INTRODUCTION

Understanding the complex interactions between the environment, population dynamics, and adaptive evolution requires detailed analysis of selection components operating at individual stages of the life cycle. Testing for nonrandom genotype frequency changes specific to each of these stages has been the focus of classic selection component analysis, based on adult-offspring samples (Bundgaard & Christiansen, 1972; Prout, 1971). In this context, viability (or zygotic) selection refers to differential survival of genotypes from the zygotic stage to adulthood, sexual selection to differential mating success, fecundity selection to differential offspring production, and gametic selection to prezygotic processes such as segregation distortion. In the case of viability selection, the often dramatic phenotypic...
changes associated with ontogeny, in addition to potentially associated changes in demography and the environment, should result in viability selection variation with age. Importantly, selective pressures operating during critical ontogenic and demographic transitions might have a disproportionate contribution to total viability selection. In the case of long-living tree species, and many other organisms producing vast offspring cohorts, most mortality occurs during the first few months after emergence, frequently exceeding 90% during the first year (e.g., Castro, Gómez, García, Zamora, & Hódar, 1999; Vizcaino-Palomar, Revuelta-Eugercios, Zavala, Alía, & González-Martínez, 2014). Although the intensity of selection is not expected to increase rapidly with early elimination rates (Haldane, 1931), the combination of strong early seedling mortality, high inter-individual variance in seed fecundity (Kang, Bila, Harju, & Lindgren, 2003), and substantial genetic variation at early fitness traits (e.g., Savolainen, Bokma, García-Gil, Komulainen, & Repo, 2004) indicates ample opportunities for strong viability selection at early tree life stages (Hufford & Hamrick, 2003; Petit & Hampe, 2006; Savolainen, Pyhäjärvi, & Knürr, 2007).

Methods for detecting loci associated with viability selection during specific developmental stages are thus essential to understand the evolution of traits that are potential major contributors to total lifetime fitness and therefore important determinants of species niches and range limits under present and future climatic conditions (Donohue, 2014; Donohue, Rubio de Casas, Burghardt, Kovach, & Willis, 2010; Jackson, Betancourt, Booth, & Gray, 2009). Genomic scans for selection are a widely used tool for detecting adaptive loci, based on the comparison of observed versus expected patterns of genetic variation under particular demographic and selective models (Jensen, Foll, & Bernatchez, 2016). Several test statistics are available to identify signatures of selection from polymorphism data taken at a single point in time, such as site frequency spectra, patterns of linkage disequilibrium, or interpopulation differentiation data (reviewed in Bank, Ewing, Ferrer-Admettla, Foll, & Jensen, 2014). These approaches are not suitable for investigating viability selection separately, however, because they are designed to detect footprints of historical selective processes, with target patterns of variation reflecting the cumulative effects of demographic and selective processes over generations, including fecundity, sexual, and gametic selection. Alternative methods use genomic time series data, tracking allele frequency trajectories over multiple generation samples, within and/or among populations (Bollback, York, & Nielsen, 2008; Feder, Kryazhimskiy, & Plotkin, 2014; Ferrer-Admettla, Leuenberger, Jensen, & Wegmann, 2016; Foll, Shim, & Jensen, 2015; Gompert, 2016; Malaspinas, Malaspinas, Evans, & Slatkin, 2012; Mathieson & McVean, 2013; Nishino, 2013; Schraiber, Evans, & Slatkin, 2016; Shim, Laurent, Matuszewski, Foll, & Jensen, 2016). The temporal dimension usually allows more powerful joint inference of selection coefficients and demographic parameters such as effective population size and gene flow, in the necessary disentanglement of the confounding effects of demography (versus selection) on allele frequency change across generations (Malaspinas, 2016). Even if these methods were powerful enough to infer selection over as few as two consecutive generations, however, they would still be measuring total selection and thus failing to isolate signatures of viability selection.

Genomic scans for viability selection acting on specific developmental stages require methods capable of testing allele frequency change caused by mortality during the target ontogenic period, within a single generation. For this purpose, alogistically appealing approach for long-living tree species would consist of simultaneously sampling for genotyping several cohorts of different ages coexisting nearby within the same population. This rapid cross-sectional sampling scheme has been used frequently in studies of neutral molecular variation across tree life stages (e.g., Alvarez-Buylla, Chaos, Piñero, & Garay, 1996). But this strategy is questionable, at least for viability selection inference, because the ontogenetic staging of cohorts recruited at different times being necessarily asynchronous, observed allele frequency differences will be inflated by cross-cohort differences in parental structure (e.g., due to masting and/or interannual pollination variation) and by temporal variation in selection.

Longitudinal cohort studies circumvent this problem, by sequentially sampling a target cohort of same-age individuals multiple times as they age and die, and are more suitable for analysis of viability selection (e.g., Christiansen, Frydenberg, & Simonsen, 1978; Edwards & Heath, 1983). The few available genome-wide analyses of selection based on longitudinal cohort studies have investigated contemporary allele frequency change caused by mortality under experimental selective regimes (Anderson, Lee, & Mitchell-Olds, 2014; Gompert et al., 2014; Pespeni et al., 2013), after mass die-offs caused by toxins in the wild (De Wit, Rogers-Bennett, Kudela, & Palumbi, 2014), and in response to unknown natural selective pressures operating during critical ontogenic and demographic transitions (Bourret, Dionne, & Bernatchez, 2014) or during invasion (Monnahan, Colicchio, & Kelly, 2015). All of these studies formally tested the null neutral hypothesis that observed temporal allele frequency changes were caused by sampling and stochastic mortality alone, while two of the studies additionally estimated selection coefficients. In particular, studies where it was feasible to track and genotype exhaustively every single individual in the (small) experimental cohort did conduct selection coefficient estimation, based on observed absolute differences in survival probability between alternative homozygotes (Anderson et al., 2014; Gompert et al., 2014). The other studies focused on big natural or experimental populations, in which individual tagging and exhaustive sampling were unfeasible, so instead independent random samples were taken from the large target cohorts before and after the selective episode of interest, and no attempts were made to estimate selection coefficients from such incomplete longitudinal samples (Bourret et al., 2014; De Wit et al., 2014; Monnahan et al., 2015; Pespeni et al., 2013). The null hypothesis of genotypic-independent mortality was still tested, either based on permutation analyses of genotypes among temporal samples (De Wit et al., 2014; Pespeni et al., 2013), homogeneity tests of genotypic proportions (Monnahan et al., 2015), or through a direct application of outlier tests designed for the analysis of historic adaptive differentiation.
among populations (Bourret et al., 2014). None of these studies evaluated the expected statistical performance of the employed methods with simulations, as would be required to determine their error rates and reliability (Lotterhos & Schaal, 2014).

As in the case of many tree species, natural populations of numerous other plant and animal taxa are large, with vast seedling or larval cohorts in which individual tagging and tracking may be difficult and costly. Further methods to infer viability selection from independent random samples in longitudinal cohort mortality studies would thus be broadly useful, facilitating the investigation of adaptive processes at particular ontogenetic stages and contemporary selective episodes. For instance, monitoring contemporary selection in exploited fisheries or forests during critical developmental or demographic stages could provide valuable real-time information to orient adaptive conservation management. Measuring stage-specific contemporary selection should also be central in studies tracking ongoing global-change effects on the evolutionary ecology of wild or managed populations. We propose here tools to address this kind of practical questions, introducing tests of neutrality and a Bayesian scheme to estimate viability selection coefficients from independent random genotypic samples, taken sequentially from a large cohort within a generation. Using simulations, we analyze the accuracy and expected error rates of these methods under contrasting sampling and selective scenarios. Using SNP markers within a large number of candidate genes, we apply the approach to investigate viability selection operating at early recruitment stages in a relict Mediterranean maritime pine (Pinus pinaster Ait.) population.

