

SUPPLEMENTARY FIGURES

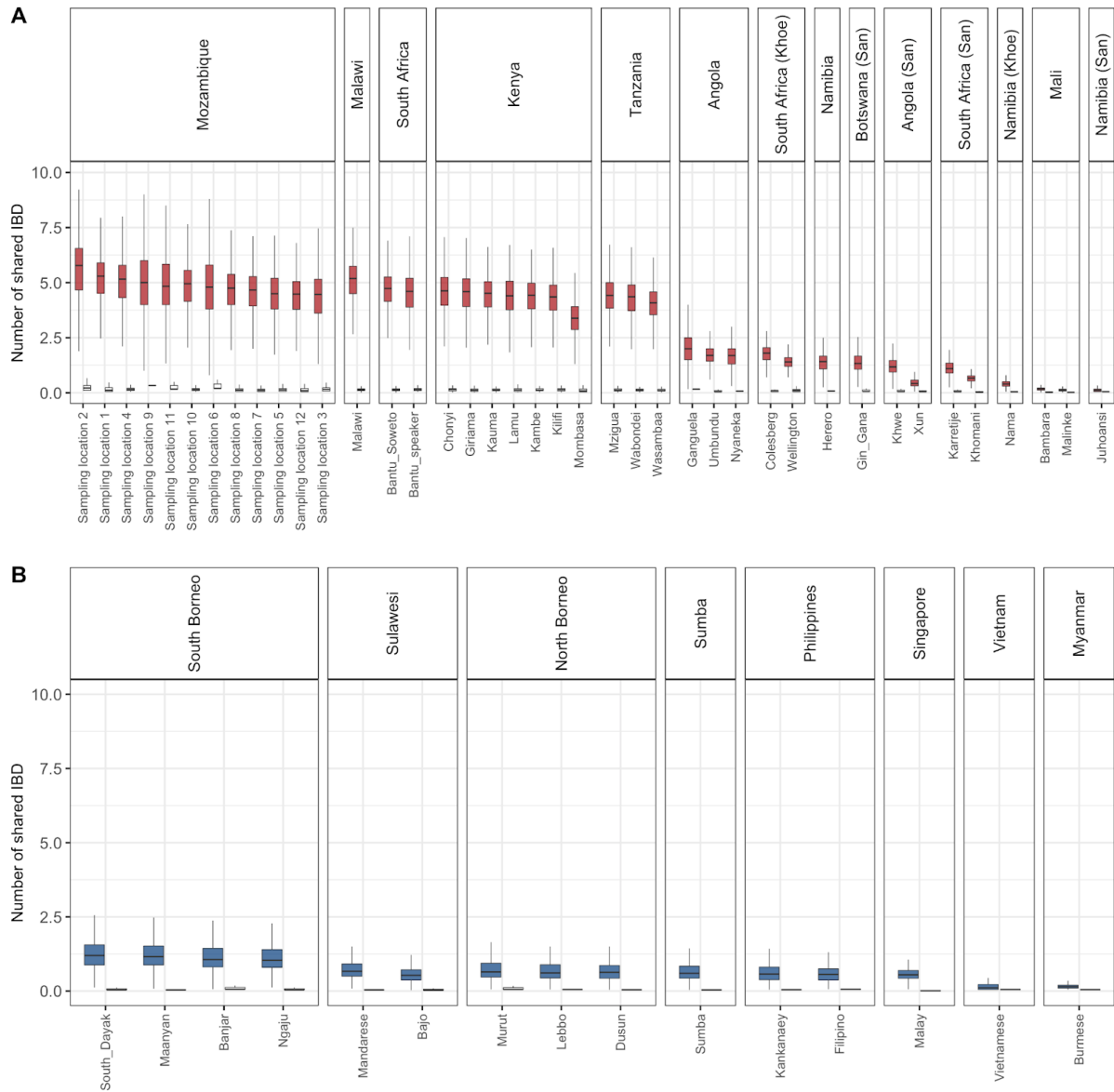


Figure S1. Boxplot of the distribution of the average number of IBD segments shared, in the density dataset, between a Malagasy individual and an individual from Africa in red (*upper panel*) or from Asia in blue (*lower panel*). Related to Figure 1. For each population IBDs are divided into 2 length bin $<5\text{cm}$ (left) and $>5\text{cm}$ (right). Comparisons were made using the density dataset, comprising 311,411 SNPs and 3168 individuals. In this study, we classified the Khoehoe-San populations based on the designations previously proposed, when referring collectively to the traditional hunter-gatherers of southern Africa as San; and when referring to traditional pastoralist groups as Khoehoe. Regarding the linguistic component of these populations, we notice that the Khwe population speak a click language from the Khoehoe subfamily.

