Biological Conservation 261 (2021) 109261

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Proportional sampling strategy often captures more genetic diversity when population sizes vary

Kaylee Rosenberger, Emily Schumacher, Alissa Brown, Sean Hoban*

Morton Arboretum, 4100 IL-53, Lisle, IL 60532, United States of America

ARTI CLE I NFO

Keywords:
Ex situ conservation
Botanic garden
Seed bank
Conservation planning
Simulation
Molecular ecology
Demography

ABSTRACT

Collecting and conserving genetic diversity from plant populations for ex situ collections is a major objective of seed banks and botanic gardens. However, current guidelines for collecting germplasm to preserve plant species ex situ might not adequately capture the genetic diversity present in real plant population systems. Here, we tested a previously unexplored challenge for sampling practices: when populations of a rare plant species vary in size. We hypothesized that sampling proportional to the population size would capture more genetic diversity than sampling an equal number of individuals from every population. Using simulations of a hypothetical rare species, we tested equal and proportional strategies and calculated how many alleles were captured by either strategy. The effects of migration rate, recent bottlenecks, and sampling intensity on genetic diversity capture were also examined. We found that when population sizes differ (e.g., one population 3 times the size of median population size), proportional sampling captures more genetic diversity under constant size populations. The relatively modest improvement (1–5% more allelic diversity for most cases) was observed across all parameters of migration and intensity tested–except when there were recent bottlenecks. We also created simulations tailored to three IUCN Red List threatened oaks (*Quercus oglethorpensis, Q. engelmanii*, and *Q. acerifolia*) and found similar results for these detailed case studies as for our generic simulations. We conclude that sampling proportional to population size may often be a useful sampling strategy to create genetically diverse ex situ plant populations, ultimately resulting in more efficient use of resources.

1. Introduction

Determining an appropriate sampling strategy has been a topic of great interest and importance in population genetics, evolutionary biology, and conservation science (Nei et al., 1975; Hamilton, 1994; Braasch et al., 2021). An inappropriate sampling strategy can lead to false signals of adaptation (Selmoni et al., 2020), incorrect inference of admixture (Toyama et al., 2020), and misidentification of distinct stocks and gene flow (Ostergren et al., 2020), among other problems that affect conservation decisions (Meirmans, 2015). In recent years, increasing attention has been given to sampling strategies for establishing or supplementing ex situ conservation populations (Kashimshetty et al., 2017; Bragg et al., 2019; Hoban et al., 2020), complementing many years of work on managing ex situ populations, e.g., captive breeding (Ralls and Ballou, 1986; Lacy, 1987). The pace of anthropogenic change and corresponding losses of habitat and populations (Ceballos et al., 2017), combined with highly limited conservation resources, means that optimizing sampling strategies for ex situ conservation is an urgent and valuable research agenda (Cavender et al., 2015). This knowledge is of particular value to the approximately 3000 botanic gardens and arboreta worldwide who play a key role in conserving and sustainably using plant biodiversity, including genetic diversity (Oldfield, 2009; Mounce et al., 2017).

Botanic gardens hold germplasm (e.g., seed and tissue banks) and living collections (e.g., individuals and populations of whole living plants) that are irreplaceable resources for scientific research, education, and conservation. In addition to safeguarding plants ex situ and thus forestalling extinction, these collections contribute to action such as producing material for restoration, and educating and inspiring the public (Fant et al., 2016; Smith, 2018). Indeed, because botanic gardens conserve 30% of all plant species diversity, and 41% of threatened plant taxa (Mounce et al., 2017), and hold unique horticultural and taxonomic knowledge, botanic gardens may be the foremost mechanism for preventing plant extinctions (Westwood et al., 2021). Botanic gardens also seek to help meet international commitments and treaties such as the Global Strategy for Plant Conservation (GSPC), Sustainable

* Corresponding author.
E-mail addresses: eschumacher@mortonarb.org (E. Schumacher), alissabrown@mortonarb.org (A. Brown), shoban@mortonarb.org (S. Hoban).

https://doi.org/10.1016/j.biocon.2021.109261
Received 3 February 2021; Received in revised form 29 June 2021; Accepted 13 July 2021
Available online 28 July 2021

Development Goals, and Convention on Biological Diversity, which all have targets for ex situ conservation. However, Cavender et al. (2015) stated that the U.S. was less than halfway toward achieving GSPC Target 8 (75% of plant species conserved ex situ) in 2015, while Hoban et al. (2020) showed that some of even the most well-protected rare plants do not achieve GSPC Target 9 (70% of species' genetic diversity is conserved ex situ, https://www.cbd.int/gspc/targets.shtml). The quality and diversity of collections varies immensely among species (Beckman et al., 2019). This further emphasizes the importance of effective, efficient sampling to fully conserve plant species and their genetic diversity.

The ultimate goal of ex situ conservation is often reintroduction or supplementation of wild populations. Long-term persistence and adaptation of wild and restored populations, especially during rapid climate and environmental change, will require high genetic diversity. High genetic diversity allows recovery of populations after extreme events, and persistence in the face of new pests and disease (Reusch et al., 2005; Reynolds et al., 2012; Morikawa and Palumbi, 2019). Genetic diversity is correlated to fitness, and alleviates inbreeding depression (Spielman et al., 2004; Griffith et al., 2014). Recent work also shows that genetic diversity is important to the services provided to society, e.g., “nature's contribution to people (NCP)” (Stange et al., 2020) and to both the stability of ecosystems and community structure (Raffard et al., 2017). Genetic diversity is even needed to maintain botanic garden populations over multiple generations; genetic collapse can occur in garden populations as well as wild ones (Fant et al., 2016; Westwood et al., 2021). Aside from conservation purposes, genetic diversity in botanic gardens also helps showcase to the public the remarkable trait diversity within species and provides breeders and researchers with diverse material to work with. Collections must therefore represent a sufficient amount of genetic diversity (Guerrant et al., 2004), typically understood as capturing 95% of the alleles present in wild populations (Brown and Marshall, 1995), although other thresholds have been proposed (Lawrence, 2002; reviewed in Lockwood et al., 2007).

Ex situ collections are often established from seed or cuttings from the wild. Therefore, genetic diversity ex situ is almost always a reduced subset of that in situ; the amount conserved is strongly determined by sampling size and strategy (Guerrant et al., 2004). The first generalized ‘rule of thumb’ for sampling a minimum number of plants to capture genetic diversity recommended sampling about 50 individuals per population, based on early analytical models (Marshall and Brown, 1975) showing that 30 to 60 random samples should capture 95% of alleles greater than 0.05 frequency (note they aimed to capture only common alleles). This is similar to recommendations in population genetics for sampling a majority of alleles and for estimating genetic differentiation (Sjogren and Wyoni, 1994; Hale et al., 2012). Numerous guidance documents in forestry, agriculture and conservation have been based on this minimum number for several decades (see Table 1 in Hoban and Strand, 2015). However, empirical studies of collections have started to reveal that many current living collections underrepresent the genetic diversity of wild populations, with recent research showing that few species' collections conserve 95% of diversity even while holding more than 50 samples, likely due to overrepresenting maternal lines in sampling (Hoban et al., 2020). This suggests that sample strategies still need improvement.

Recent research is building on and improving the aforementioned generic minimum sampling guidance. Several studies have shown it is possible to create genetically representative collections by tailoring sampling to species biology and geographic distribution. Hoban and Strand (2015) found that a more genetically diverse sample can be obtained when sample strategies account for species biological traits (such as dispersal distance and life span), history, and other factors (2015). Hoban (2019) and Hoban and Schlarbaum (2014), using simulations, determined that population structure and migration rates between populations also strongly impact the minimum sampling needed. Interestingly, parallel work in landscape genomics has shown that sampling should be tailored to species' dispersal, with more samples needed in widely dispersing species (Selmoni et al., 2020). Meanwhile, Griffith et al. (2017), using an empirical study in cycads, showed that sampling across multiple years may be needed for infrequently flowering species. These studies have provided important guidance, especially by demonstrating that minimum sampling for most species will exceed 50 samples (often 100 or more are needed, especially when samples do not come from individual genets), and yet there remain notable knowledge gaps in sampling design due to the small number of empirical studies and relatively simple simulations used.

Here, we addressed a major knowledge gap for seed collectors to create and maintain genetically rich and representative collections of plants. We focused on allocating sampling effort–how much to sample under different conditions–when in situ population sizes differ. Prior simulation work on this topic has always assumed equal population sizes and equal sampling in populations across a species distribution (Hoban and Schlarbaum, 2014; Hoban, 2019), a strong simplification of reality. In fact, simulations in population and conservation genetics often assume equal population sizes (e.g., number of individuals on each side of a barrier in Landguth et al., 2010; bottleneck population in Chikhi et al., 2010; carrying capacity in Lotterhos and Whitlock, 2014). However, many species have populations that vary greatly in size, such as a few dozen individuals in one population to a few hundred or thousand in another population. Examples of such plants in the United States include *Erigeron maguirei, Penstemon penlandii, Astragalus osterhoutii, Quercus boyntonii,* and *Quercus oglethorpensis* (U.S. Fish and Wildlife Service, 1992; U.S. Fish and Wildlife Service, 1995; Kenny et al., 2016; Beckman, 2017a, b). Since conservation organizations are strongly limited in funds for field collection, proper allocation of effort to collect maximum genetically diverse collections may help optimize use of limited funds.

This study focused on determining the strategy to implement when population sizes vary significantly for a rare species. We tested whether a *proportional* sampling strategy (sample more in big populations, less in small ones) captures more genetic diversity than sampling *an equal amount* from every population. We predicted that proportional sampling can be more efficient because large populations tend to hold more genetic diversity (Kimura and Crow, 1964; Lira-Noriega and Manthley, 2013). A proportional strategy may reach these additional alleles because it samples more individuals from larger populations and fewer individuals from smaller populations. While this proposal is intuitive, this may not be the case; alternatively, a proportional strategy may perform worse, because small populations may hold unique alleles (private to one population) and insufficient attention to these populations may miss them (Lesica and Allendorf, 1995). The relationship between population size, allelic diversity, and the appropriate sampling strategy may also depend on historical population size changes, or bottlenecks (Hoban et al., 2013; Hoban et al., 2014). Many rare species have had recent population reductions which could also affect diversity levels and therefore capture of diversity (Miller et al., 2019; Walker et al., 2020).

