Supplementary text

The probability of coalescence in the presence of variable population size

The probability that two chromosomes with compatible genetic backgrounds $\mathcal{G}_1^{(i)}$ and $\mathcal{G}_1^{(j)}$ (assuming K = 1) chosen randomly at time zero have distinct ancestors T generations ago is

$$\prod_{t=1}^{T}(1-p_c(t)) \approx \exp[-\phi(\mathcal{G}_1^{(i)},\mathcal{G}_1^{(j)})\sum_{t=1}^{T}\frac{1}{N(t)}]$$

where

$$p_c(t)=\frac{1}{N(t)}\phi(\mathcal{G}_1^{(i)},\mathcal{G}_1^{(j)})$$

and

$$\phi(\mathcal{G}_1^{(i)},\mathcal{G}_1^{(j)})=\frac{P(\mathcal{G}_1^{(a)})}{f(\boldsymbol{d})P(\mathcal{G}_1^{(i)})P(\mathcal{G}_1^{(j)})}$$

Note that the definition above has made use of the fact that $\boldsymbol{d}$ and $\mathcal{G}_1^{(a)}$ can be determined by $\mathcal{G}_1^{(i)}$ and $\mathcal{G}_1^{(j)}$. Let $N \equiv N(0)$ and rescale time in units of N generations, so that $\tau = t/N$ and $\mathcal{T} = T/N$. Following Tavaré's approach (2004), we first define the relative size function

$$g(\tau)=\frac{N(\lceil N\tau\rceil)}{N}$$

where $\lceil x\rceil$ is the smallest integer greater than x. By replacing summation with integration, we yield

$$\exp[-\phi(\mathcal{G}_1^{(i)},\mathcal{G}_1^{(j)})\sum_{t=1}^{T}\frac{1}{N(t)}] \approx \exp[-\phi(\mathcal{G}_1^{(i)},\mathcal{G}_1^{(j)})\int_0^{\mathcal{T}}\frac{d\tau}{g(\tau)}]$$

Thus, the waiting time to the next coalescent event follows a non-homogeneous exponential distribution.

Factors affecting accuracy of the coalescent model

The coalescent model relies on two key assumptions. First, the population size is sufficiently large that the genetic composition of the population, as characterised by the distribution of $\boldsymbol{d}$, stays close to Eq. (2). Second, the number of individuals in each genetic background (i.e., $Nf(\boldsymbol{d})P(\mathcal{G})$; see Eq. (4)) is sufficiently large that the continuous time approximation is valid. As pointed out previously (Zeng and Charlesworth, 2011), the tendency for the coalescent model to produce somewhat higher $E(T_n)$, the mean total branch length for a sample of size n, compared to forward simulations (e.g., Figure 3) is probably caused by the use of small population sizes in the forward simulations (typically N = 5,000). Specifically, when $Nf(\boldsymbol{d})P(\mathcal{G})$ is small, both the probability of multiple coalescent events occurring in one generation and the probability of having coalescent events where more than two ancestral lineages find their common ancestor in the previous generation are non-negligible. This will make the actual rate of coalescence higher than assumed by the coalescent model. In addition, in small populations, genetic drift induces significant linkage disequilibrium between selected sites, and the interference between selection at different sites may further reduce $E(T_n)$ (Hill and Robertson, 1966). Because forward simulations are extremely time-consuming even for moderately large populations (e.g., N > 5,000), verifying these intuitions systematically by simulating large populations is computationally prohibitive, although some evidence has been obtained from a small-scale experiment (Zeng and Charlesworth, 2011).

Diagnostic methods can be constructed to ensure that these assumptions are

approximately valid. The treatments below assume that the population of interest has a constant size N, and that there is one type of selected sites (i.e., $K = 1$). The total deleterious mutation rate and the selection coefficient are denoted by U and s, respectively, with $\lambda = U/s$. In an infinite population, the number of deleterious variants on a random chromosome, d, follows a Poisson distribution with mean λ whose probability density function is denoted by $f(d)$.

Firstly, for the distribution of d in a finite population to be close to $f(d)$, two conditions should be met: (*i*) the size of the mutation-free class is sufficiently large; (*ii*) selection is sufficiently strong. These are to ensure that the effects of Muller's ratchet can be ignored (Gordo *et al.*, 2002), and that deleterious mutations are under the control of selection rather than drift. The size of the mutation-free class can be approximated by $Ne^{-\lambda}$ (Charlesworth *et al.*, 1993). Since the selection coefficient is small, both conditions should be met by requiring

$$Ne^{-\lambda}s > 1 \qquad \text{(S1)}$$

For $Nf(d)P(\mathcal{G})$, it is necessary to consider the expected number of recombination events in the history of a sample, because each recombination event creates a new subset of individuals amongst those with the same number of deleterious mutations. For instance, the recombination event in Figure 2 creates two ancestral lineages with genetic backgrounds $\mathcal{G}_1^{(3)}$ and $\mathcal{G}_1^{(4)}$. $P(\mathcal{G}_1^{(3)})$ is the proportion of individuals with one deleterious mutation in [1, 5], a subset amongst those with one mutations; similarly, $\mathcal{G}_1^{(4)}$ encompasses individuals with one mutation in [1, 5] and one in [6, 10], a subset amongst all individuals with two mutations. Intuitively, $Nf(d)P(\mathcal{G})$