2 | MATERIAL AND METHODS

2.1 | Demographic model

We assume a cohort of diploid individuals, of which a set of \( n_0, n_1, \ldots, n_T \) are sampled randomly at an initial reference time step \( t = 0 \) and at \( T \) subsequent time steps, with \( T \geq 1 \). The cohort size throughout the sampling period \( (N_0, N_1, \ldots, N_T) \) is assumed to be unknown. We assume that the sampled cohort is coetaneous, and that no new individuals join it after \( t = 0 \), neither through reproduction nor through migration. The cohort is followed over time to determine whether genotypic-dependent mortality occurs throughout one or several demographic or environmental episodes of interest, temporarily delimited by the sampling intervals. For this purpose, the temporal samples are genotyped at \( L \) biallelic loci, yielding the \( \mathbf{y} = (y^{11}_0, y^{12}_0, y^{22}_0) \) vector of observed genotypic counts, where \( y^{11}_0, y^{12}_0 \), and \( y^{22}_0 \) are the number of individuals in the sample collected at time \( t \) that are homozygous for the first allele, heterozygous, and homozygous for the second allele at locus \( i \), respectively.

We consider as point of reference for measuring genotypic frequency change the initial genotypic frequencies in the zygote pool originating the target cohort \( \mathbf{p}_0 \), immediately after gamete fusion. The \( N_0 \) genotypes in the cohort at \( t = 0 \) can then be assumed to be multinomial draws from the initial zygote pool, while the \( n_0 \) genotypes in the initial temporal sample are in turn drawn from a finite population (cohort) of size \( N_0 \). This two-step sampling is equivalent to one-step multinomial sampling of \( n_0 \) genotypes from the initial pool of genotypes, no matter whether the \( n_0 \) individuals are sampled from the cohort with replacement or not (Nei & Tajima, 1981; Waples, 1989). This equivalence holds as well for subsequent samples \( (t \geq 1) \) in the absence of selection. If selection operates after the initial reference sample is collected at \( t = 0 \), however, then expected genotypic frequencies for \( t > 0 \) will diverge from \( \mathbf{p}_0 \).

2.2 | Quasi-exact and Monte Carlo tests of neutrality

We propose a quasi-exact test and a fast Monte Carlo approximation to test the null hypothesis of genotypic-independent mortality (i.e., neutrality) in the cohort. The question is whether observed genotype frequency changes among temporal samples are greater than those expected from random mortality and sampling alone. For the \( i \)-th locus, we used the standardized variance in allele frequency for measuring the observed allele frequency change between time steps \( t - 1 \) and \( t \) (Pollak, 1983; see Waples, 1989 for the two-allele expression)

\[
\Delta_{\text{obs}} = \frac{(f_y - f_y^0)^2}{(f_y + f_y^0)/2 - (f_y + f_y^0)/2}^2
\]

where \( f_y = (2y^{11}_t + 2y^{12}_t)/(2n_0) \) and \( f_y^0 = (2y^{11}_0 + 2y^{12}_0)/(2n_0) \). We allow for differences in sample size across loci to account for possible differences in missing data. Under the null (neutral) hypothesis, the \( n_i \) genotypes are multinomial draws from the initial zygote pool \( \mathbf{p}_0 \), where \( n_i \) will diverge from \( p_i \) at the expected genotypic frequencies for \( n_0 \). We propose a quasi-exact test of neutrality based on a single indicator variable \( I_{ijkm} \), where \( I_{ijkm} = f_y^{i(km)}/f_y \), given \( f_y \), \( I_{ijkm} \), and the first and second allele frequencies for the initial and final sample.

\[
\text{Pr}(\Delta_{\text{obs}} \geq \Delta_{\text{obs}}(\mathbf{p}_0) \mid \mathbf{p}_0) = \sum_{n_{t-1}^{(0)}} \sum_{n_{t-1}^{(0)}} \sum_{n_{T-1}} \sum_{n_{T-1}} I_{ijkm} \cdot \text{Mult}(i,j,n_t - i - j,\mathbf{p}_0) \cdot \text{Mult}(k,m,n_{t-1} - k - m,\mathbf{p}_0)
\]

where \( I_{ijkm} \) is an indicator function having the value 1 if \( \Delta_{\text{obs}} \geq \Delta_{\text{obs}}(\mathbf{p}_0) \) and the value 0 otherwise. The \( \Delta_{\text{obs}}(\mathbf{p}_0) \) values are calculated as for \( \Delta_{\text{obs}} \), but replacing \( f_y \) with \( f_y = (2i + j)/2n_0 \) and \( f_y^0 = (2i + j)/2n_0 \). The \( \mathbf{p}_0 \) frequencies are unknown, so we estimated \( \mathbf{p}_0 = (p^{11}_0, p^{12}_0, p^{22}_0) \) under the null hypothesis that all temporal samples are drawn from the same initial pool and under the prior assumption that all three possible genotypes are present in the cohort (even if any of them remains undetected; see Rannala & Mountain, 1997), as

\[
p^{11}_0 = (y^{11}_0 + 1)/3, \quad p^{12}_0 = (y^{12}_0 + 1)/3, \quad p^{22}_0 = (y^{22}_0 + 1)/3
\]

where \( y^{11}_0, y^{12}_0, \) and \( y^{22}_0 \) are the observed counts of genotypes 11, 12, and 22 in the total pooled sample, respectively, and \( n_0 \) is the
pooled sample size. We will refer to the test in Equation 2 as the quasi-exact neutrality test, to reflect the fact that, although it is nonparametric and based on exhaustive combinatorial enumeration, it relies on the estimation of initial allele frequencies. Unlike homogeneity tests used in adult-offspring selection component analysis (Christiansen & Frydenberg, 1973; Monnahan et al., 2015), our test does not require Hardy–Weinberg equilibrium assumptions. And in contrast to homogeneity tests specifically designed for the analysis of a sequentially sampled cohort (Christiansen et al., 1978; Edwards & Heath, 1983), our quasi-exact test is not based on parametric asymptotic approximations, expected to break down for low genotypic counts and rare alleles (e.g., Cressie & Read, 1989).

The quasi-exact neutrality test becomes however extremely time-consuming for large sample sizes. Therefore, we also employed a fast approximation, by approaching the expected neutral distribution of $\Delta_t$ with a random subsample of the possible outcomes of the compound binomial draws. In particular, we used the following Monte Carlo simulation algorithm:

1. Randomly draw the number of individuals from each genotypic class in the simulated samples at time steps $t$ ($y^\text{sim}_{t-1}$) from a multinomial distribution with three classes with probabilities $[p_{11}^{(t)}, p_{12}^{(t)}, p_{22}^{(t)}]$ and $n_t$ and $n_{t-1}$ trials, respectively.
2. Compute the standardized allelic frequency increment in the simulated sample ($\Delta^\text{sim}_t$) as in Equation 1, using allelic frequencies calculated from the simulated genotypic counts $y^\text{sim}_t$ and $y^\text{sim}_{t-1}$.
3. Repeat steps (1) and (2) 10,000 times and calculate the $p$-value as the proportion of simulated replicates where $\Delta^\text{sim}_t > \Delta^\text{obs}_t$.

