FILE S1: THE FULL CONDITIONAL CALCULATION

In the main text, we focused primarily on the non-conditional approximation to the coalescence probabilities, which led to our simple expression for the coalescence probabilities, Eq. (15). In this Supplementary File, we show how this approximation can be relaxed in our lineage-structure framework by carrying out the full conditional calculation for some of the simplest possible cases. We use this to understand the structure of the conditional results and discuss the validity of the non-conditional approximation. We note that the full conditional result can also be obtained from the sum of ancestral paths approach by keeping the higher order terms in Eq. (56) of Appendix A, as described in File S4, and the validity of the non-conditional approximation can be directly assessed with that approach.

We begin by considering the full conditional result for the probability that two individuals both sampled from class k coalesce in class $k - 2$. From Appendix A of the main text, we have

$$P_c^{k,k \to k-2} = I_x^{k-2} \int Q_{k,k}^{k-2}(t_1, t_2) \exp\left[-s(k-2)|t_1 - t_2|\right] dt_1 dt_2. \tag{S.1}$$

In order to evaluate this integral, we need to determine the probability distribution of mutant timings $Q_{k,k}^{k-2}(t_1, t_2)$. The time t_1 is the sum of the time for one individual to have mutated from class $k - 2$ to class $k - 1$ plus the time for it to have mutated from class $k - 1$ to class k, and analogously for t_2. However, in order for the two lineages to coalesce in class $k - 2$, they must *not* have coalesced in class $k - 1$. To illustrate the main point, we neglect the distortion in the mutant timings due to the fact that individuals did not coalesce in class k and focus only on the distortions due to the fact that coalescence did not occur in class $k - 1$; if desired, the former distortion can also be included using analogous methods. We refer to the probability distribution of the times when these individuals mutated from class $k - 1$ to class k conditional on them not having coalesced in class $k - 1$ as $Q_{k,k}^{k-1}(t_1, t_2 | nc)$. The distribution of the times for these individuals to then have mutated from class $k - 2$ to class $k - 1$ is then given by

$$Q_{1step}^{k-2}(t_1, t_2) = [s(k - 1)]^2 e^{-s(k-1)(t_1 + t_2)}. \tag{S.2}$$

Thus the distribution of t_1 and t_2 is given by

$$Q_{k,k}^{k-2}(t_1, t_2) = Q_{k,k}^{k-1}(t_1, t_2 | nc) \star Q_{1step}^{k-2}(t_1, t_2), \tag{S.3}$$

where $\star$ indicates a convolution. Note that much of the time when the individuals did coalesce in class $k - 1$, they did so because t_1 happened to be close to t_2 (since this increases the chance the two individuals mutated from the same lineage). Thus in $Q_{k,k}^{k-1}(t_1, t_2 | nc)$, t_1 and t_2 are on average further apart than in $Q_{k,k}^{k-1}(t_1, t_2)$, and t_1 and t_2 are no longer independent random variables.

A. M. Walczak *et al.*

We now need to calculate $Q_{k,k}^{k-1}(t_1, t_2|nc)$. We have

$$Q_{k,k}^{k-1}(t_1, t_2|nc) = \frac{Q_{k,k}^{k-1}(t_1, t_2) - Q_{k,k}^{k-1}(t_1, t_2|c)P_c^{k,k\to k-1}}{1 - P_c^{k,k\to k-1}}, \tag{S.4}$$

where $Q_{k,k}^{k-1}(t_1, t_2|c)$ is the distribution of timings of mutations from class $k-1$ to k given that the lineages *do* coalesce in class $k-1$. Applying the general probability identity $P(t_1, t_2|c) = \frac{1}{P(c)}P(c|t_1, t_2)P(t_1, t_2)$, and reading off the coalescence probability given t_1 and t_2 from Eq. (13), we find that

$$Q_{k,k}^{k-1}(t_1, t_2|c) = \frac{I_x^{k-1}}{P_c^{k,k\to k-1}}Q_{k,k}^{k-1}(t_1, t_2)e^{-s(k-1)|t_1-t_2|}. \tag{S.5}$$

We therefore find

$$Q_{k,k}^{k-1}(t_1, t_2|nc) = \frac{1}{1 - P_c^{k,k\to k-1}}\left[(sk)^2 e^{-sk(t_1+t_2)} - I_x^{k-1}(sk)^2 e^{-2k(t_1+t_2)}e^{-s(k-1)|t_1-t_2|}\right]. \tag{S.6}$$

