Supplementary Figures: Investigating mito-nuclear interactions in admixed human populations

Figure S1. Scatter plots of mtDNA copy number (y-axis) against autosomal ancestry fraction (x-axis). Rows and columns are split by the geographic origin of the mtDNA haplogroup and source populations, respectively. The top left to bottom right diagonal, where the nuclear and mtDNA ancestry are similar, shows positive correlations between ancestry and mtDNA copy number, while the other panels show negative correlations, consistent with our predictions in Fig. 1A. Beta coefficients and two-sided P-values are shown.

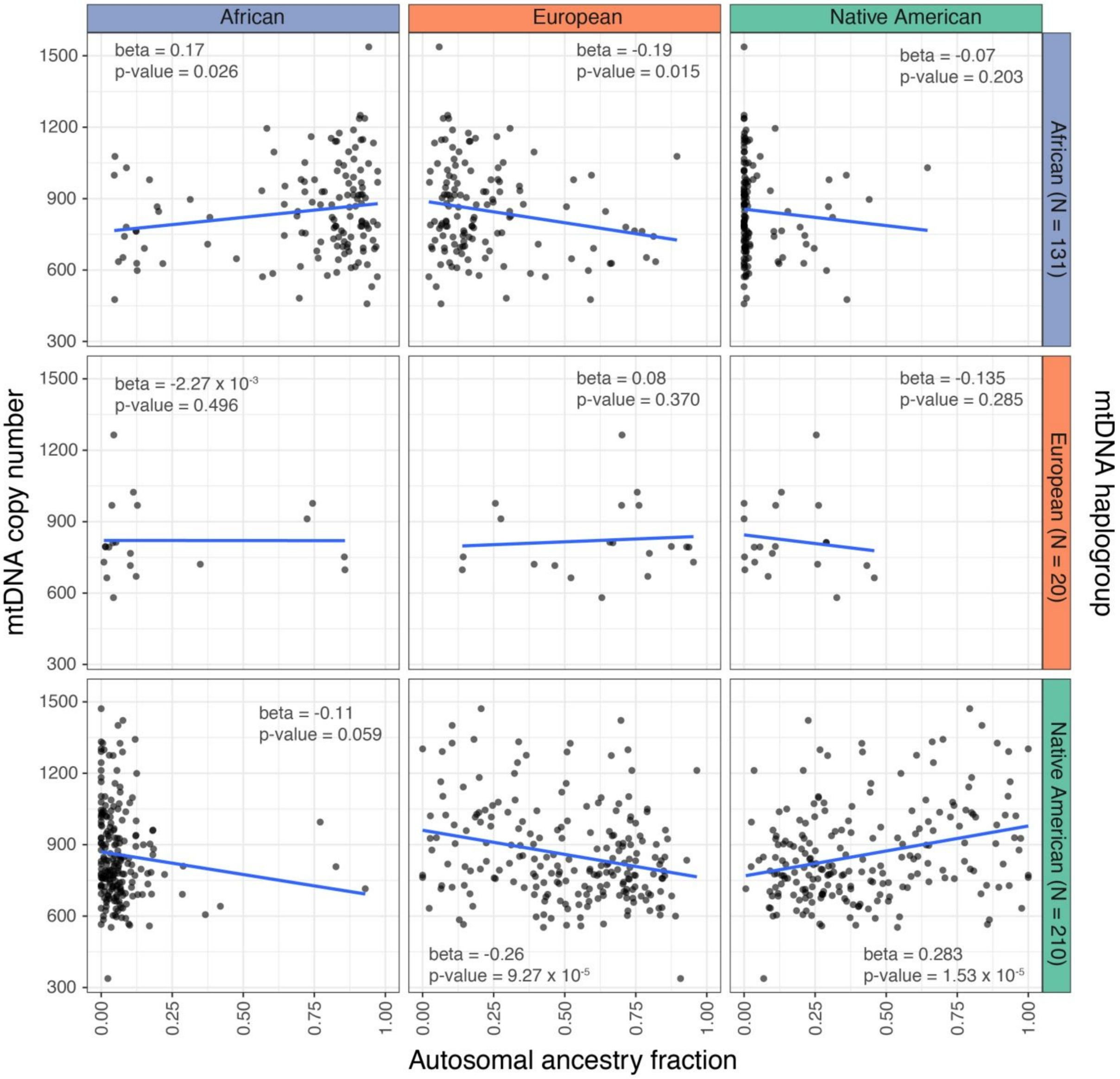


Figure S2. Estimated female and male composition in the ancestral populations for each admixed group. ACB: African Caribbeans from Barbados; ASW: African Americans from South Western USA; CLM: Colombians from Colombia; MXL: Mexicans from Los Angeles, USA; PEL: Peruvians from Lima, Peru; PUR: Puerto Ricans from Puerto Rico.

