

Supplementary text

Contents

S.1	Definition of the backward mutation rate	2
S.2	Calculating the Green's matrix	2
S.3	The intensity matrix for calculating the total branch length of a sample size of three	2
S.4	The intensity matrix for calculating the SFS for a sample size of three . .	3
S.5	A non-equilibrium phase-type model	3
S.6	Coalescent simulation with stochastic allele frequency trajectories	6
S.7	The second moment of the mean time to absorption	7
S.8	Approximating the expected pairwise coalescence time under the model of recent balancing selection	7

S.1 Definition of the backward mutation rate

Consider the equilibrium model of balancing selection. The equilibrium frequency of the selected variant A_i is \hat{p}_i ($i = 1, 2$). During reproduction, an A_i allele mutates to A_j with probability u_{ij} per generation. Forward in time, after reproduction, the frequency of A_i becomes $\hat{p}_i(1 - u_{ij}) + \hat{p}_j u_{ji}$, where the second term is the frequency of mutants amongst all A_i alleles. Going backward in time, the backward mutation rate is the probability that an A_i allele in the post-reproduction population is descendant from an A_j parent via mutation:

$$v_{ij} = \frac{\hat{p}_j u_{ji}}{\hat{p}_i(1 - u_{ij}) + \hat{p}_j u_{ji}}$$

Assuming that u_{ij} is much smaller than both \hat{p}_i and \hat{p}_j , $v_{ij} \approx u_{ji} \hat{p}_j / \hat{p}_i$, which is the same as (6) in [Kaplan *et al.* \(1988\)](#).

S.2 Calculating the Green's matrix

Let $\{X_t\}_{t \geq 0}$ be a Markov jump process with sub-intensity matrix \mathbf{S} . That is, $X_t = i$ means that the process is in state i at time t . Let $Z_j = \int_0^\infty 1(X_t = j) dt$ denote the time spent in state j before absorption, where $1(X_t = j) = 1$ if $X_t = j$, and 0 otherwise. The expected time spent in state j given initial state i is then

$$\begin{aligned} u_{ij} &= \mathbb{E}[Z_j | X_0 = i] \\ &= \mathbb{E}\left[\int_0^\infty 1(X_t = j) dt | X_0 = i\right] \\ &= \int_0^\infty \mathbb{E}[1(X_t = j) | X_0 = i] dt \\ &= \int_0^\infty \Pr(X_t = j | X_0 = i) dt \\ &= \int_0^\infty (e^{\mathbf{S}t})_{ij} dt \\ &= \left(\int_0^\infty e^{\mathbf{S}t} dt\right)_{ij}. \end{aligned}$$

To complete the proof, we note that

$$\int_0^\infty e^{\mathbf{S}t} dt = [\mathbf{S}^{-1} e^{\mathbf{S}t}]_0^\infty = -\mathbf{S}^{-1}$$

where we have used the property that $(e^{\mathbf{S}t})_{ij} \rightarrow 0$ for $t \rightarrow \infty$.

S.3 The intensity matrix for calculating the total branch length of a sample size of three

\mathbf{S}_2 and \mathbf{s}_2 in (12) are the same as the corresponding elements defined in (3).

$$\mathbf{S}_3 = \begin{pmatrix} -3M_{21} - \frac{3}{\hat{p}_2} & 3M_{21} & 0 & 0 \\ M_{12} & -M_{12} - 2M_{21} - \frac{1}{\hat{p}_2} & 2M_{21} & 0 \\ 0 & 2M_{12} & -2M_{12} - M_{21} - \frac{1}{\hat{p}_1} & M_{21} \\ 0 & 0 & 3M_{12} & -3M_{12} - \frac{3}{\hat{p}_1} \end{pmatrix} \quad (\text{S1})$$

and

$$\mathbf{S}_{32} = \begin{pmatrix} \frac{3}{\hat{p}_2} & 0 & 0 \\ 0 & \frac{1}{\hat{p}_2} & 0 \\ 0 & \frac{1}{\hat{p}_1} & 0 \\ 0 & 0 & \frac{3}{\hat{p}_1} \end{pmatrix}. \quad (\text{S2})$$

S.4 The intensity matrix for calculating the SFS for a sample size of three

The sub-matrices in (12) for the model leading to Table 1 are given below.