Plugging this into our convolution formula for $Q_{k,k}^{k-2}(t_1, t_2)$ and evaluating the integrals by separating out the possible time orderings, we find

$$Q_{k,k}^{k-2}(t_1, t_2) = \frac{k^2\left[s(k-1)\right]^2}{1 - P_c^{k,k\to k-1}}e^{-s(k-1)(t_1+t_2)}\left[\left(1 - e^{-st_1}\right)\left(1 - e^{-2t_2}\right) - \frac{I_x^{k-1}}{k-2}B\right], \tag{S.7}$$

where we have defined

$$\begin{aligned}
B &= \frac{1}{(k-2)}\left[1 - e^{-2s\min(t_1,t_2)} - \frac{2}{k}\left(1 - e^{-sk\min(t_1,t_2)}\right)\right.\\
&\quad \left. + \frac{1}{k}\left(1 - e^{-2k|t_1-t_2|}\right)\left(e^{-2s\min(t_1,t_2)} - e^{-sk\min(t_1,t_2)}\right)\right].
\end{aligned} \tag{S.8}$$

We can now use this expression in Eq. (S.1) to calculate the coalescence probability $P_c^{k,k\to k-2}$. Since the result is tedious and does not further illuminate the structure of the full conditional calculation, we do not do so explicitly here, but the integrals are straightforward to evaluate with the methods we have used above.

To motivate the validity of the non-conditional approximation, we need to consider the full calculation going back one additional step. Thus we consider the probability that two individuals both sampled from class k coalesce in class $k-3$, $P_c^{k,k\to k-3}$. This will be given by

$$P_c^{k,k\to k-3} = \int Q_{k,k}^{k-3}(t_1, t_2)\frac{x^2}{h_{k-3}^2}f_{k-3}(x)e^{-s(k-3)|t_1-t_2|}dt_1 dt_2 dx, \tag{S.9}$$

where here $Q_{k,k}^{k-3}(t_1, t_2)$ is the distribution of the time at which the ancestors of the two sampled individuals originally mutated from class $k-3$ to class $k-2$, conditional on them not coalescing in classes $k-2$ or $k-1$.

We can calculate $Q_{k,k}^{k-3}(t_1, t_2)$ in the same way we calculated $Q_{k,k}^{k-2}(t_1, t_2)$. Explicitly,

$$Q_{k,k}^{k-3}(t_1, t_2) = Q_{k,k}^{k-2}(t_1, t_2|nc) \star Q_{1step}^{k-3}(t_1, t_2), \tag{S.10}$$

where analogously to the expression in the previous step

$$Q_{k,k}^{k-2}(t_1, t_2|nc) = \frac{1}{1 - P_c^{k,k \to k-2}} \left[Q_{k,k}^{k-2}(t_1, t_2) - Q_{k,k}^{k-2}(t_1, t_2|c) P_c^{k,k \to k-2} \right]. \tag{S.11}$$

We note that $Q_{k,k}^{k-2}(t_1, t_2)$ is the expression in Eq. (S.7) we calculated above. As before, we have

$$Q_{k,k}^{k-2}(t_1, t_2|c) P_c^{k,k \to k-2} = I_x^{k-2} Q_{k,k}^{k-2}(t_1, t_2) e^{-s(k-2)|t_1-t_2|}, \tag{S.12}$$

hence we can write

$$Q_{k,k}^{k-2}(t_1, t_2|nc) = \frac{Q_{k,k}^{k-2}(t_1, t_2)}{1 - P_c^{k,k \to k-2}} \left[1 - I_x^{k-2} e^{-s(k-2)|t_1-t_2|} \right]. \tag{S.13}$$

Plugging the above expression back into Eq. (S.10), we obtain

$$\begin{aligned} Q_{k,k}^{k-3}(t_1, t_2) &= \frac{s^2(k-1)^2 k^2 s^2 (k-2)^2}{(1 - P_c^{k,k \to k-1})(1 - P_c^{k,k \to k-2})} e^{-s(k-2)(t_1+t_2)} \int_0^{t_2} \int_0^{t_1} e^{s(k-2)(y+z)} e^{s(k-1)(y+z)} \\ &\times \left[1 - I_x^{k-2} e^{-s(k-z)|y-z|} \right] \left[(1 - e^{-sy})(1 - e^{-sz}) - \frac{I_x^{k-1}}{k-2} B \right]. \end{aligned} \tag{S.14}$$

