

Figure S1: The expected pairwise coalescent time, in units of $2N$ generation, on the background of a derived neutral allele with frequency f in the population. The dashed $x = y$ line shows our approximation $\mathbb{E}[T_2] = 2Nf$. The red dots show the mean of our pairwise coalescent simulations featuring an explicit stochastic trajectories. The solid line shows the analytical expectation from GRIFFITHS (2003), equation 93, $\mathbb{E}[T_2] = \int_0^1 x/(f(x + (1-x)f))(2-f + 2(1-f)\log(1-f)/f)dx$.

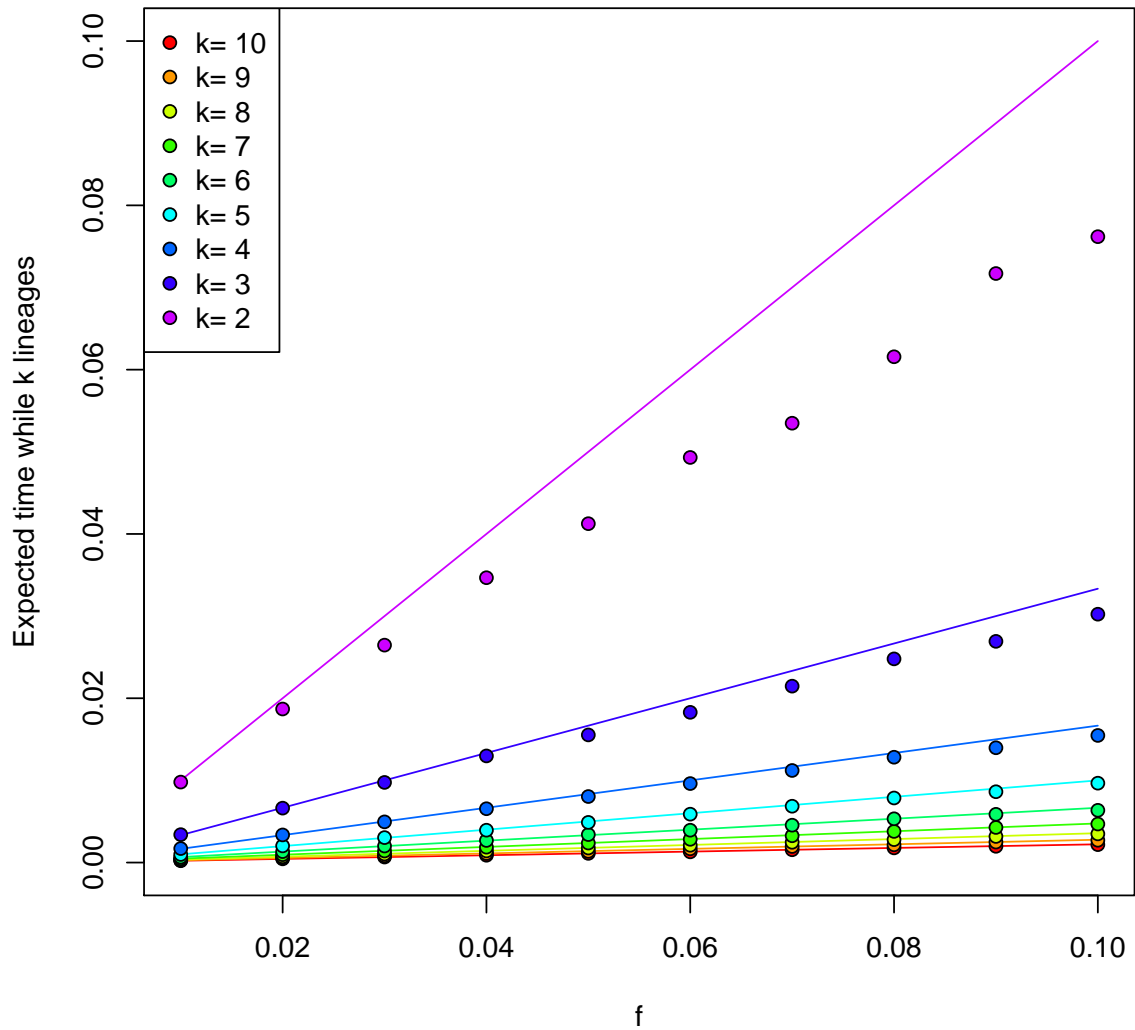


Figure S2: Expected inter-coalescent time intervals for 10 lineages sampled in the current day on the background of a derived neutral allele with frequency f in the population. The colored dots give means of T_k from our simulations using stochastic trajectories. The solid lines show the expected coalescent times under our approximation $\mathbb{E}[T_k] = 2Nf/\binom{k}{2}$. Note the good agreement except for $k = 2$. Presumably our approximation over estimates this time as it fails to acknowledge that the allele is derived, and hence is decreasing in frequency backward in time.