$$\mathbf{S}_3 = \begin{pmatrix} -3M_{21} - \frac{3}{\hat{p}_2} & 3M_{21} & 0 & 0 \\ M_{12} & -M_{12} - 2M_{21} - \frac{1}{\hat{p}_2} & 2M_{21} & 0 \\ 0 & 2M_{12} & -2M_{12} - M_{21} - \frac{1}{\hat{p}_1} & M_{21} \\ 0 & 0 & 3M_{12} & -3M_{12} - \frac{3}{\hat{p}_1} \end{pmatrix}. \quad (\text{S3})$$

$$\mathbf{S}_{32} = \begin{pmatrix} \frac{3}{\hat{p}_2} & 0 & 0 & 0 \\ 0 & \frac{1}{\hat{p}_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\hat{p}_1} & 0 \\ 0 & 0 & 0 & \frac{3}{\hat{p}_1} \end{pmatrix}. \quad (\text{S4})$$

$$\mathbf{S}_2 = \begin{pmatrix} -2M_{21} - \frac{1}{\hat{p}_2} & M_{21} & M_{21} & 0 \\ M_{12} & -M_{12} - M_{21} & 0 & M_{21} \\ M_{12} & 0 & -M_{12} - M_{21} & M_{21} \\ 0 & M_{12} & M_{12} & -2M_{12} - \frac{1}{\hat{p}_1} \end{pmatrix}. \quad (\text{S5})$$

$$\mathbf{s}_2^T = \left(\frac{1}{\hat{p}_2} \quad 0 \quad 0 \quad \frac{1}{\hat{p}_1} \right). \quad (\text{S6})$$

S.5 A non-equilibrium phase-type model

Consider a continuous time Markov chain with finite state space $\{1, 2, \dots, K, K+1\}$, where states $1, \dots, K$ are transient, and state $K+1$ is absorbing. It is assumed that the time interval $[0, \infty)$ is subdivided into H non-overlapping epochs. The duration of epoch h is $[t_{h-1}, t_h)$, where $1 \leq h \leq H$, $t_0 = 0$, and $t_H = \infty$. The intensity matrix for epoch h is constant and takes the form:

$$\mathbf{\Lambda}_h = \begin{pmatrix} \mathbf{S}_h & \mathbf{s}_h \\ \vec{0} & 0 \end{pmatrix} \quad (\text{S7})$$

where \mathbf{S}_h the K -by- K sub-intensity matrix, and \mathbf{s}_h is the K -by-1 exit rate vector.

Define

$$\begin{cases} d_h = t_h - t_{h-1} \\ \mathbb{h}(t) = \min\{h : 1 \leq h \leq H \text{ and } t_{h-1} \leq t < t_h\} \\ d_{\mathbb{h}(t)} = t - t_{\mathbb{h}(t)-1} \end{cases} \quad (\text{S8})$$

The transition probability between time 0 and time t is given by:

$$\mathbf{P}(t) = \left[\prod_{h=1}^{\mathbb{h}(t)-1} \mathbf{P}_h(d_h) \right] \mathbf{P}_{\mathbb{h}(t)}(d_{\mathbb{h}(t)}) \quad (\text{S9})$$

where $\mathbf{P}_h(t)$ is the transition matrix for epoch h . Note that the matrices do not commute. So the multiplication should be carried out from left to right, according to the chronological order of the epochs. From standard Markov chain theory, we know that:

$$\mathbf{P}_h(t) = \begin{pmatrix} e^{\mathbf{S}_h t} & \vec{1} - e^{\mathbf{S}_h t} \vec{1} \\ \vec{0} & 1 \end{pmatrix}. \quad (\text{S10})$$

Define

$$\mathbf{S}(t) = \left[\prod_{h=1}^{\mathbb{h}(t)-1} e^{\mathbf{S}_h d_h} \right] e^{\mathbf{S}_{\mathbb{h}(t)} d_{\mathbb{h}(t)}}. \quad (\text{S11})$$

We can rewrite (S9) in a more compact form:

$$\mathbf{P}(t) = \begin{pmatrix} \mathbf{S}(t) & \vec{1} - \mathbf{S}(t) \vec{1} \\ \vec{0} & 1 \end{pmatrix}. \quad (\text{S12})$$

