

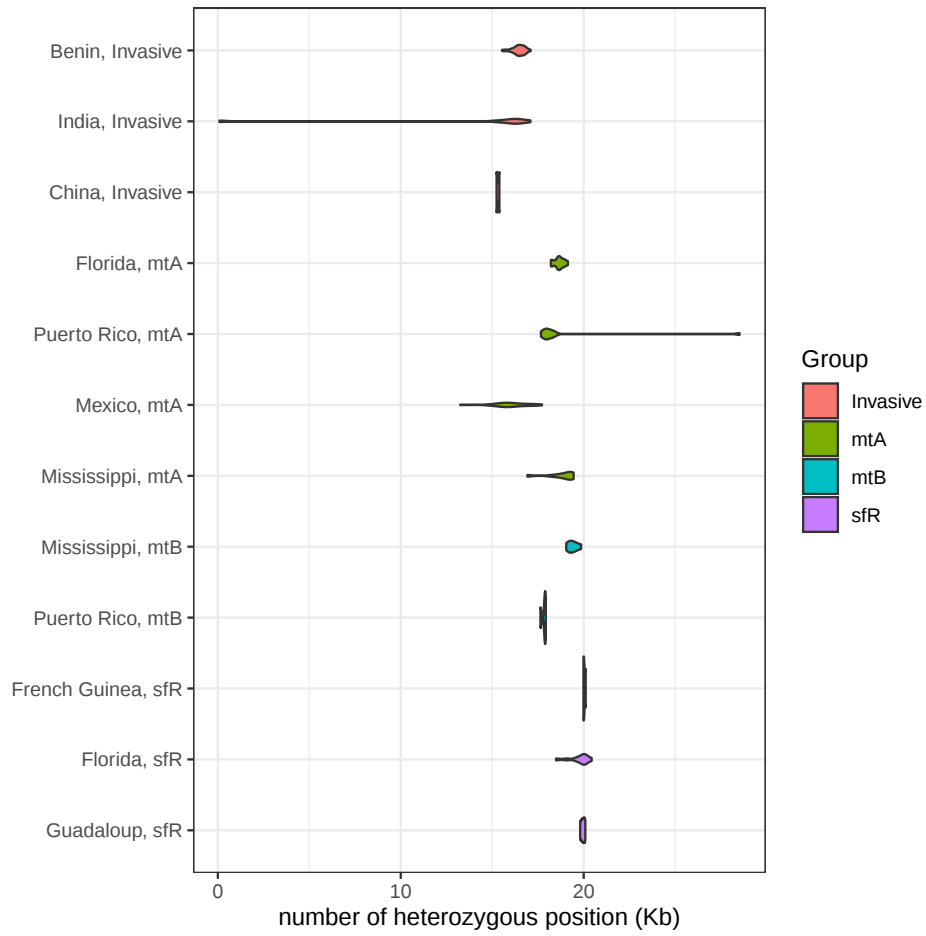


Figure S1. The average number of heterozygous positions calculated for each population