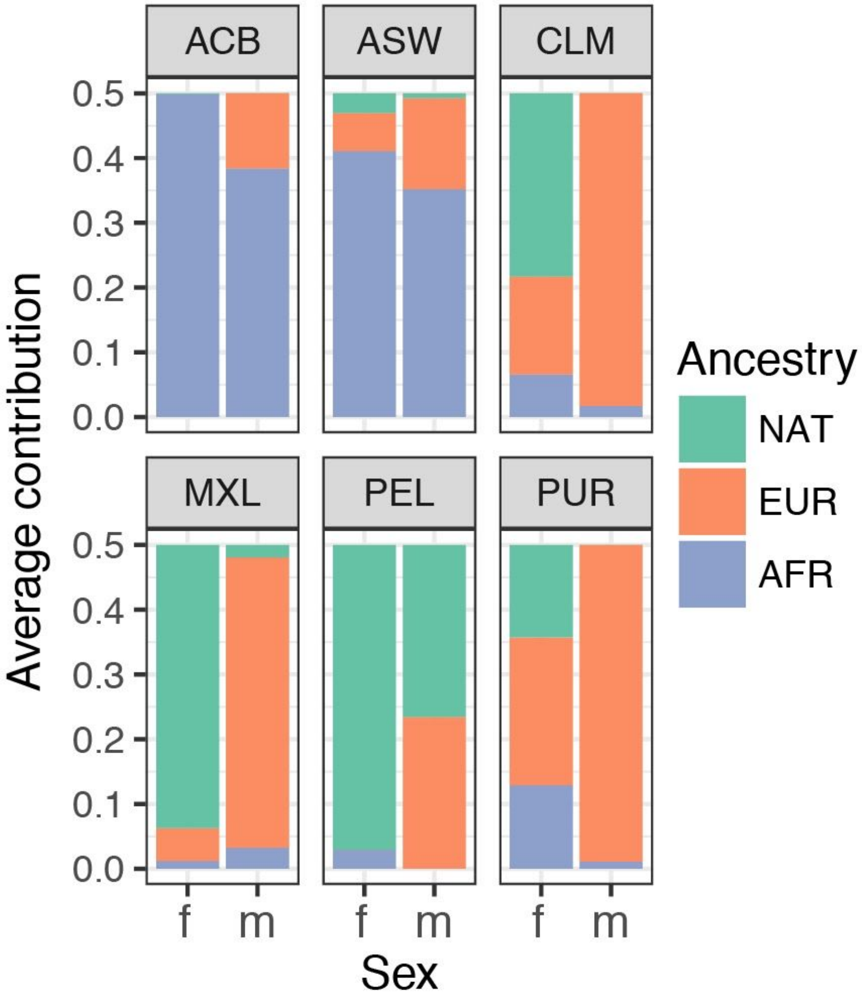


Figure S3. The simulated amount of drift in local ancestry (for 1,000 independent loci) is similar to the observed amount of drift in local ancestry at autosomal loci in both PUR and CLM. CLM: Colombians from Colombia; PUR: Puerto Ricans from Puerto Rico.

