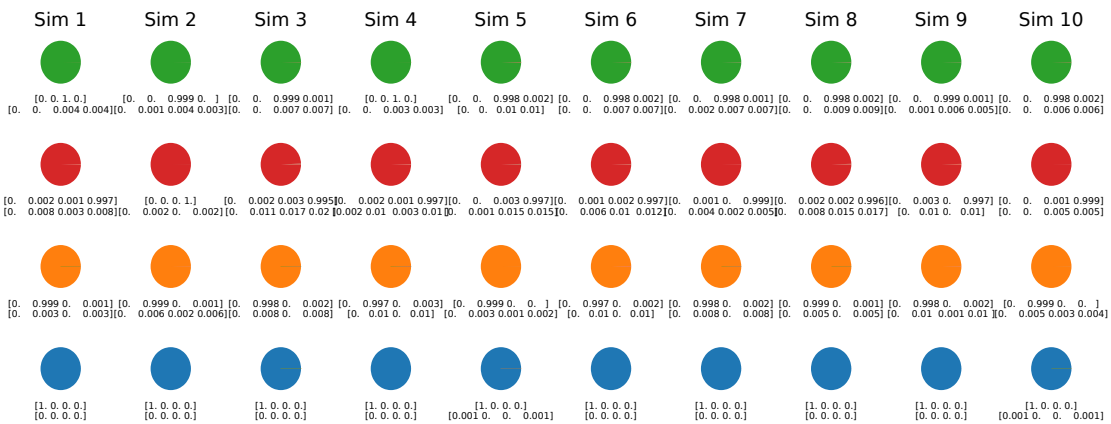
K=2

K=3

K=4

Figure S7: **ADMIXTURE's mean ancestry estimates for 10 simulation replicates under the coalescent split scenario.** Each pie chart shows the mean ancestry estimates across samples for a particular simulation replicate. The chart is followed by two vectors (enlarge for text). The first gives the mean ancestry estimate across samples, the second gives the standard deviation. Each are rounded to three decimal places.
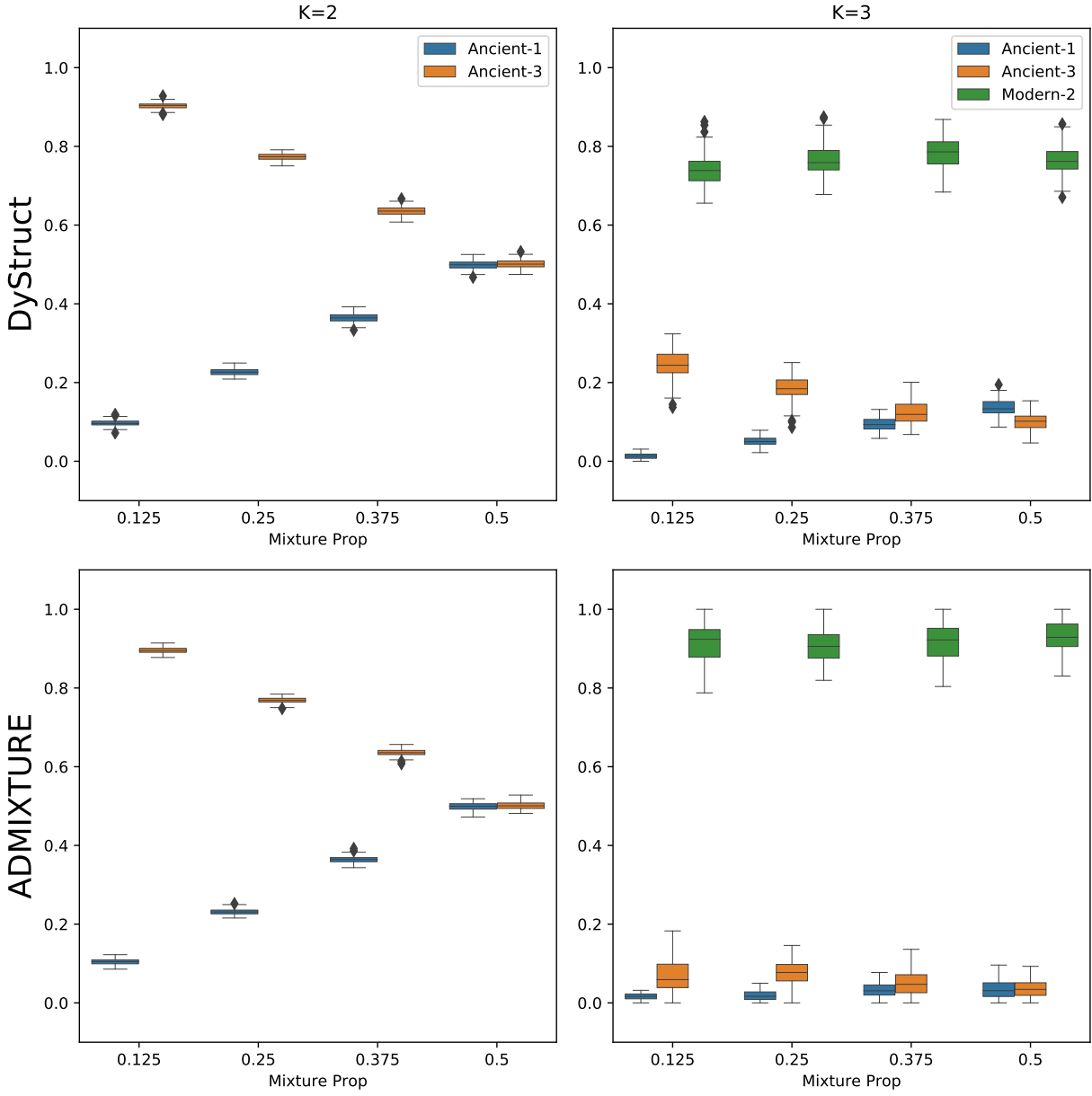
Figure S8: **Boxplots of ancestry estimates across samples for simulation 1 of the coalescent merger scenario.** Only ancestry estimates for the admixed modern population are displayed. Legend numbers correspond to labels on Figure 5 and Figure S3. Ancestry estimates are stable across samples.