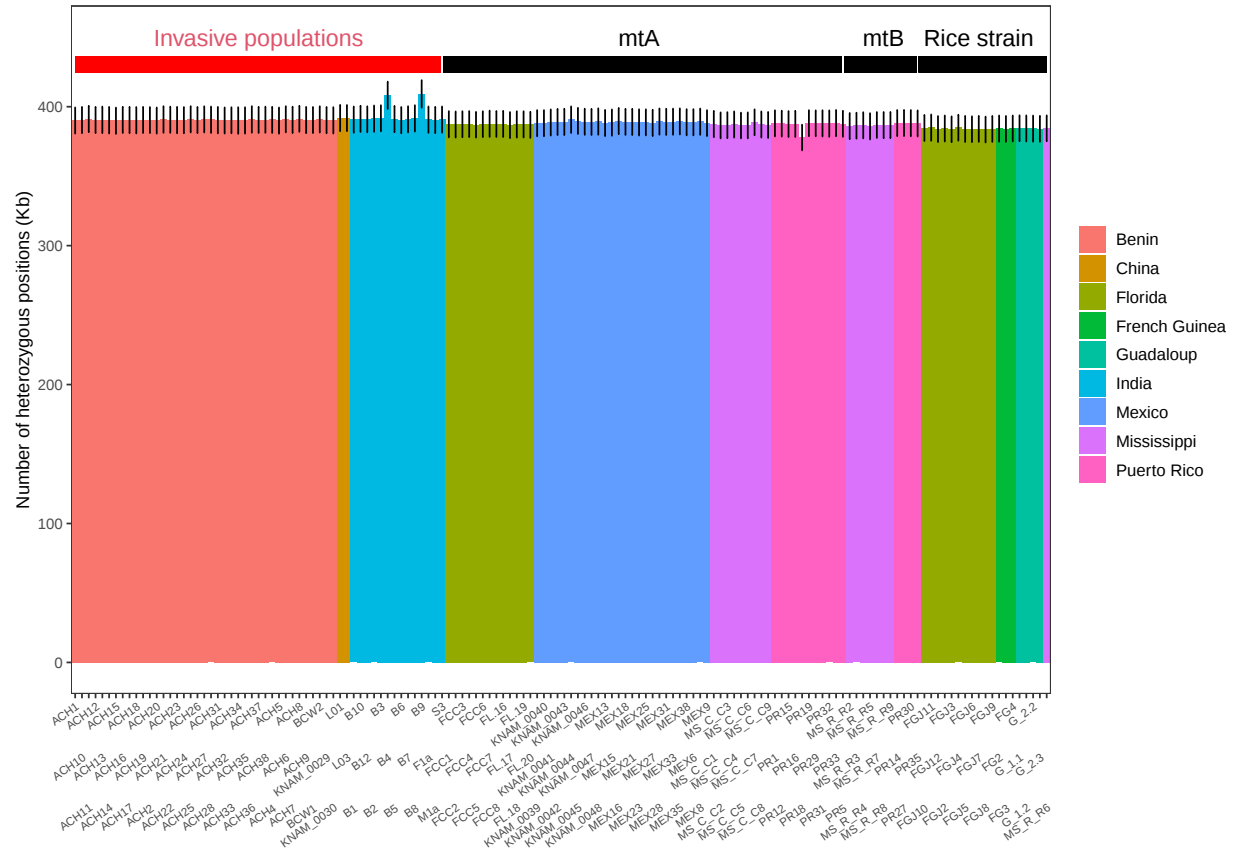


Figure S2. The number of homozygous positions for each sample. The error bars indicate 95% confidence intervals with 1,000 bootstrapping replications resampled from 100kb windows.