### 2.3 Bayesian inference of selection coefficients

Under selection, expected genotypic frequencies for $t > 0$ will diverge from $p_0$. Denote the relative fitness of each genotype as $w_r^{11} = 1 + 2s_h$ and $w_r^{22} = 1 + 2h_i$, where $s_h$ is the selection coefficient for locus $l$ and time step $t$, and $h_i$ is the heterozygous effect for locus $l$. The expected genotypic frequencies after selection for $0 < t \leq T$ are then (Gillespie, 2004)

$$E[p_{r1}^{(t)}|p_{r0}, s_h, h_i] = \frac{p_{r0}^{11} w_r^{11}}{\bar{w}_r}$$

$$E[p_{r2}^{(t)}|p_{r0}, s_h, h_i] = \frac{p_{r0}^{22} w_r^{22}}{\bar{w}_r}$$

$$E[p_{r12}^{(t)}|p_{r0}, s_h, h_i] = \frac{p_{r0}^{12} w_r^{12}}{\bar{w}_r}$$

where $\bar{w}_r = w_r^{11} p_{r0}^{11} + w_r^{12} p_{r0}^{12} + w_r^{22} p_{r0}^{22}$ is the mean fitness of the cohort.

Given the above assumptions and the unknown initial genotypic frequencies $p_0$, selection coefficients $s_h$, and dominance coefficient $h_i$, the probability of observing the genotypic count $y_{0t}$ at locus $l$ in the initial sample is given by

$$Pr(y_{0t}|p_0) = \frac{n_t}{y_{0t}^{11} y_{0t}^{12} y_{0t}^{22}} \frac{(p_{01}^{11})^{y_{0t}^{11}} (p_{01}^{12})^{y_{0t}^{12}} (p_{01}^{22})^{y_{0t}^{22}}}{\pi^{y_{0t}^{11}} \bar{w}_r^{y_{0t}^{12}} \bar{w}_r^{y_{0t}^{22}}}.$$  (4)

while the probability of observing genotypic count $y_{it}$ for $t > 0$ can be written

$$Pr(y_{it}|p_{0t}, s_{it}, h_i) = \frac{n_t}{y_{it}^{11} y_{it}^{12} y_{it}^{22}} \frac{(p_{it}^{11})^{y_{it}^{11}} (p_{it}^{12})^{y_{it}^{12}} (p_{it}^{22})^{y_{it}^{22}}}{\pi^{y_{it}^{11}} \bar{w}_r^{y_{it}^{12}} \bar{w}_r^{y_{it}^{22}}}.$$  (5)

where $s_{it}$ is the vector of size $t$ with the values of the selection coefficients for locus $l$ at previous time steps (from 0 to $t - 1$). In computing $Pr(y_{it}|p_{0t}, s_{it}, h_i)$ for the $i$-th time step ($i > 0$), the vector of expected genotypic frequencies are calculated sequentially from $t = 1$ to $i$ using Equation 3, conditional on initial genotypic frequencies $p_{0t}$, selection coefficient values at previous ($t < i$) time steps $s_{it}$, and dominance coefficient $h_i$. Note that multinomial Equation 5 assumes either that individual sampling is nondestructive (i.e., with replacement) for $t > 0$ or that it is destructive but from a cohort of large enough size ($N_l > n_l$), so that the multinomial approximates sufficiently well the actually hypergeometric sampling process.

The likelihood for locus $l$ considering all time steps is then

$$Pr(y_l|p_0, s_l, h_l) = Pr(y_0|p_0) \pi_{t=1}^{T} Pr(y_{it}|p_{0t}, s_{it}, h_i)$$  (6)

We employed uninformative prior distributions ($\pi$) for all parameters, namely a flat Dirichlet prior for the genotypic frequencies at each locus and time step, $p_0 = \text{Dir}(a = 1)$, a uniform prior on the interval $(-0.5, 0.5)^6$ for the selection coefficient of each locus and time step ($s_{it}$), and a uniform prior on the interval (0, 1) for the dominance coefficient of each locus ($h_i$).

Given the $y_i$ vector of genotypic counts for locus $l$, the joint posterior distribution over parameter set $\Theta_l = (p_{00}, s_l, h_l)$ is given by Bayes’ rule:

$$f(\Theta_l|y_l) \propto Pr(y_l|p_0, s_l, h_l)/f(p_0, s_l, h_l)$$  (7)

where the $f$ functions on the right-hand of the equation are the prior distributions. For each locus, the joint posterior distribution of Equation 7 was estimated using the MCMC algorithm described in Supporting information Appendix S1.

### 2.4 Simulation study of methods performance

Using the Monte Carlo simulations detailed in Supporting information Appendix S2, we calculated the power and Type I error rate of the quasi-exact neutrality test (Equation 2) and its Monte Carlo fast approximation. We also computed the expected bias, accuracy (root mean square error, RMSE), and credible interval noncoverage rate (NCR) of Bayesian estimates of selection coefficients (Equation 7). We considered a cohort of $N$ individuals, from which two temporal samples of size $n$ were taken and genotyped at one locus with initial minor allele frequency MAF and selection coefficient $s$. We assumed codominance ($h = 0.5$) and investigated the effect of $s$ estimates of variable levels of $N$ (from $10^5$ to $10^6$), of $n$ (ranging from $10^4$ to $10^5$), of $s$ ($0$, $-0.01$, and $-0.1$), and of MAF from (0.05 to 0.5). Exhaustive sampling scenarios (i.e., $n = N$) were included to assess the performance of the methods in an extreme reference setting without random sampling effects, even if the presented methods are actually inappropriate in that case, as the independent multinomial (or hypergeometric
2.5 | Empirical study site, field sampling, and laboratory analysis

We studied a small native population of *P. pinaster* at Fuencaliente, South-Central Spain (38°25′N, 4°15′W), comprising around 300 adult individuals over an isolated 0.1-km² steep outcrop of quartzite, at one of the Iberian sites of the species with harshest edaphoclimatic conditions (Charco, Venturas, Gil, & Nanos, 2017; Ramírez-Valiente & Robledo-Arnuncio, 2014). The poor rocky soil, steep south-facing slope (averaging 40%), high summer temperatures (with maxima occasionally exceeding 40°C), and low summer precipitation (averaging 40%), translated into very similar RMSE in general, except for the lowest sample size considered (n = 100), for which the smaller RMSE of the median was more evident (white versus gray bars in Figure 1). Irrespective of whether it was calculated with respect to the median or to the mean, the corresponding variances were similar and comparatively high, translating into very similar RMSE in general, except for the lowest sample size considered (n = 100), for which the smaller RMSE of the median was more evident (white versus gray bars in Figure 1). Irrespective of whether it was calculated with respect to the median or to the mean, the RMSE of ̂s was not affected by minor allele frequency variation between MAF = 0.3–0.5, but it increased markedly as the minor allele became more rare (for MAF = 0.1 and especially for MAF = 0.05; Figure 1). The increase in the RMSE of ̂s with decreasing MAF was largely driven by an increment of the variance, as the bias of ̂s either tended to actually decrease with decreasing MAF (when calculated relative to the median) or it showed a nonmonotonic relationship with MAF (when calculated relative to the mean). Overall, considering as reference minimum RMSE of ̂s the one obtained when exhaustively sampling the entire cohort (RMSE ≤ 0.01 for n = N = 10,000), the RMSE increased substantially for n < N, remaining at or below 0.1 only in the case of the largest random sample size considered (n = 1,000) with MAF ≥ 0.1, or in the case of a smaller sample (n = 500) with MAF ≥ 0.3. Uncertainty estimates of the method remained good in all scenarios considered, however, as the noncoverage rate (NCR) of the 95% credibility intervals (CI) of ̂s remained around or below the nominal value of 0.05 even for the lowest sample size and MAF value (Figure 1). In other words, in the absence of selection the method would produce 95% CI for ̂s correctly including zero as frequently as expected.

