

Supplementary File 1 Simulation of genetic polymorphism data

The neutral coalescent model simulation software MS (Hudson 2002) was used to produce sequence level single nucleotide polymorphism (SNP) data based on scaled population parameters under the infinite sites model of mutation. The parameters θ and ρ are the scaled population mutation and recombination rate, respectively (Hudson 1983). Values for parameters θ and ρ were chosen to mimic the simulation of 1Mb region of the genome. Using $\theta=400$ and $\rho=400$, with a region of 1Mb and an original (pre-bottlenecks) effective population (N_e) size of 10,000 the corresponding mutation rate $\mu=10^{-8}$, and a recombination rate of 1cM/Mb. We simulated data representing 1Mb chromosome segments for a total of 16 demographic scenarios, 13 single, non-splitting, and three multiple, splitting populations, with each of the multiple population scenarios comprising of a three-population demographic event. Each of the 10 scenarios was independently replicated 100,000 times with results averaged across the replicates. The genetic data produced by MS software allows us to evaluate the properties of population demographic estimators such as N_e , T_F and T_{LD} using populations simulated under different demographic scenarios. The following procedures were applied to each replicate to extract genetic data then used to calculate the relevant population demographic estimators.

- 1) A sample of 500 individuals was randomly selected from the population, with each individual represented by two chromosome segments. For multiple population scenarios a sample of 500 was selected for each of the diverged populations. Simulated genetic data comprised of sequence level single nucleotide polymorphisms (SNPs) segregating within the population of 500 individuals.

- 2) Any SNPs with a minor allele frequency (MAF) less than 5% were removed from the datasets. Additionally, for the multiple population scenarios any SNPs not segregating in all three populations were removed.
- 3) Estimates of N_e and divergence time (T) are based on the recombination distance between SNPs. SNP pairs were binned into one of 50 recombination distance categories with incremental upper boundaries of 0.005cM up to 0.25cM. We did not include the first category, LD observations where pairs of SNPs were separated by $<0.005\text{cM}$ (see **Materials and Methods** in main text). For each population scenario separately and pair of SNPs separated by $<0.25\text{cM}$, we calculated linkage disequilibrium (LD) levels by the correlation (r_{LD}) and squared correlation (r^2_{LD}) in genotype frequencies.
- 4) Calculation of estimators, N_e (all population scenarios), T_F and T_{LD} (multiple population scenarios only) followed procedures detailed in **Materials and Methods** (main text).
- 5) Estimators were calculated for each replication individual and the means and standard errors calculated from the 100,000 independent replications.

Supplementary References

- Hudson, R.R. 1983. Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology* **23**(2): 183-201.
- . 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**(2): 337-338.
- Stringer, C. 2002. Modern human origins: progress and prospects. *Philos Trans R Soc Lond B Biol Sci* **357**(1420): 563-579.
- Sved, J.A., McRae, A.F., and Visscher, P.M. 2008. Divergence between human populations estimated from linkage disequilibrium. *Am J Hum Genet* **83**(6): 737-743.