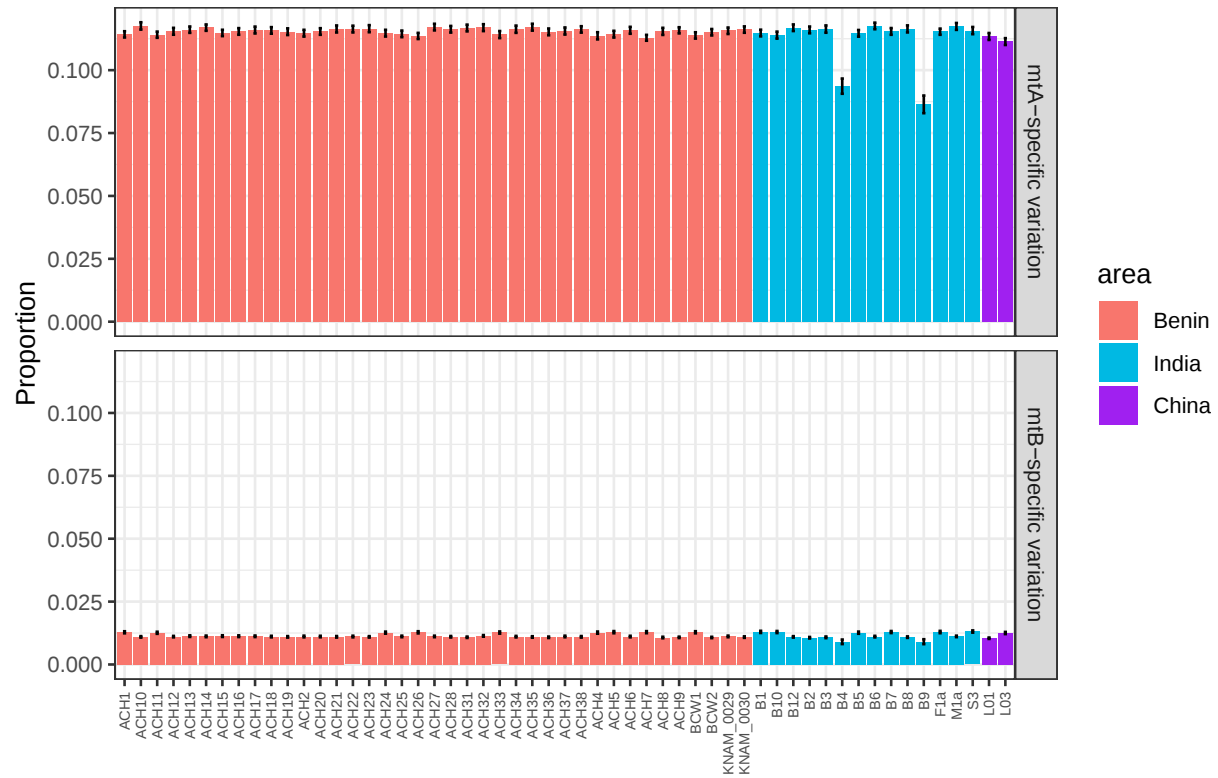


Figure S3. The proportion of mtA-specific (upper) or mtB-specific variations among total SNVs was calculated for each invasive sample. The error bars indicate 95% confidence intervals calculated from resamplings in a 100kb window with 1,000 replications