We could evaluate the integrals in the above expression for $Q_{k,k}^{k-3}(t_1, t_2)$ in the same way that we did in our calculation for $Q_{k,k}^{k-2}(t_1, t_2)$. We would then substitute this result for $Q_{k,k}^{k-3}(t_1, t_2)$ into an analogous calculation of $Q_{k,k}^{k-4}(t_1, t_2)$, and so on. In this way we can build up the full conditional results. The most useful way to go about this is to separate the results into powers of I_x, which is a small parameter related to the coalescent probability in each step. We see from the expression for $Q_{k,k}^{k-3}(t_1, t_2)$ that there is a term in $(I_x)^0$, which is exactly the non-conditional approximation. There are two terms involving $(I_x)^1$, and a single term involving $(I_x)^2$. In general, in the expression for $Q_{k,k}^{k-\ell}(t_1, t_2)$, we will have one $(I_x)^0$ term (which equals the result in the non-conditional approximation) plus ℓ terms proportional to I_x, $\binom{2}{\ell}$ terms proportional to $(I_x)^2$, and so on. Fortunately, the dependence on the population parameters is entirely contained within these powers of I_x. That is, the coefficients of these various powers of I_x depend *only* on k and ℓ, and not at all on the population parameters N, s, and U_d. Thus we could simply calculate a table of coefficients once, and then would be able to understand all the distributions of mutant timings (and from this all the coalescent probabilities).

In practice, it is easier to make these full conditional calculations within the sum of ancestral paths approach. As we show in File S4, that approach leads naturally to a power series in I_x of exactly the form described above, in which the leading order term is the non-conditional approximation and the additional terms represent the conditional corrections. This calculation shows that provided $I_x \ll 1$, which is true provided our usual condition that $Nh_k sk \gg 1$ holds, these higher order terms are all small, and our non-conditional approximation is valid.

These full conditional results are, however, very complex and unilluminating. Therefore we focus here on

A. M. Walczak *et al.*

understanding the general structure of these results, and on showing why the non-conditional approximation is good description of the distribution of mutation timings. We can see that at each step back through the fitness distribution, the probability distribution of times shifts from the non-conditional results by a factor which is roughly proportional to the coalescence probability at that step. That is, in general we have

$$Q_{k,k}^{k-\ell}(t_1, t_2) = \frac{1}{1 - P_c^{k,k \to k-\ell}} \left[Q_{k,k}^{k-\ell}(t_1, t_2) - P_c^{k,k \to k-\ell} Q_{k,k}^{k-2}(t_1, t_2|c) \right]. \tag{S.15}$$

The first term in square brackets reflects the fact that the probability distribution at a given step conditional on non-coalescence at that step is almost equal to the unconditional probability distribution at that step. The second term represents the correction: note that it is proportional to the coalescence probability in that step, $P_c^{k,k \to k-\ell}$. The nature of the correction can be seen by plugging in the distribution of times conditional on coalescence, giving

$$Q_{k,k}^{k-\ell}(t_1, t_2) = \frac{Q_{k,k}^{k-\ell}(t_1, t_2)}{1 - P_c^{k,k \to k-\ell}} \left[1 - I_x^{k-\ell} e^{-s(k-\ell)|t_1 - t_2|} \right]. \tag{S.16}$$

We see that the correction acts to reduce the probability that $|t_1 - t_2|$ is small — that is, it makes it more likely that t_1 and t_2 are further apart, because this is more likely to be the case given that coalescence did not occur.

Since at each step the shift in the distribution of mutant timings is proportional to the coalescence probability, and the coalescence probability at each step is small, it seems clear that the non-conditional approximation where we simply ignore this shift in mutant timings is reasonable. However there is one potential caveat we must consider: although the shift in the distribution of mutation timings due to conditioning on non-coalescence is small *in each step*, we typically take many steps before the lineages coalesce. In fact, since the shift in mutation timings is proportional to the coalescence probability, and we typically go back a number of steps of order one over the coalescence probability, in principle the shifts in mutation timings could add up to a substantial shift.

Fortunately, there are three factors which prevent this from happening. First, the shift in mutation timings at each step is always to reduce the probability of times t_1 and t_2 where $|t_1 - t_2| \lesssim \frac{1}{(k-\ell)s}$. Since at each step ℓ is increasing, and the range of separations between mutation timings at which coalescence can happen is also increasing, the shifts in mutation timings from many steps ago are not a huge factor in determining coalescence probabilities in a particular step. That is, though the shifts in mutation timings add up over many steps, the shifts most relevant to the coalescent probability in a given step do not. Second, the coalescence probabilities at each step are different. This reduces the chance that we take enough steps to shift the overall mutation timings substantially by the time we coalesce. Finally, and most importantly, we will see that the there is a substantial probability that the ancestors of the two individuals sampled do

not coalesce until they are in the most-fit class. This means that the total sum of coalescence probabilities (and hence the total possible weight in the shift of mutation timings) remains small even in the worst case where the two lineages do not coalesce for the maximum possible number of steps. The non-conditional approximation will always be good in the regime where this is true. All of these heuristic conclusions are reflected in the fact that the full conditional result we calculate in the sum of ancestral paths approach is equal to the non-conditional result plus corrections that are small provided $I_x \gg 1$.

A. M. Walczak *et al.*