The probability that the process jumps to the absorbing state in the time interval $[t, t+dt)$ is given by:

$$f(t)dt = \sum_{i=1}^K \alpha_i \sum_{j=1}^K s_{ij}(t) s_j(t) dt = \boldsymbol{\alpha} \mathbf{S}(t) \mathbf{s}(t) dt \quad (\text{S13})$$

where $\boldsymbol{\alpha}$ is the initial probability vector, $s_{ij}(t)$ are elements of $\mathbf{S}(t)$, and $s_j(t)$ are elements of $\mathbf{s}_{\mathbb{h}(t)}$, the exit rate vector at time t . The Laplace transform of $f(t)$ is defined as:

$$\mathcal{L}(z) = \int_0^\infty e^{-zt} \boldsymbol{\alpha} \mathbf{S}(t) \mathbf{s}(t) dt \quad (\text{S14})$$

for $z \geq 0$. Noting that $\mathbf{s}_h = -\mathbf{S}_h \vec{1}$ and substituting (S11) into (S14) leads to:

$$\mathcal{L}(z) = -\boldsymbol{\alpha} \sum_{h=1}^H \left[\prod_{i=1}^{h-1} e^{\mathbf{S}_i d_i} \right] \left[\int_{t_{h-1}}^{t_h} e^{-(z\mathbf{I} - \mathbf{S}_h)t} dt \right] e^{-\mathbf{S}_h t_{h-1}} \mathbf{S}_h \vec{1}, \quad (\text{S15})$$

where \mathbf{I} is the identity matrix. To evaluate the integral, we define $\mathbf{A}_h(z) = \mathbf{A}_h = -(z\mathbf{I} - \mathbf{S}_h)$. Because all eigenvalues of \mathbf{A}_h have strictly negative real parts (Hobolth et al., 2019), $\lim_{t \rightarrow \infty} e^{\mathbf{A}_h t} = 0$. We obtain:

$$\int_{t_{h-1}}^{t_h} e^{\mathbf{A}_h t} dt = \mathbf{A}_h^{-1} (e^{\mathbf{A}_h t_h} - e^{\mathbf{A}_h t_{h-1}}). \quad (\text{S16})$$

Taking the derivative with respect to z , we obtain:

$$\frac{d}{dz} \int_{t_{h-1}}^{t_h} e^{\mathbf{A}_h t} dt = \mathbf{A}_h^{-1} [(\mathbf{A}_h^{-1} - t_h \mathbf{I}) e^{\mathbf{A}_h t_h} - (\mathbf{A}_h^{-1} - t_{h-1} \mathbf{I}) e^{\mathbf{A}_h t_{h-1}}]. \quad (\text{S17})$$

Noting that the mean time to absorption is given by $-\frac{d\mathcal{L}(z)}{dz}|_{z=0}$ and that $\mathbf{A}_h(0) = \mathbf{S}_h$, we have:

$$\mathbb{E}[\mathcal{T}] = \boldsymbol{\alpha} \sum_{h=1}^H \left[\prod_{i=1}^{h-1} e^{\mathbf{S}_i d_i} \right] [(\mathbf{S}_h^{-1} - t_h \mathbf{I}) e^{\mathbf{S}_h d_h} + t_{h-1} \mathbf{I} - \mathbf{S}_h^{-1}] \vec{1}. \quad (\text{S18})$$

Expanding the equation and removing terms that cancel each other, we arrive at Theorem 1. To facilitate further discussion, we state this Theorem in a slightly different way:

Corollary 1. *Let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$, where α_i is the probability that the initial state is i and $\sum_{i=1}^K \alpha_i = 1$. Let \mathcal{T} be a random variable representing the time to absorption. We have:*

$$\mathbb{E}[\mathcal{T}] = \boldsymbol{\alpha} \mathbf{U} \vec{1} \quad (\text{S19})$$

where

$$\begin{cases} \mathbf{U} = \sum_{h=1}^H \left[\prod_{i=1}^{h-1} e^{\mathbf{S}_i d_i} \right] \mathbf{U}_h \\ \mathbf{U}_h = e^{\mathbf{S}_h d_h} \mathbf{S}_h^{-1} - \mathbf{S}_h^{-1} \end{cases} \quad (\text{S20})$$

and $e^{\mathbf{S}_h d_h} = 0$ if $d_h = \infty$.

We have also derived an expression for the second moment of \mathcal{T} in Theorem 2 in Supplementary Text S.7.

Let $u_{ij,h}$ represent the elements of \mathbf{U}_h . $u_{ij,h}$ is the amount of time the process spends in state j during $[t_{h-1}, t_h]$ given that it is in state i at time t_{h-1} . That is, \mathbf{U}_h is the Green's matrix for the h -th epoch. Also note that element i in the vector $\boldsymbol{\alpha} \prod_{j=1}^{h-1} e^{\mathbf{S}_j d_j}$ gives the probability that the process is in state i at time t_{h-1} . Thus, Corollary 1 shows that, under this stepwise model, the Green's matrix for the entire process \mathbf{U} is the weighted average of the Green's matrices of all the constituent epochs.