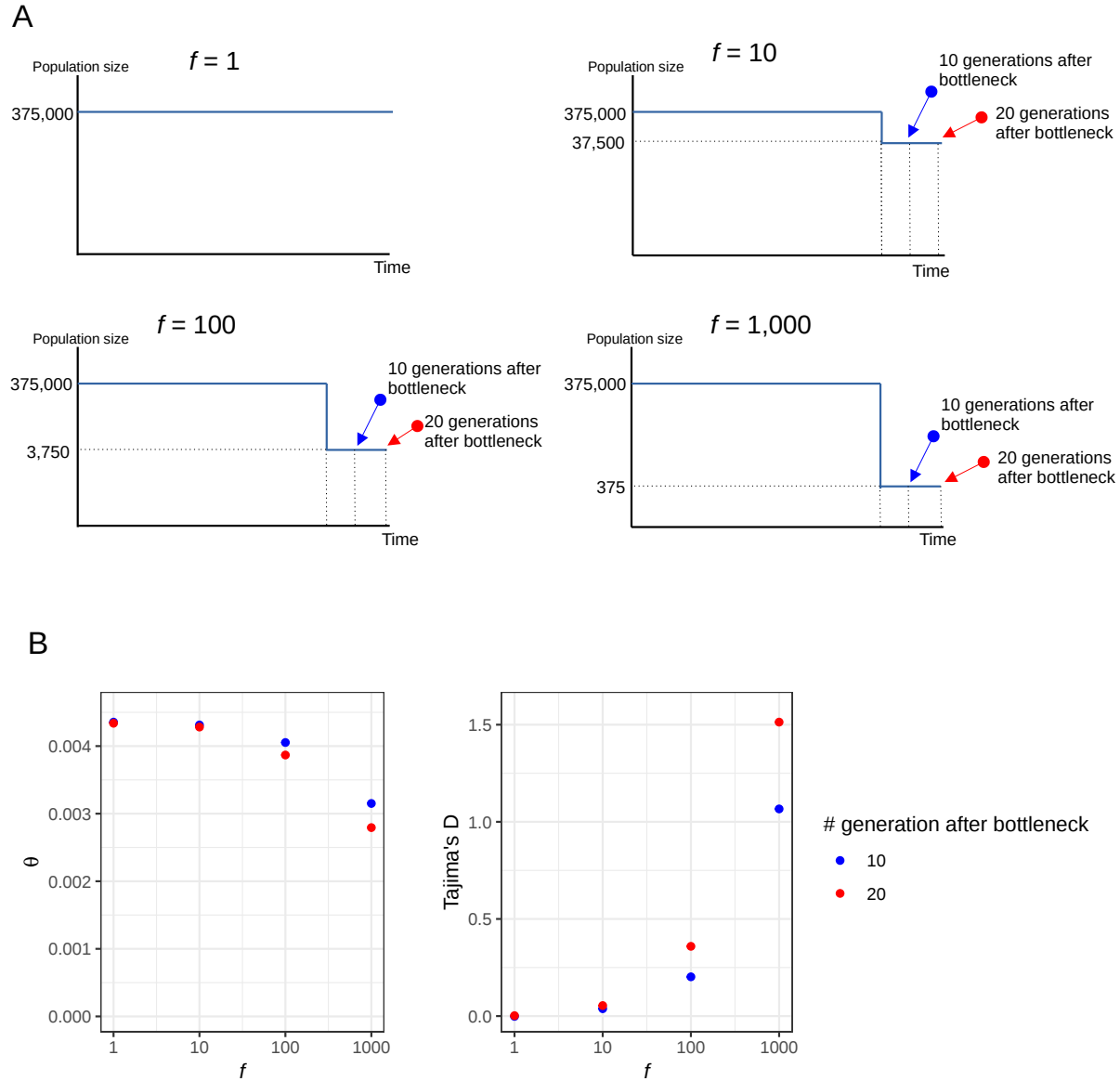


Figure S4. Coalescent simulation with a recent bottleneck and no lag phase. A. The demographic model employed in the simulation involves varying the strength of the bottleneck. The initial effective population size was 375,000, and this population experienced a bottleneck with varying strengths. f represents the relative effective population sizes of the original to bottlenecked populations. B. The plots show correlations between f and Watterson's θ (left) and between f and Tajima's D (right).