The simulation results revealed qualitatively similar trends in the case of a strongly selected locus (s = −0.1) than in the case of a neutral one, with analogous effects of sample size and minor allele frequency on estimation errors (Figure 2 versus Figure 1). Although the (absolute) bias and RMSE of ̂s were slightly lower than in the absence of selection for all the scenarios considered, the RMSE of ̂s was around or below its assumed absolute value (0.1) only when n = 1,000 and MAF ≥ 0.1, or when n = 500 and MAF ≥ 0.3 (Figure 2).
The NCR remained around or below its nominal value in all cases, meaning that the method would produce 95% CI for $\hat{s}$ correctly including the value $-0.1$ as frequently as expected.

In the case of weak selection ($s = -0.01$), simulations showed again similar trends in the accuracy of $\hat{s}$ as a function of $n$ and MAF as in the other selective scenarios, with absolute values for the RMSE of $\hat{s}$ generally slightly larger than those obtained for strong selection, and similar to those obtained under neutrality (Figure 3 versus Figures 1 and 2). With respect to the low assumed value ($s = -0.01$), however, the relative RMSE of $\hat{s}$ became much larger for weak than for strong selection, reaching minimum values of 50–100% when the cohort was exhaustively sampled, and exceeding 1,000% even when $n = 1,000$ and MAF $\geq 0.1$, or when $n = 500$ and MAF $\geq 0.3$ (values relative to the mean; Figure 3 shows absolute values). The NCR remained close to its nominal value in all cases (Figure 3), with wide CIs reflecting the uncertainty in $s$ estimation (Figure 4).

3.2 | Expected power and error rates of neutrality tests

For a given sample size $n$, the RMSE and NCR of $\hat{s}$ were weakly sensitive to the cohort size $N$, irrespective of the assumed values of $s$ and MAF (Figures 1–3 versus Supporting information Figures S2–S10). The exhaustive sampling scenarios ($N = n$) were exceptional, however, as the RMSE and (especially) the NCR of $\hat{s}$ decreased relative to scenarios with the same $n$ but with $n < N$ (Figures 1–3 and Supporting information Figures S2–S10).

FIGURE 1 Effect of sample size ($n$) and minor allele frequency on selection coefficient estimates for a neutral locus ($s = 0$). RMSE is the root mean square error, and NCR the noncoverage rate of 95% credible intervals (the dotted line shows the nominal 5% value). Bias and RMSE were measured with respect to the mean (white bars) or with respect to the median (gray bars). Based on 1,000 Monte Carlo replicates per scenario, assuming a cohort of size $N = 10,000$ and a biallelic locus for a given sample size $n$, the RMSE and NCR of $\hat{s}$ were weakly sensitive to the cohort size $N$, irrespective of the assumed values of $s$ and MAF (Figures 1–3 versus Supporting information Figures S2–S10). The exhaustive sampling scenarios ($N = n$) were exceptional, however, as the RMSE and (especially) the NCR of $\hat{s}$ decreased relative to scenarios with the same $n$ but with $n < N$ (Figures 1–3 and Supporting information Figures S2–S10).
approximation for \( n > 500 \), as the quasi-exact test became prohibitively time-consuming in the latter cases.

The expected Type I error (false positive) rate of the neutrality test (assuming \( s = 0 \)) was zero in the extreme exhaustive sampling scenario, and ranged between 3 and 6% for the different values of sample size \( n \) and minor allele frequency MAF considered (Figure 5, top), that is, close to the nominal 5% value in all cases. The expected false positive rate of the neutrality test was quite similar to the NCR of 95% CI of \( \hat{s} \) obtained with the Bayesian inference method when \( s = 0 \), although the NCR tended to be very slightly larger for the smaller sample sizes and lower minor allele frequencies in the simulations (Figure 1 and also plotted for reference in Figure 5, top).

The power of the neutrality test was moderate-to-low for a strongly selected locus (\( s = -0.1 \)), increasing with raising \( n \) and MAF (Figure 5, middle). Power reached up to 90–100% in the reference scenario with exhaustive sampling, decreasing to 20–50% for random samples of size \( n = 500–1,000 \), provided MAF \( \geq 0.3 \), down to lows of 5–15% for the smallest samples (\( n = 100 \)) or for larger samples when MAF \( \leq 0.1 \) (Figure 5, middle). We compared the power of the neutrality test with that of the Bayesian approach, defined as the proportion of times the 95% CI of Bayesian \( \hat{s} \) estimates did not include zero (when \( s \neq 0 \)). The power of the Bayesian approach also increased with \( n \) and MAF, with the neutrality test having slightly greater power in general, especially for the smallest sample size, but slightly lower power for larger samples with MAF \( \leq 0.1 \) (Figure 5, middle).

The neutrality test most often failed to reject the null neutral hypothesis for loci under weak selection (\( s = -0.01 \), with a very low power of 4–6% for all simulated values of \( n \) and MAF (Figure 5, bottom), of the same order than the expected false positive rate for neutral loci.

**FIGURE 2** Effect of sample size (\( n \)) and minor allele frequency on selection coefficient estimates for a locus under strong selection (\( s = -0.1 \)). RMSE is the root mean square error, and NCR the noncoverage rate of 95% credible intervals (the dotted line shows the nominal 5% value). Bias and RMSE were measured with respect to the mean (white bars) or with respect to the median (gray bars). Based on 1,000 Monte Carlo replicates per scenario, assuming a cohort of size \( N = 10,000 \) and a locus with two codominant alleles.
Power was even zero in the theoretical exhaustive sampling scenario, suggesting that the violation of the independent sampling assumption precluded the statistical detection of weak selection. The power of the Bayesian approach was slightly below that of the neutrality test for all simulated weak selection scenarios (Figure 5, bottom).

For a given sample size \( n \), changes in the cohort size \( N \) did not substantially alter the power and false positive rate of either the neutrality test or the Bayesian approach, except for the extreme reference scenarios with exhaustive sampling \( (N = n) \), when the model assumptions are violated and power generally dropped relative to cases with the same \( n \) but \( N > n \), while the expected false positive rate remained invariably close to nominal levels (Figure 5 versus Supporting information Figures S11-S13).

### 3.3 Genomic analysis of early selection in *Pinus pinaster*

The 384-plex SNP assay produced reliable seedling genotypes at 356 loci, of which 302 were polymorphic. The genomic analysis of selection was conducted on 237 (78%) polymorphic loci having MAF > 0.05, which exhibited a substantial range of temporal allele frequency change, somewhat wider during the first (winter) than during the second (spring-summer) target periods (Figure 6). The allele frequency change was not larger than expected by random mortality and sampling for most loci, as indicated by the quasi-exact neutrality test (Figure 6). In particular, there was no significant evidence of genotypic-dependent mortality during the second study period for
any locus. During the first period, however, allele frequencies at 5 loci showed temporal changes significantly larger than expected from random mortality and sampling (Figure 6 and Table 1). The same set of loci was identified with the quasi-exact test and the fast Monte Carlo approximation (results not shown). Bayesian estimates of selection coefficients were significantly different from zero for the five loci.
that exhibited exceptional allele frequency change, with the strength of selection, based on median point estimates, ranging from 0.29 to 0.47 (Table 1). The loci potentially under selection are found in genes related with water stress response (m650), abiotic stress responses (m1115), disease resistance (m1496), or more general functions such as cellular signaling (m102) and protein assembly (m698).