decreases as the recombination rate increases, since the focal region will be divided into smaller sub-intervals with more recombination events. The expected number of recombination events can be predicted as follows. Take an estimate of $E(T_n)$ produced by the coalescent model. Note that T_n is scaled by N. Unlike in the main text, it is *not* expressed relative to its neutral expectation here and in Supplementary Table S3. Using the theory of Hudson and Kaplan (1985), the expected number of recombination events is

$$R_e^* = \rho E(T_n)$$

where $\rho = NR$ and R is the recombination rate. To be conservative, the following calculation uses estimates of $E(T_n)$ obtained from the left-hand end of the focal region [$E(T_n)$ is larger at the end points (e.g., Figure 3)]. Assume that there are on average $R_e = \lceil R_e^* \rceil$ recombination events. Since the recombination rate is uniform across the region, we can assume that the R_e breakpoints are uniformly distributed, so that they divide the focal region into $R_e + 1$ equal-sized sub-intervals. Thus, amongst individuals with d deleterious mutations, the number of individuals in the least likely genetic background whereby all d mutations are situated within one of the $R_e + 1$ sub-intervals is

$$h(d) = Nf(d)\left(\frac{1}{1+R_e}\right)^d$$

Since $f(d)$ is maximised for d close to λ. Thus, with

$$h(\lceil \lambda \rceil + 1) > 10 \qquad \text{(S2)}$$

the second assumption of the coalescent model should be approximately valid over the bulk of the distribution of d.

Thus, Eqs. (S1) and (S2) define two diagnostic criteria, and when they are both satisfied, we expect the coalescent model to provide good approximations. Note that these criteria depend solely on data generated by the coalescent model. To assess the performance of these criteria, data have been generated by both coalescent and forward simulations. A total of 203 sets of parameters have been considered. To make the assessment conservative, parameters were chosen from regions in the parameter space that are more likely to result in breakdowns of the coalescent model (i.e., when selection is weak with $\gamma \leq 7.5$). Forward simulations have been carried out with N = 5,000. Local genealogies were recorded for the left-hand end only in both types of simulations, since recording local genealogies for many sites is computationally expensive. Let $E(X)^{forw}$ and $E(X)^{coal}$ denote estimates obtained from the two types of simulations, where $X = T_n$ or ξ_n. We define the relative deviation, $R_d(X)$, as

$$R_d(X) = \frac{E(X)^{forw} - E(X)^{coal}}{E(X)^{forw}}$$

$E(T_n)$ values obtained from the left-hand end were used to calculate Eqs. (S1) and (S2). The diagnostic criteria are said to score a *true positive* when both Eqs. (S1) and (S2) are satisfied, and $|R_d(T_n)| < 0.05$. A less stringent criterion of $|R_d(T_n)| < 0.1$ has also been tested. Note that $|R_d(T_n)| < 0.05$ is likely to be rather conservative, since the difference between $E(T_n)^{forw}$ and $E(T_n)^{coal}$ is likely to be exaggerated due to the use of a small N in the forward simulations. In contrast, the diagnostic criteria are said to produce a *false positive* if both Eqs. (S1) and (S2) are satisfied, but $|R_d(T_n)| \geq 0.05$ (or $|R_d(T_n)| \geq 0.1$). The true and false positive rates are summarised in Supplementary Tables S2. Note that the difference in true positive rates between $|R_d(T_n)| < 0.05$ and

$|R_d(T_n)| < 0.1$ is mainly due to *false negatives* in the latter case (i.e., $|R_d(T_n)| < 0.1$ but the diagnostic criteria suggest a breakdown). Also reported in the table are the frequency of cases where $|R_d(\xi_n)| < |R_d(T_n)|$. The full data set can be found in Supplementary Tables S3.

References

Charlesworth B, Morgan MT, Charlesworth D (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134:** 1289-1303.

Gordo I, Navarro A, Charlesworth B (2002). Muller's ratchet and the pattern of variation at a neutral locus. *Genetics* **161:** 835-848.

Hill WG, Robertson A (1966). The effect of linkage on limits to artificial selection. *Genet Res* **8:** 269-294.

Hudson RR, Kaplan NL (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111:** 147-164.

Tavaré S (2004). Ancestral inference in population genetics. In: Picard J (ed) *Lectures on Probability Theory and Statistics. Ecole d'Etés de Probabilité de Saint-Flour XXXI -- 2001*. Springer Verlag: New York. Vol. 1837, pp 1–188.

Zeng K, Charlesworth B (2011). The joint effects of background selection and genetic recombination on local gene genealogies. *Genetics* **189:** 251-266.