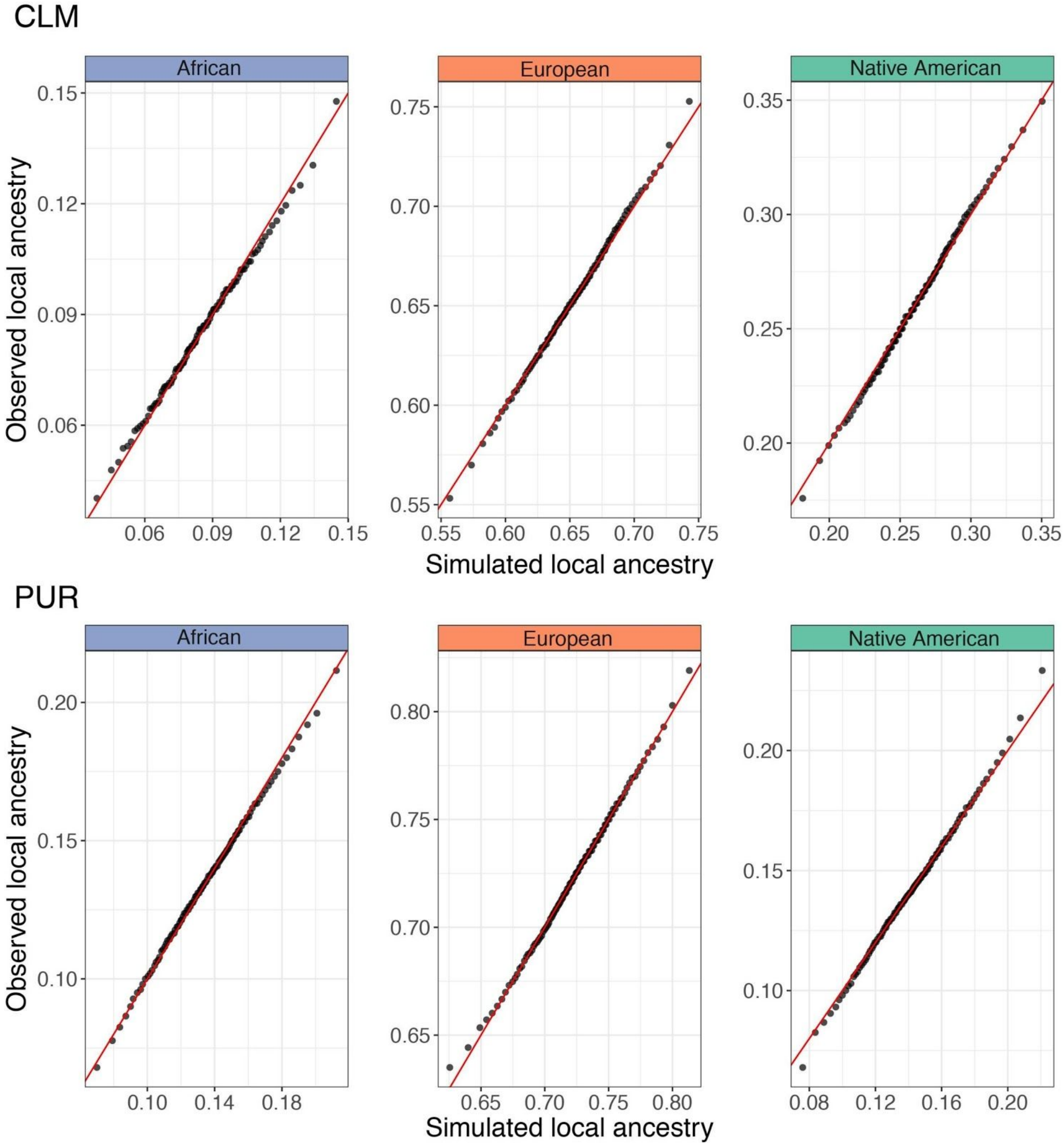


Figure S4. Results of simulations showing that drift since admixture is not sufficient to account for the increase in frequency of Native American mtDNA haplogroup, and concomitant decrease in European mtDNA haplogroup, in Colombians and Puerto Ricans. The violin plots show ancestry for 10,000 independent loci evolving neutrally since admixture. Thus, the median of the distribution represents the average ancestry contribution from each source population (African, European, and Native American), and the width of the distribution represents the amount of drift in ancestry since admixture. The red points indicate the observed frequency of mtDNA haplogroups. CLM: Colombians from Colombia; PUR: Puerto Ricans from Puerto Rico.

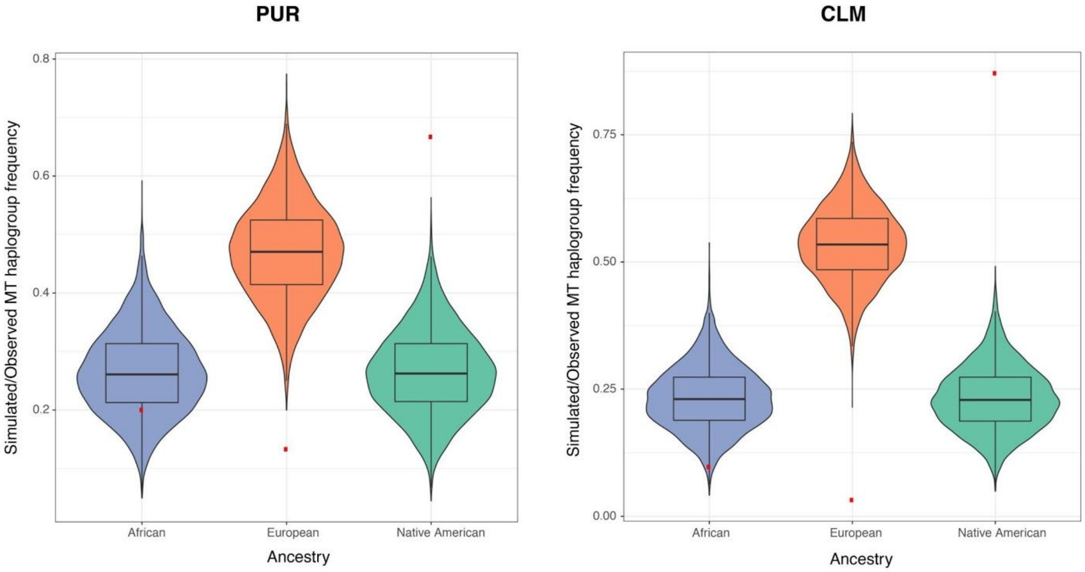


Figure S5: ADMIXTURE bar plot for values of K from 2 to 5. Each vertical line on the x-axis represents an individual and the different colors represent different inferred ancestral components. The height of each bar represents the estimated fraction of the genome from each component. The panels from left to right show individuals who were characterized as admixed, African, European, or Native American. The panels from top to bottom show bar plots for increasing values of K.