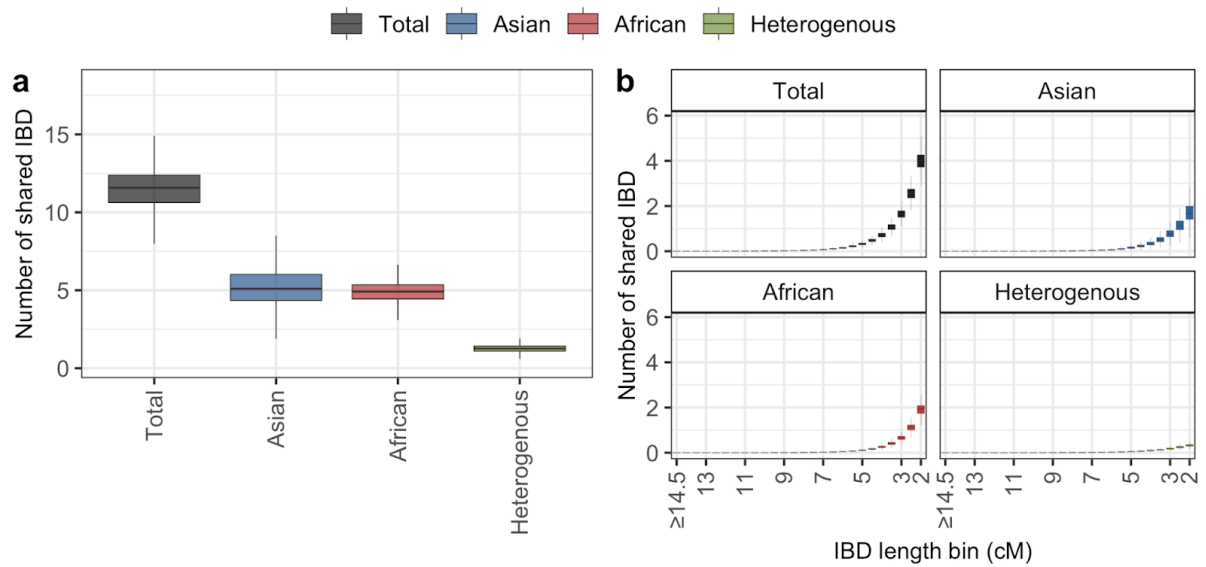


Figure S2. Distribution of the average number of shared IBD segments between two Malagasy individuals, according to the inferred continental ancestry of each IBD segment. A) boxplot of the global number of shared IBD B) Distribution of shared IBD according their size. Related to Figure 1. We estimated the ancestry of each IBD segment by mapping its coordinates to local ancestry assignments from sample haplotypes. The ancestry proportion was calculated using the number of Asian (ASI) sites divided by the total number of SNPs spanning the IBD region in both haplotypes. IBD segments were classified as African (AFR SNPs $\geq 90\%$), Heterogenous ($10\% < \text{ASI SNPs} < 90\%$) or Asian (ASI SNPs $\geq 90\%$).

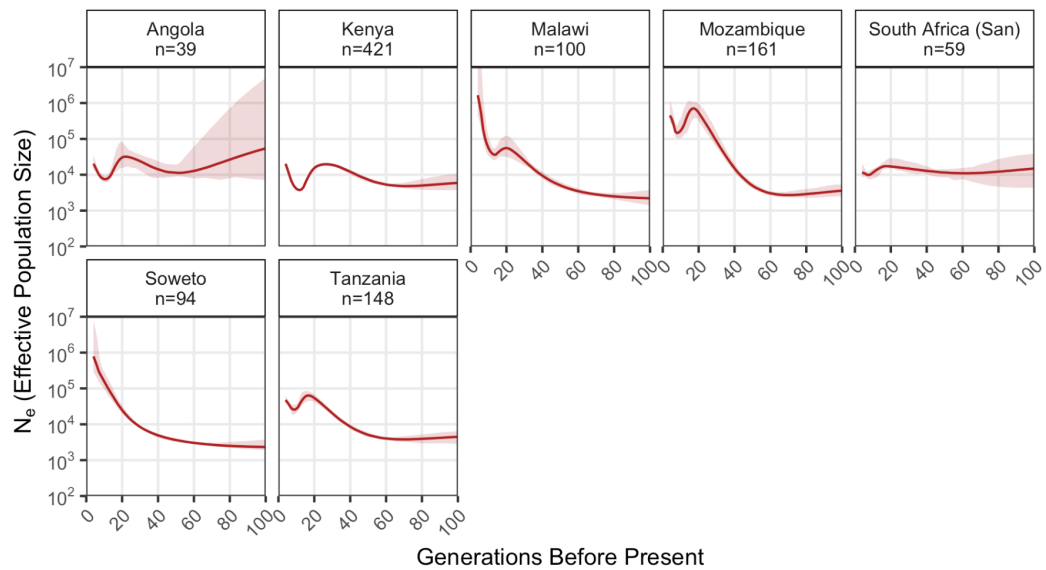
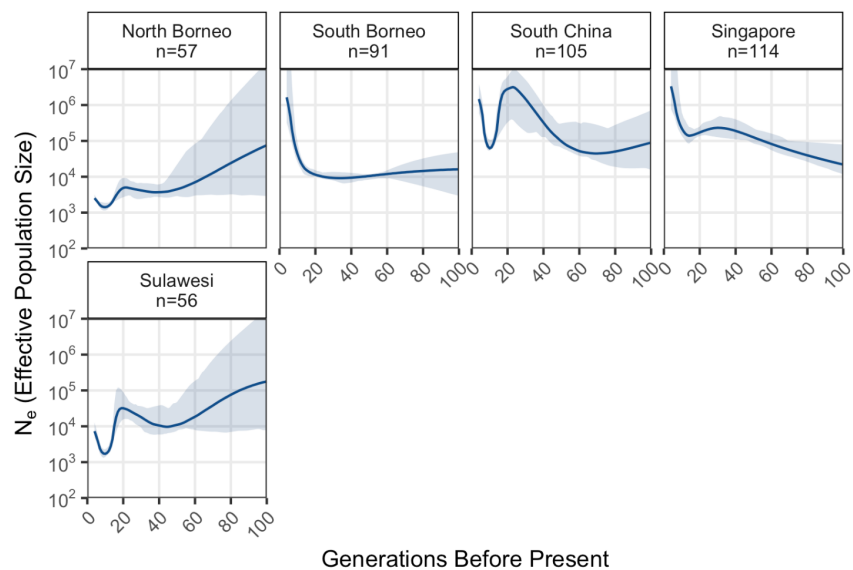
A**B**