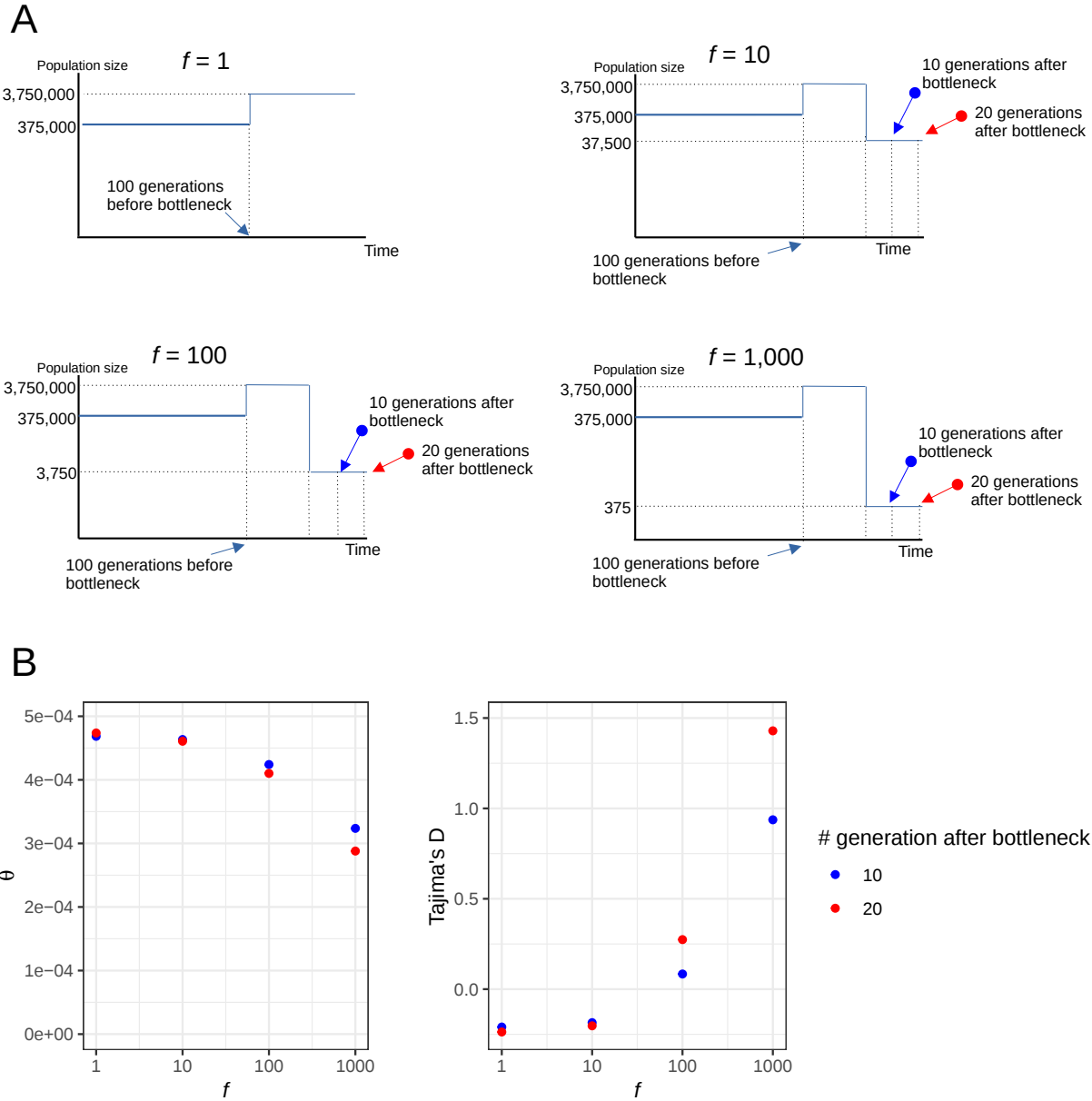


Figure S5. Coalescent simulation with a recent bottleneck preceded by an extreme population growth A. The demographic models used in the simulation involve a 10-fold increase in effective population size 100 generations ago, followed by a bottleneck event 10 generations ago with varying strength. The initial effective population size was 10^6 . f is the relative effective population sizes of the original to bottlenecked populations. B. The plots show correlations between f and Watterson's θ (left) and between f and Tajima's D.

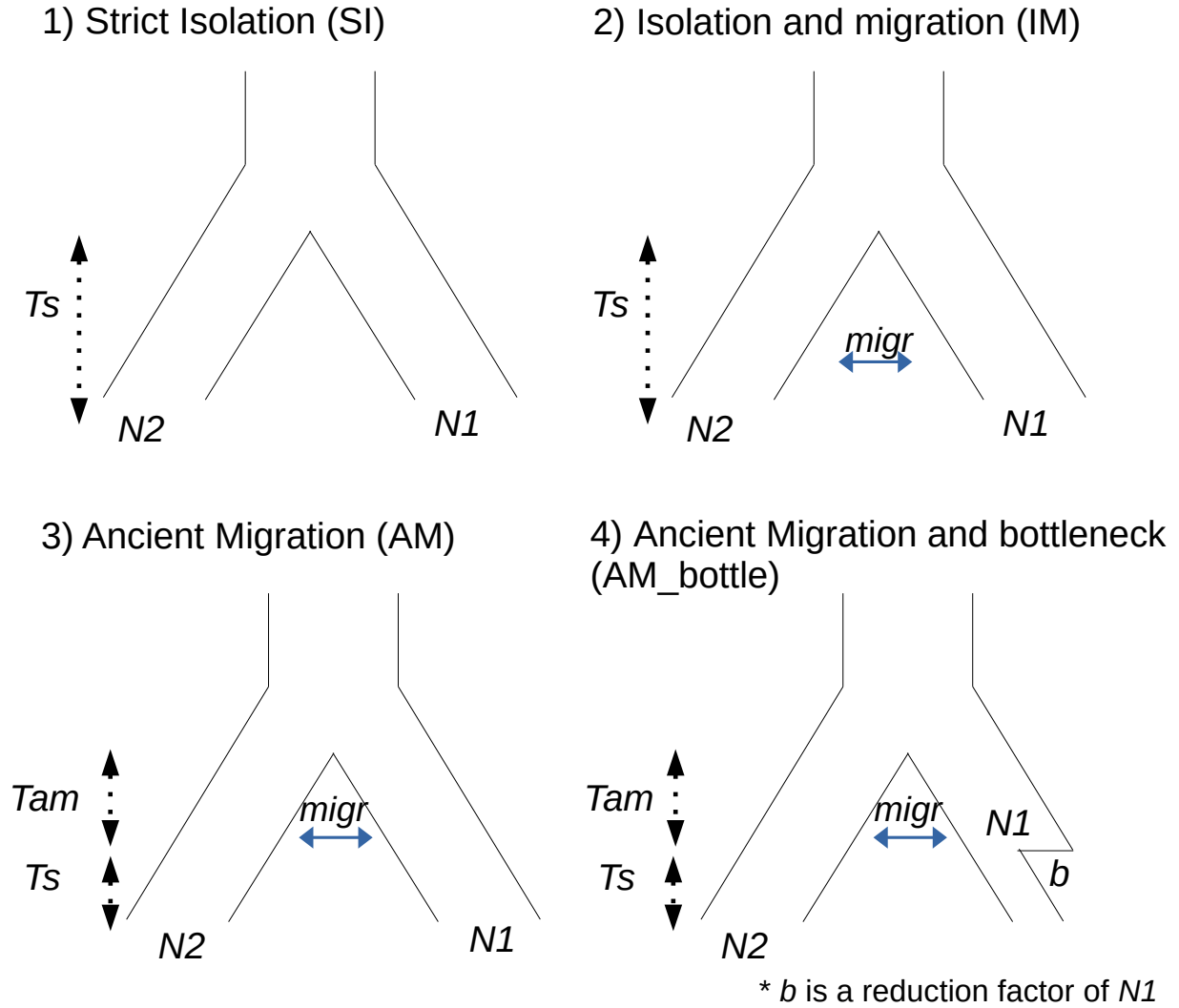
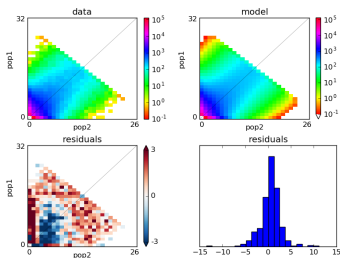