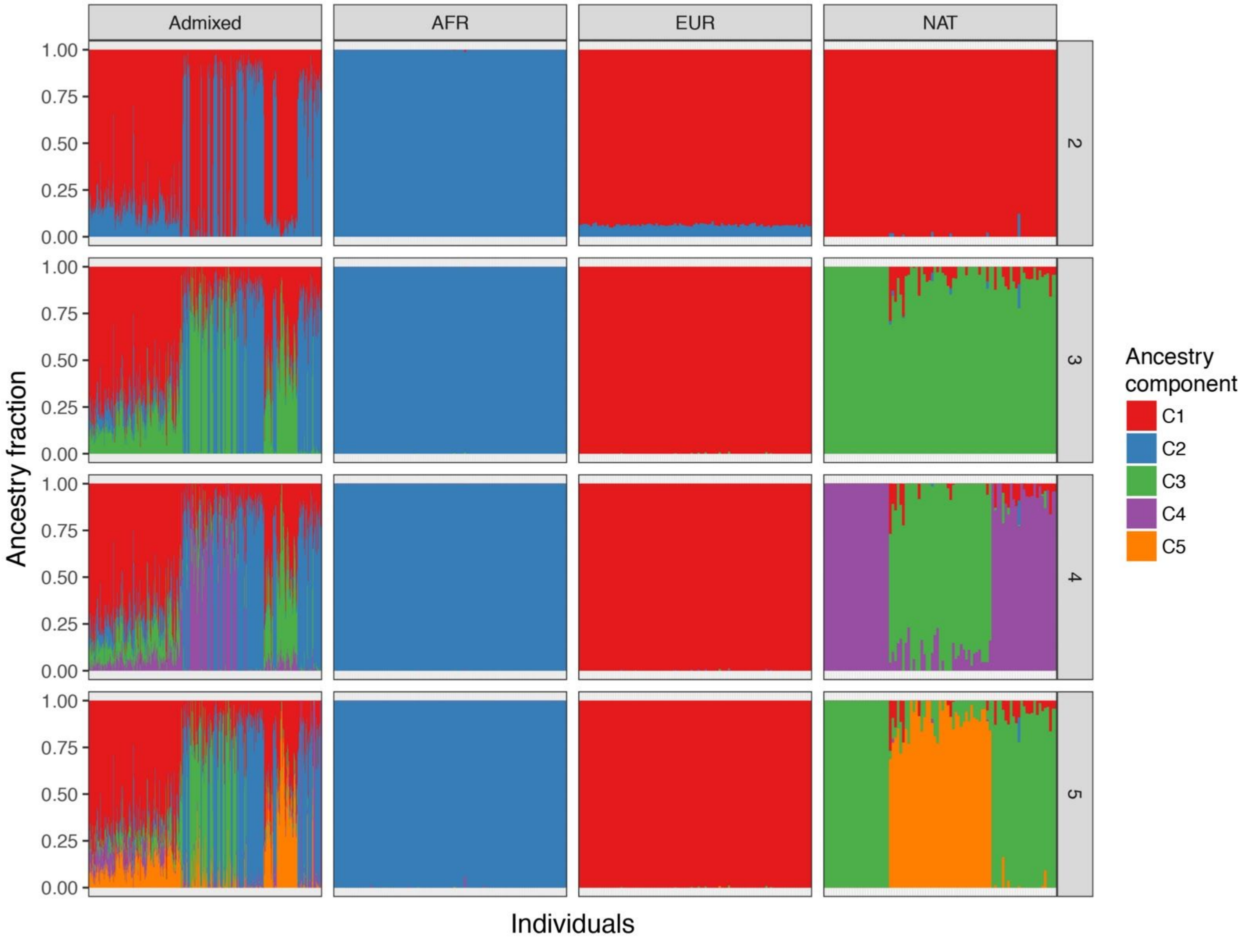


Figure S6. ADMIXTURE cross-validation error for values of K from 1 to 5. Cross-validation error is lowest for k = 3.