Testing these hypotheses requires simulations, a useful tool in ecology, evolutionary biology, conservation, epidemiology and other fields (Balkenhol and Landguth, 2011; Hoban et al., 2012; Hoban, 2014). Simulations are useful for quantifying and uncovering ecological processes that would be difficult or impossible to do empirically (Peck, 2004; Hoban, 2014). Because simulations replicate a rules-based representation of the world and can be run via many repetitions to see different results, they can be useful in predicting outcomes or responses in a quantifiable manner (Peck, 2004; Lotterhos et al., 2018). We first used a set of simulations representing a ‘model’ species, in a fully crossed experimental design where we modify parameters representing biological traits applicable to real taxa (Hoban and Schlarbaum, 2014). We focused primarily on the degree of population size variance, but also modified migration rate, sampling intensity, and historical population size. We then simulated more realistic scenarios for three case study species–*Quercus acerifolia, Quercus oglethorpensis, and Quercus*

engelmannii–in which parameters match closely the geographic distribution and abundance of the species. They also differ in overall abundance- one of these species has a much larger total census size while the other two are rare. We chose these species because they are all IUCN Red List Threatened, they have observational data on variation in sizes of populations, and *Quercus* (oaks) all have recalcitrant seed which means they must be maintained ex situ as living individuals, not seeds. The use of case studies to complement a robust simulation study is an important aspect of testing simulation results (Heller et al., 2013; Hoban and Strand, 2015; Ostergren et al., 2020).

In summary, we tested two sampling strategies: sampling proportional to the population size and sampling equally from every population. We hypothesized that:

1) When a species' population sizes vary strongly, a proportional strategy captures more genetic diversity than an equal strategy.
2) The relative advantage of a proportional strategy will decrease as population sizes become similar.
3) The effect will be stronger when migration rates are lower and sampling intensity lower.

2. Methods

2.1. Overview

Simulations were used to assess the genetic diversity captured by different sampling strategies on populations of varying sizes. The procedure has four steps which apply both to our generic simulations and our case studies: 1) We created parameter files to represent a hypothetical rare species for our first set of simulations, a similar approach as in prior work (Hoban and Schlarbaum, 2014; Hoban and Strand, 2015; Hoban, 2019). 2) Parameter files were then used by the simulation software to create outputs of realistic genotypes from the parameters given, representing the entire population system. 3) We sampled from the simulated output, representing sampling seed in the wild, using two strategies: equally across all populations, and proportional to the population size. 4) We analyzed the samples and the entire simulated population to calculate the proportion of the total alleles captured by each strategy. The proportion of alleles captured describes the genetic conservation success of the strategy.

2.2. Generic simulations with fully crossed design and sampling from simulation outputs

We ran backwards-in-time simulations in the software Simcoal 2 (Version 2.1.2, Laval and Excoffier, 2004) to simulate genetic datasets based on 5 unequal populations. Simcoal 2 uses the coalescent theory to model genealogies–the history of a sample of genes from the current population (Excoffier, 2003). It has been used in simulating genetic diversity in hundreds of population and conservation genetic studies including many studies of rare species (e.g., Rodriguez et al., 2011; Lozier, 2013). We defined a 'scenario' as simulations in which population size differs. Each scenario consisted of 5 populations summing to 1500 individuals; however, each scenario had varying population sizes–from extremely different population sizes, to moderately different population sizes, to equal population sizes. Scenarios 1–8 had varied population sizes with the general pattern of one small population, three medium-sized populations, and one large population. Scenario 9 had equal population sizes. Table 1 shows the population sizes of each scenario, and Fig. 1 shows a visual representation of some of the scenarios used.

Parameters represent values of the species' traits; we focused primarily on varying population sizes but varied some other parameters. In the first set of simulations, the following parameters were held constant across all scenarios and categories: 0 growth rate and a historical event 10,000 generations in the past where all populations merged to allow coalescence. Then, two types of bottleneck simulations were implemented: in bottleneck type 1, deme size was reduced by 10 × in each population (each population was formerly ten times its current size). In bottleneck type 2, all historical populations were initially the same size (3000 individuals) and then constricted to yield the current unequal population sizes due to differing degrees of constriction (i.e., present day scenario 1 with 30, 100, 100, 100, 1170, and scenario 9 with 300, 300, 300, 300, 300). In all bottleneck simulations, the total historical species size was 15,000 individuals. Bottleneck events occurred 5 generations in the past and were instantaneous. A visual representation of each type of bottleneck is shown in Fig. 2. We also defined the type of genetic marker used in the parameter files–20 independently assorting (unlinked) microsatellite loci with mutation rates of 0.0005 applied to all markers, to allow comparability to previous work and because microsatellite markers are still the most common marker used in rare plants. Additionally, for all scenarios 1–9 we simulated both low migration (0.001) and high migration (0.01) which can represent species with a narrow range that is animal dispersed vs. wind dispersed (Nybom, 2004), with equal migration among all populations. Genetic differentiation, measured by Nei's Fst, was used to assess whether the scenarios resulted in realistic levels of genetic structure. In cases with low migration rates and high Fst, populations should be more genetically differentiated and more locally rare alleles may be present (Slatkin, 1985). Nei's Fst was calculated using the function pairwise.neifst() in the R package hierfstat (version 0.5–7), which estimates pairwise Fst by dividing the amount of gene diversity among populations by the total gene diversity (Goudet, 1995).

Table 1
Population sizes for each scenario.

Scenario	Population 1	Populations 2, 3, and 4	Population 5
1	30	100	1170
2	40	150	1010
3	50	200	850
4	100	200	800
5	150	200	750
6	200	250	550
7	200	300	400
8	290	300	310
9	300	300	300

For each scenario, migration rate, and bottleneck type, we ran 100 replicates of the simulation to incorporate stochasticity of the outcome of the simulation. The output files of the simulation contain the genotypes of individuals of the population in Arlequin format. Arlequin files were converted to Genepop files, and Genepop files were converted to genind objects using the R package *Adegenet* (Jombart et al., 2020).

To 'simulate' sampling from the populations, we created R scripts to randomly select a number of individuals (according to the proportion–high or low %) from each population (custom written R scripts are available with comments and a readme file at (https://github.com/kayleejorose/Morton-REU). For example, based on Scenario 1 in Table 2, the code randomly selects 3, 10, 10,10, and 117 individuals from their respective populations for the proportional strategy. Note that this simulates selection of an individual to be placed in the ex situ collection, representing a grafted cutting or a single outcrossing seed. Caveats to this simulated sampling are presented in Section 4.4 of the Discussion.

We tested both strategies, equal and proportional, on every simulation replicate across each of the 9 scenarios, for both high and low migration, and for both bottleneck types. Additionally, we tested two sampling intensities, high and low, on each replicate of each scenario. Two sampling intensities were tested to represent the effort invested for a certain amount of genetic diversity captured. For high intensity sampling, we sampled 10% of every population, or 30 individuals per population when sampling equally. For low intensity sampling, we sampled 5% of every population, or 15 individuals per population when

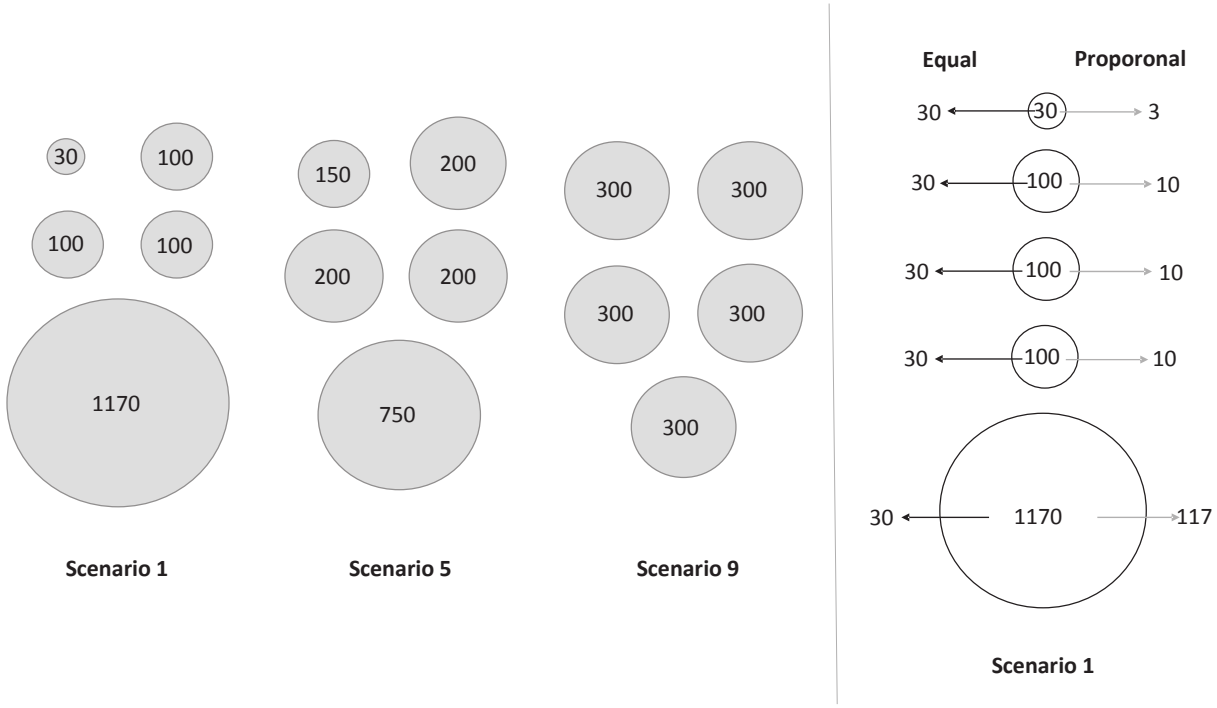

Fig. 1. a) Representation of three of the nine scenarios. Scenarios range from extremely varied population sizes to equal population sizes. b) An example of the sampling strategies used, showing high intensity sampling. For the proportional strategy, we sampled a percentage of the population size (10% for high intensity, 5% for low). For the equal strategy, an equal number of individuals were sampled from each population (30 for high intensity, 15 for low). These numbers were chosen such that between the two strategies, the same total sample size would be obtained (150 individuals for high intensity, 75 for low).

sampling equally. If the proportion resulted in a decimal value (only in a few cases), we rounded up to the next whole individual. The total sample size for each strategy is equal (150 individuals for high intensity, 75 individuals for low intensity). Fig. 1b shows an example of high intensity sampling on the populations of scenario 1. In summary, we simulated nine scenarios of populations with constant size, each with high and low migration, and each sampled at low and high intensity, with each tested via proportional and equal sampling, resulting in 18 parameter files sampled four ways. We followed the same procedure for simulating nine scenarios for each of two types of historical bottlenecks, resulting in another 36 parameter files sampled four ways.

