



# Optimal sampling of seeds from plant populations for *ex-situ* conservation of genetic biodiversity, considering realistic population structure



Sean Hoban<sup>a,\*</sup>, Scott Schlarbaum<sup>b</sup>

<sup>a</sup> National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, USA

<sup>b</sup> Department of Forestry, Wildlife and Fisheries, University of Tennessee, Knoxville, USA

## ARTICLE INFO

### Article history:

Received 17 April 2014

Received in revised form 9 June 2014

Accepted 15 June 2014

### Keywords:

Biodiversity  
*Ex situ* sampling  
 Genetic diversity  
*Juglans cinerea*  
 Living collection  
 Minimum sample size  
 Optimization  
 Sample strategy  
 Seed bank  
 Simulations

## ABSTRACT

As habitat loss accelerates, there is pressing need to preserve plant genetic diversity in *ex situ* conservation collections. Population structure (i.e. subdivision), which is common in plants, may be an important consideration when planning such collections because it results in locally restricted alleles or traits, which have high conservation, ecological, or economic value. Nonetheless, common protocols for *ex situ* collections do not consider population structure. To help inform collection decisions, we utilize computer simulations with different levels of realistic hierarchical population structure to evaluate the expected performance of an array of sample sizes and several spatial distributions of sampled populations. We quantify how population structure affects the expected probability of capturing alleles (especially rare alleles). We also test the effect of family-level structure due to pollen pool composition (e.g., sibling clusters). Our findings suggest that when range-wide population structure exists, the spatial distribution of sampled populations is crucial: sampling one population per region (dispersed sampling) captured up to 175% more alleles than sampling all populations in one region (constrained sampling), and nearly as much as sampling all existing populations. The spatial effect is strongest for poorly-connected (low gene flow) species. Under realistic population structure, moderate sampling (25–30 individuals per population) from few but widely-spaced populations performs optimally; this differs from previous recommended guidelines that do not consider structure. There is smaller effect of the pollen pool composition on collection performance. We conclude that seed collection plans should incorporate spatial considerations, especially for poorly-connected taxa. Our simulation approach can be extended to particular species and other spatial patterns. We use the butternut tree as a case study for collection planning.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Many rare and/or economically valuable plant species (including forest trees, crop-wild relatives, and endemics) are declining due to biotic and abiotic pressures, which may lead to local or species-wide extinction. *In situ* preservation is sometimes impractical or impossible due to changing climate, invasive species, habitat degradation, and/or lack of management resources. *Ex situ* collections are an important option for preserving biodiversity for later reintroduction when biotic or abiotic environmental pressures relax and suitable locations become available for reintroduction (Schoen and Brown, 2001). Such collections require high diversity to enable future reintroduced populations to adapt, and for future

use in agriculture or pharmaceuticals (Frankham, 2010; Neale and Kremer, 2011; Pautasso, 2011). Exhaustive sampling to capture all diversity, however, is unfeasible even for rare species (Richards et al., 2007). In practice, the amount of resources available for collections, or jurisdiction/political boundaries, may preclude capturing an ideal amount of diversity. Collectors must determine a sampling methodology, including how many samples to collect, and from which locations. This task must balance oversampling, which wastes limited resources, against undersampling, which can fail to sufficiently provision a species' future. Collectors therefore need tools to quantify the expected amount of genetic diversity captured per collection unit (essentially, expected return on investment). This information (i.e. identifying particular strategies, and recognizing points of diminishing returns along a continuum of sample size investments) helps to ensure a given proportion of diversity is obtained even when collecting resources are limited.

\* Corresponding author. Tel.: +1 865 974 9195.

E-mail address: [shoban@alumni.nd.edu](mailto:shoban@alumni.nd.edu) (S. Hoban).

Two main approaches have previously been used for planning optimal *ex situ* collections: the probability approach and the resampling approach. Probability equations have the advantage of simplicity and can quickly calculate the necessary sample size to capture alleles at particular frequencies, e.g., >0.05 (Brown and Marshall, 1995; Lawrence et al., 1995). These equations have general applicability across plant taxa, and therefore have influenced many sampling guidelines, including those used by the United States' Bureau of Land Management and Forest Service (Rogers and Montalvo, 2004; Engels et al., 2008; BLM, 2012). Several authors have used the probability approach to estimate the number of samples needed to capture >95% of relevant diversity. However, their recommendations differ based on differing assumptions. Assuming all populations contain exactly the same genetic composition, i.e. no population structure, Lawrence et al. (1995) recommended sampling ~170 individuals in total, spread over all populations. Alternately, Brown and Marshall (1995) recommended 50 samples per population and sampling every population by assuming all populations have completely non-overlapping genetic composition, i.e., extremely high population structure.