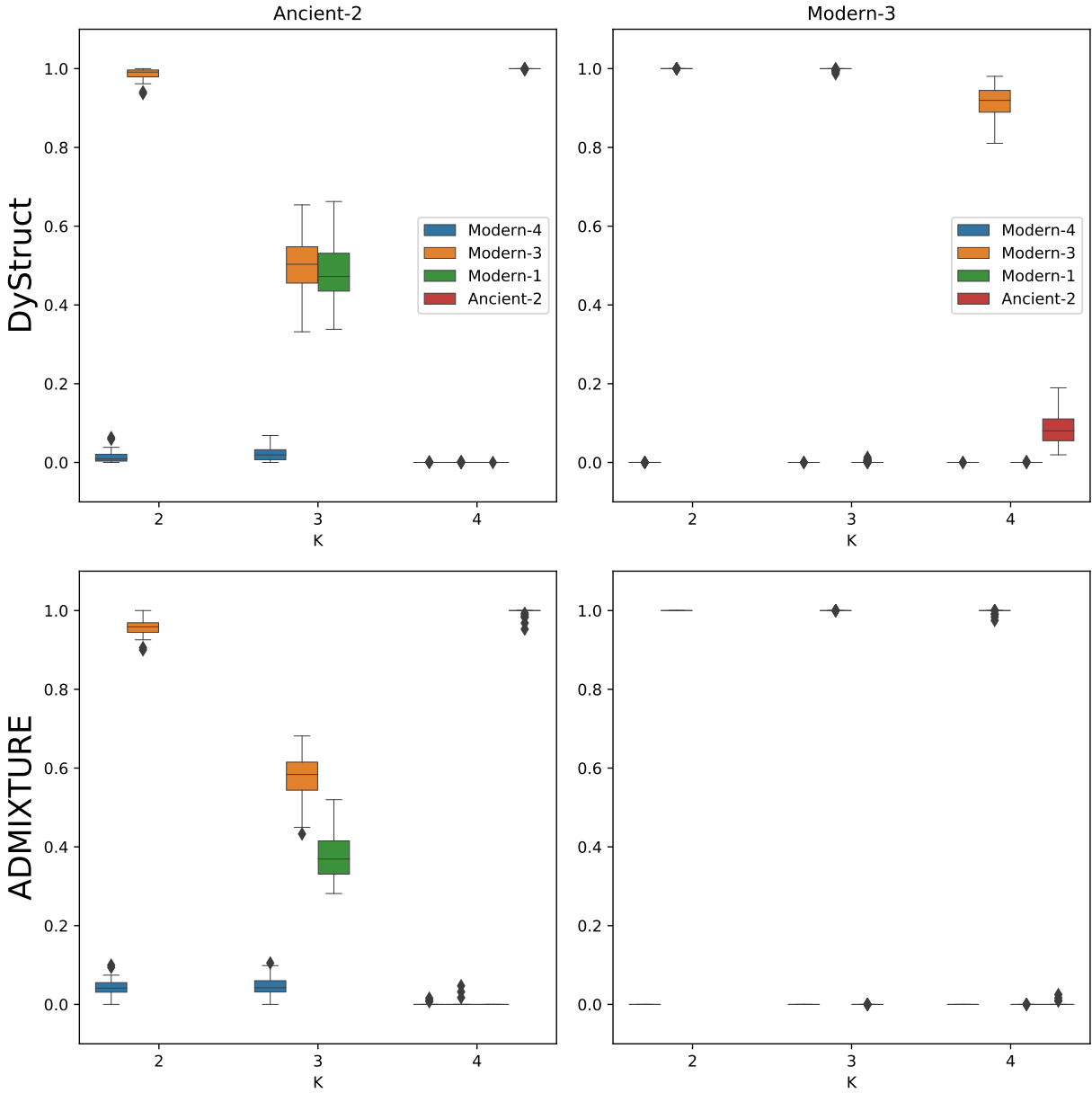
Figure S9: **Boxplots of ancestry estimates across samples for simulation 1 of the coalescent split scenario.** Only ancestry estimates for the ancient population and one of the modern populations are displayed. Legend numbers correspond to labels on Figure 5 and Figure S3. Ancestry estimates are stable across samples.