2.3. Statistical analysis

Using data from the generic, non-bottleneck simulations, we constructed a generalized linear model (GLM) to evaluate: which sampling strategy captures more genetic diversity; identify for which scenarios the strategies are significantly different; and to determine whether migration rate and sampling intensity impacted those results. The response variable is the number of alleles captured (i.e., number of successes) and the number of alleles not captured (number of failures), which is modeled using the binomial distribution with a logit link function. Scenario (1–9), migration rate (low, high), sampling intensity (low, high), and sampling strategy (equal, proportional) were included as categorical predictors and main effects in the GLM. Additionally, we included interactions among the model predictors to determine how these variables affect one another. For example, we included an interaction between scenario and sampling strategy to determine how variation among population sizes influences whether proportional sampling performs better than equal sampling (one of our research hypotheses). We included interactions for: scenario and strategy; scenario and migration rate; and a four-way interaction between scenario, migration rate, sampling strategy, and sampling intensity (shown in Supplemental Eq. (S1)). We performed pairwise contrasts on the GLM output to answer questions about interactive effects. The GLM was run using the glm function in base R, and pairwise contrasts were evaluated using the lsmeans package (Lenth, 2016; R Core Team, 2021).

Using the simulations with historical bottlenecks, we constructed a GLM like the one described for the non-bottleneck simulations, with one difference. We included bottleneck type (1 or 2) as a categorical variable, which acted as both a main effect and an interactive effect (a 4-way interaction between scenario, strategy, migration rate, and bottleneck type).

To visualize which sampling strategy captured more genetic diversity, we calculated and plotted the proportion of the total alleles captured in the sample by each strategy for any given scenario (across all simulation replicates). This proportion represents the genetic conservation success–a higher proportion represents more genetic diversity conserved in the sample. The assumption that the number of alleles should be the primary focus of ex situ collections will be visited in the Discussion. We also calculated the total number of alleles in each scenario and identified their frequency in their populations to determine how each sampling strategy captured 'rare' alleles compared to more common alleles.

2.4. Case studies: Q. acerifolia, Q. oglethorpensis, and Q. engelmannii equal and proportional sampling simulations and analyses

We chose three oak species for our case studies: *Quercus acerifolia, Quercus oglethorpensis,* and *Quercus engelmannii.* These three oak species are rare (to varying degrees) and have unequal population sizes. Detailed information about each species' population sizes was obtained from the IUCN Red List (https://www.iucnredlist.org/, accessed 21 September 2020). *Quercus acerifolia* has about 600 known individuals total across its distribution. There are 4 populations with sizes 288, 147, 52, and 29 individuals (Wenzell et al., 2016). These values were rounded

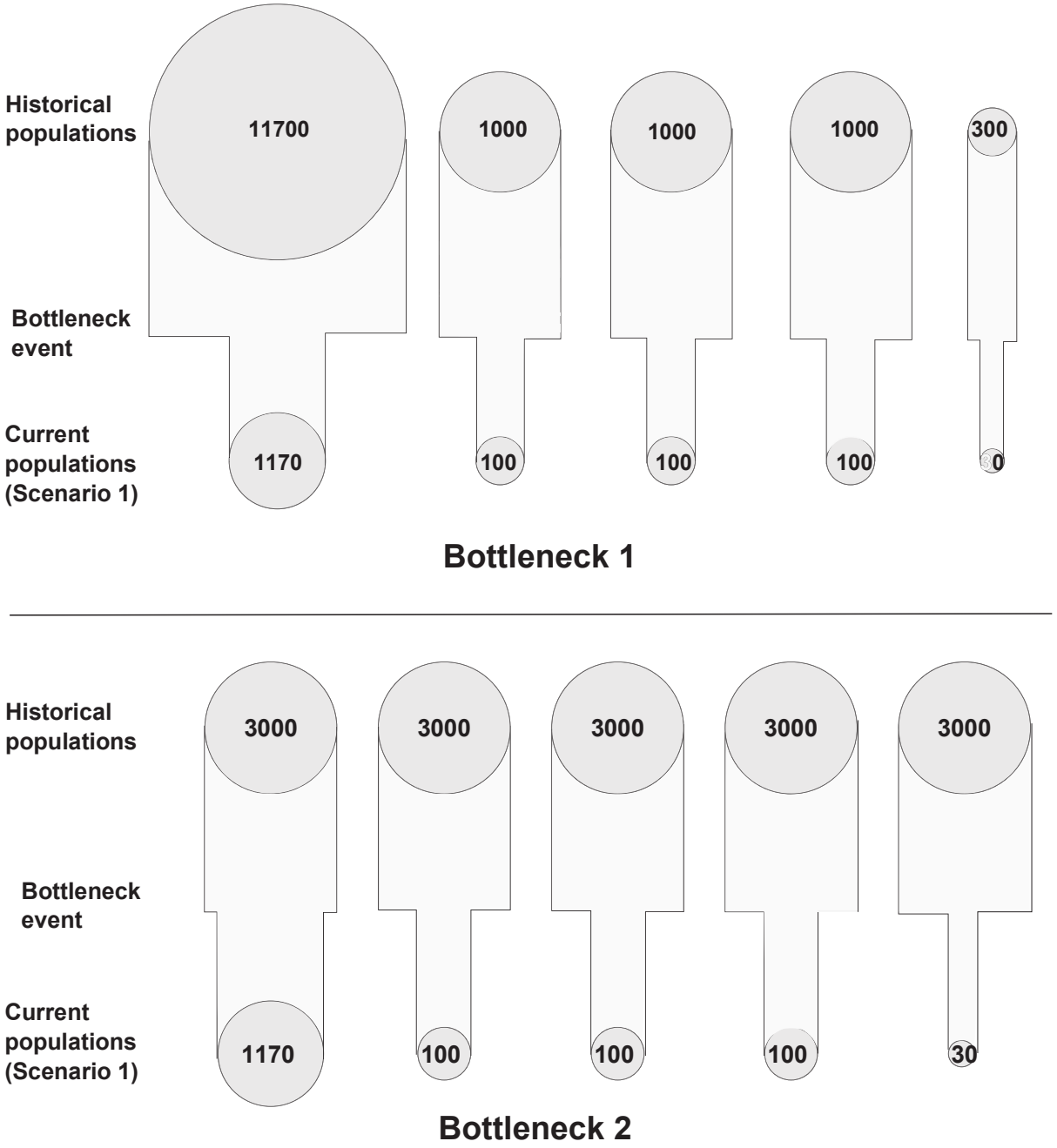

Fig. 2. Visual representation of the two types of bottlenecks we used for simulation. Scenario 1 (which has unequal population sizes) is shown as an example. a) In bottleneck type 1, all historical populations experienced bottlenecks which constricted population sizes 10 ×. This represents an event that impacts all populations equally (e.g., climate change). b) In bottleneck type 2, all historical populations were the same size (3000 individuals), but populations have experienced bottlenecks of varying degrees, leading to current unequal population sizes. This represents an event that impacts certain populations more than others (e.g., habitat destruction in one area).

Table 2
Population sizes for each species used for simulation, and percent sampled from each population (with the total number of individuals sampled in parentheses).

Species	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	High intensity sampling	Low intensity sampling
Q. acerifolia	300	150	50	30	–	12% (64 individuals)	6% (32 individuals)
Q. oglethorpensis	600	200	100	50	50	10% (100 individuals)	5% (50 individuals)
Q. engelmannii	18,500	1200	200	100	–	2% (400 individuals)	1% (200 individuals)

to multiples of 10 to simplify simulations and sampling. *Quercus oglethorpensis* has around 1000 total mature individuals (Beckman, 2017a, b). Due to the large number of 'sites' described on the Red List, and based on firsthand knowledge of the species, we aggregated these into 5 larger populations to simplify the simulation. *Quercus engelmannii* had the largest total size, estimated at around 200,000 individuals

(Beckman, 2017a, b). We simplified the simulation by using 10% of the total size for our simulations, or 20,000 individuals total. The proportion of individuals in each population was provided in the Red List. 93% of the total individuals reside in one population, with 6%, 0.5%, and 0.1% in the remaining populations (Beckman, 2017a, b). The population sizes we used for simulations are described in Table 2.

We ran simulations for each case study species to test whether sampling proportional to the population size captures more genetic diversity than sampling an equal number of individuals from each population, using a similar framework as described in detail above. However, different from the simulations above, we did not vary most of the parameter values used for simulation of our case study species. Parameter values were approximated to fit each species–no high or low migration scenarios were created since migration values were chosen realistically for each case study species. The only parameter values that were varied in these simulations were historical events. Historical bottleneck simulations were added, as described above in Section 2.2 of the methods. To simulate sampling, we used a similar R script to the above simulations. Because the population sizes for each species were very different, we sampled a different total number of individuals for each species (e.g., not 5% and 10% as above), described in Table 2. We couldn't exceed 2% for high intensity sampling for *Q. engelmannii* due to the large difference between the largest population and the smallest population, and because we aimed for realism in these numbers based on realistic seed sampling for rare plants. Note that these sample numbers are consistent with the recommendations from Hoban (2019)–a larger number of samples (but lower percentage) for species/populations with larger census size. For each case study species, we ran 100 replicates of the simulation.

Similar to the methods in Section 2.2, we ran Nei's pairwise Fst on case study populations to determine if our parameter files represented known or realistic levels of genetic differentiation between populations of these species. We then calculated allelic richness within the resulting sample populations and then ran a GLM to compare the outcome of proportional and equal sampling on these species. The GLM used the same response variable as described for the generic simulations: number of alleles captured (successes) and number of alleles not captured (failures). As with the generic simulations, we modeled the response using a binomial distribution with the logit link function. For the categorical predictor variables, we used species, sampling strategy, and intensity as the main effects. We also included interactive effects: species and sampling strategy; species and intensity; and a three-way effect of species, strategy, and intensity. We performed pairwise contrasts on the model results to evaluate how different combinations of sampling intensity and strategy lead to differences in the number of alleles captured for the three oak species. We visualized these results using the proportion of alleles captured, representing the genetic conservation success.