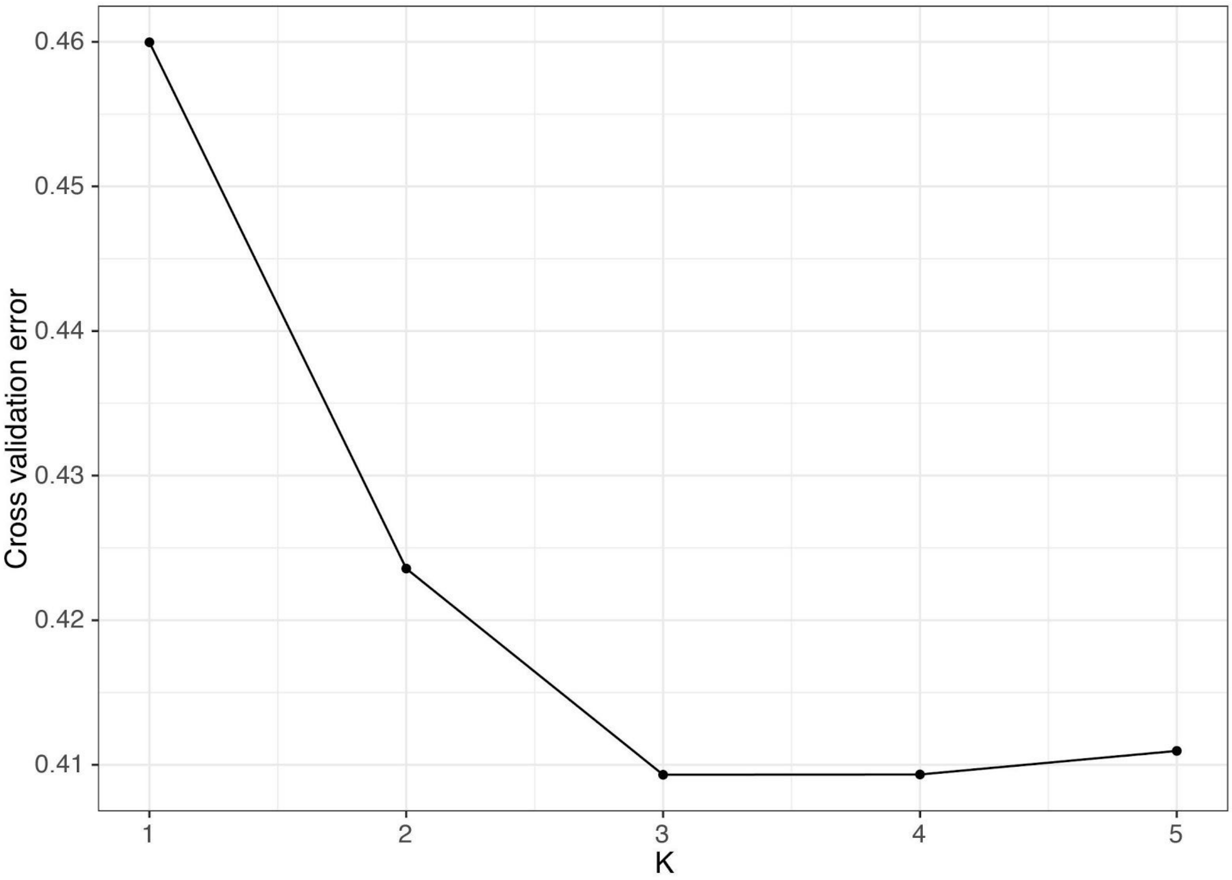


Figure S7. Breakdown of mtDNA haplogroup frequency observed in each admixed population. ACB: African Caribbeans from Barbados; ASW: African Americans from South Western USA; CLM: Colombians from Colombia; MXL: Mexicans from Los Angeles, USA; PEL: Peruvians from Lima, Peru; PUR: Puerto Ricans from Puerto Rico.

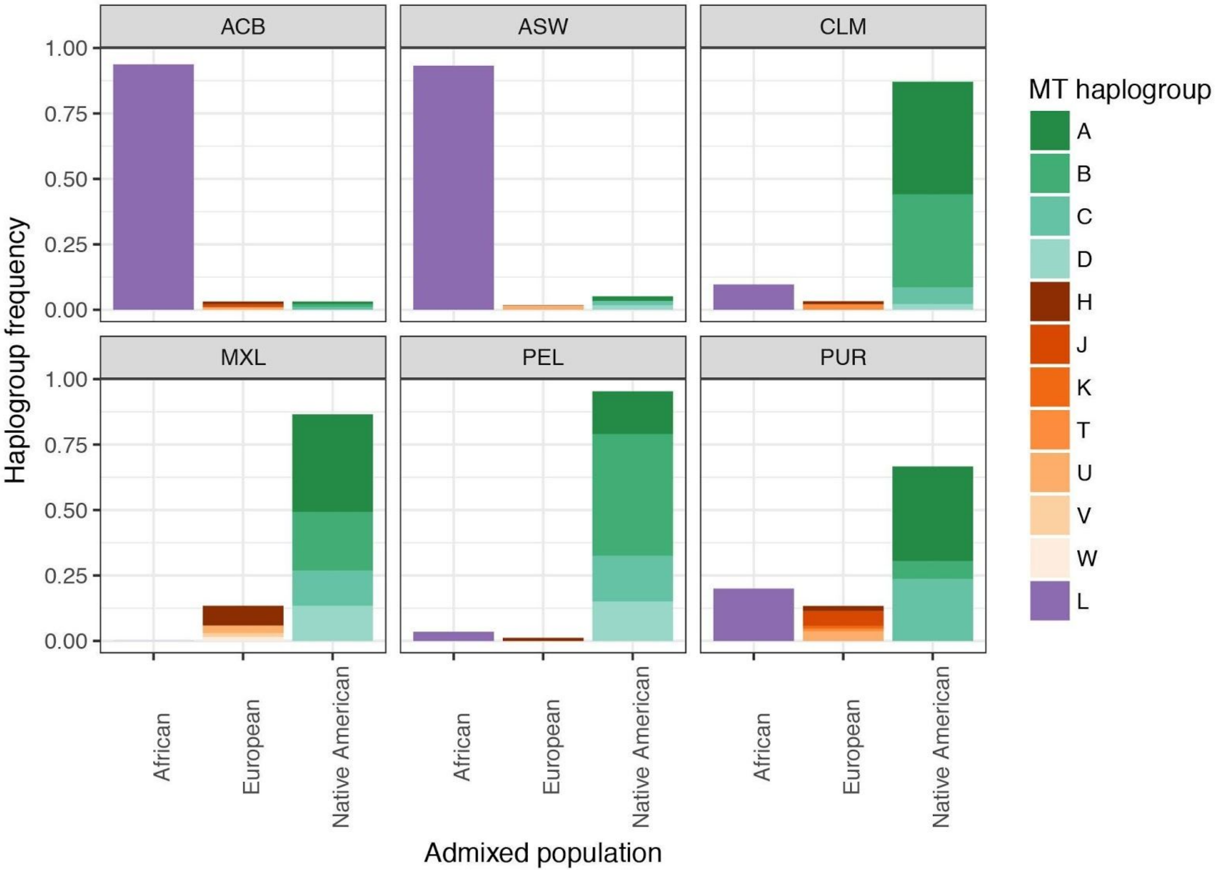


Figure S8. Plot showing concordance between mtDNA copy number calculated from high-coverage and low-coverage alignments. The source of the DNA sample (where available) is indicated. LCL - lymphoblastoid cell line