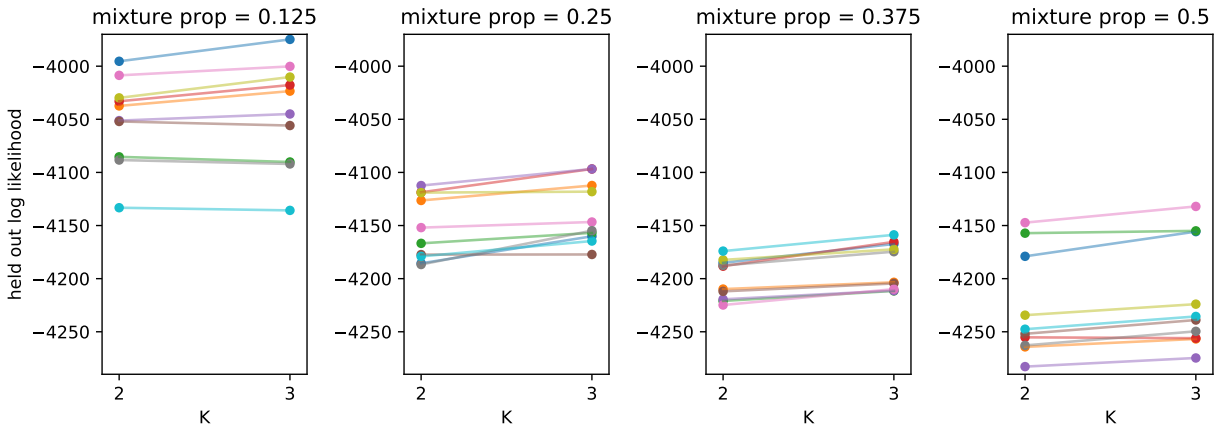
Figure S10: **Model fit across K for the coalescent merger simulations.** Value of the log likelihood (y-axis) on held out genotypes across K (x-axis) and different mixture proportions. 10 simulation replicates were performed for each mixture proportion. The best supported K is the one with the highest held out log likelihood. However, both $K = 2$ and $K = 3$ appear equally well supported by the model.
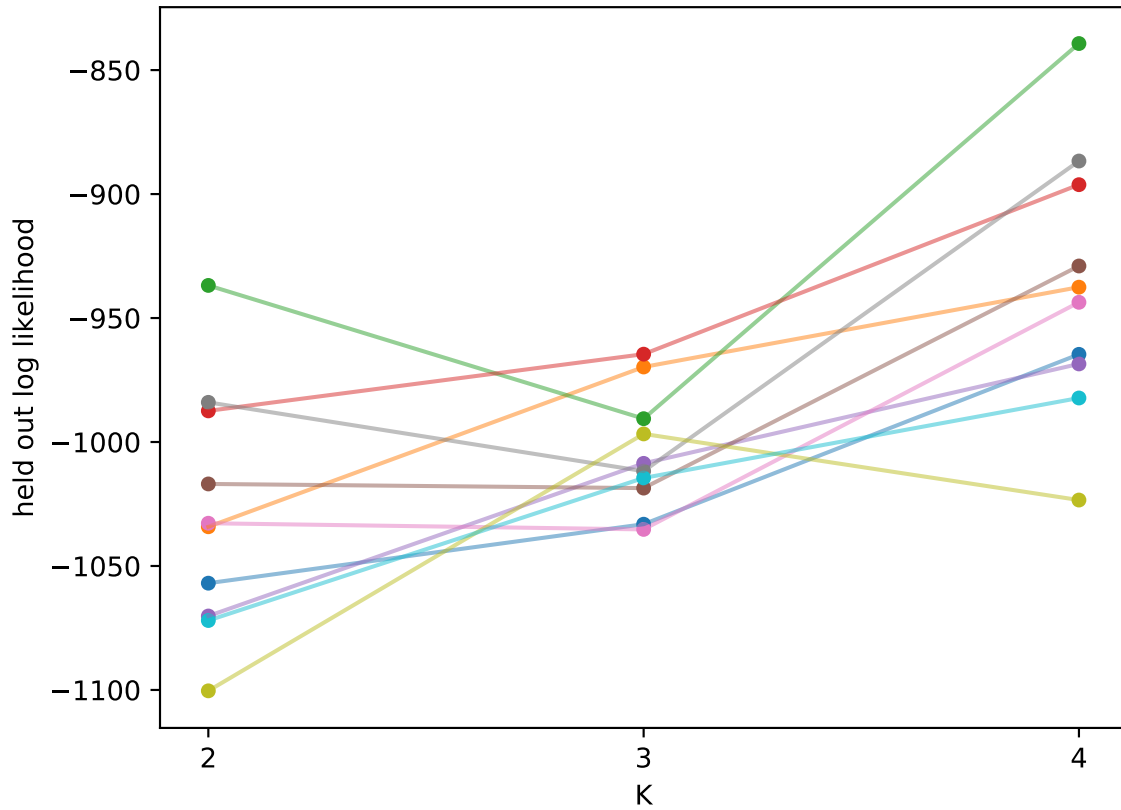

Figure S11: **Model fit across K for the coalescent split simulations.** Value of the log likelihood (y-axis) on held out genotypes across K (x-axis) for the coalescent split simulations.