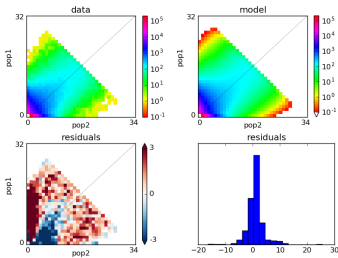
Figure S6. **Demographic models used in diffusion approximation.** 1) The SI model assumes a split of an ancestral population into two populations. 2) The IM model includes bidirectional gene flow between sibling populations. 3) The AM model involves a cessation of gene flow. 4) The AM_bottle model includes a bottleneck at the cessation of gene flow. $N1$ and $N2$ are the population sizes of splitted population. Ts is the split time for SI and IM and the time without gene flow in AM and AM_bottle. $migr$ is the rate of gene flow. $Tam + Ts$ is the time of split between $N1$ and $N2$ in AM and AM_bottle. b is a reduction factor of $N1$, such that the population size changes from $N1$ to $N1/b$ after bottleneck.

A. Benin

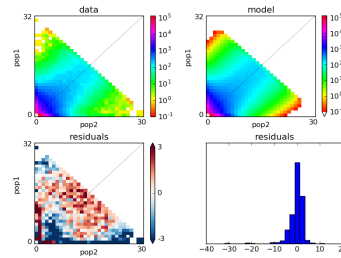
Florida



Mississippi

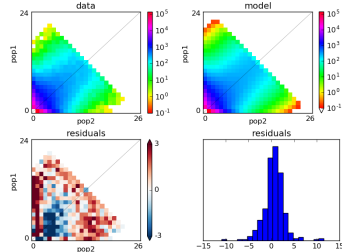


Puerto Rico

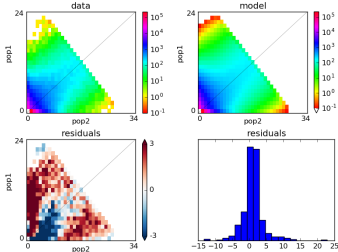


B. India

Florida



Mississippi



Puerto Rico

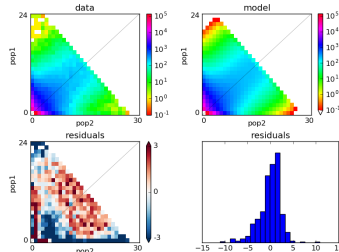


Figure S7. The joint site frequency spectrum of real data (upper left) alongside the prediction with estimated demographic parameters (upper right), the residuals between real data and prediction (lower right), and the histogram of residuals (lower left), calculated from the SNV dataset from either (A) Benin or (B) India and one native population (Florida, Mississippi, or Puerto Rico).