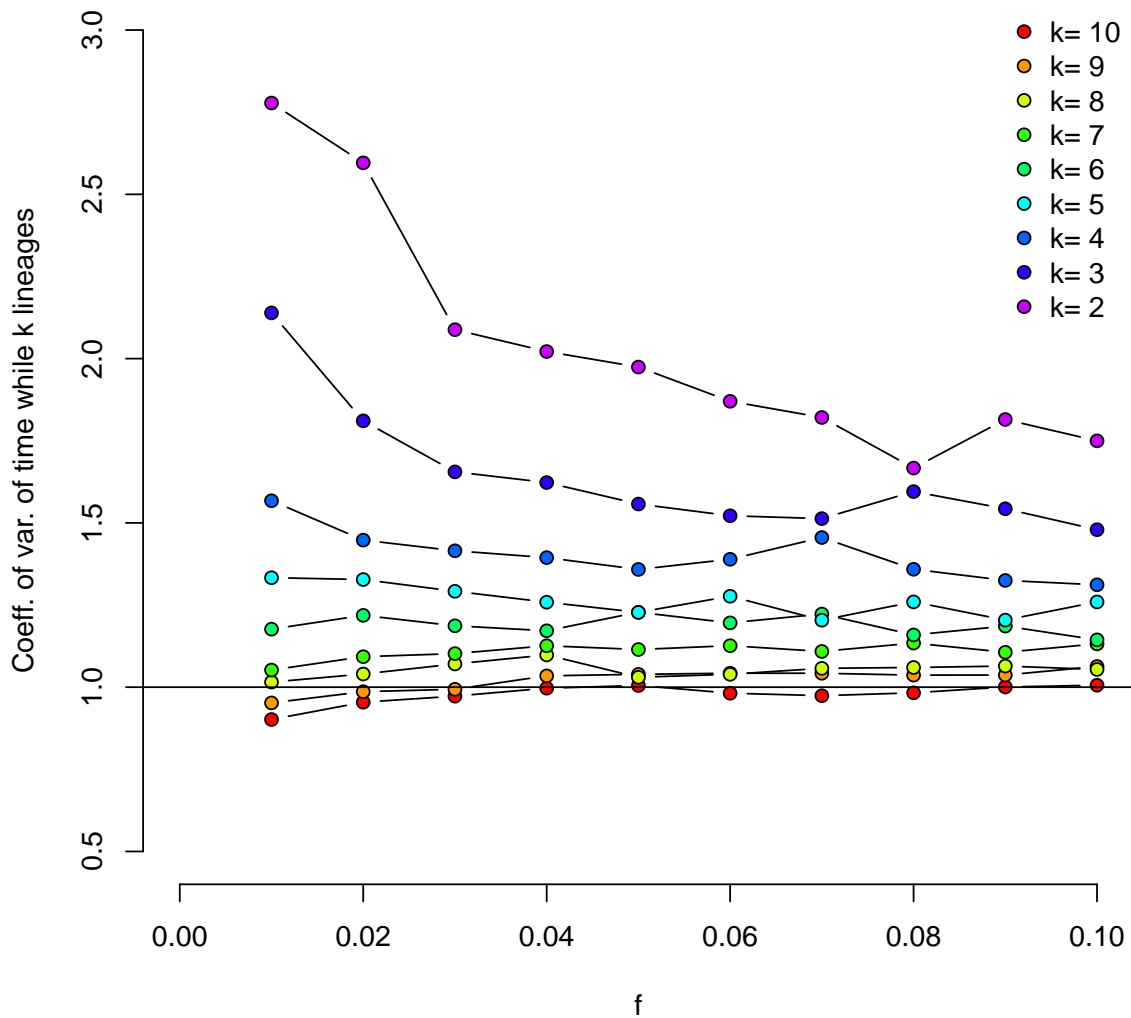


Figure S3: The coefficient of variation (CV) of the inter-coalescent time intervals for 10 lineages sampled in the current day on the background of a derived neutral allele with frequency f in the population. The colored dots give the CVs of T_k from our simulations using stochastic trajectories. Under our approximation the coalescent time intervals are exponential, and so have $CV=1$. The deeper coalescent time-intervals (low k s) are more variable than our predictions presumably because neutral trajectories are highly variable so increasing the variance of the coalescent times. More recent time-intervals (high k) are in better agreement with our approximation, as the trajectory with not have strayed far from a frequency f a short way back in time.

File S1

Simulation Code

Available for download at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178962/-/DC1