For the case study simulations using historical bottlenecks, we constructed the same GLM as described for the non-bottleneck case study simulations, with the addition of bottleneck type (1 or 2) as a categorical variable. Bottleneck type serves as a main effect and a 4-way interactive effect between strategy, intensity, species, and bottleneck type.

3. Results

3.1. Overview

As a reminder, for the constant population size simulations, each combination of scenario and migration rate ($9 \times 2 = 18$ simulation models) was sampled in four different ways (low/high intensity with equal/proportional strategy). For historical bottleneck simulations, we simulated each combination of scenario, migration rate, and bottleneck type (i.e., 36 simulation models), then sampled using the four approaches. The realism of each scenario's migration rate was assessed for genetic differentiation, as measured by Nei's pairwise Fst. In simulations with constant population sizes, Nei's pairwise Fst ranged from 0.011 to 0.093 in high migration scenarios while in lower migration scenarios ranged from 0.075 up to 0.405 in the most genetically distinct populations (Table 3, Supplemental Table S1), all of which are within expected ranges of differentiation for plant species (Nybom, 2004). Nei's pairwise Fst in bottleneck simulations ranged from 0.043–0.12 in bottleneck type 1 and 0.041–0.279 in bottleneck type 2 (Supplemental Tables S2, S3). In low migration combinations, more alleles total were present (168), than in high migration combinations (161), whereas in both bottleneck types the total number of alleles were similar (Bottleneck 1, high migration: 199.9; bottleneck 1, low migration: 199.88; bottleneck 2, high migration: 199.84; bottleneck 2, low migration: 199.83).

3.2. Genetic diversity capture results for either equal or proportional sampling strategies

Whether sampling with low intensity (5%) or high intensity (10%), the proportional strategy captured more alleles across scenarios 1–5 (where there was a greater variance between population size). However, when sampling with low intensity, fewer alleles were captured overall. The high intensity sampling strategy captured >93% alleles on average, while the low intensity sampling strategy captured about 90%. In simulations with historical bottlenecks, a higher proportion of alleles was captured even when sampling with low intensity, as shown in Fig. 3, panels E and F. On average> 97% of the alleles were captured for both bottleneck simulations. Note that we have only included figures for low migration and low sampling intensity for both bottleneck simulations.

For most combinations (compare white and blue bars), the proportional sampling strategy performs significantly better when population sizes vary (in scenarios 1–5, Supplemental Table S17). For example, for scenario 1, low migration and low sample intensity (panel D), the proportional strategy captured >90% of the alleles, while the equal strategy captured ~86%. As the population sizes become more equal (i.e., in scenarios 6–9), the difference of proportion of alleles captured between strategies becomes non-significant. In addition, when migration rates are high, less diversity is captured for all but one scenario in comparison to low migration rates (compare panel A to panel B, and panel C to panel D; Supplement Table S17). However, in both bottleneck simulations, proportional or equal sampling did not have a significant effect on allelic capture (compare panels E and F) and allelic capture was high for all

Table 3
Mean Nei's Fst values calculated for 100 simulation replicates for scenarios 1, 5, and 9 with high migration rates (0.01 per generation) and low migration (0.001 per generation) rates as well as Fsts for case studies.

Scenario	Scenario 1		Scenario 5		Scenario 9		Case studies		
	High migration	Low migration	High migration	Low migration	High migration	Low migration	*Quercus acerifolia*	*Quercus engelmannii*	*Quercus olgethorpensis*
Mean Pairwise Fst	0.051	0.250	0.019	0.126	0.015	0.098	0.309	0.113	0.239
Maximum Pairwise Fst	0.093	0.405	0.028	0.172	0.018	0.116	0.473	0.214	0.393
Minimum Pairwise Fst	0.019	0.115	0.012	0.080	0.011	0.079	0.151	0.016	0.092

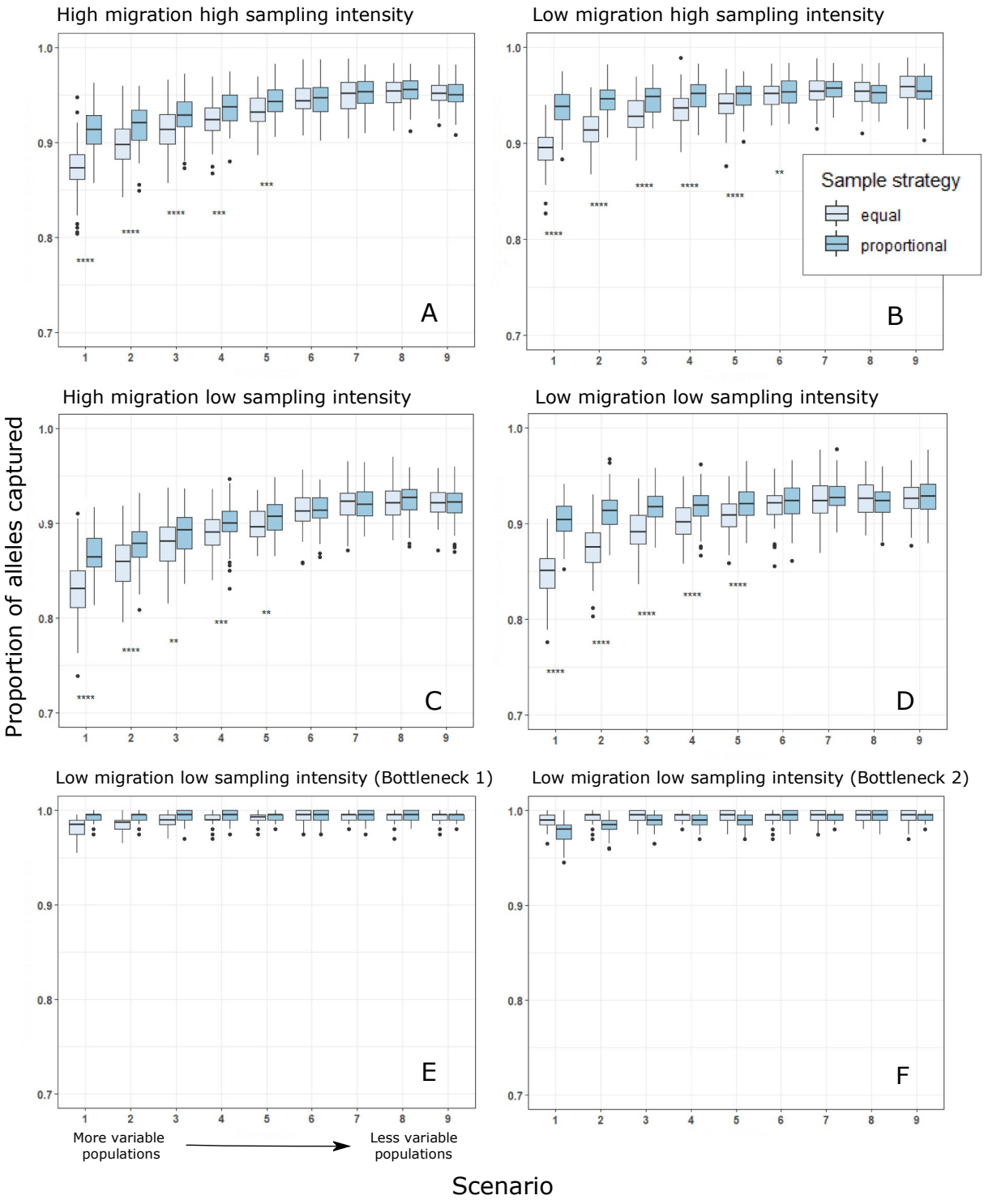

Fig. 3. Results for all combinations of migration rates and sampling intensity for our main simulations with constant population sizes (panels A-D). Each combination is represented by a different panel. Panels E and F show the results for low migration and low sampling intensity for each of the bottleneck simulations. Significant results are indicated by stars, with ** $= p \leq 0.01$, *** $= p \leq 0.001$ and **** $= p \leq 0.0001$.

scenarios. Proportional sampling appears to perform better for bottleneck type 1 in a few scenarios and equal sampling appears to perform better for bottleneck type 2, but these differences are not significant. Only sampling intensity exhibited a significant effect on allelic capture in bottleneck scenarios. We also observe an unexpected phenomenon, that, across all combinations, a lower proportion of alleles is captured in scenarios with unequal populations (scenario 1–5) when using either sampling strategy. In contrast, scenarios with relatively equal population sizes capture a greater proportion of alleles (scenario 6–9) with both strategies. This trend is especially apparent in combinations with high migration, visualized by the upward trend in proportion of alleles captured across scenarios from left to right (see panels A and C). In other words, both sampling strategies, equal and proportional, capture less diversity when population sizes are unequal compared to when populations are all of relatively equal sizes.

3.3. Case studies: Q. acerifolia, Q. oglethorpensis, and Q. engelmannii equal and proportional sampling strategy genetic diversity capture

All case study species exhibited similar results to the generic simulation. Case study Fst values ranged from low in *Quercus engelmannii* with a minimum pairwise Fst of 0.016 and maximum pairwise Fst of 0.214 to much higher in *Q. acerifolia* which ranged from 0.151 to 0.473 (Table 3). Overall, these values are realistic for many species depending on life history traits and fragmentation history of populations (Gregorius et al., 2007; Borkowski et al., 2017; Di Pietro et al., 2020). For all three species, with constant size populations, sampling proportional to the population size captured significantly more genetic diversity than sampling an equal number of individuals from each population. Additionally, when sampling species with high intensity, a greater proportion of alleles is captured compared to low intensity sampling. Results for low intensity sampling for all three species are shown in Fig. 4. For the bottleneck simulations, each species exhibited a different result. For *Q. acerifolia*, proportional sampling captured more alleles than equal sampling for bottleneck 1; however, equal sampling captured more alleles for bottleneck 2. Similarly for *Q. oglethorpensis*, proportional sampling performed better for bottleneck 1, whereas there was no difference between sampling strategies for bottleneck 2. For bottlenecks 1 and 2, both sampling strategies captured all alleles for *Q. engelmannii*, so there were no statistical differences. Fig. 4, panel A shows the results of case study species with constant population sizes, and Fig. 4, panels B and C show the results of the simulations with historical bottlenecks (for high intensity sampling–see Supplemental Fig. S1).