As noted in the main text, the expectation of both L_{n_1, n_2} and $\phi^{(n_1, n_2)}$ can be written in the form $\boldsymbol{\alpha} \mathbf{U} \mathbf{D}$. Let Y represent either of these two random variables. Corollary 1 tells us that:

$$\mathbb{E}[Y] = \sum_{h=1}^H \mathbb{E}[Y_h] \quad (\text{S21})$$

where

$$\mathbb{E}[Y_h] = \boldsymbol{\alpha} \left[\prod_{i=1}^{h-1} e^{\mathbf{S}_i d_i} \right] \mathbf{U}_h \mathbf{D} \quad (\text{S22})$$

which is the expected contribution from epoch h .

We have so far assumed that the state space is the same across epochs. This restriction can be relaxed. Let the size of the state space in epoch h be K_h . Let $\mathbf{E}_{h-1,h}$ be a K_{h-1} -by- K_h matrix that defines the mapping of the states from epoch $h-1$ to epoch h ($h = 1, \dots, H$ and $\mathbf{E}_{01} = \mathbf{I}$, the identity matrix). Corollary 1 holds if we replace $\prod_{i=1}^{h-1} e^{\mathbf{S}_i d_i}$ by $(\prod_{i=1}^{h-1} \mathbf{E}_{i-1,i} e^{\mathbf{S}_i d_i}) \mathbf{E}_{h-1,h}$. For (S21), we additionally need to replace \mathbf{D} by an epoch-specific \mathbf{D}_h .

S.6 Coalescent simulation with stochastic allele frequency trajectories

The theory developed in the main text assumes that the allele frequencies of the variants at the selected locus change deterministically over time. In reality, frequencies of the selected variants fluctuate because of random genetic drift. To investigate the effects of stochastic allele frequency fluctuation on the accuracy of our model predictions, we conducted simulations with stochastic allele frequency trajectories. Each simulation replicate contained two steps: (1) forward simulation to obtain allele frequency trajectories for the selected variants given the demographic history; (2) coalescent simulation for a sample of n alleles at a linked neutral site, conditioning on the trajectories obtained in step 1. The forward simulation step was based on a custom R script, which is available on request. The coalescent simulation step was performed using `mbs` (Teshima and Innan, 2009).

Forward simulations

Consider a Wright-Fisher population. The variants at the focal selected site are represented by A_1 and A_2 , respectively. The fitnesses of A_1A_1 , A_1A_2 , and A_2A_2 are denoted by w_{11} , w_{12} , and w_{22} , respectively. The model has the following life cycle: mutation, random mating, selection, and sampling. The mutation rate from A_1 to A_2 is u_{12} per generation, and that in the opposite direction is u_{21} . Note that reversible mutation between A_1 and A_2 is included in the life cycle to make the model consistent with `mbs`.

Let the frequency of A_2 in generation t be p_2 . After mutation, it becomes $p_2^* = p_2(1 - u_{21}) + (1 - p_2)u_{12}$. After random mating and selection, its frequency becomes

$$p_2^{**} = p_2^* + \frac{(1 - p_2^*)p_2^*(w_{2.} - w_{1.})}{\bar{w}} \quad (\text{S23})$$

where $w_{1.} = (1 - p_2^*)w_{11} + p_2^*w_{12}$, $w_{2.} = (1 - p_2^*)w_{12} + p_2^*w_{22}$, and $\bar{w} = (1 - p_2^*)w_{1.} + p_2^*w_{2.}$. Let the population size of generation $t + 1$ be N_{t+1} . We draw a random number X from the binomial distribution with parameters $2N_{t+1}$ and p_2^{**} . The frequency of A_2 in generation $t + 1$ is therefore $X/(2N_{t+1})$. Thus, this approach can accommodate arbitrary changes in population size.

Models with strong balancing selection and changes in population size

This set of simulations are intended to check whether the results presented in Figures 5a and 5b are robust to stochastic fluctuation in the frequencies of the selected variants. We assume overdominance with $w_{11} = 1 - s_1$, $w_{12} = 1$, and $w_{22} = 1 - s_2$. The sample size was $n = 20$. Because L (the total branch length) is insensitive to the equilibrium frequencies of A_1 and A_2 when the sample size is not small (see Figure 4), we assumed a symmetric selection model with $s_1 = s_2 = s$ (i.e., the equilibrium frequencies of A_1 and A_2 are 50%). To simulate the population expansion model presented in 5a, we assumed that $N_{e,1} = 20,000$ (the effective population size of the current epoch) and $N_{e,2} = 2,000$ (the effective population size of the ancestral epoch). For the population reduction model in 5b, we used $N_{e,1} = 2,000$ and $N_{e,2} = 20,000$.