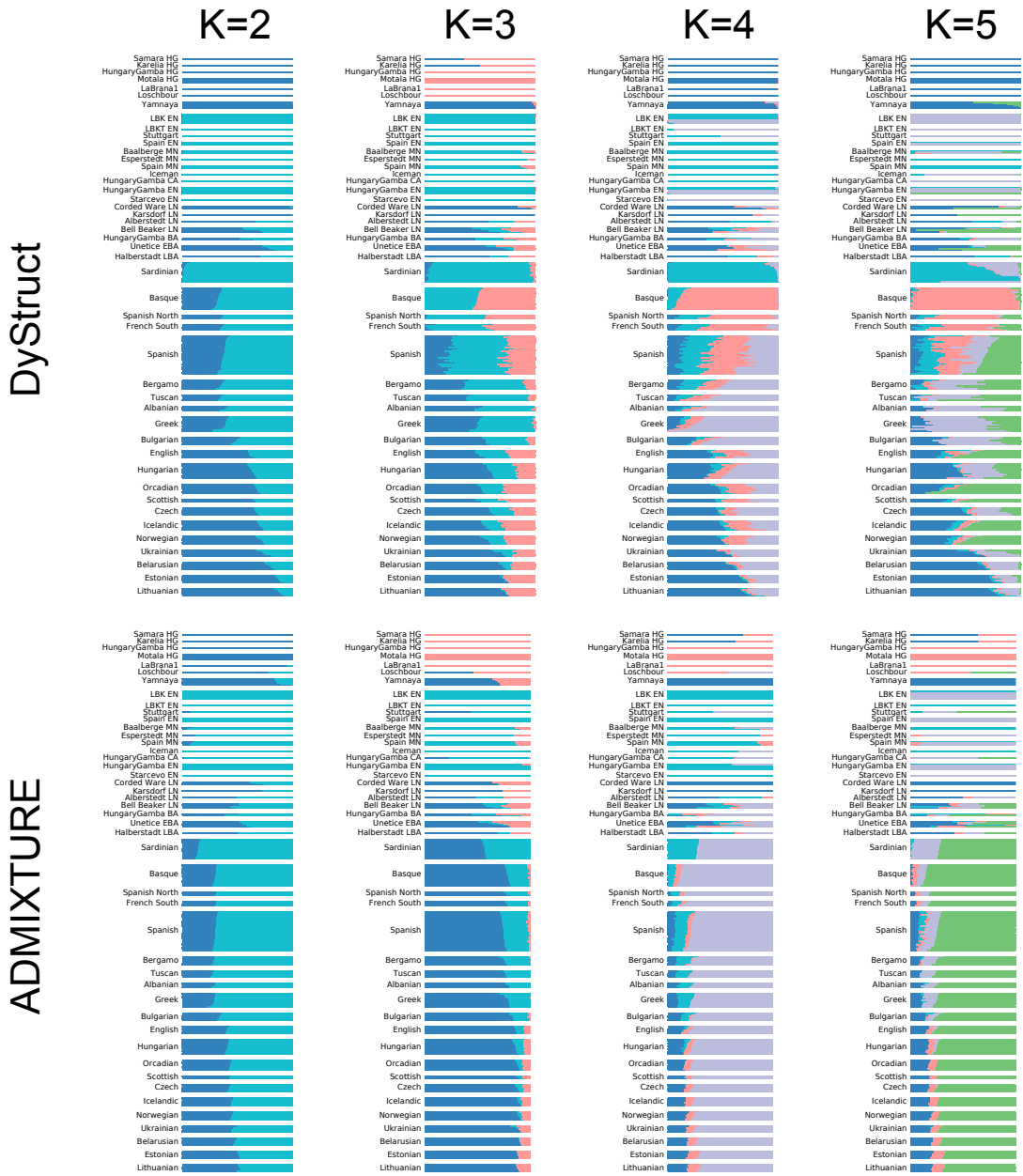
Figure S12: **Stacked bar plots of ancestry estimates from DyStruct and ADMIXTURE on a subset of ancient samples and modern Europeans.**
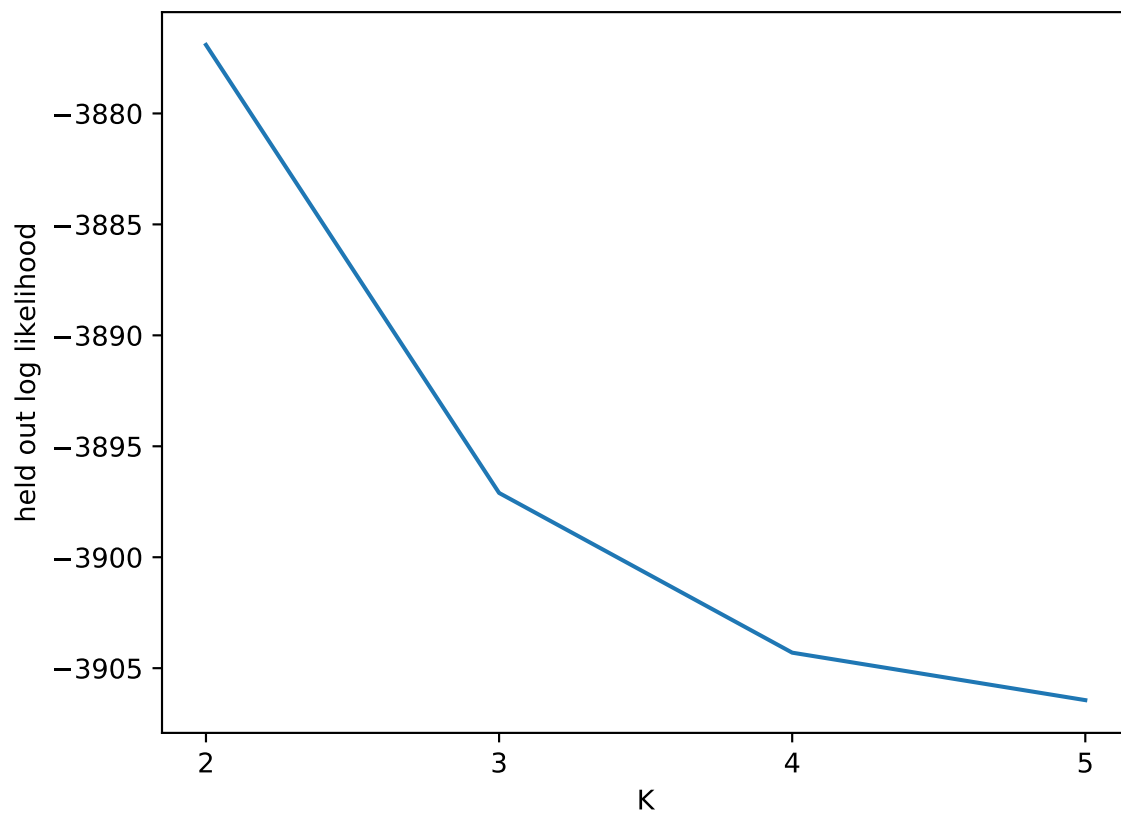