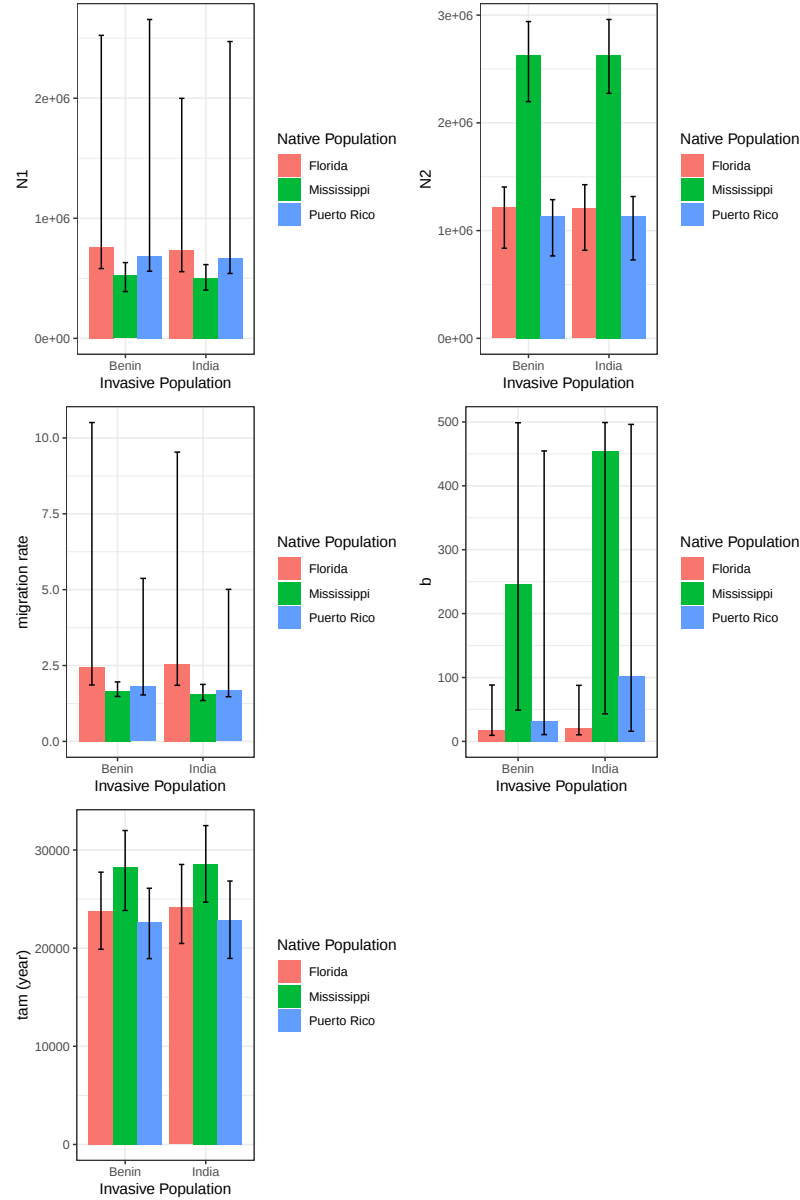


Figure S8. The estimated parameters of the best models determined by the likelihood ratio test (See Table S1). $N1$ and $N2$ represent the population sizes of invasive and native populations, respectively. The parameter b is a reduction factor for $N1$, indicating that the population size changes from $N1$ to $N1/b$ after a bottleneck. The migration rate ($migr$) is the rate of gene flow between $N1$ and $N2$. tam is the time between the split (between $N1$ and $N2$) and the cessation of gene flow (please see Fig. S6). Error bars indicate 95% bootstrapping confidence intervals, which were calculated using scaffold-resampling with 1,000 replications.