For both demographic models, we assumed that the frequency of A_2 at the beginning of the simulations was 50%. We then allowed the population to evolve forward in time for $50N_{e,2}$ generations. The population size was then changed to $N_{e,1}$. We let the population evolve for another $2N_{e,1}t$ generations, where t is the time parameter shown in Figure 5.

If either A_1 or A_2 was lost before the end of the simulation, the process was restarted. The allele frequency trajectory obtained was then passed onto **mbs** to obtain simulated sequence polymorphism data at a linked neutral site. L was estimated using S/θ , where S is the observed number of segregating sites, $\theta = 2N_{e,1}u$, and u is the neutral mutation rate.

Recent balancing selection and selective sweeps

Here we set out to test whether phase-type theory can produce accurate predictions about the SFS. As in Figure 10 in the main text, we considered overdominance with fitnesses $w_{11} = 1 - s_1$, $w_{12} = 1$, and $w_{22} = 1 - s_2$. The corresponding sweep model has fitnesses $w_{11} = 1$, $w_{12} = 1 + s_1$, and $w_{22} = 1 + 2s_1$. In the forward simulations, we assumed that A_2 was the mutant allele, and it appeared as a single copy in the population when it first arose. We then allowed its frequency to evolve. We set $t = 0$ when the frequency of A_2 exceeded $\epsilon = 1/\gamma_1$ for the first time, where $\gamma_1 = 2Ns_1$ and N is the effective population size. We allowed the forward simulation to continue until, time t (in units of $2N$ generations). The trajectory of A_2 was then used by **mbs** for obtaining samples at a linked neutral site. The unfolded SFS as defined in the main text at the neutral site was estimated using data from 10,000 replicates.

S.7 The second moment of the mean time to absorption

The second moment of \mathcal{T} is given by $\frac{d^2 \mathcal{L}(z)}{dz^2} \Big|_{z=0}$. The second derivative with respect to z for the integral in (S16) reads:

$$\frac{d^2}{dz^2} \int_{t_{h-1}}^{t_h} e^{\mathbf{A}_h t} dt = \mathbf{A}_h^{-1} \sum_{k=0}^1 (-1)^k e^{\mathbf{A}_h t_{h-k}} [\mathbf{A}_h^{-2} + (\mathbf{A}_h^{-1} - t_{h-k} \mathbf{I})^2] \quad (\text{S24})$$

Substituting (S24) into (S15) leads to the following result.

Theorem 2. *The second moment of the mean time to absorption, $\mathbb{E}[\mathcal{T}^2]$, is given by:*

$$\alpha \sum_{h=1}^H \left[\prod_{i=1}^{h-1} e^{\mathbf{S}_i d_i} \right] \sum_{k=0}^1 (-1)^{k+1} e^{\mathbf{S}_h (t_{h-k} - t_{h-1})} [\mathbf{S}_h^{-2} + (\mathbf{S}_h^{-1} - t_{h-k} \mathbf{I})^2] \vec{1}. \quad (\text{S25})$$

S.8 Approximating the expected pairwise coalescence time under the model of recent balancing selection

As in the main text, we assume that a new allele A_2 has arisen by mutation, and has spread to a frequency \tilde{p}_2 that is close to its equilibrium value under balancing selection, which is $\hat{p}_2 = s_1/(s_1 + s_2)$ with heterozygote advantage. Providing that the recombination rate is not too high relative to the strength of selection, the expected coalescence time for a pair of A_2 alleles sampled at frequency \tilde{p}_2 can be obtained from Equations 9, 10, 11a and A1-A3 of Charlesworth (2020b), where $\Delta\pi$ in his Equation 11a is equivalent to the reduction in the mean pairwise coalescence time relative to the neutral value of $2N_e$ generations. To obtain $\Delta\pi$, \tilde{p}_2 replaces q_2 in Equations 9, 10 and A1-A3 of Charlesworth (2020b), where the selection parameters in Equations A1-A3 are $\gamma = 2N_e s_1$, $a = 1$, and

$b = -(s_1 + s_2)/s_1$. At the time when \tilde{p}_2 is reached, the values of the expected coalescent times (on the timescale of $2N_e$ generations) for a pair of A_1 alleles is approximately equal to 1.