4 | DISCUSSION

Motivated by the need to quantify contemporary selection acting on specific ontogenetic stages, we have proposed and evaluated a neutrality test and a Bayesian method to infer selection coefficients using random genotypic samples drawn from a longitudinal cohort mortality study. The approach should be of interest for researchers investigating selective mortality in large natural or experimental cohorts of any organism in which individual tagging and tracking are unfeasible. We have applied the proposed methods in a natural maritime pine seedling cohort, detecting five loci undergoing exceptional allele frequency change during a short early establishment stage with high elimination rates.

4.1 | Measuring selection from cohort mortality studies

Our model represents a first attempt to explicitly estimate contemporary selection coefficients from random genotypic samples taken sequentially from a large cohort mortality study, within a generation. Even if genetic cohort studies are free from genetic drift effects complicating selection inference from multigeneration time series (e.g., Feder et al., 2014), our simulation analysis shows for the first time that distinguishing within-generation allele frequency changes caused by selective mortality from the stochastic noise produced by sampling and genotype-independent mortality remains an inherently difficult statistical exercise. A reassuring aspect of the results, however, is that the method does not tend to mistake neutral loci as being under selection more frequently than expected, since the false positive rate of the neutrality test and the NCR of Bayesian $s$ estimates (when $s = 0$) remain close to nominal values, independently of sample size and minor allele frequency. Results are also promising in terms of the power to discriminate loci under strong selection, which should be achievable with moderate or even high confidence by sufficiently increasing the sample size, provided the target locus is not close to fixation (Figure 5, middle). Moreover, median-based point estimates of the selection coefficient obtained with the Bayesian method can be expected to be minimally biased for strongly selected loci, although, as a consequence of their large variance, large samples will be necessary to reduce the relative RMSE below 100% (Figure 2). Prospects for detecting and quantifying weak selection are bleaker, as the probability to discriminate a weakly selected locus is similar to the probability of mistaking a neutral locus as nonneutral (Figure 5), and as the RMSE of Bayesian...
4.2 | Model assumptions and perspectives

Unlike statistical methods to infer selection from multigeneration genetic time series data, our model does not require assumptions on (or joint estimation of) historical demographic and reproductive processes that may erase, alter, or mimic footprints of selection on genotypic frequency change across generations, such as fluctuations in effective population size and gene flow among divergent populations. Nor does the model assume Hardy–Weinberg equilibrium within the sampled cohort, as genotypic frequencies are estimated directly, rather than based on expected random-mating proportions. The method relies however on some potentially limiting assumptions that should be noted. First, samples taken from the target cohort are assumed to be free of immigrants and recruits newly established after the initial sampling step, either because they do not exist or because they can be discriminated during sampling (as they could represent a confounding source of temporal genotypic frequency change). This requirement will be easily met in experimental studies, but also in natural populations of many organisms, such as plant species for which both seed dispersal and germination occur in discrete synchronous periods, or animal species with synchronous reproduction and temporally coherent juvenile cohorts (e.g., anadromous fishes). Otherwise, special caution should be taken, for instance when studying seedling cohorts in plants with spread germination periods, since genetically determined emergence phenology (Donohue, 2005) might inflate temporal variation in genotypic frequencies. The model also assumes a cohort of same-age individuals, considering as reference population the initial zygotic pool that originated such cohort. If multiple age classes were present, then the total cohort itself (at the time of the initial sample) could be considered as reference population, and our methods should presumably be robust to this reference change as long as the cohort is large.

Another assumption of the model is that sampling is random, that is, that the probability of sampling a genotype does only depend on its frequency in the cohort at the time of collection. This assumption could be violated if both individuals and genotypes are spatially structured in the study area and sampling is biased toward high-density groups, especially if there are temporal changes in local density (e.g., due to stochastic plant mortality or animal movement). To avoid this problem, individual sampling should be proportional to local abundance at the time of sampling. The random sampling assumption would also be violated if there were genetically determined phenotypic differences among individuals that could bias sampling between temporal samples (e.g., sex-biased sampling between life stages in anadromous salmons; Bourret et al., 2014).

If in practice individuals were sampled without replacement, the model would then also be assuming that the cohort under study is large, so that the distribution of observed genotypic counts after selection is approximately multinomial, remaining therefore largely independent of the unknown cohort size $N$. It would be possible to drop the large cohort size assumption in cases where sampling is destructive, by assuming a hypergeometric distribution instead of a multinomial in Equation 5. The latter would require however joint estimation of the unknown cohort size $N$, which in practice results in important numerical optimization problems, while yielding very similar estimates of the parameters of interest ($s$ and $h$) and largely uninformative posterior distributions of $N$, which suggests that the multinomial approximation under destructive sampling is good unless $N$ is small (results not shown). The presented simulation results showed that exhaustively sampling a small cohort, which violates the independent sampling assumption, tends to hamper the statistical detection of selection (tests become nonsignificant), especially for weak selection, even if the RMSE of the (nonsignificantly different from zero) selection coefficient estimates actually decrease. This is not, however, a drawback of our random-sampling oriented methods, as they are not intended for studying a cohort small enough that every individual could be tracked, in which case more straightforward and powerful approaches should be employed (e.g., Anderson et al., 2014; Gompert et al., 2014).

In summary, as guidance for empirical studies, our approach should work best when large independent random genotypic samples are sequentially taken from a vast cohort of same-age individuals, in which immigrants or new recruits arriving after the time of the first temporal sample either do not exist or can be easily discriminated. Ideally, the cohort should consist of same-age individuals, but the presence of multiple age classes is unlikely to compromise inference if the cohort is large. As a rule of thumb, loci under strong selection should be detectable with acceptable power, and the corresponding selection coefficients estimated with acceptable accuracy, for sample sizes in the order of $10^3$ genotypes randomly taken from a cohort of size in the order of $10^4$ individuals or more.

Several further tests and developments of the presented model are possible. We did not investigate the effect of the number of temporal samples on selection estimates. The model assumes two or more temporal samples, collected before and after one or more consecutive periods with potentially variable selection. Although separate selection coefficients are estimated for each period, they are not independent, because the expected genotypic frequencies at the end of each time interval are calculated conditional on initial frequencies and selection coefficients during that and preceding periods. Preliminary simulation results (not shown) indicate that adding temporal samples does not significantly affect the accuracy of per-period selection coefficient estimates, assuming codominance during simulations and inference. Future studies may address how unknown and contrasting heterozygous effects, and their joint estimation, affect selection estimates, an underexplored issue in most statistical methods inferring selection, including those based on time series data (but see Teshima, Coop, & Przeworski, 2006; Shim et al.,...
Although estimating unknown heterozygous effects should complicate inference of selection coefficients (Shim et al., 2016), adding temporal samples might improve their joint estimation if the heterozygous effect remained approximately constant across sampling periods.

Our model neither makes assumptions nor provides information on whether correlated selection among loci occurs or not (e.g., due to genetic linkage or epistasis). Estimated selection coefficients represent differences in expected single-locus marginal fitness, that is, difference in relative fitness between the alternative homozygotes in a locus, averaged over genomic backgrounds (Ewens, 2004; Gompert et al., 2014). The ability of the model to detect selection will thus depend on the size of expected locus-specific effects, including direct (for causal genetic variants) and/or indirect (for loci correlated with causal variants) effects. The method can be expected to have rather low power in the case of polygenic selection acting on many small-effect loci, as is the norm in single-locus-based genomic scan methods. In that case, attempting the explicit estimation of selection coefficients may not prove a rewarding statistical exercise, and alternative multivariate approaches testing nonrandom temporal covariation in allele frequencies among loci might be more helpful for detecting traces of selection (e.g., Bourret et al., 2014; Laporte et al., 2016).