4. Discussion

4.1. Summary

Obtaining genetically diverse seed collections to form ex situ populations in arboreta and botanic gardens is an important step to safeguard plant species against extinction. It is increasingly known that the generalized sampling guideline typically applied to all species (50 samples per population) does not capture equal amounts of genetic diversity for every species, and likely would not capture a *sufficient* amount of genetic diversity for many species (Hoban and Schlarbaum, 2014; Hoban et al., 2020). A recent meta-analysis showed that current collections represent less genetic diversity than is found in wild populations (Wei and Jiang, 2020), while a recent empirical study showed that collections could be 40% more effective with an improved sampling design (Hoban et al., 2020). Therefore, it is important to develop and practice sampling strategies that maximize the amount of genetic diversity captured, especially considering scarce resources for conservation. In this paper, we provide further support to the general hypothesis that *a species-tailored sampling guideline can help capture more genetic diversity in* ex situ *collections than a generalized approach*. As opposed to other work focused on the minimal sampling size, we focus on allocation of effort among populations across species' geographic distributions. Prior work has assumed species have populations of equal sizes, but this may not be the case for many species (see Discussion in Hoban and Schlarbaum, 2014; Hoban and Strand, 2015). This is the first work to address how to sample genetic diversity from populations with unequal population sizes. We hypothesized that sampling proportional to the population size would capture more genetic diversity than sampling an equal number of individuals from every population.

4.2. Overall genetic diversity capture in generic simulations

Our main hypothesis was partially supported. When population sizes vary to a substantial degree (scenarios 1–5), a proportional strategy captures significantly more genetic diversity than sampling an equal number of individuals from every population, as long as populations have not undergone historical bottlenecks (especially type 2 bottlenecks). After scenario 5, the difference between population sizes is not drastic enough to cause a significant result; in such cases, sampling equally or proportionally has the same outcome. This is expected, because as population sizes become more homogenous, the sample sizes approach equality, even when sampling proportionally. Therefore, when population sizes are quite similar (in our results when the largest population was no more than 3 times the median population size) or have experienced certain historical bottlenecks, sampling an equal number of individuals from every population will suffice.

We also found that migration has a relatively small but noticeable impact on the outcome but does not change the pattern associated with sampling strategy. Specifically, a lower proportion of alleles (e.g., 1–3%) was captured in high migration scenarios. This may be because a high migration system ($m = 0.01$, Fst range from 0.007 to 0.131) acts more like a single, large population, with a single large effective population size (Ryman et al., 2019; Hoban et al., 2021), which can maintain more rare alleles. Examining the allele counts in different frequency categories, we do find more rare alleles in the high migration scenario (Supplemental Table S4 and S7). As these are harder to capture, we expect, overall, any sampling strategy to perform worse in the high migration strategy when population sizes differ. Interestingly, this is the opposite of previous results in constant size populations–in constant size populations more alleles are generally captured in situations of high migration (Hoban, 2019; Hoban and Schlarbaum, 2014). This further emphasizes the importance of considering the size of populations being sampled from.

These results held for sampling with low intensity, although low intensity sampling captured fewer alleles than high intensity sampling. The difference in the proportion of alleles captured between low and high sampling intensities was not very large–only a few percent. This is because of known 'diminishing returns' when sampling–most alleles are captured with relatively low effort, and it is only when seeking to capture approximately 95% or more of the alleles present that quite large sample sizes are needed (Brown and Marshall, 1995). Seed sampling efforts should consider logistical and practical aspects of sampling (time, resources, and the species' biological limitations e.g., fruiting season and frequency), when deciding if sampling with low intensity may be 'good enough.' Readers interested in more thorough investigations optimizing the minimum sample size can see Hale et al. (2012) and McGlaughlin et al. (2015). We note that although a few percent seems small, this has the outcome that several sampling strategies can exceed the recommended 95% of alleles while other strategies fall below this conservation threshold.

Finally, we observed an unexpected and intriguing result. In scenarios where the population sizes were unequal (Scenarios 1 to 5), *either* sampling strategy captured noticeably less genetic diversity compared to scenarios where population sizes are approximately equal (Scenarios 6 to 9), even though the total species abundance was constant at 1500 individuals (compare the height of box plots on the left vs. right of each plot–on the left all boxplots are lower and the percentage rises across scenarios, *even though sample size and the sum of all populations is constant*). This trend is especially apparent in scenarios with high migration (panels A and C) but is noticeable in all simulations and for low and high sample intensities. For example, in scenario 1 in Fig. 4, both strategies capture <87% alleles, while scenario 8 captures ~93% alleles with either strategy. Upon closer inspection of results, scenario 1 had fewer alleles total compared to scenario 9 (where all populations were equal sizes); however, scenario 1 had more *rare* alleles (Supplemental Tables S4-S9). This is likely due to scenarios with a large variance in size (e.g., Scenarios 1 to 5) also having one very large population (see Table 1). In large populations, rare alleles are less likely to be lost to drift (Wright, 1931, 1939; Young et al., 1996; Honnay and Jacquemyn, 2007). We confirm this by examining allele frequency histograms and find that Scenario 1, for example, has many more alleles (below a frequency of about 0.02) than Scenario 9 (Supplemental Figs. S2, S3). This is in spite of there being fewer total alleles in Scenario 1 (Supplemental Table S4),

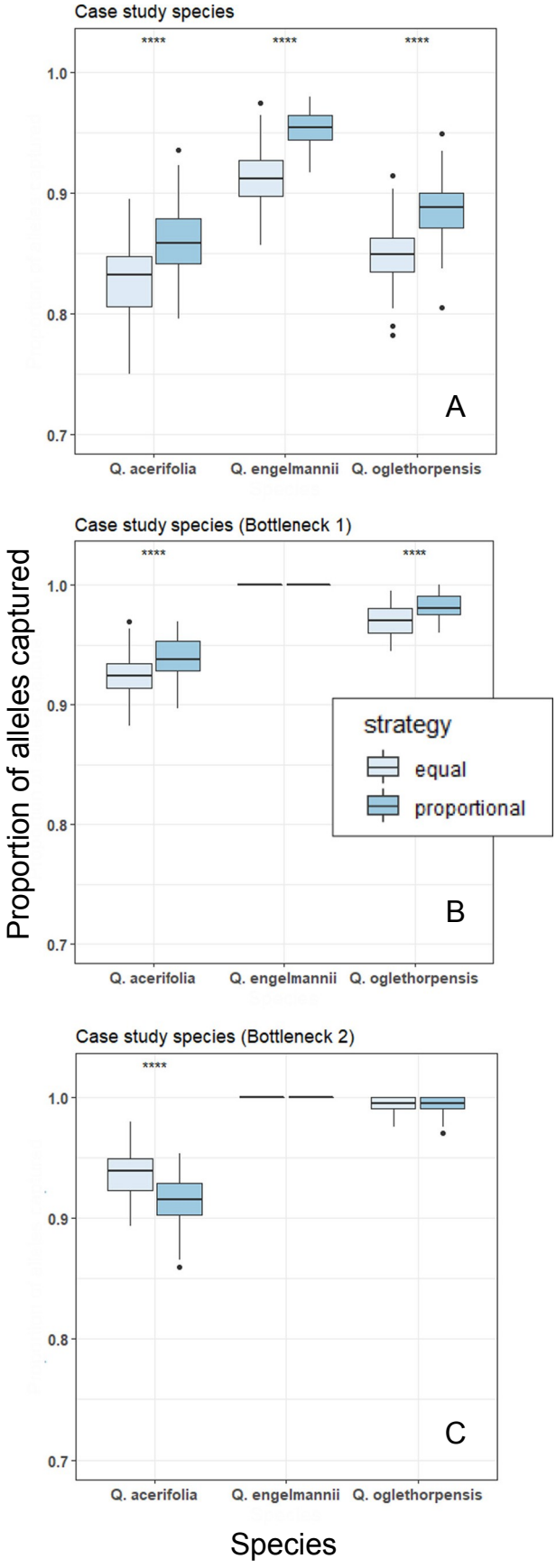

Fig. 4. Results for all case study species with low intensity sampling. Panel A shows the results for simulations without historical bottlenecks, and panels B and C show results for the sets of simulations with bottlenecks. Significant results are indicated with **** $= p \leq 0.0001$.

as expected based on lower global Ne in a metapopulation with varying population sizes (Nunney, 1999, Whitlock and Barton, 1997). In summary, in spite of harboring fewer total alleles in their lower global Ne, Scenarios 1 to 5 have more alleles in the rare, and especially very rare, frequency bins due to single large populations. Since rare alleles are more difficult to capture in sampling, this explains the lower overall allelic diversity captured in scenario 1. Still, proportional sampling helps–proportional sampling consistently captures alleles at lower frequencies (Supplemental Fig. S3).

The practical outcome of this observation of more rare alleles in these variable population systems is that sampling strategies in realistic situations of varying population sizes may capture less diversity than expected from previous simulation-based studies (e.g., Hoban, 2019) which simulated species with populations of equal sizes. This further emphasizes a need for exploring sampling strategies with simulations, which can uncover results we cannot predict a priori, and for increasing realism in simulations (Lotterhos et al., 2018). When considering sampling guidelines, collectors may need to increase sample intensity collecting from species with unequal population sizes in order to achieve their goals for collection.

Proportional sampling did not, however, capture significantly more diversity when bottleneck scenarios were implemented in populations' histories. In either bottleneck scenario, there was no significant difference in genetic diversity captured by equal or proportional sampling–– though proportional sampling appears to perform a little better for type 1 and equal sampling for type 2. Populations that have experienced historical reductions in population size due to bottlenecks have likely lost many rare alleles, which makes it easier to capture all alleles (Luikart et al., 1998). This was supported by our examination of allele frequency categories, which had very few rare alleles in bottleneck types (Supplemental Tables S12-S15). While numerous rare species have recently undergone bottlenecks from human activities (Miller et al., 2019; Walker et al., 2020), there is still little known about how most species' demography has changed over time (Convention on Biological Diversity, 2010; Pope et al., 2015). With either sampling strategy, >97% of diversity was captured in bottleneck scenarios, so equal or proportional sampling from each population may suffice. If a species is being sampled with little knowledge of its demographic history, as is the case with most rare species, we still recommend sampling proportional to population size.