Figure S13: **Model fit across K for subset of ancient samples and modern Europeans.** Value of the log likelihood (y-axis) on held out genotypes across K (x-axis).
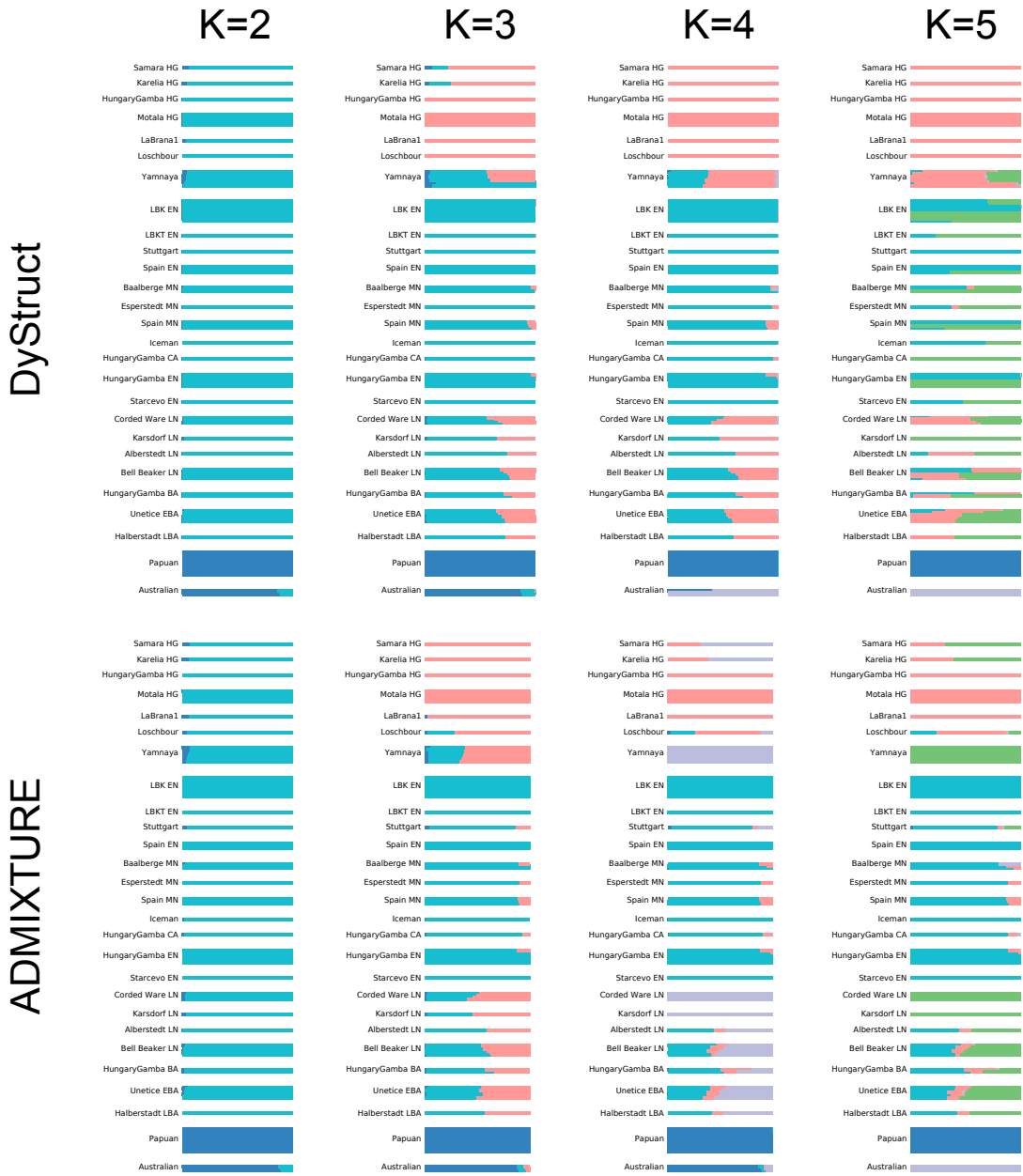
Figure S14: **Stacked bar plots of ancestry estimates from DyStruct and ADMIXTURE on a