4.3 | Contemporary selection on maritime pine recruits

We investigated for the first time whether genotypically based selection could determine nonrandom mortality during a critical early developmental stage in a tree species. In line with the model assumptions, seedling emergence is expected to be largely synchronous in the species, and samples were obtained over the entire population area, proportionally to local abundance and nondestructively whenever possible (Unger et al., 2016). Despite the vast cohort size at the initial seed-crop stage (also in line with the model assumptions), and even if collected seedlings represented a considerable portion of surviving individuals at the time of sampling (with virtually exhaustive sampling during the last period), the low emergence rates and strong early mortality in the area did not allow large samples, which according to the simulations likely reduced substantially the power of the selection analysis, though without increasing the false positive rate. Still, consistently with recent genomic scans of selection based on prospective cohort studies (De Wit et al., 2014; Gompert et al., 2014; Pespeni et al., 2013), our results provide evidences of selective elimination during particular life stages inducing significant allele frequency changes within a single generation. Selection at early life stages can define adaptive potential across the life cycle, as it changes the genetic background available for selection at subsequent life stages, especially in novel environments or extreme climatic events, and when pleiotropy is present (Donohue, 2014).

We used candidate-gene marker loci, so positive tests must necessarily correspond to putatively functional SNPs. It is however noteworthy that three of the five identified loci potentially under selection during the first 4 months after seedling emergence have been related with functions that could influence survival precisely at early stages. In particular, the putative ADP-ribosylation factor (SNP m650; Table 1) has shown altered expression levels in P. pinaster seedlings subject to experimental water deficit (Dubos et al., 2003). The SNP m102 encodes proteins related to the protein-kinase family, associated with general cellular signaling functions increased during pine seedling development (Ávila, Pérez-Rodríguez, & Cánovas, 2006). And the putative NAC transcription factor encoded by m698 belongs to a family apparently involved in abiotic stress responses of P. pinaster seedlings (Pascual, Cánovas, & Ávila, 2015). Of the two other identified loci, the putative protein encoded by m1496 is related to the TIR/NBS/LRR disease resistance protein family, found in Arabidopsis and other plant genomes including pines (Meyers, Morgante, & Michelmore, 2002), and the 6-phosphogluconate dehydrogenase protein-coding gene (m1115) is associated with general abiotic stress responses in several model plant species (Hou, Huang, Yu, & Zhang, 2007).

The estimated average absolute strength of selection experienced by the five loci over the 4-month period was rather high (0.362), though of the same order than the one estimated for loci undergoing exceptional temporal allele frequency change in another short-term cohort mortality study, which focused on rapid adaptation of stick insects to a sudden host shift (Gompert et al., 2014). These values should obviously not be regarded as unbiased estimates of the average strength of genome-wide selection, since they correspond to a small subset of sampled loci with the strongest evidence of selection, and estimates of locus-specific size effects based on crossing significance thresholds are expected to be upwardly biased (Göring, Terwilliger, & Blangero, 2001; Ioannidis, 2008; Williams & Haines, 2011). Moreover, the observed exceptional allele frequency changes were probably determined by both selective mortality and sampling noise, with their relative contributions being unknown (Gompert et al., 2014).

Apart from statistical and stochastic factors, two main biological reasons could help explain the significant signals of strong stage-specific selection observed in our field cohort study. First, drought stress is a major driver of pine early seedling mortality in the Iberian Peninsula (Castro, Zamora, Hódar, & Gómez, 2004; Vizcaíno-Palomar et al., 2014), and recent mortality could be genotype-dependent because the relict P. pinaster population might be currently undergoing rapid genetic adaptation to new drier conditions. Recent climate change is actually resulting in a higher warming rate across Spain than the global average, accompanied by reductions in relative humidity and rainfall (Vicente-Serrano, Rodríguez-Camino, Domínguez-Castro, El Kenawy, & Azorín-Molina, 2017). The potential selective pressures exerted by ongoing climate change would be greatly enhanced by the human-mediated severe range contraction suffered by the studied population in historic times (Charco et al., 2017), which has enforced a strong habitat shift, from the nonmarginal historic one to which the current population could still be genetically adapted, to the only recently occupied extremely marginal microenvironment. An alternative biological explanation consistent
with our strong selection estimates is that viability selection (measured by survival) tends to be stronger when measured over shorter periods, which could result from selection over longer time intervals tempering bursts of strong selection, via reversal or stasis periods (Hoekstra et al., 2001).

Beyond increasing sample sizes for enhanced statistical power, further insights could be gained by replicating our study on different cohorts, newly established over subsequent seed dispersal seasons. It would then be possible to test whether the same set of loci tend to exhibit nonrandom temporal allele frequency changes for cohorts established under similar seasonal environmental conditions. This would help supporting the true association of detected loci with selective mortality under a particular environment (Chanock et al., 2007). If in addition reverse locus-specific allele frequency changes were consistently detected in cohorts established under contrasting environmental conditions (e.g., dry versus wet), it would suggest that fluctuating selection is acting at early developmental stages, a process potentially contributing to the high within-population genetic variation of trees (Yeaman & Jarvis, 2006). It would be particularly interesting to track cohorts established under the typically rare combination of favorable environmental conditions allowing a successful transition to the sapling stage (Jackson et al., 2009), as it may reveal the otherwise hidden selective patterns that are mostly shaping the adaptive genetic variation of individuals reaching reproduction. Finally, by extending the tracking period of successfully recruited sapling cohorts, it would be possible to study which life stages are under the most intense natural selection, whether pleiotropy is present across life stages, and whether the strength of selection changes with the temporal scale of measurement (Donohue, 2014; Hoekstra et al., 2001). The methods presented here should provide a useful tool for addressing these fundamental questions in many species with large offspring cohorts and may open the way for further statistical advances for genomic analyses of selection on cohort mortality studies.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Economy, Industry and Competitiveness and the European Regional Development Fund (projects CGL2009-09428 and CGL2015-64164-R to JJRA and PhD grant BES-2010-031797 to GMU). We specially thank Santiago C. González-Martínez for providing early access to SNP marker information and protocols, and for genotyping assistance. We also thank Myriam Heuertz and Fernando del Caño for laboratory and field assistance, and three anonymous reviewers for helpful comments. Most computations in this study were conducted at CESGA high performance computing facilities, with funds from INIA’s Information Technology Unit.

CONFLICT OF INTEREST

None declared.

DATA ARCHIVING STATEMENT

Data for this study are available at the Dryad Digital Repository: https://doi.org/10.5061/dryad.3p5n8bj.

ORCID

Juan J. Robledo-Arnuncio http://orcid.org/0000-0002-3909-8928

REFERENCES

Alvarez-Buylla, E. R., Chaos, Á., Piñero, D., & Garay, A. A. (1996). Demographic genetics of a pioneer tropical tree species: Patch dynamics, seed dispersal, and seed banks. Evolution, 50, 1155–1166. https://doi.org/10.1111/j.1558-5646.1996.tb02356.x

Anderson, J. T., Lee, C. R., & Mitchell-Olds, T. (2014). Strong selection genome-wide enhances fitness trade-offs across environments and episodes of selection. Evolution, 68, 16–31. https://doi.org/10.1111/evo.12259

Ávila, C., Pérez-Rodríguez, J., & Cánovas, F. M. (2006). Molecular characterization of a receptor-like protein kinase gene from pine (Pinus sylvestris L.). Planta, 224, 12–19. https://doi.org/10.1007/s00425-005-0184-x

Bank, C., Ewing, G. B., Ferrer-Admettla, A., Foll, M., & Jensen, J. D. (2014). Thinking too positive? Revisiting current methods of population genetic selection inference. Trends in Genetics, 30, 540–546. https://doi.org/10.1016/j.tig.2014.09.010

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B: Statistical Methodology, 57, 289–300.