The resampling approach guides collections by stochastically resampling from a previously gathered genetic dataset (Caujapé-Castells and Pedrola-Monfort, 2004; Gapare et al., 2008; Namoff et al., 2010). This approach has the advantage of species-specific recommendations, based on empirical data. Genetic datasets are typically small, however, representing only a portion of the species' range (e.g., 100 individuals across five sites), and may be spatially or ecologically biased (Bamberg et al., 2010). As such, the dataset is unlikely to contain all of the species' alleles (especially rare ones), and therefore this approach may overestimate the performance of a proposed strategy. Still, if a large dataset is available, it may help guide collections in the focal species or closely related species (Brown and Hardner, 2000; Rozale et al., 2001; Neel and Cummings, 2003). For example, Gapare et al. (2008) used this approach in Sitka spruce [*Picea sitchensis* (Bong.) Carr.], and recommend gathering 150–180 samples from every population. Nonetheless, most collections are designed with little or no genetic data from the target species (Lockwood et al., 2007).

Consequently, the probability approach is simple but lacks realism, and is thus unable to provide quantitative guidelines about number of samples or spatial distribution of sampled populations under realistic genetic structure. The resampling approach allows realism including possibly incorporating spatial structure, but requires a large genetic dataset and may suffer from bias in that genetic dataset. Neither approach considers family-level structure due to pollination (although Sapra et al. (2003) consider selfing).

Our work aims to improve upon these approaches in three ways. (I) We explicitly account for realistic population genetic structure at regional scales, an important but neglected consideration in sampling strategies (Brown and Hardner, 2000; Brown and Marshall, 1995). Population structure, due to gene flow or selection, can result in some alleles that are restricted to one or few populations or regions. Such locally unique alleles (and traits) are high priority for collections (Bataillon et al., 1996; Schoen and Brown, 2001). Population structure also means that other alleles are redundant, i.e. present in more than one population, increasing their chance of being sampled (Brown and Marshall, 1995). To incorporate population structure, we present and utilize a simulation-based approach, inspired by the work of Bataillon et al. (1996) in designing crop germplasm core collections. Empirically collected genetic data is not necessary for this approach, only some knowledge of species' biology to parameterize the simulation models (though see Discussion). (II) In contrast to some previous work, we do not fix the amount of genetic diversity that the sampler wishes to capture, e.g., the goal is not always to capture 95% of existing alleles. Previous goals of very high allelic capture have

led to unfeasibly (for many species and agencies) large required number of samples, as discussed above. Considering the pragmatic goal of capturing substantial diversity with reasonable sample effort, we aim to calculate expected diversity conserved by an array of sample size investments and spatial arrangements. This “sliding scale” provides information on amount of diversity per investment of available resources, under realistic structure. (III) We consider composition of the donor pollen pool, which varies among species and locations. This may be an additional consideration because, if the pollen pool is small, seeds on a maternal plant will contain similar alleles (due to sharing paternal ancestry), which may affect the utility of sampling multiple seeds per maternal plant (Brown and Hardner, 2000; Sork et al., 2002).

We use our simulation approach to quantify sampling strategy performance in three hypothetical species with different population structure and population sizes. We assess three spatial strategies for each species: sampling all populations, sampling only one population per region (dispersed), and sampling all populations in only one region (constrained). We test a sliding scale of sampling sizes (number of samples from 5 to 50, per population). Furthermore we simulate seed formation under various pollen pool compositions (number of fathers), to assess sample sizes when collecting seeds from maternal parents. We quantify performance of strategies by calculating the proportion of alleles captured in various allelic categories (see Table 1). Our specific goals are: (1) quantify how performance of a given sample size varies for different, realistic population structures (e.g., different species) and spatial strategies; and (2) quantify how the size of the paternal pool affects the utility of sampling multiple seeds from each maternal parent. Based on results, we discuss sampling guidelines in general, and apply our methodology to a case study in an endangered North American tree, butternut (*Juglans cinerea*), whose recent *in situ* decline from an exotic fungal disease makes germplasm conservation especially urgent.

## 2. Methods

Our approach has three steps (summarized in Fig. 1):

- (1) Use stochastic genetic simulations to recreate, *in silico*, a system of populations separated by defined genetic distances. The number of populations, population size, number of loci, and genetic distances (in the form of migration rates) are set as parameters, so simulations can be tailored to approximate particular taxa.
- (2) Sample the *in silico* dataset to test various spatial sampling strategies (arrangement of chosen populations) for vegetative sampling (e.g., this step assumes one sample per individual, as in a cutting).
- (3) Create paternal pools with different numbers of fathers and then create simulated seeds, to test performance when sampling multiple seeds from each maternal plant, under different pollen pool compositions.

### Step 1- Creating populations

Following Bataillon et al. (1996), we define a model of a hierarchically subdivided plant species (e.g., several ecoregions), a model that may result in alleles that are spatially segregated due to low migration, high genetic drift, or natural selection. Populations within an ecoregion are connected by migration  $m1$ , and connected to other regions by migration  $m2$ . Migration between regions is lower than amongst populations within a region, so  $m1 > m2$ . We create four regions to represent the range of the hypothetical species, with four populations of equal size  $N$  within each region (total 16 populations). We also created simulations having three regions and three populations per region, to test the effect of range size.

Download English Version:

<https://daneshyari.com/en/article/6300203>

Download Persian Version:

<https://daneshyari.com/article/6300203>

[Daneshyari.com](https://daneshyari.com)