subset of ancient samples and modern Oceanians.**
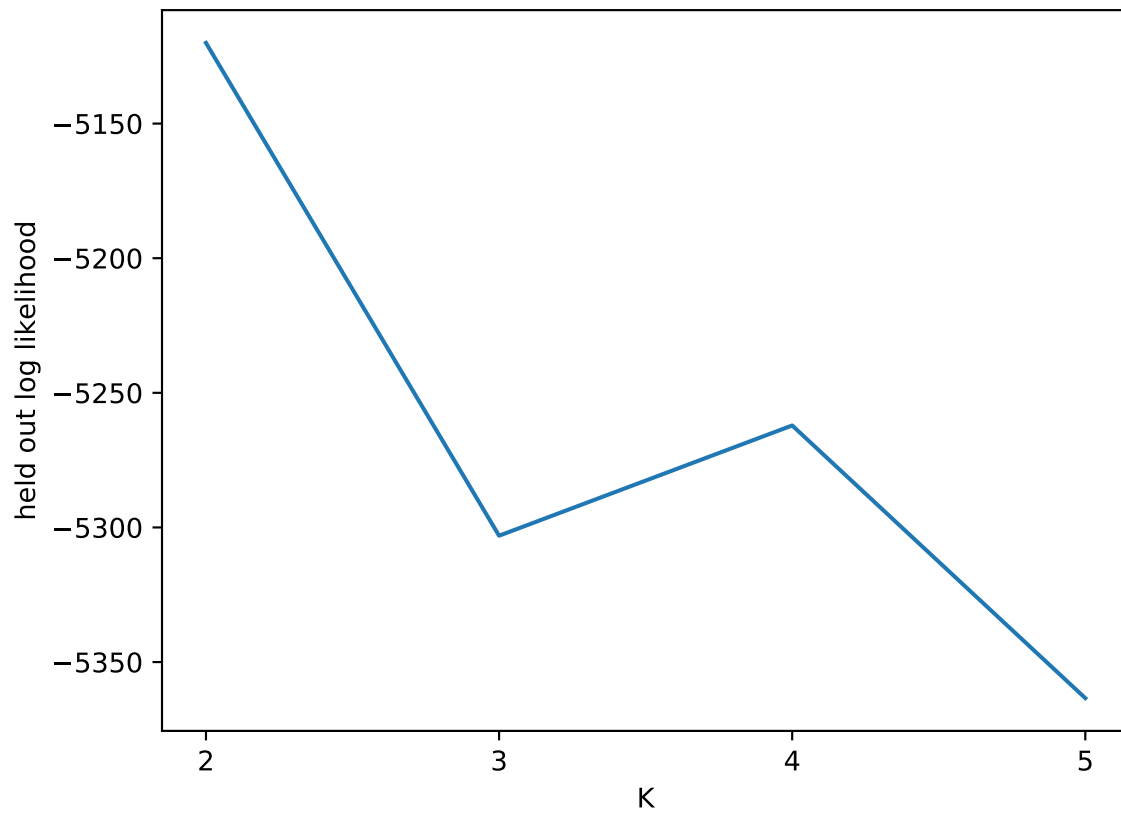
Figure S15: **Model fit across K for subset of ancient samples and modern Oceanians.** Value of the log likelihood (y-axis) on held out genotypes across K (x-axis).