Bollback, J. P., York, T. L., & Nielsen, R. (2008). Estimation of 2Nė from temporal allele frequency data. Genetics, 179, 497–502. https://doi.org/10.1534 genetics.107.085019

Bourret, V., Dionne, M., & Bernatchez, L. (2014). Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: Polygenetic multilocus analysis surpasses genome scan. Molecular Ecology, 23, 4444–4457. https://doi.org/10.1111/mec.12798

Budde, K. B., Heuertz, M., Hernández-Serrano, A. P., Pausas, J. G., Vendramin, G. G., Verdú, M., & González-Martínez, S. C. (2014). In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (Pinus pinaster). New Phytologist, 201, 230–241. https://doi.org/10.1111/nph.12483

Bundgaard, J., & Christiansen, F. B. (1972). Dynamics of polymorphisms: I. Selection components in an experimental population of Drosophila melanogaster. Genetics, 71, 439–460.

Castro, J., Gómez, J. M., García, D., Zamora, R., & Hódar, J. A. (1999). Seed predation and dispersal in relict Scots pine forests in southern Spain. Plant Ecology, 145, 115–123. https://doi.org/10.1023/A:1009865709906

Castro, J., Zamora, R., Hódar, J. A., & Gómez, J. M. (2004). Seedling establishment of a boreal tree species (Pinus sylvestris) at its southernmost distribution limit: Consequences of being in a marginal Mediterranean habitat. Journal of Ecology, 92, 266–277. https://doi.org/10.1111/j.0022-0477.2004.00870.x

Chancerel, E., Lepoittevin, C., Le Provost, G., Lin, Y. C., Jaramillo-Correa, J. P., Eckert, A. J., ... Plomion, C. (2011). Development and implementation of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. BMC Genomics, 12, 368. https://doi.org/10.1186/1471-2164-12-368

Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., ... Collins, F. S. (2007). Replicating genotype-phenotype associations. Nature, 447, 655–660. https://doi.org/10.1038/447655a
Charjo, J., Venturas, M., Gil, L., & Nanos, N. (2017). Effective seed dispersal and fecundity variation in a small and marginal population of Pinus pinaster Ait. Growing in a harsh environment: Implications for conservation of forest genetic resources. Forests, 8, 312. https://doi.org/10.3390/f8090312

Christiansen, F. B., & Frydenberg, O. (1973). Selection component analysis of natural populations using population samples including mother-offspring combinations. Theoretical Population Biology, 4, 425–445. https://doi.org/10.1016/0040-5809(73)90019-1

Christiansen, F. B., Frydenberg, O., & Simonsen, V. (1978). Genetics of Zaoreus populations: X. Selection component analysis of the EstII polymorphism using samples of successive cohorts. Hereditas, 87, 129–150.

Cressie, N., & Read, T. R. C. (1989). Pearson’s $X^2$ and the loglikelihood ratio statistic $G^2$: A comparative review. International Statistical Review, 57, 19–43. https://doi.org/10.2307/1403582

De Wit, P., Rogers-Bennett, L., Kudela, R. M., & Palumbi, S. R. (2014). Forensic genomics as a novel tool for identifying the causes of mass mortality events. Nature Communications, 5, 3652. https://doi.org/10.1038/ncomms4652

Donohue, K. (2005). Seeds and seasons: Interpreting germination timing in the field. Seed Science Research, 15, 175–187. https://doi.org/10.10117/s5RR2005208

Donohue, K. (2014). Why ontogeny matters during adaptation: developmental niche construction and pleiotropy across the life cycle in Arabidopsis thaliana. Evolution, 68, 32–47. https://doi.org/10.1111/evol.12284

Donohue, K., Rubio de Casas, R., Burghardt, L., Kovach, K., & Willis, C. G. (2010). Germination, postgermination adaptation, and species ecological ranges. Annual Review of Ecology, Evolution, and Systematics, 41, 293–319. https://doi.org/10.1146/annurev-ecolsys-102209-144715

Dubos, C., Le Provost, G., Pot, D., Salin, F., Lalane, C., Madur, D., ... Plomion, C. (2003). Identification and characterization of water-stress-responsive genes in hydroponically grown maritime pine (Pinus pinaster) seedlings. Tree Physiology, 23, 169–179. https://doi.org/10.1093/treephys/23.3.169

Edwards, J. P., & Heath, D. J. (1983). Dynamics of an enzyme polymorphism in the isopod, Sphaeroma rugicuca (Leach). I. Temporal variation in genotype frequencies. Heredity, 51, 467–475. https://doi.org/10.1038/hdy.1983.57

Ewens, W. J. (2004). Mathematical population genetics. I. Theoretical introduction. New York: Springer Verlag. https://doi.org/10.1007/978-0-387-21222-9

Feder, A. F., Kryazhimskiy, S., & Plotkin, J. B. (2014). Identifying signatures of selection in genetic time series. Genetics, 196, 509–522. https://doi.org/10.1534/genetics.113.158220

Ferrer-Admetlla, A., Leuenberger, C., Jensen, J. D., & Wegmann, D. (2016). An approximate Markov model for the Wright-Fisher diffusion and its application to time series data. Genetics, 203, 831–846. https://doi.org/10.1534/genetics.115.184598

Foll, M., Shim, H., & Jensen, J. D. (2015). WFABC: A Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Molecular Ecology Resources, 15, 87–98. https://doi.org/10.1111/1755-0999.12280

Gillespie, J. H. (2004). Population genetics: A concise guide. Baltimore: The Johns Hopkins University Press.

Gompert, Z. (2016). Bayesian inference of selection in a heterogeneous environment from genetic time-series data. Molecular Ecology, 25, 121–134. https://doi.org/10.1111/mec.13323

Gompert, Z., Comeault, A. A., Farkas, T. E., Feder, J. L., Parchman, T. L., Buerkle, C. A., & Nosil, P. (2014). Experimental evidence for ecological selection on genome variation in the wild. Ecology Letters, 17, 369–379. https://doi.org/10.1111/ele.12238

Göring, H. H. H., Terwilliger, J. D., & Blanqero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. The American Journal of Human Genetics, 69, 1357–1369. https://doi.org/10.1086/324471

Haldane, J. B. S. (1931). A mathematical theory of natural and artificial selection. Part VII. Selection intensity as a function of mortality rate. Mathematical Proceedings of the Cambridge Philosophical Society, 27, 131–136. https://doi.org/10.1017/S0305004100009427

Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hoang, A., Hill, C. E., ... Kingsolver, J. G. (2001). Strength and tempo of directional selection in the wild. Proceedings of the National Academy of Sciences of the United States of America, 98, 9157–9160. https://doi.org/10.1073/pnas.161281098

Hou, F. Y., Huang, J., Yu, S. L., & Zhang, H. S. (2007). The 6-phosphogluconate dehydrogenase genes are responsive to abiotic stresses in rice. Journal of Integrative Plant Biology, 49, 655–663. https://doi.org/10.1111/j.1744-7909.2007.00460.x