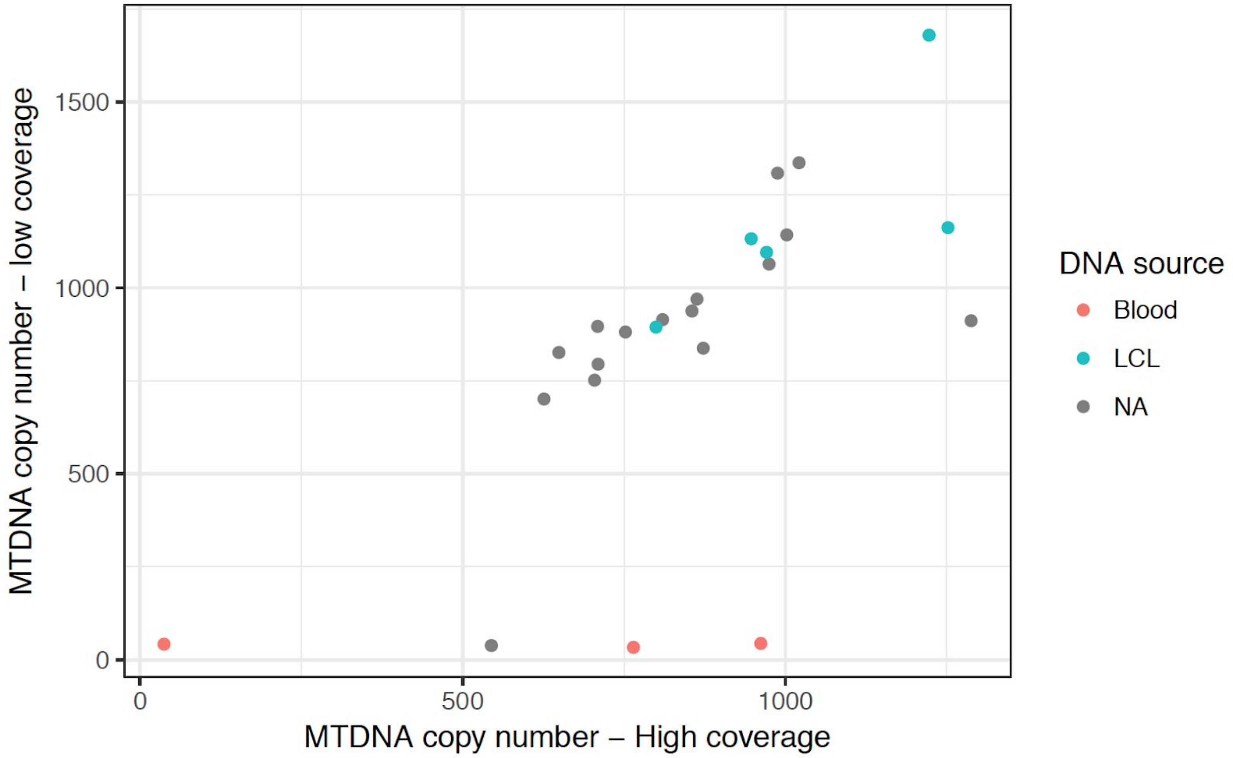


Figure S9. Distribution of mtDNA copy number calculated using low-coverage sequence alignments. There is a clear separation between samples sequenced from lymphoblastoid cell lines (LCLs) and peripheral blood mononuclear cells (PBMCs). In two cases, samples from lymphoblastoid cells appear to be mislabeled as blood.

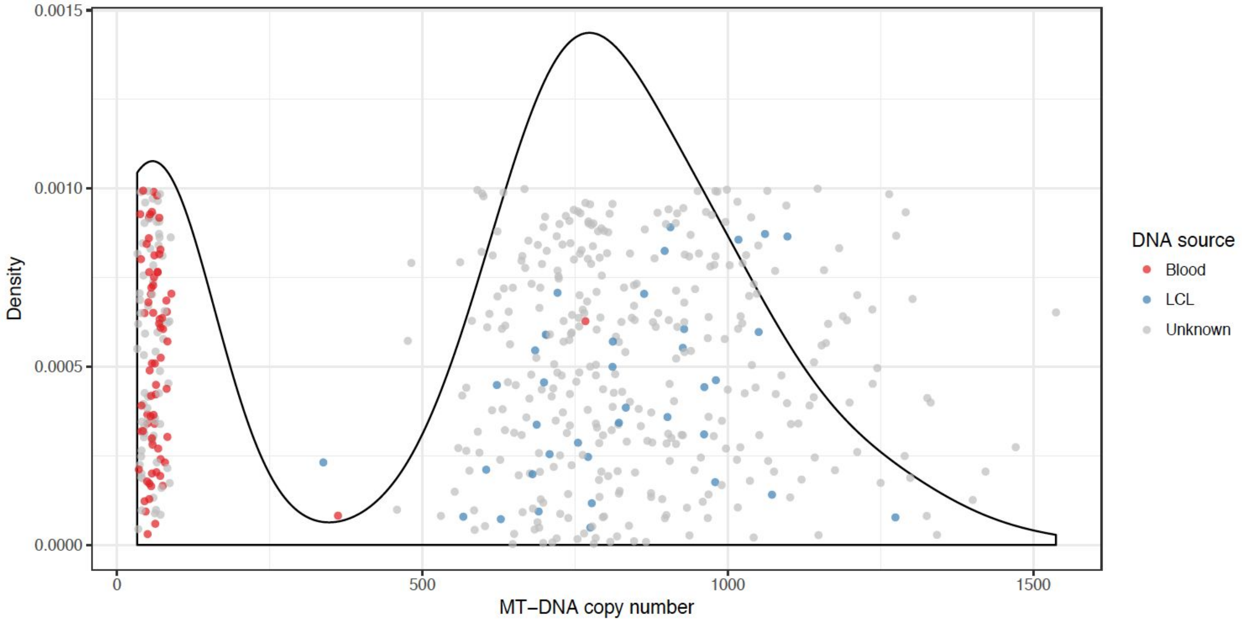


Figure S10. Global ancestry estimated using ADMIXTURE is highly correlated with global ancestry calculated from masked RFMix local ancestry calls. The red line represents y=x and the blue line is the line of best fit.