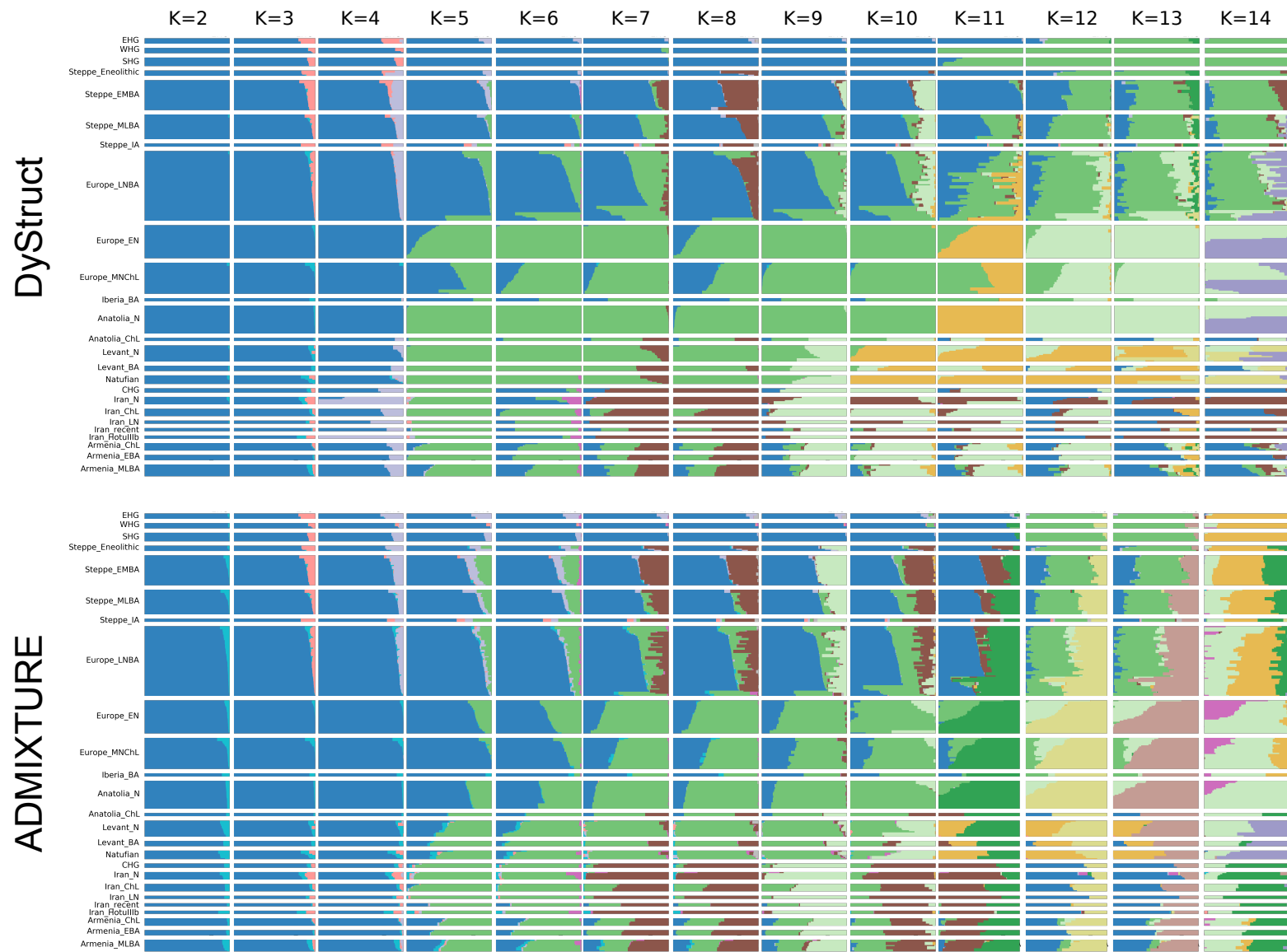
Figure S16: **Stacked bar plots of ancestry estimates from DyStruct and ADMIXTURE on ancient samples in the Lazaridis dataset**.
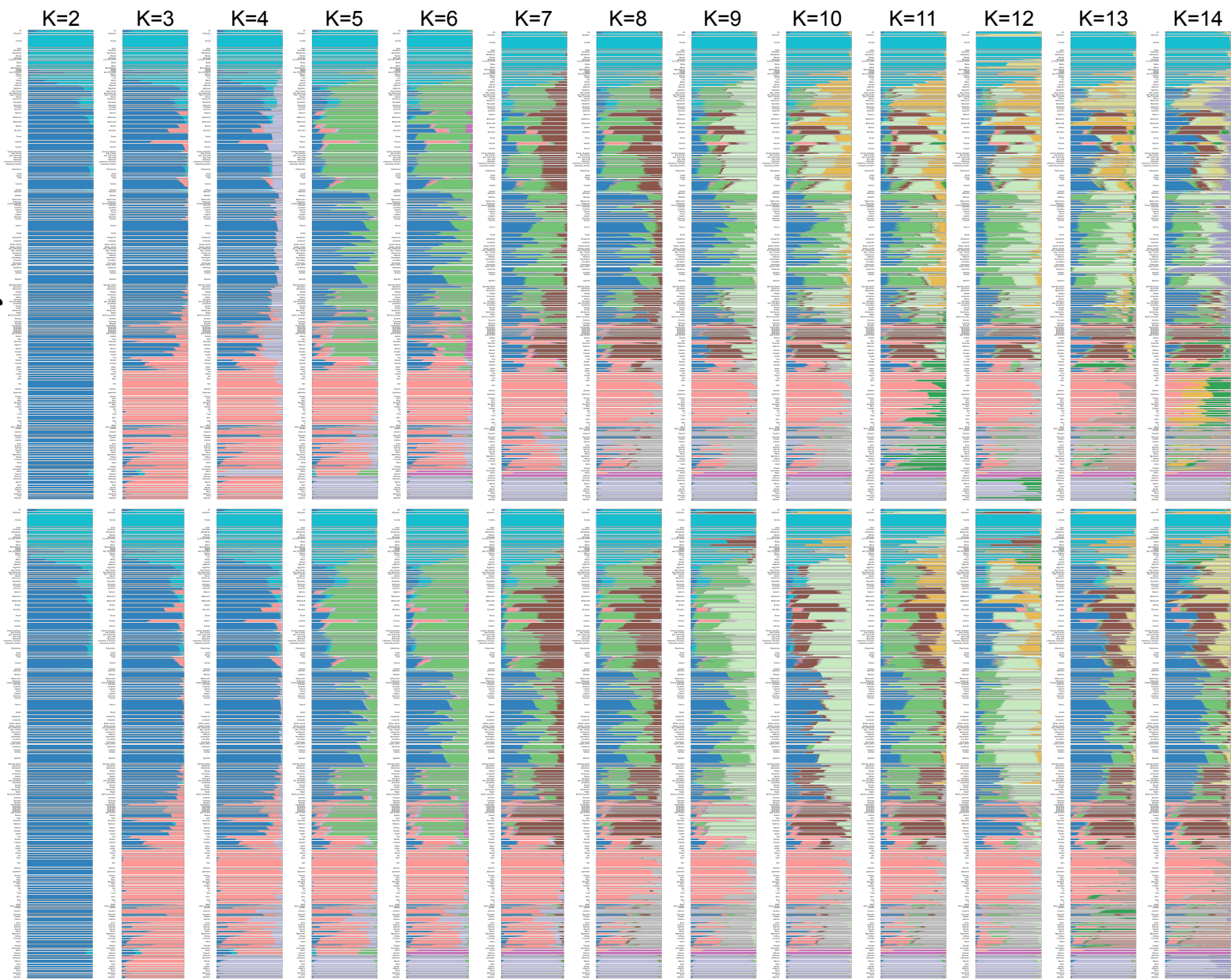
Figure S17: **Stacked bar plots for DyStruct and