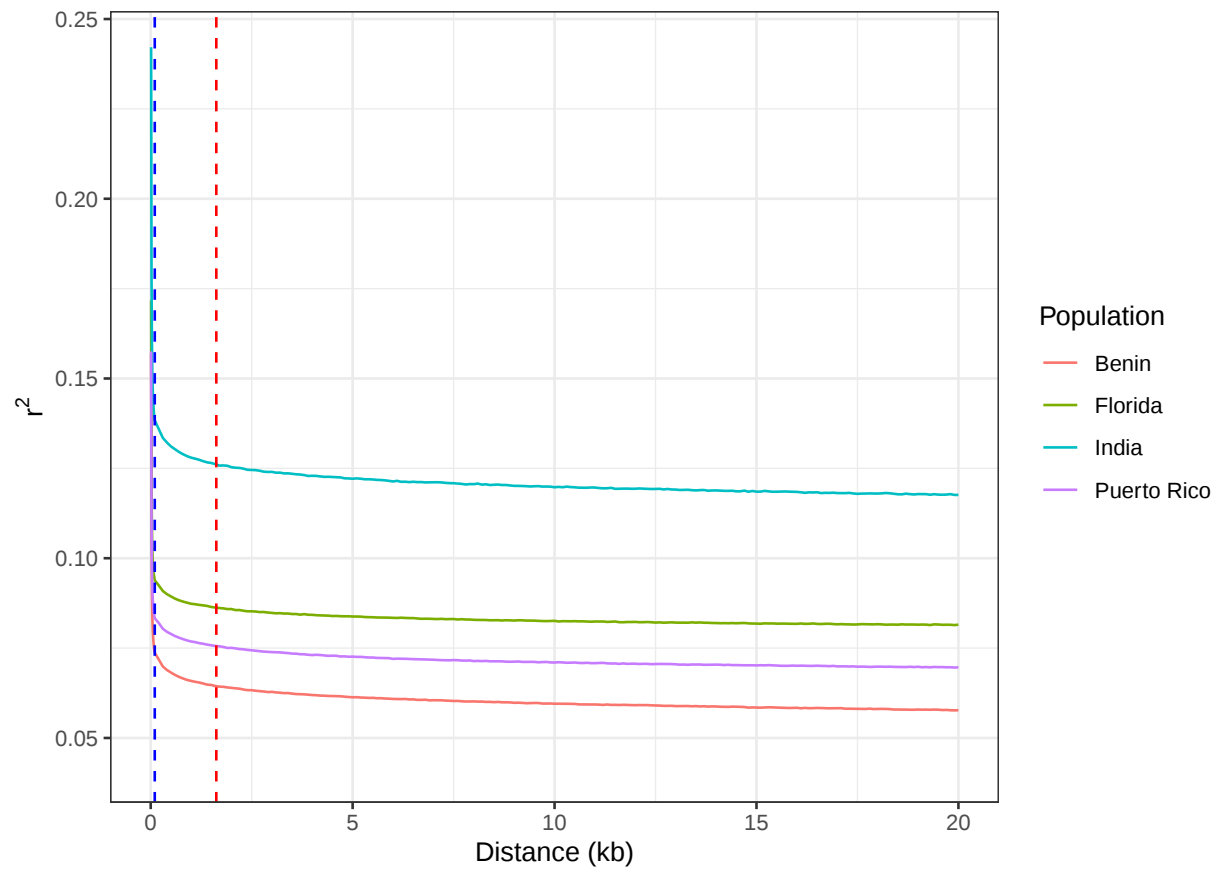


Figure S9. The decay curve of linkage disequilibrium. The dotted red and blue vertical bar indicate the average SNV density in the genome and a distance of 100bp, respectively.

Table S1. Log-likelihood ratio test to determine the best model for each SNV dataset, comprising an invasive population from Benin or India and a native population from Florida, Mississippi, or Puerto Rico. The SI, IM, AM, and AM_bottle models represent Strict Isolation, Isolation-with-Migration, Ancient Migration, and Ancient Migration and Bottleneck, respectively. The D statistic is two times the difference in likelihood between the two models, while the degree of freedom corresponds to the difference in the number of parameters between the models. P-values were calculated by fitting the D statistic with the degree of freedom to the chi-square distribution.

Population		Alternative		Null		D statistic	Degree of freedom	P-value
Invasive	Native	Model	Likelihood	Model	Likelihood			
Benin	Florida	AM_bottle	-2272.2	AM	-4544.4	4544.4	1	~ 0
		AM	-4544.4	IM	-7257.6	5426.4	1	~ 0
		IM	-7257.6	SI	-31798.2	49081.2	1	~ 0
	Mississippi	AM_bottle	-4173.1	AM	-7415.8	6485.4	1	~ 0
		AM	-7415.8	IM	-9969.5	5107.4	1	~ 0
		IM	-9969.5	SI	-48179.5	76420	1	~ 0
	Puerto Rico	AM_bottle	-2814.0	AM	-4838.4	4048.8	1	~ 0
		AM	-4838.4	IM	-7552.8	5428.8	1	~ 0
		IM	-7552.8	SI	-29882.5	44659.4	1	~ 0
India	Florida	AM_bottle	-1896.0	AM	-4181.1	4570.2	1	~ 0
		AM	-4181.1	IM	-6866.2	5370.2	1	~ 0
		IM	-6866.2	SI	-27694.8	41657.2	1	~ 0
	Mississippi	AM_bottle	-3597.9	AM	-6723.8	6251.8	1	~ 0
		AM	-6723.8	IM	-9464.1	5480.6	1	~ 0
		IM	-9464.1	SI	-42211.2	65494.2	1	~ 0
	Puerto Rico	AM_bottle	-2231.7	AM	-4362.9	4262.4	1	~ 0
		AM	-4362.9	IM	-7029.9	5334	1	~ 0
		IM	-7029.9	SI	-25887.0	37714.2	1	~ 0