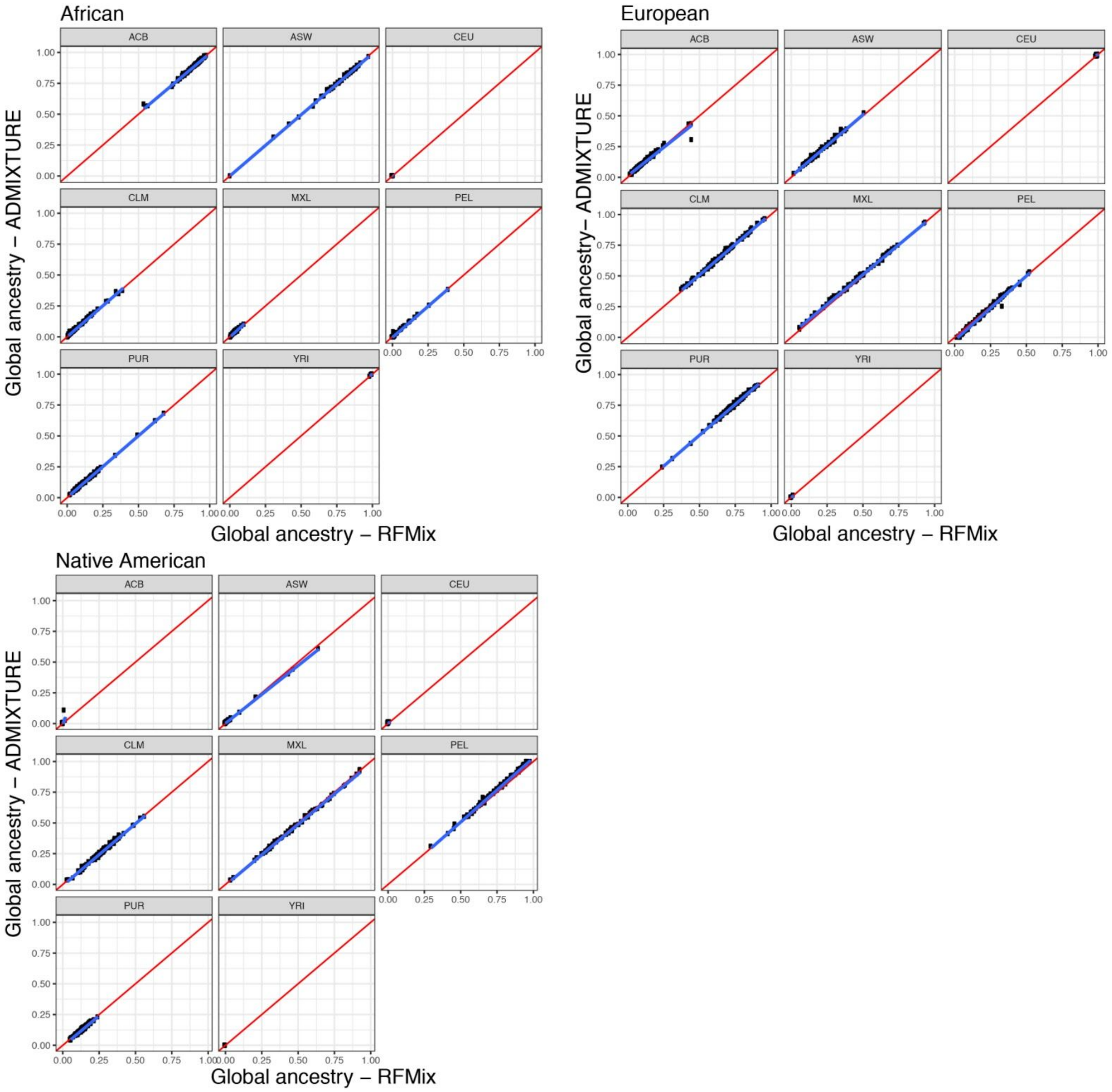


Figure S11. Estimates of male and female contributions from source populations with bootstrapped confidence intervals. ACB: African Caribbeans from Barbados; ASW: African Americans from South Western USA; CLM: Colombians from Colombia; MXL: Mexicans from Los Angeles, USA; PEL: Peruvians from Lima, Peru; PUR: Puerto Ricans from Puerto Rico.