It is interesting to compare our findings to the oft-cited '50 sample' recommendation for capturing 95% of alleles. Our main (non bottleneck) simulations had different total number of samples (150 total for high intensity, 75 for low intensity), as did our case studies (from 400 to 32). Examining Figs. 3 and 4, it is clear that our sampling usually did not capture 95% of the alleles (note that we consider all alleles, not just common ones)- especially for highly unequal population sizes or low migration situations. Only sample sizes approaching those of *Q. engelmanii* (200 to 400) were often sufficient to do so. For a multi-population system, it is clear that much more than 50 samples *total* are needed, probably several hundred total, though something less than 50 samples per population, as suggested by Hoban (2019). However, under bottleneck populations, 50 samples often seems sufficient, emphasizing that the impact of bottlenecks on minimum sample size needs more research.

4.3. Case study species patterns in comparison to generic simulations

The purpose of simulating case study species (*Q. acerifolia, Q. oglethorpensis,* and *Q. engelmannii)* was to determine if the main hypothesis is supported when applied to more realistic scenarios with parameter values based on real species. Each case study species has populations that vary in size and number of populations from the simple situation we simulated above (which was designed to test hypotheses in a fully crossed design). For all three species, sampling proportional to the population size captured significantly more genetic diversity than sampling equally among all populations in non-bottleneck scenarios, while under bottleneck type 1, proportional sampling performs better and under bottleneck type 2, equal sampling performs better. In other words, we reached the same conclusion as we did when simulating a 'hypothetical' species. Thus, the general results and our suggestions likely apply to other rare species with unequal population sizes across their distribution. The case studies also suggest that some real species are more similar to scenario 1 than to scenario 8 or 9. For most case study simulations, sampling with high intensity captured more genetic diversity than low intensity sampling. However, even when sampling with low intensity (as low as 1% in the case of *Q. engelmannii)* a greater amount of genetic diversity was captured when sampling proportionally for all species (upwards of 85%).

Tree species were chosen for our case studies because 10% of all tree species are threatened with extinction (Cavender et al., 2015). Furthermore, trees adapt slower to climate change than other species due to their longevity and relatively small population sizes (Enquist et al., 2019), so conservation efforts are increasingly more important (Cavender et al., 2015). Oaks in particular are keystone species, serving as the foundation for diverse ecosystems, and at least one quarter of oak species within the Americas are Threatened under the IUCN Red List (Beckman et al., 2019). Results from our case studies show that this advice is directly applicable and can inform increasing efforts to coordinate on sampling *Quercus* and other tree genera globally (https://www.bgci.org/our-work/projects-and-case-studies/a-global-conservation-consortium-for-oak/, accessed 18 January 2021). Future simulations could be created for all rare species in a genus, if sufficient demographic and geographic knowledge (number of populations, population size, and distance among populations) are available for designing the simulations.

There are very few comparable studies to ours that exist. One previous study, Hoban et al. (2018), investigated how to sample across the range of a widespread common species, *Fraxinus excelsior*. They tested allocation of samples to the 'core' and 'margin' of the range and found that it was most efficient to sample more populations in the core (larger populations) and less in the margin (small populations). This performed better than all other strategies including equal allocation of effort to core and margin. They emphasize that it was not very effective to ignore small populations entirely. While the parameters of this simulation matched a particular very common species (>1000 populations simulated with >1 million individuals), the general conclusion to sample more in large populations matches our results here for rare species. These results support the general supposition in the literature that small populations do retain conservation value and may hold unique alleles (Lesica and Allendorf, 1995).

4.4. Caveats

Simulations represent simplified versions of reality. These tools provide insight on systems that are difficult to quantify with empirical studies, by allowing otherwise infeasible numbers of replicates and fine scale combinations of parameters (e.g., 9 fully crossed scenarios here) but they are not perfect. The main assumption in simulation studies is that the model well-represents real processes (Peck, 2004; Hoban, 2014; Hoban and Strand, 2015). Here, we used fairly simple geographic ranges as well as case studies with realistic parameter values to verify our results from our first set of simulations. The parameter values used in our first set of simulations represent a hypothetical *rare* species with 1500 individuals spread over 5 populations. We assume that populations can be identified, and that the population size can be estimated, though sometimes this is difficult such as for continuously distributed species (Waples and Gaggiotti, 2006; Hoban et al., 2021). Simplifications of reality include that populations are not experiencing colonization or geographic range shifts over time and are undergoing constant migration–our simulations are not representative of common species spread over a large area. We also assume that demographic processes within populations are similar, including seed and pollen dispersal distances,

reproductive success, and sex ratios, though we acknowledge that such processes can vary among populations and would impact genetic diversity (Hamrick and Godt, 1996; Nybom, 2004; De Kort et al., 2021). Further research is needed to determine whether these aspects will impact the choice of proportional or constant seed sampling effort. Also, in simulating our case study species, it was necessary to simplify the parameters used for simulation by aggregating localities or sites into larger populations for *Q. oglethorpensis,* and by simulating 10% of the total census size of *Q. engelmannii.*

Additionally, our sampling itself was simplified for simulation. Our strategies assume collectors take only one 'seed' from each maternal plant (see Hoban et al., 2018 for details), or equivalent to taking a cutting from a plant for grafting or tissue propagation. In reality, this may not be feasible. Often, collectors take multiple samples/seeds from each maternal plant either because only a few plants are reproductive at any given time, or because resources don't allow for extensive and widespread sampling (Hoban and Strand, 2015). Simulations could be performed to examine the genetic diversity captured when sampling many seeds per plant. However, it is recommended to sample from as many unique plants as possible to obtain more genetically diverse seed samples (Hoban, 2019; Hoban et al., 2020). In addition, all populations were visited, and sampling was random within populations, which may not be possible in reality due to logistical constraints like time, funding, or land access. Of note for practical conservation efforts, Hoban and Strand (2015) found that random sampling outperforms spatially biased sampling, so in cases where random sampling is not feasible, sampling efforts should be increased.

We also note that we simulated markers under neutral evolution, e.g., no adaptive pressure, an important caveat. It may be argued that ex situ collections should focus on adaptive genetic diversity. However, lower levels of neutral genetic diversity are correlated to increased levels of inbreeding depression and loss of fitness (Forstmeier et al., 2012; Ruiz-López et al., 2012), reduction in adaptive potential that is needed during times of environmental change, and decreased ability to support ecosystem functions and for keystone species to survive extreme weather (Stange et al., 2020; Des Roches et al., 2021). In addition, a focus on 'gene targeted conservation,' or optimization of conservation for alleles of known or suspected adaptive advantage, has several downsides, including ignoring tradeoffs among traits, reducing diversity at other loci, increasing relatedness among populations, and the risk that the environment does not change as predicted or presents challenges that were not anticipated (Jump et al., 2009; Kardos and Shafer, 2018). Lastly, neutral genetic diversity and adaptive genetic diversity seem to lead to similar prioritization outcomes for conservation in some cases (Fernandez-Fournier et al., 2021), though not in others (Reeves et al., 2012). Preserving neutral genetic diversity likely remains a worthwhile approach for a robust conservation collection.

Finally, for our simulations, we used microsatellite-like genetic markers. There are some limitations associated with using microsatellite markers in real studies, including the risk of null alleles and homoplasy (Lemopoulos et al., 2019). Also, some populations may have limited microsatellite variation due to prior bottlenecks (Lemopoulos et al., 2019). Therefore, if the history of a species has resulted in extremely low genetic diversity, using other markers, like several thousand SNPs, might be more useful. SNPs can identify more genetic variation in low diversity populations due to the large number of SNPs present in the genome. We used microsatellites for computational efficiency and because they remain a common marker in conservation genetics and molecular ecology for managing ex situ populations (Puckett, 2017; Guzmán et al., 2020; Vashistha et al., 2020). Also, they represent loci with numerous alleles including rare alleles. Future work should replicate the simulations with SNP markers to test whether the same results are generated, though when looking at SNPs, haplotypes are a useful metric rather than SNPs because there will be many more alleles (Reeves and Richards, 2017). We also assume that the number of alleles is the primary measure of conserving genetic diversity in ex situ collections. Alternative goals exist, as early work on this topic noted (Brown and Marshall, 1995; Richards et al., 2007). Minimization of kinship is commonly used for managing collections, while inbreeding or heterozygosity can also be used. However, Brown and Marshall (1995) points out that alleles are ultimately the target of selection and should likely be conserved ex situ. It would be better to conserve traits, but there are few traits whose genetic architecture and conservation importance are generally understood for numerous plant species.

4.5. Utility in seed collections for ex situ conservation

Utilizing information about the species of conservation interest can lead to a more informed and tailored sampling strategy to efficiently capture the most genetic diversity of a species, though model realism is an important consideration. In this project, we found that for populations without bottlenecks allocating sampling efforts proportional to population sizes will capture more genetic diversity than sampling an equal number from every population when population sizes of a rare species vary greatly (when one population is three times the size of median population size); these results were reiterated in three case study species. Sampling proportionally was not significantly better when populations have undergone bottlenecks, emphasizing a need for further analyses of population history to understand when this strategy is appropriate. The results of this project directly and immediately contribute to sampling guidelines for other rare species when populations differ strongly in size. Our results build on and complement past studies that have established recommended minimum sampling numbers for a variety of simulated and empirical situations (Hoban and Schlarbaum, 2014; Hoban, 2019; Hoban et al., 2020; Wei and Jiang, 2021). Creating sampling guidelines tailored to species traits remains important to accomplish the goals of individual gardens as well as global conservation goals for conserving species– and we note that results should apply to animals (e.g., coral, see Afiq-Rosli et al., 2019; or tortoises, see Quinzin et al., 2019) as well as plant ex situ collections.

CRediT authorship contribution statement

Conceptualization- SH, KR, ES; Data Curation- all authors; Formal analysis- all authors, led by KR; Methodology- all authors, led by SH and KR; Project administration- SH and KR; Visualization- KR, AB, ES; Writing- all authors, led by SH and KR.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge funding from NSF REU Grant 1851961 and from a Center for Tree Science Fellowship to KR. SH was also partially supported in this work by IMLS grant MA-30-18-0273-18 and MG-245575-OMS-20. AB was supported by NSF ABI award 1759759.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2021.109261. All code to run and analyze simulations, along with a readme file, can be found at https://github.com/kayleejorose/Morton-REU

References

Afiq-Rosli, L., Huang, D., Toh, T.C., Taira, D., Ng, C.S.L., Song, T., Chou, L.M., 2019. Maximising genetic diversity during coral transplantation from a highly impacted source reef. Conserv. Genet. 20 (3), 629–637.

Balkenhol, N., Landguth, E.L., 2011. Simulation modelling in landscape genetics: on the need to go further. Mol. Ecol. 20 (4), 667–670.