In addition, the possibility that a recombination event introduces the neutral site from an A_1 allele onto an A_2 background, thereby reducing the initial divergence at the neutral site between an A_1 and A_2 pair, is modelled by using Equation A3a of [Charlesworth \(2020b\)](#) with q_2 replaced with $1-p_2$ and q with $1-\epsilon$, to yield a probability of an A_1 to A_2 recombination event of P_{r1} . In addition, the selection parameters a and b should be replaced with $a+b$, and $-b$, respectively. It is assumed that such a recombination event is followed by coalescence with a non-recombined neutral site associated with A_2 , with a coalescence time equal to the duration of sweep, t_s , as given by [23](#) with $p_2 = \tilde{p}_2$. The divergence between an A_1 and A_2 pair at the time of sampling is then given by $1-P_{r1}(1-t_s)$.

A simple way to obtain the pairwise coalescence times at an arbitrary time after the allele frequency \tilde{p}_2 has been reached is to consider the recursion relations for the corresponding pairwise expected diversity measures with a neutral mutation rate of u under the infinite sites mutation model and assuming that the frequency of A_2 remains close to its equilibrium value. The scaled mutation rate in the absence of selection, $\theta = 2N_e u$, is sufficiently small that second-order terms in θ can be neglected ([Malécot, 1969](#), p. 40; [Wiehe and Stephan, 1993](#), Equation 6a). Writing π_{ij} for the expected diversity for a pair of alleles A_i and A_j , and using primes for their values in a new generation, and neglecting second-order terms, we have:

$$\pi'_{11} = \left[1 - \left(2u + 2r\hat{p}_2 + \frac{1}{2N_e\hat{p}_1} \right) \right] \pi_{11} + r\hat{p}_1\pi_{12} + 2u \quad (\text{S26a})$$

$$\pi'_{12} = 2r\hat{p}_2\pi_{11} + [1 - (2u + r)] \pi_{12} + 2r\hat{p}_1\pi_{22} + 2u \quad (\text{S26b})$$

$$\pi'_{22} = r\hat{p}_2\pi_{12} + \left[1 - \left(2u + 2r\hat{p}_1 + \frac{1}{2N_e\hat{p}_2} \right) \right] \pi_{22} + 2u \quad (\text{S26c})$$

The coefficients of the π_{ij} in these equations provide the corresponding coefficients for the recursions of the deviations of the π_{ij} from their equilibrium values, thereby eliminating the term in $2u$ on the right-hand sides of the equations. If the π_{ij} are scaled relative to their expected value 2θ in the absence of selection, and u is set arbitrarily close to zero, solving for equilibrium gives π_{ij} values relative to 2θ that are equivalent to the equilibrium coalescent times given by [\(8\)](#), as can be verified by direct calculation.

By setting u to zero in [\(S26\)](#), and using the scaled the π_{ij} , we thus obtain a recursion for the deviations from equilibrium of the corresponding expected pairwise coalescence times on the timescale of $2N_e$ generations. While it is possible in principle to diagonalize the relevant matrix, and express the solution for an arbitrary time after reaching \tilde{p}_2 in term of its eigenvalues and eigenvectors, in practice it is simpler to iterate the matrix with assigned numerical values of the parameters. In order to save computation time, a relatively small value of N_e can be used, and the recombination parameters rescaled accordingly to represent a much larger N_e with the same value of $\rho = 2N_e r$. The initial relative values of π_{11} , π_{12} , and π_{22} are 1, 1, and $1 - \Delta\pi$.

References

Charlesworth, B., 2020b How good are predictions of the effects of selective sweeps on

- levels of neutral diversity? *Genetics* **216**: 1217–1238.
- Hobolth, A., A. Siri-Jegousse, and M. Bladt, 2019 Phase-type distributions in population genetics. *Theoretical Population Biology* **127**: 16–32.
- Kaplan, N. L., T. Darden, and R. R. Hudson, 1988 The coalescent process in models with selection. *Genetics* **120**: 819–829.
- Malécot, G., 1969 *The mathematics of heredity*. Freeman.
- Teshima, K. M. and H. Innan, 2009 mbs: modifying Hudson’s ms software to generate samples of dna sequences with a biallelic site under selection. *BMC Bioinformatics* **10**: 166.
- Wiehe, T. H. and W. Stephan, 1993 Analysis of a genetic hitchhiking model, and its application to dna polymorphism data from *drosophila melanogaster*. *Mol Biol Evol* **10**: 842–854.