Hufford, K. M., & Hamrick, J. L. (2003). Viability selection at three early life stages of the tropical tree, Platypodium elegans (Fabaceae, Papilionoideae). Evolution, 57, 518–526. https://doi.org/10.1111/j.0014-3820.2003.tb01543.x

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. Epidemiology, 19, 640–648. https://doi.org/10.1097/EDE.0b013e31818131e7

Jackson, S. T., Betancourt, J. L., Booth, R. K., & Gray, S. T. (2009). Ecology and the ratchet of events: Climate variability, niche dimensions, and species distributions. Proceedings of the National Academy of Sciences of the United States of America, 106, 19685–19692. https://doi.org/10.1073/pnas.0901464106

Jaramillo-Correa, J.-P., Rodríguez-Quiñó, I., Grivet, D., Lepoittevin, C., Sebastiani, F., Heuertz, M., ... González-Martínez, S. C. (2015). Molecular proxies for climate maladaptation in a long-lived tree (Pinus pinaster Aiton, Pinaceae). Genetics, 199, 793–807. https://doi.org/10.1534/genetics.114.173252

Jensen, J. D., Foll, M., & Bernatchez, L. (2016). The past, present and future of genomic scans for selection. Molecular Ecology, 25, 1–4. https://doi.org/10.1111/mec.13493

Juez, L., González-Martínez, S. C., Nanos, N., de-Lucas, A. I., Ordóñez, C., del Peso, C., Bravo, F. (2014). Can seed production and restricted dispersal limit recruitment in Pinus pinaster Aiton from the Spanish Northern Plateau? Forest Ecology and Management, 313, 329–339. https://doi.org/10.1016/j.foreco.2013.10.033

Kang, K.-S., Bila, A. D., Harju, A. M., & Lindgren, D. (2003). Estimation of fertility variation in forest tree populations. Forestry, 76, 329–344. https://doi.org/10.1093/forestry/76.3.329

Laporte, M., Pavey, S. A., Rougeux, C., Pierron, F., Lauzon, M., Budzinski, H., ... Bernatchez, L. (2016). RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. Molecular Ecology, 25, 219–237. https://doi.org/10.1111/mec.13466

Lotterhos, K. E., & Schaal, S. M. (2014). Genome scans for the contemporaneous response to selection in quantitative traits. Molecular Ecology, 23, 4435–4437. https://doi.org/10.1111/mec.12853

Malaspina, A. S. (2016). Methods to characterize selective sweeps using time serial samples: An ancient DNA perspective. Molecular Ecology, 25, 24–41. https://doi.org/10.1111/mec.13492

Malaspina, A. S., Malaspina, O., Evans, S. N., & Slatkin, M. (2012). Estimating allele age and selection coefficient from time-series data. Genetics, 192, 599–607. https://doi.org/10.1534/genetics.112.140939

Mathieson, I., & McVean, G. (2013). Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics, 193, 973–984. https://doi.org/10.1534/genetics.112.147611
Meyers, B. C., Morgante, M., & Michelmore, R. W. (2002). TIR-X and TIR-NBS proteins: Two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. The Plant Journal, 32, 77–92. https://doi.org/10.1046/j.1365-313X.2002.01404.x

Monnahan, P. J., Colicchio, J., & Kelly, J. K. (2015). A genomic selection component analysis characterizes migration-selection balance. Evolution, 69, 1713–1727. https://doi.org/10.1111/evo.12698

Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. Genetics, 98, 625–640.

Nishino, J. (2013). Detecting selection using time-series data of allele frequencies with multiple independent reference loci. G3: Genes Genomes, Genetics, 3, 2151–2161. https://doi.org/10.1534/g3.113.008276

Pascual, M. B., Cánovas, F. M., & Ávila, C. (2015). The NAC transcription factor family in maritime pine (Pinus pinaster): Molecular regulation of two genes involved in stress responses. BMC Plant Biology, 15, 254. https://doi.org/10.1186/s12870-015-0640-0

Pespeni, M. H., Sanford, E., Gaylord, B., Hill, T. M., Hosfelt, J. D., Jaris, H. K., … Palumbi, S. R. (2013). Evolutionary change during experimental ocean acidification. Proceedings of the National Academy of Sciences of the United States of America, 110, 6937–6942. https://doi.org/10.1073/pnas.1220673110

Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution, and Systematics, 37, 187–214. https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. Genetics, 104, 531–548.

Prout, T. (1971). The relation between fitness components and population prediction in Drosophila. I: The estimation of fitness components. Genetics, 68, 127–149.

Ramírez-Valiente, J. A., & Robledo-Arnuncio, J. J. (2014). Adaptive consequences of human-mediated introgression for indigenous tree species: The case of a relict Pinus pinaster population. Tree Physiology, 34, 1376–1387. https://doi.org/10.1093/treephys/tpu097

Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America, 94, 9197–9201. https://doi.org/10.1073/pnas.94.17.9197

Savolainen, O., Bomka, F., García-Gil, R., Komulainen, P., & Repo, T. (2004). Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of Pinus sylvestris to climatic changes. Forest Ecology and Management, 197, 79–89. https://doi.org/10.1016/j.foreco.2004.05.006

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. Annual Review of Ecology, Evolution, and Systematics, 38, 595–619. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646

Schräber, J. G., Evans, S. N., & Slatkin, M. (2016). Bayesian inference of natural selection from allele frequency time series. Genetics, 203, 493–511. https://doi.org/10.1534 genetics.116.187278

Shim, H., Laurent, S., Matuszewski, S., Foll, M., & Jensen, J. D. (2016). Detecting and quantifying changing selection intensities from time-sampled polymorphism data. G3: Genes Genomes, Genetics, 6, 893–904. https://doi.org/10.1534/g3.115.023200

Teshima, K. M., Coop, G., & Przeworski, M. (2006). How reliable are empirical genomic scans for selective sweeps? Genome Research, 16, 702–712. https://doi.org/10.1101/gr.5105206

Unger, G. M., Heuertz, M., Vendramin, G. G., & Robledo-Arnuncio, J. J. (2016). Assessing early fitness consequences of exotic gene flow in the wild: A field study with Iberian pine relicts. Evolutionary Applications, 9, 367–380. https://doi.org/10.1111/evo.12333

Vicente-Serrano, S. M., Rodríguez-Camino, E., Domínguez-Castro, F., El Kenawy, A., & Azorín-Molina, C. (2017). An updated review on recent trends in observational surface atmospheric variables and their extremes over Spain. Geographical Research Letters, 43, 209–232.

Vizcaíno-Palomar, N., Revuelta-Eugercios, B., Zavala, M. A., Alía, R., & González-Martínez, S. C. (2014). The role of population origin and microenvironment in seedling emergence and early survival in Mediterranean maritime pine (Pinus pinaster Aiton). PLoS ONE, 9, e109132. https://doi.org/10.1371/journal.pone.0109132

Waples, R. S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics, 121, 379–391.

Williams, S. M., & Haines, J. L. (2011). Correcting away the hidden heritability. Annals of Human Genetics, 75, 348–350. https://doi.org/10.1111/j.1469-1809.2011.00640.x

Yeaman, S., & Jarvis, A. (2006). Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. Proceedings of the Royal Society B, 273, 1587–1593.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

_How to cite this article:_ Robledo-Arnuncio JJ, Unger GM. Measuring viability selection from prospective cohort mortality studies: A case study in maritime pine. _Evol Appl._ 2019;12:863–877. [https://doi.org/10.1111/evo.12729](https://doi.org/10.1111/evo.12729)