Beckman, E. (2017a). Quercus engelmannii. The IUCN Red List of Threatened Species.

Beckman, E., (2017b). Quercus oglethorpensis. The IUCN Red List of Threatened Species.

Beckman, E., Meyer, A., Denvir, A., Gill, D., Man, G., Pivorunas, D., Shaw, K., Westwood, M., 2019. Conservation Gap Analysis of Native U.S. Oaks. The Morton Arboretum, Lisle, IL.

Borkowski, D.S., Hoban, S.M., Chatwin, W., Romero-Severson, J., 2017. Rangewide population differentiation and population substructure in Quercus rubra L. Tree Genetics and Genomes 13 (3), 67.

Braasch, J., Di Santo, L., Tarble, Z., Prasifka, J., & Hamilton, J. (2021). Testing for evolutionary change in restoration: a genomic comparison between ex situ, native and commercial seed sources of Helianthus maximiliani. bioRxiv.

Bragg, J.G., Cuneo, P., Sherieff, A., Rossetto, M., 2019. Optimizing the genetic composition of a translocation population: incorporating constraints and conflicting objectives. Mol. Ecol. Resour. 20 (1), 54–65.

Brown, A.H.D., Marshall, D.R., 1995. A Basic Sampling Strategy: Theory and Practice. Collecting Plant Genetic Diversity: Technical Guidelines. CAB International, Wallingford, pp. 75–91.

Cavender, N., Westwood, M., Bechtoldt, C., Donnelly, G., Oldfield, S., Gardner, M., Rae, D., McNamara, W., 2015. Strengthening the conservation value of ex situ tree collections. ORYX 49 (3), 416–424.

Ceballos, G., Ehrlich, P., Dirzo, R., 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. PNAS 114 (30), E6089–E6096.

Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B., Beaumont, M.A., 2010. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification. Genetics Society of America 186, 983–995.

Convention on Biological Diversity, 2010. Global biodiversity outlook (GBO-3). In: Convention on Biological Diversity.

De Kort, H., Prunier, J.G., Ducatez, S., Honnay, O., Baguette, M., Stevens, V.M., Blanchet, S., 2021. Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and animal populations. Nat. Commun. 12 (1), 1–11.

Des Roches, S., Pendleton, L.H., Shapiro, B., Palkovacs, E.P., 2021. Conserving intraspecific variation for nature's contributions to people. Nature Ecology & Evolution 5 (5), 574–582.

Di Pietro, R., Di Marzio, P., Antonecchia, G., Conte, A.L., Fortini, P., 2020. Preliminary characterization of the Quercus pubescens complex in southern Italy using molecular markers. Acta Botanica Croatica 79 (1), 0.

Enquist, B., Feng, X., Boyle, B., Maitner, B., Newman, E.A., Jorgensen, P.M., Roehrdanz, P.R., Thiers, P.M., Burger, J.R., Corlett, R.T., Couvreur, T.L.P., Dauby, G., Donoghue, J.C., Foden, W., Lovett, J.C., Marquet, P.A., Merow, C., Midgley, G., Morueta-Holme, N., Neves, D.M., Oliveira-Filho, A.T., Kraft, N.J.B., Park, D.S., Peet, R.K., Pillet, M., Serra-Diaz, J.M., Sandel, B., Schildhauer, M., Šímová, I., Violle, C., Wieringa, J.J., Wiser, S.K., Hannah, L., Svenning, J.C., McGill, B.J., 2019. The commonness of rarity: global and future distribution of rarity across land plants. Sci. Adv. 5, 1–13.

Excoffier, L., 2003. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. http://cmpg.unibe.ch/software/simcoal/.

Fant, J.B., Havens, K., Kramer, A.T., Walsh, S.K., Callicrate, T., Lacy, R.C., Maunder, M., Meyer, A.H., Smith, P.P. and, 2016. What to do when we can't bank on seeds: what botanic gardens can learn from the zoo community about conserving plants in living collections. Am. J. Bot. 103 (9), 1–3.

Forstmeier, W., Schielzeth, H., Mueller, J.C., Ellegren, H., Kempenaers, B., 2012. Heterozygosity–fitness correlations in zebra finches: microsatellite markers can be better than their reputation. Mol. Ecol. 21 (13), 3237–3249.

Goudet, J., 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. J. Hered. 86 (6), 485–486.

Gregorius, H.R., Degen, B., König, A., 2007. Problems in the analysis of genetic differentiation among populations–a case study in Quercus robur. Silvae Genetica 56 (1–6), 190–199.

Griffith, M.P., Calonje, M., Meerow, A.W., Francisco-Ortega, J., Knowles, L., Aguilar, R., Tut, F., Sanchez, V., Meyer, A., Noblick, L.R., Magellan, T.M., 2017. Will the same ex situ protocols give similar results for closely related species? Biodivers. Conserv. 26, 2951–2966.

Griffith, M.P., Calonje, M., Meerow, A.W., Tut, F., Kramer, A.T., Hird, A., Magellan, T.M., Husby, C.E., 2014. Can a botanic garden Cycad collection capture the genetic diversity in a wild population? Int. J. Plant Sci. 176 (1), 1–10.

Guerrant, E.O., Fiedler, P., Havens, K., Maunder, M., 2004. Revised genetic sampling guidelines for conservation collections of rare and endangered plants: supporting species survival in the wild. In: Ex Situ Plant Conservation: Supporting Species Survival in the Wild. Island Press, pp. 419–441.

Guzmán, F.A., Moore, S., de Vicente, M.C., Jahn, M.M., 2020. Microsatellites to enhance characterization, conservation and breeding value of Capsicum germplasm. Genet. Resour. Crop. Evol. 67 (3), 569–585.

Hale, M.L., Burg, T.M., Steeves, T.E., 2012. Sampling for Microsatellite-based Population Genetic Studies: 25 to 30 Individuals Per Population is Enough to Accurately Estimate Allele Frequencies.

Hamilton, M.B., 1994. Ex situ conservation of wild plant species: time to reassess the genetic assumptions and implications of seed banks. Conserv. Biol. 8 (1), 39–49.

Hamrick, J.L., Godt, M.W., 1996. Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. B Biol. Sci. 351 (1345), 1291–1298.

Heller, R., Chikhi, L., Siegismund, H.R., 2013. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. PLoS One 8 (5), E62992.

Hoban, S., Bertorelle, G., Gaggiotti, O.E., 2012. Computer simulations: tools for population and evolutionary genetics. Nat. Rev. Genet. 13 (2), 110–122.

Hoban, S., 2014. An overview of the utility of population simulation software in molecular ecology. Mol. Ecol. 23 (10), 2383–2401.

Hoban, S., 2019. New guidance for ex situ gene conservation: sampling realistic population systems and accounting for collection attrition. Biol. Conserv. 235, 199–208.

Hoban, S., Arntzen, J., Bruford, M., Godoy, J., Hoelzel, A., Segelbacher, G., Vila, C., Bertorelle, G., 2014. Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. Evol. Appl. 7 (9), 984–998.

Hoban, S., Kallow, S., Trivedi, C., 2018. Implementing a new approach to effective conservation of genetic diversity, with ash (Fraxinus excelsior) in the UK as a case study. Biol. Conserv. 225, 10–21.

Hoban, S., Callicrate, T., Clark, J., Deans, S., Dosmann, M., Fant, J., Gailing, O., Havens, K., Hipp, A., Kadav, P., Kramer, A., Lobdell, M., Magellan, T., Meerow, A., Meyer, A., Pooler, M., Sanchez, V., Spence, E., Thompson, P., Toppila, R., Walsh, S., Westwood, M., Wood, J., Griffith, M., 2020. Taxonomic similarity does not predict necessary sample size for ex situ conservation: a comparison among five genera. Proc. R. Soc. B Biol. Sci. 287 (1926).

Hoban, S., Gaggiotti, O., Bertorelle, G., 2013. The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation-based study. Mol. Ecol. 22 (13), 3444–3450.

Hoban, S., Paz-Vinas, I., Aitken, S., Bertola, L.D., Breed, M.F., Bruford, M.W., Funk, W.C., Grueber, C.E., Heuertz, M., Hohenlohe, P., Hunter, M.E., Jaffe, R., Fernandez, M.L., Mergeay, J., Moharrek, F., O'Brien, D., Segelbacher, G., Vernesi, C., Laikre, L., 2021. Effective population size remains a suitable, pragmatic indicator of genetic diversity for all species, including forest trees. Biol. Conserv. 253, 108906.

Hoban, S., Schlarbaum, S., 2014. Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. Biol. Conserv. 177, 90–99.

Hoban, S., Strand, A., 2015. Ex situ seed collections will benefit from considering spatial sampling design and species' reproductive biology. Biol. Conserv. 187, 182–191.

Honnay, O., Jacquemyn, H., 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. Conserv. Biol. 21 (3), 823–831.

Jombart, T., Kamvar, Z.N., Collins, C., Lu štrik, R., Beugin, M., Knaus, B.J., Solymos, P., Mikryukov, V., Schliep, K., Mai é, T., Morkovsky, L., Ahmed, I., Cori, A., Calboli, F., Ewing, R., Michaud, F., DeCamp, R., Courtiol, A., 2020. adegenet: exploratory analysis of genetic and genomic data [Data Collection]. In: The Comprehensive R Archive Network. https://cran.r-project.org/web/packages/adegenet/.

Jump, A.S., Marchant, R., Pe ñuelas, J., 2009. Environmental change and the option value of genetic diversity. Trends Plant Sci. 14 (1), 51–58.

Kardos, M., Shafer, A.B., 2018. The peril of gene-targeted conservation. Trends Ecol. Evol. 33 (11), 827–839.

Kashimshetty, Y., Pelikan, S., Rogstad, S.H., 2017. Effective seed harvesting strategies for the ex situ genetic diversity conservation of rare tropical tree populations. Biodivers. Conserv. 26, 1311–1331.

Kenny, L., Wenzell, K., Beckman, E., 2016. Quercus boyntonii. The IUCN Red List of Threatened Species.

Kimura, M., Crow, J.F., 1964. The number of alleles that can be maintained in a finite population. Genetics 49 (4), 725–738.

Lacy, R.C., 1987. Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. Conserv. Biol. 1 (2), 143–158.

Landguth, E.L., Cushman, S.A., Schwartz, M.K., McKelvey, K.S., Murphy, M., Luikarts, G., 2010. Quantifying the lag time to detect barriers in landscape genetics. Mol. Ecol. 19, 4179–4191.