ADMIXTURE on modern samples from the Lazaridis dataset**. Enlarge for labels.
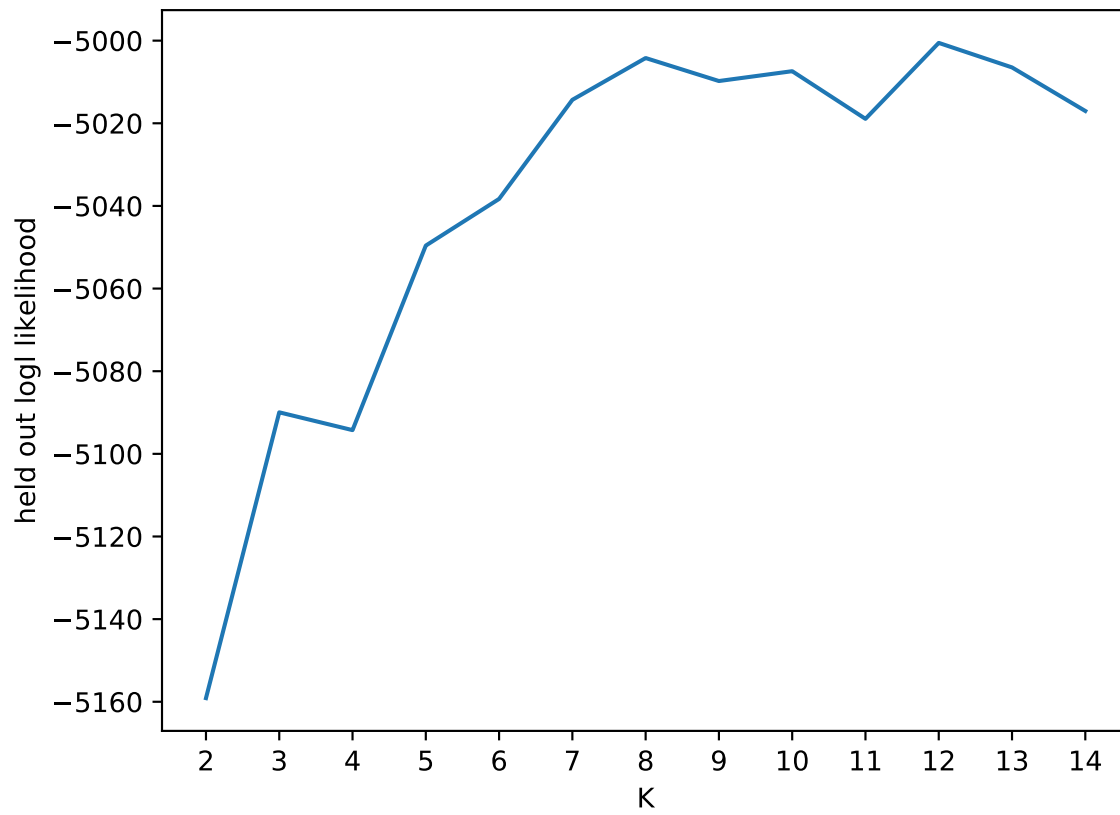
Figure S18: **Model fit across K for the Lazaridis dataset.** Value of the log likelihood (y-axis) on held out genotypes across K (x-axis).