Supplementary Data

Supplementary Text

To demonstrate the performance of PReFerSim we compared its output to predictions made from existing computational approaches. First, we compared the site frequency spectra (SFSs) from PReFerSim to the SFS generated from diffusion theory as computed using the Dadi software package (Gutenkunst *et al.*, 2009; Figure 1). Specifically, this comparison used a population expansion model where an ancestral population of size 10,000 individuals expanded to 40,000 individuals 3000 generations ago. We used the gamma distribution of selective effects inferred from the African American data in Boyko *et al.* (2008). Additionally, the population scaled mutation rate (θ) was set to 2000, when scaled by the ancestral population size.

Second, we compared the predictions for the ages of deleterious alleles obtained from PReFerSim to the ages predicted by diffusion theory. Specifically, we compared estimates of allele ages (in number of generations) of a 1% frequency variant in a population with a constant size of 10,000 individuals, with theoretical estimates from Table 3 of Maruyama (1974). Overall, we found that the predictions between PReFerSim and theory were reasonably comparable.

It is noteworthy that we found a relatively large standard deviation for the age of neutral alleles. This is consistent with previous theoretical results (Li, 1975). Intuitively, the large standard deviation in the age of neutral alleles reflects the fact that a neutral allele could have a low frequency because it appeared recently or because it arose further back in time and decreased from a higher frequency (Slatkin and Rannala, 2000). On the other hand, deleterious alleles under stronger selective pressure have a lower probability of having reached a higher frequency and then decreased in frequency (as seen in the mean age of deleterious alleles in Supplementary Table 1). This effect is the main contributor to the smaller standard deviation seen for alleles under selection.

Supplementary Tables

Supplementary Table 1: Comparison of theoretical estimates of allele age (generations) with the mean (and standard deviation) of allele age estimates from PReFerSim.

4Ns	0	-10	-50	-100
Theoretical estimates*	1840	1180	632	452
PReFerSim estimates	1850.62 (5252.34)	1201.57 (2030.91)	657.61 (653.80)	468.77 (377.34)

References

Boyko,A.R. *et al.* (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet*, **4**, e1000083.

Gutenkunst,R.N. *et al.* (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.*, **5**, e1000695.

Li,W.H. (1975) The first arrival time and mean age of a deleterious mutant gene in a finite population. *Am. J. Hum. Genet.*, **27**, 274–286.

Maruyama,T. (1974) The age of an allele in a finite population. *Genet. Res.*, **23**, 137–143.

Slatkin,M. and Rannala,B. (2000) Estimating allele age. *Annu. Rev. Genomics Hum. Genet.*, **1**, 225–249.