Figure S3. Estimated effective population size (N_e) over the last 100 generations for the African populations (A) and Asian populations (B) analysed in this study. Related to Figure 2. The line in each plot shows the estimated effective population size, and the coloured regions are bootstrap 95% confidence intervals. The y-axes (effective population size) are plotted on a logarithmic scale.

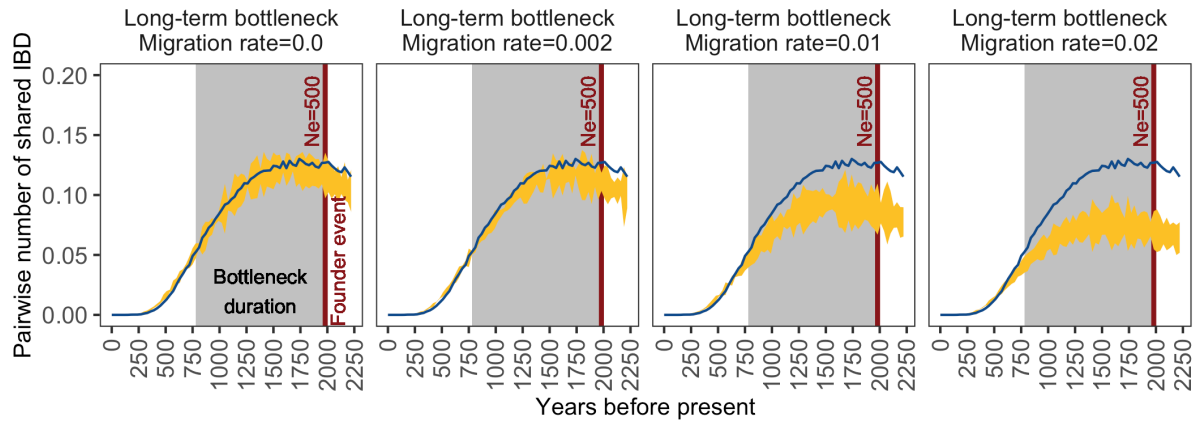


Figure S4. Effect of gene flow during the long-term bottleneck on the IBD segments of Asian origin shared in Madagascar. Related to Figure 3. Each panel shows the result for a Long-term bottleneck scenario according to different simulated migration rates. The x-axis represents the age of IBD haplotypes (in years before present). The number of segments shared per pair is shown on the y-axis. Yellow curves show the mean pairwise IBD-sharing at each generation ± 5 standard deviations, based on 3 simulated replicas. Blue lines in each panel represent the observed pairwise IBD-sharing of Asian segments in the Malagasy population.

SUPPLEMENTARY TABLES

ABC algorithm	Ne_startBOT			Ne_endBOT			Tdur_BOT			Mig_BOT		
	2.5%	Mean	97.5%	2.5%	Mean	97.5%	2.5%	Mean	97.5%	2.5%	Mean	97.5%
Neural Network	471	487	500	365	367	368	66	66	66	0.0021	0.0021	0.0021
Local Linear Regression	109	930	3560	505	615	790	46	65	74	0.0000	0.0063	0.0197
Rejection	52	534	3608	165	1603	8457	22	52	73	0.0004	0.0071	0.0176

Table S1. Estimated demographic parameters under the Bottleneck event model based on ABC method (tolerance rate of 5 %). Related to figure 3.