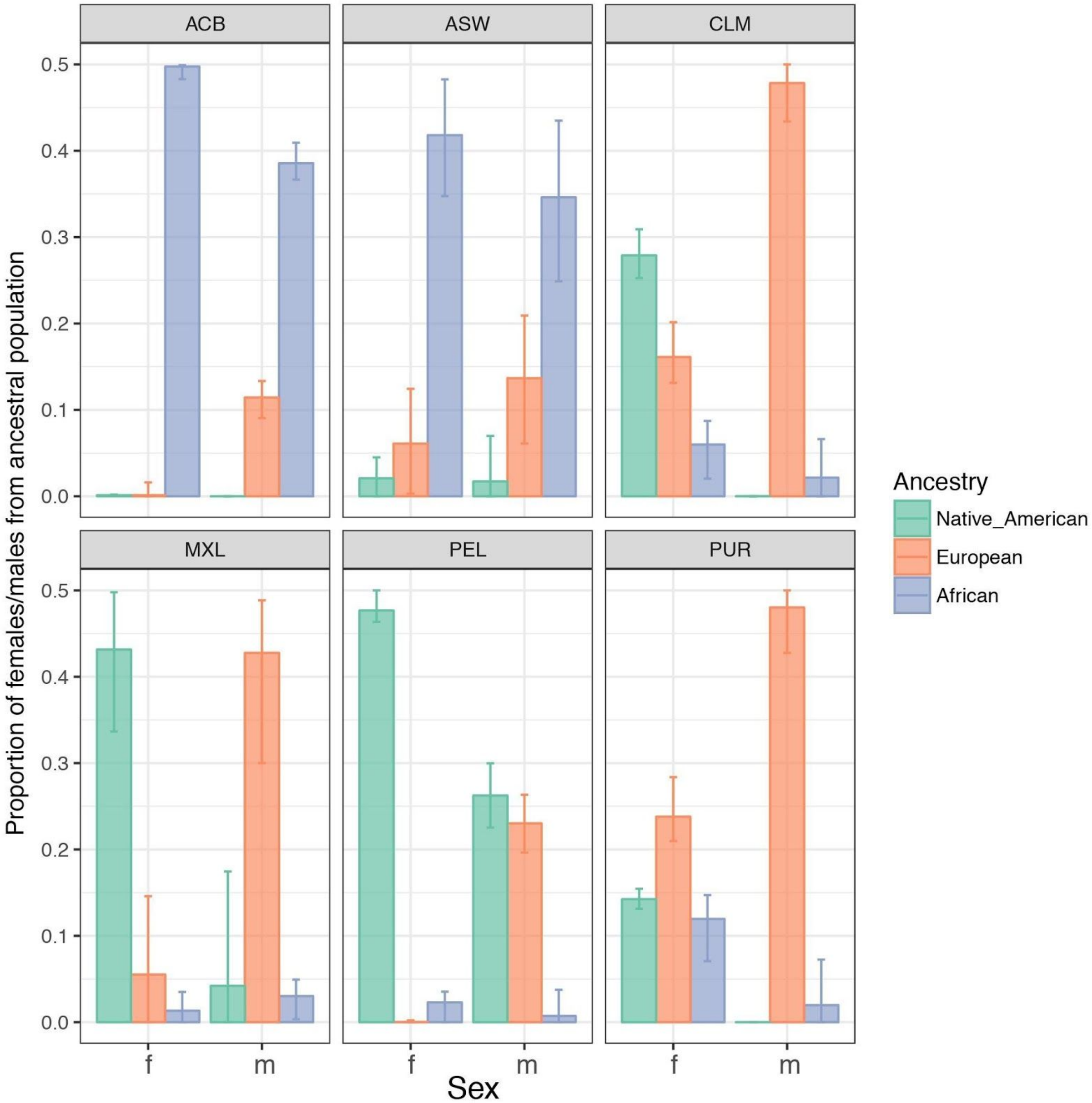
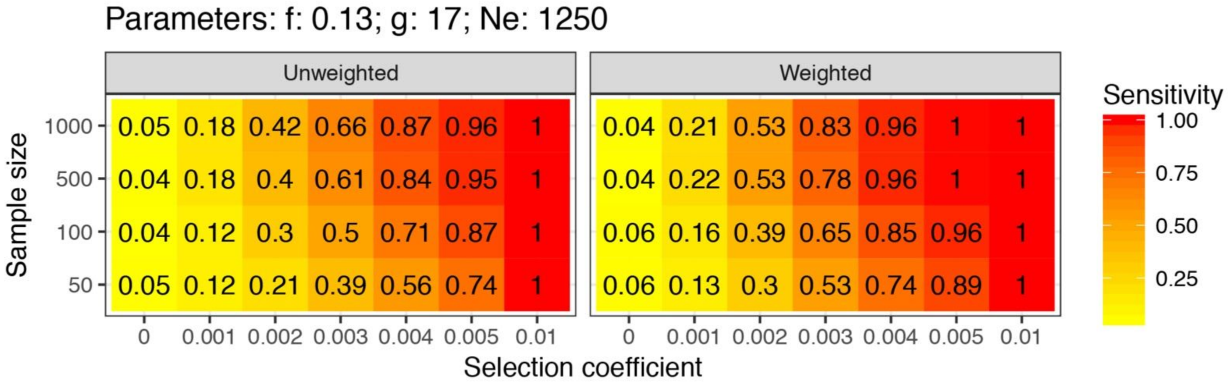


Figure S12. Sensitivity to detect selection with mean deviation in local ancestry. Simulations were run for varying strengths of selection (x-axis) and a range of sample sizes (y-axis). The numbers within each tile shows the proportion of simulations in which selection was detected, i.e., the mean deviation in local ancestry was significantly greater than zero at the 95% level of significance. Neutral simulations were run with a selection coefficient of zero. The left and right panels show results for the unweighted and weighted mean deviation in ancestry, respectively. The parameters were chosen to match the degree of drift experienced by Puerto Ricans.



Unweighted

Sample size	0	0.001	0.002	0.003	0.004	0.005	0.01
1000	0.05	0.18	0.42	0.66	0.87	0.96	1
500	0.04	0.18	0.4	0.61	0.84	0.95	1
100	0.04	0.12	0.3	0.5	0.71	0.87	1
50	0.05	0.12	0.21	0.39	0.56	0.74	1

Weighted

Sample size	0	0.001	0.002	0.003	0.004	0.005	0.01
1000	0.04	0.21	0.53	0.83	0.96	1	1
500	0.04	0.22	0.53	0.78	0.96	1	1
100	0.06	0.16	0.39	0.65	0.85	0.96	1
50	0.06	0.13	0.3	0.53	0.74	0.89	1