FILE S5: THE CORRESPONDENCE BETWEEN STEPTIMES AND REAL TIMES

In this Supplementary File, we calculate the correspondence between steptimes and the actual times measured in generations. Our goal is to calculate the probability distribution of real coalescence times, $\psi(t|k,k',\ell)$, given that individuals were initially in classes k and k' and coalesced in class $k-\ell$.

To begin, we neglect the coalescence time within class $k-\ell$, and consider the distribution of the time at which an ancestor of one of the two sampled individuals first mutated from class $k-\ell$ to class $k-\ell+1$. We refer to this as $\psi_1(t|k,k',\ell)$. We first calculate the joint distribution of the times at which both ancestors mutated out of the class, $R^{k-\ell}_{k,k'}(t_1,t_2)$. Conditional on coalescence in class $k-\ell$, $R^{k-\ell}_{k,k'}(t_1,t_2)$, is given by the probability of t_1 and t_2 and coalescence divided by the total probability of coalescence. That is,

$$R(t_1,t_2) = \frac{P(coal|t_1,t_2)P(t_1,t_2)}{P(coal)}. \tag{S.62}$$

Substituting in the relevant expressions from the main text, this gives

$$R^{k-\ell}_{k,k'}(t_1,t_2) = \frac{1}{A^{k,k'}_{\ell}} Q^{k-\ell}_{k,k'}(t_1,t_2) e^{-s(k-\ell)|t_1-t_2|}. \tag{S.63}$$

The time at which the first ancestor mutated out of class $k-\ell$ is the longer of the two times t_1 and t_2,

$$\psi(t|k,k',\ell) = \left[\int_0^t R^{k-\ell}_{k,k'}(t_1,t)dt_1 + \int_0^t R^{k-\ell}_{k,k'}(t,t_2)dt_2\right]. \tag{S.64}$$

Substituting in our expression for $R^{k-\ell}_{k,k'}(t_1,t_2)$ and carrying out the integrals as in File S3, we find

$$\psi_1(t|k,k',\ell) = s\pi_d e^{-s(k'+k)t}(e^{st}-1)^{\pi_d-1}\binom{k'+k}{\pi_d}, \tag{S.65}$$

where we have used $\pi_d = k'-k+2\ell$.

We can alternatively calculate $\psi_1(t|k,k',\ell)$ using our sum of ancestral paths approach. As before, we imagine two individuals sampled from classes k and k' and condition on them coalescing in class $k-\ell$. Consider a case where $k \neq k'$. Then the first event in the history of these two individuals must be a deleterious mutation. Since these mutations happen at rate sk and sk' in each lineage, the distribution of times since this mutation occurred in one of the two ancestral lineages is

$$P(t) = s(k+k')e^{-s(k+k')t}. \tag{S.66}$$

With probability $\frac{k'}{k+k'}$, this mutation is in the lineage sampled from class k', in which case the two lineages are now in classes k and $k'-1$. Alternatively, the mutaion occurred in the lineage sampled from k and the lineages are in classes $k-1$ and k'.

We can now consider the time to the next event backwards in time. If the two lineages are in the same class (but not yet in class $k-\ell$), the distribution of times to the next deleterious mutation event is somewhat shorter, because we are conditioning on coalescence not occuring. However, provided that $2sk_1 \gg \frac{1}{Nh_k}$ (the condition we are already making elsewhere), this shortening of the time will be a small correction and neglecting it is a good approximation.

Making this approximation, the rate at which the next deleterious mutation event occurs when the two lineages are in classes k_1 and k_2 is just $s(k_1+k_2)$. Regardless of the order in which these mutations happen between the two lineages, this sum is simply decreased by s at each step. This will continue until the both ancestral lineages are in class $k-\ell$. Therefore, the distribution of times until the original mutation out of class $k-\ell$ is given by:

$$\psi_1(t|k',k,\ell) = s(k'+k)e^{-s(k'+k)t} \star s(k'+k-1)e^{-s(k'+k-1)t} \star \ldots \star s(2k-2\ell+1)e^{-s(2k-2\ell+1)t}. \tag{S.67}$$

This can be written as

$$\psi_1(t|k',k,\ell) = \lambda_0 e^{-\lambda_0 t} \star \lambda_1 e^{-\lambda_1 t} \star \ldots \star \lambda_{k'-k+2\ell-1} e^{-\lambda_{k'-k+2\ell-1} t}, \tag{S.68}$$

where we have defined:

$$\lambda_i = s(k'+k-i). \tag{S.69}$$

We can compute this convolution as in File S2 (compare to Eq. (S.17) for $Q^{2k-2\ell}_{k+k'}(t)$). We find

$$\psi_1(t|k,k',\ell) = s\pi_d e^{-s(k'+k)t}(e^{st}-1)^{\pi_d-1}\binom{k'+k}{\pi_d}, \tag{S.70}$$

identical to the result of our lineage structure calculation above.

Distribution of Coalescence Times: To calculate the correspondence between steptimes and real times, we now need to add the time it takes two individuals two coalesce in class $k-\ell$, which we refer to as $\psi_2(t|k,k',\ell)$, to the time it took them both to get to that class, $\psi_1(t|k,k',k-\ell)$. The rate of coalescence once in class $k-\ell$ is $\frac{1}{Nh_{k-\ell}}$, so we have

$$\psi_2(t|k',k,\ell) = \left(2s(k-\ell)+1/Nh_{k-\ell}\right) e^{-[2s(k-\ell)+1/Nh_{k-l}]t}. \tag{S.71}$$

Putting this together, the full distribution of times since coalescence is

$$\psi(t|k',k,\ell) = \psi_1(t|k',k,\ell) \star \psi_2(t|k',k,\ell). \tag{S.72}$$

Carrying out this convolution (and expanding the binomial factor $(e^{st}-1)^{\pi_d-1}$ in ψ_1), we find

$$\psi(t|k',k,\ell) = \sum_{i=0}^{\pi_d-1} s\pi_d(-1)^{\pi_d-i-1}\binom{\pi_d-1}{i}\binom{k'+k}{\pi_d}\frac{B}{A-B}\left(e^{-sBt}-e^{-sAt}\right), \tag{S.73}$$

where we have defined $A \equiv k' + k - i$ and $B \equiv 2\,(k - \ell) + \frac{1}{Nsh_{k-\ell}}$.