Laval, G., Excoffier, L., 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics 20 (15), 2485–2487.

Lawrence, M.J., 2002. A comprehensive collection and regeneration strategy for *ex situ* conservation. Genetics Resources and Crop Evolution 49, 199–209.

Lemopoulos, A., Prokkola, J.M., Uusi-Heikkil ä, S., Vasemagi, A., Huusko, A., Hyvarinen, P., Koljonen, M.-L., Koskiniemi, J., Vainikka, A., 2019. Comparing RADseq and microsatellites for estimating genetic diversity and relatedness-implications for brown trout conservation. Ecology and Evolution 9, 2106–2120.

Lenth, R.V., 2016. Least-squares means: the R package lsmeans. J. Stat. Softw. 69 (1), 1–33. https://doi.org/10.18637/jss.v069.i01.

Lesica, P., Allendorf, F.W., 1995. When are peripheral populations valuable for conservation. Conserv. Biol. 9 (4), 753–760.

Lira-Noriega, A., Manthley, J.D., 2013. Relationship of genetic diversity and niche centrality: a survey and analysis. Evolution 68 (4), 1082–1093.

Lockwood, D.R., Richards, C.M., Volk, G.M., 2007. Probabilistic models for collecting genetic diversity: comparisons, caveats, and limitations. Crop Sci. 47 (2), 861–866.

Lotterhos, K., Moore, J., Stapleton, A., 2018. Analysis validation has been neglected in the Age of Reproducibility. PLoS Biol. 16 (12).

Lotterhos, K.E., Whitlock, M.C., 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol. Ecol. 23 (9), 2178–2192.

Lozier, J.D., 2013. Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in north American bumble bees using RAD sequencing. Mol. Ecol. 23 (4), 788–801.
Luikart, G., Allendorf, F.W., Cornuet, J.M., Sherwin, W.B., 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. 89 (3), 238–247.
Marshall, D.R., Brown, A.H.D., 1975. Optimum Sampling Strategies in Conservation. *International Biological Programme 2: Crop Genetic Resources for Today and Tomorrow*. Cambridge University Press.peery.
McGlaughlin, M.E., Riley, L., Brandsrud, M., Arcibal, E., Helenurm, M.K., Helenurm, K., 2015. How much is enough? Minimum sampling intensity required to capture extant genetic diversity in ex situ seed collections: examples from the endangered plant *Sibara filifolia* (Brassicaceae). Conserv. Genet. 16, 253–266.
Meirmans, P.G., 2015. Seven common mistakes in population genetics and how to avoid them. Mol. Ecol. 24 (13), 3223–3231.
Miller, B.P., Symons, D.R., Barrett, M.D., 2019. Persistence of rare species depends on rare events: demography, fire response and phenology of two plant species endemic to a semiarid Banded Iron Formation range. Aust. J. Bot. 67 (3), 268–280.
Morikawa, M.K., Palumbi, S.R., 2019. Using naturally occurring climate resilient corals to construct bleaching-resistant nurseries. PNAS 116 (21), 10586–10591.
Mounce, R., Smith, P., Brockington, S., 2017. Ex situ conservation of plant diversity in the world's botanic gardens. Nature Plants 3, 795–802.
Nei, M., Maruyama, T., Chakraborty, R., 1975. The bottleneck effect and genetic variability in populations. Evolution 29 (1), 1–10. https://doi.org/10.2307/2407137.
Nunney, L., 1999. The effective size of a hierarchically structured population. Evolution 53 (1), 1–10.
Nybom, H., 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol. Ecol. 13 (5), 1143–1155.
Oldfield, S.F., 2009. Botanic gardens and the conservation of tree species. Trends Plant Sci. 14 (11), 581–583.
Ostergren, J., Palm, S., Gilbey, J., Dannewitz, J., 2020. Close relatives in population samples: evaluation of the consequences for genetic stock identification. Mol. Ecol. Resour. 20 (2), 498–510.
Peck, S.L., 2004. Simulation as experiment: a philosophical reassessment for biological modeling. TRENDS in Ecology and Evolution 19 (10), 530–534.
Pope, L. C., Liggins, L., Keyse, J., Carvalho, S. B.,& Riginos, C. (2015). Not the time or the place: the missing spatio-temporal link in publicly available genetic data.
Puckett, E.E., 2017. Variability in total project and per sample genotyping costs under varying study designs including with microsatellites or SNPs to answer conservation genetic questions. Conserv. Genet. Resour. 9 (2), 289–304.
Quinzin, M.C., Sandovaul-Castillo, J., Miller, J.M., Beheregaray, L.B., Rusello, M.A., Hunter, E.A., Gibbs, J.P., Tapia, W., Villalva, F., Caccone, A., 2019. Genetically informed captive breeding of hybrids of an extinct species of Galapagos tortoise. Conserv. Biol. 1–11.
R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Raffard, A., Lecerf, A., Cote, J., Buoro, M., Lassus, R., Cucherousset, J., 2017. The functional syndrome: linking individual trait variability to ecosystem functioning. Proc. R. Soc. B 284 (1868).
Ralls, K., Ballou, J., 1986. Captive breeding programs for populations with a small number of founders. Trends in Ecology and Evolution 1 (1), 19–22.
Reeves, P.A., Richards, C.M., 2017. Capturing haplotypes in germplasm core collections using bioinformatics. Genet. Resour. Crop. Evol. 64 (8), 1821–1828.
Reeves, P.A., Panella, L.W., Richards, C.M., 2012. Retention of agronomically important variation in germplasm core collections: implications for allele mining. Theor. Appl. Genet. 124 (6), 1155–1171.
Reusch, T.B.H., Ehlers, A., Hammerli, A., Worm, B., 2005. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. PNAS 102 (8), 2826–2831.
Reynolds, L.K., McGlathery, K.J., Waycott, M., 2012. Genetic diversity enhances restoration success by augmenting ecosystem services. PLoS One 7 (6), e38397.
Richards, C.M., Antolin, M.F., Reilley, A., Poole, J., Walters, C., 2007. Capturing genetic diversity of wild populations for ex situ conservation: Texas wild rice (Zizania texana) as a model. Genet. Resour. Crop. Evol. 54 (4), 837–848.
Rodriguez, R., Ramirez, O., Valdiosera, C.E., Garcia, N., Alda, F., Madurell-Malapeira, J., Marmi, J., Doadrio, I., Willerslev, E., Gotherstrom, A., Arsuaga, J.L., Thomas, M.G., Lalueza-Fox, C., Dalen, L., 2011. 50,000 years of genetic uniformity in the critically endangered Iberian lynx. Mol. Ecol. 20 (18), 3785–3795.
Ruiz-López, M.J., Gañan, N., Godoy, J.A., Del Olmo, A., Garde, J., Espeso, G., Vargas, A., Martinez, F., Roldán, E.R., Gomendio, M., 2012. Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. Conserv. Biol. 26 (6), 1121–1129.
Ryman, N., Laikre, L., Hossjer, O., 2019. Do estimates of contemporary effective population sizes tell us what we want to know? Mol. Ecol. 28 (8), 1904–1918.
Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., Joost, S., 2020. Sampling strategy optimization to increase statistical power in landscape genomics: a simulation-based approach. Mol. Ecol. Resour. 20 (1), 154–169.
Sjogren, P., Wyoni, P., 1994. Conservation genetics and detection of rare alleles in finite populations. Conserv. Biol. 8 (1), 267–270.
Slatkin, M., 1985. Rare alleles as indicators of gene flow. Evolution 39 (1), 53–65.
Smith, P., 2018. The challenge for botanic garden science. Plants People Planet 1 (1), 38–43.
Spielman, D., Brook, B., Frankham, R., 2004. Most species are not driven to extinction before genetic factors impact them. PNAS 101 (42), 15261–15264.
Stange, M., Barrett, R.D.H., Hendry, A.P., 2020. The importance of genomic variation for biodiversity, ecosystems and people. Nat. Rev. Genet. 22 (2), 89–105.
Toyama, K.S., Crochet, P.-A., Leblois, P., 2020. Sampling schemes and drift can bias admixture proportions inferred by STRUCTURE. Mol. Ecol. Resour. 20 (6), 1769–1785.
U.S. Fish and Wildlife Service (1992). Osterhout Milkvetch (*Astragalus osterhoutii*) and Penland beardtongue (*Penstemon penlandii*) recovery plan. U.S. Fish and Wildlife Service. Denver, Colorado. 16 pp.
U.S. Fish and Wildlife Service (1995). Maguire daisy (*Erigeron maguirei*) recovery plan. U. S. Fish and Wildlife Service, Denver, Colorado. 13 pp.
Vashistha, G., Deepika, S., Dhakate, P.M., Khudsar, F.A., Kothamasi, D., 2020. The effectiveness of microsatellite DNA as a genetic tool in crocodilian conservation. In: Conservation Genetics Resources, pp. 1–12.
Walker, B.E., Leão, T.C., Bachman, S.P., Bolam, F.C., Nic Lughadha, E., 2020. Caution needed when predicting species threat status for conservation prioritization on a global scale. Front. Plant Sci. 11, 520.
Waples, R.S., Gaggiotti, O., 2006. INVITED REVIEW: what is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol. Ecol. 15 (6), 1419–1439.
Wei, X., Jiang, M., 2020. Meta-Analysis of Genetic Representativeness of Plant Populations under Ex Situ Conservation in Contrast to Wild Source Populations. Conservation Biology.
Wei, X., Jiang, M., 2021. Meta-analysis of genetic representativeness of plant populations under ex situ conservation in contrast to wild source populations. Conserv. Biol. 35 (1), 12–23.
Wenzell, K., Kenny, L., Beckman, E., 2016. Quercus acerifolia. *The IUCN Red List of Threatened* Species.
Westwood, M., Cavender, N., Meyer, A., Smith, P., 2021. Botanic garden solutions to the plant extinction crisis. Plants, People, Planet 3 (1), 22–32.
Whitlock, M.C., Barton, N.H., 1997. The effective size of a subdivided population. Genetics 146 (1), 427–441.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16 (2), 97.
Wright, S., 1939. The distribution of self-sterility alleles in populations. Genetics 24 (4), 538.
Young, A., Boyle, T., Brown, T., 1996. The population genetic consequences of habitat fragmentation for plants. Trends in Ecology and Evolution 11 (10), 413–418.